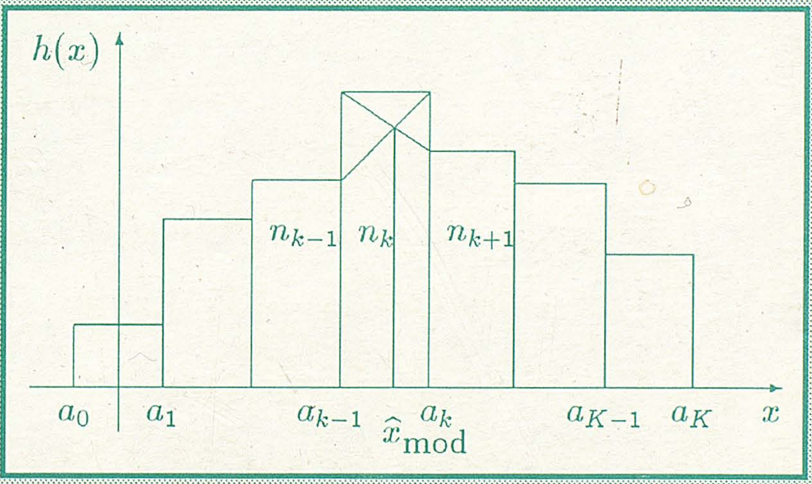
Ե. Հարությունյան, Տ. Ղազանչյան,
Ն. Մեսրոպյան, Դ. Ասատրյան, Մ. Հարությունյան,
Մ. Սահակյան, Հ. Շահումյան

ՀԱՎԱՆԱԿԱՆՈՒԹՅՈՒՆ և ԿԻՐԱՌԱԿԱՆ ՎԻՃԱԿԱԳՐՈՒԹՅՈՒՆ



[illegible]
26. X. 2000
[illegible]

Եվգենի Հարությունյան, Տաթևիկ Ղազանչյան,
Նաիրա Մեսրոպյան, Դավիթ Ասատրյան, Մարիամ Հարությունյան,
Մելս Սահակյան, Հարություն Շահումյան

ՀԱՎԱՆԱԿԱՆՈՒԹՅՈՒՆ և ԿԻՐԱՌԱԿԱՆ ՎԻՃԱԿԱԳՐՈՒԹՅՈՒՆ

ՏՆՏԵՍԱԳԵՏՆԵՐԻ ու ԳՈՐԾԱՐԱՐՆԵՐԻ ՀԱՄԱՐ

Թույլատրված է ՀՀ Կրթության և գիտության նախարարության կողմից,
որպես բարձրագույն ուսումնական հաստատությունների դասագիրք

ՀՀ ԳԱԱ «Գիտություն»
հրատարակչություն

Երևան 2000

ԴՏՀ 51(07)
ԳՄԴ 22.1g73
Հ 371

Դասագիրքը ստեղծված է Ե. Ա. Հարությունյանի ընդհանուր խմբագրությամբ:

Այս գրքի ստեղծմանն ու հրատարակմանը աջակցել է «Եվրասիա» հիմնադրամը՝ Ամերիկայի Միացյալ Նահանգների Միջազգային զարգացման գործակալության (USAID) տրամադրած միջոցների հաշվին:

Հ 371 Հավանականություն և կիրառական վիճակագրություն (տնտեսագետների ու գործարարների համար) /Ե. Հարությունյան, Տ. Ղազանչյան, Ն. Մեսրոպյան և ուրիշներ - Եր.: «Գիտություն», 2000. - 298 էջ:

Դասագիրքը նվիրված է հավանականության տեսության և մաթեմատիկական վիճակագրության հիմունքների շարադրմանը՝ առավելապես նկատի ունենալով տնտեսագիտական կիրառությունները: Պարունակում է հիմնական և հատուկ դասընթացների նյութեր բուհերի ուսանողների, ինչպես նաև մագիստրատուրայում ու ասպիրանտուրայում սովորողների համար: Կարող է ծառայել ինքնուրույն ուսումնասիրության համար, օժանդակել ժամանակակից հաշվային (կոմպյուտերային) ծրագրաշարերի արդյունավետ օգտագործմանը:

Օգտակար կլինի նաև այն մասնագետներին, որոնք նպատակ ունեն տիրապետելու մի շարք ավելի խորացված վիճակագրական եղանակների: Մաթեմատիկական ապացույցները տրվում են սահմանափակ ծավալով (ավելի մանր շարվածքով): Բազմաթիվ կիրառական օրինակների վերլուծությունն ուղեկցվում է թվային լուծումներով: Վարժություններն ամփոփված են զուգահեռաբար հրատարակվող խնդրագրքում:

$$Հ\ \frac{1602010000}{703(02)-2000}2000$$

ISBN 5-8080-0453-5 ԳՄԴ 22.1g73

Evgueni Haroutunian, Tatevik Kazanchyan,
Naira Mesropian, David Asatryan, Mariam Harutyunyan,
Mels Sahakyan, Harutyun Shahumyan

Probability and Applied Statistics
for Economics and Business

"Gitutiun" Publishing House of NAS of RA
Yerevan 2000

Евгений Арутюнян, Татевик Казанчян,
Наира Месропян, Давид Асатрян, Мариам Арутюнян,
Мелс Саакян, Арутюн Шаумян

Вероятность и прикладная статистика
для экономистов и предпринимателей

Издательство "Гитутюн" НАН РА
Ереван 2000

Խմբագիրներ՝
Ռուբեն Համբարձումյան, Բորիս Նահապետյան,
Եվգենի Հարությունյան, Մամիկոն Գինովյան,
Ալեքսան Սիմոնյան, Դավիթ Ասատրյան,
Աշոտ Հարությունյան

Editors
Ruben Ambartsumian, Boris Nahapetyan,
Evgueni Haroutunian, Mamikon Ginovian,
Alexan Simonian, David Asatryan,
Ashot Harutyunyan

Редакторы
Рубен Амбарцумян, Борис Нахапетян,
Евгений Арутюнян, Мамикон Гиновян,
Алексан Симонян, Давид Асатрян,
Ашот Арутюнян

Նվիրվում է

Հայաստանում հավանականության տեսության և մաթեմատիկական վիճակագրության ներդրման ու դասավանդման անխոնջ նախագնացների՝

Գոհար Համբարձումյանի,
Սարգիս Թումանյանի

և

սույն գրքի ստեղծմանն իր ավանդը բերած, ինֆորմատիկայի և ծրագրավորման հայաստանյան առաջին սերնդի խանդավառ մասնագետ ու մանկավարժ

Լոսա Խզմասանի

հիշատակին։

Բովանդակություն

Առաջաբան

Գիտությունը նման է ժամացույցի՝ օգտագործելու համար միշտ պետք է լարել:

Հակոբ Պարոնյան

Ես փախա դպրոցից այն բանից հետո, երբ գլուխս սկսեցին լցնել հանրահաշվով, ոչինչ չասելով, որ այդ առարկան ինչ համար հետաքրքիր դառնա:

Բեռնարդ Շոու

Հաշվարային (կոմպյուտերային) տեխնիկայի բուռն զարգացման ներկա ժամանակաշրջանում հնարավոր է դառնում էապես ընդլայնել կիրառական բազմապիսի խնդիրների լուծման նպատակով մաթեմատիկական եղանակների օգտագործման սահմանները: Դա արտահայտվում է նաև մաթեմատիկական վիճակագրության ավելի ու ավելի արդյունավետ ներդրմամբ: Վիճակագրական մեթոդներն ունեն ամենօրյա կիրառություններ: Օրինակ, շուկայագիտական (մարքեթինգային) հետազոտությունները չեն կարող իրականացվել առանց անհրաժեշտ վիճակագրական եղանակների կիրառման:

Գոյություն ունեն վիճակագրական տվյալների ձեռքբերման երկու միջոց. հաշվեհամար (որի ընթացքում կատարվում է հետազոտման առարկա բազմության (հանույթի) բոլոր տարրերի տվյալների գրանցում, բնակչության ուսումնասիրման դեպքում այն կոչվում է մարդահամար), և ընտրանքային հետազոտություն, երբ գրանցվում են հատուկ ձևով հանույթից վերցված սահմանափակ թվով տարրերի տվյալները, ինչը, բնականաբար, ավելի մատչելի է և՛ ժամանակի, և՛ աշխատանքի ծավալների առումներով: Տարածված է այն տեսակետը, որ հաշվեհամարը տալիս է ուսումնասիրվող հանույթի մասին սպառիչ տեղեկություններ: Սակայն մեծ աշխատանքն ունի իր բարդությունները, հաճախ բավարար թվով մասնագետ վիճակագիրների բացակայության պատճառով ներգրավվում են ոչ լրիվ իրավասու անձինք, որոնց թեր աշխատանքը բերում է ընդհանուր արդյունքների անճշտությունների: Տեղին է մեջբերել հետևյալ միտքը ճանաչված ամերիկյան գիտնականներ Ֆ. Մոստելլերի, Ռ. Ռուռկեի, Ջ. Թոմասի «Հավանականություն» հանրահայտ հանրամատչելի գրքից. «ԱՄՆ մարդահամարի արդյունքներում սխալների հայտնաբերման և ուղղման նպատակով ԱՄՆ ազգային մարդահամարի անցկացման վարչությունն օգտագործում է հատուկ ընտրանքային հարցումներ: Նման հարցումները լայնորեն կիրառվում են և մասամբ փոխարինում են լրիվ մարդահամարը, որն զգալիորեն ավելի աշխատատար է և չի ապահովում ստացվող տեղեկությունների ճշգրտորեն ավելի բարձր արժեք: *Ընտրանքային ուսումնասիրության հիման վրա հնարավոր է անել ավելի մանրամասն և նույնիսկ ավելի որակյալ եզրակացություններ, քան ամբողջ մեծ հանույթի ուսումնասիրման հիման վրա»:*

Դա բացատրվում է ժամանակակից մաթեմատիկական վիճակագրության մեթոդների տվյալներից բոլոր հնարավոր տեղեկությունները քաղելու հզորությամբ:

Այժմեական համընդհանուր օգտագործման հաշվարային ծրագրաշարերի և մասնագիտացված մաթեմատիկական ծրագրաշարերի անհրաժեշտ բաղկացուցիչ մասեր են կազմում առավել հաճախ օգտագործվող վիճակագրական ծրագրերը:

Միաժամանակ ստեղծվում և լայնորեն օգտագործվում են հատուկ, նորագույն կատարելագործված եղանակները ներառող վիճակագրական ծրագրաշարեր, մասնավորապես, անձնական օգտագործման հաշվարկների համար: Այդ ծրագրաշարերը մեծապես նպաստում են վիճակագրական եղանակների լայն կիրառմանը, և, հետևաբար, ծանոթությունը դրանց կառուցվածքին և հնարավորություններին անհրաժեշտ է: Գրքի մի առանձին գլուխ նվիրված է դրանցից առավել լայնորեն կիրառվող մի քանիսի ներկայացմանը:

Սակայն հարկավոր է գիտակցել, որ առանց մաթեմատիկական վիճակագրության գաղափարները, հիմնական սկզբունքներն ու գործնական եղանակները բավարար չափով յուրացնելու հնարավոր չէ հաջողությամբ լուծել լուրջ կիրառական խնդիրներ:

Քանի որ մաթեմատիկական վիճակագրության հիմքը հավանականության տեսությունն է, վիճակագրության կողմից առաջարկվող եղանակներն ու գործիքները հասկանալու, գրագիտորեն ընտրելու և օգտագործելու համար հարկավոր է տիրապետել այդ տեսության հիմնական գաղափարներին և սկզբունքներին:

Գիրքը պարունակում է այն կարևոր գիտելիքները, որոնց իմացությունը ճանապարհ է բացում ժամանակակից վիճակագրական եղանակների և հաշվարկային ծրագրաշարերի արդյունավետ օգտագործման համար:

Վիճակագրության ուսուցման միջազգային գործընթացը ներկայումս շարունակվում է նոր սոցիալական պայմաններում: Այն կրում է մաթեմատիկական առարկաների ուսուցման հանրայնացման բարեփոխական շարժման ազդեցությունը: Ընդսմին առարկայի բնույթը պահանջում է ուսուցման ժամանակակից բովանդակություն և տեխնոլոգիա: Վիճակագրության ուսուցման «պայուսակի» էական վերանայումը հիմնվում է բովանդակության, դասավանդման և տեխնոլոգիայի խիստ համագործակցության վրա: Վիճակագրություն դասավանդող մասնագետները պետք է քաջատեղյակ լինեն ուսումնական տեխնոլոգիայի նորույթներին:

Իհարկե, ճշգրիտ որևէ գիտություն ուսումնասիրելու, յուրացնելու լավագույն միջոցը վարժությունների լուծումն է: Այդ պատճառով գրքում բերված են բազմաթիվ՝ մանրամասնորեն վերլուծված, առավելապես կիրառական բնույթի օրինակներ, իսկ զուգահեռ հրատարակվող խնդրագրքում առաջարկվում է լուծել բազմապիսի խնդիրներ, որոնցից շատերը կապված են իրական իրադրությունների հետ, օգտագործում են տվյալներ տնտեսության տարբեր բնագավառներից: Վարժությունների ընտրության ընթացքում օգտագործվել են նաև բազմաթիվ աղբյուրներ (դասագրքեր, խնդրագրքեր և այլն):

Դասագրքի և խնդրագրքի ստեղծմանը մասնակցել են մեծ թվով մասնագետներ, հեղինակներ, խմբագիրներ (մասնագիտական և լեզվական), հաշվարկային ձևավորումը իրագործողներ: Զուգորդվել են մանկավարժական և կիրառական վիճակագրական աշխատանքների մեծ փորձ ունեցող մասնագետների և երիտասարդ գիտաշխատողների ջանքերը:

Հեղինակները ստեղծել են գրքի հետևյալ մասերը. Եվգենի Հարությունյանը՝ գլուխներ 1 (առաջին կեսը), 6, 7, 8 և համապատասխան վարժությունները, Տաթևիկ Ղազանչյանը՝ գլուխներ 1 (երկրորդ կեսը), 2, 5 ու վարժությունների մի մասը, Մարիամ Հարությունյանը՝ գլուխներ 3 և 4 ու դրանց վարժությունները, Նաիրա Մեսրոպյանը՝ գլուխ 9 և դրա վարժությունները, Դավիթ Ասատրյանը՝ գլուխներ 10, 11, 12, 13 և վարժությունները, Մելս Սահակյանը՝ գլուխ 14, Հարություն Շահումյանը՝ գլուխ 15, Անահիտ Ղազարյանը՝ 1 և 2 գլուխների վարժությունների մի մասը: Մասնագիտական խմբագրումը կատարել են. Եվգենի Հարությունյանը՝ բոլոր գլուխները, Ռուբեն Համբարձումյանը՝ գլուխ 1, մի շարք առանցքային խորհուրդներով հանդերձ, Բորիս Նահապետյանը՝ գլուխներ 1, 2, 3, 4, 5, Դավիթ Ասատրյանը՝ գլուխներ 6, 7, 8, 9, 14, Մամիկոն Գինովյանը՝ գլուխներ 10, 11, 12, 13, Ալեքսան Սիմոնյանը՝ գլուխ 9, Աշոտ

Հարությունյանը՝ գլուխներ (մասամբ) 6, 7, 8, 14, 15:

Համակարգչային ձևավորման պատասխանատուն էր Աշոտ Հարությունյանը, բազմաթիվ անգամ (որոշ գլուխների դեպքում 30-ի հասնող) վերամշակումներն ու ուղղումները գրանցելու և վերատպելու մեծ աշխատանքը համբերատար կատարել են Լյուսյա Խզմալյանը, Անահիտ Ղազարյանը, Մարիամ Հարությունյանը, Դավիթ Ասատրյանը, մասնակցել են նաև Արմեն Սողոյանը, Հարություն Շահումյանը, Լև Թադևոսյանը: Շարադրանքի լեզվական խմբագրումը իրագործել են Ազատուհի Սահակյանը և Մուշեղ Համբարձումյանը:

Մեծ ուշադրություն է պահանջում տերմինաբանական աշխատանքը, դրան այս կամ այն չափով մասնակցել են բոլոր հեղինակներն ու խմբագիրները: Զգալի օգնության համար (որի շնորհիվ տերմինները մանրազնին քննարկվեցին և դրանց մի մասը հաստատվեց Լեզվի բարձրագույն խորհրդի կողմից) շնորհակալ ենք ՀՀ Կառավարությանն առընթեր Լեզվի պետական տեսչության պետ Լևոն Գալստյանին, նույն տեսչության աշխատակից Եսայի Թադևոսյանին, Լեզվի բարձրագույն խորհրդի նախագահ, ակադեմիկոս Գևորգ Ջահուկյանին, նույն խորհրդի պատասխանատու քարտուղար, պրոֆեսոր Հովհաննես Բարսեղյանին, խորհրդի անդամներ, լեզվաբաններ Լիանա Հովսեփյանին և Լիլիթ Բրուտյանին: Կարևոր էր Մաթեմատիկական տերմինների հանձնաժողովի նախագահ Թովմաս Թովմասյանի, հանձնաժողովի անդամներ Մելս Սահակյանի և ակադեմիկոս Ռուբեն Համբարձումյանի ներդրումը::

Հարկ ենք համարում անդրադառնալ հայերեն նոր տերմինների ստեղծման ու եղածների հղկման գործում մեր կողմից՝ որպես ուղենիշային ընդունած սկզբունքներին: Նախևառաջ, տերմինի հայացումը համարել ենք խիստ ցանկալի, սակայն ոչ առաջին պատահած, լավ չկշռադատված, չհիմնավորված տարբերակով: Նախընտրելի է հայերեն համարժեքները չհորինել այն դեպքերում, երբ բառարաններից կարելի է գտնել գուցե և սակավ օգտագործվող, բայց իմաստով հարմար բառեր: Տերմինը թարգմանելիս նախապատվությունը տվել ենք հայերենում հնարավոր ամենահակիրճ տարբերակին: Ձգտել ենք, որ նոր տերմինը հնարավորինս լավ արտահայտի հասկացության իմաստը և լինի հարմար ածանցյալ տերմիններ կազմելու: Հիմնվել ենք ռուսերեն, անգլերեն, ֆրանսերեն, գերմաներեն, հունարեն համարժեքների համեմատական վերլուծության վրա: Տերմիններին վերաբերող որոշակի հղումներ կան նաև տեքստի համապատասխան տեղերում, իսկ գրքի վերջում զետեղված է հայերեն նոր (կամ վերանայված) տերմինների ու բառակապակցությունների ցանկը, դրանց անգլերեն և ռուսերեն համարժեքների և որոշ բացատրությունների հետ միասին:

Գիտակցելով, որ չնայած մեր բոլոր ջանքերին, աշխատանքը զուրկ չէ թերություններից, շնորհակալ կլինենք ընթերցողների դիտողությունների ու առաջարկությունների համար:

Եվգենի Հարությունյան

Ներածություն

Հետազոտության հավանականային եղանակներն ունեն շատ լայն, համարյա համապարփակ նշանակություն:

Վիկտոր Համբարձումյան

Հավանականության տեսությունն անհրաժեշտ է դասավանդել, որովհետև այն կարևոր դեր է կատարում սովորողների մտածողության զարգացման գործում: Դրա եզրակացությունները կիրառություն են գտնում առօրյա կյանքում, գիտությունում, տեխնիկայում և այլուր: Այն ունի կարևոր, ոչնչի հետ չհամեմատվող նշանակություն մաթեմատիկական կրթության համար:

Ալֆրեդ Ռենյի

«Վիճակագրություն» տերմինը հիմնականում օգտագործվում է երկու իմաստով: Առաջինն՝ ըստ բառի կազմության, արտացոլում է թվային տվյալների միջոցով «վիճակների նկարագրության» գաղափարը: Այս դեպքում անընդհատ կատարելագործվում են տվյալներ հավաքելու, խմբավորելու և պահպանելու հնարավորությունները: Նշված իմաստի տեսակետից կազմվում են ածականներ, որոնք ցույց են տալիս կիրառական այն ոլորտը, որի տվյալների նկարագրությունն է ներկայացվում: Օրինակներ են. **ազգաբնակչության վիճակագրությունը**, որն անվանում են նաև **ժողովրդագրություն** (այն բնութագրում է մարդկանց բաշխումն ըստ ժամանակամիջոցների, սեռի, տարիքի, բնակավայրի, ծննդավայրի, ծնելիության, մահացության, ամուսնությունների, կրթության աստիճանի, հանցագործությունների հաճախության և այլն), **օդերևութաբանական վիճակագրությունը** (ուսումնասիրում է տարբեր ժամանակաշրջաններում ջերմաստիճանների բաշխման, տեղումների, քամու ուղղության և արագության վերաբերյալ տվյալներ), **տնտեսական, երկրաբանական, միջազգային, բժշկական, առողջապահական, շրջակա միջավայրի վիճակագրությունները**:

Տերմինի երկրորդ իմաստն արտահայտում է բառի գիտա-մեթոդաբանական կողմը: Վիճակագրության՝ որպես գիտության, կարևոր մասը կազմում է մաթեմատիկական վիճակագրությունը, որը մաթեմատիկայի ճյուղ է և վերաբերում է այսպես կոչված «ստոխաստիկ մաթեմատիկային», որն զբաղվում է պատահական երևույթների օրինաչափություններով, և որի «միջուկն» է հավանականության տեսությունը:

Մաթեմատիկայի մյուս բաղկացուցիչ մասերի նման հավանականության տեսությունը ստեղծվել և զարգացվել է մարդկության կենսական խնդիրների լուծման նպատակով: Այդ գիտական ուղղությունն ուսումնասիրում է հատուկ դասի՝ *զանգվածային պատահական* երևույթների օրինաչափությունները: Այսպես, օրինակ, ապահովագրված օբյեկտի (տուն, ավտոմեքենա և այլն) վնասվելը՝ տարերային աղետի կամ այլ պատճառի հետևանքով, «պատահականության» արդյունք է, սակայն մեծ թվով այդպիսի օբյեկտների ընդհանուր վիճակի մասին կարելի է անել բավականին վստահելի եզրահանգումներ, որոնք, ի դեպ, ապահովագրական գործակալությունների գոյատևման գրավականն են:

Հավանականության տեսությունը հետազոտում է հավանականային մոդելներ, որոնք նկարագրում են մեծ թվով պատահական գործոնների համատեղ ազդեցության տակ գտնվող իրական երևույթների և համակարգերի վարքի օրինաչափությունները: Մաթեմատիկական վիճակագրության միջոցները թույլ են տալիս հավանականության տեսության մոդելների հավաքածուից ընտրել հետազոտողի ունեցած վիճակագրական տվյալներին որոշ իմաստով լավագույն ձևով համապատասխանող տարբերակը:

Զանգվածային պատահական երևույթները մշտապես և ամենուրեք ուղեկցում են մեզ: Նշենք՝ իրադարձության պատահական լինելը չի նշանակում, որ այն պատճառաբանված չէ: Իհարկե, ամեն մի երևույթ սերտորեն կապված է շրջապատի ուղեկցող եղելությունների հետ և կարող է բացատրվել այդ կապերի միջոցով: Սակայն, եթե երևույթը պայմանավորված է մեծ թվով գործոններով, որոնք ժամանակի տարբեր պահերին կարող են արտահայտվել տարբեր ձևով, ապա դրանց գումարային ազդեցությունը տվյալ դեպքի վրա իրականում միարժեքորեն գուշակելի չէ:

Դեռ XVII դարի սկզբին նպատակ է դրվել լուծել զանգվածային պատահական երևույթներին վերաբերող խնդիրներ: Գալիլեո Գալիլեյը հետազոտում էր ֆիզիկական չափումների սխալները: Փորձեր են արվել ստեղծել մարդկանց ապահովագրման ընդհանուր տեսություն՝ հիմնված հիվանդությունների և դժբախտ պատահարների օրինաչափությունների վերլուծության վրա:

Դասական հավանականության տեսության հիմնադիրները՝ Բլեզ Պասկալը, Պիեռ Ֆերման, Քրիստիան Հյուգենսը, XVII դարի կեսերին սկսել են պատահական տարրեր պարունակող խաղերի (զառերով, խաղաքարտերով և այլն) ուսումնասիրությունից: XVIII դարի սկզբին Յակոբ Բեռնուլին ձևակերպեց փորձում պատահույթի հավանականությունը՝ որպես այդ պատահույթին նպաստող ելքերի (շանսերի) թվի հարաբերությունը բոլոր հնարավորների թվին: 1812-ին Պիեռ Լապլասը ավելի խիստ ձևակերպեց պատահույթի հավանականության սահմանումը, պահանջելով բոլոր ելքերի հավասարահնարավորությունը: Նա ապացուցեց հավանականության տեսության սահմանային թեորեմներից պարզագույնը: Էմիլ Բորելը ձևակերպեց բազմության չափի գաղափարի օգտագործման անհրաժեշտությունը:

XX դարի սկզբին մի քանի նշանավոր մաթեմատիկոսներ առաջարկեցին հավանականության տեսության աքսիոմատիկ հիմնավորման մի քանի տարբերակներ: Սակայն բոլորի կողմից ընդունվեց և այժմ անվիճելի է խոշորագույն մաթեմատիկոս Անդրեյ Կոլմոգորովի՝ 1933 թ. հրատարակված «Հավանականության տեսության հիմնական գաղափարները» գրքում շարադրված՝ հավանականության տեսության կառուցման տեսակետը: Դրանից հետո, ունենալով խիստ և կայուն մաթեմատիկական հիմք, հավանականության տեսությունը դարձավ ավելի բուռն զարգացող գիտական ուղղություն:

Հավանականության տեսության զարգացման գործում մեծ է արևմտաեվրոպական և ռուսական գիտնականների վաստակը: Մի շարք նոր կիրառական ուղղությունների հիմնադրման և լայն օգտագործման, բազմապիսի ընթերցողների պահանջներին համապատասխանող հարուստ գրականության ստեղծման տեսակետից վերջին կես դարում նկատելի են Ամերիկայի Միացյալ Նահանգների գիտնականների նվաճումները:

XX դարի երկրորդ կեսից ի վեր հավանականության տեսության բնագավառում արդյունավետ հետազոտություններ են կատարվում նաև Հայաստանում: 1958թ. սեպտեմբերին Երևանում տեղի ունեցավ Հավանականության տեսության և մաթեմատիկական վիճակագրության համամիութենական խորհրդակցություն, որի կազմակերպիչները էին դոցենտներ Գոհար Համբարձումյանը և Սարգիս Թումանյանը: Ակտիվորեն մասնակցում էր, բացման խոսքով և գիտական զեկուցումով հանդես եկավ Հայաստանի Գիտությունների ակադեմիայի պրեզիդենտ, ակադեմիկոս Վիկտոր

Համբարձումյանը: Նա մասնավորապես նշեց. «Հավանականության տեսությունը ... դարձել է գիտության ընդարձակ, չափազանց բազմակողմանի ոլորտ, որն ունի մեծ թվով կիրառություններ: Այդ չափազանց կարևոր ուղղությանը մենք մինչ այժմ կարողացել ենք հատկացնել դեռ շատ փոքր ուժեր»:

Վերջին տասնամյակներում մեր երկրում արդեն զարգանում են այդ գիտության մի շարք մաթեմատիկական և որոշակի կիրառական ուղղվածություն ունեցող ճյուղեր: Կարևոր աշխատանքներ են տարվում տնտեսագիտական վիճակագրության ասպարեզում: Այս դասագիրքն ու կից խնդրագիրքը կոչված են նպաստելու հանրապետությունում այդ գործընթացի ընդլայնմանը և կատարելագործմանը: Հուսով ենք, որ գրքերը օգտակար կլինեն զանազան բնագավառներում վիճակագրական հետազոտություններ կատարողներին, ուսանողներին, մագիստրատուրայում և ասպիրանտուրայում սովորողներին, գործարարներին, ֆինանսական և բանկային մասնագետներին:

Բաժին Ա
Հավանականության տեսության սկզբունքներ
Գլուխ 1
Պատահույթներ, հավանականային տարածություն

Ես կարծում եմ, որ առարկայի ուշադիր ուսումնասիրության ընթացքում ընթերցողը կնկատի, որ գործ ունի ոչ միայն խաղի հետ, այլ որ այստեղ շատ հետաքրքիր և խորը տեսության հիմք է դրվում:

Քրիստիան Հյուգենս

Դիպվածը զառերով խաղացող է: *Թոմաս Ֆուլեր*

1.1. Գործողություններ պատահույթների հետ

Հավանականության տեսության ուսումնասիրման առարկան պատահական (անվանում են նաև ստոխաստիկ) երևույթների վիճակագրական հատկություններն են: Տրամաբանական միջոցներով տեսությունը գործում է ստոխաստիկ մոդելների հետ, որոնք որոշակի ձևական գաղափարների և աքսիոմների միջոցով ներկայացնում են պատահական երևույթների էական կողմերը:

Տեսական գիտելիքները կիրառելիս հարկավոր է կարողանալ վերացական տեսության պնդումները «թարգմանել» առարկայական, բովանդակային տերմինների և գաղափարների լեզվի: Նոր գիտության դրույթները յուրացնելու համար պետք է սովորել այդպիսի թարգմանությունը վարժ կատարել: Դա հատկապես կարևոր է նաև մեր առարկան՝ հավանականության տեսությունը և վիճակագրությունն ուսումնասիրելիս:

Հավանականության տեսության գաղափարների շարքում առաջինը **վիճակագրական փորձի** գաղափարն է, որն արտացոլում է որոշակի պայմաններում բազմիցս կրկնելի այնպիսի գործողությունների համալիր, որոնց իրականացման արդյունքում կարող է տեղի ունենալ հնարավոր ելքերից մեկն ու մեկը: Այդպիսի փորձերի ավանդական օրինակներ են մետաղադրամի նետումը, զառի նետումը, զանազան գնդակներ պարունակող սափորից գնդակի հանումը և այլն: Այդ և նման օրինակները հարմար են հավանականության տեսության ուսուցման ժամանակ, հատկապես սկզբնական փուլում:

Վիճակագրական փորձի բոլոր հնարավոր պարզագույն ելքերի բազմությունը կոչվում է **տարրական պատահույթների տարածություն**: Դրա տարրը՝ **տարրական պատահույթը**, նշանակում են ω, իսկ ամբողջ տարածությունը՝ Ω: Դիտարկում են Ω-ի ենթաբազմությունների որոշակի $\mathcal{F}$ դաս, որի տարրերը կոչվում են **պատահույթներ** և նշանակվում են $\mathcal{A}, \mathcal{B}, \mathcal{C}$ և այլն: Պատահույթը կարող է փորձի արդյունքում տեղի ունենալ կամ տեղի չունենալ: Ասում են, որ $\mathcal{A}$ պատահույթը տեղի է ունեցել (հանդես է եկել, իրականացել է), եթե փորձի ելքը՝ ω-ն, $\mathcal{A}$-ի տարրն է՝ $\omega \in \mathcal{A}$: $\mathcal{F}$-ի մեջ միշտ մտնում են երկու առանձնահատուկ պատահույթներ. Ω-ն, որը դիտարկվում է նաև որպես պատահույթ, որն անպայման տեղի է ունենում և կոչվում է **հավաստի**, և պատահույթը, որը տվյալ փորձում չի իրականանում, նշանակվում է $\varnothing$ և կոչվում է **անհնար**:

<u>*Օրինակ*</u> 1: Նետվում է *կանոնավոր մետաղադրամը*: Մեկ նետումը փորձն է: Մետաղադրամն ընկավ «զինանշանով վեր» և մետաղադրամն ընկավ «թվով վեր» տարրական պատահույթներն են:

Օրինակ 2: *Զառի նետման փորձը*: Նշանակենք ω_n, $n = \overline{1,6}$, կանոնավոր զառի նետումից հետո վերևի նիստում n կետեր հանդես գալու տարրական պատահույթը, ուրեմն $\Omega = \{\omega_1, \omega_2, \cdots, \omega_6\}$, իսկ $\mathcal{F}$ դասը պարունակում է Ω-ի բոլոր հնարավոր ենթաբազմությունները՝

$\mathcal{F} = \{\varnothing, \{\omega_{n_1}\}, \cdots, \{\omega_{n_1}, \omega_{n_2}\}, \cdots, \{\omega_{n_1}, \omega_{n_2}, \omega_{n_3}\}, \cdots, \{\omega_{n_1}, \omega_{n_2}, \omega_{n_3}, \omega_{n_4}, \omega_{n_5}\}, \cdots, \Omega\}$, որտեղ n_1-ը, ..., n_5-ը ընդունում են 1-ից 6 արժեքները և իրար հավասար չեն: Եթե «հանդես կգա զույգ թիվ» պատահույթը նշանակենք $\mathcal{A}$, իսկ «ելքը չի գերազանցի 4» պատահույթը՝ $\mathcal{B}$, ապա կարող ենք գրել $\mathcal{A} = \{\omega_2, \omega_4, \omega_6\}$, $\mathcal{B} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$:

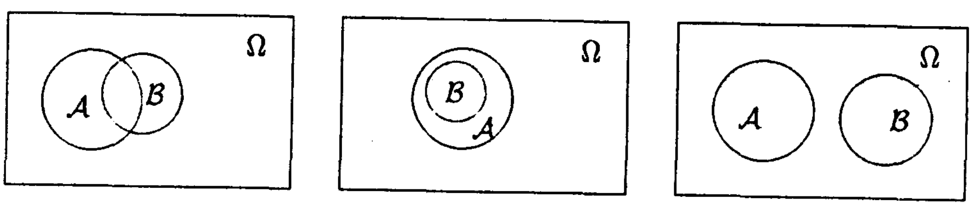
Պատահույթների միջև սահմանվում են տարբեր հարաբերություններ և գործողություններ, որոնք հեշտ կյուրացվեն, եթե նկատենք, որ դրանք համընկնում են բազմությունների միջև համապատասխան հարաբերությունների և գործողությունների հետ:

$\mathcal{A}$ պատահույթը կոչվում է $\mathcal{B}$-ի **մասնավոր դեպք**, եթե $\mathcal{A}$ պատահույթը իրականանալիս միաժամանակ հանդես է գալիս $\mathcal{B}$ պատահույթը. գրում են $\mathcal{A} \subset \mathcal{B}$:
Երկու պատահույթներ **նույնն** են, ասում են նաև իրար **հավասար են**՝ $\mathcal{A} = \mathcal{B}$, այն և միայն այն դեպքում, երբ միաժամանակ $\mathcal{A} \subset \mathcal{B}$ և $\mathcal{A} \supset \mathcal{B}$:

$\mathcal{A}$ և $\mathcal{B}$ պատահույթների **միավորում**՝ $\mathcal{A} \cup \mathcal{B}$, անվանվում է այն պատահույթը, որը տեղի է ունենում, երբ իրականանում է նրանցից գոնե մեկը:

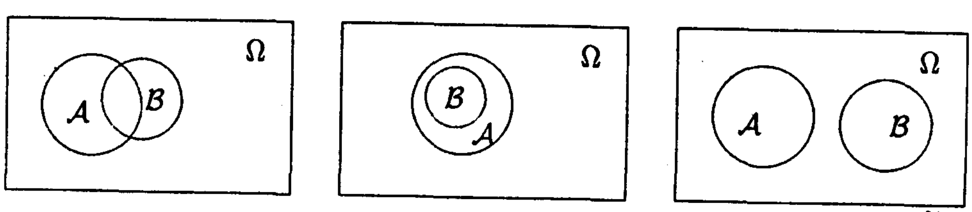
Օրինակ 3: Զառի նետման փորձում (տե՛ս օրինակ 2) $\mathcal{A} \cup \mathcal{B} = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_6\}$:

Եթե փորձը կետը «պատահականորեն» Ω ուղղանկյան մեջ նետելն է, իսկ $\mathcal{A}$-ն և $\mathcal{B}$-ն՝ շրջանների մեջ ω կետի ընկնելը՝ համապատասխան պատահույթի իրականանալը, ապա $\mathcal{A} \cup \mathcal{B}$ պատահույթը կպատկերվի նկար 1-ում գունանշված տիրույթով: Այդպիսի պատկերումները կոչվում են **Վենի գծապատկերներ**:



Նկար 1: $\mathcal{A}$ և $\mathcal{B}$ պատահույթների միավորման տարբեր դեպքերի՝ Վենի գծապատկերները:

$\mathcal{A}$ և $\mathcal{B}$ պատահույթների **հատումը** (կամ **արտադրյալը**) նշանակվում է $\mathcal{A} \cap \mathcal{B}$ կամ ավելի հաճախ $\mathcal{AB}$, հանդես է գալիս այն և միայն այն դեպքում, երբ $\mathcal{A}$-ն և $\mathcal{B}$-ն համատեղ են իրականանում:



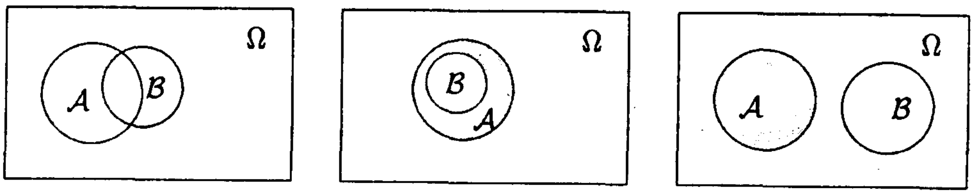
Նկար 2: $\mathcal{A}$ և $\mathcal{B}$ պատահույթների հատման դեպքեր:

Օրինակ 4: Զառի նետման փորձում՝ $\mathcal{A} \cap \mathcal{B} = \{\omega_2, \omega_4\}$:

Կարևոր են այդ **գործողությունների** հետևյալ **հատկությունները**.
համաչափությունը (սիմետրիկությունը)՝ $A\cup B=B\cup A$, $A\cap B=B\cap A$,
զուգորդականությունը՝ $(A\cup B)\cup C=A\cup(B\cup C)$, $(A\cap B)\cap C=A\cap(B\cap C)$,
բաշխականությունը՝ $(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$:

Ապացուցումները դժվար չէ ստանալ սահմանումներից:

$A-B$ պատահույթը (A-ի և B-ի **տարբերությունը**) իրականանում է այն և միայն այն դեպքում, երբ A-ն տեղի է ունենում, իսկ B-ն՝ ոչ:



Նկար 3: $A-B$ պատահույթի դեպքերի Վենի գծապատկերները:

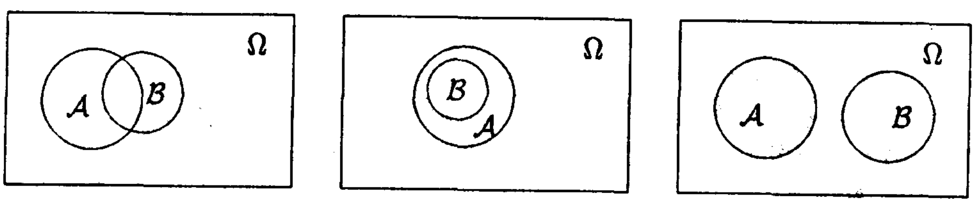
Որոշ դեպքերում հարմար է գրել այսպես՝ $A-B=\{\omega:\omega\in A,\ \omega\notin B\}$: Այս արտահայտության մեջ երկկետը կարդացվում է այսպես. «այն ω-ները, որոնք բավարարում են երկկետից հաջորդող պնդումներին», տվյալ դեպքում՝ ω-եր, որոնց համար A-ն տեղի է ունենում, իսկ B-ն՝ ոչ: Նման ձևով, օրինակ,

$$A\cap B=\{\omega:\omega\in A,\ \omega\in B\}:$$

Երբեմն պետք է լինում նաև **համաչափ տարբերության գործողությունը** (նկար 4)

$$A\triangle B=(A-B)\cup(B-A):$$

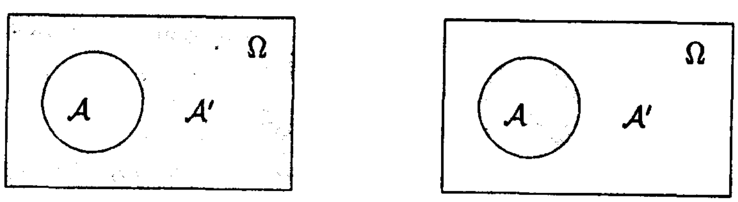
$A\triangle B$ պատահույթը տեղի ունի այն և միայն այն դեպքում, երբ տեղի ունի կամ A-ն, կամ B-ն, բայց ոչ երկուսը միասին:



Նկար 4: $A\triangle B$ պատահույթի Վենի գծապատկերները:

A պատահույթի **հակադիր** պատահույթը՝ A'-ը (ասում են՝ A-ն տեղի չի ունեցել, կամ անվանում են A-ի **ժխտում**, նաև A-ի **լրացում** կամ **լրույթ**), սահմանվում է հետևյալ կերպ՝

$$A'=\Omega-A:$$



Նկար 5: Հակադիր պատահույթներ:

$\mathcal{A}$ և $\mathcal{B}$ պատահույթները կոչվում են **անհամատեղելի**, եթե $\mathcal{A} \cap \mathcal{B} = \emptyset$:

<u>*Օրինակ*</u> 5: 1. $\mathcal{A}$-ն և $\mathcal{A}'$-ը անհամատեղելի են, որովհետև հնարավոր չէ, որ միաժամանակ $\mathcal{A}$-ն տեղի ունենա և տեղի չունենա:

<u>*Օրինակ*</u> 6: Անհամատեղելի են $\mathcal{A}$ և $\mathcal{B}$ պատահույթները 1-ից 4 նկարների աջ կողմից պատկերված դեպքերում:

Հարկ է լինում դիտարկել պատահույթների որոշակի հատկություններով օժտված ընտանիքներ, դասեր:

Ասում են, որ $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ պատահույթները կազմում են **լրիվ խումբ**, եթե

$$\mathcal{A}_{n_1} \cap \mathcal{A}_{n_2} = \emptyset, \quad n_1 \neq n_2, \quad n_1, n_2 = \overline{1, N}\,^{\dagger}, \quad \text{և} \quad \bigcup_{n=1}^{N} \mathcal{A}_n = \Omega:$$

<u>*Օրինակ*</u> 7: Զառի նետման փորձում $\{\omega_1, \omega_2\}$, $\{\omega_3, \omega_4, \omega_5\}$, $\{\omega_6\}$ պատահույթները կազմում են լրիվ խումբ:

Շատ կարևոր է σ-հանրահաշիվ կոչվող ընտանիքը:

Ω-ի ենթաբազմությունների $\mathcal{F}$ դասը կոչվում է σ-**հանրահաշիվ**, եթե այն բավարարում է հետևյալ երեք պայմաններին. 1. $\Omega \in \mathcal{F}$, 2. եթե $\mathcal{A} \in \mathcal{F}$, ապա $\mathcal{A}' \in \mathcal{F}$, 3. եթե $\mathcal{A}_n \in \mathcal{F}$, $n = 1, 2, \ldots$, ապա $\bigcup\limits_{n=1}^{\infty} \mathcal{A}_n \in \mathcal{F}$:

Այսուհետև պատահույթներ կանվանենք միայն $\mathcal{F}$ σ-հանրահաշվի տարրերը:

Նշենք, որ այդ հատկություններից հետևում է, որ σ-հանրահաշվին պատկանում են ոչ միայն լրացման և հաշվելի թվով միավորման գործողությունների, այլ նաև *բոլոր դիտարկված գործողությունների հաշվելի թվով կիրառման արդյունքները*:

1.2. Հավանականության գաղափարը

Պատահույթի **հավանականությունը** (բովանդակային իմաստով) նրա իրականացման հնարավորության աստիճանի թվային գնահատականն է: Հավաստի պատահույթի հավանականությունը մեծագույնն է, այն ընդունում են հավասար 1-ի, իսկ անհնարինինը (փոքրագույնինը)՝ հավասար 0-ի:

Հավանականության մաթեմատիկական սահմանումը ժամանակի ընթացքում աստիճանաբար ընդլայնվում ու հղկվում էր և իր՝ ներկայումս ամենուրեք ընդունված աքսիոմատիկ տեսքը ստացավ Ա. Ն. Կոլմոգորովի ձևակերպմամբ: Այն է.

P հավանականությունը սահմանվում է որպես ֆունկցիա, որն ամեն մի $\mathcal{A}$ պատահույթին ($\mathcal{A} \in \mathcal{F}$) համապատասխանեցնում է $\mathrm{P}\{\mathcal{A}\}$ թիվը՝ $\mathcal{A}$-ի հավանականությունը, այնպես որ տեղի ունենան հետևյալ հատկությունները (աքսիոմները).

1. $\mathrm{P}\{\mathcal{A}\} \geq 0$, 2. $\mathrm{P}\{\Omega\} = 1$,

3. եթե $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n, \ldots$ պատահույթները անհամատեղելի են՝ $\mathcal{A}_{n_1} \cap \mathcal{A}_{n_2} = \emptyset$, երբ $n_1 \neq n_2$, ապա (**գումարականության (ադիտիվության) աքսիոմը**)

$$\mathrm{P}\{\bigcup_{n=1}^{\infty} \mathcal{A}_n\} = \sum_{n=1}^{\infty} \mathrm{P}\{\mathcal{A}_n\}:$$

[†] Նշենք, որ $n = 1, 2, \ldots, N$ գրառման փոխարեն հաճախ օգտագործվում է ավելի կարճ գրառում՝ $n = \overline{1, N}$:

Մասնավորապես, տեղի ունի նաև անհամատեղելի պատահույթների վերջավոր գումարականությունը՝

$$\mathbf{P}\{\bigcup_{n=1}^{N} \mathcal{A}_n\} = \sum_{n=1}^{N} \mathbf{P}\{\mathcal{A}_n\}:$$

$\{\Omega,\mathcal{F},\mathbf{P}\}$ եռյակը կոչվում է **հավանականային տարածություն**: Սա պատահական ելքերով փորձի մաթեմատիկական մոդելն է, հավանականության տեսության առաջին հիմնական օբյեկտը: Հավանականության այս ընդհանուր մաթեմատիկական սահմանումը այն «հիմքն» է, որի վրա կառուցվում է հավանականության տեսության շենքը: Սակայն այդ սահմանումը, արտացոլելով հավանականության ընդհանուր հատկությունները, իսկ որոշակի գործնական իրավիճակներում հավանականությունների թվային արժեքների որոշումը պահանջում է լրացուցիչ հետազոտություն: Այդ նպատակին կարող են ծառայել հետևյալ «մասնավոր» սահմանումները:

Հավանականության որոշման եղանակը, որը պատմականորեն եղել է առաջինը, հիմնվում է պատահույթների հավասարահնարավորության գաղափարի վրա և կոչվում է **դասական սահմանում**:

> Դիցուք Ω-ն կազմված է N հատ տարրական պատահույթներից.
> $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$: $\mathcal{F}$ σ-հանրահաշիվը այս դեպքում ընդգրկում է Ω-ի բոլոր ենթաբազմությունները: Ենթադրենք, որ տարրական պատահույթները հավասարահնարավոր են, «հավասարահավանական» են՝ $\mathbf{P}\{\omega_n\} = 1/N$, $n = \overline{1,N}$: Նշանակենք $n(\mathcal{A})$-ով այն տարրական պատահույթների թիվը, որոնք $\mathcal{A}$-ի մասնավոր դեպքերն են: $\mathcal{A}$-ի **հավանականությունը** հավասար է $\mathcal{A}$-ին նպաստող հավասարահնարավոր տարրական պատահույթների $n(\mathcal{A})$ թվի հարաբերությանը տարրական պատահույթների ընդհանուր N թվին.
>
> $$\mathbf{P}\{\mathcal{A}\} = n(\mathcal{A})/N:$$

Դասական սահմանումը կիրառելիս հաճախ օգտվում են **համակցաբանության** (կոմբինատորիկայի) գաղափարներից և արդյունքներն արտահայտող բանաձևերից:

N հատ տարրեր պարունակող $\mathcal{X} = \{x_1, \ldots, x_n, \ldots, x_N\}$ բազմության տարրերով հնարավոր է կազմել տարբեր **համակցություններ** (կոմբինացիաներ): Սովորաբար հետաքրքրվում են դրանց քանակներով: Հաշվարկը հիմնվում է **համակցաբանության հիմնական՝ բազմապատկման, սկզբունքի** վրա:

> Եթե հարկավոր է կատարել K տեսակի գործողություններ, ընդ որում առաջին տեսակի գործողությունը հնարավոր է կատարել M_1 տարբեր ձևերով, այնուհետև երկրորդը՝ M_2 ձևերով, և այլն, վերջին՝ K-րդ տեսակի գործողությունը հնարավոր է կատարել M_K ձևերով, ապա K *տեսակի գործողությունների շարքը հնարավոր է կատարել*
>
> $$M = M_1 \cdot M_2 \cdot \ldots \cdot M_K$$
>
> *տարբեր ձևերով:*

Ապացույց: Եթե $K = 2$, ապա կարելի է բոլոր հնարավոր դեպքերի թիվը ստանալ՝ հաշվի առնելով, որ ամեն մի առաջին տեսակի գործողություն կարող է զուգորդվել երկրորդ տեսակի ցանկացած գործողությամբ: Ուրեմն գործողությունների զույգերի թիվը կլինի $M_1 \cdot M_2$: $K = 3$ դեպքում կարող ենք առաջին երկու գործողությունները դիտարկել որպես մեկ բարդ գործողություն (որոնց թիվն է $M_1 \cdot M_2$) և գալ նախորդ դեպքին: Այնուհետև կիրառում ենք մակածության (ինդուկցիայի) եղանակը:

N տարրեր պարունակող $\mathcal{X}$ բազմության K հատ տարրերով կազմենք կարգավորված $(x_{n_1}, x_{n_2}, \ldots, x_{n_K})$ համակցություններ (գրված վեկտորների ձևով)՝ թույլատրելով, որ տարրերը կրկնվեն, այլ կերպ ասած՝ $x_{n_k} \in \mathcal{X}$, $k = \overline{1, K}$: Քանի՞ տարբեր վեկտորներ կան: Պատասխանը՝

$$M = N^K,$$

ստացվում է բազմապատկման սկզբունքի կիրառումով, քանի որ բոլոր k-երի համար էլ ունենք ընտրության N տարբերակներ: Նշենք, որ N-ը և K-ն կարող են լինել կամայական ամբողջ դրական թվեր:

Այժմ նույն N տարրանոց $\mathcal{X}$ բազմությունից վերցնենք K հատ ($K \leq N$) **չկրկնվող** տարրերով կարգավորված համակցություն $(x_{n_1}, x_{n_2}, \ldots, x_{n_K})$, $x_{n_k} \neq x_{n_{k'}}$, եթե $k \neq k'$: Այդպիսի համակցությունը կոչվում է **կարգադրություն:**

N տարրերից K-ական ($K \leq N$) **կարգադրությունների թիվը** նշանակվում է $A_N^{K\dagger}$: Կիրառելով բազմապատկման սկզբունքը, ստանում ենք՝

$$A_N^K = N(N-1)\ldots(N-K+1):$$

Իսկապես, առաջին տարրը ընտրելիս կա N հնարավորություն, իսկ այնուհետև հաջորդ տարրերի հնարավոր տարբերակների թիվը ամեն անգամ մեկով պակասում է, քանի որ ընտրությունն արդեն կարելի է կատարել միայն մնացած տարրերից:

Տարադրությունը կոչվում է **տեղափոխություն,** եթե համակցությունում մասնակցում են $\mathcal{X}$-ի բոլոր տարրերը:

Երբ $K = N$, տարադրությունների թիվը համընկնում է N տարրերի **տեղափոխությունների** P_N թվի հետ՝

$$P_N = A_N^N = N(N-1)\ldots 1 = N!:$$

Դիտարկենք այժմ այն դեպքը, երբ $\mathcal{X}$ բազմությունից վերցված K տարրերի համակցության մեջ մեզ հետաքրքրում է մասնակցող տարրերի միայն ցուցակը, և կարևոր չէ նրանց կարգը: Այդպիսի համակցությունները կոչվում են **զուգորդություններ**: Դրանք կարելի է դիտարկել նաև որպես $\mathcal{X}$ բազմության K տարրանոց ենթաբազմություններ:

N տարրերից K-ական ($K \leq N$) **զուգորդությունների թիվը** նշանակում են $C_N^{K\ddagger}$: Այն հավասար է՝

$$C_N^K = A_N^K / P_K:$$

Տեղադրելով համապատասխան արժեքները կստանանք՝

$$C_N^K = \frac{N(N-1)\cdots(N-K+1)}{K!} = \frac{N!}{K!(N-K)!}, \quad C_N^K = C_N^{N-K}:$$

Դիտարկենք ավելի ընդհանուր դեպք, երբ բազմության տարրերը բաժանվում են $r \geq 2$ ենթաբազմությունների՝ առաջինում K_1 տարր, երկրորդում՝ K_2, $\ldots$, r-րդում՝ K_r $(K_1 + K_2 + \ldots + K_r = N)$: N տարրերից $K_1, K_2, \ldots, K_r$ տարրերով զուգորդությունների թիվը կլինի՝

$$M = \frac{N!}{K_1! K_2! \cdots K_r!}:$$

† Կարդացվում է՝ «ա էնից կաական»:

‡ Կարդացվում է՝ «ցե էնից կաական»:

Ապացուցումը կատարվում է մաթեմատիկական մակածության եղանակով և բացատրվում է հետևյալ հավասարության միջոցով՝
$$M = C_N^{K_1} C_{N-K_1}^{K_2} C_{N-K_1-K_2}^{K_3} \cdots C_{N-K_1-K_2-\ldots-K_{r-2}}^{K_{r-1}} :$$

Օրինակ 8: Դիցուք $\mathcal{X} = (x_1, x_2, x_3)$: Երեք տարրերից երկուական տարադրությունները վեց հատ են ($A_3^2 = 3 \times 2 = 6$), դրանք են՝ (x_1, x_2), (x_1, x_3), (x_2, x_3), (x_2, x_1), (x_3, x_1), (x_3, x_2) : Երեքից երկուական զուգորդությունները երեքն են ($C_3^2 = 3$). (x_1, x_2), (x_1, x_3), (x_2, x_3) : Երեք տարրերի տեղափոխությունները նույնպես 6 հատ են՝ (x_1, x_2, x_3), (x_1, x_3, x_2), (x_2, x_1, x_3), (x_2, x_3, x_1), (x_3, x_1, x_2), (x_3, x_2, x_1) :

Օրինակ 9: Հայաստանի քաղաքացու անձնագրի համարը կազմված է երկու լատինական տառերից և յոթ թվանշաններից: Քանի՞ տարբեր անձնագիր կարող է լինել, եթե օգտագործվեն բոլոր 26 լատինական տառերը և 0-ից մինչև 9-ը թվանշանները:

Լուծում: $N = 26^2 \cdot 10^7 = 676 \cdot 10^7$:

Օրինակ 10: *Դրոշների ցուցադրումը*:

ա) Դիցուք տարբեր գույների N հատ դրոշներ ցուցադրվում են K ձողերի վրա: Ենթադրենք, որ ձողերն այնքան երկար են, որ նույնիսկ բոլոր դրոշները հնարավոր է տեղավորել մեկ ձողի վրա՝ վերից վար: Ամեն մի դրոշ կարող է գտնվել կամայական ձողի վրա՝ մյուսների նկատմամբ կամայական դիրքում: Քանի՞ ձևով կարելի է տեղավորել դրոշները:

Լուծում: Ցույց տանք, որ $M = K(K+1)\cdots(K+N-1)$: Սկզբից ընտրենք առաջին դրոշի տեղը՝ ունենք K հնարավորություն: Երկրորդ դրոշի համար կան արդեն մեկով ավել տարբերակներ, քանի որ այն կարող է գտնվել $K-1$ ազատ ձողերի վրա, կամ արդեն տեղադրված դրոշից վեր, կամ վար: Այսպիսով, ամեն հաջորդ դրոշ ունի նախորդից մեկով ավել դիրքերի տարբերակներ:

բ) Լուծել նույն խնդիրը, երբ դրոշները միևնույն գույնի են:

Լուծում: Տեղադրումների թիվը կլինի նախորդ խնդրի տեղադրումների թվից $N!$ անգամ փոքր, քանի որ N դրոշների գույների տարբեր լինելը տալիս է $N!$ տեղափոխությունների հնարավորություն: Այսինքն, N միագույն դրոշները կարելի է տեղադրել K ձողերի վրա
$$M = K(K+1)\cdots(K+N-1)/N!$$
տարբեր ձևերով:

գ) Դիտարկենք մի այլ դեպք, երբ N_1 դրոշները կարմիր են, իսկ $N_2 = N - N_1$ դրոշները՝ սպիտակ:

Լուծում: Կրկնելով նախորդ խնդրի դատողությունները, կստանանք՝
$$M = K(K+1)\cdots(K+N-1)/N_1!N_2! :$$

Օրինակ 11: *Կանոնավոր մետաղադրամի* մեկ նետման փորձի ելքերն են՝ ω_1 (զինանշան) և ω_2 (թիվ): Կառուցենք հավանականային տարածության Ω, $\mathcal{F}$, $\mathbf{P}$ բաղադրիչները: Տարրական պատահույթների տարածությունն ընդգրկում է ընդամենը երկու տարր՝ $\Omega = \{\omega_1, \omega_2\}$, σ-հանրահաշիվը կազմված է չորս տարրից՝ $\mathcal{F} = \{\varnothing, \{\omega_1\}, \{\omega_2\}, \Omega\}$: Քանի որ մետաղադրամը կանոնավոր է և, հետևաբար, երկու կողմերը հավասարահավանական են, ապա $\mathbf{P}$ հավանականությունը $\mathcal{F}$-ի տարրերի վրա ընդունում է հետևյալ արժեքները՝ $\mathbf{P}\{\varnothing\} = 0$, $\mathbf{P}\{\omega_1\} = \mathbf{P}\{\omega_2\} = 1/2$, $\mathbf{P}\{\Omega\} = 1$:

Օրինակ 12: *Զառի նետման փորձը*: Տարրական ելքերը հավասարահավանական են՝ $\mathbf{P}\{\omega_n\} = 1/6$, $n = \overline{1,6}$, իսկ, օրինակ, «զույգ թվի հանդես գալը» պատահույթի հավանականությունը ըստ դասական սահմանման կլինի $\mathbf{P}\{\mathcal{A}\} = 3/6 = 1/2$, քանի որ $\mathcal{A} = (\omega_2, \omega_4, \omega_6)$, $n(\mathcal{A}) = 3$:

Օրինակ 13: Ութ կառավարիչներից և երկու հաշվապահներից հարկավոր է պատահական ձևով կազմել չորս հոգանոց հանձնաժողով: Գտնել հանձնաժողովը միայն կառավարիչներով համալրելու հավանականությունը:

Լուծում : $N = C_{10}^4$, $n(\mathcal{A}) = C_8^4$, $\mathbf{P}\{\mathcal{A}\} = \dfrac{C_8^4}{C_{10}^4} = \dfrac{8 \times 7 \times 6 \times 5}{10 \times 9 \times 8 \times 7} = 1/3$:

<u>*Օրինակ*</u> 14: Քարտերի վրա գրված են 1, 2, 3, 4, 5 թվերը: Պատահականորեն վերցվում են հաջորդաբար 3 քարտ, որոնք շարվում են ձախից աջ: Գտնել այդ ձևով ստացված եռանիշ թվի զույգ լինելու հավանականությունը:

Լուծում : $N = A_5^3$, $n(\mathcal{A}) = 2A_4^2$, $\mathbf{P}\{\mathcal{A}\} = \dfrac{2A_4^2}{A_5^3} = \dfrac{2 \times 4 \times 3}{5 \times 4 \times 3} = 2/5$:

Դիտողություն. Այս օրինակը կարելի է նաև լուծել՝ պարզապես հաշվի առնելով, որ թվի զույգ լինելու պարազան կախված է միայն վերջին նիշից, որի զույգ լինելուն հավասարահնարավոր հինգից նպաստում են երկու դեպքը, ուրեմն $\mathbf{P}\{\mathcal{A}\} = 2/5$:

Այժմ դիտարկենք մի այլ հավանականային տարածությունների դաս, որում հավանականությունը որոշվում է «երկրաչափական սահմանման» միջոցով:

Դիցուք Ω-ն n-չափանի վեկտորական տարածության՝ $\mathcal{R}^n$-ի, սահմանափակ բազմություն է, որի լեբեգյան չափն է $\mathrm{mes}\Omega$ (միաչափ դեպքում՝ երկարությունը, $\mathcal{R}^2$-ում՝ մակերեսը, $\mathcal{R}^3$-ում՝ ծավալը), իսկ $\mathcal{F}$-ը Ω-ի այն $\mathcal{A}$ ենթաբազմությունների դասն է, որոնք ունեն լեբեգյան չափ՝ $\mathrm{mes}\mathcal{A}$: Այդ դեպքում $\mathcal{A}$-ի հավանականությունը՝ $\mathbf{P}\{\mathcal{A}\}$-ն, ըստ **երկրաչափական սահմանման** հավասար է.

$$\mathbf{P}\{\mathcal{A}\} = \frac{\mathrm{mes}\mathcal{A}}{\mathrm{mes}\Omega}:$$

Դիտարկենք երկու դասական օրինակներ:

<u>*Օրինակ*</u> 15: *Հանդիպման խնդիրը*: Երկու ընկեր պայմանավորվել են հանդիպել որոշակի վայրում՝ ժամը 12-ի և 13-ի միջև: Առաջին եկողը մյուսին սպասում է 20 րոպե, որից հետո հեռանում է: Գտնել նրանց իրար հանդիպելու հավանականությունը, եթե նրանցից ամեն մեկի գալու պահը պատահական է այդ 60 րոպեի ընթացքում և անկախ է մյուսի գալու պահից:

Լուծում: Նշանակենք ընկերներից մեկի գալու պահը x-ով, իսկ մյուսինը՝ y-ով: Ներկայացնելով x-ը և y-ը հարթության համապատասխան առանցքների վրա՝ կստանանք, որ նրանց գալու պահերի (x, y) կետերը լրացնում են 60 միավոր կողմ ունեցող քառակուսին՝ $\Omega = \{(x,y) : 0 \le x \le 60,\ \ 0 \le y \le 60\}$: Որպեսզի նրանք հանդիպեն, անհրաժեշտ է և բավարար, որ նրանց ժամանելու պահերի տարբերությունը մեծ չլինի 20-ից՝ $|x - y| \le 20$ (տե՛ս նկար 6): Նշանակենք հանդիպմանը նպաստող (x, y) տարրական պատահույթների բազմությունը $\mathcal{A}$-ով (նկարում նշված է մուգ գույնով), $\mathcal{A} = \{(x,y) : |x - y| \le 20\}$: Ըստ երկրաչափական սահմանման

$$\mathbf{P}\{\mathcal{A}\} = \frac{\mathrm{mes}\mathcal{A}}{\mathrm{mes}\Omega} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}:$$

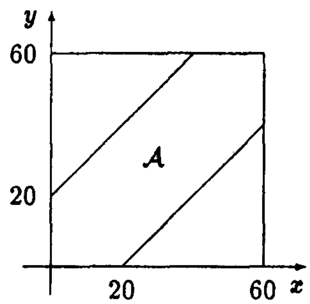
<u>*Օրինակ*</u> 16: *Բյուֆոնի խնդիրը*: Հարթության վրա գծված են զուգահեռ ուղիղներ՝ իրարից հավասար $2a$ հեռավորության վրա: Պատահականորեն հարթության վրա նետվում է $2l$ $(l < a)$ երկարության ասեղ: Հարկավոր է գտնել որևէ ուղղի հետ ասեղի հատվելու հավանականությունը:

Լուծում: Նշանակենք x-ով ասեղի կենտրոնի և մոտակա ուղղի հեռավորությունը, իսկ φ-ով՝ ասեղի և այդ ուղղի կազմած անկյունը : x-ը և φ-ն որոշում են ասեղի դիրքը մոտակա ուղղի նկատմամբ: Ասեղի բոլոր հնարավոր դիրքերին համապատասխանող (φ, x) զույգերը լրացնում են մի ուղղանկյուն.

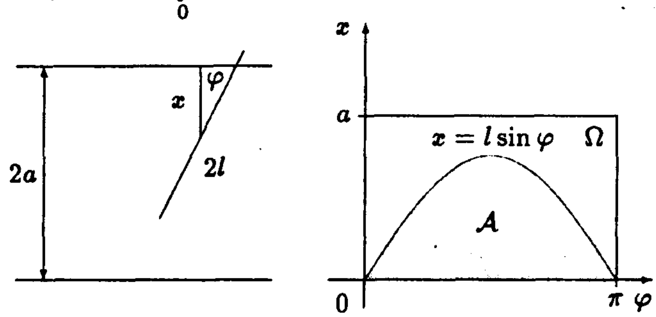
$$\Omega = \{(\varphi, x) : 0 \le \varphi \le \pi,\ \ 0 \le x \le a\}:$$

Նկարից երևում է. որպեսզի ասեղը հատվի որևէ ուղղի հետ, անհրաժեշտ է և բավարար, որ $0 \le x \le l\sin\varphi$: Այսինքն, $\mathcal{A} = \{(\varphi, x) : 0 \le x \le l\sin\varphi\}$: Երկրաչափական սահմանման համաձայն ստանում ենք՝

$$\mathbf{P}\{\mathcal{A}\} = \frac{mes\mathcal{A}}{mes\Omega} = \frac{1}{a\pi}\int_0^{\pi} l\sin\varphi d\varphi = \frac{2l}{a\pi}:$$



Նկար 6: Հանդիպման խնդիրը:



Նկար 7: Բյուֆոնի խնդիրը:

Հավանականության գաղափարը խորը տեսական և կիրառական իմաստ է ստանում հետևյալ կարևորագույն, փորձնականորեն բազմիցս ստուգված օրինաչափության հիման վրա:

Դիցուք իրարից անկախ փորձերը կրկնվում են միևնույն պայմաններում այնպես, որ մեզ հետաքրքրող $\mathcal{A}$ պատահույթի հանդես գալու հնարավորությունը ամեն մի փորձում նույնն է: Նշանակենք N-ով փորձերի թիվը, իսկ $n_N(\mathcal{A})$-ով՝ այն փորձերի թիվը, որոնցում $\mathcal{A}$-ն տեղի է ունեցել: $n_N(\mathcal{A})$-ի հարաբերությունը N-ին անվանում են $\mathcal{A}$-ի **հարաբերական հաճախություն**: Եթե փորձերի թիվը բավականին մեծ է, ապա կարելի է նկատել, որ N փորձերի հաջորդականությունը բազմիցս կրկնելիս $n_N(\mathcal{A})/N$ հարաբերական հաճախությունը ցուցաբերում է կայունություն, «քիչ» է տարբերվում որոշ հաստատուն $\mathbf{P}\{\mathcal{A}\}$ արժեքից՝

$$\mathbf{P}\{\mathcal{A}\} \sim \frac{n_N(\mathcal{A})}{N}:$$

Հետևաբար, բնական է ընդունել $n_N(\mathcal{A})/N$ հարաբերությունը որպես $\mathbf{P}(\mathcal{A})$ հավանականության մոտավոր արժեք: Այս հանգամանքը կարելի է արտահայտել նաև սահմանի գաղափարի օգնությամբ (որը, իհարկե, ճշգրտման կարիք ունի)՝

$$\mathbf{P}\{\mathcal{A}\} = \lim_{N\to\infty} \frac{n_N(\mathcal{A})}{N}: \qquad (1)$$

Այդ օրինաչափությունը ստուգվել է բազմաթիվ փորձարկումների իրականացման արդյունքում: Օրինակ, բազմիցս համոզվել են, որ մետաղադրամի նետման փորձերի թվի աճի հետ զինանշանի հաճախությունն ավելի ու ավելի է մոտենում 1/2-ին:

$n_N(\mathcal{A})/N$ հարաբերությունը բավականաչափ մեծ N-երի համար կարելի է ընդունել որպես $\mathcal{A}$ պատահույթի հավանականության՝ $\mathbf{P}(\mathcal{A})$-ի, մոտավոր արժեք: **Հավանականության որոշման այդ մոտեցումը** կոչվում է **վիճակագրական**:

Օրինակ, եթե արհեստավարժ բասկետբոլիստի 100 տուգանային նետումներից 65-ում գնդակն ընկնում է զամբյուղը, մենք ասում ենք, որ նա գրավում է զամբյուղը տուգանային նետումների ժամանակ $\mathbf{P}\{\mathcal{A}\} = 65/100 = 13/20$ հավանականությամբ:

Իհարկե, անհայտ հավանականության որոշման վիճակագրական մոտեցումը կիրառելի է առավել լայն ոլորտի իրադրություններում, սակայն դրա իրագործումը կապված է ծավալուն գործողությունների հետ և ոչ միշտ է հնարավոր:

Հնարավոր չէ խիստ մաթեմատիկորեն հիմնավորել՝ կիրառելի՞ է, թե՞ ոչ վիճակագրական սահմանումը տվյալ իրավիճակում: Որոշ դեպքերում դա պարզ է, օրինակ, կանոնավոր մետաղադրամի, կամ զառի նետումները, անփոփոխ պայմաններում զանգվածային արտադրության արտադրանքի ստուգումը և այլն: Այլ իրավիճակներում դասական, երկրաչափական և վիճակագրական սահմանումները կարելի է օգտագործել որոշ վերապահումներով, մոտավորապես:

Երբեմն հավանականությունը պետք է գնահատել այնպիսի իրադրությունում, որտեղ նշված դասական, վիճակագրական, երկրաչափական մոտեցումներն իրագործել հնարավոր չէ: Դա կարող է լինել, մասնավորապես, երբ նախնական փորձարկումների տվյալները բացակայում են: Այդ դեպքերում միակ հնարավորությունն է գնահատել հավանականությունը առողջ դատողությամբ կամ փորձագետների կարծիքների համադրումով: Դա, այսպես կոչված, **սուբյեկտիվ հավանականությունների** մոտեցումն է, ըստ որի որևէ պատահույթին վերագրվում է հավանականություն՝ ելնելով մեր ունեցած փաստերից և տեղեկություններից: Որոշ դեպքերում սուբյեկտիվ մոտեցումն օգտագործվում է, երբ ուզում ենք գնահատել այնպիսի մի պատահույթի հավանականությունը, որը երբեք դեռ չի իրագործվել: Օրինակ, Հայաստանի Հանրապետությունում կին նախագահ ընտրվելու հավանականությունը: Այդպիսի դեպքեր դեռ չեն եղել, և այդ հավանականությունը որոշելու համար մենք կարող ենք միայն հենվել կարծիքների համադրումից ստացվող գնահատականների վրա:

1.3. Հավանականության հատկությունները

Նախորդ ենթաբաժնում ձևակերպված աքսիոմներից հետևում են հավանականության հետևյալ **հատկությունները**՝

1. $\mathrm{P}\{\varnothing\} = 0$:
2. $\mathrm{P}\{A'\} = 1 - \mathrm{P}\{A\}$:
3. Եթե $A \subset B$, ապա $\mathrm{P}\{A\} \le \mathrm{P}\{B\}$:
4. $0 \le \mathrm{P}\{A\} \le 1$:
5. «**Գումարման թեորեմը**»: Կամայական A և B պատահույթների համար
$$\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\} - \mathrm{P}\{AB\} : \qquad (2)$$
6. **Անընդհատության հատկությունը**: Եթե $A_1, A_2, \ldots, A_n, \ldots$ հաջորդականության պատահույթները այնպիսին են, որ ամեն հաջորդը նախորդի մասնավոր դեպքն է՝ $(A_1 \supset A_2 \supset A_3 \supset \cdots)$ և $A = \bigcap_{n=1}^{\infty} A_n$, ապա $\mathrm{P}\{A\} = \lim_{n\to\infty} \mathrm{P}\{A_n\}$:

Տանք այս հատկությունների մի մասի կրճատ ապացուցումները:

2. $\Omega = A \cup A'$, ուրեմն $1 = \mathrm{P}\{\Omega\} = \mathrm{P}\{A \cup A'\} = \mathrm{P}\{A\} + \mathrm{P}\{A'\}$, որտեղից $\mathrm{P}\{A'\} = 1 - \mathrm{P}\{A\}$:

3. Ներկայացնելով այսպես՝ $B = A \cup (B - A)$, և հաշվի առնելով, որ A-ն և $(B-A)$-ն անհամատեղելի են, կստանանք $\mathrm{P}\{B\} = \mathrm{P}\{A\} + \mathrm{P}\{B - A\}$, և քանի որ $\mathrm{P}\{B - A\} \ge 0$, ապա $\mathrm{P}\{A\} \le \mathrm{P}\{B\}$:

4. Հատկությունը հետևում է նախորդ հատկությունից, քանի որ $\varnothing \subset A \subset \Omega$:

5. Հատկությունն ապացուցելու համար ներկայացնենք՝ $A \cup B = A \cup (B - A)$, $B = (AB) \cup (B - A)$, որտեղ աջում անհամատեղելի պատահույթների միավորումներ են, ուրեմն ըստ ադիտիվության աքսիոմի՝ $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B - A\}$ և $\mathrm{P}\{B\} = \mathrm{P}\{AB\} + \mathrm{P}\{B - A\}$: Արտահայտենք $\mathrm{P}\{B - A\}$-ն երկրորդ հավասարությունից և տեղադրենք նախորդ հավասարության մեջ, կստանանք (2)-ը:

6. Հատկությունը հետևում է երրորդ (գումարականության) աքսիոմից:

Նույն ձևով ապացուցվում է, որ եթե $B_1, B_2, \ldots, B_n, \ldots$ հաջորդականությունը աճող է՝ $B_1 \subset B_2 \subset B_3 \subset \cdots$ և $B = \bigcup_{n=1}^{\infty} B_n$, ապա $\mathbf{P}\{B\} = \lim_{n\to\infty} \mathbf{P}\{B_n\}$:

Գումարման թեորեմից մաթեմատիկական մակածության օգնությամբ ստացվում է, որ ցանկացած $A_1, A_2, \ldots . A_N$ պատահույթների համար՝

$$\mathbf{P}\{\bigcup_{n=1}^{N} A_n\} = \sum_{n=1}^{N} \mathbf{P}\{A_n\} - \sum_{n_1=1}^{N-1} \sum_{n_2=n_1+1}^{N} \mathbf{P}\{A_{n_1}A_{n_2}\} +$$

$$+ \sum_{n_1=1}^{N-2} \sum_{n_2=n_1+1}^{N-1} \sum_{n_3=n_2+1}^{N} \mathbf{P}\{A_{n_1}A_{n_2}A_{n_3}\} + \ldots + (-1)^{N-1}\mathbf{P}\{A_1A_2\cdots A_N\} : \qquad (3)$$

Մասնավորապես, $N = 3$ դեպքում՝

$$\mathbf{P}\{A_1 \cup A_2 \cup A_3\} = \sum_{i=1}^{3} \mathbf{P}\{A_i\} - \sum_{i=1}^{2} \sum_{i<j\leq 3} \mathbf{P}\{A_iA_j\} + \mathbf{P}\{A_1A_2A_3\}:$$

Օրինակ 17: *Համընկնման խնդիրը*: Դիցուք ռազմական ուսումնարանի մի խումբ սովորողներ թողել են իրենց գրատները (շինելները) հանդերձարանի կախիչների վրա: Պարապմունքից հետո նրանք շտապում են վերցնել մեկական պատահական գրատ: Գտնել գոնե մեկ գրատն իր իսկական տիրոջն ընկնելու հավանականությունը:

Լուծում: Համարակալենք գրատները $1, 2, \ldots, N$: Նույն համարները տանք համապատասխան սովորողներին: Նշանակենք A_n-ով n-րդ գրատը իր տիրոջն ընկնելու պատահույթը: Պետք է գտնենք $A = \bigcup_n A_n$ պատահույթի հավանականությունը: Օգտվենք (3) բանաձևից: Գրատների պատահական բաշխման տարրական ելքը նկարագրվում է $(k_1, k_2, \ldots, k_N)$ թվերի տեղափոխությամբ, որտեղ k_n-ը n-րդ սովորողին ընկած գրատի համարն է: Բոլոր տեղափոխությունների թիվը $N!$ է: Եթե մեկ գրատն ընկել է իր տիրոջը, օրինակ՝ n-րդը, ապա $k_n = n$, և հնարավոր տեղափոխությունների թիվը կլինի $(N-1)!$: Ուրեմն՝

$$\mathbf{P}\{A_n\} = (N-1)!/N! = 1/N:$$

Եթե համընկել են երկու՝ n_1-րդ և n_2-րդ սովորողների գրատները, ապա մնում է $(N-2)!$ հնարավոր տեղափոխություններ, և

$$\mathbf{P}\{A_{n_1}A_{n_2}\} = (N-2)!/N!:$$

Նման ձևով՝

$$\mathbf{P}\{A_{n_1}A_{n_2}\cdots A_{n_m}\} = (N-m)!/N!:$$

Հիշելով, որ N համարներից երկու հատը կարելի է վերցնել C_N^2 տարբերակներով, ստանում ենք՝

$$\sum_{i_1=1}^{N-1} \sum_{i_2=i_1}^{N} \mathbf{P}\{A_{n_1}A_{n_2}\} = C_N^2 \frac{(N-2)!}{N!} = 1/2! :$$

Նմանապես՝

$$\sum_{n_1=1}^{N-2} \sum_{n_2=n_1+1}^{N-1} \sum_{n_3=n_2+1}^{N} \mathbf{P}\{A_{n_1}A_{n_2}A_{n_3}\} = C_N^3 \frac{(N-3)!}{N!} = 1/3!,$$

և այլն: Տեղադրելով բանաձևի մեջ՝ կստանանք՝ $\mathbf{P}\{A\} = 1 - 1/2! + 1/3! - \cdots (-1)^{N-1}1/N!$:

Աչից ստացվել է $1 - e^x$ ֆունկցիայի $x = -1$ կետում շարքի մասնակի գումարը, ուրեմն բավականաչափ մեծ N-երի համար կարող ենք $\mathbf{P}\{A\}$ հավանականությունը համարել մոտավորապես հավասար $1 - 1/e \approx 0.64$:

1.4. Պայմանական հավանականություն, պատահույթների անկախություն

Վերհիշենք կանոնավոր զառի նետման օրինակը: Դիցուք A պատահույթը զառի վրա 1 միավորի ստացվելն է՝ $A = \{\omega_1\}$: Դասական սահմանման համաձայն $\mathbf{P}\{A\} = 1/6$:

Այժմ ենթադրենք՝ հայտնի դարձավ, որ ստացված թիվը կենտ է, այսինքն՝ տեղի է ունեցել $B = \{\omega_1, \omega_3, \omega_5\}$ պատահույթը: Այդ դեպքում A-ին կնպաստի մեկ ելք հնարավոր երեք ելքերից, և $\mathcal{A}$-ի հավանականությունը կդառնա հավասար $1/3$: Նշանակենք այդպիսի **պայմանական հավանականությունը** $\mathrm{P}\{\mathcal{A}|\mathcal{B}\}$: Այս օրինակում $\mathrm{P}\{\mathcal{A}|\mathcal{B}\} = 1/3$, իսկ $\mathrm{P}\{\mathcal{A}|\mathcal{B}'\} = 0$:

Դիտարկենք ավելի ընդհանուր օրինակ: Դիցուք Ω-ն կազմված է N հավասարահնարավոր տարրական պատահույթներից: $\mathcal{A}$ պատահույթին նպաստում են r ելքեր, $\mathcal{B}$ պատահույթին՝ m ելքեր, իսկ $\mathcal{AB}$ պատահույթին՝ k ելքեր: Նախորդ օրինակի նման, համաձայն դասական սահմանման՝

$$\mathrm{P}\{\mathcal{A}|\mathcal{B}\} = \frac{k}{m} = \frac{k/N}{m/N} = \frac{\mathrm{P}\{\mathcal{AB}\}}{\mathrm{P}\{\mathcal{B}\}}, \text{ իսկ } \mathrm{P}\{\mathcal{B}|\mathcal{A}\} = \frac{k}{r} = \frac{k/N}{r/N} = \frac{\mathrm{P}\{\mathcal{AB}\}}{\mathrm{P}\{\mathcal{A}\}} :$$

Կարող է օգտակար լինել նաև գծապատկերներով բացատրությունը: Դիտարկենք նկար 8-ի Ω տարրական պատահույթների տարածությունը, $\mathcal{A}$, $\mathcal{B}$ և $\mathcal{AB}$ պատահույթները: Նրանց հավանականությունները կլինեն՝

$$\mathrm{P}\{\mathcal{A}\} = \frac{mes\mathcal{A}}{mes\Omega}, \quad \mathrm{P}\{\mathcal{B}\} = \frac{mes\mathcal{B}}{mes\Omega}, \quad \mathrm{P}\{\mathcal{AB}\} = \frac{mes(\mathcal{AB})}{mes\Omega}:$$

Եթե տեղի է ունեցել $\mathcal{B}$ պատահույթը, մենք կարող ենք անտեսել $\mathcal{B}$-ից դուրս գտնվող տարրական պատահույթները և դիտարկել $\mathcal{B}$-ն, որպես **տարրական պատահույթների նոր տարածություն** Ω^*:

Պայմանական հավանականությունները կլինեն՝ $\mathrm{P}\{\mathcal{B}|\mathcal{B}\} = \frac{mes\mathcal{B}}{mes\Omega^*} = \frac{mes\mathcal{B}}{mes\mathcal{B}} = 1,$

$$\mathrm{P}\{\mathcal{A}|\mathcal{B}\} = \frac{mes(\mathcal{AB})}{mes\Omega^*} = \frac{mes(\mathcal{AB})}{mes\Omega} : \frac{mes\mathcal{B}}{mes\Omega} = \frac{\mathrm{P}\{\mathcal{AB}\}}{\mathrm{P}\{\mathcal{B}\}} = \frac{\mathrm{P}\{\mathcal{A}^*\}}{\mathrm{P}\{\mathcal{B}\}} :$$

Նկար 8: Պայմանական հավանականային տարածություն:

Այժմ անցնենք ընդհանուր սահմանմանը:

> Դիցուք տրված է $\{\Omega, \mathcal{F}, \mathrm{P}\}$ հավանականային տարածությունը, և դիցուք $\mathcal{A}$-ն և $\mathcal{B}$-ն կամայական պատահույթներ են: Եթե $\mathrm{P}\{\mathcal{B}\} > 0$, ապա $\mathcal{A}$-ի **պայմանական հավանականությունը** $\mathcal{B}$-ի տեղի ունենալու պայմանով, ըստ սահմանման, ընդունվում է հավասար $\mathrm{P}\{\mathcal{AB}\}$-ի և $\mathrm{P}\{\mathcal{B}\}$-ի հարաբերությանը՝
>
> $$\mathrm{P}\{\mathcal{A}|\mathcal{B}\} = \frac{\mathrm{P}\{\mathcal{AB}\}}{\mathrm{P}\{\mathcal{B}\}} :$$

Պայմանական հավանականությունը՝ $\mathrm{P}\{\mathcal{A}|\mathcal{B}\}$-ն, կարդացվում է այսպես. «$\mathcal{A}$-ի ըստ $\mathcal{B}$-ի պայմանական հավանականությունը»:

> Պայմանական հավանականության սահմանումից հետևում է **բազմապատկման կանոնը**, այն է՝
>
> $$\mathrm{P}\{\mathcal{AB}\} = \mathrm{P}\{\mathcal{A}\}\mathrm{P}\{\mathcal{B}|\mathcal{A}\} = \mathrm{P}\{\mathcal{B}\}\mathrm{P}\{\mathcal{A}|\mathcal{B}\}, \qquad (4)$$
>
> այսինքն, երկու պատահույթների համատեղ հավանականությունը հավասար է նրանցից մեկի հավանականության և մյուսի՝ ըստ առաջինի պայմանական հավանականության արտադրյալին: (4) նույնությունը ճիշտ է նաև, եթե $\mathrm{P}\{\mathcal{AB}\} = 0$:

Օրինակ 18: Ամսագրի գովազդային բաժնի ղեկավարը ամսագրում զետեղված մի որոշակի գովազդը բաժանորդի կողմից կարդալու հավանականությունը գնահատում է հավասար 0.4-ի, իսկ դրանից հետո գովազդված իրը գնվելու հավանականությունը՝ հավասար 0.1-ի: Գտնենք ընթերցողի կողմից գովազդը կարդալու և գովազդված իրը գնելու հավանականությունը:

Լուծում: Նշանակենք $\mathcal{A}$ = {ընթերցողը կկարդա գովազդը}, $\mathcal{B}$ = {կգնի ապրանքը}:

$$\mathbf{P}\{\mathcal{AB}\} = \mathbf{P}\{\mathcal{A}\}\mathbf{P}\{\mathcal{B}|\mathcal{A}\} = 0.4 \times 0.1 = 0.04:$$

$\mathcal{A}$ և $\mathcal{B}$ պատահույթները կոչվում են **անկախ**, եթե

$$\mathbf{P}\{\mathcal{AB}\} = \mathbf{P}\{\mathcal{A}\}\mathbf{P}\{\mathcal{B}\}:$$

Անկախ պատահույթների համար՝ $\mathbf{P}\{\mathcal{B}|\mathcal{A}\} = \mathbf{P}\{\mathcal{B}\}$ և $\mathbf{P}\{\mathcal{A}|\mathcal{B}\} = \mathbf{P}\{\mathcal{A}\}$:

Այսինքն՝ անկախությունը նշանակում է, որ պատահույթներից մեկի տեղի ունենալը չի փոխում մյուսի հավանականությունը:

$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ պատահույթները կոչվում են **փոխադարձաբար (խմբովին) անկախ**, եթե կամայական $k \le N$ և $1 \le n_1 < n_2 < \cdots < n_k \le N$ համար

$$\mathbf{P}\{\mathcal{A}_{n_1}\mathcal{A}_{n_2}\ldots\mathcal{A}_{n_k}\} = \mathbf{P}\{\mathcal{A}_{n_1}\}\mathbf{P}\{\mathcal{A}_{n_2}\}\ldots\mathbf{P}\{\mathcal{A}_{n_k}\}:$$

$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ պատահույթները կոչվում են **զույգ առ զույգ անկախ**, եթե ցանկացած $1 \le n_1 < n_2 \le N$ համար

$$\mathbf{P}\{\mathcal{A}_{n_1}\mathcal{A}_{n_2}\} = \mathbf{P}\{\mathcal{A}_{n_1}\}\mathbf{P}\{\mathcal{A}_{n_2}\}:$$

Զույգ առ զույգ անկախությունից չի հետևում խմբովին անկախությունը, այսինքն՝ հնարավոր է, որ պատահույթները լինեն զույգ առ զույգ անկախ, բայց խմբովին անկախ չլինեն: Դա ցույց է տալիս

Օրինակ 19: *Բեռնշտեյնի օրինակը*: Հարթության վրա նետվում է կանոնավոր քառանիստը, որի երեք նիստերը ներկված են համապատասխանաբար կարմիր, կապույտ և կանաչ, իսկ չորրորդը՝ բոլոր երեք գույներով: Նշանակենք $\mathcal{A}_1$-ով քառանիստի ներքևի նիստի վրա կարմիր գույնի առկայության պատահույթը, և $\mathcal{A}_2$-ով, $\mathcal{A}_3$-ով՝ համապատասխանաբար կապույտ և կանաչ գույների առկայության պատահույթները: Համոզվենք, որ $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ պատահույթները զույգ առ զույգ անկախ են, սակայն խմբովին անկախ չեն:

Լուծում: Իսկապես՝ $\mathbf{P}\{\mathcal{A}_1\} = \mathbf{P}\{\mathcal{A}_2\} = \mathbf{P}\{\mathcal{A}_3\} = 1/2$, $\mathbf{P}\{\mathcal{A}_1\mathcal{A}_2\} = 1/4 = \mathbf{P}\{\mathcal{A}_1\}\mathbf{P}\{\mathcal{A}_2\}$,

$$\mathbf{P}\{\mathcal{A}_1\mathcal{A}_3\} = \mathbf{P}\{\mathcal{A}_1\}\mathbf{P}\{\mathcal{A}_3\}, \quad \mathbf{P}\{\mathcal{A}_2\mathcal{A}_3\} = \mathbf{P}\{\mathcal{A}_2\}\mathbf{P}\{\mathcal{A}_3\}:$$

Սակայն՝ $\mathbf{P}\{\mathcal{A}_1\mathcal{A}_2\mathcal{A}_3\} = 1/4 \ne \mathbf{P}\{\mathcal{A}_1\}\mathbf{P}\{\mathcal{A}_2\}\mathbf{P}\{\mathcal{A}_3\} = 1/8$:

Ընդհանուր դեպքում կամայական $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ պատահույթների համար բազմապատկման կանոնը հետևյալն է՝

$$\mathbf{P}\{\mathcal{A}_1\mathcal{A}_2\cdots\mathcal{A}_N\} = \mathbf{P}\{\mathcal{A}_1\}\mathbf{P}\{\mathcal{A}_2|\mathcal{A}_1\}\mathbf{P}\{\mathcal{A}_3|\mathcal{A}_1\mathcal{A}_2\}\cdots\mathbf{P}\{\mathcal{A}_N|\mathcal{A}_1\cdots\mathcal{A}_{N-1}\}:$$

Նկատենք, որ աջից պատահույթների հերթականությունը կարելի է փոխել, այսինքն՝ մեկի հավանականությունը պետք է բազմապատկվի մեկ ուրիշի ըստ նախորդի պայմանական հավանականությամբ, այնուհետև բազմապատկվի երրորդի՝ ըստ նախորդ երկուսի պայմանական հավանականությամբ և այլն:

Օրինակ 20: Սափորում կան a սպիտակ և b սև գնդակներ ($a \ge 1$, $b \ge 1$): Սափորից պատահականորեն առանց վերադարձի վերցվում է երկու գնդակ: Գտնել երկրորդ գնդակի սպիտակ լինելու հավանականությունը, եթե առաջին գնդակը սպիտակ է եղել:

Լուծում: Առաջին գնդակի սպիտակ լինելու պատահույթը նշանակենք A_1-ով, իսկ երկրորդ գնդակինը՝ A_2-ով: Պարզ է, որ $\mathbf{P}\{A_2|A_1\} = (a-1)/(a+b-1)$: Քանի որ

$$\mathbf{P}\{A_1A_2\} = a(a-1)/(a+b)(a+b-1), \quad \mathbf{P}\{A_1\} = a/(a+b),$$

դժվար չէ տեսնել, որ (4)-ը տեղի ունի:

Օրինակ 21: Հաջորդաբար նետվում են երկու կանոնավոր զառեր: Նշանակենք A_1-ով առաջին զառի վրա հայտնված թվի զույգ լինելու պատահույթը, A_2-ով՝ երկրորդ զառի թիվը 3-ին պատիկ լինելու պատահույթը, A_3-ով՝ առաջին և երկրորդ զառերի վրա հայտնված թվերի գումարը զույգ լինելու պատահույթը: Ցույց տալ, որ A_1, A_2, A_3 պատահույթները խմբովին անկախ են:

Լուծում: Երկու զառի նետման 36 դեպքերը կարելի է ներկայացնել հետևյալ ձևով՝ $(1,1)$, $(1,2)$,..., $(1,6)$,...,$(6,5)$, $(6,6)$: A_1 պատահույթին նպաստում են դրանցից 18-ը: A_2 պատահույթին՝ 12-ը: A_3 պատահույթին՝ $(1,1)$, $(1,3)$,..., $(6,4)$, $(6,6)$ դեպքերը, որոնց թիվը հավասար է 18-ի: Ուրեմն՝

$$\mathbf{P}\{A_1\} = 18/36 = 1/2, \quad \mathbf{P}\{A_2\} = 12/36 = 1/3, \quad \mathbf{P}\{A_3\} = 18/36 = 1/2 :$$

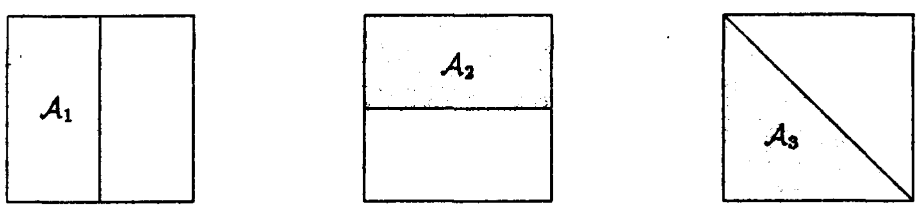
A_1 և A_2 պատահույթների անկախությունն ստուգենք՝ հաշվելով հավանականությունները: A_1A_2 պատահույթը տեղի ունի հետևյալ 6 դեպքերում՝ $(2,3)$, $(2,6)$, $(4,3)$, $(4,6)$, $(6,3)$, $(6,6)$: Տեսնում ենք, որ $\mathbf{P}\{A_1A_2\} = 6/36 = 1/6 = \mathbf{P}\{A_1\}\mathbf{P}\{A_2\}$, ուրեմն A_1-ը և A_2-ը անկախ են: Համոզվենք, որ անկախ են նաև A_1-ը և A_3-ը: Նրանք միաժամանակ տեղի ունեն հետևյալ 9 դեպքում՝ $(2,2)$, $(2,4)$, $(2,6)$, $(4,2)$, $(4,4)$, $(4,6)$, $(6,2)$, $(6,4)$, $(6,6)$: Հետևաբար,

$$\mathbf{P}\{A_1A_3\} = 9/36 = 1/4 = \mathbf{P}\{A_1\}\mathbf{P}\{A_3\} :$$

A_2A_3 պատահույթին նպաստում է հետևյալ 6 դեպքը՝ $(1,3)$, $(2,6)$, $(3,3)$, $(4,6)$, $(5,3)$, $(6,6)$, որտեղից՝ $\mathbf{P}\{A_2A_3\} = 6/36 = 1/6 = \mathbf{P}\{A_2\}\mathbf{P}\{A_3\}$: Վերջապես, $A_1A_2A_3$ պատահույթին նպաստում է միայն 3 դեպք՝ $(2,6)$, $(4,6)$, $(6,6)$, հետևաբար, քանի որ $3/36 = 1/12 = (1/2) \times (1/3) \times (1/2)$, ապա՝ $\mathbf{P}\{A_1A_2A_3\} = \mathbf{P}\{A_1\}\mathbf{P}\{A_2\}\mathbf{P}\{A_3\}$: Այսպիսով, A_1, A_2, A_3 պատահույթները խմբովին անկախ են:

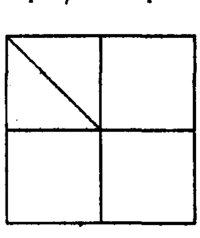
Օրինակ 22: Կառուցենք օրինակ, որից երևում է, որ վերջին հավասարության տեղի ունենալուց չի հետևում A_1, A_2, A_3 պատահույթների զույգ առ զույգ անկախությունը:

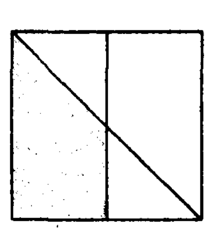
Լուծում: Դիտարկենք նկար 9-ում պատկերված պատահույթները: A_1-ը՝ «քառակուսու վրա պատահականորեն նետված կետը գտնվում է նրա ձախ կեսում»:



Նկար 9: A_1, A_2 և A_3 պատահույթները:

A_2-ին համապատասխանում է վերին կեսը, A_3-ին՝ անկյունագծից ձախ և ներքև գտնվող կեսը: $A_1A_2A_3$ պատահույթին համապատասխանում է նկար 10ա-ում նշված տիրույթը, որի մակերեսը հավասար է քառակուսու մակերեսի $1/8$ մասին:





Նկար 10ա: $A_1A_2A_3$ պատահույթը: Նկար 10բ: A_1A_3 պատահույթը:

Այստեղից՝ $\mathrm{P}\{A_1A_2A_3\} = 1/8 = \mathrm{P}\{A_1\}\mathrm{P}\{A_2\}\mathrm{P}\{A_3\}$: Սակայն A_1A_3 պատահույթին համապատասխանում է 10ք նկարում նշված տիրույթը, որի մակերեսը հավասար է քառակուսու մակերեսի 3/8-ին:

Եզրակացություն. քանի որ $\mathrm{P}\{A_1A_3\} = 3/8 \neq \mathrm{P}\{A_1\}\mathrm{P}\{A_3\}$, ապա A_1 և A_3 պատահույթները անկախ չեն:

1.5. Լրիվ հավանականության և Բայեսի բանաձևերը

Դիցուք A-ն որևէ պատահույթ է, B_1-ը, B_2-ը, $\ldots, B_N$-ը լրիվ խումբ կազմող պատահույթներ են: Ենթադրենք, որ հայտնի են B_n-երի $\mathrm{P}\{B_n\} > 0$ հավանականությունները և A-ի ըստ B_n-երի պայմանական հավանականությունները՝

$$\mathrm{P}\{A|B_1\}, \ldots, \mathrm{P}\{A|B_N\}:$$

Հարկավոր է գտնել A-ի հավանականությունը: Քանի որ $\bigcup_n B_n = \Omega$, ապա՝

$$A = A\Omega = A(\bigcup_n B_n) = \bigcup_n (AB_n):$$

$B_1, \ldots, B_N$ պատահույթների անհամատեղելիությունից հետևում է $AB_1, AB_2, \ldots, AB_N$ պատահույթների անհամատեղելիությունը, որտեղից՝

$$\mathrm{P}\{A\} = \sum_n \mathrm{P}\{AB_n\}:$$

Օգտվելով բազմապատկման կանոնից, կստանանք՝

$$\mathrm{P}\{A\} = \sum_{n=1}^{N} \mathrm{P}\{B_n\}\mathrm{P}\{A|B_n\}:$$

Այս՝ հաճախ օգտագործվող բանաձևը կոչվում է **լրիվ հավանականության բանաձև**:

B_n պատահույթները կոչվում են **վարկածներ**:

Այժմ նույն պայմաններում ենթադրենք նաև, որ A պատահույթը տեղի է ունեցել:

Հարկավոր է գտնել B_n-ի ըստ A-ի պայմանական հավանականությունը: Լուծումը գտնելու համար կիրառենք բազմապատկման կանոնը՝

$$\mathrm{P}\{AB_n\} = \mathrm{P}\{A\}\mathrm{P}\{B_n|A\} = \mathrm{P}\{B_n\}\mathrm{P}\{A|B_n\},$$

որտեղից (եթե $\mathrm{P}\{A\} > 0$)՝

$$\mathrm{P}\{B_n|A\} = \frac{\mathrm{P}\{B_n\}\mathrm{P}\{A|B_n\}}{\mathrm{P}\{A\}}, \quad n = \overline{1,N}:$$

$\mathrm{P}\{A\}$-ն արտահայտելով ըստ լրիվ հավանականության բանաձևի, ստանում ենք՝

$$\mathrm{P}\{B_n|A\} = \frac{\mathrm{P}\{B_n\}\mathrm{P}\{A|B_n\}}{\sum_{k=1}^{N} \mathrm{P}\{B_k\}\mathrm{P}\{A|B_k\}}, \quad n = \overline{1,N}:$$

Այս բանաձևերը կրում են **Բայեսի բանաձևեր** անվանումը:

Դրանք նույնպես շատ հաճախ են օգտագործվում:

<u>*Օրինակ*</u> 23: Արկղում կա թենիսի 10 գնդակ, այդ թվում՝ 6 նոր և 4 արդեն օգտագործված: Խաղի համար արկղից պատահականորեն վերցվում է երկու գնդակ, որոնք խաղից հետո վերադարձվում են արկղ: Հաջորդ խաղի համար արկղից նորից է վերցվում երկու պատահական գնդակ: Պահանջվում է գտնել այդ երկու գնդակների նոր լինելու հավանականությունը:

Լուծում: Նշանակենք մեզ հետաքրքրող պատահույթը $\mathcal{A}$-ով: Դիտարկենք հետևյալ դեպքերը (վարկածները).

$\mathcal{B}_1$ - առաջին խաղի համար վերցված երկու գնդակն էլ նոր էին,

$\mathcal{B}_2$ - առաջին խաղի համար վերցված գնդակներից մեկը նոր էր, մյուսը՝ օգտագործված,

$\mathcal{B}_3$ - առաջին խաղի երկու գնդակներն էլ արդեն օգտագործված էին:

Հաշվենք համապատասխան հավանականությունները՝

$$\mathbf{P}\{\mathcal{B}_1\} = C_6^2/C_{10}^2 = 1/3, \quad \mathbf{P}\{\mathcal{B}_2\} = C_6^1C_4^1/C_{10}^2 = 8/15, \quad \mathbf{P}\{\mathcal{B}_3\} = C_4^2/C_{10}^2 = 2/15,$$
$$\mathbf{P}\{\mathcal{A}|\mathcal{B}_1\} = C_4^2/C_{10}^2 = 2/15, \quad \mathbf{P}\{\mathcal{A}|\mathcal{B}_2\} = C_5^2/C_{10}^2 = 2/9, \quad \mathbf{P}\{\mathcal{A}|\mathcal{B}_3\} = \mathbf{P}\{\mathcal{B}_1\} = 1/3:$$

Լրիվ հավանականության բանաձևի օգնությամբ կստանանք՝

$$\mathbf{P}\{\mathcal{A}\} = \sum_{n=1}^{3} \mathbf{P}\{\mathcal{B}_n\}\mathbf{P}\{\mathcal{A}|\mathcal{B}_n\} = \frac{1}{3} \times \frac{2}{15} + \frac{8}{15} \times \frac{2}{9} + \frac{2}{15} \times \frac{1}{3} = \frac{28}{135} = 0.2074:$$

<u>*Օրինակ*</u> 24: Գործարանն օգտվում է բաղադրատարրեր մատակարարող երեք՝ Ա, Բ, Գ ֆիրմաներից: Ա ֆիրմայի մատակարարման ծավալը կազմում է ընդհանուրի 50%-ը, իսկ Բ և Գ ֆիրմաներինը՝ համապատասխանաբար 30% և 20%: Հայտնի է, որ Ա ֆիրմայի մատակարարած բաղադրատարրերի 10%-ը խոտանված է: Համապատասխանաբար Բ ֆիրմայի խոտանը կազմում է 5%, Գ-ինը՝ 6%: Պատահականորեն վերցված բաղադրամասը խոտանված է: Ինչի՞ է հավասար այն Ա ֆիրմայից ստացված լինելու հավանականությունը:

Լուծում: Նշանակենք $\mathcal{A}$-ով վերցված բաղադրատարրի խոտանված լինելու պատահույթը: Երեք վարկածներն են՝ $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$՝ բաղադրամասը պատրաստված էր համապատասխանաբար Ա, Բ, Գ ֆիրմաների կողմից: Հարկավոր է հաշվել $\mathbf{P}\{\mathcal{B}_1|\mathcal{A}\}$ հավանականությունը: Բայեսի բանաձևի համաձայն՝

$$\mathbf{P}\{\mathcal{B}_1|\mathcal{A}\} = \mathbf{P}\{\mathcal{B}_1\}\mathbf{P}\{\mathcal{A}|\mathcal{B}_1\}/\left(\sum_{n=1}^{3}\mathbf{P}\{\mathcal{B}_n\}\mathbf{P}\{\mathcal{A}|\mathcal{B}_n\}\right):$$

Անհրաժեշտ հավանականությունները տրված են՝

$$\mathbf{P}\{\mathcal{B}_1\} = 0.5, \quad \mathbf{P}\{\mathcal{B}_2\} = 0.3, \quad \mathbf{P}\{\mathcal{B}_3\} = 0.2,$$
$$\mathbf{P}\{\mathcal{A}|\mathcal{B}_1\} = 0.1, \quad \mathbf{P}\{\mathcal{A}|\mathcal{B}_2\} = 0.05, \quad \mathbf{P}\{\mathcal{A}|\mathcal{B}_3\} = 0.06,$$

տեղադրելով, կստանանք՝

$$\mathbf{P}\{\mathcal{B}_1|\mathcal{A}\} = \frac{0.5 \times 0.1}{0.5 \times 0.1 + 0.3 \times 0.05 + 0.2 \times 0.06} = \frac{0.05}{0.077} \approx 0.65:$$

Գլուխ 2
Պատահական մեծություն և բաշխման ֆունկցիա

հավանականության տեսությունը, ըստ էության, ողջամիտ ընդհանուր իմաստություն է՝ վերածված հաշվի:

Պիեռ Սիմոն Լապլաս

2.1. Պատահական մեծություններ

Ինչպես գիտենք նախորդ գլխից, պատահական ելքերով փորձը նկարագրվում է հավանականային տարածության $\{\Omega, \mathcal{F}, \mathbf{P}\}$ մաթեմատիկական մոդելով: Սակայն կամայական կիրառական հետազոտություններ կատարելիս բնական է փորձերի արդյունքները նկարագրել թվային տվյալների միջոցով: Պատճառն այն է, որ թվերի հետ աշխատելը հարմար է, և միաժամանակ հնարավորություն է ստեղծվում առավելագույնս օգտվել մաթեմատիկական եղանակներից: Պատահական փորձի նկարագրումը թվային տվյալներով կատարվում է **պատահական մեծության** գաղափարի միջոցով: Նախ, կշարադրենք այդ կարևոր գաղափարի բովանդակային նկարագրությունը, ապա կտանք պատահական մեծության խիստ մաթեմատիկական սահմանումը:

Փորձի ամեն մի ω ելքին ինչ-որ ձևով համապատասխանեցնենք $X(\omega)$ իրական թիվ: Այլ կերպ ասած, ω ելքի ֆիզիկական կամ այլ նկարագրության փոխարեն ներկայացնենք այն $X(\omega)$ թվի միջոցով: Այսպիսով, հավանականային տարածության մոդելի առաջին բաղադրիչը՝ կամայական Ω բազմությունը, փոխարինվում է **իրական թվերի $\mathcal{R}$ բազմությունով** կամ նրա որևէ ենթաբազմությունով: Հաճախ փորձի արդյունքներն արդեն թվային են, այդ դեպքում $\Omega = \mathcal{R}$, և արտապատկերումը կարող է լինել նույնական՝ $X(x) = x$ կամ մեկ այլ տեսակի: Մոդելի հաջորդ բաղադրիչին՝ $\mathcal{F}$ σ-հանրահաշվին, համապատասխանեցվում է մի որոշակի՝ **բորելյան դաշտ** կոչվող σ-հանրահաշիվ, որը պարունակում է իրական առանցքի բոլոր հնարավոր միջակայքերը և դրանցից վերջավոր կամ անվերջ թվով գործողությունների միջոցով ստացվող բազմությունները: Ինչու՞ ենք վերցնում միջակայքերը: Շատ հաճախ պետք է լինում իմանալ, թե ինչ հավանականությամբ է X պատահական մեծությունը արժեքներ ընդունում որևէ (x_1, x_2) միջակայքից: Եթե այդ հավանականությունները հայտնի լինեն կամայական իրական x_1-ի և x_2-ի համար, ապա փորձի մասին կունենանք սպառիչ տեղեկություններ: Դա այդպես է, քանի որ այլ պատահույթների համար, որոնք ստացվում են միջակայքերից՝ որոշակի գործողությունների միջոցով (այսինքն՝ բորելյան դաշտի տարրեր են), հնարավոր է հավանականությունները հաշվարկել, կիրառելով նախորդ գլխում ուսումնասիրված հատկությունները: Այսպիսով, մենք կստանանք փորձի հավանականային $\mathbf{P}$ «նկարագրության» փոխարեն նոր՝ $\mathbf{P}_X$ նկարագրությունը:

Եթե տրված են կամայական միջակայքերից X պատահական մեծության արժեքներ ընդունելու հավանականությունները, ապա ասում են, որ տրված է պատահական մեծության **բաշխման օրենքը**: Գոյություն ունեն բաշխման օրենքը ներկայացնելու մի քանի միջոցներ կամ, կարելի է ասել, «գործիքներ»: Պարզվում է, օրինակ, որ բավական է իմանալ հավանականությունները միայն $(-\infty, x)$ տեսքի անվերջ միջակայքերի համար, քանի որ կամայական վերջավոր միջակայք կարելի է ներկայացնել որպես նշված տեսքի անվերջ միջակայքերի տարբերություն՝

$$[x_1, x_2) = (-\infty, x_2) - (-\infty, x_1),$$

և հավանականությունն էլ կգրվի համապատասխանաբար որպես տարբերություն՝

$$\mathbf{P}\{x_1 \le X < x_2\} = \mathbf{P}\{-\infty < X < x_2\} - \mathbf{P}\{-\infty < X < x_1\}:$$

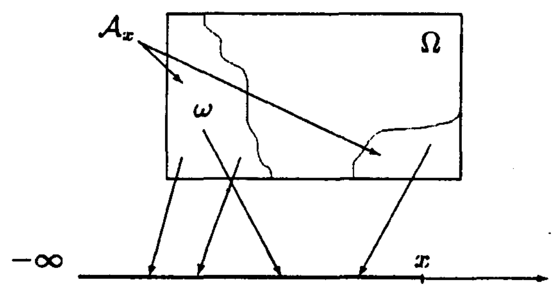
Կարելի է գրել նաև ավելի կարճ՝

$$\mathbf{P}\{-\infty < X < x\} = \mathbf{P}\{X < x\}:$$

Այսպիսով, բոլոր x-երի համար $\mathbf{P}\{X < x\}$ հավանականություններն ունենալը նույնպես X պատահական մեծության բաշխման օրենքը տալու միջոց է:

Անցնենք խիստ ձևակերպումներին:

> Դիցուք տրված է $\{\Omega, \mathcal{F}, \mathbf{P}\}$ հավանականային տարածությունը: Ω-ի վրա որոշված, իրական արժեքներ ընդունող
>
> $$X = X(\omega),\ \omega \in \Omega,$$
>
> չափելի ֆունկցիան անվանում են **պատահական մեծություն**: Ֆունկցիան կոչվում է **չափելի**, եթե կամայական իրական x-երի համար $(-\infty, x)$ միջակայքերի $\mathcal{A}_x = \{\omega : X(\omega) < x\}$ նախապատկերները պատկանում են $\mathcal{F}$-ին:



Նկար 1: $(-\infty, x)$ միջակայքի $\mathcal{A}_x$ նախապատկերը:

Օրինակ 1: Առաջին գլխի օրինակ 10-ում մետաղադրամի նետման տարրական պատահույթները նշանակել ենք ω_1, ω_2: Դիցուք խաղացողը ω_1-ի հանդես գալու դեպքում շահում է 20 դրամ, իսկ ω_2-ի դեպքում տանուլ է տալիս 20 դրամ: Այլ կերպ ասած, ω_1-ին համապատասխանեցվում է 20 թիվը, $X(\omega_1) = 20$, իսկ ω_2-ին՝ -20-ը, $X(\omega_2) = -20$: Ասում են, որ X պատահական մեծությունը՝ խաղացողի շահումը, ընդունում է 20 և -20 արժեքները, տվյալ դեպքում իրար հավասար՝ $1/2$, հավանականություններով:

> Քանի որ $\{\omega : X(\omega) < x\}$ բազմությունը պատկանում է $\mathcal{F}$-ին, այսինքն՝ պատահույթ է, ապա որոշված է նրա հավանականությունը՝ $\mathbf{P}\{\omega : X(\omega) < x\}$: Այդ հավանականությունը, որպես ֆունկցիա x-ից, $x \in \mathcal{R}$, կոչվում է X պատահական մեծության **բաշխման ֆունկցիա** և նշանակվում է $F_X(x)$.
>
> $$F_X(x) = \mathbf{P}\{\omega : X(\omega) < x\} = \mathbf{P}\{X < x\}:$$

Եթե պարզ է, թե ինչ պատահական մեծության մասին է խոսքը, ապա $F_X(x)$-ի փոխարեն երբեմն գրվում է $F(x)$:

Բաշխման ֆունկցիան որոշված է կամայական պատահական մեծության համար և սպառիչ տեղեկություններ է տալիս նրա մասին: Սակայն բաշխման օրենքը նկարագրելու համար կան նաև այլ միջոցներ, որոնց օգտագործումը որոշ խնդիրներում ավելի հարմար է:

2.2. Բաշխման ֆունկցիայի հատկությունները

Բաշխման ֆունկցիան օժտված է հետևյալ հիմնական հատկություններով.
1. Եթե $x_1 < x_2$, ապա $F_X(x_1) \le F_X(x_2)$ (**միընթացությունը**),
2. $\lim\limits_{x\to-\infty} F_X(x) = 0, \ \lim\limits_{x\to+\infty} F_X(x) = 1,$
3. $F_X(x-0) = F_X(x)$ (**ձախից անընդհատությունը**):

Ապացուցենք այդ հատկությունները:

1. Դիցուք $x_1 < x_2$: Քանի որ $\{\omega : X(\omega) < x_1\} \subset \{\omega : X(\omega) < x_2\}$, ապա համաձայն հավանականության միընթացության հատկության՝ $\mathbf{P}\{X < x_1\} \le \mathbf{P}\{X < x_2\}$: Նկատի ունենալով, որ $\mathbf{P}\{X < x_1\} = F_X(x_1)$ և $\mathbf{P}\{X < x_2\} = F_X(x_2)$, ստանում ենք՝ $F_X(x_1) \le F_X(x_2)$:

2. Ապացուցենք, որ $\lim\limits_{x\to-\infty} F_X(x) = 0$: Վերցնենք որևէ նվազող հաջորդականություն $x_1 > x_2 > \ldots > x_n > \ldots$, որը ձգտում է $-\infty$: Նշանակենք $A_n = \{\omega : X(\omega) < x_n\}$: $A_1, A_2, \ldots, A_n, \ldots$ պատահույթների հաջորդականությունն այնպիսին է, որ $A_n \supseteq A_{n+1}$, $n = 1, 2, \ldots$ և $\bigcap\limits_n A_n = \varnothing$: Հավանականության անընդհատության հատկությունից հետևում է, որ $\lim\limits_{n\to\infty} \mathbf{P}\{A_n\} = \mathbf{P}\{\varnothing\} = 0$, ուստի $\lim\limits_{n\to\infty} F_X(x_n) = 0$: Քանի որ $x_1, x_2, \ldots, x_n, \ldots$ հաջորդականությունն ընտրվել էր կամայականորեն, ապա $\lim\limits_{x\to-\infty} F_X(x) = 0$:

Նման ձևով ապացուցվում են $\lim\limits_{x\to\infty} F_X(x) = 1$ և բաշխման ֆունկցիայի ձախից անընդհատության հատկությունները:

Կարելի է ապացուցել, որ ձևակերպված երեք հատկություններով օժտված կամայական $F(x)$ ֆունկցիան որևէ X պատահական մեծության բաշխման ֆունկցիա է՝ $F_X(x) = F(x)$:

Նշենք նաև բաշխման ֆունկցիայի այլ կարևոր հատկությունները, որոնք բխում են հիմնական հատկություններից.

4. $0 \le F_X(x) \le 1$, 5. $\mathbf{P}\{a \le X < b\} = F_X(b) - F_X(a)$,
6. $\mathbf{P}\{X = x\} = F_X(x+0) - F_X(x)$,
7. $F_X(x)$-ը կարող է ունենալ ոչ ավել, քան հաշվելի թվով խզման կետեր:

Հաջորդ ենթաբաժիններում կուսումնասիրենք մի շարք որոշակի դասերի պատահական մեծությունները և նրանց բաշխման ֆունկցիաները:

2.3. Ընդհատ բաշխումներ

Առանձնացնում են պատահական մեծությունների երկու առավել կարևոր տիպեր՝ **ընդհատ (դիսկրետ)** և **անընդհատ**: Դիսկրետ տիպի պատահական մեծությունը կարող է ընդունել միայն իրարից անջատված արժեքներ, անընդհատ տիպի պատահական մեծությունը՝ որևէ հատվածի կամ հատվածների բոլոր արժեքները:

X պատահական մեծությունը կոչվում է **ընդհատ** (կամ ասում են պատահական մեծությունը ունի **ընդհատ բաշխում**), եթե այն ընդունում է վերջավոր կամ հաշվելի թվով արժեքներ՝ $x_1, x_2, \ldots, x_k, \ldots$, համապատասխանաբար $p_k = \mathbf{P}\{X = x_k\}$, $k = 1, 2, \ldots$, հավանականություններով: X ընդհատ պատահական մեծության բաշխման օրենքը տրվում է հետևյալ տեսքի **բաշխման աղյուսակով**՝

X	x_1	x_2	$\ldots$	x_k	$\ldots$
p	p_1	p_2	$\ldots$	p_k	$\ldots$

$p_k \ge 0, \ k = 1, 2, \ldots, \ \sum\limits_k p_k = 1$:

Ընդհատ X պատահական մեծության բաշխման ֆունկցիան կարելի է կառուցել բաշխման աղյուսակի օգնությամբ՝

$$F_X(x) = \sum_{k:x_k<x} p_k,$$

այն աստիճանաձև է, դրա խզման կետերն են $x_1, x_2, \dots, x_k, \dots$, խզման մեծությունները այդ կետերում հավասար են պատահական մեծության համապատասխան արժեքների հավանականություններին՝

$$F_X(x_k+0) - F_X(x_k) = p_k, \quad k = 1, 2, \dots :$$

Օրինակ 2: Կանոնավոր մետաղադրամի մեկ նետման ժամանակ զինանշանը կարող է հանդես գալ (1 անգամ) կամ հանդես չգալ (հանդես գալ 0 անգամ): Դիցուք X-ը մետաղադրամի մեկ նետման ժամանակ զինանշանի հանդես գալու թիվն է: Կազմենք X-ի բաշխման աղյուսակը և կառուցենք $F_X(x)$-ը:

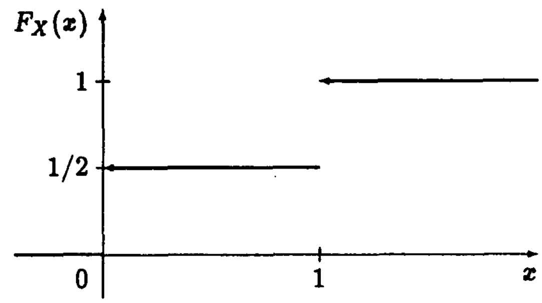
Լուծում: X-ի հնարավոր արժեքներն են 0-ն և 1-ը: $\mathbf{P}\{X = 0\} = \mathbf{P}\{X = 1\} = 1/2$: Հետևաբար X-ի բաշխման աղյուսակն է՝

X	0	1
p	1/2	1/2

:

Գտնենք $F_X(x)$ բաշխման ֆունկցիան և կառուցենք նրա գծապատկերը (նկար 2)՝

$$F_X(x) = \begin{cases} 0, & \text{երբ } x \le 0, \\ 1/2, & \text{երբ } 0 < x \le 1, \\ 1, & \text{երբ } x > 1 : \end{cases}$$



Նկար 2: Մետաղադրամի մեկ նետման դեպքում զինանշանի հանդես գալու թվի բաշխման ֆունկցիայի գծապատկերը:

Օրինակ 3: Կանոնավոր մետաղադրամը նետվում է երկու անգամ՝ $\Omega = \{ԹԹ, ԶԹ, ԹԶ, ԶԶ\}$: X-ը զինանշանի հանդես գալու թիվն է: Գտնենք X-ի բաշխման աղյուսակը և բաշխման ֆունկցիան:

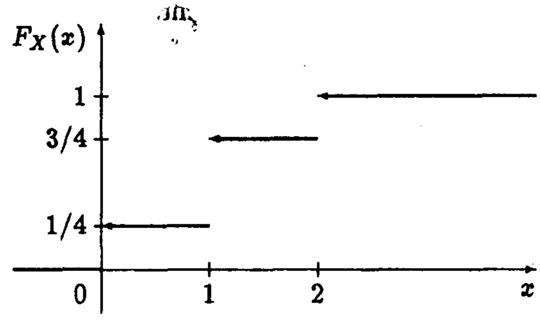
Լուծում: X-ի հնարավոր արժեքներն են 0-ն, 1-ը և 2-ը, $\mathbf{P}\{X = 0\} = \frac{1}{4}$, $\mathbf{P}\{X = 1\} = \frac{1}{2}$, $\mathbf{P}\{X = 2\} = \frac{1}{4}$: Բաշխման աղյուսակն է՝

X	0	1	2
p	1/4	1/2	1/4

:

Գտնենք բաշխման ֆունկցիան և կառուցենք նրա գծապատկերը (նկար 3)՝

$$F_X(x) = \begin{cases} 0, & \text{երբ } x \le 0, \\ 1/4, & \text{երբ } 0 < x \le 1, \\ 3/4, & \text{երբ } 1 < x \le 2, \\ 1, & \text{երբ } x > 2 : \end{cases}$$



Նկար 3: Օրինակ 3-ի բաշխման ֆունկցիայի գծապատկերը:

Կարևորագույն ընդհատ բաշխումներից մեկը N և p պարամետրերով **երկանդամային բաշխումն** է (N-ը բնական թիվ է, $0 \le p \le 1$): Կարճ գրում են՝ $X \sim \mathcal{B}(N, p)$:

Դիտարկենք մի որոշակի փորձի N անկախ կրկնությունների հաջորդականությունը: Ամեն փորձում $\mathcal{A}$ պատահույթը տեղի է ունենում միևնույն p հավանականությամբ և տեղի չի ունենում $q = 1 - p$ հավանականությամբ: Փորձերի անկախ լինելը նշանակում է, որ ամեն փորձում $\mathcal{A}$-ի տեղի ունենալու հավանականությունը կախված չէ մյուս փորձերի արդյունքներից: Մեզ հետաքրքրում է N փորձերում $\mathcal{A}$ պատահույթի n անգամ հանդես գալու հավանականությունը, $0 \le n \le N$, որը կնշանակենք $P_N(n)$-ով: Սկզբից դիտարկենք $n = N$ դեպքը, այսինքն գտնենք $\mathcal{A}$ պատահույթի բոլոր փորձերում իրականանալու $P_N(N)$ հավանականությունը: Քանի որ փորձերն անկախ են, ապա հավանականությունները բազմապատկվում են և

$$P_N(N) = \mathbf{P}\{\underbrace{\mathcal{A}\cap\mathcal{A}\cap\ldots\cap\mathcal{A}}_{N \text{ անգամ}}\} = p^N :$$

$\mathcal{A}$-ի ոչ մի անգամ տեղի չունենալու $P_N(0)$ հավանականությունը հավասար է $\mathcal{A}'$-ը N անգամ իրականանալու հավանականությանը, այն է՝ q^N-ին: Երբ սկզբից $\mathcal{A}$-ն տեղի է ունենում n անգամ, իսկ հետո $\mathcal{A}'$-ը՝ $(N-n)$ անգամ, ապա հաշվի առնելով փորձերի անկախությունը՝ այդ դեպքի հավանականությունը կլինի

$$\mathbf{P}\{\underbrace{\mathcal{A}\cap\cdots\cap\mathcal{A}}_{n \text{ անգամ}}\cap\underbrace{\mathcal{A}'\cap\cdots\cap\mathcal{A}'}_{(N-n) \text{ անգամ}}\} = p^n \times q^{N-n} :$$

Որպեսզի գտնենք $P_N(n)$-ը, մնում է նկատել, որ բոլոր այն հաջորդականությունները, որոնցում $\mathcal{A}$-ն իրականացել է n անգամ, իսկ $(N-n)$ անգամ տեղի չի ունեցել, ունեն միևնույն հավանականությունը՝ $p^n \times q^{N-n}$ և տարբերվում են միայն $\mathcal{A}$ և $\mathcal{A}'$ պատահույթների տեղի ունենալու հերթականությամբ: Այդպիսի հաջորդականությունների թիվը C_N^n է: Հետևաբար, ստանում ենք **Բեռնուլիի բանաձևը**՝

$$P_N(n) = C_N^n p^n q^{N-n}, \quad n = \overline{0,N} :$$

Քանի որ N փորձերում նշված դեպքերից մեկն ու մեկն անպայման կիրականանա, ապա այս հավանականությունների գումարը պետք է լինի հավասար 1-ի:

Իսկապես, Նյուտոնի երկանդամի բանաձևի օգնությամբ ստանում ենք՝

$$\sum_{n=0}^{N} P_N(n) = \sum_{n=0}^{N} C_N^n p^n q^{N-n} = (p+q)^N = 1 :$$

Դիցուք X-ը N անկախ փորձերում որևէ $\mathcal{A}$ պատահույթի իրականացումների թիվն է, ասում են՝ «հաջողությունների» թիվը: Եթե փորձերից յուրաքանչյուրում $\mathbf{P}\{\mathcal{A}\} = p$, ապա X-ը ունի **երկանդամային** (կամ **Բեռնուլիի**) **բաշխում**՝

$$\mathbf{P}\{X = n\} = P_N(n) = C_N^n p^n (1-p)^{N-n}, \quad n = \overline{0,N} :$$

Օրինակ 4: Հաշվապահությունում ստուգվող փաստաթղթերի 7%-ը թերի է ձևակերպված: Ստուգվող փաստաթղթերից կամայականորեն ընտրվում են 4-ը: X-ը նրանց միջև ճիշտ ձևակերպված փաստաթղթերի թիվն է: Ինչպիսի՞ բաշխում ունի X-ը:

Լուծում: Կատարվում է 4 փորձ: Փաստաթղթի թերի լինելու հավանականությունը՝ $p = 0.07$: Ուստի X-ն ունի երկանդամային բաշխում՝ $P_4(n) = C_4^n \cdot (0.93)^n \cdot (0.07)^{4-n}, n = 0,1,2,3,4$:

Օրինակ 5: Մետաղադրամը նետում են 5 անգամ: Գտնենք զինանշանի հանդես գալու թիվը ներկայացնող X պատահական մեծության բաշխման աղյուսակը:

Լուծում: $P_5(0) = (1/2)^5 = 1/32$, $P_5(1) = C_5^1(1/2)(1/2)^4 = 5/32$,

$P_5(2) = C_5^2(1/2)^2(1/2)^3 = 5/16$, $P_5(3) = C_5^3(1/2)^3(1/2)^2 = 5/16$,

$P_5(4) = C_5^4(1/2)^4(1/2) = 5/32$, $P_5(5) = 1/32$:

Ստացանք՝

X	0	1	2	3	4	5
p	1/32	5/32	10/32	10/32	5/32	1/32

:

Որոշ իրադրություններում կարևոր է գտնել «հաջողությունների ամենահավանական թիվը», այսինքն այն n^*-ը, որի դեպքում $P_N(n)$ հավանականությունը մեծագույնն է: n^*-ը գտնելու համար դիտարկենք երկու հաջորդական հավանականությունների հարաբերությունը.

$$\frac{P_N(n+1)}{P_N(n)} = \frac{C_N^{n+1} p^{n+1} q^{N-n-1}}{C_N^n p^{n+1} q^{N-n}} = \frac{(N-n)p}{(n+1)q}:$$

Հարաբերությունը մեծ է մեկից, այսինքն՝ $P_N(n+1)$-ը մեծ է $P_N(n)$-ից, եթե

$$(N-n)p > (n+1)q, \quad \text{կամ} \quad n < Np - q:$$

Իսկ

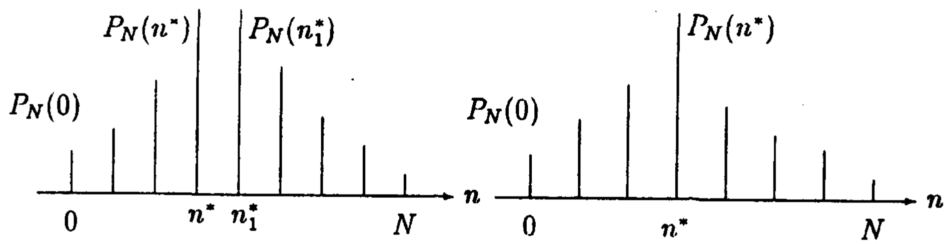
$$P_N(n+1) = P_N(n), \text{ եթե } n = Np - q, \text{ և } P_N(n+1) < P_N(n), \text{ եթե } n > Np - q:$$

Տեսնում ենք, որ $P_N(n)$ հավանականությունը n-ի մեծացման հետ նախ աճում է, հասնում իր մեծագույն արժեքին, ապա նվազում է:

Եթե $(Np - q)$ թիվն ամբողջ է, ապա $P_N(n)$-ը մեծագույն արժեքը ընդունում է n-ի հետևյալ երկու արժեքների դեպքում՝

$$n^* = Np - q \text{ և } n_1^* = n^* + 1 = Np - q + 1 = Np + p:$$

Իսկ եթե $(Np - q)$-ն ամբողջ չէ, ապա $n^* = [Np - q] + 1$ ($[.]$-ը թվի ամբողջ մասն է):



Նկար 4: Հավանականությունների երկանդամային բաշխման օրենքի դեպքերի գծապատկերներ:

<u>*Օրինակ*</u> 6: Ներդրումային ընկերությունն ունի 500 աշխատակից: Աշխատակցի տվյալ օրը ուշանալու կամ աշխատանքի չգալու հավանականությունը 0.02 է: Գտնել ժամանակին աշխատանքի եկող աշխատակիցների ամենահավանական թիվը:

Լուծում: Խնդրի պայմանները թույլ են տալիս օգտվել Բեռնուլիի բանաձևից: Այստեղ $N = 500$, $q = 0.02$, $p = 0.98$: Քանի որ $Np - q = 490 - 0.02 = 489.98$ թիվը ամբողջ չէ, ապա ամենահավանական արժեքը միակն է՝ 490:

λ պարամետրով **Պուասոնի բաշխման** համար ընդունված է կարճ գրառում՝ $\Pi(\lambda)$: N-ի մեծ արժեքների դեպքում $P_N(n)$ հավանականությունների Բեռնուլիի բանաձևով հաշվարկումը դժվարանում է, ուստի նպատակահարմար է ասիմպտոտական մոտավոր բանաձևերի օգտագործումը:

Այն դեպքում, երբ N-ը մեծ է, իսկ p հավանականությունը՝ փոքր, Բեռնուլիի բանաձևը կարելի է փոխարինել (մոտարկել) **Պուասոնի բանաձևով**՝

$$P_N(n) \approx \frac{(Np)^n}{n!} e^{-Np}, \quad n = 0, 1, 2, \ldots:$$

Հաջորդ թեորեմը հիմնավորում է այդ մոտարկումը:

Պուասոնի թեորեմը: Եթե $N \to \infty$, իսկ $p \to 0$ այնպես, որ $Np \to \lambda$, $0 < \lambda < \infty$, ապա

$$\lim_{N\to\infty} P_N(n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, \ldots: \qquad (1)$$

Ապացույց: Նշանակենք $\lambda_N = Np$ կամ $p = \frac{\lambda_N}{N}$: Թեորեմի պայմանի համաձայն $\lambda_N \to \lambda$, երբ $N \to \infty$: Ներկայացնենք $P_N(n)$-ը հետևյալ տեսքով.

$$P_N(n) = C_N^n p^n q^{N-n} = \frac{N(N-1)\dots(N-(n-1))}{n!}\left(\frac{\lambda_N}{N}\right)^n\left(1-\frac{\lambda_N}{N}\right)^{N-n} =$$

$$= \frac{\lambda_N^n}{n!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\left(1-\frac{\lambda_N}{N}\right)^N\left(1-\frac{\lambda_N}{N}\right)^{-n}:$$

Երբ $N \to \infty$, ապա $\lambda_N^n \to \lambda^n$,

$$\left(1-\frac{\lambda_N}{N}\right)^N \to e^{-\lambda},\ \left(1-\frac{\lambda_N}{N}\right)^{-n} \to 1,\ \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right) \to 1:$$

Հետևաբար, ստանում ենք (1)-ը:

Հարկավոր է նշել, որ Պուասոնի թեորեմից կարելի է օգտվել նաև, երբ p-ն մոտ է 1-ին, քանի որ այդ դեպքում $q = (1-p)$-ն մոտ է 0-ին, և մենք հնարավորություն ենք ստանում գտնել A պատահույթի n անգամ տեղի չունենալու հավանականության մոտավոր արժեքը:

Եթե X պատահական մեծության համար

$$p_n = \mathrm{P}\{X = n\} = \frac{\lambda^n}{n!}e^{-\lambda},\quad n = 0,1,2,\dots,\quad \lambda > 0,$$

ապա ասում են, որ այն բաշխված է λ պարամետրով **Պուասոնի օրենքով** (կամ **բաշխման օրենքը պուասոնյան է**): Գրում են $X \sim \Pi(\lambda)$:

Համոզվենք, որ $\sum_{n=0}^{\infty} p_n = 1$: Իրոք՝ $\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1$:

Պուասոնյան բաշխումը հաճախ է հանդիպում պատահական ընթացքների, զանգվածային սպասարկման և հուսալիության տեսություններում:

<u>Օրինակ</u> 7: *Վիճակախաղի տոմսերի խնդիրը*: Վիճակախաղի քանի՞ տոմս պետք է գնել, որպեսզի շահելու հավանականությունը լինի տրված P թվից ոչ պակաս, $0 < P < 1$:

Լուծում: Դիցուք վիճակախաղի տոմսերի ընդհանուր թիվը K է, իսկ k-ն շահող տոմսերի թիվն է: Հետևաբար մեկ տոմսի շահող լինելու հավանականությունը $p = k/K$ է: Տոմսերի ստուգումը կարելի է դիտարկել որպես անկախ փորձեր՝ շահելու p հավանականությամբ: Եթե գնված տոմսերի թիվը՝ N-ը, բավականաչափ մեծ է, ապա կարելի է, օգտվելով Պուասոնի թեորեմից, կիրառել (1) բանաձևը,

$$P_N(n) \approx \frac{\lambda^n}{n!}e^{-\lambda},$$

որտեղ n-ը շահումների թիվն է, $\lambda = N\frac{k}{K}$: Ոչ մի շահում չունենալու հավանականությունը կլինի $P_N(0) = e^{-\lambda}$: Գոնե մեկ տոմսի շահելու հավանականությունը հավասար է $1 - P_N(0) = 1 - e^{-\lambda}$: Հետևաբար, որոնելի գոնե մեկ շահում ապահովող տոմսերի նվազագույն թիվը կարելի է գտնել

$$P \le 1 - e^{-\lambda} = 1 - \exp\{-Nk/K\},\ \text{կամ}\ \exp\{-Nk/K\} \le 1 - P$$

պայմանից: Ստանում ենք՝ $N \ge (-K/k)\log(1-P)$: Օրինակ, $K = 10000$, $k = 100$, $P = 0.95$, ապա գոնե մեկ շահում ապահովելու համար հարկավոր է գնել առնվազն 300 տոմս, իսկ եթե P-ն վերցնենք ավելի մեծ՝ $P = 0.99$, ապա արդեն հարկավոր է գնել առնվազն 460 տոմս:

Ասում են, որ X պատահական մեծության **բաշխումը երկրաչափական** է p պարամետրով $(0 \le p \le 1)$ և գրում են՝ $X \sim \mathcal{G}(p)$, եթե

$$p_n = \mathrm{P}\{X = n\} = q^{n-1}p,\quad q = 1-p,\quad n = 1,2,\dots:$$

Պարզ է, որ $\sum_{n=1}^{\infty} p_n = 1$, քանի որ $\sum_{n=1}^{\infty} q^{n-1}p = \frac{p}{1-q} = 1$:

Դիցուք անկախ փորձերը, որոնցից յուրաքանչյուրում A պատահույթի հանդես գալու հավանականությունը p է, կատարվում են մինչև A պատահույթի առաջին անգամ հանդես գալը: Նշանակենք X-ով այդ փորձերի քանակը: Դժվար չէ համոզվել, որ X-ը բաշխված է երկրաչափական օրենքով:

<u>*Օրինակ*</u> 8: Զառը նետվում է մինչև 6-ի հանդես գալը: X-ը նետումների թիվն է: Այդ դեպքում X-ը բաշխված է երկրաչափական օրենքով՝

$$p_1 = \frac{1}{6},\ p_2 = \frac{5}{6}\cdot\frac{1}{6},\ p_3 = \left(\frac{5}{6}\right)^2\cdot\frac{1}{6},\ p_4 = \left(\frac{5}{6}\right)^3\cdot\frac{1}{6},\ldots,\ p_n = \left(\frac{5}{6}\right)^{n-1}\cdot\frac{1}{6},\ldots:$$

<u>*Օրինակ*</u> 9: Առկա է N հատ էլեկտրական լամպ, որոնցից յուրաքանչյուրը p հավանականությամբ, $0 < p < 1$, կարող է ունենալ թերություն: Լամպերը հաջորդաբար ստուգվում են, մինչև որ հայտնաբերվի անթերի լամպը: Լամպը ներպտուտակվում է կոթառի մեջ, և այնուհետև միացվում է հոսանքը: Հոսանքի միացումից թերություն ունեցող լամպն անմիջապես այրվում է, որից հետո փոխարինվում է նորով: Դիցուք X-ը փորձարկված լամպերի քանակն է: Գտնել X-ի բաշխումը:

Լուծում: X-ի հնարավոր արժեքներն են $1, 2, \ldots, N$: $\{X = n\}$, $1 \le n \le N$, պատահույթը նշանակում է, որ n-րդ լամպը չունի թերություն, իսկ նախկինում փորձարկված բոլոր լամպերն ունեն: Հետևաբար՝

$$\mathbf{P}\{X = n\} = p^{n-1}q,\ \ 1 \le n \le N-1,\ \ \mathbf{P}\{X = N\} = p^{N-1}:$$

Այսպիսով, երբ $1 \le n \le N-1$, հավանականությունները տրվում են երկրաչափական բաշխման բանաձևով, իսկ $\mathbf{P}\{X = N\}$ հավանականությունն ապահովում է $\sum\limits_n \mathbf{P}\{X = n\} = 1$ պայմանը:

Ասում են, որ X պատահական մեծության **բաշխումը հիպերերկրաչափական** է N, M, K պարամետրերով, $N = 1, 2, \ldots$, $M = \overline{0, N}$, $K = \overline{0, N}$, եթե

$$p_{N,M,K}(n) = \mathbf{P}\{X = n\} = C_K^n C_{N-K}^{M-n} / C_N^M,\ n = \max(0, M-N+K), \ldots, \min(M, K):$$

Այս բաշխման կիրառությունների մասին պատկերացում է տալիս հետևյալ իրավիճակը. N հատ շինվածքից բաղկացած խմբաքանակում K հատը առաջին տեսակի է, իսկ մնացած $(N-K)$ հատը՝ երկրորդ տեսակի: Ստուգման համար այդ խմբաքանակից պատահականորեն վերցնում են M հատ շինվածք: Նշանակենք X-ով վերցվածների մեջ առաջին տեսակի շինվածքների քանակը: Այդ դեպքում X-ի բաշխումը հիպերերկրաչափական է:

<u>*Օրինակ*</u> 10: Գործարարն ունի երկու ընկերությունների 7 հատ բաժնետոմս (չորսը՝ առաջին ընկերության, երեքը՝ երկրորդ ընկերության): Տվյալ օրը բոլոր բաժնետոմսերի գինը նույնն է: Գործարարը պատահականորեն ընտրում է 4 բաժնետոմս, որոնք պետք է վաճառքի դրվեն հաջորդ օրը: Գտնել գործարարի շահույթ ստանալու հավանականությունը, եթե հաջորդ օրն առաջին ընկերության բաժնետոմսերի գինը բարձրանա 10%-ով, իսկ երկրորդ ընկերության բաժնետոմսերի գինը իջնի 20%-ով:

Լուծում: Գործարարը շահույթ կստանա միայն այն դեպքում, եթե առաջին ընկերության վաճառված բաժնետոմսերի թիվը լինի 4 կամ 3: Նկատելով, որ առաջին ընկերության վաճառված բաժնետոմսերի թիվը բաշխված է հիպերերկրաչափական օրենքով, այդ դեպքերի հավանականությունները համապատասխանաբար կլինեն $\dfrac{C_4^4 \cdot C_3^0}{C_7^4}$ և $\dfrac{C_4^3 \cdot C_3^1}{C_7^4}$: Հետևաբար, գործարարի՝ շահույթ ստանալու հավանականությունը կլինի՝

$$\frac{C_4^4 \cdot C_3^0}{C_7^4} + \frac{C_4^3 \cdot C_3^1}{C_7^4} = \frac{13 \cdot 3! \cdot 4!}{7!} = \frac{13}{35} \approx 0.37:$$

2.4. Անընդհատ բաշխումներ

X պատահական մեծությունը կոչվում է **բացարձակ անընդհատ** (կամ ասում են X պատահական մեծությունն ունի **անընդհատ բաշխում**), եթե գոյություն ունի այնպիսի ոչ բացասական $f_X(x)$ ֆունկցիա, որ կամայական x-ի համար $F_X(x)$ բաշխման ֆունկցիան ներկայացվում է հետևյալ կերպ՝

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du :$$

Հեշտ է տեսնել, որ $\int_{-\infty}^{\infty} f_X(u)du = 1$: Իրոք՝

$$\int_{-\infty}^{\infty} f_X(u)du = \lim_{x\to\infty} \int_{-\infty}^{x} f_X(u)du = \lim_{x\to\infty} F_X(x) = 1:$$

$f_X(x)$ ֆունկցիան կոչվում է X պատահական մեծության հավանականությունների բաշխման **խտության ֆունկցիա** կամ ավելի կրճատ՝ **բաշխման խտություն**: Եթե պարզ է, թե որ պատահական մեծության մասին է խոսքը, $f_X(x)$-ի փոխարեն կարելի է գրել $f(x)$:

Եթե X-ը բացարձակ անընդհատ է, ապա $F_X(x)$ բաշխման ֆունկցիան անընդհատ է, և կամայական իրական x-ի համար $\mathbf{P}\{X = x\} = 0$:

Խտության ֆունկցիայի երկու հատկություններ հիմնական են՝

$$1.\ f_X(x) \geq 0, \qquad 2.\ \int_{-\infty}^{\infty} f_X(u)du = 1 :$$

Նշենք, որ կամայական $f(x)$, $x \in \mathcal{R}_1$ ֆունկցիա, որը օժտված է նշված երկու հատկություններով, որևէ X պատահական մեծության բաշխման խտություն է՝ $f_X(x) = f(x)$:

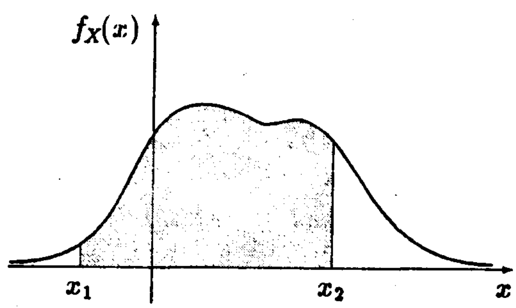
Նշենք նաև $f_X(x)$ բաշխման խտության հետևյալ կարևոր հատկությունները.

3. $f_X(x)$-ի անընդհատության կետերում $f_X(x) = F'_X(x)$,
4. կամայական $x_1 < x_2$ զույգի համար $\mathbf{P}\{x_1 \leq X < x_2\} = \int_{x_1}^{x_2} f_X(u)du$:

Ապացույց: Երրորդ հատկությունը ինտեգրալի ընդհանուր հատկությունների հետևանքն է: Ապացուցենք չորրորդ հատկությունը: Ունենք՝

$$\mathbf{P}\{x_1 \leq X < x_2\} = F_X(x_2) - F_X(x_1) = \int_{-\infty}^{x_2} f_X(u)du - \int_{-\infty}^{x_1} f_X(u)du = \int_{x_1}^{x_2} f_X(u)du :$$

Ըստ $\mathbf{P}\{x_1 \leq X < x_2\}$ հավանականության երկրաչափական մեկնաբանման՝ այն հավասար է նկար 5-ում ստվերագծված տիրույթի մակերեսին:



Նկար 5: $X \in [x_1, x_2)$ պատահույթի հավանականությունը:

Դիտարկենք անընդհատ բաշխումների կարևորագույն դասերը:

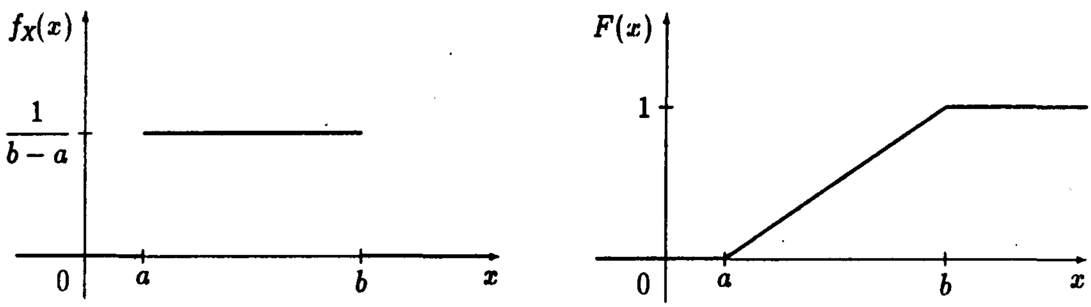
$[a,b]$ միջակայքում **հավասարաչափ օրենքով բաշխված** պատահական մեծության (կարճ գրում են՝ $X \sim \mathcal{U}(a,b)$) խտության ֆունկցիան է (տե՛ս նկար 6)՝

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{երբ } x \in (u,b), \\ 0, & \text{երբ } x \notin (a,b) : \end{cases}$$

Պարզ է, որ $\int\limits_{-\infty}^{\infty} f_X(u)du = 1$: Իսկապես, $\int\limits_{-\infty}^{\infty} f_X(u)du = \int\limits_{a}^{b} f_X(u)du = \dfrac{1}{b-a}\int\limits_{a}^{b} du = 1$:

Բաշխման ֆունկցիան տրվում է հետևյալ բանաձևով և ներկայացված է նկար 6-ում՝

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(u)du = \begin{cases} 0, & \text{երբ } x \le a, \\ \dfrac{x-a}{b-a}, & \text{երբ } a < x \le b, \\ 1, & \text{երբ } x > b : \end{cases}$$



Նկար 6: Հավասարաչափ օրենքի խտության և բաշխման ֆունկցիաները:

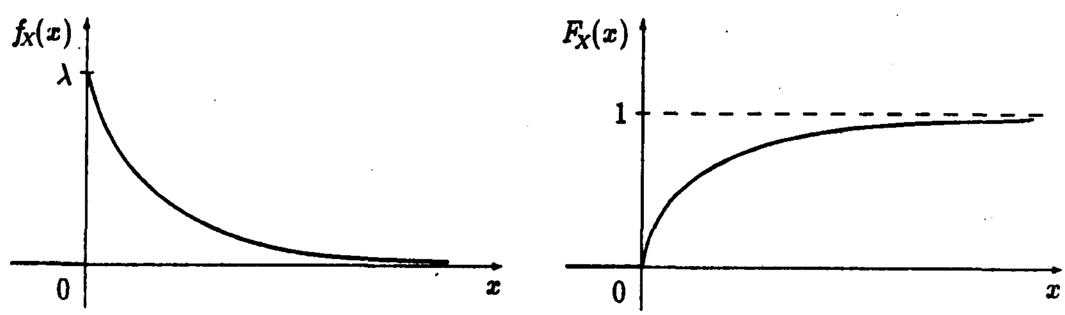
$\lambda > 0$ պարամետրով **ցուցչային օրենքով** բաշխված պատահական մեծության խտության ֆունկցիան է (տե՛ս նկար 7)՝

$$f_X(x) = \begin{cases} 0, & \text{երբ } x \le 0 \\ \lambda e^{-\lambda x}, & \text{երբ } x > 0 : \end{cases}$$

Ցույց տանք, որ $\int\limits_{-\infty}^{\infty} f_X(x)dx = 1$: Իսկապես՝ $\int\limits_{-\infty}^{\infty} f_X(x)dx = \int\limits_{-\infty}^{0} 0dx + \int\limits_{0}^{\infty} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$:

Ինտեգրելով խտությունը, կստանանք բաշխման ֆունկցիան (տե՛ս նկար 7)՝

$$F_X(x) = \begin{cases} 0, & \text{երբ } x \le 0, \\ 1 - e^{-\lambda x}, & \text{երբ } x > 0 : \end{cases}$$



Նկար 7: Ցուցչային օրենքի խտության և բաշխման ֆունկցիաների տեսքը:

Ցուցչային բաշխումը հաճախ է հանդիպում զանգվածային սպասարկման և հուսալիության տեսություններում:

$\lambda > 0$ և $\alpha > 0$ պարամետրերով **Վեյբուլի օրենքով** (ընդունված է կարճ գրառում՝ $X \sim \mathcal{W}(\lambda, \alpha)$) բաշխված պատահական մեծության խտության ֆունկցիան է՝

$$f_X(x) = \begin{cases} \dfrac{\alpha}{\lambda^\alpha} x^{\alpha-1} \exp\left\{-\left(\dfrac{x}{\lambda}\right)^\alpha\right\}, & \text{երբ } x > 0, \\ 0, & \text{երբ } x \le 0: \end{cases}$$

Համոզվենք, որ $\int\limits_{-\infty}^{\infty} f_X(x)dx = 1$: Կատարելով փոփոխականի փոխարինում՝ $(x/\lambda)^\alpha = u$, $x^{\alpha-1}dx = \dfrac{\lambda^\alpha}{\alpha}du$, կստանանք՝ $\int\limits_{-\infty}^{\infty} f_X(x)dx = \int\limits_{0}^{\infty} e^{-u}du = 1$:

Վեյբուլի բաշխումամբ է բնութագրում սարքերի հուսալիության ժամկետը:

$c > 0$ և $\alpha > 0$ պարամետրերով **Պարետոյի օրենքով** բաշխված պատահական մեծության խտության ֆունկցիան է՝

$$f_X(x) = \begin{cases} \alpha c^\alpha / x^{\alpha+1}, & \text{երբ } x > c, \\ 0, & \text{երբ } x \le c: \end{cases}$$

Պարետոյի բաշխումը կիրառվում է որպես վիճակագրական դիտումների մոդել, օրինակ՝ եկամուտի բաշխման, քաղաքի բնակչության քանակի, լեզվի բառերի հաճախության համար:

---

m, σ^2 պարամետրերով, $-\infty < m < \infty$, $\sigma > 0$, **նորմալ** (կամ **Գաուսի**) **բաշխման օրենքի** համար ընդունված է կարճ գրառում՝ $X \sim \mathcal{N}(m, \sigma^2)$: Խտության ֆունկցիան է՝

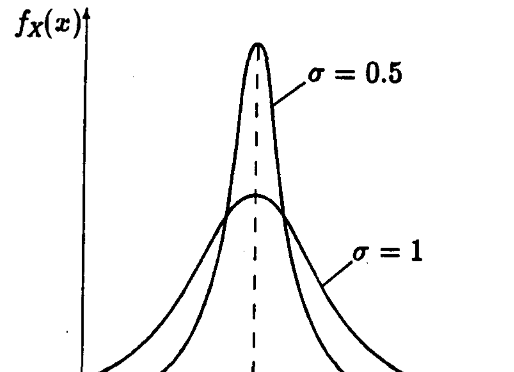
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty:$$

Նկար 8-ում պատկերված են խտության ֆունկցիայի գծապատկերները m պարամետրի միևնույն և σ^2 պարամետրի տարբեր արժեքների համար:

Կատարելով փոփոխականի փոխարինում՝ $(x-m)/\sigma = u$, $dx = \sigma du$, կհամոզվենք, որ՝

$$\int\limits_{-\infty}^{\infty} f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\{-u^2/2\}\, du = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1:$$

Այստեղ մենք օգտվել ենք մաթեմատիկական անալիզի դասընթացից հայտնի Պուասոնի ինտեգրալի արժեքից՝ $\int\limits_{-\infty}^{\infty} \exp\{-u^2/2\}\, du = \sqrt{2\pi}$:



Նկար 8: Նորմալ բաշխման խտության ֆունկցիայի գծապատկերները:

Եթե $m = 0$ և $\sigma^2 = 1$, այսինքն՝ $X \sim \mathcal{N}(0, 1)$, X-ի **բաշխումը** անվանում են **կանոնական նորմալ**:

Նորմալ բաշխումը մեծ կարևորություն ունի և հաճախ է կիրառվում տարբեր բնազավառներում: Դա բացատրվում է նրանով, որ որոշակի պայմանների դեպքում մեծ թվով պատահական մեծությունների գումարի բաշխումը մոտ է նորմալ բաշխմանը:

Եթե N-ը մեծ է, իսկ p-ն մոտ չէ 0-ին կամ 1-ին, $P_N(n)$ երկանդամային հավանականությունը հաշվելու համար օգտագործում են **Լապլասի բանաձևը**.

$$P_N(n) \approx \frac{\varphi(x)}{\sqrt{Npq}},$$

որտեղ

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \text{ և } x = \frac{n - Np}{\sqrt{Npq}} :$$

Աղյուսակում, որը բերված է հավելվածում, տրված են $\varphi(x)$ ֆունկցիայի արժեքները: Նշենք, որ $\varphi(x)$-ը զույգ ֆունկցիա է, և այդ պատճառով աղյուսակում տրվում են նրա արժեքները միայն $x > 0$ համար, և քանի որ $\varphi(x)$ ֆունկցիան շատ արագ է ձգտում զրոյի, ընդունում են, որ $\varphi(x) = 0$, երբ $x > 4$: Լապլասի բանաձևը տալիս է լավ մոտարկում, երբ p-ն մոտ է 1/2-ին: Լապլասի մոտարկման բանաձևի հիմնավորումը տալիս է

Մուավրի-Լապլասի մասնակի սահմանային թեորեմը: Եթե A պատահույթի հավանականությունը N անկախ փորձերից յուրաքանչյուրում նույնն է՝ հավասար p-ի, և եթե $N \to \infty$, ապա կամայական C դրական թվի համար հավասարաչափ ըստ բոլոր $|x| \le C$,

$$P_N(n) = \frac{1}{\sqrt{Npq}} \varphi(x)(1 + o(1)), \quad \text{որտեղ } x = \frac{n - Np}{\sqrt{Npq}},$$

$o(1)$-ով նշանակված է մի մեծություն, որը ձգտում է 0-ի, երբ $N \to \infty$:

Հաջորդ թեորեմն օգտակար է Բեռնուլիի փորձերում A պատահույթի n_1-ից մինչև n_2 անգամ հանդես գալու $\mathbf{P}_N\{n_1 \le n \le n_2\}$ հավանականությունը գտնելու համար:

Մուավրի-Լապլասի ինտեգրալ սահմանային թեորեմը: Եթե N անկախ փորձերում A պատահույթի հավանականությունը՝ p-ն, նույնն է, ապա հավասարաչափ a-ի և b-ի նկատմամբ $-\infty \le a \le b \le \infty$, երբ $N \to \infty$,

$$\mathbf{P}_N\{a\sqrt{Npq} + Np \le n \le b\sqrt{Npq} + Np\} =$$

$$= \mathbf{P}_N\{a \le \frac{n - Np}{\sqrt{Npq}} \le b\} \to \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left\{-\frac{x^2}{2}\right\} dx:$$

Նշենք, որ այս թեորեմը տալիս է կիրառությունների համար բավարար մոտարկում, երբ (Npq)-ն մեծ է 9-ից: Աջից գտնվող ինտեգրալի արժեքը կարելի է գտնել **Լապլասի**՝

$$x \ge 0, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left\{-\frac{t^2}{2}\right\} dt, \quad \Phi(-x) = -\Phi(x),$$

ֆունկցիայի Ա.2. աղյուսակի օգնությամբ (տե՛ս գրքի վերջում) և հետևյալ ներկայացման միջոցով, որը ճիշտ է և՛ դրական, և՛ բացասական a-երի ու b-երի դեպքում.

$$\int_a^b \varphi(x)dx = \int_0^b \varphi(x)dx - \int_0^a \varphi(x)dx = \Phi(b) - \Phi(a) :$$

$X \sim \mathcal{N}(m, \sigma^2)$ պատահական մեծության համար արտահայտենք (x_1, x_2) միջակայքի մեջ ընկնելու հավանականությունը Լապլասի ֆունկցիայի միջոցով՝

$$\mathbf{P}\{x_1 < X < x_2\} = \frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} \exp\left\{-\frac{(t-m)^2}{2\sigma^2}\right\} dt = \frac{1}{\sqrt{2\pi}}\int_{(x_1-m)/\sigma}^{(x_2-m)/\sigma} \exp\left\{-\frac{t^2}{2}\right\} dt =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x_2-m)/\sigma} \exp\left\{-\frac{t^2}{2}\right\} dt - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x_1-m)/\sigma} \exp\left\{-\frac{t^2}{2}\right\} dt = \tag{2}$$

$$= \Phi\left((x_2-m)/\sigma\right) - \Phi\left((x_1-m)/\sigma\right) :$$

Oգտվելով այս առնչությունից և $\Phi(x)$ ֆունկցիայի աղյուսակից՝ կարելի է որոշել $X \sim N(m, \sigma^2)$ պատահական մեծության՝ տրված միջակայքից արժեք ընդունելու հավանականությունը, այդ թվում $(m-\sigma, m+\sigma)$, $(m-2\sigma, m+2\sigma)$, $(m-3\sigma, m+3\sigma)$ միջակայքերի համար: Պարզվում է, որ այդ հավանականությունները, համապատասխանաբար, հավասար են՝ 0.68269, 0.95450, 0.99730: Քանի որ վերջին թիվը բավականաչափ մոտ է 1-ին, կիրառական խնդիրներում ընդունված է **«երեք սիգմաների կանոնը»**, որի համաձայն նորմալ բաշխված պատահական մեծության հնարավոր արժեքների տիրույթը գործնական տեսակետից համարվում է $(m-3\sigma, m+3\sigma)$ միջակայքը և ոչ թե ամբողջ իրական առանցքը:

<u>*Օրինակ*</u> 11: Հայտնի է, որ $X \sim \mathcal{N}(0, 1.44)$: Ճի՞շտ է արդյոք, որ $\mathbf{P}(|X| > 3) \geq 0.002$:

Լուծում: Այս հավանականությունը հնարավոր է գտնել $\Phi(x)$-ի աղյուսակի օգնությամբ: Համաձայն (2)-ի՝

$$\mathbf{P}(|X| > 3) = 1 - \mathbf{P}(-3 < X < 3) = 1 - \Phi\left((3-0)/1.2\right) + \Phi\left((-3-0)/1.2\right) =$$

$$= 1 - 2 \cdot \Phi(2.5) = 1 - 2 \cdot 0.4938 = 0.0124 :$$

Կարելի է այն գնահատել, նաև օգտվելով «երեք սիգմաների» կանոնից՝ $\mathbf{P}(|X| \leq 3\sigma) = \mathbf{P}(|X| \leq 3.6) = 0.9973$: Հետևաբար՝ $\mathbf{P}(|X| > 3) = 1 - \mathbf{P}(|X| \leq 3) \geq 1 - 0.9973 = 0.0027$: Երկու եղանակով էլ ստացվեց դրական պատասխան:

$c > 0$ և $-\infty < a < \infty$ պարամետրերով **Կոշիի բաշխման** խտության ֆունկցիան է՝

$$f_X(x) = \frac{1}{\pi} \cdot \frac{c}{c^2 + (x-a)^2}, \quad -\infty < x < \infty :$$

Կարելի է ցույց տալ, որ $m = 0$ և $\sigma^2 = 1$ պարամետրերով երկու անկախ գաուսյան բաշխված պատահական մեծությունների հարաբերությունը ենթարկվում է $a = 0$ և $c = 1$ պարամետրերով Կոշիի բաշխմանը:

$a > 0$, $\lambda > 0$ պարամետրերով **գամմա բաշխման** (ընդունված է կարճ գրառում՝ $X \sim \Gamma(a, \lambda)$) խտության ֆունկցիան է՝

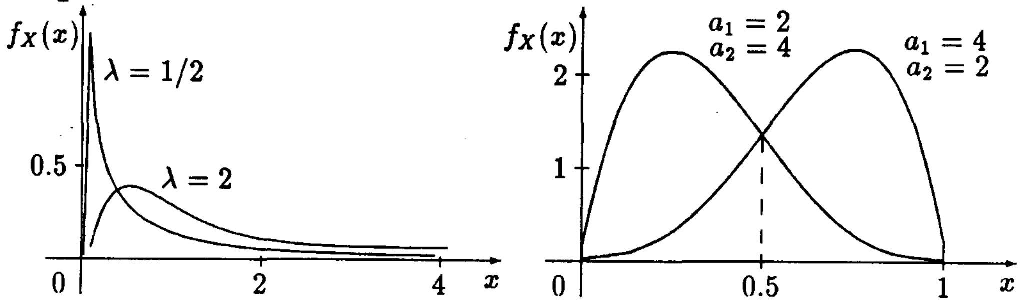
$$f_X(x) = \begin{cases} \dfrac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \exp\{-ax\}, & \text{երբ } x > 0, \\ 0, & \text{երբ } x \leq 0, \end{cases}$$

որտեղ $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} \exp\{-x\}\, dx$ հայտնի «գամմա» ֆունկցիան է:

Ստուգենք, որ $\int_{-\infty}^{\infty} f_X(x)dx = 1$: Կատարենք $ax = u$ փոփոխականի փոխարինում, կստանանք՝

$$\int_{-\infty}^{\infty} f_X(x)dx = \frac{a^\lambda}{\Gamma(\lambda)} \frac{1}{a^\lambda} \int_0^\infty u^{\lambda-1} \exp\{-u\} du = 1 :$$

Գամմա բաշխումը հաճախ է հանդիպում հատկապես մաթեմատիկական վիճակագրությունում: Նկատենք, որ $\lambda = 1$ դեպքում ստանում ենք a պարամետրով ցուցչային բաշխումը:



Նկար 9: Գամմա և բետա բաշխումների խտության ֆունկցիաները:

$a_1 > 0$ և $a_2 > 0$ պարամետրերով **բետա բաշխման** (ընդունված է կարճ գրառում՝ $X \sim \beta(a_1, a_2)$) խտության ֆունկցիան է՝

$$f_X(x) = \begin{cases} \dfrac{x^{a_1-1}(1-x)^{a_2-1}}{B(a_1,a_2)}, & \text{երբ } 0 < x < 1, \\ 0, & x - \text{ի մյուս արժեքների դեպքում}, \end{cases}$$

որտեղ

$$B(a_1, a_2) = \int_0^1 x^{a_1-1}(1-x)^{a_2-1}dx$$

հայտնի «բետա» ֆունկցիան է: Նշենք, որ $B(a_1, a_2) = \Gamma(a_1)\Gamma(a_2)/\Gamma(a_1 + a_2)$: Երբ $a_1 = a_2 = 1$, բետա բաշխումը վերածվում է [0,1] միջակայքում հավասարաչափ բաշխման:

2.5. Բազմաչափ պատահական մեծություններ

Միևնույն հավանականային տարածության վրա կարելի է դիտարկել մի քանի պատահական մեծություններ: Դրա անհրաժեշտությունը ծագում է, երբ օբյեկտները բնութագրվում են մի քանի հատկանիշներով: Օրինակ, որևէ բնակավայրի ընտանիքների սննդի, հագուստի, տրանսպորտի ծախսերի բնութագրիչները կարող են ուսումնասիրվել որպես մեկ պատահական վեկտոր:

Դիցուք $\{\Omega, \mathcal{F}, \mathcal{P}\}$-ն հավանականային տարածություն է և X_1-ը, X_2-ը, ..., X_M-ը, Ω-ի վրա որոշված պատահական մեծություններ են: M հատ պատահական մեծությունների կարգավորված համախմբությունը կոչվում է **M-չափանի պատահական վեկտոր** կամ **M-չափանի պատահական մեծություն** և նշանակվում է $\mathbf{X} = (X_1, X_2, \ldots, X_M)$:

$\mathbf{X}$-ը արտապատկերում է Ω-ն $\mathcal{R}^M$-ի մեջ: Դիցուք x_1-ը, x_2-ը, ..., x_M-ը իրական թվեր են: Քանի որ

$$\{\omega : X_1(\omega) < x_1,\ \ X_2(\omega) < x_2, \ldots, X_M(\omega) < x_M\} = \bigcap_{m=1}^{M} \{X_m < x_m\}$$

բազմությունը պատկանում է $\mathcal{F}$-ին, այսինքն պատահույթ է, ապա որոշված է նրա հավանականությունը:

M-չափանի պատահական վեկտորի բաշխման ֆունկցիան նշանակվում է $F_{\mathbf{X}}(x_1, x_2, \ldots, x_M)$՝

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_M) = \mathbf{P}\{\omega : X_1(\omega) < x_1,\ X_2(\omega) < x_2,\ \ldots,\ X_M(\omega) < x_M\} = \\ = \mathbf{P}\{X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_M < x_M\}:$$

Պատահական վեկտորի բաշխման ֆունկցիան օժտված է հետևյալ հատկություններով.

1. $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)$-ը միընթաց չնվազող է ըստ յուրաքանչյուր x_m-ի, $m=\overline{1,M}$:
2. $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)\to 0$, երբ գոնե մեկ x_m-ը ձգտում է $-\infty$, և $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)\to 1$, եթե բոլոր x_m-րը ձգտում են ∞, $m=\overline{1,M}$:
3. $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)$-ը ձախից անընդհատ է ըստ յուրաքանչյուր x_m-ի, $m=\overline{1,M}$:
4. Նշանակենք և կոչենք **տարբերական օպերատոր**՝

$$\Delta_{a_m,b_m}F_{\mathbf{X}}(x_1,x_2,\dots,x_M)=$$
$$=F_{\mathbf{X}}(x_1,x_2,\dots,x_{m-1},b_m,x_{m+1},\dots,x_M)-F_{\mathbf{X}}(x_1,x_2,\dots,x_{m-1},a_m,x_{m+1},\dots,x_M):$$

Կամայական $a=(a_1,a_2,\dots,a_M)$ և $b=(b_1,b_2,\dots,b_M)$, $a_m\le b_m$, $m=\overline{1,M}$, վեկտորների համար

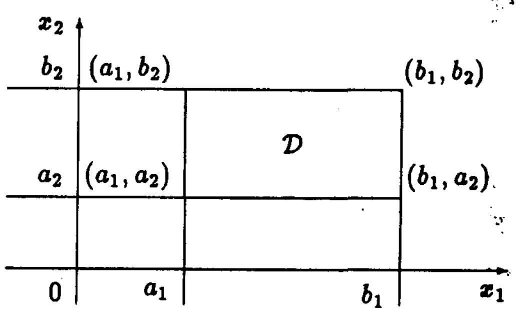
$$\Delta_{a_1,b_1}\Delta_{a_2,b_2}\dots\Delta_{a_M,b_M}F_{\mathbf{X}}(x_1,x_2,\dots,x_M)\ge 0,$$

որտեղ ձախից M-չափանի՝ (a_m,b_m), $m=\overline{1,M}$, կողմերով, ուղղանկյան մեջ **X** վեկտորի արժեք ընդունելու հավանականությունն է:

Ի տարբերություն միաչափ դեպքի, այստեղ 1-3 հատկություններին ավելացել է 4-րդ հատկությունը, որի ապացուցումը պարզության համար անցկացնենք $M=2$ դեպքում՝

$$\Delta_{a_1,b_1}\Delta_{a_2,b_2}F_{\mathbf{X}}(x_1,x_2)=\Delta_{a_1,b_1}(F_{\mathbf{X}}(x,b_2)-F_{\mathbf{X}}(x,a_2))=$$
$$=\Delta_{a_1,b_1}F_{\mathbf{X}}(x,b_2)-\Delta_{a_1,b_1}F_{\mathbf{X}}(x,a_2)=$$
$$=F_{\mathbf{X}}(b_1,b_2)-F_{\mathbf{X}}(a_1,b_2)-F_{\mathbf{X}}(b_1,a_2)+F_{\mathbf{X}}(a_1,a_2)=$$
$$=\mathbf{P}\{a_1\le X_1<b_1,X_2<b_2\}-\mathbf{P}\{a_1\le X_1<b_1,X_2<a_2\}=$$
$$=\mathbf{P}\{a_1\le X_1<b_1,a_2\le X_2<b_2\}\ge 0:$$

Վերջին հավանականությունը (X_1,X_2) վեկտորի նկար 10-ում ստվերագծված ուղղանկյան մեջ ընկնելու հավանականությունն է:



Նկար 10: Երկչափ պատահական վեկտորի արժեքների ուղղանկյուն տիրույթը:

Կարելի է ապացուցել, որ 1-ից 4 հատկություններով օժտված կամայական $F(x_1,x_2,\dots,x_M)$, $x_1,x_2,\dots,x_M\in\mathcal{R}$, ֆունկցիա որևէ **X** պատահական վեկտորի բաշխման ֆունկցիա է՝ $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)=F(x_1,x_2,\dots,x_M)$:

Օրինակ 12: Դիցուք (X,Y) պատահական վեկտորի բաշխման ֆունկցիան է՝

$$F_{\mathbf{X}}(x,y)=\begin{cases}1-2^{-x}-2^{-y}+2^{-x-y}, & \text{եթե } x\ge 0,\ y\ge 0,\\ 0, & \text{մնացած } x\text{-երի և } y\text{-երի համար:}\end{cases}$$

Գտնել (X,Y) վեկտորի $x=1$, $x=2$, $y=3$, $y=5$ ուղիղներով սահմանափակված S ուղղանկյանը պատկանելու հավանականությունը:

Լուծում: Նախևառաջ ստուգենք, որ $F_{\mathbf{X}}(x,y)$-ը իսկապես բաշխման ֆունկցիա է, այսինքն՝ այն բավարարում է 1 – 4 պայմաններին: Քանի որ 2^{-u}, $u>0$ ցուցչային ֆունկցիան նվազող է և ձգտում է զրոյի, երբ $u\to\infty$, ապա 1 - 3 հատկությունները տեղի ունեն: Ստուգենք չորրորդը: Ունենք՝

$$F_{\mathbf{X}}(b_1,b_2)-F_{\mathbf{X}}(a_1,b_2)-F_{\mathbf{X}}(b_1,a_2)+F_{\mathbf{X}}(a_1,a_2)=1-2^{-b_1}-2^{-b_2}+2^{-b_1-b_2}-1+2^{-a_1}+2^{-b_2}-$$
$$-2^{-a_1-b_2}-1+2^{-b_1}+2^{-a_2}-2^{-b_1-a_2}+1-2^{-a_1}-2^{-a_2}+2^{-a_1-a_2}=$$
$$=+2^{-b_1-b_2}-2^{-a_1-b_2}-2^{-b_1-a_2}+2^{-a_1-a_2}=(2^{-b_1}-2^{-a_1})(2^{-b_2}-2^{-a_2})\geq 0:$$

Այժմ օգտվենք հետևյալ բանաձևից՝ $\mathbf{P}\{(X,Y)\in S\}=F_{\mathbf{X}}(2,5)-F_{\mathbf{X}}(1,5)-F_{\mathbf{X}}(2,3)+F_{\mathbf{X}}(1,3):$
Հաշվենք բաշխման ֆունկցիայի արժեքները նշված կետերում.

$$F_{\mathbf{X}}(2,5)=\frac{128-32-4+1}{128}=\frac{93}{128},\; F_{\mathbf{X}}(1,5)=\frac{64-32-2+1}{64}=\frac{31}{64},$$
$$F_{\mathbf{X}}(1,3)=\frac{16-8-2+1}{16}=\frac{7}{16},\; F_{\mathbf{X}}(2,3)=\frac{32-8-4+1}{32}=\frac{21}{32}:$$

Այսպիսով՝ $\mathbf{P}\{(X,Y)\in S\}=\frac{93}{128}-\frac{31}{64}-\frac{21}{32}+\frac{7}{16}=\frac{3}{128}:$

Օրինակ 13: $F_{\mathbf{X}}(x_1,x_2,\ldots,x_M)$ ֆունկցիան կարող է բավարարել 1–3 պայմաններին, բայց չորրորդ պայմանին չբավարարի և, հետևաբար, չլինի բաշխման ֆունկցիա:

Լուծում: Դիցուք

$$F_{\mathbf{X}}(x_1,x_2)=\begin{cases}1, & \text{եթե } x_1>0,\; x_2>0,\; x_1+x_2>1,\\ 0, & \text{մնացած դեպքերում}:\end{cases}$$

Դժվար չէ համոզվել, որ $F_{\mathbf{X}}(x_1,x_2)$ ֆունկցիան բավարարում է 1-3 պայմաններին: Ցույց տանք, որ այդ ֆունկցիան չի բավարարում չորրորդ պայմանին: Վերցնելով $a_1=1/2,\ b_1=1,\ a_2=1/2,\ b_2=1$՝ կստանանք՝

$$\Delta_{a_1,b_1}\Delta_{a_2,b_2}F_{\mathbf{X}}(x_1,x_2)=F_{\mathbf{X}}(b_1,b_2)-F_{\mathbf{X}}(a_1,b_2)-F_{\mathbf{X}}(b_1,a_2)+F_{\mathbf{X}}(a_1,a_2)=$$
$$=F_{\mathbf{X}}(1,1)-F_{\mathbf{X}}(1/2,1)-F_{\mathbf{X}}(1,1/2)+F_{\mathbf{X}}(1/2,1/2)=1-1-1+0=-1<0:$$

Նշենք բազմաչափ բաշխման ֆունկցիայի ևս երկու կարևոր հատկություններ՝

5. $F_{X_1,\ldots,X_m,\ldots,X_j,\ldots,X_M}(x_1,\ldots,x_m,\ldots,x_j,\ldots,x_M)=$
$=F_{X_1,\ldots,X_j,\ldots,X_m,\ldots,X_M}(x_1,\ldots,x_j,\ldots,x_m,\ldots,x_M)$ (**համաչափության հատկությունը**),
6. $\lim\limits_{x_M\to\infty}F_{X_1,X_2,\ldots,X_M}(x_1,x_2,\ldots,x_M)=F_{X_1,X_2,\ldots,X_{M-1}}(x_1,x_2,\ldots,x_{M-1})$ (**համաձայնեցվածության հատկությունը**):

Վերջին հատկությունը հաճախ է օգտագործվում, քանի որ հնարավոր է դարձնում վեկտորի բաշխման ֆունկցիայից ստանալ կամայական ենթավեկտորների բաշխման ֆունկցիաները:

Դիտարկենք բաշխումների երկու առավել կարևոր դասերը՝ ընդհատ և անընդհատ: Պարզության համար դիտարկենք $M=2$ դեպքը:

$\mathbf{X}=(X_1,X_2)$ պատահական **վեկտորը** կոչվում է **ընդհատ** (կամ ասում են $\mathbf{X}$-ը ունի **ընդհատ բաշխում**), եթե այդ վեկտորի բաղադրիչները ընդհատ պատահական մեծություններ են: $\mathbf{X}=(X_1,X_2)$ ընդհատ պատահական վեկտորի բաշխման օրենքը ներկայացվում է հետևյալ **երկչափ բաշխման աղյուսակով**՝

$X_2 \backslash X_1$	x_{11}	x_{12}	$\ldots$	x_{1k}	$\ldots$	x_{1K}	
x_{21}	p_{11}	p_{21}	$\ldots$	p_{k1}	$\ldots$	p_{K1}	$p_{\cdot 1}$
x_{22}	p_{12}	p_{22}	$\ldots$	p_{k2}	$\ldots$	p_{K2}	$p_{\cdot 2}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_{2l}	p_{1l}	p_{2l}	$\ldots$	p_{kl}	$\ldots$	p_{Kl}	$p_{\cdot l}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_{2L}	p_{1L}	p_{2L}	$\ldots$	p_{kL}	$\ldots$	p_{KL}	$p_{\cdot L}$
	$p_{1\cdot}$	$p_{2\cdot}$	$\ldots$	$p_{k\cdot}$	$\ldots$	$p_{K\cdot}$	1

որտեղ (x_{1k}, x_{2l}), $k = \overline{1,K}$, $l = \overline{1,L}$, (X_1, X_2) պատահական վեկտորի արժեքներն են (արժեքների թիվը կարող է լինել վերջավոր կամ հաշվելի), իսկ համապատասխան զույգերի համատեղ հավանականություններն են՝ $p_{kl} = \mathbf{P}\{X_1 = x_{1k}, X_2 = x_{2l}\}$, ընդ որում $p_{kl} \geq 0$ և $\sum\limits_{k,l} p_{kl} = 1$: Նշենք նաև, որ

$$\mathbf{P}\{X_1 = x_{1k}\} = \sum_l p_{kl} = p_{k.}, \quad \mathbf{P}\{X_2 = x_{2l}\} = \sum_k p_{kl} = p_{.l}, \quad k = \overline{1,K}, \ l = \overline{1,L}:$$

Երկչափ պատահական վեկտորի բաշխման աղյուսակում աջից ավելացված է մի սյունակ, որում գրված են X_2-ի **մասնատեղ բաշխման հավանականությունները**, իսկ ներքևից մի տող, որում գրված են X_1-ի մասնատեղ բաշխման հավանականությունները: Այդ գրառման հետ կապված մասնատեղ բաշխման հավանականությունները երբեմն կոչում են **լուսանցքային** (marginal) **հավանականություններ**:

<u>*Օրինակ*</u> 14: $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի բաշխման աղյուսակն է՝

$X_2 \backslash X_1$	0	1
−1	0.1	0.2
0	0.2	0.3
1	0	0.2

: Գտնենք X_1-ի և X_2-ի բաշխման աղյուսակները:

Լուծում: X_1-ի հնարավոր արժեքներն են 0 և 1: Գտնենք համապատասխան լուսանցքային հավանականությունները՝ $\mathbf{P}\{X_1 = 0\} = 0.1 + 0.2 + 0 = 0.3$, $\mathbf{P}\{X_1 = 1\} = 0.2 + 0.3 + 0.2 = 0.7$:

Հետևաբար՝ X_1-ի բաշխման աղյուսակն է՝

X_1	0	1
p	0.3	0.7

:

X_2-ի հնարավոր արժեքներն են -1, 0 և 1: Գտնենք համապատասխան հավանականությունները՝ $\mathbf{P}\{X_2 = -1\} = 0.1 + 0.2 = 0.3$, $\mathbf{P}\{X_2 = 0\} = 0.2 + 0.3 = 0.5$, $\mathbf{P}\{X_2 = 1\} = 0 + 0.2 = 0.2$:

Հետևաբար՝ X_2-ի բաշխման աղյուսակն է՝

X_2	−1	0	1
p	0.3	0.5	0.2

:

Այսպիսով՝ լուսանցքային հավանականություններով հանդերձ երկչափ $\mathbf{X}$ պատահական վեկտորի բաշխման աղյուսակը կլինի՝

$X_2 \backslash X_1$	0	1	
−1	0.1	0.2	0.3
0	0.2	0.3	0.5
1	0	0.2	0.2
	0.3	0.7	1

<u>*Օրինակ*</u> 15: Հետազոտենք բազմակի փորձերի շարքը: Ամեն մի փորձում հնարավոր են M ելքեր՝ A_1, A_2, …, A_M, $M \geq 2$, համապատասխանաբար $p_1, p_2, \ldots, p_M$ հավանականություններով, $0 \leq p_m \leq 1$, $\sum\limits_{m=1}^{M} p_m = 1$: Գտնենք N անկախ փորձերի արդյունքում A_m պատահույթների n_m անգամ, $m = \overline{1,M}$, տեղի ունենալու $P_N(n_1, \ldots, n_M)$ հավանականությունը:

Լուծում: Հաշվի առնելով, որ

$$C_N^{n_1} C_{N-n_1}^{n_2} \ldots C_{N-n_1-n_2-\cdots-n_{M-1}}^{n_M} = \frac{N!}{n_1! n_2! \ldots n_M!},$$

դժվար չէ համոզվել, որ

$$P_N(n_1, \ldots, n_M) = \frac{N!}{n_1! n_2! \ldots n_M!} p_1^{n_1} p_2^{n_2} \ldots p_M^{n_M}:$$

Դիցուք X_m-ը N անկախ բազմակի փորձերում A_m պատահույթի իրականացումների թիվն է, $m = \overline{1,M}$: Եթե $\mathbf{P}\{A_m\} = p_m$, $m = \overline{1,M}$, ապա $\mathbf{X} = (X_1, X_2, \ldots, X_M)$ պատահական վեկտորն ունի **բազմանդամային բաշխում**՝

$$\mathbf{P}\{X_1 = n_1, X_2 = n_2, \ldots, X_M = n_M\} = P_N(n_1, \ldots, n_M) = \frac{N!}{n_1! n_2! \ldots n_M!} p_1^{n_1} p_2^{n_2} \ldots p_M^{n_M}:$$

Օրինակ 16: Հատվածը բաժանված է 4 մասի՝ 1, 2, 3, 4 համամասնությամբ: Հատվածի վրա պատահականորեն նետվում է 8 կետ: Գտնել առաջին մասի մեջ 3 կետ, երկրորդ մասի մեջ 2 կետ և չորրորդ մասի մեջ 3 կետ ընկնելու հավանականությունը:

Լուծում: Նշանակենք A_k-ով k-րդ մասի մեջ մեկ կետ ընկնելու պատահույթը, $k = \overline{1,4}$: Հատվածների երկարությունների հարաբերությունից, հավանականության երկրաչափական սահմանման համաձայն, գտնում ենք՝ $p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.3$, $p_4 = 0.4$: Որոնելի հավանականությունը կլինի՝

$$\mathbf{P}_8(3,2,0,3) = \frac{8!}{3!2!0!3!}(0.1)^3(0.2)^2(0.3)^0(0.4)^3 \approx 0.0013:$$

$\mathbf{X} = (X_1, X_2)$ պատահական վեկտորը կոչվում է **բացարձակ անընդհատ** (ասում են նաև՝ X-ի **բաշխումը անընդհատ** է), եթե գոյություն ունի այնպիսի ոչ բացասական $f_{\mathbf{X}}(x_1, x_2)$ ֆունկցիա, որ $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի $F_{\mathbf{X}}(x_1, x_2)$ բաշխման ֆունկցիան ներկայացվում է հետևյալ կերպ՝

$$F_{\mathbf{X}}(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{\mathbf{X}}(u_1, u_2)du_1du_2:$$

Իհարկե՝ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, u_2)du_1du_2 = 1$, քանի որ $F_{\mathbf{X}}(\infty, \infty) = 1$:

$f_{\mathbf{X}}(x_1, x_2)$ ֆունկցիան կոչվում է $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի հավանականությունների **բաշխման խտության ֆունկցիա** (կամ **բաշխման խտություն**): Նկատենք, որ $f_X(x_1, x_2)$-ի անընդհատության կետերում՝

$$f_{\mathbf{X}}(x_1, x_2) = \frac{\partial^2 F_{\mathbf{X}}(x_1, x_2)}{\partial x_1 \partial x_2}:$$

Եթե հայտնի է $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի $f_{\mathbf{X}}(x_1, x_2)$ բաշխման խտությունը, ապա կարելի է գտնել X_1 և X_2 պատահական մեծությունների համապատասխան $f_{X_1}(x_1)$ և $f_{X_2}(x_2)$ լուսանցքային խտության ֆունկցիաները:

Ըստ բազմաչափ բաշխման ֆունկցիայի համաձայնեցվածության հատկության՝

$$F_{X_1}(x_1) = \lim_{x_2\to\infty} F_{\mathbf{X}}(x_1, x_2) = \lim_{x_2\to\infty}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{\mathbf{X}}(u_1, u_2)du_1du_2 = \int_{-\infty}^{x_1}\int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, u_2)du_1du_2,$$

$$F_{X_2}(x_2) = \lim_{x_1\to\infty} F_{\mathbf{X}}(x_1, x_2) = \lim_{x_1\to\infty}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{\mathbf{X}}(u_1, u_2)du_1du_2 = \int_{-\infty}^{x_2}\int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, u_2)du_1du_2:$$

Դիֆերենցելով առաջին հավասարությունն ըստ x_1-ի, իսկ երկրորդը՝ ըստ x_2-ի, կստանանք՝

$$f_{X_1}(x_1) = F'_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, u_2)du_2, \quad f_{X_2}(x_2) = F'_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, x_2)du_1:$$

Դիտարկենք անընդհատ բազմաչափ պատահական վեկտորների կարևորագույն տեսակները:

Ասում են, որ $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորը **հավասարաչափ** է **բաշխված** հարթության $\mathcal{D}$ տիրույթում, եթե նրա խտության ֆունկցիան է՝

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} c, & \text{եթե } (x_1, x_2) \in \mathcal{D}, \\ 0, & \text{եթե } (x_1, x_2) \notin \mathcal{D}: \end{cases}$$

Քանի որ $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2)dx_1dx_2 = 1$, ապա $c = 1/S_{\mathcal{D}}$, որտեղ $S_{\mathcal{D}}$-ն $\mathcal{D}$ տիրույթի մակերեսն է:

Օրինակ 17: $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորը հավասարաչափ բաշխված է $(0,0)$, $(0,1)$, $(1,0)$ գագաթներ ունեցող $\mathcal{D}$ եռանկյան ներսում (տե՛ս նկար 11): Գտնենք $f_{X_1}(x_1)$ և $f_{X_2}(x_2)$ լուսանցքային խտությունները:

Լուծում: Քանի որ $S_{\mathcal{D}} = 1/2$, ապա $c = 2$, և $\mathbf{X}$-ի խտության ֆունկցիան է՝

$$f_{\mathbf{X}}(x_1,x_2) = \begin{cases} 2, & \text{եթե } (x_1,x_2) \in \mathcal{D}, \\ 0, & \text{եթե } (x_1,x_2) \notin \mathcal{D}: \end{cases}$$

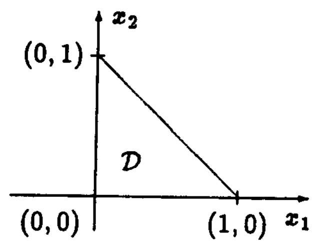
Ուստի՝

$$f_{X_1}(x_1) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2)dx_2 =$$

$$= 2\int\limits_{0}^{1-x_1} dx_2 = 2(1-x_1), \quad 0 \le x_1 \le 1,$$

$$f_{X_2}(x_2) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2)dx_1 =$$

$$= 2\int\limits_{0}^{1-x_2} dx_1 = 2(1-x_2), \quad 0 \le x_2 \le 1:$$



Նկար 11:

Պարզ է, որ $[0,1]$ միջակայքից դուրս $f_{X_1}(x_1)$-ը և $f_{X_2}(x_2)$-ը հավասար են 0-ի:

Ասում են, որ $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորը բաշխված է ըստ **երկչափ նորմալ** (կամ **Գաուսի**) **օրենքի**՝ m_1, m_2, σ_1^2, σ_2^2, r պարամետրերով, $-\infty < m_1, m_2 < \infty$, $\sigma_1 > 0$, $\sigma_2 > 0$, $|r| < 1$, եթե նրա խտության ֆունկցիան է՝

$$f_{\mathbf{X}}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2r\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\}:$$

Ընդունված է կարճ գրառում՝ $\mathbf{X} \sim \mathcal{N}(m_1, m_2, \sigma_1^2, \sigma_2^2, r)$:

Օրինակ 18: Դիցուք $\mathbf{X} \sim \mathcal{N}(0,0,1,1,r)$: Գտնենք $f_{X_1}(x_1)$ և $f_{X_2}(x_2)$ խտությունները:

Լուծում: $\mathbf{X}$-ի խտության ֆունկցիան է՝

$$f_{\mathbf{X}}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[x_1^2 - 2rx_1x_2 + x_2^2\right]\right\}:$$

Հետևաբար՝

$$f_{X_1}(x_1) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2)dx_2 = \frac{1}{2\pi\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{x_1^2 - 2rx_1x_2 + x_2^2}{2(1-r^2)}\right\}dx_2 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{(x_2-rx_1)^2 + x_1^2 - r^2x_1^2}{2(1-r^2)}\right\}dx_2 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{(x_2-rx_1)^2}{2(1-r^2)} - \frac{x_1^2}{2}\right\}dx_2 =$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x_1^2}{2}\right\}\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2}\left(\frac{x_2-rx_1}{\sqrt{1-r^2}}\right)^2\right\}dx_2:$$

Վերջին ինտեգրալի ընդինտեգրալ արտահայտությունը $m = rx_1$, $\sigma^2 = 1 - r^2$ պարամետրերով նորմալ բաշխման խտության ֆունկցիան է, ուստի այդ ինտեգրալը հավասար է 1-ի: Այստեղից՝

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x_1^2}{2}\right\}:$$

Այսպիսով, $X_1 \sim \mathcal{N}(0,1)$, նման ձևով կստանանք, որ $X_2 \sim \mathcal{N}(0,1)$:

Կարելի է ցույց տալ, որ եթե $\mathbf{X} \sim \mathcal{N}(m_1,m_2,\sigma_1^2,\sigma_2^2,r)$, ապա $X_1 \sim \mathcal{N}(m_1,\sigma_1^2)$ և $X_2 \sim \mathcal{N}(m_2,\sigma_2^2)$:

2.6. Պատահական մեծությունների անկախությունը, պայմանական բաշխումներ

X_1, X_2, ..., X_M պատահական մեծությունները կոչվում են **անկախ**, եթե $\mathbf{X} = (X_1,X_2,\dots,X_M)$ պատահական վեկտորի $F_{\mathbf{X}}(x_1,x_2,\dots,x_M)$ բաշխման ֆունկցիան հավասար է $X_1, X_2,\dots,X_M$ բաղադրիչների բաշխման ֆունկցիաների արտադրյալին՝

$$F_{\mathbf{X}}(x_1,x_2,\dots,x_M) = F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_M}(x_M)\ :$$

Ընդհատ X_1, X_2, ..., X_M, պատահական մեծությունների անկախության համար անհրաժեշտ է և բավարար, որպեսզի X_m պատահական մեծությունների x_{mk}, $k = \overline{1,K_m}$, $m = \overline{1,M}$ արժեքների բոլոր հնարավոր տարբերակների դեպքում՝

$$\begin{aligned}&\mathbf{P}\{X_1 = x_{1k}, X_2 = x_{2k},\dots,X_M = x_{Mk}\} = \\ &= \mathbf{P}\{X_1 = x_{1k}\}\mathbf{P}\{X_2 = x_{2k}\}\dots\mathbf{P}\{X_M = x_{Mk}\}\ :\end{aligned}$$

Անընդհատ X_1, X_2, ..., X_M պատահական մեծությունների անկախության համար անհրաժեշտ է և բավարար, որպեսզի կամայական x_1, x_2, ..., x_M իրական թվերի համար $\mathbf{X} = (X_1,X_2,\dots,X_M)$ վեկտորի խտության ֆունկցիան հավասար լինի բաղադրիչների խտության ֆունկցիաների արտադրյալին՝

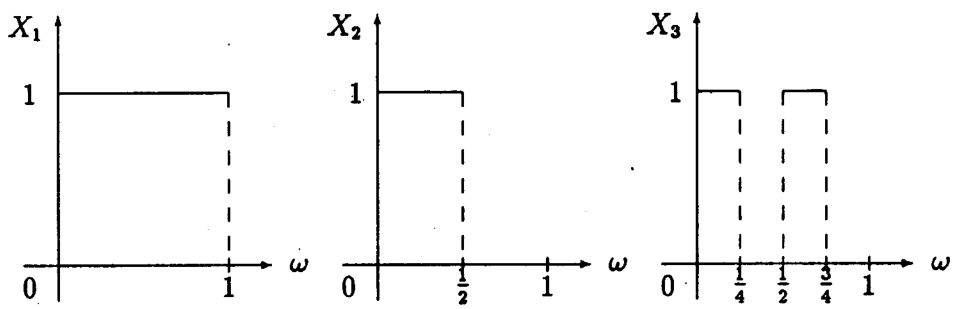
$$f_{\mathbf{X}}(x_1,x_2,\dots,x_M) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_N}(x_M)\ :$$

Օրինակ 19: Դիցուք $\Omega = [0,1]$: Դիտարկենք X_1, X_2, X_3 ընդհատ պատահական մեծությունները՝

$$X_1(\omega) = 1, \quad \omega \in [0,1],$$

$$X_2(\omega) = \begin{cases} 1, & \omega \in [0,1/2], \\ 0, & \omega \notin [0,1/2], \end{cases} \qquad X_3(\omega) = \begin{cases} 1, & \omega \in [0,1/4] \cup [1/2,3/4], \\ 0, & \text{հակառակ դեպքում}\ : \end{cases}$$

Ցույց տալ, որ X_1, X_2, X_3 պատահական մեծություններն անկախ են:



Նկար 12: X_1, X_2, X_3 պատահական մեծությունները:

Լուծում: Նախ կազմենք դիտարկվող պատահական մեծությունների բաշխման աղյուսակները՝

X_1	1
p	1

,

X_2	0	1
p	1/2	1/2

,

X_3	0	1
p	1/2	1/2

:

Այնուհետև նկատենք, որ $\mathbf{P}\{X_1 = 1, X_2 = 1, X_3 = 1\} = 1/4$: Հետևաբար՝

$$\mathbf{P}\{X_1 = 1, X_2 = 1, X_3 = 1\} = \mathbf{P}\{X_1 = 1\}\mathbf{P}\{X_2 = 1\}\mathbf{P}\{X_3 = 1\}:$$

Ստուգենք մեկ ուրիշ դեպք՝ $\mathbf{P}\{X_1 = 1, X_2 = 0, X_3 = 1\} = 1/4$, տեսնում ենք, որ

$$\mathbf{P}\{X_1 = 1, X_2 = 0, X_3 = 1\} = \mathbf{P}\{X_1 = 1\}\mathbf{P}\{X_2 = 0\}\mathbf{P}\{X_3 = 1\}:$$

Մյուս դեպքերը ստուգվում են նույն ձևով, ուրեմն X_1, X_2, X_3 պատահական մեծությունները անկախ են:

Օրինակ 20: $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորը հավասարաչափ բաշխված է $(0,0)$, $(0,2)$, $(2,0)$, $(2,2)$ գագաթներ ունեցող քառակուսու ներսում: Ցույց տալ, որ X_1 և X_2 պատահական մեծությունները անկախ են:

Լուծում: Քանի որ $S_{\mathcal{D}} = 4$, հետևաբար $c = 1/4$ և $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի խտության ֆունկցիան է՝

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} 1/4, & \text{եթե } (x_1, x_2) \in \mathcal{D}, \\ 0, & \text{եթե } (x_1, x_2) \notin \mathcal{D}: \end{cases}$$

Գտնենք $f_{X_1}(x_1)$-ը և $f_{X_2}(x_2)$-ը՝

$$f_{X_1}(x_1) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2)dx_2 = \frac{1}{4}\int\limits_0^2 dx_2 = \frac{1}{2}, \quad x_1 \in [0,2],$$

$$f_{X_2}(x_2) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2)dx_1 = \frac{1}{4}\int\limits_0^2 dx_1 = \frac{1}{2}, \quad x_2 \in [0,2]:$$

Այսպիսով, քանի որ $f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, հետևաբար X_1-ը և X_2-ը անկախ են:

Երկչափ $\mathbf{X} = (X_1, X_2)$ պատահական վեկտորի համար ներմուծենք **պայմանական** բաշխման գաղափարը: Դիտարկենք սկզբից ընդհատ բաշխում ունեցող վեկտորը, որի բաշխման օրենքը տրված է երկչափ բաշխման աղյուսակով, $\mathbf{P}\{X_1 = x_{1k}, X_2 = x_{2l}\} = p_{kl}$, $k = \overline{1,K}$, $l = \overline{1,L}$, հավանականություններով:

Երբեմն հարկ է լինում դիտարկել X_2 պատահական մեծության պայմանական բաշխումը, երբ հայտնի է, որ, ասենք, $X_1 = x_{1k}$: Դիցուք $\mathbf{P}\{X_1 = x_{1k}\} = p_{k.} = \sum\limits_{l=1}^{L} p_{kl} > 0$: Նշանակենք $\mathbf{P}\{X_2 = x_{2l} | X_1 = x_{1k}\} = p_{l|k}$:

Պարզ է, որ $p_{l|k} = p_{kl}/p_{k.}$ և $\sum\limits_l p_{l|k} = 1$, $k = \overline{1,K}$, $l = \overline{1,L}$:

Երբ $p_{.l} > 0$, նույն ձևով նշանակելով $\mathbf{P}\{X_1 = x_{1k} | X_2 = x_{2l}\} = p_{k|l}$, կարող ենք գտնել, որ $p_{k|l} = p_{kl}/p_{.l}$, $\sum\limits_k p_{k|l} = 1$, $k = \overline{1,K}$, $l = \overline{1,L}$: Այդ հավանականությունները նույնպես հարմար է ներկայացնել աղյուսակով, օրինակ, $p_{l|k}$ հավանականություններինը կլինի՝

$X_2 \backslash X_1$	x_{11}	x_{12}	$\dots$	x_{1k}	$\dots$	x_{1K}
x_{21}	$p_{1\|1}$	$p_{2\|1}$	$\dots$	$p_{k\|1}$	$\dots$	$p_{K\|1}$
x_{22}	$p_{1\|2}$	$p_{2\|2}$	$\dots$	$p_{k\|2}$	$\dots$	$p_{2\|K}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_{2l}	$p_{1\|l}$	$p_{2\|l}$	$\dots$	$p_{k\|l}$	$\dots$	$p_{K\|l}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_{2L}	$p_{1\|L}$	$p_{2\|L}$	$\dots$	$p_{k\|L}$	$\dots$	$p_{K\|L}$

:

Այսպիսի **մատրիցը** կոչվում է **ստոխաստիկ**, նրա ամեն մի տողի տարրերի գումարը հավասար է մեկի: Եթե գրվի $p_{k|l}$ հավանականությունների մատրիցը, ապա նրա սյուներով գումարները կլինեն հավասար մեկի:

Օրինակ 21: Օրինակ 14-ի տվյալներով կազմել X_2-ի ըստ X_1-ի պայմանական բաշխման աղյուսակը:

Լուծում: Կատարելով պարզ հաշվարկներ, օրինակ, $p_{-1|0} = 0.1/0.3 = 1/3$, կստանանք X_2

պատահական մեծության պայմանական բաշխման աղյուսակը X_1-ի տրված արժեքներից կախված՝

$X_2 \backslash X_1$	0	1
−1	1/3	2/7
0	2/3	3/7
1	0	2/7

:

Անցնենք անընդհատ երկչափ վեկտորի դեպքին: Ելնելով պայմանական հավանականության սահմանումից, կարելի է ցույց տալ, որ

$$f_{X_2}(x_2|x_1) = f_{\mathbf{X}}(x_1, x_2)/f_{X_1}(x_1), \ -\infty < x_1 < \infty, \ -\infty < x_2 < \infty,$$

հարաբերությունը կարելի է վերցնել որպես X_2 պատահական մեծության ըստ X_1-ի տրված x_1 արժեքի **պայմանական խտության ֆունկցիա**:

X_2 պատահական մեծության **պայմանական բաշխման ֆունկցիան** ըստ $\{X_1 = x_1\}$ պայմանի բնական է սահմանել

$$F_{X_2}(x_2|x_1) = \frac{\mathbf{P}(X_1 < x_1, X_2 = x_2)}{\mathbf{P}(X_1 = x_1)}, \ -\infty < x_1 < \infty, \ -\infty < x_2 < \infty,$$

հարաբերությամբ, իհարկե, եթե $\mathbf{P}(X_2 = x_2)$ հավասար չէ 0: Եթե (X_1, X_2) պատահական վեկտորը անընդհատ է, ապա

$$F_{X_2}(x_2|x_1) = \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{x_2} f_{\mathbf{X}}(x_1, u)du, \ -\infty < x_1 < \infty, \ -\infty < x_2 < \infty :$$

Ածանցելով այս հավասարությունն ըստ x_2-ի, կգանք վերը տրված պայմանական խտության ֆունկցիայի սահմանմանը: Կամայական $-\infty < b_1 < b_2 < \infty$ համար $\mathbf{P}\{b_1 < X_2 < b_2 | X_1 = x_1\}$ պայմանական հավանականությունը կարելի է ստանալ որպես ինտեգրալ՝

$$\mathbf{P}\{b_1 < X_2 < b_2 | X_1 = x_1\} = \int_{b_1}^{b_2} f_{X_2}(x_2|x_1)dx_2:$$

Նման ձևով կամայական $-\infty < a_1 < a_2 < \infty$ համար

$$\mathbf{P}\{a_1 < X_1 < a_2 | X_2 = x_2\} = \int_{a_1}^{a_2} f_{X_1}(x_1|x_2)dx_1,$$

որտեղ $f_{X_1}(x_1|x_2) = f_{\mathbf{X}}(x_1, x_2)/f_{X_2}(x_2)$:

<u>*Օրինակ*</u> 22: Վերադառնանք օրինակ 17-ին և գտնենք $f_{X_2}(x_2|x_1)$ պայմանական խտության ֆունկցիան:

Լուծում: $(0, 1)$ կետը $(1, 0)$ կետին միացնող ուղղի հավասարումն է՝ $x_1 + x_2 = 1$: Տրված x_1-ի համար x_2-ի հնարավոր արժեքներն են՝ $0 < x_2 < 1 - x_1$: Հետևաբար՝

$$f_{X_2}(x_2|x_1) = \begin{cases} 1/(1 - x_1), & \text{եթե } 0 < x_2 < 1 - x_1, \\ 0, & \text{եթե } x_2 < 0, \ x_2 > 1 - x_1 : \end{cases}$$

Ստուգման համար համոզվենք, որ $\int_{-\infty}^{\infty} f_{X_2}(x_2|x_1)dx_2 = 1/(1 - x_1) \int_0^{1-x_1} dx_2 = 1$:

2.7. Ֆունկցիաներ պատահական մեծություններից

Տարբեր պատահական մեծություններ կարող են կապված լինել ֆունկցիոնալ կապով: Սկզբից դիտարկենք ընդհատ պատահական մեծությունների դեպքը: Դիցուք X-ը ընդհատ պատահական մեծություն է, իսկ $y = g(x)$ որևէ ֆունկցիա է: Դիտարկենք $Y = g(X)$ պատահական մեծությունը: Այն նույնպես ընդհատ է: Ենթադրենք, որ X-ի բաշխման աղյուսակն է՝

X	x_1	x_2	$\dots$	x_K
p	p_1	p_2	$\dots$	p_K

:

Այդ դեպքում Y պատահական մեծության հնարավոր արժեքներն են

$$y_1 = g(x_1), \quad y_2 = g(x_2), \quad \dots \quad , \quad y_K = g(x_K) :$$

Եթե բոլոր $y_1, y_2, \dots, y_K$ արժեքները տարբեր են, ապա $\{X = x_k\}$ և $\{Y = y_k\}$, $k = \overline{1,K}$, պատահույթները համարժեք են: Հետևաբար՝

$$\mathbf{P}\{X = x_k\} = \mathbf{P}\{Y = y_k\} = p_k, \quad k = \overline{1,K},$$

և ուրեմն Y պատահական մեծության բաշխման աղյուսակը կլինի՝

Y	y_1	y_2	$\dots$	y_K
p	p_1	p_2	$\dots$	p_K

:

Իսկ եթե $y_1, y_2, \dots, y_N$ արժեքների մեջ կան համընկնողներ, ապա յուրաքանչյուր իրար հավասար արժեքների խմբին աղյուսակում կհատկացվի մեկ տեղ, և համապատասխան հավանականությունները կգումարվեն:

Օրինակ 23: X ընդհատ պատահական մեծությունը տրված է հետևյալ բաշխման աղյուսակով՝

X	-1	0	2
p	0.1	0.6	0.3

:

Գտնել $Y = 3X + 1$ պատահական մեծության բաշխումը:

Լուծում: Y-ի հնարավոր արժեքներն են՝ $y_1 = 3\cdot(-1)+1 = -2, y_2 = 3\cdot0+1 = 1, y_3 = 3\cdot2+1 = 7$: Հետևաբար, Y պատահական մեծության բաշխման աղյուսակը կլինի՝

Y	-2	1	7
p	0.1	0.6	0.3

:

Օրինակ 24: X ընդհատ պատահական մեծության բաշխման աղյուսակն է՝

X	-1	0	1
p	0.1	0.7	0.2

:

Գտնել $Y = X^2$ պատահական մեծության բաշխումը:

Լուծում: Y-ի հնարավոր արժեքներն են՝ $y_1 = (-1)^2 = 1, \quad y_2 = 0^2 = 0, \quad y_3 = 1^2 = 1$: Այստեղ $y_1 = y_3$ և $P\{Y = 1\} = P\{X = -1\} + P\{X = 1\} = 0.1 + 0.2 = 0.3$: Հետևաբար, Y պատահական մեծության բաշխման աղյուսակն է՝

Y	0	1
p	0.7	0.3

:

Դիցուք (X,Y)-ը երկչափ ընդհատ պատահական վեկտոր է և $z = g(x,y)$-ը՝ որևէ ֆունկցիա: Դիտարկենք $Z = g(X,Y)$ պատահական մեծությունը: Այն նույնպես ընդհատ է: Եթե (x_i, y_j) զույգերը, $i = \overline{1,I}$, $j = \overline{1,J}$, (X,Y) պատահական վեկտորի հնարավոր արժեքներն են և $\mathbf{P}\{X = x_i, Y = y_j\} = p_{ij}$, ապա

$$\mathbf{P}\{Z = z\} = \sum_{i,j:g(x_i,y_j)=z} p_{ij} :$$

Օրինակ 25: (X,Y) երկչափ ընդհատ պատահական վեկտորը տրված է հետևյալ բաշխման աղյուսակով՝

$X \backslash Y$	0	1	2
-1	0.1	0.2	0.1
1	0.2	0.1	0.3

Գտնել $Z = X + Y$ պատահական մեծության բաշխումը:

Լուծում: Z պատահական մեծության հնարավոր արժեքներն են՝ -1, 0, 1, 2, 3: Z-ը ընդունում է -1 արժեքն այն դեպքում, երբ $X = -1$, $Y = 0$, ուրեմն՝ $\mathbf{P}\{Z = -1\} = \mathbf{P}\{X = -1, Y = 0\} = 0.1$: Իսկ 1 արժեքը Z-ը ընդունում է, եթե (X, Y) վեկտորը ընդունում է $(-1, 2)$ արժեքը, կամ եթե (X, Y)-ը ընդունում է $(1, 0)$ արժեքը: Հետևաբար՝

$$\mathbf{P}\{Z = -1\} = \mathbf{P}\{X = -1, Y = 2\} + \mathbf{P}\{X = 1, Y = 0\} = 0.1 + 0.2 = 0.3 :$$

Նույն ձևով հաշվվում են Z-ի մյուս արժեքների համապատասխան հավանականությունները: Z պատահական մեծության համար ստանում ենք հետևյալ բաշխման աղյուսակը՝

Z	-1	0	1	2	3
p	0.1	0.2	0.3	0.1	0.3

Տեսնենք՝ ինչպես է արտահայտվում $Y = g(X)$ պատահական մեծության բաշխման ֆունկցիան X անընդհատ պատահական մեծության բաշխման ֆունկցիայի միջոցով:

Եթե $g(x)$-ն անընդհատ, միընքացորեն աճող ֆունկցիա է և, հետևաբար, գոյություն ունի հակադարձ (նույնպես միընքացորեն աճող) ֆունկցիա $x = g^{-1}(y)$, ապա $Y = g(X)$ պատահական մեծության բաշխման ֆունկցիան է՝

$$F_Y(y) = F_X(g^{-1}(y)) : \tag{3}$$

Իրոք. $F_Y(y) = \mathbf{P}\{Y < y\} = \mathbf{P}\{g(X) < y\} = \mathbf{P}\{X < g^{-1}(y)\} = F_X(g^{-1}(y))$:

Եթե $g(x)$-ը անընդհատ, միընքացորեն նվազող ֆունկցիա է, ապա հակադարձ ֆունկցիան նույնպես միընքացորեն նվազող է: $Y = g(X)$ պատահական մեծության բաշխման ֆունկցիան այս դեպքում կլինի՝

$$F_Y(y) = 1 - F_X(g^{-1}(y)) : \tag{4}$$

Իրոք. $F_Y(y) = \mathbf{P}\{Y < y\} = \mathbf{P}\{g(X) < y\} = \mathbf{P}\{X > g^{-1}(y)\} =$
$= 1 - \mathbf{P}\{X \le g^{-1}(y)\} = 1 - \mathbf{P}\{X < g^{-1}(y)\} = 1 - F_X(g^{-1}(y))$,

քանի որ X-ը անընդհատ է, ապա $\mathbf{P}\{X = x\} = 0$:

Այժմ պարզենք, թե ինչպես կարելի է գտնել $Y = g(X)$ պատահական մեծության բաշխման խտությունը X պատահական մեծության $f_X(x)$ բաշխման խտության միջոցով, երբ $y = g(x)$ ֆունկցիան նաև դիֆերենցելի է:

Օգտվելով (3)-ից, կունենանք՝

$$f_Y(y) = F_Y'(y) = \frac{dF_X(g^{-1}(y))}{dg^{-1}(y)} \cdot \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy} : \tag{5}$$

Օգտվելով (4)-ից, կգտնենք՝

$$f_Y(y) = F_Y'(y) = -\frac{dF_X(g^{-1}(y))}{dg^{-1}(y)} \cdot \frac{dg^{-1}(y)}{dy} = -f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy} : \tag{6}$$

Նկատենք, որ այստեղ $\dfrac{dg^{-1}(y)}{dy} < 0$:

Միավորելով (5)-ը և (6)-ը՝ ստանում ենք, որ եթե X-ը անընդհատ պատահական մեծություն է, իսկ $y = g(x)$ ֆունկցիան միընքաց է և դիֆերենցելի, ապա $Y = g(X)$ պատահական մեծության բաշխման խտությունն է՝

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{dg^{-1}(y)}{dy}\right| : \tag{7}$$

Օրինակ 26: Դիցուք $Y = aX + b$, $a \neq 0$, որտեղ X-ը անընդհատ պատահական մեծություն է: Գտնենք Y-ի խտությունը:

Լուծում: $y = ax + b$ ֆունկցիայի հակադարձը կլինի $x = (y - b)/a$: Հետևաբար (7)-ից կստանանք

$$f_{aX+b}(y) = f_X((y-b)/a)/|a| : \qquad (8)$$

Օրինակ 27: (8) բանաձևի օգնությամբ ապացուցել, որ եթե $X \sim \mathcal{N}(m, \sigma^2)$ և $Y = aX + b$, ապա $Y \sim \mathcal{N}(ma + b, \sigma^2 a^2)$:

Լուծում: Կիրառենք (8) բանաձևը՝

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma\,|a|} \exp\left\{-\frac{1}{2}\cdot\frac{((y-b)/a - m)^2}{\sigma^2}\right\} = \frac{1}{\sqrt{2\pi}\sigma\,|a|} \exp\left\{-\frac{1}{2}\cdot\frac{(y-(ma+b))^2}{\sigma^2 a^2}\right\} :$$

Օրինակ 28: Դիցուք X-ը անընդհատ պատահական մեծություն է և $y = g(x) = x^2$: Գտնենք $Y = X^2$ պատահական մեծության $F_Y(y)$ բաշխման ֆունկցիան և $f_Y(y)$ բաշխման խտությունը:

Լուծում: Քանի որ $g(x) = x^2$ ֆունկցիան միընթաց չէ, մենք չենք կարող կիրառել (7) բանաձևը: Ընտրենք այլ ճանապարհ: Եթե $y \le 0$, ապա $F_Y(y) = \mathbf{P}\{Y < y\} = \mathbf{P}\{X^2 < y\} = 0$ և $f_Y(y) = F_Y'(y) = 0$: Եթե $y > 0$, ապա՝

$$F_Y(y) = \mathbf{P}\{Y < y\} = \mathbf{P}\{X^2 < y\} = \mathbf{P}\{-\sqrt{y} < X < \sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

որտեղից՝

$$f_Y(y) = f_X(\sqrt{y})\cdot\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\cdot\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})) : \qquad (9)$$

Օրինակ 29: Օգտվելով (9) բանաձևից՝ գտնենք $Y = X^2$ պատահական մեծության բաշխման խտությունը, եթե $X \sim \mathcal{N}(0, 1)$:

Լուծում:

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y}{2}\right\} + \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y}{2}\right\}\right) = \frac{1}{\sqrt{2\pi y}}\exp\left\{-\frac{y}{2}\right\}, \text{ եթե } y > 0,$$

$$f_Y(y) = 0, \text{ եթե } y \le 0 :$$

Հավանականային մոդելների կառուցման ժամանակ հաճախ օգտագործում են ոչ գծային ֆունկցիաներ նորմալ բաշխված պատահական մեծություններից: Դիտարկենք մի այդպիսի հաճախ օգտագործվող պատահական մեծություն:

Ոչ բացասական Y պատահական մեծության **բաշխման օրենքը լոգարիթմնորեն նորմալ** է, եթե $X = \ln Y$ պատահական մեծության բաշխման օրենքը նորմալ է: Այլ կերպ ասած, $Y = e^X$, որտեղ $X \sim \mathcal{N}(m, \sigma^2)$:

Օրինակ 30: Գտնել լոգարիթմնորեն նորմալ բաշխված Y պատահական մեծության բաշխման խտությունը:

Լուծում: Այստեղ $y = g(x) = e^x$ և $x = g^{-1}(y) = \ln y$: Համաձայն (7)-ի կստանանք՝

$$f_Y(y) = 0, \text{ երբ } y < 0,$$

$$f_Y(y) = f_X(\ln y)\cdot\frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y}\exp\left\{-\frac{(\ln y - m)^2}{2\sigma^2}\right\}, \text{ երբ } y > 0 :$$

Լոգարիթմնորեն նորմալ բաշխված են հետևյալ տնտեսագիտական հատկանիշները՝ բոլոր աշխատակիցների խմբի աշխատավարձը, տրված չափի ընտանիքների խմբում մեկ շնչի եկամուտը և այլն:

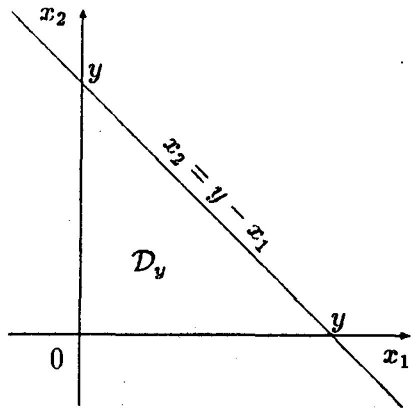
Օգտագործվում են նաև ֆունկցիաներ պատահական վեկտորներից:

Օրինակ 31: Դիցուք $\mathbf{X} = (X_1, X_2)$-ը երկչափ անընդհատ պատահական վեկտոր է, իսկ $f_X(x_1, x_2)$-ը նրա բաշխման խտությունն է: Լուծենք $Y = g(X_1, X_2)$ պատահական մեծության բաշխման խտությունը գտնելու խնդիրը մի շատ կարևոր մասնավոր դեպքի համար, երբ $Y = g(X_1, X_2) = X_1 + X_2$: Նախ գտնենք Y-ի բաշխման ֆունկցիան:

Լուծում:
$$F_Y(y) = \mathbf{P}\{Y < y\} = \mathbf{P}\{X_1 + X_2 < y\} = \iint\limits_{\mathcal{D}_y} f_{\mathbf{X}}(x_1, x_2)dx_1dx_2,$$

որտեղ ինտեգրումը կատարվում է $\mathcal{D}_y = \{(x_1, x_2) : x_1 + x_2 < y\}$ տիրույթով, որը նկար 13-ում ստվերանշված է:

Տրված x_1 արժեքի դեպքում x_2-ը ընդունում է արժեքներ $(-\infty, y - x_1)$ միջակայքից, իսկ x_1-ը փոփոխվում է $-\infty$-ից ∞: Հետևաբար,
$$F_Y(y) = \int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{y-x_1} f_{\mathbf{X}}(x_1, x_2)dx_2\right)dx_1 :$$



Նկար 13:

Ներսի ինտեգրալում կատարենք $x_2 = t - x_1$ փոփոխականի փոխարինում և փոխենք ինտեգրման կարգը: Կստանանք՝
$$F_Y(y) = \int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{y} f_{\mathbf{X}}(x_1, t - x_1)dt\right)dx_1 = \int\limits_{-\infty}^{y}\left(\int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, t - x_1)dx_1\right)dt :$$

Որեմն Y պատահական մեծության բաշխման խտությունն է՝
$$f_Y(t) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, t - x_1)dx_1:$$

Եթե ինտեգրման կարգը փոխենք, ապա նման ձևով կստանանք $f_Y(t) = \int\limits_{-\infty}^{\infty} f_{\mathbf{X}}(t - x_2, x_2)dx_2$:

Եթե X_1 և X_2 պատահական մեծություններն անկախ են, ապա՝
$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2),$$

հետևաբար, ստանում ենք երկու անկախ պատահական մեծությունների գումարի խտությունը, արտահայտված գումարելիների խտությունների միջոցով՝
$$f_{X_1+X_2}(t) = \int\limits_{-\infty}^{\infty} f_{X_1}(x_1)f_{X_2}(t - x_1)dx_1 = \int\limits_{-\infty}^{\infty} f_{X_1}(t - x_2)f_{X_2}(x_2)dx_2 : \qquad (10)$$

(10) բանաձևերը կոչվում են **համադրման (կոմպոզիցիայի) բանաձևեր**:

Օրինակ 32: Դիցուք X_1 և X_2 պատահական մեծություններն անկախ են և բաշխված են λ պարամետրով ցուցչային միևնույն օրենքով, այսինքն՝

$$f_{X_1}(x) = f_{X_2}(x) = \begin{cases} 0, & \text{երբ } x \le 0, \\ \lambda e^{-\lambda x}, & \text{երբ } x > 0: \end{cases}$$

Գտնել $Y = X_1 + X_2$ պատահական մեծության բաշխման խտությունը:

Լուծում: Եթե $y \le 0$, ապա $f_{X_1+X_2}(y) = 0$, իսկ երբ $y > 0$, ապա

$$f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x)dx = \int_{0}^{\infty} \lambda e^{-\lambda x} f_{X_2}(y-x)dx :$$

Եթե $x > y$, ապա $f_{X_2}(y-x) = 0$, հետևաբար $y > 0$ դեպքում՝

$$f_{X_1+X_2}(y) = \int_0^y \lambda e^{-\lambda x} \lambda e^{-\lambda(y-x)}dx = \lambda^2 e^{-\lambda y} \int_0^t dx = \lambda^2 y e^{-\lambda y} :$$

$\lambda > 0$ պարամետրով ցուցչային բաշխում ունեցող N անկախ պատահական մեծությունների գումարի բաշխումը կոչվում է $(N-1)$-րդ կարգի **Էրլանգի բաշխում**, որի բաշխման խտությունն է՝

$$f_{N,\lambda}(x) = \begin{cases} 0, & \text{երբ } x \le 0, \\ \dfrac{\lambda(\lambda x)^{N-1} e^{-\lambda x}}{(N-1)!}, & \text{երբ } x > 0: \end{cases}$$

Վիճակագրությունում հաճախ կիրառվում են նաև ֆունկցիաներ, նորմալ բաշխված անկախ պատահական մեծություններից: Եթե X_1, X_2, ..., X_N պատահական մեծություններն անկախ են, ապա անկախ են նաև $Y_1 = g_1(X_1)$, $Y_2 = g_2(X_2)$, ..., $Y_N = g_N(X_N)$ պատահական մեծությունները:

Դիցուք X_1-ը, X_2-ը, ..., X_N-ը անկախ պատահական մեծություններ են, ընդ որում $X_n \sim \mathcal{N}(0,1)$, $n = \overline{1,N}$: Դիտարկենք

$$\chi^2(N) = X_1^2 + X_2^2 + \cdots + X_N^2$$

պատահական մեծությունը, դրա բաշխումը կոչվում է N ազատության աստիճաններով **χ^2-բաշխում**:

Դիցուք $N = 2$: Գտնենք $\chi^2(2) = X_1^2 + X_2^2$ պատահական մեծության բաշխման խտությունը, որը նշանակենք $f_2(t)$-ով: Քանի որ X_1-ը և X_2-ը անկախ են, ապա անկախ են նաև X_1^2 և X_2^2: Քանի որ X_1^2-ն և X_2^2-ն միատեսակ են բաշխված, ապա օգտվելով (10) բանաձևերից և օրինակ 29-ից՝ $t > 0$ համար կստանանք՝

$$f_2(t) = \int_{-\infty}^{\infty} f_1(x) f_1(t-x)dx = \int_0^{\infty} \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{x}{2}\right\} f_1(t-x)dx :$$

Եթե $t - x < 0$, ապա $f_1(t-x) = 0$, հետևաբար

$$f_2(t) = \int_0^t \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{x}{2}\right\} \frac{1}{\sqrt{2\pi(t-x)}} \exp\left\{-\frac{t-x}{2}\right\} dx = \frac{1}{2\pi} \exp\left\{-\frac{t}{2}\right\} \int_0^t \frac{dx}{\sqrt{tx - x^2}} =$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{t}{2}\right\} \int_0^t \frac{dx}{\sqrt{(t/2)^2 - (x-t/2)^2}} = \frac{1}{2\pi} \exp\left\{-\frac{t}{2}\right\} 2\arcsin 1 = \frac{1}{2} \exp\left\{-\frac{t}{2}\right\} :$$

Իսկ $t \le 0$ դեպքում, բնականաբար, $f_2(t) = 0$:

Այսպիսով, 2 ազատության աստիճաններով χ^2-բաշխման խտության ֆունկցիան է՝

$$f_2(x) = \begin{cases} 1/2 \exp\{-x/2\}, & \text{եթե } x > 0, \\ 0, & \text{եթե } x \le 0: \end{cases}$$

Կարելի է ցույց տալ, որ N ազատության աստիճաններով χ^2-բաշխման խտության ֆունկցիան է՝

$$f_N(x) = \begin{cases} \dfrac{1}{2^{N/2}\Gamma(N/2)} x^{(N/2)-1} \exp\{-x/2\}, & \text{եթե } x > 0, \\ 0, & \text{եթե } x \le 0 : \end{cases}$$

Դժվար չէ համոզվել, որ $\chi^2(N)$ բաշխումը $\Gamma(1/2, N/2)$ բաշխումն է:

Դիցուք $X_0, X_1, X_2, \ldots, X_N$ պատահական մեծությունները անկախ են և $X_n \sim \mathcal{N}(0, \sigma^2)$, $n = \overline{0, N}$: Դիտարկենք հետևյալ պատահական մեծությունը՝

$$\mathrm{t}(N) = X_0 \left(\sqrt{\frac{1}{N} \sum_{n=1}^{N} X_n^2} \right)^{-1} :$$

Այս պատահական մեծության բաշխումը կոչվում է N ազատության աստիճաններով **Ստյուդենտի բաշխում** կամ t-բաշխում: Կարելի է ցույց տալ, որ t-բաշխումը կախված չէ σ պարամետրից, և նրա խտությունն ունի հետևյալ տեսքը՝

$$f_{\mathrm{t}(N)}(t) = \frac{\Gamma\left((N+1)/2\right)}{\sqrt{N\pi}\,\Gamma\left(N/2\right)} \left(1 + \frac{t^2}{N}\right)^{-(N+1)/2} :$$

Դիցուք $X_1, X_2, \ldots, X_{N_1}, X_{N_1+1}, \ldots, X_{N_1+N_2}$ պատահական մեծությունները անկախ են և $X_n \sim \mathcal{N}(0, \sigma^2)$, $n = \overline{1, N_1 + N_2}$: Դիտարկենք հետևյալ պատահական մեծությունը՝

$$\mathcal{F}(N_1, N_2) = \frac{1}{N_1} \sum_{n=1}^{N_1} X_n^2 \left(\frac{1}{N_2} \sum_{n=N_1+1}^{N_1+N_2} X_n^2 \right)^{-1} :$$

Այս պատահական մեծության բաշխումը կոչվում է N_1 և N_2 ազատության աստիճաններով **Ֆիշերի բաշխում** կամ $\mathcal{F}$**-բաշխում**: Ֆիշերի բաշխման խտությունն է՝

$$f_{N_1,N_2}(x) = \begin{cases} \dfrac{\Gamma((N_1+N_2)/2) N_1^{N_1/2} N_2^{N_2/2} x^{(N_1/2)-1}}{\Gamma(N_1/2)\Gamma(N_2/2)(N_2 + N_1 x)^{(N_1+N_2)/2}}, & \text{երբ} \quad x > 0, \\ 0, & \text{երբ} \quad x \le 0 : \end{cases}$$

N_1 և N_2 ազատության աստիճաններով Ֆիշերի բաշխումը կախված չէ σ պարամետրից:

χ^2, Ստյուդենտի և Ֆիշերի բաշխումները լայնորեն կիրառվում են վիճակագրությունում՝ պարամետրերի կետային և միջակայքային գնահատման, վարկածների ստուգման տեսություններում, ռեգրեսիոն և հարաբերակցային վերլուծություններում և այլն: Կիրառությունների համար այդ բաշխման օրենքներով որոշվող հավանականությունները կարելի է ստանալ, օգտվելով աղյուսակներից, որոնք բերված են գրքի վերջում:

Գլուխ 3
Պատահական մեծությունների թվային բնութագրիչներ

Հետաքրքիր է, որ Հյուգենսի մոտ հիմնական գաղափարը սպասելին է, ոչ թե հավանականությունը:

Ես ցանկանում եմ ընդգծել, որ էնտրոպիան և ինֆորմացիան դասում եմ հավանականության տեսության հիմնական գաղափարների շարքին:

Ալֆրեդ Ռենյի

3.1. Կենտրոնական դիրքի բնութագրիչներ

Այս գլխում մենք կդիտարկենք պատահական մեծությունների հիմնական, լայնորեն օգտագործվող թվային բնութագրիչները: Բաշխման ֆունկցիան սպառիչ տեղեկություններ է պարունակում պատահական մեծության մասին, սակայն հաճախ բավական է իմանալ պատահական մեծության մեկ կամ մի քանի էական հատկությունները բնութագրող տվյալներ, որոնք ավելի ամփոփ պատկերացում են տալիս մեզ հետաքրքրող հատկանիշների մասին: Երբեմն պետք է լինում իմանալ մի «միջին» թիվ, որի շուրջ խմբավորվում են պատահական մեծության հնարավոր արժեքները: Այսպես, օրինակ, աշխատակիցների աշխատավարձի ուսումնասիրման ժամանակ առաջին հերթին հետաքրքրվում են միջին աշխատավարձի չափով: Հաճախ կարևորվում է պատահական մեծությունների արժեքների ցրվածության աստիճանը այդ միջինի նկատմամբ: Մեծ նշանակություն ունեն այն բնութագրիչները, որոնք արտացոլում են պատահական մեծությունների ստոխաստիկ փոխկապվածության աստիճանը: Ցանկալի է նաև, որ թվային բնութագրիչները պատահական մեծության բաշխման օրենքից ստացվեն հնարավորին չափ հեշտ հաշվարկների միջոցով: Մեծ մասամբ բնութագրիչները հետևյալ երկու խմբերից են: Առաջին խումբը կազմում են այն բնութագրիչները, որոնց սահմանումը օգտագործում է պատահական մեծության բաշխման **մոմենտների** գաղափարը: Մի ուրիշ խմբի բնութագրիչները կառուցվում են այսպես կոչված **քանորդիչների** օգնությամբ:

Պատահական մեծության կենտրոնական դիրքի ամենակարևոր բնութագրիչներից մեկը սպասելին է: Սպասելիի գաղափարը երևան է եկել Ք. Հյուգենսի «Մոլեխաղերում հաշվարկների մասին» աշխատության մեջ, որը հրապարակվել է 1657 թվականին և եղել է հավանականության տեսության առաջին երկը:

X **ընդհատ պատահական մեծության սպասելին** կամ **միջին արժեքը**, որը նշանակում են $\mathbf{E}(X)$, հավասար է նրա x_n արժեքների և համապատասխան $p_n = \mathbf{P}\{X = x_n\}$ հավանականությունների արտադրյալների գումարին՝

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} x_n p_n, \tag{1}$$

եթե շարքը բացարձակ զուգամետ է, այսինքն՝ $\sum_{n=1}^{\infty} |x_n| p_n < \infty$: Եթե (1) շարքը բացարձակ զուգամետ չէ, ապա X պատահական մեծությունը սպասելի չունի:

<u>*Օրինակ*</u> 1: Կանոնավոր մետաղադրամը նետում են մինչև զինանշանի առաջին անգամ հանդես գալը: Հարկավոր է գտնել նետումների թվի սպասելին:

Լուծում: Նետումների թիվը նշանակենք X: Այս պատահական մեծությունը, ինչպես գիտենք, բաշխված է երկրաչափական օրենքով և ընդունում է $1, 2, \ldots, n, \ldots$ արժեքները $\mathbf{P}\{X = n\} = (1/2)^n$ հավանականություններով: Կստանանք՝ $\mathbf{E}(X) = 2$, քանի որ

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} n(1/2)^n = \sum_{n=1}^{\infty} (n-1)(1/2)^n + \sum_{n=1}^{\infty} (1/2)^n = \sum_{n=0}^{\infty} n(1/2)^{n+1} + 1 = (1/2)\mathbf{E}(X) + 1 :$$

Սահմանման համաձայն՝ վերջավոր N թվով արժեքներ ընդունող պատահական մեծության **սպասելին** վերածվում է վերջավոր գումարի՝

$$\mathbf{E}(X) = x_1p_1 + x_2p_2 + \ldots + x_Np_N = \sum_{n=1}^{N} x_np_n : \qquad (2)$$

Մասնավորապես, եթե պատահական մեծության արժեքները հավասարահավանական են՝ $p_n = 1/N, n = \overline{1, N}$, ապա այդ պատահական մեծության սպասելին համընկնում է նրա արժեքների **միջին թվաբանականի հետ**՝

$$\mathbf{E}(X) = \frac{1}{N} \sum_{n=1}^{N} x_n :$$

<u>*Օրինակ*</u> 2: Ըստ ավտոտեսչության տվյալների՝ ավտոմեքենայի՝ և մեկ տարի անվթար երթևեկելու հավանականությունը 0.992 է: Ապահովագրական ընկերությունը առաջարկում է ավտոմեքենան ապահովագրել 1000000 դրամ գումարով, եթե անձը կատարի մուծում 10000 դրամի չափով: Պահանջվում է գտնել ապահովագրական ընկերության եկամուտի սպասելին մեկ տարով մեկ մեքենա ապահովագրելիս (կողմնակի ծախսերը հաշվի չեն առնվում):

Լուծում: Տարեկան եկամուտի հավանականությունների բաշխման աղյուսակն է՝

X	10000	−990000
p	0.992	0.008

իսկ մեկ մեքենայի ապահովագրումից ընկերության միջին տարեկան եկամուտը ըստ (2)-ի կկազմի

$$\mathbf{E}(X) = 10000 \cdot 0.992 - 990000 \cdot 0.008 = 2000(\text{դրամ}):$$

Սպասելի եկամուտը դրական է, դա հնարավորություն է տալիս ընկերությանը շարունակել աշխատանքը:

<u>*Օրինակ*</u> 3: *A պատահույթի հայտիչի սպասելին*: Մեկ փորձում A պատահույթի հանդես գալու հավանականությունը p է: Գտնել մեկ փորձում A պատահույթի հանդես գալու թվի սպասելին:

Լուծում: Նշանակենք I_A մեկ փորձում A-ի հանդես գալու թիվը: Այդ պատահական մեծությունը կոչվում է A-ի հայտիչ և ընդունում է երկու արժեք՝ 1 (A-ն տեղի է ունեցել) p հավանականությամբ և 0 (A-ն տեղի չի ունեցել) q հավանականությամբ: Որոնելի սպասելին կլինի՝

$$\mathbf{E}(I_A) = q \cdot 0 + p \cdot 1 = p :$$

Այս կարևոր օրինակի արդյունքը հետագայում կօգտագործվի ավելի բարդ պատահական մեծությունների բնութագրիչների հաշվարկման ժամանակ, ուստի ընդգծենք, որ՝

մեկ փորձում A պատահույթի հանդես գալու «անգամների» սպասելին հավասար է այդ պատահույթի p հավանականությանը:

:

Այժմ սահմանենք անընդհատ պատահական մեծության սպասելին:

X **անընդհատ պատահական մեծության** $\mathrm{E}(X)$ **սպասելին** սահմանվում է որպես նրա x արժեքների և խտության $f_X(x)$ ֆունկցիայի արտադրյալի ինտեգրալ.

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx,$$

եթե ինտեգրալը բացարձակ զուգամետ է, այսինքն՝ $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$: Հակառակ դեպքում պատահական մեծությունը սպասելի չունի:

Օրինակ 4: Գտնենք $[a,b]$ միջակայքում **հավասարաչափ բաշխված** X պատահական մեծության սպասելին:

Լուծում: Ունենք՝

$$\mathrm{E}(X) = \frac{1}{b-a}\int_a^b x dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2} :$$

Օրինակ 5: Դիցուք X պատահական մեծության խտությունն է (տե՛ս նկար 1)

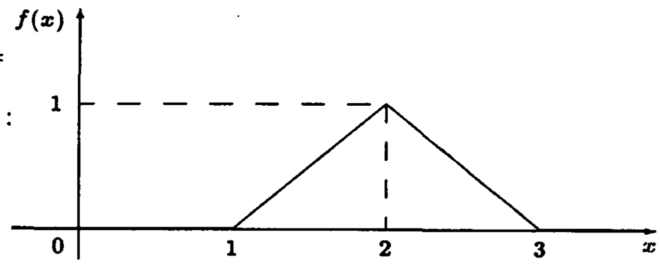
$$f_X(x) = \begin{cases} 0, & \text{երբ } x \leq 1, x \geq 3, \\ x-1, & \text{երբ } 1 \leq x \leq 2, \\ -x+3, & \text{երբ } 2 \leq x \leq 3 : \end{cases}$$

Այս պատահական մեծությունն ունի **Սիմփսոնի (կամ եռանկյունային) բաշխում**: Գտնենք սպասելին:

Լուծում:

$$\mathrm{E}(X) = \int_1^2 x(x-1)dx + \int_2^3 x(3-x)dx =$$

$$= (x^3/3 - x^2/2)|_1^2 + (3x^2/2 - x^3/3)|_2^3 = 2 :$$



Նկար 1: Սիմփսոնի բաշխման օրենքի խտության ֆունկցիան:

Սպասելին օժտված է հետևյալ հատկություններով (C-ն հաստատուն է):

1. $\mathrm{E}(C) = C$: 2. $\mathrm{E}(CX) = C\mathrm{E}(X)$:
3. $\mathrm{E}(X_1 + X_2 + ... + X_N) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + ... + \mathrm{E}(X_N)$:
4. Եթե X-ի բոլոր արժեքները ոչ բացասական են՝ $X \geq 0$, ապա $\mathrm{E}(X) \geq 0$:
4'. Եթե $X \geq Y$, ապա $\mathrm{E}(X) \geq \mathrm{E}(Y)$:
5. **Անկախ պատահական մեծությունների** արտադրյալի սպասելին հավասար է պատահական մեծությունների սպասելիների արտադրյալին՝

$$\mathrm{E}(X_1X_2...X_N) = \mathrm{E}(X_1)\mathrm{E}(X_2)...\mathrm{E}(X_N) :$$

Ապացուցենք նշված հատկությունները: Հաստատուն պատահական մեծությունը 1 հավանականությամբ ընդունում է C արժեքը: Հետևաբար՝

$$\mathrm{E}(C) = C \cdot 1 = C:$$

2-րդ, 3-րդ և 5-րդ հատկությունները կապացուցենք միայն ընդհատ պատահական մեծությունների համար, քանի որ անընդհատ պատահական մեծությունների համար դատողությունները նման են: Ունենք՝

$$\mathbf{E}(CX) = \sum_{n=1}^{\infty} Cx_n p_n = C\sum_{n=1}^{\infty} x_n p_n = C\mathbf{E}(X):$$

Երրորդ և հինգերորդ հատկությունները մենք կապացուցենք միայն երկու պատահական մեծությունների համար, քանի որ ընդհանուր դեպքը հեշտությամբ բերվում է այս դեպքին: X և Y պատահական մեծությունների համատեղ հավանականությունները նշանակենք

$$p_{nm} = P\{X = x_n, Y = y_m\},$$

իսկ մասնատեղ բաշխումները՝

$$\mathbf{P}(X = x_n) = p_n, \mathbf{P}\{Y = y_m\} = q_m:$$

Քանի որ

$$\sum_{m=1}^{\infty} p_{nm} = p_{n.}, \quad \sum_{n=1}^{\infty} p_{nm} = q_{m.}, \tag{3}$$

ուստի՝

$$\mathbf{E}(X+Y) = \sum_n \sum_m (x_n + y_m)p_{nm} = \sum_n x_n \sum_m p_{nm} + \sum_m y_m \sum_n p_{nm} =$$
$$= \sum_n x_n p_{n.} + \sum_m y_m q_{.m} = \mathbf{E}(X) + \mathbf{E}(Y):$$

Չորրորդ հատկությունը հետևում է սպասելիի սահմանումից: Ապացուցենք $4'$ հատկությունը: Եթե $X \geq Y$, ապա $X - Y \geq 0$, ուրեմն $\mathbf{E}(X) - \mathbf{E}(Y) = \mathbf{E}(X - Y) \geq 0$, որտեղից հետևում է $\mathbf{E}(X) \geq \mathbf{E}(Y)$:

X և Y ընդհատ պատահական մեծությունների անկախությունը նշանակում է, որ բոլոր (n, m) զույգերի համար $p_{nm} = p_n q_m$: 5-րդ հատկությունը այդ սահմանման հետևանքն է՝

$$\mathbf{E}(XY) = \sum_n \sum_m x_n y_m p_{nm} = \sum_n \sum_m x_n y_m p_n q_m =$$
$$= \sum_n x_n p_n \sum_m y_m q_m = \mathbf{E}(X)\mathbf{E}(Y):$$

Նշենք սպասելիի ևս մեկ հատկություն՝

> Կամայական X պատահական մեծության համար սպասելիից շեղման սպասելին հավասար է 0-ի՝
> $$\mathbf{E}(X - \mathbf{E}X) = 0:$$

Իսկապես, օգտվելով 1 և 3 հատկություններից, ստանում ենք.

$$\mathbf{E}(X - \mathbf{E}(X)) = \mathbf{E}(X) - \mathbf{E}(\mathbf{E}(X)) = \mathbf{E}(X) - \mathbf{E}(X) = 0:$$

Հաճախ հարկավոր է լինում հաշվել պատահական մեծությունից ֆունկցիայի սպասելին: Դիցուք $Y = g(X)$, այսինքն՝ Y պատահական մեծությունը ֆունկցիոնալ կապի մեջ է X պատահական մեծության հետ:

> Եթե X-ը **ընդհատ** է և ընդունում է x_n արժեքները p_n հավանականություններով, $n = 1, 2, \ldots$, ապա $Y = g(X)$ պատահական մեծության սպասելին է՝
> $$\mathbf{E}(Y) = \sum_{n=1}^{\infty} g(x_n)p_n, \text{ եթե } \sum_{n=1}^{\infty} |g(x_n)|p_n < \infty:$$
> Իսկ եթե X անընդհատ պատահական մեծության բաշխման խտության **ֆունկցիան** $f_X(x)$-ն է, ապա Y-ի սպասելին ստացվում է ինտեգրալի միջոցով՝
> $$\mathbf{E}(Y) = \int_{-\infty}^{\infty} g(x)f_X(x)dx, \text{ եթե } \int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty:$$

Ե՛վ տեսության մեջ, և՛ կիրառություններում կարևոր դեր են կատարում X պատահական մեծության $g_r(X) = X^r$, և $g_r^0(X) = (X - \mathrm{E}(X))^r$, $r = 1, 2, \ldots$, ֆունկցիաների սպասելիները: Դրանք կոչվում են, համապատասխանաբար, r-րդ **կարգի սկզբնական մոմենտներ** կամ ուղղակի **մոմենտներ**՝

$$m_r = \mathrm{E}(X^r),\ r = 1, 2, \ldots$$

և r-րդ **կարգի կենտրոնական մոմենտներ**՝

$$\mu_r = \mathrm{E}(X - \mathrm{E}(X))^r,\ r = 1, 2, \ldots :$$

Կատարելով սպասելիի նշանի տակ երկանդամի r-րդ աստիճանի վերածումը $r+1$ գումարելիների գումարի՝ կարելի է արտահայտել կենտրոնական մոմենտները սկզբնական մոմենտների միջոցով.

$$\begin{aligned} \mu_2 &= m_2 - m_1^2, \\ \mu_3 &= m_3 - 3m_1m_2 + 2m_1^3, \\ \mu_4 &= m_1 - 4m_1m_3 + 6m_2m_1^2 - 3m_1^4, \end{aligned} \tag{4}$$

և այլն: Նշենք, որ չորրորդից բարձր կարգի մոմենտները օգտագործվում են հիմնականում տեսական խնդիրներում:

N-չափանի $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ **պատահական մեծության (վեկտորի) սպասելին** սահմանվում է որպես բաղադրիչների սպասելիների վեկտոր՝

$$\mathrm{E}(\mathbf{X}) = (\mathrm{E}(X_1), \mathrm{E}(X_2), \ldots, \mathrm{E}(X_N)) :$$

Օրինակ 6: Գտնենք բաշխման աղյուսակով տրված (X, Y) ընդհատ պատահական վեկտորի սպասելին՝

$X \backslash Y$	−1	0	1
1	0.15	0.30	0.35
2	0.05	0.05	0.10

:

Լուծում: Նախ հարկավոր է ստանալ X-ի և Y-ի մասնատեղ բաշխման օրենքները՝

$$\mathbf{P}\{X = 1\} = \mathbf{P}\{X = 1, Y = -1\} + \mathbf{P}\{X = 1, Y = 0\} + \mathbf{P}\{X = 1, Y = 1\} =$$
$$= 0.15 + 0.3 + 0.35 = 0.8,$$
$$\mathbf{P}\{X = 2\} = 0.2,\ \mathbf{P}\{Y = -1\} = 0.2,\ \mathbf{P}\{Y = 0\} = 0.35,\ \mathbf{P}\{Y = 1\} = 0.45,$$

որոնք լրացվում են աղյուսակի «լուսանցքներում»՝

$X \backslash Y$	−1	0	1	
1	0.15	0.30	0.35	0.80
2	0.05	0.05	0.10	0.20
	0.20	0.35	0.45	1

:

Այժմ դժվար չէ հաշվել սպասելիները:

$$\mathrm{E}(X) = 1 \cdot 0.8 + 2 \cdot 0.2 = 1.2,\ \ \mathrm{E}(Y) = -1 \cdot 0.2 + 0 \cdot 0.35 + 1 \cdot 0.45 = 0.25 :$$

Ծանոթանանք տեսական և կիրառական մեծ կարևորություն ունեցող **պայմանական սպասելիի** գաղափարին: Դիտարկենք (X, Y) ընդհատ պատահական վեկտորը, որի բաղադրիչները ընդունում են, համապատասխանաբար, $x_n, n = 1, 2, \ldots$, և $y_m, m = 1, 2, \ldots$, արժեքները: Գիտենք (հիշենք 2.6 ենթաբաժնի սահմանումները), որ եթե $q_{\cdot m} > 0$, ապա կամայական n-ի ու m-ի համար

$$p_{n|m} = p_{nm}/q_{\cdot m} \geq 0,\ \ \text{և}\ \sum_n p_{n|m} = 1 :$$

Ուստի սահմանենք X պատահական մեծության $\{Y = y_m\}$ պայմանով սպասելին:

X ընդհատ պատահական մեծության **պայմանական սպասելին ըստ** $\{Y = y_m\}$ **պայմանի** նշանակվում է $\mathbf{E}(X|y_m)$ կամ $\mathbf{E}(X|Y = y_m)$ և սահմանվում է այսպես՝

$$\mathbf{E}(X|y_m) = \sum_{n=1}^{\infty} x_n p_{n|m} = (1/q_m) \sum_{n=1}^{\infty} x_n p_{nm},\ m = 1, 2, \ldots :$$

Նույն ձևով կարելի է սահմանել Y պատահական մեծության պայմանական սպասելին, ըստ $\{X = x_n\}$ պայմանի՝

$$\mathbf{E}(Y|x_n) = (1/p_n) \sum_{m=1}^{\infty} y_m p_{nm},\ n = 1, 2, \ldots :$$

X **պատահական մեծության ըստ** Y**-ի պայմանական սպասելի** կոչվում է $\mathbf{E}(X|Y)$ *պատահական մեծությունը*, որն ընդունում է $\mathbf{E}(X|y_m)$ արժեքները (y_m-երի տեղի ունենալու դեպքերում) q_m հավանականություններով $m = 1, 2, \ldots$:

Հաշվենք նրա սպասելին՝

$$\mathbf{E}\left[\mathbf{E}(X|Y)\right] = \sum_m \mathbf{E}(X|y_m) q_m = \sum_m q_m (1/q_m) \sum_n x_n p_{nm} =$$
$$= \sum_n x_n \sum_m p_{nm} = \sum_n x_n p_n = \mathbf{E}(X) :$$

$$\mathbf{E}[\mathbf{E}(X|Y)] = \mathbf{E}(X) :$$

Այս բանաձևը կոչվում է **լրիվ սպասելիի բանաձև**:

Պատահական մեծության կենտրոնական դիրքի բնութագրիչներ են նաև պատահական մեծության միջին երկրաչափականը, միջին ներդաշնակը, միջին քառակուսայինը, կիսողը և մոդը:

Եթե X դրական պատահական մեծության համար $\mathbf{E}(\ln X)$-ը գոյություն ունի, ապա **միջին երկրաչափականն** է՝

$$x_{\mathrm{G}} = \exp \mathbf{E}(\ln X),$$

որտեղ ցուցչային և լոգարիթմական ֆունկցիաները դիտարկվում են ըստ $e = 2.71828\ldots$ հիմքի:

Կարելի է ապացուցել, որ կամայական X պատահական մեծության համար

$$x_{\mathrm{G}} \le \mathbf{E}(X):$$

Միջին երկրաչափականը կիրառվում է մեծությունների փոփոխման թափի (տեմպի) հաշվման ժամանակ, օրինակ՝ բնակչության թվի ուսումնասիրելիս կամ գների ցուցիչների (ինդեքսների) հաշվարկի ժամանակ:

Երբ $p_n = 1/N$, $n = \overline{1, N}$, ապա միջին **երկրաչափականը հավասար է** $x_1, x_2, \ldots, x_N$ արժեքների միջին երկրաչափականին՝

$$x_{\mathrm{G}} = \exp\left\{(1/N) \sum_{n=1}^{N} \ln x_n\right\} = x_1^{1/N} \cdot x_2^{1/N} \cdot \ldots \cdot x_N^{1/N} = \sqrt[N]{x_1 x_2 \cdot \ldots \cdot x_N} :$$

Եթե X դրական պատահական մեծության համար $\mathbf{E}(1/X) < \infty$, ապա **միջին ներդաշնակը**՝ x_{H}-ը, սահմանվում է այսպես՝

$$x_{\mathrm{H}} = 1/\mathbf{E}(1/X) :$$

Կարելի է ապացուցել, որ կամայական դրական պատահական մեծության միջին ներդաշնակը փոքր է նրա միջին երկրաչափականից և առավել ևս միջին թվաբանականից: Տնտեսագետները միջին ներդաշնակը կիրառում են որոշ ցուցիչների հաշվարկների համար:

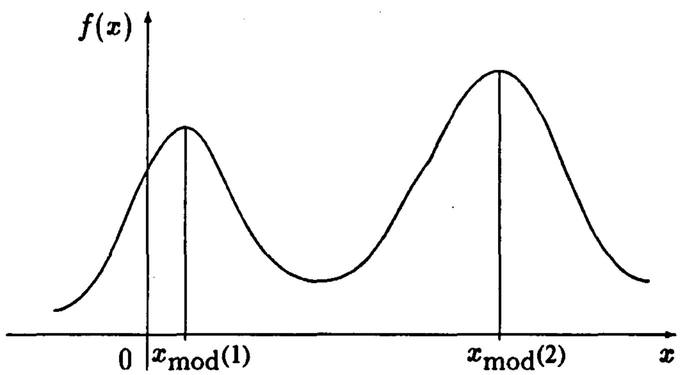
Եթե X պատահական մեծության $\mathrm{E}X^2 < \infty$, ապա **միջին քառակուսայինը** սահմանվում է հետևյալ կերպ՝

$$x_{\mathrm{Q}} = \sqrt{\mathrm{E}X^2}:$$

Բնութագրիչների հաջորդ խումբը սահմանվում է առանց սպասելիի գաղափարի օգտագործման:

X պատահական մեծության **մոդը**՝ x_{mod}-ը, պատահական մեծության այն հնարավոր արժեքն է, որում խտության ֆունկցիան (անընդհատ դեպքում) կամ հավանականությունը (ընդհատ դեպքում) իր շրջակայքում մաքսիմալ է:

Մոդը կարող է գոյություն չունենալ, կարող է լինել միակը (**միամոդալ բաշխում**) և կարող է միակը չլինել (**բազմամոդալ բաշխում**):



Նկար 2: Երկմոդալ բաշխման խտության ֆունկցիայի գծապատկեր:

X **անընդհատ պատահական մեծության կիսողը**՝ x_{med}-ը, այն արժեքն է, որի դեպքում տեղի ունի հետևյալ հավասարությունը՝

$$\mathrm{P}\{X < x_{\mathrm{med}}\} = \mathrm{P}\{X > x_{\mathrm{med}}\} = 0.5:$$

Ընդհատ պատահական մեծության դեպքում կարող է գոյություն չունենալ արժեք, որը ճշգրիտ բավարարի նշված պահանջին: Այդ պատճառով՝

ընդհատ պատահական մեծության կիսողը ըստ սահմանման հավասար է $F_X(x_{k_0}) < 0.5$ և $F_X(x_{k_0+1}) \geq 0.5$ պայմաններին բավարարող x_{k_0} և x_{k_0+1} հարևան արժեքների կիսագումարին՝ $x_{\mathrm{med}} = 1/2(x_{k_0} + x_{k_0+1})$:

Եթե X միամոդալ պատահական մեծության բաշխումը համաչափ է $x = m$ ուղղի նկատմամբ (անընդհատ դեպքում՝ $f_X(m - x) = f_X(m + x)$), ապա պատահական մեծության միջինը, կիսողը և մոդը համընկնում են՝

$$\mathrm{E}(X) = x_{\mathrm{med}} = x_{\mathrm{mod}} = m :$$

Ինչպես կտեսնենք 3.4. ենթաբաժնում, այս պնդումը, մասնավորապես, ճիշտ է նորմալ պատահական մեծության դեպքում:

3.2. Ցրվածության բնութագրիչներ

Տարբեր բաշխումներ ունեցող պատահական մեծությունների սպասելիները կարող են համընկնել:

Օրինակ 7: Դիցուք X և Y ընդհատ պատահական մեծությունները տրված են բաշխման աղյուսակներով՝

X	−0.5	−0.1	0	0.1	0.5
P	0.1	0.2	0.4	0.2	0.1

,

Y	−10	−5	0	5	10
Q	0.3	0.1	0.2	0.1	0.3

:

Երկուսի սպասելիներն էլ հավասար են 0-ի՝

$$\mathrm{E}(X) = -0.5\cdot 0.1 - 0.1\cdot 0.2 + 0\cdot 0.4 + 0.1\cdot 0.2 + 0.1\cdot 0.5 = 0$$

$$\mathrm{E}(Y) = -10\cdot 0.3 - 5\cdot 0.1 + 0\cdot 0.2 + 5\cdot 0.1 + 10\cdot 0.3 = 0 :$$

Սակայն X պատահական մեծությունն ընդունում է արժեքներ, որոնք մոտ են սպասելիին, իսկ Y պատահական մեծությունը կարող է մեծ հավանականությամբ ընդունել արժեքներ, որոնք զգալիորեն շեղված են միջին արժեքից:

Հաճախ կարևոր է լինում իմանալ պատահական մեծության արժեքների սփռվածությունը, ցրվածության աստիճանը սպասելիի նկատմամբ: Կիրառվում են պատահական մեծության արժեքների ցրվածության մի քանի բնութագրիչներ, որոնցից ամենատարածվածը ցրվածքն է, դրանից ստացվող միջին քառակուսային շեղումը նույնպես հաճախ է օգտագործվում:

Ինչպես նշել ենք, կամայական X պատահական մեծության սպասելիից շեղումների սպասելին՝ $\mathrm{E}(X - \mathrm{E}(X))$ հավասար է 0-ի և որպես սփռվածության բնութագրիչ չի կարող ծառայել: Ուստի բնական է օգտագործել երկրորդ կենտրոնական մոմենտը:

Պատահական մեծության **ցրվածք** կոչվում է սպասելիից շեղումների քառակուսիների սպասելին՝

$$\mathrm{D}(X) = \mathrm{E}[X - \mathrm{E}(X)]^2 : \tag{5}$$

Կարելի է նաև դիտարկել բացարձակ շեղումների սպասելին՝ $\mathrm{E}|X - \mathrm{E}(X)|$, սակայն դա հազվադեպ է օգտագործվում, քանի որ նրա հետ աշխատելը հարմար չէ:

Ցրվածքի չափի միավորը համընկնում է պատահական մեծության չափի միավորի քառակուսու հետ, այդ պատճառով հաճախ ավելի հարմար է հետևյալ բնութագրիչը, որի հաշվարկը, սակայն, մի փոքր ավելի բարդ է:

Ցրվածքից քառակուսի արմատը կոչվում է **պատահական մեծության միջին քառակուսային շեղում**

$$\sigma_X = \sqrt{\mathrm{D}(X)} :$$

Օրինակ 8: Հաշվենք օրինակ 7-ում տրված X և Y պատահական մեծությունների ցրվածքները և միջին քառակուսային շեղումները:

Լուծում: $\mathrm{D}(X) = (-0.5-0)^2\cdot 0.1 + (-0.1-0)^2\cdot 0.2 + (0-0)^2\cdot 0.4 +$

$$+(0.1-0)^2\cdot 0.2 + (0.5-0)^2\cdot 0.1 = 0.054,$$

$\mathrm{D}(Y) = (-10-0)^2 \cdot 0.3 + (-5-0)^2 \cdot 0.1 + (0-0)^2 \cdot 0.2 + (5-0)^2 \cdot 0.1 + (10-0)^2 \cdot 0.3 = 65$:

Համապատասխանաբար, $\sigma_X = \sqrt{0.054} \approx 0.23238$, $\sigma_Y = \sqrt{65} \approx 8.06226$:

<u>*Օրինակ*</u> 9: Տրված է X անընդհատ պատահական մեծության բաշխման խտության ֆունկցիան՝

$$f_X(x) = \begin{cases} 0, & \text{երբ } x < -1, x > 0, \\ 3x^2, & \text{երբ } -1 \le x \le 0 : \end{cases}$$

Գտնենք այդ պատահական մեծության ցրվածքը:

Լուծում: Նախ հաշվենք սպասելին՝ $\mathrm{E}(X) = \int\limits_{-\infty}^{\infty} x f_X(x)dx = \int\limits_{-1}^{0} 3x^3 dx = \frac{3}{4}x^4\Big|_{-1}^{0} = -3/4$:

Այնուհետև հաշվենք ցրվածքը՝ $\mathrm{D}(X) = \int\limits_{-1}^{0} (x+3/4)^2 3x^2 dx = 3\int\limits_{-1}^{0} (x^4 + (3/2)x^3 + (9/16)x^2)dx =$

$= 3(x^5/5 + 3x^4/8 + 3x^3/16)\Big|_{-1}^{0} = 0.0375$:

Սպասելիի հատկություններից և (5) սահմանումից արտածվում են **ցրվածքի հատկությունները**.

1. $\mathrm{D}(X) \ge 0$, ընդ որում հավասար է 0-ի այն և միայն այն դեպքում, երբ X պատահական մեծությունը հաստատուն է՝ $X \equiv C$:
2. $\mathrm{D}(CX) = C^2\mathrm{D}(X)$:
3. $\mathrm{D}(X \pm C) = \mathrm{D}(X)$:
4. $\mathrm{D}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$:
5. Եթե X-ը և Y-ը *անկախ* են, ապա $\mathrm{D}(X \pm Y) = \mathrm{D}(X) + \mathrm{D}(Y)$:

Ապացուցենք այդ հատկությունները:

1. $(X - \mathrm{E}(X))^2$ շեղման քառակուսին բացասական չէ, իսկ ոչ բացասական պատահական մեծության սպասելին ոչ բացասական է:

2. $\mathrm{D}(CX) = \mathrm{E}[CX - C\mathrm{E}(X)]^2 = \mathrm{E}C^2[X - \mathrm{E}(X)]^2 = C^2\mathrm{E}[X - \mathrm{E}(X)]^2 = C^2\mathrm{D}(X)$:

3-րդ հատկությունը ստացվում է գումարի և հաստատունի սպասելիի հատկություններից՝

$$\mathrm{D}(X \pm C) = \mathrm{E}[(X \pm C) - \mathrm{E}(X \pm C)]^2 = \mathrm{E}[X - \mathrm{E}(X)]^2 = \mathrm{D}(X)$$

4-րդ հատկությունը (4) բանաձևի այլ գրառում է:

Ապացուցենք 5-ը: Ըստ սպասելիի 3-րդ հատկության՝ $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$, ուրեմն՝

$$[X + Y - \mathrm{E}(X+Y)]^2 = \{[X - \mathrm{E}(X)] + [Y - \mathrm{E}(Y)]\}^2 =$$
$$= [X - \mathrm{E}(X)]^2 + [Y - \mathrm{E}(Y)]^2 + 2[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)] :$$

Հետևաբար,

$$\mathrm{D}(X+Y) = \mathrm{E}[X - \mathrm{E}(X)]^2 + \mathrm{E}[Y - \mathrm{E}(Y)]^2 + 2\mathrm{E}[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)] =$$
$$= \mathrm{D}(X) + \mathrm{D}(Y) + 2\mathrm{E}[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)] :$$

Քանի որ X-ը և Y-ը անկախ են, ուստի անկախ են նաև $[X - \mathrm{E}(X)]$ և $[Y - \mathrm{E}(Y)]$ պատահական մեծությունները, և ըստ սպասելիի 5-րդ հատկության՝

$$\mathrm{E}[X - \mathrm{E}(X)][Y - \mathrm{E}(Y)] = \mathrm{E}[X - \mathrm{E}(X)]\mathrm{E}[Y - \mathrm{E}(Y)] = 0 :$$

Նկատենք նաև, որ

$$\mathrm{D}(X - Y) = \mathrm{D}(X + (-Y)) = \mathrm{D}(X) + \mathrm{D}(-Y) = \mathrm{D}(X) + \mathrm{D}(Y):$$

3-րդ հատկությունը հետևում է նաև 5-րդից, քանի որ հաստատուն պատահական մեծությունն անկախ է կամայական պատահական մեծությունից, իսկ ըստ հատկություն 1-ի հաստատունի ցրվածքը հավասար է 0-ի:

Օրինակ 10: Ձեռնարկությունը պատրաստվում է վաճառել նոր արտադրանք: Մարքեթինգի բաժինը կատարել է վերլուծություն, որից հետևում է, որ եկամուտը կկազմի 4.5 միլիոն դրամ, եթե արտադրանքն ունենա բարձր պահանջարկ, 1.2 միլիոն դրամ, եթե վաճառված արտադրանքի քանակը լինի ոչ շատ մեծ, և կորուստը կլինի 2.3 միլիոն դրամ, եթե վաճառքի մակարդակը լինի շատ ցածր: Այս երեք հնարավորությունների հավանականություններն են, համապատասխանաբար, $0.32, 0.51, 0.17$: Պահանջվում է հաշվել եկամուտի սպասելին և միջին քառակուսային շեղումը:

Լուծում: Հաշվենք միջինը, ցրվածքը՝

$$\mathbf{E}(X) = 4.5\cdot 0.32 + 1.2\cdot 0.51 - 2.3\cdot 0.17 = 1.44 + 0.612 - 0.391 = 1.661(\text{միլիոն դրամ}),$$

$$\mathbf{D}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = 20.25\cdot 0.32 + 1.44\cdot 0.51 + 5.29\cdot 0.7 - (1.661)^2 =$$
$$= 6.48 + 0.7344 + 0.8993 - (1.661)^2 = 8.1137 - 2.7589 = 5.3548,$$

և միջին քառակուսային շեղումը՝ $\sigma_X = \sqrt{5.3548} = 2.314$(միլիոն դրամ) :

Միջին քառակուսային շեղման հատկությունները հետևում են ցրվածքի համապատասխան հատկություններից:

Ինչպես արդեն նշել ենք, պատահական մեծության սպասելիի և միջին քառակուսային շեղման չափողականություններն արտահայտվում են պատահական մեծության չափի միավորներով, իսկ ցրվածքը չափվում է քառակուսի միավորներով: Օրինակ, եթե X-ը բերքահավաքի քանակն է՝ արտահայտված ցենտներներով, ապա նրա սպասելին և միջին քառակուսային շեղումը նույնպես արտահայտվում են ցենտներներով, իսկ ցրվածքը՝ «ցենտներների քառակուսով»: Կիրառական հարցերում դա անհարմար է, այդ պատճառով երբեմն դիտարկվող պատահական մեծությունը վերածում են որևէ կանոնաձև (ստանդարտ) տեսքի, որպեսզի նրանց բնութագրիչները անկախ լինեն չափողականությունից՝ **անչափում** լինեն: Այդպիսին է, օրինակ, նորմավորված և կենտրոնավորված պատահական մեծությունը:

> X պատահական մեծության հետևյալ ձևափոխությունը՝
> $$Y = (X - \mathbf{E}(X))/\sigma_X$$
> կոչվում է X-ի **կանոնաձևում**: Կանոնաձև պատահական մեծությունը անչափում է, նրա սպասելին հավասար է զրոյի, իսկ ցրվածքը՝ մեկի:

Իսկապես,

$$\mathbf{E}(Y) = \mathbf{E}\left[(X - \mathbf{E}(X))/\sigma_X\right] = \mathbf{E}(X - \mathbf{E}(X))/\sigma_X = 0,$$

$$\mathbf{D}(Y) = \mathbf{D}(X - \mathbf{E}(X))/\sigma_X^2 = \mathbf{D}(X)/\sigma_X^2 = 1 :$$

Տնտեսագիտությունում օգտագործվում է նաև պատահական մեծության փոփոխման (վարիացիոն) գործակցի գաղափարը:

> X պատահական մեծության V_X **փոփոխման գործակից** կոչվում է նրա միջին քառակուսային շեղման հարաբերությունը սպասելիին՝ տոկոսներով (ենթադրվում է, որ $\mathbf{E}(X) \neq 0$)՝
> $$V_X = \frac{\sqrt{\mathbf{D}(X)}}{\mathbf{E}(X)}\cdot 100\% = \frac{\sigma(X)}{\mathbf{E}(X)}\cdot 100\% :$$

Ինչպես երևում է սահմանումից, փոփոխման գործակիցն անչափում է:

Լայնորեն կիրառվում են $F_X(x)$ բաշխման օրենքի q քանորդչի՝ u_q, ինչպես նաև Q-տոկոսային կետի՝ w_Q, գաղափարները: Քանորդիչների օգնությամբ կազմվում են բաշխման օրենքի մի խումբ այլ բնութագրիչներ:

X **անընդհատ պատահական մեծության** q **մակարդակի քանորդիչ** (կրճատ՝ q-**քանորդիչ**) կոչվում է այդ պատահական մեծության այնպիսի հնարավոր u_q արժեքը, որի համար $\{X < u_q\}$ պատահույթի հավանականությունը հավասար է տրված q $(0 < q < 1)$ թվին, այսինքն՝

$$F(u_q) = \mathbf{P}\{X < u_q\} = q : \tag{6}$$

Մասնավորապես, կիսողը 0.5-քանորդիչն է:

Դիսկրետ պատահական մեծության $F_X(x)$ բաշխման ֆունկցիան x-ի մեծանալու հետ փոփոխվում է թռիչքներով և, հետևաբար, գոյություն ունեն q-ի այնպիսի արժեքներ, որոնց համար չի գտնվի (6)-ին բավարարող u_q: Այդ պատճառով՝

ընդհատ պատահական մեծության q-**քանորդիչը** սահմանվում է որպես u_q թիվ, որը հավասար է պատահական մեծության այն երկու հարևան $x_{k(q)}$ և $x_{k(q)+1}$ արժեքների կիսագումարին՝ $u_q = (x_{k(q)} + x_{k(q)+1})/2$, որոնց համար $F(x_{k(q)}) < q$, իսկ $F(x_{k(q)+1}) \geq q$:

Հաճախ քանորդչի փոխարեն օգտագործում են նրա հետ կապված տոկոսային կետի գաղափարը:

X **անընդհատ պատահական մեծության** Q-**տոկոսային կետ** $(0 < Q < 100)$ կոչվում է նրա այն w_Q արժեքը, որի համար $\{X > w_Q\}$ պատահույթի հավանականությունը հավասար է $Q/100$, այսինքն՝

$$1 - F(w_Q) = \mathbf{P}\{X > w_Q\} = Q/100 :$$

Ընդհատ պատահական մեծության դեպքում սահմանումը ճշտվում է քանորդչի օրինակով:

Քանորդչի և տոկոսային կետի սահմանումներից հետևում է, որ՝

$$u_q = w_{100(1-q)} :$$

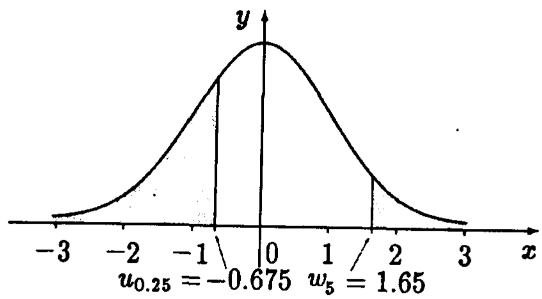
Դիտողություն: Այս սահմանումները կիրառելի են, եթե $f_X(x)$ խտությունը խիստ դրական է X պատահական մեծության հնարավոր արժեքների տիրույթում: Հակառակ դեպքում պետք է արվեն որոշ ճշգրտումներ:

Երբ $q = 0.25,\ 0.5,\ 0.75$, քանորդիչներն անվանում են **քառորդիչներ**, $q = 0.1,\ 0.2,\ 0.3, \ldots, 0.9$ քանորդիչները՝ **տասնորդիչներ**:
$(u_{0.75} - u_{0.25})/2$ տարբերությունը կիրառվում է որպես ցրվածության աստիճանի բնութագրիչ և կոչվում է **քառորդչային կիսահեռավորություն**:
Նման ձևով սահմանվում են **հարյուրորդիչները** $(q = k/100, k = 1, 2, \ldots, 99)$ և **հազարորդիչները** $(q = k/1000, k = 1, 2, \ldots, 999)$:

Բնակչության եկամուտների և աշխատավարձերի բաշխումն ուսումնասիրելու համար դիտարկվում են q **մակարդակի քանորդչային** $K(q)$ **տարբերակման (դիֆերենցիացիայի) գործակիցները**՝

$$K(q) = (u_{1-q})/u_q, \ 0 < q \leq 0.25 :$$

Քանորդիչները ու տոկոսային կետերը կիրառվում են նաև ուսումնասիրվող հատկանիշի փոփոխման իրական սահմանները հայտնաբերելու համար: Օրինակ, 0.005 և 0.995 մակարդակի հազարորդիչներով երբեմն որոշում են, համապատասխանաբար, աշխատավարձերի նվազագույն և առավելագույն մակարդակները:



Նկար 3: Կանոնածև նորմալ բաշխման խտության ֆունկցիան և համապատասխան 0.25-քանորդիչն ու 5-տոկոսային կետը:

3.3. Պատահական մեծության մոմենտները, անհամաչափության և կուտակվածության գործակիցները

Պատահական մեծության բնութագրիչները, որոնք կդիտարկենք այս բաժնում, շատ կարևոր դեր են կատարում և՛ տեսական հետազոտություններում, և՛ կիրառություններում: 3.1 ենթաբաժնում մենք արդեն սահմանել ենք սկզբնական m_r և կենտրոնական μ_r մոմենտները.

$$m_r = m_r(X) = \mathbf{E}(X^r), \ \mu_r = \mu_r(X) = \mathbf{E}[X - \mathbf{E}(X)]^r, \ r = 1, 2, \ldots :$$

Մասնավորապես, եթե $\sum\limits_k |x_k|^r p_k < \infty$, ապա ընդհատ պատահական մեծության r-րդ կարգի սկզբնական մոմենտը կհաշվարկվի հետևյալ բանաձևի միջոցով՝ $m_r = \sum\limits_k x_k^r p_k$, իսկ անընդհատ պատահական մեծության r-րդ կարգի սկզբնական մոմենտը՝ $f_X(x)$ խտության ֆունկցիայի օգնությամբ՝

$$m_r = \int\limits_{-\infty}^{\infty} x^r f_X(x)dx, \text{ եթե } \int\limits_{-\infty}^{\infty} |x|^r f_X(x)dx < \infty:$$

Կամայական պատահական մեծության 0 կարգի սկզբնական մոմենտը հավասար է 1-ի: Առաջին կարգի սկզբնական մոմենտը պատահական մեծության սպասելին է՝

$$m = m_1 = \mathbf{E}(X) :$$

X **պատահական մեծության ֆունկցիայի**՝ $Y = g(X)$-ի, սկզբնական և կենտրոնական մոմենտներն են, համապատասխանաբար՝

$$m_r = m_r(Y) = \mathbf{E}\left(g(x)^r\right),$$
$$\mu_r = \mu_r(Y) = \mathbf{E}\left(g(x) - \mathbf{E}g(X)\right)^r, \ r = 0, 1, 2, \ldots :$$

Կամայական պատահական մեծության 1 կարգի կենտրոնական մոմենտը հավասար է 0-ի: Երկրորդ կարգի կենտրոնական մոմենտը ցրվածքն է՝

$$\mu_2 = \mathbf{E}[X - \mathbf{E}(X)]^2 = \mathbf{D}(X) :$$

Երբեմն օգտակար են նաև պատահական մեծության բացարձակ մոմենտները:

> X պատահական մեծության r-րդ կարգի **բացարձակ սկզբնական մոմենտ** կոչվում է $|X|^r$ պատահական մեծության սպասելին՝
> $$c_r = \mathbf{E}|X|^r,\ r = 1,2,\ldots :$$
> X պատահական մեծության r-րդ կարգի **բացարձակ կենտրոնական մոմենտ** կոչվում է $|X - \mathbf{E}(X)|^r$ պատահական մեծության սպասելին՝
> $$\nu_r = \mathbf{E}|X - \mathbf{E}(X)|^r,\ r = 1,2,\ldots :$$

Մոմենտները օժտված են հետևյալ հատկություններով:

> 1. Եթե գոյություն ունի r-րդ կարգի մոմենտը, ապա գոյություն ունեն r-ից փոքր բոլոր կարգերի մոմենտները:
> 2. Եթե պատահական մեծության բաշխման խտության ֆունկցիան համաչափ է սպասելիի նկատմամբ, ապա բոլոր կենտ կարգի կենտրոնական մոմենտները հավասար են 0-ի:

Ապացուցենք երկրորդ հատկությունը անընդհատ պատահական մեծությունների համար: Դիցուք X պատահական մեծության բաշխման խտությունը համաչափ է $x = \mathbf{E}(X)$ ուղղի նկատմամբ: Խտության համաչափությունը նշանակում է, որ կամայական $u \in \mathcal{R}$ համար $f(\mathbf{E}(X) - u) = f(\mathbf{E}(X) + u)$: Այդ դեպքում կենտ կենտրոնական մոմենտը հավասար է 0-ի: Իսկապես՝

$$\mu_{2r+1} = \int_{-\infty}^{\infty} (x - \mathbf{E}(X))^{2r+1} f_X(x)dx =$$

$$= \int_{-\infty}^{\mathbf{E}(X)} (x - \mathbf{E}(X))^{2r+1} f_X(x)dx + \int_{\mathbf{E}(X)}^{\infty} (x - \mathbf{E}(X))^{2r+1} f_X(x)dx:$$

Կատարելով փոփոխականի փոխարինումներ, համապատասխանաբար՝

$$x - \mathbf{E}(X) = y,\ x - \mathbf{E}(X) = -u,$$

կստանանք

$$\mu_{2r+1} = \int_{-\infty}^{0} y^{2r+1} f_X(\mathbf{E}(X) + y)dy - \int_{-\infty}^{0} u^{2r+1} f_X(\mathbf{E}(X) + u)du = 0 :$$

Որպես օրինակ կարելի է դիտարկել նորմալ բաշխումը, որի բաշխման խտությունը համաչափ է սպասելիի նկատմամբ:

> Նորմալ բաշխված պատահական մեծության բոլոր կենտ կարգի կենտրոնական մոմենտները հավասար են 0-ի:

Հավանականության տեսության կիրառություններում հաճախ օգտագործվում են անհամաչափության և կուտակվածության Ֆիշերի գործակիցները, որոնք պատկերացում են տալիս բաշխման խտության կորի ձևի մասին: Այդ բնութագրիչները անչափում են:

> X պատահական մեծության **անհամաչափության** (կամ **թեքվածության**) **Ֆիշերի գործակիցը** հավասար է երրորդ կարգի կենտրոնական մոմենտի հարաբերությանը միջին քառակուսային շեղման խորանարդին՝
> $$\beta(X) = \mu_3(X)/\sigma_X^3 : \qquad (7)$$

Նշենք անհամաչափության գործակցի կարևոր հատկությունները:

1. Եթե $C > 0$, ապա $\beta(CX) = \beta(X)$, եթե $C < 0$, ապա $\beta(CX) = -\beta(X)$:
2. Անհամաչափության գործակիցը հավասար է 0-ի, եթե բաշխման խտության ֆունկցիան համաչափ է սպասելիի նկատմամբ:
3. Եթե բաշխման խտությունը համաչափ չէ, ընդ որում «երկար մասը» ընկած է սպասելիից աջ, ապա $\beta > 0$: Հակառակ դեպքում $\beta < 0$ (տե՛ս նկար 4):

X պատահական մեծության **կուտակվածության Ֆիշերի գործակիցն** է՝

$$\gamma(X) = \mu_4(X)/\sigma_X^4 - 3\, : \tag{8}$$

Կուտակվածության գործակիցն օժտված է հետևյալ հատկություններով:

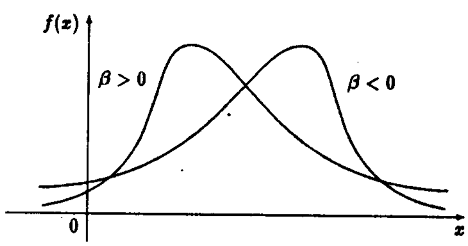
1. X պատահական մեծության գծային ձևափոխությունը չի փոխում նրա կուտակվածության գործակիցը՝
$$\gamma(aX + b) = \gamma(X)\, :$$
2. Նորմալ բաշխված պատահական մեծության կուտակվածության գործակիցը հավասար է 0-ի:
3. Եթե X պատահական մեծության բաշխումը միամոդալ է, և $f_X(x)$ ֆունկցիան ավելի «սրագագաթ» է, քան միևնույն ցրվածքով նորմալ բաշխման խտության ֆունկցիան, ապա $\gamma(X) > 0$, հակառակ դեպքում $\gamma(X) < 0$ (տե՛ս նկար 5):

Սահմանափակվենք երկրորդ հատկության ապացուցումով: Եթե Y-ը նորմալ բաշխված X պատահական մեծության կանոնաձևումն է, ապա՝

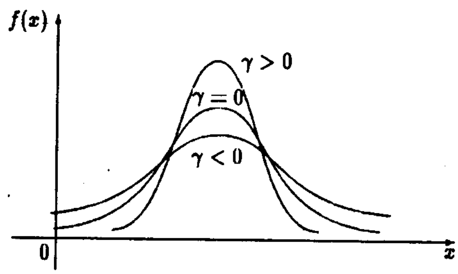
$$\mu_4(Y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^4 e^{-\frac{1}{2}t^2}dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^3 de^{-\frac{1}{2}t^2} = \frac{3}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2}dt = 3\mathrm{D}(Y),$$

հետևաբար՝

$$\gamma(X) = \mu_4(Y) - 3 = 0\, :$$



Նկար 4: Անհամաչափության գործակցի դեպքերը:



Նկար 5: Կուտակվածության գործակցի դեպքերը:

Օրինակ 11: Պատահական մեծությունը տրված է հետևյալ բաշխման խտության ֆունկցիայի միջոցով՝

$$f_X(x) = \begin{cases} 0, & \text{երբ } x < -2,\ x > 0, \\ (-1/4)x^3, & \text{երբ } -2 \le x \le 0\, : \end{cases}$$

Գտնենք անհամաչափության և կուտակվածության գործակիցները:

Լուծում: Նախ հաշվենք առաջին չորս սկզբնական մոմենտները՝

$$m_1 = -1/4 \int_{-2}^{0} x^4 dx = -\frac{1}{4}\frac{x^5}{5}\Big|_{-2}^{0} = -1.6, \; m_2 = -1/4 \int_{-2}^{0} x^5 dx = -\frac{1}{4}\frac{x^6}{6}\Big|_{-2}^{0} = \frac{8}{3} \approx 2.67 :$$

Նույն ձևով կարելի է համոզվել, որ $m_3 = -32/7 \approx 4.57$, $m_4 = 8$:

Հաշվենք կենտրոնական մոմենտները

$$\mu_2 = m_2 - m_1^2 \approx 2.67 - (1.6)^2 \approx 0.11,$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 \approx -4.57 + 3 \cdot 1.6 \cdot 2.67 - 2 \cdot (1.6)^3 \approx 0.054,$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 \approx 0.1024 :$$

Այստեղից հետևում է, որ $\sigma_X = \sqrt{\mathbf{D}(X)} \approx 0.33$, $\sigma_X^3 \approx 0.036$, $\sigma_X^4 \approx 0.0121$: Այժմ հեշտությամբ կգտնենք, որ $\beta(X) = \frac{\mu_3}{\sigma_X^3} \approx 1.5$, $\gamma(X) = \frac{\mu_4}{\sigma_X^4} - 3 \approx 5.46$:

3.4. Կարևորագույն բաշխումների թվային բնութագրիչները

Դիտարկենք մեզ արդեն հայտնի ընդհատ պատահական մեծությունների բաշխման օրենքները և արտածենք դրանց հիմնական թվային բնութագրիչների կապը համապատասխան բաշխումների պարամետրերի հետ:

Սկսենք երկանդամային բաշխումից:

Երկանդամային բաշխում ունեցող պատահական մեծության՝ $X \sim \mathcal{B}(N,p)$, հիմնական *բնութագրիչները* արտահայտվում են N և p պարամետրերի միջոցով հետևյալ կերպ՝

$$\mathbf{E}(X) = Np, \; \mathbf{D}(X) = Npq, \; \sigma_X = \sqrt{Npq}, \; \beta(X) = \frac{q-p}{\sqrt{Npq}}; \; \gamma(X) = \frac{1-6pq}{\sqrt{Npq}} :$$

Ինչպես, դիտարկենք $\mathcal{A}$ պատահույթի մեկ փորձում տեղի ունենալու թվին հավասար $I_{\mathcal{A}}$ պատահական մեծությունը: Այն կոչվում է $\mathcal{A}$-ի հայտիչ (տե՛ս օրինակ 3) և ունի հետևյալ բաշխումը՝

$$\mathbf{P}(I_{\mathcal{A}} = 1) = p, \; \mathbf{P}(I_{\mathcal{A}} = 0) = q :$$

Պարզ է, որ $\mathbf{E}(I_{\mathcal{A}}) = 1 \cdot p + 0 \cdot q = p$: Ներկայացնենք N փորձերում $\mathcal{A}$-ի տեղի ունենալու թիվը՝ X-ը, որպես տարբեր փորձերում դրա տեղի ունենալու թվերի՝ $I_{\mathcal{A},n}, n = \overline{1,N}$, գումար՝ $X = \sum_{n=1}^{N} I_{\mathcal{A},n}$:

Այստեղից, քանի որ $\mathbf{E}(I_{\mathcal{A},n}) = \mathbf{E}(I_{\mathcal{A}}) = p$, $n = \overline{1,N}$, $\mathbf{E}(X) = \sum_{n=1}^{N} \mathbf{E}(I_{\mathcal{A},n}) = Np$: $I_{\mathcal{A}}$ հայտիչի ցրվածքը՝ $\mathbf{D}(I_{\mathcal{A}})$-ն, հավասար է pq: Իսկապես՝

$$\mathbf{D}(I_{\mathcal{A}}) = q(0-p)^2 + pq^2 = pq(q+p) = pq :$$

Փորձերն անկախ են, հետևաբար անկախ են $I_{\mathcal{A},n}$ հայտիչները, և գումարի ցրվածքը հավասար է ցրվածքների գումարին՝

$$\mathbf{D}(X) = \mathbf{D}\left(\sum_{n=1}^{N} I_{\mathcal{A},n}\right) = \sum_{n=1}^{N} \mathbf{D}(I_{\mathcal{A},n}) = Npq :$$

Օրինակ 12: Մարզի մասնավոր շինարարական աշխատանքների 18.6%-ը կատարում է «Շինարար» ընկերությունը: Ստուգման են ենթարկվում մարզի իրարից անկախ պատահականորեն վերցված 25 շինարարական օբյեկտներ: Հարկավոր է գտնել X-ի բաշխման օրենքի սպասելին և միջին քառակուսային շեղումը, եթե X-ով նշանակված է 25 օբյեկտից «Շինարար» ընկերության միջոցով կառուցված օբյեկտների թիվը:

Լուծում: X-ի բաշխումը երկանդամային է՝ $N = 25$, $p = 0.186$, $q = 0.814$ պարամետրերով: Օգտվելով սպասելիի և միջին քառակուսային շեղման բանաձևերից՝ կստանանք՝

$$\mathbf{E}(X) = Np = 25 \cdot 0.186 = 4.65, \; \sigma_X = \sqrt{Npq} = \sqrt{25 \cdot 0.186 \cdot 0.814} = 1.95 :$$

Գտնենք Պուասոնի բաշխման թվային բնութագրիչները:

$\lambda > 0$ պարամետրով *Պուասոնի բաշխում* ունեցող պատահական մեծության՝ $X \sim \Pi(\lambda)$, բնութագրիչներն են՝ $\mathbf{E}(X) = \lambda,\ \mathbf{D}(X) = \lambda,\ \sigma_X = \sqrt{\lambda},\ \beta(X) = 1/\sqrt{\lambda},\ \gamma(X) = 1/\lambda$:

Քանի որ պուասոնյան X պատահական մեծության բաշխումն է՝

$$\mathbf{P}_m = \mathbf{P}\{X = m\} = \frac{\lambda^m}{m!}e^{-\lambda},\ m = 0, 1, 2, ...,$$

ապա նրա սպասելիին բանաձևի համաձայն կլինի՝

$$\mathbf{E}(X) = \sum_{m=0}^{\infty} m\mathbf{P}_m = \sum_{m=1}^{\infty} m\frac{\lambda^m}{m!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} :$$

Կատարելով փոփոխականի $m - 1 = k$ փոխարինումը՝ կստանանք՝

$$\mathbf{E}(X) = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}e^{\lambda} = \lambda :$$

Հաշվենք ցրվածքը՝

$$\mathbf{D}(X) = \sum_{m=0}^{\infty} m^2\frac{\lambda^m}{m!}e^{-\lambda} - \lambda^2 = \lambda e^{-\lambda}\sum_{m=1}^{\infty} m\frac{\lambda^{m-1}}{(m-1)!} - \lambda^2 =$$

$$= \lambda e^{-\lambda}\sum_{m=1}^{\infty}[(m-1)+1]\frac{\lambda^{m-1}}{(m-1)!} - \lambda^2 = \lambda e^{-\lambda}\sum_{m=1}^{\infty}(m-1)\frac{\lambda^{m-1}}{(m-1)!} +$$

$$+ \lambda e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} - \lambda^2 = \lambda^2 e^{-\lambda}\sum_{m=2}^{\infty}\frac{\lambda^{m-2}}{(m-2)!} + \lambda e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} - \lambda^2 =$$

$$= \lambda^2 e^{-\lambda}e^{\lambda} + \lambda e^{-\lambda}e^{\lambda} - \lambda^2 = \lambda :$$

Հաշվենք անհամաչափության գործակիցը: Նախ գտնենք μ_3-ը սկզբնական մոմենտների միջոցով՝

$$m_1 = \mathbf{E}(X) = \lambda,\ m_2 = \mathbf{E}(X^2) = \lambda^2 + \lambda,$$

$$m_3 = \mathbf{E}(X^3) = \sum_{m=0}^{\infty} m^3\frac{\lambda^m}{m!}e^{-\lambda} = \lambda\sum_{m=1}^{\infty} m^2\frac{\lambda^{m-1}}{(m-1)!}e^{-\lambda} = \lambda\sum_{k=0}^{\infty}(k+1)^2\frac{\lambda^k}{k!}e^{-\lambda} =$$

$$= \lambda\sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} + 2\lambda\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} + \lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda} =$$

$$= \lambda\mathbf{E}(X^2) + 2\lambda\mathbf{E}(X) + \lambda = \lambda(\lambda^2 + \lambda) + 2\lambda^2 + \lambda = \lambda^3 + 3\lambda^2 + \lambda :$$

Երրորդ կարգի կենտրոնական մոմենտը հավասար է

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3 = \lambda^3 + 3\lambda^2 + \lambda - 3(\lambda^2 + \lambda)\lambda + 2\lambda^3 = \lambda :$$

Նմանաբար կգտնենք, որ $\mu_4 = \lambda(1 + 3\lambda)$: Անհամաչափության գործակիցը կգտնենք (7) բանաձևի միջոցով՝ $\beta(X) = \mu_3(X)/\sigma_X^3 = \lambda/\sqrt{\lambda^3} = 1/\sqrt{\lambda} > 0$, իսկ կուտակվածության գործակիցը՝ (8) բանաձևի միջոցով՝ $\gamma(X) = \frac{\mu_4}{\sigma_X^4} - 3 = \frac{\lambda(1+3\lambda)}{\lambda^2} - 3 = \frac{1}{\lambda}$:

Քանի որ λ պարամետրը դրական է, ապա Պուասոնի բաշխման անհամաչափության և կուտակվածության գործակիցները դրական են:

<u>*Օրինակ*</u> 13: Լվացքատուն մեկ ժամում միջինում այցելում են 5 պատվիրատու: Լվացքատունը մեքենաների վերանորոգման պատճառով մեկ ժամով փակվում է: Որոշել այդ ընթացքում լվացքատանը 8 հաճախորդի մոտենալու հավանականությունը, եթե ընդունենք, որ հաճախորդների թիվը ներկայացնող X պատահական մեծությունը բաշխված է Պուասոնի օրենքով:

Լուծում: Քանի, որ $\mathbf{E}(X) = \lambda = 5$, որոնելի հավանականությունը կլինի՝

$$\mathbf{P}\{X = 8\} = \frac{5^8e^{-5}}{8!} = \frac{390625 \cdot 0.006738}{40320} = 0.0653 :$$

Հաջողության p հավանականությամբ **երկրաչափական բաշխում** ունեցող պատահական մեծության սպասելին գտնելու համար հիշենք, որ

$$\mathbf{P}_k = \mathbf{P}\{X = k\} = q^{k-1}p,\ k = 1, 2, \dots,\ 0 < p < 1,\ q = 1 - p :$$

Սպասելիի բանաձևից՝ օգտվելով երկրաչափական պրոգրեսիայի հատկություններից, կստանանք՝

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} kq^{k-1}p = p/(1-q)^2 = 1/p :$$

Հաշվենք նաև ցրվածքը՝

$$\mathbf{D}(X) = \sum_{k=1}^{\infty} k^2q^{k-1}p - (1/p)^2 = 2q + p/p^2 - 1/p^2 = q/p^2 :$$

Հաջողության p հավանականությամբ *երկրաչափական օրենքով* բաշխված $X \sim \mathcal{G}(p)$ պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X) = \frac{1}{p},\ \mathbf{D}(X) = \frac{1-p}{p^2},\ \sigma_X = \frac{\sqrt{1-p}}{p},\ \beta(X) = \frac{2-p}{\sqrt{1-p}},\ \gamma(X) = 6 + \frac{p^2}{1-p} :$$

Հիպերերկրաչափական բաշխում ունեցող պատահական մեծության թվային *բնութագրիչներն են*

$$\mathbf{E}(X) = KM/N,\ \mathbf{D}(X) = \frac{MK(N-K)(N-M)}{N^2(N-1)},$$

$$\beta(X) = \frac{(N-2K)(N-2M)\sqrt{N-1}}{\sqrt{KM(N-K)(N-M)(N-2)}},$$

$$\gamma(X) = \frac{N^2(N-1)}{M(N-2)(N-3)(N-K)}\times$$

$$\times\left(\frac{N(N+1) - 6N(N-M)}{K(N-K)} + \frac{3M(N-M)(N+6)}{N^2} - 6\right) :$$

Այժմ դիտարկենք մեզ արդեն ծանոթ կարևորագույն անընդհատ պատահական մեծությունների թվային բնութագրիչները:

$[a, b]$ հատվածի վրա հավասարաչափ բաշխված X պատահական մեծության համար կատարենք հետևյալ պարզ հաշվարկները (տե՛ս օրինակ 4)՝

$$\mathbf{E}(X) = \frac{1}{b-a}\int_a^b x dx = \frac{b+a}{2},$$

$$\mathbf{D}(X) = \frac{1}{b-a}\int_a^b x^2 dx - \left(\frac{b+a}{2}\right)^2 = \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12} :$$

Հավասարաչափ բաշխումը համաչափ է սպասելիի նկատմամբ, հետևաբար, կենտ աստիճանի կենտրոնական մոմենտները հավասար են 0-ի, ուրեմն և անհամաչափության գործակիցը նույնպես հավասար է 0-ի: Կուտակվածության գործակիցը հաշվելու համար գտնենք չորրորդ կարգի կենտրոնական մոմենտը՝

$$\mu_4 = \frac{1}{b-a}\int_a^b (x - \frac{a+b}{2})^4 dx =$$

$$= \frac{1}{b-a}\left(\frac{x^5}{5} - \frac{a+b}{2}x^4 + \frac{(a+b)^2}{2}x^3 - \frac{(a+b)^3}{4}x^2 + \frac{(a+b)^4}{16}x\right)\Big|_a^b = \frac{(b-a)^4}{80},$$

որտեղից՝

$$\gamma(X) = \frac{\mu_4}{\sigma_X^4} - 3 = \frac{(b-a)^4 \cdot 144}{80 \cdot (b-a^4)} - 3 = -1.2 :$$

Այսպիսով՝

> $[a,b]$ հատվածի վրա *հավասարաչափ բաշխված* պատահական մեծության (գրում են՝ $X \sim \mathcal{U}(a,b)$) թվային *բնութագրիչներն են*
>
> $$\mathbf{E}(X) = x_{\text{med}} = \frac{b+a}{2},\ \mathbf{D}(X) = \frac{(b-a)^2}{12},\ \sigma_X = \frac{b-a}{2\sqrt{3}},\ \beta(X) = 0,\ \gamma(X) = -1.2 :$$

Այժմ դիտարկենք ցուցչային բաշխումը:

> λ պարամետրով *ցուցչային բաշխման բնութագրիչներն են*
>
> $$\mathbf{E}(X) = \frac{1}{\lambda},\ \mathbf{D}(X) = \frac{1}{\lambda^2},\ \sigma_X = \frac{1}{\lambda},\ \beta(X) = 2,\ \gamma(X) = 6 :$$

Իրոք,

$$\mathbf{E}(X) = \lambda \int_0^\infty x e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda},$$

իսկ

$$\mathbf{D}(X) = \lambda \int_0^\infty x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} :$$

Միջին քառակուսային շեղումը հավասար է սպասելիին՝ $\sigma_X = \sqrt{\mathbf{D}(X)} = \dfrac{1}{\lambda}$:
Նկատենք, որ կուտակվածության և անհամաչափության գործակիցները կախված չեն λ-ից:

Ինչպիսի՞ն են **նորմալ բաշխման բնութագրիչները**:

> *Նորմալ բաշխված* $X \sim \mathcal{N}(m, \sigma^2)$ պատահական մեծության սպասելին m պարամետրն է, միջին քառակուսային շեղումը՝ σ պարամետրը, իսկ կուտակվածության և անհամաչափության գործակիցները հավասար են 0-ի՝
>
> $$\mathbf{E}(X) = x_{\text{mod}} = x_{\text{med}} = m,\ \mathbf{D}(X) = \sigma^2,\ \sigma_X = \sigma,\ \beta(X) = 0,\ \gamma(X) = 0 :$$

Նորմալ բաշխված պատահական մեծության խտությունն է՝

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}, -\infty < x < \infty :$$

Ըստ սպասելիի սահմանման՝

$$\mathbf{E}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} dx :$$

Կատարենք փոփոխականի փոխարինում՝ $t = \dfrac{x-m}{\sigma}$, կստանանք

$$\mathbf{E}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\sigma t + m) e^{-t^2/2} \sigma dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-t^2/2} dt + m \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt :$$

Առաջին ինտեգրալը հավասար է 0-ի, քանի որ ինտեգրալի նշանի տակ կենտ ֆունկցիա է, երկրորդ ինտեգրալը հավասար է 1-ի, քանի որ դա կանոնավոր նորմալ պատահական մեծության խտության ինտեգրալն է: Հետևաբար, $\mathbf{E}(X) = m$:

Նորմալ բաշխումը համաչափ է $x = m$ ուղղի նկատմամբ, ուստի՝ $\mathbf{E}(X) = x_{\text{mod}} = x_{\text{med}} = m$: Իսկապես, կամայական $x_1 = m - t, x_2 = m + t$ կետերի համար $(x_1 - m)^2 = (m - t - m)^2 = t^2$ և $(x_2 - m)^2 = (m + t - m)^2 = t^2$, ուստի՝

$$f_X(x_1) = f_X(x_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2} :$$

Գտնենք ցրվածքը՝

$$\mathbf{D}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-m)^2 \exp\{-(x-m)^2/2\sigma^2\}\, dx :$$

Եթե նորից կատարենք փոփոխականի նույն $t = \dfrac{x-m}{\sigma}$ փոխարինումը, ապա արդյունքում կստանանք՝

$$\mathbf{D}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \sigma^2 t^2 e^{-t^2/2} \sigma dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt :$$

Մասերով ինտեգրենք ստացված ինտեգրալը՝ ընդունելով $t = u, te^{-t^2/2}dt = dv$: Ստանում ենք

$$\mathbf{D}(X) = \frac{\sigma^2}{\sqrt{2\pi}} (-te^{-t^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2/2} dt) = \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2,$$

իսկ միջին քառակուսային շեղումը՝

$$\sigma_X = \sqrt{\mathbf{D}(X)} = \sigma :$$

Կենտրոնական մոմենտների սահմանումից կարելի է նորմալ բաշխման համար ստանալ հետևյալ առնչությունը՝

$$\mu_k = (k-1)\sigma^2 \mu_{k-2}, \ k = 3, 4, \ldots :$$

Այստեղից՝

$$\mu_{2k+1} = 0, \ \mu_{2k} = \sigma^{2k}(2k-1)!, \ k = 1, 2, \ldots,$$

իսկ կենտ կարգի կենտրոնական մոմենտները հավասար են 0-ի, քանի որ $\mu_1 = 0$, նաև $\beta(X) = 0$, իսկ $\mu_4 = 3\sigma^4$, որտեղից $\gamma(X) = 0$:

Ստորև առանց ապացուցման կներկայացնենք գլուխ 2-ում սահմանված մի քանի կարևոր բաշխման օրենքների թվային բնութագրիչների արժեքները՝ արտահայտված համապատասխան պարամետրերով:

N ազատության աստիճանով $\chi^2(N)$ *բաշխում* ունեցող պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X) = N, \ \mathbf{D}(X) = 2N,$$

$$x_{\text{mod}} = N - 2, \ N \geq 2, \ \beta(X) = 2^{3/2} N^{-1/2}, \ \gamma(X) = 3 + 12/N :$$

$a > 0, \lambda > 0$ պարամետրերով *Գամմա օրենքով* բաշխված $X \sim \Gamma(a, \lambda)$, պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X) = \lambda/a, \ \mathbf{D}(X) = \lambda/a^2,$$

$$x_{\text{mod}} = (\lambda - 1)/a, \ \lambda \geq 1, \ \beta(X) = 2\lambda^{-1/2}, \ \gamma(X) = 3 + 6/\lambda :$$

N ազատության աստիճանով *Ստյուդենտի օրենքով բաշխված* $X \sim t(N)$ պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X) = 0, \ \mathbf{D}(X) = N/(N-2), \ N > 2,$$

$$x_{\text{mod}} = 0, \ \beta(X) = 0, \ \gamma(X) = 0 :$$

$C > 0, -\infty < a < \infty$ պարամետրերով *Կոշիի օրենքով* բաշխված պատահական մեծության մոմենտները գոյություն չունեն, իսկ $x_{\text{mod}} = x_{\text{med}} = a$:

$(N-1)$ կարգի *Էրլանգի բաշխման* օրենքին ենթարկվող պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X)=N/\lambda,\ \mathbf{D}(X)=N/\lambda^2,$$
$$x_{\text{mod}}=(N-1)/\lambda,\ \beta(X)=2N^{-1/2},\ \gamma(X)=3+6/N:$$

N_1, N_2 ազատության աստիճաններով *Ֆիշերի բաշխում* ունեցող $(X\sim\mathcal{F}(N_1,N_2))$ պատահական մեծության թվային *բնութագրիչներն են* (գոյություն ունեն պարամետրերի փակագծերում նշված արժեքների դեպքում)

$$\mathbf{E}(X)=\frac{N_2}{N_2-2},\ (N_2>2),\ \mathbf{D}(X)=\frac{2N_2^2(N_1+N_2-2)}{N_1(N_2-2)^2(N_2-4)},\ (N_2>4),$$
$$x_{\text{mod}}=\frac{N_2(N_1-2)}{N_1(N_2+2)},\ (N_1>2),\ \gamma(X)=\frac{(2N_1+N_2-2)\sqrt{8(N_2-4)}}{(N_2-6)\sqrt{N_1+N_2-2}},\ (N_2>6):$$

Լոգարիթմորեն նորմալ բաշխման օրենքին ենթարկվող պատահական մեծության՝ $Y=e^X$, $X\sim\mathcal{N}(m,\sigma^2)$, *բնութագրիչները* արտահայտվում են $r=e^m$, $w=e^{\sigma^2}$ պարամետրերի միջոցով՝

$$\mathbf{E}(Y)=r\sqrt{w},\ \mathbf{D}(Y)=r^2w(w-1),\ y_{\text{mod}}=r/w,\ y_{\text{med}}=r,$$
$$\beta(Y)=(w+2)(w-1)^{1/2},\ \gamma(Y)=w^4+2w^3+3w^2-3:$$

$\lambda>0$ և $\alpha>0$ պարամետրերով *Վեյբուլի* $\mathcal{W}(\lambda,\alpha)$ *բաշխման* օրենքին ենթարկվող պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X)=\lambda\Gamma[(\lambda+1)/\lambda],\ \mathbf{D}(X)=\lambda^2\left(\Gamma[(\lambda+2)/\lambda]-(\Gamma[(\alpha+1)/\alpha])^2\right),$$
$$x_{\text{mod}}=\begin{cases}\lambda(1-1/\alpha)^{1/\alpha}, & \alpha>1,\\ 0, & \alpha\le 1:\end{cases}$$

$c>0$, $\alpha>0$ պարամետրերով *Պարետոյի բաշխման* օրենքին ենթարկվող պատահական մեծության թվային *բնութագրիչներն են*

$$x_{\text{mod}}=c,\ x_{\text{med}}=2^{1/\alpha}c,$$
$$\mathbf{E}(X)=\frac{\alpha}{\alpha-1}c,\ \alpha>1;\ \mathbf{D}(X)=\frac{\alpha}{(\alpha-1)^2(\alpha-2)}c^2,\ \alpha>2:$$

$a_1>0$, $a_2>0$ պարամետրերով *բետա բաշխման* $\beta(a_1,a_2)$ օրենքին ենթարկվող պատահական մեծության *բնութագրիչներն են*

$$\mathbf{E}(X)=\frac{a_1}{a_1+a_2},\ \mathbf{D}(X)=\frac{a_1a_2}{(a_1+a_2)^2(a_1+a_2+1)},$$
$$x_{\text{mod}}=\frac{a_1-1}{a_1+a_2-2},\ a_1>1, a_2>1,\ \beta(X)=\frac{2(a_2-a_1)(a_1+a_2+1)^{1/2}}{(a_1+a_2+2)(a_1a_2)^{1/2}},$$
$$\gamma(X)=\frac{3(a_1+a_2)(a_1+a_2+1)(a_1+1)(2a_2-a_1)}{a_1a_2(a_1+a_2+2)(a_1+a_2+3)}+\frac{a_1(a_1-a_2)}{a_1+a_2}:$$

3.5. Պատահական մեծությունների փոխկապվածության չափի բնութագրիչները

X, Y պատահական մեծությունները կարող են կապված լինել ֆունկցիոնալ կապով: Օրինակ, $Y = aX + b$, $Y = X^2$, $Y = \exp(2X)$ և այլն: Այդպիսի կապը պատահական չէ՝ դետերմինացված է: Փոխկապվածությունը կարող է լինել նաև ստոխաստիկ, երբ մի պատահական մեծության պայմանական բաշխման օրենքը պատահականորեն փոփոխվում է՝ մյուս պատահական մեծության արժեքներից կախված: Ստոխաստիկ փոխկապվածության բնութագրիչներից է երկու պատահական մեծությունների համացրվածքը:

X և Y պատահական մեծությունների **համացրվածքը** (կովարիացիան) նշանակվում է $\sigma_{XY} = \text{cov}(X, Y)$ և ըստ սահմանման հավասար է՝

$$\sigma_{XY} = \text{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}(X))(Y - \mathbf{E}(Y)) : \qquad (9)$$

Այն կարելի է հաշվարկել նաև հետևյալ բանաձևով, որը դժվար չէ ստանալ (9)-ից՝

$$\sigma_{XY} = \mathbf{E}(X, Y) - \mathbf{E}(X)\mathbf{E}(Y) :$$

Համացրվածքն օժտված է մի շարք **հատկություններով**, որոնք հետևում են սպասելիի հատկություններից:

1. Եթե X-ը և Y-ը **անկախ** են, ապա $\text{cov}(X, Y) = 0$, եթե $\text{cov}(X, Y) \neq 0$, ապա X, Y պատահական մեծություններն անկախ չեն:
2. $\text{cov}(X, Y) = \text{cov}(Y, X)$:
3. $\text{cov}(X, X) = \mathbf{D}(X)$:
4. Կամայական a և b հաստատունների համար $\text{cov}(X + a, Y + b) = \text{cov}(X, Y)$:
5. $\text{cov}(aX + bY, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$:

$\mathbf{X} = (X_1, X_2, ..., X_N)$ պատահական վեկտորի **համացրվածքային մատրից** կոչվում է $\sigma_{n_1,n_2} = \text{cov}(X_{n_1}, X_{n_2})$ տարրեր ունեցող Σ մատրիցը՝

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & ... & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & ... & \sigma_{2N} \\ ... & ... & ... & ... \\ \sigma_{N1} & \sigma_{N2} & ... & \sigma_{NN} \end{pmatrix}. :$$

Համացրվածքի հատկություններից բխում է, որ համացրվածքային մատրիցը համաչափ է ($\sigma_{n_1 n_2} = \sigma_{n_2 n_1}$) և $\sigma_{nn} = \mathbf{D}(X_n) = \sigma_n^2, n = \overline{1, N}$:

Համացրվածքային մատրիցի որոշիչը կոչվում է **ընդհանրացված ցրվածք**, այն կարելի է կիրառել որպես **N-չափանի պատահական վեկտորի ցրվածության բնութագրիչ**:

<u>*Օրինակ*</u> 14: Արտադրանքի որակը բնութագրվում է երկու պարամետրով՝ X և Y: (X, Y) երկչափ պատահական վեկտորի բաշխման օրենքը տրված է աղյուսակում: Հարկավոր է գտնել (X, Y) պատահական վեկտորի Σ համացրվածքային մատրիցը:

Լուծում: Նախ հաշվենք $\mathbf{E}(X)$-ը, $\mathbf{E}(Y)$-ը և $\mathbf{E}(XY)$-ը՝

$\mathbf{E}(X) = 1 \cdot 0.8 + 2 \cdot 0.2 = 1.2,$

$\mathbf{E}(Y) = -1 \cdot 0.2 + 0 \cdot 0.35 + 1 \cdot 0.45 = 0.25,$

$\mathbf{E}(XY) = \sum_{i=1}^{2} \sum_{j=1}^{3} x_i y_j p_{ij} =$

$= 1 \cdot (-1) \cdot 0.15 + \ldots + 2 \cdot 1 \cdot 0.1 = 0.3 :$

X \ Y	−1	0	1	
1	0.15	0.30	0.35	0.8
2	0.05	0.05	0.10	0.2
	0.20	0.35	0.45	1

Համացրվածքային մատրիցի տարրերը կհաշվենք (9) բանաձևի օգնությամբ՝

$$\sigma_{XY} = \sigma_{YX} = \mathrm{cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0.3 - 1.2 \cdot 0.25 = 0 :$$

$$\sigma_{XX} = \mathbf{D}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = 1^2 \cdot 0.8 + 2^2 \cdot 0.2 - 1.2^2 = 0.16 :$$

$$\sigma_{YY} = \mathbf{D}(Y) = \mathbf{E}(Y^2) - \mathbf{E}^2(Y) = (-1)^2 \cdot 0.2 + 0^2 \cdot 0.35 + 1^2 \cdot 0.45 - 0.25^2 = 0.5875 :$$

Այսպիսով,

$$\Sigma = \begin{pmatrix} 0 & 0.16 \\ 0.5875 & 0 \end{pmatrix} :$$

Համացրվածքը գծայնորեն կախված է պատահական մեծությունների չափման սանդղակից: Իսկապես, եթե X-ից անցնենք նոր պատահական մեծության $X_1 = aX$, որտեղ a-ն հաստատուն է, ապա $\mathrm{cov}(X_1, Y) = a \cdot \mathrm{cov}(X,Y)$: Այդ պատճառով նպատակահարմար է դիտարկել կանոնաձև պատահական մեծությունները՝

$$\tilde{X} = (X - \mathbf{E}(X))/\sigma_X, \quad \tilde{Y} = (Y - \mathbf{E}(Y))/\sigma_Y,$$

որոնց համար (համոզվեք դրանում ինքնուրույն)՝

$$\mathrm{cov}(\tilde{X}, \tilde{Y}) = \mathrm{cov}(X,Y)/\sigma_X\sigma_Y :$$

X և Y պատահական մեծությունների կախվածությունը բնութագրելու համար օգտագործվում է **հարաբերակցության գործակիցը**, որը հավասար է կանոնաձևված պատահական մեծությունների համացրվածքին՝

$$\rho_{XY} = \mathrm{cov}(X,Y)/\sigma_X\sigma_Y, \text{կամ } \mathrm{cov}(X,Y) = \rho_{XY}\sigma_X\sigma_Y :$$

Անկախ պատահական մեծությունների հարաբերակցության գործակիցը հավասար է 0-ի (քանի որ այդ դեպքում համացրվածքը հավասար է 0-ի): Հակառակ պնդումը ճիշտ չէ, կախյալ պատահական մեծությունների հարաբերակցության գործակիցը կարող է հավասար լինել 0-ի: Օրինակ 14-ում $\sigma_{XY} = 0$, սակայն X-ը և Y-ը կախյալ են: Իսկապես,

$$\mathbf{P}\{X = 1\} \cdot \mathbf{P}\{Y = -1\} = 0.2 \cdot 0.8 \neq \mathbf{P}\{X = 1, Y = -1\} = 0.15 :$$

X և Y պատահական մեծությունները կոչվում են **ոչ հարաբերակցված**, երբ նրանց հարաբերակցության գործակիցը հավասար է 0-ի՝ $\rho_{XY} = 0$:

Այսպիսով, պատահական մեծությունների անկախությունից հետևում է նրանց ոչ հարաբերակցված լինելը, իսկ ոչ հարաբերակցված պատահական մեծությունները կարող են կախյալ լինել: Սակայն երկչափ նորմալ բաշխման դեպքում (տե՛ս գլուխ 2) հարաբերակցության գործակցի զրո լինելուց հետևում է, որ բաղադրիչները անկախ են:

Հարաբերակցության գործակիցն օժտված է հետևյալ **հատկություններով**՝

1. Հարաբերակցության գործակցի բացարձակ արժեքը չի կարող 1-ից մեծ լինել՝
$$|\rho_{XY}| \le 1 :$$
2. $|\rho_{XY}| = 1$ այն և միայն այն դեպքում, երբ X-ը և Y-ը գծայնորեն կապված են, այսինքն՝ $Y = aX + b$, որտեղ a և b-ն հաստատուններ են, $a \neq 0$, ընդ որում, եթե $a > 0$, ապա $\rho_{XY} = 1$, և, եթե $a < 0$, ապա $\rho_{XY} = -1$: Այդ պատճառով ρ_{XY}-ը կարելի է դիտարկել որպես X-ի և Y-ի գծային կախվածության չափ:

Ապացուցենք այդ հատկությունները:

1. Դիցուք X-ը և Y-ը այնպիսի պատահական մեծություններ են, որ $\mathbf{E}(X) = 0$, $\mathbf{E}(Y) = 0$: Դիտարկենք $(X + tY)^2$ պատահական մեծությունը: Պարզ է, որ ըստ սպասելիի 4 հատկության այդ պատահական մեծության սպասելին բացասական չէ՝

$$\mathbf{E}(X+tY)^2=\mathbf{E}(X^2)+2t\mathbf{E}(XY)+t^2\mathbf{E}(Y^2)\geq 0:$$

Այս քառակուսի անհավասարությունը ճիշտ է t-ի բոլոր արժեքների համար, ուստի նրա որոշիչը դրական չէ՝

$$\mathbf{E}^2(XY)-\mathbf{E}(X^2)\mathbf{E}(Y^2)\leq 0,$$

այսինքն՝ $|\mathbf{E}(XY)|/(\sqrt{\mathbf{E}(X^2)}\sqrt{\mathbf{E}(Y^2)})\leq 1$: Քանի որ $\mathbf{E}(X)=0$, $\mathbf{E}(Y)=0$, ապա

$$\mathrm{cov}(X,Y)=\mathbf{E}(XY),\ \mathbf{D}(X)=\mathbf{E}(X^2),\ \mathbf{D}(Y)=\mathbf{E}(Y^2),$$

և հետևաբար վերջին անհավասարությունը կարելի է գրի առնել այսպես՝ $|\rho_{XY}|=\dfrac{|\mathrm{cov}(X,Y)|}{\sigma(X)\sigma(Y)}\leq 1$:

2. Եթե $Y=aX+b$ $(a\neq 0)$, ապա

$$\mathrm{cov}(X,Y)=\mathrm{cov}(X,aX+b)=a\cdot\mathrm{cov}(X,X)+b\cdot\mathrm{cov}(X,1)=a\mathbf{D}(X),\ \mathbf{D}(Y)=a^2\mathbf{D}(X),$$

որտեղ

$$\rho_{XY}=\frac{|\mathrm{cov}(X,Y)|}{\sqrt{\mathbf{D}(X)}\sqrt{\mathbf{D}(Y)}}=\frac{a\mathbf{D}(X)}{|a|\mathbf{D}(X)}=\begin{cases}1, & a>0,\\ -1, & a<0:\end{cases}$$

Հակառակ պնդման ապացուցումը մեծածավալ լինելու պատճառով բաց ենք թողնում:

<u>Օրինակ</u> 15: Ընդհատ X և Y պատահական մեծությունների բաշխման օրենքը տրված է աղյուսակով՝

X \ Y	-1	0	1	
0	0.10	0.15	0.20	0.45
1	0.15	0.25	0.15	0.55
	0.25	0.40	0.35	1

Գտնենք հարաբերակցության գործակիցը:

Լուծում:

$$\mathbf{E}(XY)=0\cdot(-1)\cdot 0.1+\ldots+1\cdot 1\cdot 0.15=0,$$

$$\mathbf{E}(X)=0\cdot 0.45+1\cdot 0.55=0.55,\ \mathbf{E}(Y)=(-1)\cdot 0.25+0\cdot 0.4+1\cdot 0.35=0.1,$$

$$\mathbf{D}(X)=0.55-0.55^2=0.2475,\ \sigma_X=\sqrt{0.2475}\approx 0.497,$$

$$\mathbf{D}(Y)=0.6-0.1^2=0.59,\ \sigma_Y=\sqrt{0.59}\approx 0.768:$$

Հետևաբար՝

$$\rho_{XY}=\frac{\mathbf{E}(XY)-\mathbf{E}(X)\mathbf{E}(Y)}{\sigma_X\sigma_Y}\approx\frac{0-0.55\cdot 0.1}{0.497\cdot 0.768}\approx -0.144:$$

$\mathbf{X}=(X_1,X_2,...,X_N)$ վեկտորի **հարաբերակցային մատրից** կոչվում է $\rho_{n_1n_2}=\rho_{X_{n_1}X_{n_2}}$ տարրերով R մատրիցը $(n_1=\overline{1,N}, n_2=\overline{1,N})$՝

$$R=\begin{pmatrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1N}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2N}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N1} & \rho_{N2} & \cdots & \rho_{NN}\end{pmatrix},$$

որը համաչափ է, և որի անկյունագծի տարրերը հավասար են 1-ի:

Ըստ հարաբերակցության գործակցի սահմանման հարաբերակցային մատրիցի տարրերը կարելի է գտնել համացրվածքային մատրիցի տարրերից՝

$$\rho_{n_1n_2}=\sigma_{n_1n_2}/(\sigma_{X_{n_1}}\cdot\sigma_{X_{n_2}}):$$

<u>Օրինակ</u> 16: Գտնենք հարաբերակցային մատրիցը ըստ օրինակ 14-ում արդեն հաշվարկված համացրվածքային մատրիցի:

Լուծում: Գտնենք $\rho_{12} = \rho_{XY}$ հարաբերակցության գործակիցը՝ $\rho_{12} = \mathrm{cov}(X,Y)/(\sigma(X)\sigma(Y)) = \sigma_{12}/(\sigma_{11}\sigma_{22}) = 0$: Այսպիսով, $R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$:

X և Y պատահական մեծությունների միջև փոխադարձ կապը կարող է առաջանալ նաև նրանից, որ գոյություն ունի այդ պատահական մեծությունների փոխկապվածությունը երրորդ՝ Z պատահական մեծության հետ: Գտնելու համար X և Y պատահական մեծությունների հարաբերակցության գործակիցը, որը «մաքրված» է Z-ի ազդեցությունից, X-ը և Y-ը փոխարինենք $X_1 = X - a_1Z - b_1$ և $Y_1 = Y - a_2Z - b_2$ պատահական մեծություններով, որոնք հարաբերակցված չեն Z-ի հետ: X և Y պատահական մեծությունները կարելի է արտահայտել $X = X' + X_1 = a_1Z + b_1 + X_1$, $Y = Y' + Y_1 = a_2Z + b_2 + Y_1$ տեսքով, որտեղ

$$X' = a_1Z + b_1,\ Y' = a_2Z + b_2,$$
$$\mathbf{E}(X_1) = \mathbf{E}(Y_1) = 0,\ a_1 = \rho_{XZ}\sigma_X/\sigma_Z,\ a_2 = \rho_{YZ}\sigma_Y/\sigma_Z,$$
$$b_1 = \mathbf{E}(X) - \rho_{XZ}\frac{\sigma_X}{\sigma_Z}\mathbf{E}(Z), b_2 = \mathbf{E}(Y) - \rho_{YZ}\frac{\sigma_Y}{\sigma_Z}\mathbf{E}(Z) :$$

Հարաբերակցության $\rho_{XY|Z}$ գործակիցը սահմանվում է այսպես՝

$$\rho_{XY|Z} = \mathrm{cov}(X_1, Y_1)/(\sigma_{X_1}\sigma_{Y_1}), \tag{10}$$

որտեղ

$$X_1 = X - \rho_{XZ}\frac{\sigma_X}{\sigma_Z}Z - \left\{\mathbf{E}(X) - \rho_{XZ}\frac{\sigma_X}{\sigma_Z}\mathbf{E}(Z)\right\},\ \sigma^2_{X_1} = \mathbf{D}(X_1),$$
$$Y_1 = Y - \rho_{YZ}\frac{\sigma_Y}{\sigma_Z}Z - \left\{\mathbf{E}(Y) - \rho_{YZ}\frac{\sigma_Y}{\sigma_Z}\mathbf{E}(Z)\right\},\ \sigma^2_{X_2} = \mathbf{D}(X_2) :$$

Ըստ համացրվածքի հատկությունների (10) բանաձևում աջ կողմի համարիչը կարելի է ձևափոխել հետևյալ կերպ՝

$$\mathrm{cov}(X_1, Y_1) = \mathrm{cov}(X - a_1Z - b_1, Y - a_2Z - b_2) =$$
$$= \mathrm{cov}(X - a_1Z, Y - a_2Z) = \sigma_X\sigma_Y\rho_{XY} - a_1\sigma_Y\sigma_Z\rho_{YZ} - a_2\sigma_X\sigma_Z\rho_{XY} + a_1a_2\sigma_Z^2 :$$

a_1-ի և a_2-ի արժեքները տեղադրելով կստանանք՝ $\mathrm{cov}(X_1, Y_1) = (\rho_{XY} - \rho_{XZ}\rho_{YZ})\sigma_X\sigma_Y$:

3.6. Էնտրոպիա և շենոնյան ինֆորմացիա

Հռչակավոր ամերիկյան գիտնական Կլոդ Էլվուդ Շենոնը 1948թ. հրապարակեց հոդված, որով հիմնադրվեց մի նոր գիտություն՝ **ինֆորմացիայի տեսությունը**: Այդ մաթեմատիկական գիտության արդյունքները, շնորհիվ իրենց ընդհանուր բնույթի, մեծ դեր են կատարում ինֆորմացիայի պահպանման և հաղորդման տեխնիկայի և բազմապիսի կիրառությունների առաջընթացում: Կարևոր դեր ունեն ինֆորմացիայի տեսության գաղափարները և եղանակները նաև վիճակագրությունում:

Այս բաժնում կուսումնասիրենք Շենոնի առաջարկած հիմնական բնութագրիչներից երկուսը՝ էնտրոպիան և միջին փոխադարձ ինֆորմացիան, որոնք լայնորեն կիրառվում են ոչ միայն կապի տեսությունում, այլ նաև գիտության տարբեր ճյուղերում, մասնավորապես վիճակագրությունում և տնտեսագիտությունում: Այստեղ կդիտարկվեն այդ գաղափարների սահմանումները և հատկությունները, կիրառություններով հետաքրքրվող ընթերցողին առաջարկում ենք կարդալ Տ. Մ. Կովերի և Ջ. Ա. Տոմասի «Ինֆորմացիայի տեսության տարրերը» գրքի «Ինֆորմացիայի տեսությունը և վիճակագրությունը», «Ինֆորմացիայի տեսությունը և ֆոնդային բորսան» գլուխները, ինչպես նաև Ռ. Ե. Բլեյհուտի «Ինֆորմացիայի տեսություն» գիրքը:

Պատահական մեծությունը, որը մի փորձի մաթեմատիկական մոդելն է, փորձի իրագործումից առաջ պարունակում է որոշակի անորոշություն, որի աստիճանը կախված է պատահական մեծության հավանականությունների բաշխումից: Հասկանալի է, որ N հավասարահավանական ելքերով փորձի անորոշությունը կախված է N-ից: Եթե $N = 1$, ապա փորձի արդյունքը պատահական չէ, և անորոշությունը պետք է հավասար լինի 0-ի: Եթե N-ը մեծանում է, ապա անորոշությունը նույնպես մեծանում է: Բնական է նաև, որ երկու անկախ փորձերի անորոշության չափը հավասար լինի այդ փորձերի անորոշությունների գումարին: Ուրեմն անորոշությունը որպես f ֆունկցիա փորձի հավասարահավանական ելքերի թվից պետք է օժտված լինի հետևյալ հատկություններով՝

1. $f(NM) = f(N) + f(M)$,
2. $f(1) = 0$,
3. $f(N) > f(M)$, երբ $N > M$:

Կարելի է ապացուցել, որ այն միակ ֆունկցիան, որը բավարարում է այս պայմաններին, լոգարիթմական ֆունկցիան է: Լոգարիթմի հիմքը էական չէ, դրանից փոխվում է միայն չափման միավորը: Ընդունված է դիտարկել երկու հիմքով լոգարիթմները, այդ դեպքում միավորը կոչվում է **բիթ**: Այսպիսով, N հավասարահավանական ելքերով փորձի անորոշությունը հավասար է $\log N$, բայց քանի որ յուրաքանչյուր ելքի հավանականությունը $1/N$ է, ապա մի ելքի անորոշությունն է $1/N \log N = -1/N \log 1/N$: Պարզ է, որ ընդհանուր դեպքում անորոշության չափը պետք է կախված լինի ելքերի հավանականություններից:

> Ընդհատ X պատահական մեծության անորոշության չափը Շենոնի կողմից անվանվել է **էնտրոպիա**, այն նշանակվում է $\mathbf{H}(X)$ և հավասար է՝
> $$\mathbf{H}(X) = -\sum_{n=1}^{N} p_n \log p_n :$$

Նկատենք, որ էնտրոպիան կախված չէ պատահական մեծության արժեքներից, այլ արտահայտվում է միայն դրանց հավանականություններով:

Օրինակ 17: Եղանակի բազմամյա դիտումներից հայտնի է, որ մի որոշակի վայրում հունիսի 16-ին անձրև գալու հավանականությունը հավասար է 0.4, համապատասխանաբար, անձրև չլինելու հավանականությունը 0.6 է: Նույն վայրում նոյեմբերի 16-ին անձրև գալու հավանականությունը հավասար է 0.65, ձյուն գալու հավանականությունը՝ 0.15, իսկ տեղումներ չլինելու հավանականությունը՝ 0.2: Հարկավոր է պարզել, թե որ օրն է եղանակը ավելի անորոշ:

Լուծում: Հաշվենք երկու օրերի համար եղանակի էնտրոպիաները՝

$$\mathbf{H}(X_1) = -0.4 \log 0.4 - 0.6 \log 0.6 \approx 0.97\text{բիթ},$$

$$\mathbf{H}(X_2) = -0.65 \log 0.65 - 0.15 \log 0.15 - 0.2 \log 0.2 \approx 1.28\text{բիթ},$$

Տեսնում ենք, որ $\mathbf{H}(X_2) > \mathbf{H}(X_1)$, նոյեմբերի 16-ին եղանակը ավելի անորոշ է, քան թե հունիսի 16-ին: Սակայն, եթե մեզ հետաքրքրում է միայն՝ տեղումներ կլինե՞ն, թե՞ ոչ, ապա

$$\mathbf{H}(X_2') = -0.8 \log 0.8 - 0.2 \log 0.2 \approx 0.72 < \mathbf{H}(X_1),$$

այսինքն՝ նոյեմբերի 16-ին տեղումներ լինելու անորոշությունը ավելի փոքր է:

Այժմ սահմանենք X պատահական մեծության պայմանական էնտրոպիան, երբ հայտնի է մեկ այլ Y պատահական մեծության ընդունած արժեքը:

X ընդհատ պատահական մեծության **պայմանական էնտրոպիան** Y պատահական մեծության պայմանում հավասար է՝

$$\mathrm{H}(X|Y) = -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_{n|m},$$

իսկ X և Y պատահական մեծությունների **համատեղ** էնտրոպիան հավասար է՝

$$\mathrm{H}(X,Y) = -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_{nm} :$$

Օրինակ 18: Տրված է (X,Y) ընդհատ պատահական վեկտորի բաշխումը՝

X \ Y	1	2	3	
1	1/8	1/8	1/8	3/8
2	1/8	0	1/8	2/8
3	1/8	1/8	1/8	3/8
	3/8	2/8	3/8	1

Համեմատել $\mathrm{H}(Y)$-ը և $\mathrm{H}(Y|X)$-ը:

Լուծում:

$$\mathrm{H}(X) = \mathrm{H}(Y) = -3/8\log 3/8 - 2/8\log 2/8 - 3/8\log 3/8 = 1.57\text{բիթ},$$

$\mathrm{H}(Y|X)$-ը գտնելու համար հաշվենք $p_{n|m}$-երը $p_{n|m} = \dfrac{p_{nm}}{q_{.m}}$ բանաձևով:

$$\mathrm{H}(Y|X) = -1/3\log 1/3 - 1/2\log 1/2 - 1/3\log 1/3 = 1.43\text{բիթ},$$

որտեղից երևում է, որ $\mathrm{H}(Y) > \mathrm{H}(Y|X)$: Դա, իհարկե, բնական է, քանի որ լրացուցիչ տեղեկությունը չի կարող մեծացնել անորոշությունը (տե՛ս ստորև երրորդ հատկությունը):

Էնտրոպիան օժտված է հետևյալ հատկություններով՝

1. $0 \le \mathrm{H}(X) \le \log N$, ընդ որում հավասարությունը ձախից տեղի ունի միայն այն դեպքում, երբ պատահական մեծության արժեքներից մեկի հավանականությունը հավասար է մեկի, իսկ մյուսներինը՝ զրոյի, աջից հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ X-ի բոլոր N արժեքները հավասարահավանական են:
2. $\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y|X) = \mathrm{H}(Y) + \mathrm{H}(X|Y)$:
3. $\mathrm{H}(X|Y) \le \mathrm{H}(X)$, ընդ որում հավասարության համար անհրաժեշտ է և բավարար, որպեսզի X և Y պատահական մեծությունները լինեն անկախ:

Իսկապես, 1 հատկության ձախ անհավասարությունն ակնհայտ է, քանի որ $0 \le p(x) \le 1$ և $\log p(x)$ չի կարող դրական լինել:

Աջ անհավասարությունն ապացուցելու համար օգտվենք անալիզից հայտնի $\ln\alpha \le \alpha - 1$ անհավասարությունից: Այն ճիշտ է կամայական α-ի համար և վերածվում է հավասարության, երբ $\alpha = 1$: Կազմենք $\mathrm{H}(X) - \log N$ տարբերությունը և գնահատենք այն.

$$\mathrm{H}(X) - \log N = \sum_{n=1}^{N} p_n\left[\log\frac{1}{p_n} - \log N\right] = \sum_{n=1}^{N} p_n\log\frac{1}{Np_n} \le$$

$$\le \sum_{n=1}^{N} p_n\left[\frac{1}{Np_n} - 1\right]\log e = 0 :$$

Ուրեմն $\mathrm{H}(X) \le \log N$, իսկ հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ

$$\frac{1}{Np_n} = 1, \text{ այսինքն, երբ } p_n = 1/N:$$

Երկրորդ հատկությունն անմիջապես հետևում է էնտրոպիայի սահմանումից՝

$$\mathbf{H}(X,Y) = -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_{nm} = -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_n p_{m|n} =$$
$$= -\sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_n - \sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_{m|n} =$$
$$= -\sum_{n=1}^{N} p_n\log p_n - \sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log p_{m|n} = \mathbf{H}(X) + \mathbf{H}(Y|X):$$

Երրորդ հատկությունը կարելի է ապացուցել առաջինի նման՝ դիտարկելով $\mathbf{H}(Y|X) - \mathbf{H}(X)$ տարբերությունը և գնահատելով այն $\ln\alpha \le \alpha - 1$ անհավասարության օգնությամբ:

2. և 3. հատկություններից ստացվում է, որ անկախ X և Y պատահական մեծությունների դեպքում՝

$$\mathbf{H}(X,Y) = \mathbf{H}(X) + \mathbf{H}(Y):$$

Հատկություն 3-ից տեսնում ենք, որ X պատահական մեծության անորոշությունը կարող է փոքրանալ, եթե մի այլ Y պատահական մեծության միջոցով տեղեկություն ստացվի X-ի մասին: Առաջանում է բնական հարց, ինչպես չափել ստացված տեղեկությունների «ծավալը», քանակը: Ստացված ինֆորմացիայի քանակը կախված է հավանականությունների փոփոխությունից:

Y պատահական մեծության y_m արժեքի՝ X պատահական մեծության x_n արժեքի վերաբերյալ տված **ինֆորմացիայի քանակը** ըստ Շենոնի հավասար է՝

$$\mathbf{I}(x_n \wedge y_m) = \log(p_{n|m}/p_{n.}):$$

Բնական է ենթադրել, որ $p_{n.} \neq 0$ և $q_{.m} \neq 0$ բոլոր n, m-երի համար, քանի որ հակառակ դեպքում համապատասխան արժեքները չէին դիտարկվի: Ըստ սահմանման x_n-ի ինֆորմացիան y_m-ի նկատմամբ հավասար է՝

$$\mathbf{I}(y_m \wedge x_n) = \log\frac{p_{m|n}}{q_{.m}} = \log\frac{p_{m|n}p_{n.}}{q_{.m}p_{n.}} = \log\frac{p_{n|m}}{p_{n.}} = \mathbf{I}(x_n \wedge y_m):$$

Ստացվում է, որ y_m-ը պարունակում է x_n-ի նկատմամբ նույնքան ինֆորմացիա, որքան x_n-ը՝ y_m-ի նկատմամբ: Այդ պատճառով $\mathbf{I}(x_n \wedge y_m)$-ը կոչվում է x_n-ի և y_m-ի **փոխադարձ ինֆորմացիայի քանակ**:

Ինֆորմացիայի քանակը հավասար է զրոյի, եթե պատահույթները վիճակագրորեն անկախ են՝ $p_{m|n} = q_m$, դրական է, եթե $p_{m|n} > q_m$ և բացասական է, երբ $p_{m|n} < q_m$: Դա բնական է, քանի որ անկախ երևույթները մեկը մյուսի մասին չեն կրում որևէ ինֆորմացիա, իսկ եթե x_n-ը բարձրացնում է y_m-ի հավանականությունը՝ $p_{m|n} > q_m$, ապա y_m-ի մասին ինֆորմացիան դրական է:

$\mathbf{I}(x_n \wedge y_m)$-ը կարելի է դիտարկել որպես պատահական մեծություն, որի սպասելին կոչվում է միջին փոխադարձ ինֆորմացիա:

X և Y պատահական մեծությունների միջին փոխադարձ ինֆորմացիայի քանակը կամ պարզապես **փոխադարձ ինֆորմացիան** նշանակվում է $\mathbf{I}(X \wedge Y)$ և սահմանվում հետևյալ կերպ՝

$$\mathbf{I}(X \wedge Y) = \sum_{n=1}^{N}\sum_{m=1}^{M} p_{nm}\log(p_{n|m}/p_{n.}):$$

Սահմանումից դժվար չէ տեսնել, որ ստացված ինֆորմացիան հավասար է երկրորդ պատահական մեծության տրված տեղեկության հետևանքով անորոշության նվազմանը՝

$$\mathbf{I}(X \wedge Y) = \mathbf{H}(X) - \mathbf{H}(X|Y) = \mathbf{H}(Y) - \mathbf{H}(Y|X) = \mathbf{H}(X) + \mathbf{H}(Y) - \mathbf{H}(XY): \quad (11)$$

Միջին փոխադարձ ինֆորմացիան օժտված է հետևյալ հատկություններով.

1. $\mathrm{I}(X \wedge Y) = \mathrm{I}(Y \wedge X)$,
2. $\mathrm{I}(X \wedge Y) \geq 0$, ընդ որում հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ X-ը և Y-ը անկախ են:
3. $\mathrm{I}(X \wedge Y) \leq \min[\mathbf{H}(X), \mathbf{H}(Y)]$,
4. $\mathrm{I}(X \wedge YZ) = \mathrm{I}(X \wedge Y) + \mathrm{I}(X \wedge Z|Y)$,
5. $\mathrm{I}(X \wedge YZ) \geq \mathrm{I}(X \wedge Y)$:

Նշենք, որ երկրորդ հատկության շնորհիվ ինֆորմացիան կարող է կատարել պատահական մեծությունների փոխկապվածության բնութագրիչի դեր:

1 և 3 հատկություններն անմիջապես հետևում են (11)-ից:

2-ը էնտրոպիայի 3 հատկության մեկ այլ գրառումն է: Իսկապես, քանի որ $\mathbf{H}(X|Y) \leq \mathbf{H}(X)$, իսկ $\mathrm{I}(X \wedge Y) = \mathbf{H}(X) - \mathbf{H}(X|Y)$, ապա $\mathrm{I}(X \wedge Y) \geq 0$, ընդ որում հավասարությունը տեղի ունի այն և միայն այն դեպքում, երբ $\mathbf{H}(X|Y) = \mathbf{H}(X)$, այսինքն, երբ X-ը և Y-ը անկախ են:

4 հատկությունը հեշտությամբ ստացվում է ինֆորմացիայի սահմանումից, եթե հաշվի առնենք, որ $p_{mk|n} = p_{m|n}p_{k|nm}$ և $p_{mk} = p_m p_{k|m}$, իսկ արտադրյալի լոգարիթմը հավասար է լոգարիթմների գումարին՝

$$\mathrm{I}(X \wedge YZ) = \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K} p_{nmk} \log \frac{p_{mk|n}}{p_{mk}} =$$

$$\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K} p_{nmk} \log \frac{p_{m|n}}{q_m} + \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{k=1}^{K} p_{nmk} \log \frac{p_{k|nm}}{p_{k|m}} =$$

$$\mathrm{I}(X \wedge Y) + \mathrm{I}(X \wedge Z|Y):$$

Հինգերորդ հատկությունը հեշտությամբ կստանանք՝ տեղադրելով $\mathrm{I}(X \wedge Z|Y) \geq 0$ 4-րդ հատկության նույնության մեջ:

Օրինակ 19: Ամենաքիչը քանի՞ հարցով կարելի է պարզել, թե որ՝ 10-ից ոչ մեծ, բնական թիվն է մտքում պահել մեր զրուցակիցը, եթե նա պատասխանում է մեր տված հարցերին միայն «այո» կամ «ոչ»:

Լուծում: X պատահական մեծությունը կարող է ունենալ 10 հավասարահավանական արժեքներ, այսինքն անորոշությունը հավասար է՝ $\mathbf{H}(X) = \log 10 \approx 3.32$ բիթ: Յուրաքանչյուր հարց կարող է ունենալ երկու պատասխան, այսինքն ամենաշատը տալիս է 1 բիթ: Եթե ենթադրենք, որ k հարցով պարզվել է մտքում պահած թիվը, և նշանակենք այդ պատահական մեծությունը Y-ով, ապա

$$\mathrm{I}(X \wedge Y) = \mathbf{H}(X):$$

Ստանում ենք, որ

$$\log 10 = \mathbf{H}(X) = \mathrm{I}(X \wedge Y) \leq \mathbf{H}(Y) \leq k,$$

այսինքն $k \geq 3.32$ կամ, քանի որ k-ն ամբողջ թիվ է, ապա $k \geq 4$:

Այժմ համոզվենք, որ 4 հարցով իսկապես կարելի է գտնել x թիվը: Դրա համար հարկավոր է ձգտել, որ հարցի պատասխանի էնտրոպիան լինի հնարավորին չափ մեծ, այսինքն մոտ 1 բիթին: Պարզ է, որ հարցի երկու պատասխանները պետք է լինեն մոտավորապես հավասարահավանական: Օրինակ, կարելի է հարցնել. «x-ը մե՞ծ է 5-ից, թե՞ ոչ»: Հաջորդ հարցը կարելի է կազմել նույն սկզբունքով:

Ընդհանուր դեպքում, եթե X-ի հնարավոր արժեքները հավասարահավանական չեն, ապա այդ արժեքների բազմությունը հարկավոր է բաժանել այնպիսի երկու մասի, որ x թվի դրանց պատկանելու հավանականությունները լինեն հնարավորին չափ մոտ:

Նշենք, որ խնդիրն ունի նաև տնտեսական կիրառություններ: Դրանք են՝ հաղորդագրությունների ռացիոնալ կոդավորումը, առարկաների տեսակավորումը ըստ այս կամ այն հատկանիշի, բառարանում բառի կամ մեծ գրադարանում գրքի որոնումը, որևէ օբյեկտ հսկելու արդյունավետ ծրագրերի ստեղծումը:

Անընդհատ պատահական մեծությունների դեպքում էնտրոպիան, պայմանական էնտրոպիան, միջին փոխադարձ ինֆորմացիան սահմանվում են համանման ձևով (գումարները փոխարինվում են ինտեգրալներով)՝

$$\mathbf{H}(X) = -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)dx,$$

$$\mathbf{H}(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_Y(y)f_X(x|y)\log f_X(x|y)dxdy,$$

$$\mathbf{I}(X\wedge Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_Y(y)f_X(x|y)\log\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}dxdy :$$

Վերը նշված հատկությունների մեծ մասը պահպանվում է նաև անընդհատ պատահական մեծությունների դեպքում: Սակայն մեկնաբանումներում հարկավոր է զգույշ լինել, քանի որ խտության ֆունկցիան պատահական մեծության գծային ձևափոխության ժամանակ փոխվում է, և, հետևաբար, փոխվում է նաև էնտրոպիան:

3.7. Կուլբակի-Լեյբլերի ինֆորմացիա (ինֆորմացիոն տարամիտություն)

Դիցուք ունենք երկու ընդհատ պատահական մեծություններ, որոնց արժեքները կազմում են նույն վերջավոր բազմությունը $x_1,\ldots,x_K$, իսկ հավանականությունների բաշխումներն են՝ $\mathbf{P}_1 = \{p_{1k}\}$ և $\mathbf{P}_2 = \{p_{2k}\}$, $k = \overline{1,K}$:

> $\mathbf{P}_1$ և $\mathbf{P}_2$ բաշխումների **ինֆորմացիոն տարամիտության** կամ **Կուլբակի-Լեյբլերի ինֆորմացիայի**՝ $\mathrm{K}(\mathbf{P}_1||\mathbf{P}_2)$, սահմանումն է՝
>
> $$\mathrm{K}(\mathbf{P}_1||\mathbf{P}_2) = \sum_k p_{1k}\log(p_{1k}/p_{2k}) :$$

Այս գաղափարը առաջարկվել է 1951 թվականին Կուլբակի և Լեյբլերի կողմից որպես $\mathbf{P}_2$ *բաշխումը* $\mathbf{P}_1$ *բաշխումից տարբերելու վիճակագրական չափանիշ*:

Տարամիտությունը կոչում են նաև ընդհանրացված ինֆորմացիա, քանի որ Շենոնի փոխադարձ ինֆորմացիան կարելի է դիտել որպես դրա մասնավոր դեպք: Իսկապես, դիցուք ունենք (X,Y) երկչափ վեկտորի երկու տարբերակ, մեկի բաշխումը p_{nm}-ն է, իսկ մյուսինը՝ $p_{n.}q_{.m}$-ը: Դժվար չէ համոզվել, որ՝

$$\mathrm{K}(\mathbf{P}_{XY}||\mathbf{P}_X\times\mathbf{P}_Y) = \sum_{nm} p_{nm}\log\frac{p_{nm}}{p_{n.}q_{.m}} = \mathbf{I}(X\wedge Y) :$$

Եթե երկու բաշխումները անընդհատ են և տրվում են համապատասխանաբար $f_1(x)$ և $f_2(x)$ խտության ֆունկցիաների միջոցով, ապա տարամիտությունը սահմանվում է ինտեգրալի միջոցով՝

$$\mathrm{K}(\mathbf{P}_1||\mathbf{P}_2) = \int_{\mathcal{X}} f_1(x)\log\frac{f_1(x)}{f_2(x)}dx :$$

Երբեմն $\mathrm{K}(\mathbf{P}_1||\mathbf{P}_2)$ տարամիտությունը վիճակագրական հետազոտություններում անվանում են «հեռավորություն» ըստ Կուլբակի-Լեյբլերի, սակայն հարկավոր է գիտակցել, որ սովորաբար մաթեմատիկայում երկու տարրերի հեռավորություն անվանում են այդ տարրերի այնպիսի ֆունկցիան, որը բացի դրական լինելուց համաչափ է այդ տարրերի նկատմամբ և երեք տարրերի նկատմամբ բավարարում է «եռանկյան անհավասարությանը»: Տարամիտությունը այդ երկու վերջին հատկություններից զուրկ է: Այդ հանգամանքը, սակայն, չի խանգարում դրա լայն օգտագործմանը մաթեմատիկական վիճակագրությունում, ինչպես նաև ինֆորմացիայի տեսությունում:

> Տարամիտությունը ընդունում է ոչ բացասական արժեքներ՝
> $$\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2) \geq 0,$$
> ընդ որում հավասար է զրոյի այն և միայն այն դեպքում, երբ $\mathbf{P}_1$ և $\mathbf{P}_2$ բաշխումները համընկնում են:

Ապացուցումը ստացվում է նորից $\ln \alpha \leq \alpha - 1$ տարրական անհավասարության օգնությամբ:

<u>*Օրինակ*</u> 20: $\mathcal{X} = \{0, 1\}$ -ի վրա դիտարկենք $\mathbf{P}_1$ և $\mathbf{P}_2$ բաշխումները՝

$$\mathbf{P}_1(0) = 1 - r,\ \mathbf{P}_1(1) = r \text{ և } \mathbf{P}_2(0) = 1 - s,\ \mathbf{P}_2(1) = s :$$

Հաշվել $\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2)$-ը և $\mathbf{K}(\mathbf{P}_2||\mathbf{P}_1)$-ը $r = 1/2$, $s = 1/4$ դեպքում:

Լուծում: Համաձայն սահմանման՝

$$\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2) = (1 - r)\log\frac{1-r}{1-s} + r\log\frac{r}{s}$$

և

$$\mathbf{K}(\mathbf{P}_2||\mathbf{P}_1) = (1 - s)\log\frac{1-s}{1-r} + s\log\frac{s}{r} :$$

Տեղադրենք $r = 1/2$, $s = 1/4$, կստանանք՝

$$\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2) = 1/2\log\frac{1/2}{3/4} + 1/2\log\frac{1/2}{1/4} = 1 - 1/2\log 3 = 0.2075\,\text{բիթ},$$

$$\mathbf{K}(\mathbf{P}_2||\mathbf{P}_1) = 3/4\log\frac{3/4}{1/2} + 1/4\log\frac{1/4}{1/2} = 3/4\log 3 - 1 = 0.1887\,\text{բիթ} :$$

Ինչպես տեսնում ենք, ընդհանրապես $\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2) \neq \mathbf{K}(\mathbf{P}_2||\mathbf{P}_1)$: Հեշտ է նաև համոզվել, որ $r = s$ դեպքում $\mathbf{K}(\mathbf{P}_1||\mathbf{P}_2) = \mathbf{K}(\mathbf{P}_2||\mathbf{P}_1) = 0$:

Գլուխ 4
Հավանականության տեսության սահմանային թեորեմները

Եկեք օգնենք իրար ավելի լավ տեսնել իրերը:

Կլոդ Մոնե

4.1. Ներածություն: Զուգամիտություն ըստ հավանականության

Այս գլխում կներկայացվեն մեծ թվերի օրենքի և կենտրոնական սահմանային թեորեմի հաճախ կիրառվող պարզագույն տարբերակները:

Մեծ թվերի օրենքը պնդում է, որ մեծաքիվ անկախ պատահական մեծությունների միջին թվաբանականը «մոտ է» այդ պատահական մեծությունների սպասելիների միջին թվաբանականին: Վաղուց արձանագրվել է, որ թեև միևնույն պայմաններում իրականացված հերթական դիտարկումների (տնտեսագիտական, ժողովրդագրական, բնագիտական, տեխնիկական և այլն) առանձին թվային արդյունքները նկատելիորեն տարբերվում են իրարից, այդուամենայնիվ մեծ թվով դիտարկումների միջինները օժտված են կայունությամբ: Այդ օրինաչափության բացատրությունը հետևյալն է: Միևնույն պայմաններում կատարվող առանձին փորձերի արդյունքները պատահականորեն տատանվում են որոշակի հաստատունի շուրջ, իսկ մեծ թվով փորձերի արդյունքների շեղումների փոխադարձ չեզոքացման հետևանքով այդ արդյունքների միջին թվաբանականը բավականին կայուն է և, փորձերի թիվը մեծացնելիս, ավելի ու ավելի է մոտենում այդ հաստատունին: Նշված երևույթը մաթեմատիկորեն ձևակերպվում և հիմնավորվում է մի խումբ թեորեմներով, որոնք մեծ թվերի օրենքի մասնավոր դեպքերն են:

Ուսումնասիրվում են նաև այն օրինաչափությունները, որոնց ենթարկվում է մեծ թվով պատահական մեծությունների նորմավորված գումարը: Պարզվում է, որ որոշ պայմանների դեպքում գումարելիների քանակը անսահմանափակորեն մեծացնելիս այդ գումարի բաշխման օրենքը «մոտենում է» նորմալ բաշխման օրենքին: Այս թեորեմների խումբը կրում է կենտրոնական սահմանային թեորեմ անվանումը:

Հայտնի են պատահական մեծությունների հաջորդականությունների զուգամիտության մի քանի տարբեր սահմանումներ: Մենք կծանոթանանք ըստ հավանականության զուգամիտության սահմանմանը :

> Պատահական մեծությունների $X_1, X_2, \ldots, X_n, \ldots$ հաջորդականությունը **զուգամիտում է ըստ հավանականության** X պատահական մեծությանը, եթե կամայական $\epsilon > 0$ համար
>
> $$\lim_{n\to\infty} \mathbf{P}\{|X_n - X| < \epsilon\} = 1,$$
>
> կամ (որ համարժեք է)՝
>
> $$\lim_{n\to\infty} \mathbf{P}\{|X_n - X| \geq \epsilon\} = 0 :$$
>
> Կրճատ գրում են $X_n \underset{n\to\infty}{\Rightarrow} X$, կամ $X_n \xrightarrow[n\to\infty]{P} X$:

4.2. Չեբիշևի անհավասարությունը

Հավանականության տեսությունում մեծ կարևորություն ունեն հետևյալ անհավասարությունները:

Եթե X պատահական մեծությունը ոչ բացասական է և ունի սպասելի, ապա կամայական α դրական թվի համար տեղի ունի **Մարկովի անհավասարությունը**՝

$$\mathbf{P}\{X < \alpha\} \geq 1 - \frac{\mathbf{E}(X)}{\alpha}, \tag{1}$$

կամ, որ նույնն է՝

$$\mathbf{P}\{X \geq \alpha\} \leq \frac{\mathbf{E}(X)}{\alpha}: \tag{2}$$

Ապացույց: Ենթադրենք, որ X-ը անընդհատ պատահական մեծություն է՝ տրված իր $f_X(x)$ խտությամբ: Ապացույցը ստացվում է հետևյալ դատողություններից՝

$$\mathbf{E}(X) = \int_0^\infty x f_X(x)dx = \int_0^\alpha x f_X(x)dx + \int_\alpha^\infty x f_X(x)dx \geq \int_\alpha^\infty x f_X(x)dx \geq \alpha \int_\alpha^\infty f_X(x)dx,$$

հետևաբար՝

$$\mathbf{P}\{X \geq \alpha\} = \int_\alpha^\infty f_X(x)dx \leq \frac{\mathbf{E}(X)}{\alpha}:$$

Ընդհատ X պատահական մեծության համար ապացույցը նման է. ուղղակի սպասելին ինտեգրալի փոխարեն արտահայտվում է գումարով:

Մարկովի անհավասարությունը հնարավորություն է տալիս որոշ դատողություններ անել դրական արժեքներ ընդունող պատահական մեծությանը վերաբերող հավանականությունների մասին, իմանալով միայն նրա սպասելին:

<u>Օրինակ</u> 1: Խնայբանկում ներդրումների միջին չափը 20 000 դրամ է: Հարկավոր է գնահատել պատահականորեն նշված հաշվում 100 000 դրամից պակաս գումար լինելու հավանականությունը:

Լուծում: Պատահականորեն նշված հաշվի վրա եղած գումարը նշանակենք X: Ըստ պայմանի՝ $X \geq 0$, $\mathbf{E}(X) = 20000$, $\alpha = 100000$: Կիրառելով Մարկովի (1) անհավասարությունը, կստանանք՝

$$\mathbf{P}\{X < 100\,000\} \geq 1 - \frac{20\,000}{100\,000} = 0.8:$$

<u>Օրինակ</u> 2: Դիցուք $\phi(x)$-ը միընթացորեն աճող դրական ֆունկցիա է: Եթե կամայական X անընդհատ պատահական մեծության համար գոյություն ունի $\mathbf{E}[\phi(X)] = m$ սպասելին, ապա

$$\mathbf{P}(X > t) \leq \frac{m}{\phi(t)}:$$

Լուծում: $\phi(x)$ ֆունկցիայի հատկություններից հետևում է՝

$$\mathbf{P}(X > t) = \int_{x>t} f_X(x)dx \leq \frac{1}{\phi(t)} \int_{x>t} \phi(x) f_X(x)dx \leq \frac{1}{\phi(t)} \int_{-\infty}^{\infty} \phi(x) f_X(x)dx \leq \frac{m}{\phi(t)},$$

քանի որ

$$\int_{-\infty}^{\infty} \phi(x) f_X(x)dx = m:$$

X պատահական մեծության ցրվածքը՝ D(X)-ը, բնութագրում է $\mathrm{E}(X)$ միջին արժեքի շուրջ X-ի հնարավոր արժեքների սփռվածության աստիճանը, այն առումով, որ ինչքան մեծ է ցրվածքը, այնքան ավելի հավանական են $\mathrm{E}(X)$-ից զգալի շեղումները: Հաջորդ՝ Չեբիշևի անունը կրող անհավասարությունը հնարավորություն է տալիս գնահատել տրված թվից մեծ շեղումների հավանականությունը, գիտենալով միայն պատահական մեծության ցրվածքը:

Եթե X պատահական մեծությունն ունի $\mathrm{E}(X)$ սպասելի և $\mathrm{D}(X)$ ցրվածք, ապա կամայական α դրական թվի համար տեղի ունի **Չեբիշևի անհավասարությունը**՝

$$\mathrm{P}\{|X-\mathrm{E}(X)| \geq \alpha\} \leq \frac{\mathrm{D}(X)}{\alpha^2} :$$

Ապացույց: $|X - \mathrm{E}(X)| \geq \alpha$ անհավասարությունից հետևում է $|X - \mathrm{E}(X)|^2 \geq \alpha^2$ անհավասարությունը: Քանի որ $(X - \mathrm{E}(X))^2 \geq 0$ և $\mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{D}(X)$, ապա Չեբիշևի անհավասարությունը բխում է (2)-ից:

Եթե բաշխումը համաչափ է $\mathrm{E}(X)$ միջինի նկատմամբ, ապա տեղի ունեն նաև միակողմանի անհավասարությունները՝

$$\mathrm{P}\{X-\mathrm{E}(X) \geq \alpha\} = \mathrm{P}\{\mathrm{E}(X)-X \geq \alpha\} \leq \frac{\mathrm{D}(X)}{2\alpha^2} :$$

Անցնելով հակադիր պատահույթի, Չեբիշևի անհավասարությունը կարելի է գրել նաև հետևյալ համարժեք ձևով՝

$$\mathrm{P}\{|X-\mathrm{E}(X)| < \alpha\} \geq 1-\frac{\mathrm{D}(X)}{\alpha^2} : \qquad (3)$$

Եթե տեղադրենք Չեբիշևի անհավասարության մեջ $\alpha = 3\sigma_X$, որտեղ $\sigma_X = \sqrt{\mathrm{D}(X)}$ միջին քառակուսային շեղումն է, ապա կստանանք՝

$$\mathrm{P}\{|X-\mathrm{E}(X)| \geq 3\sigma_X\} \leq \frac{1}{9\sigma_X^2}\mathrm{D}(X) = \frac{1}{9},$$

այսինքն, կամայական պատահական մեծության՝ իր սպասելիից $3\sigma_X$-ից ավելի շեղվելու հավանականությունը չի գերազանցում 1/9: Այլուսակից կարելի է տեղեկանալ, որ նորմալ բաշխման համար նշված հավանականությունը հավասար է 0.0027: Տեսնում ենք, որ նորմալ բաշխման դեպքում Չեբիշևի անհավասարության տված գնահատականը բավականին կոպիտ է:

<u>Օրինակ</u> 3: Միանման մանրամասեր պատրաստելու ժամանակ դրանց տրամագծի 10 մմ-ից թույլատրելի շեղումը հավասար է ± 0.1մմ: Հարկավոր է գնահատել պատահականորեն վերցված մանրամասի խոտանված լինելու հավանականությունը, եթե տրամագծերի ցրվածքը հավասար է 0.0025:

Լուծում: Նշանակենք պատահականորեն ընտրված մանրամասի տրամագիծը X-ով: Ըստ պայմանի մանրամասը խոտանված է, եթե $|X - 10| > 0.1$: Հետևաբար, օգտվելով Չեբիշևի անհավասարությունից, կստանանք՝

$$\mathrm{P}\{|X-10| > 0.1\} \leq \frac{0.0025}{0.01} = 0.25 :$$

Այսինքն, պահանջվող հավանականությունը չի գերազանցում 1/4:

<u>Օրինակ</u> 4: Կատարվում են զառի նետման 10 000 անկախ փորձեր: Գնահատենք վեց միավորի հանդես գալու հաճախությունը իր p հավանականությունից 0.01-ից ոչ ավել չափով շեղվելու հավանականությունը:

Լուծում: Ըստ պայմանի՝ $N = 10\,000$, $p = 1/6$, $q = 5/6, \alpha = 0.01$: Նշանակենք X_N-ով N փորձերում վեց միավորի հանդես գալու անգամների թիվը, $\mathbf{D}(X_N) = \frac{pq}{N}$: Կիրառելով (3) անհավասարությունը, կստանանք՝

$$\mathbf{P}\left\{\left|\frac{X_N}{10\,000} - \frac{1}{6}\right| < 0.01\right\} \geq 1 - \frac{1/6 \cdot 5/6}{10\,000 \cdot (0.01)^2} \approx 0.86 :$$

Չեբիշևի (3) անհավասարությունից N հատ պատահական մեծությունների միջինի վերաբերյալ կարելի է ստանալ հետևյալ անհավասարությունը:

Սպասելի և ցրվածք ունեցող $X_1, ..., X_N$ անկախ պատահական մեծությունների և կամայական $\epsilon > 0$ համար՝

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n)\right| < \epsilon\right\} \geq 1 - \frac{1}{N^2\epsilon^2}\sum_{n=1}^{N}\mathbf{D}(X_n) : \qquad (4)$$

Ապացույց: Սպասելիի և ցրվածքի հատկություններից ստանում ենք՝

$$\mathbf{E}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n),$$

$$\mathbf{D}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2}\sum_{n=1}^{N}\mathbf{D}(X_n) :$$

Կիրառելով Չեբիշևի անհավասարությունը, կստանանք՝

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n)\right| < \epsilon\right\} \geq 1 - \frac{1}{\epsilon^2 N^2}\sum_{n=1}^{N}\mathbf{D}(X_n) :$$

4.3. Մեծ թվերի օրենքը

Մեծ թվերի օրենքը ընդհանուր սկզբունք է, ըստ որի պատահական գործոնների համատեղ ազդեցությունը բավականին ընդհանուր պայմանների առկայության դեպքում հանգեցնում է միջինում ոչ պատահական արդյունքի:

Պատմականորեն մեծ թվերի օրենքի առաջին պարզագույն տարբերակը Բեռնուլիի թեորեմն էր (1713թ.)՝ անկախ փորձերում հաճախությունների կայունության վերաբերյալ: Մեծ թվերի օրենքի առաջին մասնավոր ձևակերպումը տվել է ֆրանսիացի մաթեմատիկոս Ս. Պուասոնը (1837թ.): Հետագայում ավելի ընդհանուր թեորեմներ են ապացուցել Պ. Լ. Չեբիշևը (1866թ.), Ա. Ա. Մարկովը, Ա. Յու. Խինչինը, Ա. Ն. Կոլմոգորովը և ուրիշներ:

Մեծ թվերի օրենքը պնդում է, որ անկախ պատահական մեծությունների թվի անսահմանափակ աճի դեպքում այդ պատահական մեծությունների միջին թվաբանականը ըստ հավանականության զուգամիտում է նրանց սպասելիների միջին թվաբանականին:

Չեբիշևի թեորեմը: Եթե $X_1, X_2, ..., X_N, ...$ անկախ պատահական մեծությունները ունեն $\mathbf{E}(X_n)$ սպասելիներ, և նրանց ցրվածքները սահմանափակ են միևնույն C հաստատունով՝ $\mathbf{D}(X_n) \leq C, n = 1, 2, ...$, ապա կամայական $\epsilon > 0$ համար՝

$$\lim_{N\to\infty}\mathbf{P}\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n)\right| < \epsilon\right\} = 1, \qquad (5)$$

կամ, այլ կերպ ասած՝

$$\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n) \underset{N\to\infty}{\Rightarrow} 0 :$$

Ապացուցում: Քանի որ $\mathrm{D}(X_n) \le C$, $n = 1, 2, \ldots$, ապա

$$\frac{1}{N}\sum_{n=1}^{N} \mathrm{D}(X_n) \le C:$$

Ըստ (4)-ի՝

$$\mathbf{P}\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N} \mathbf{E}(X_n)\right| < \epsilon\right\} \ge 1 - \frac{C}{N\epsilon^2}:$$

Անցնելով սահմանի, երբ $N \to \infty$, կստանանք (5)-ը:

Կիրառենք Չեբիշևի թեորեմը միատեսակ բաշխված պատահական մեծությունների հաջորդականության նկատմամբ: Այդ դեպքում $\mathrm{E}(X_n) = m$, $n = 1, 2, \ldots$, և ստանում ենք՝

$$\frac{1}{N}\sum_{n=1}^{N} X_n \underset{N\to\infty}{\Rightarrow} m:$$

Մեծ թվերի օրենքն ունի կիրառական մեծ նշանակություն: Այն թույլ է տալիս՝ օգտվելով տվյալների միջին թվաբանականից, պատկերացում կազմել սպասելիի մասին: Այսպես, օրինակ, որևէ պարամետրի վերաբերյալ սարքի միջոցով չափումներ կատարելիս, կարելի է մեծ թվով չափումների միջին թվաբանականի միջոցով գնահատել պարամետրի իրական արժեքը: Մեծ թվերի օրենքի կարևոր հետևանքն է

Բեռնուլիի թեորեմը: Դիցուք X_N-ը Բեռնուլիի N փորձերում p հավանականությամբ հաջողությունների թիվն է: Այդ դեպքում X_N/N հաճախությունները ըստ հավանականության ձգտում են p-ին, երբ $N \to \infty$, $X_N/N \underset{N\to\infty}{\overset{P}{\to}} p$, այսինքն, կամայական $\epsilon > 0$ համար՝

$$\mathbf{P}\left\{\left|\frac{X_N}{N} - p\right| < \epsilon\right\} \underset{N\to\infty}{\to} 1:$$

Ապացուցում: Հիշենք, որ n-րդ փորձի հաջողության I_n հայտիչն ընդունում է 0 արժեքը q հավանականությամբ և 1 արժեքը՝ p հավանականությամբ: Գիտենք, որ

$$X_N = \sum_{n=1}^{N} I_n,$$

$$\mathbf{E}(I_n) = p,\ \mathbf{D}(I_n) = pq,\ \mathbf{E}\left(\frac{X_N}{N}\right) = \mathbf{E}\left(\frac{1}{N}\sum_{n=1}^{N} I_n\right) = p:$$

Ուստի կիրառելով (5) անհավասարությունը կամայական $\epsilon > 0$ համար, ստանում ենք՝

$$\mathbf{P}\left\{\left|\frac{X_N}{N} - p\right| < \epsilon\right\} \to 1, \text{երբ } N \to \infty:$$

Ինչպես նշվել է գլուխ 1-ում, Բեռնուլիի թեորեմը հավանականության որոշման վիճակագրական մոտեցման հիմքն է:

4.4. Կենտրոնական սահմանային թեորեմը

Այս ենթաբաժնում կդիտարկվեն պատահական մեծությունների կանոնաձև գումարների բաշխման օրենքի զուգամիտության մասին պնդումները, որոնք անվանում են սահմանային թեորեմներ:

Դրանց շարքում կարևոր տեղ են զբաղեցնում այսպես կոչված կենտրոնական սահմանային թեորեմի տարբերակները, որոնք տալիս են Գաուսի բաշխմանը զուգամիտելու պայմանները: Քանի որ այդ պայմանները բավականին ընդհանուր են և իրականում հաճախ կատարվող, ապա նորմալ բաշխման օրենքը լայնորեն կիրառելի է:

Մեզ արդեն հայտնի է Մուավրի-Լապլասի ինտեգրալ սահմանային թեորեմը (տե՛ս գլուխ 2), որն ավելի ընդհանուր կենտրոնական սահմանային թեորեմների մասնավոր դեպքն է: Մենք կձևակերպենք (առանց ապացուցման) կենտրոնական սահմանային թեորեմի երկու ավելի ընդհանուր տարբերակներ:

Դիտարկենք $X_1, X_2, ..., X_n, ...$ անկախ պատահական մեծություններից կազմված $X(N) = \sum_{n=1}^{N} X_n$ գումարները, և Y_N կանոնաձև պատահական մեծությունները՝

$$Y_N = \frac{X(N) - \mathbf{E}(X(N))}{\sqrt{\mathbf{D}(X(N))}},\ N = 1, 2, ... :$$

Կենտրոնական սահմանային թեորեմ (Լյապունովի թեորեմը). Դիցուք $X_1,\ X_2, \ldots,\ X_n, \ldots$ հաջորդականությունը կազմված է երրորդ կարգի վերջավոր բացարձակ մոմենտներ ունեցող՝

$$\nu_3(n) = \mathbf{E}|X_n - \mathbf{E}(X_n)|^3,\ n = 1, 2, ...,$$

անկախ պատահական մեծություններից: Եթե

$$\lim_{N\to\infty} \frac{\sqrt[3]{\nu_3(1) + \nu_3(2) + ... + \nu_3(N)}}{\sqrt{\mathbf{D}(X_1) + \mathbf{D}(X_2) + ... + \mathbf{D}(X_N)}} = 0,$$

ապա կամայական x-ի համար $(-\infty < x < \infty)$ տեղի ունի

$$\mathbf{P}\{Y_N < x\} \underset{N\to\infty}{\to} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

զուգամիտությունը:

Հիշենք (տե՛ս գլուխ 2), որ Լապլասի ֆունկցիան (արժեքները տրված են աղյուսակում)՝

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-t^2/2} dt,\ x \geq 0,\ \Phi(-x) = -\Phi(x),$$

տալիս է՝

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2} dt = \Phi(\beta) - \Phi(\alpha) :$$

Եթե $X_1, X_2, ..., X_n, ...$ պատահական մեծությունները միանման են բաշխված, ապա թեորեմի պայմանները կարելի է թուլացնել՝ բավական է X_n-երի ցրվածքների գոյությունը:

Լինդեբերգի-Լևիի թեորեմը: Եթե $X_1, X_2, ..., X_n, ...$ վերջավոր ցրվածք ունեցող անկախ միանման բաշխված պատահական մեծությունների հաջորդականություն է՝ $\mathbf{E}(X_n) = m$, $\mathbf{D}(X_n) = \sigma^2 > 0$, $n = 1, 2, ...$, ապա կամայական x_1-ի ու x_2-ի համար $(-\infty < x_1 < x_2 < \infty)$, երբ $N \to \infty$

$$\mathbf{P}\{x_1 < Y_N < x_2\} = \mathbf{P}\left\{x_1 < \frac{X(N) - Nm}{\sigma\sqrt{N}} < x_2\right\} \to \Phi(x_2) - \Phi(x_1) :$$

Լինդեբերգի-Լևիի թեորեմից մեծ N-երի դեպքում ստանում ենք անկախ միանման բաշխված պատահական մեծությունների կանոնաձև գումարի բաշխման ֆունկցիայի մոտավոր բանաձևը՝

$$\mathbf{P}\{\alpha < Y_N < \beta\} \approx \Phi(\beta) - \Phi(\alpha) :$$

Այստեղից կարող ենք ստանալ՝

$$\mathbf{P}\{a < X(N) < b\} = \mathbf{P}\left\{\frac{a - Nm}{N\sigma^2} < Y_N < \frac{b - Nm}{N\sigma^2}\right\} \approx$$
$$\approx \Phi\left(\frac{b - Nm}{\sigma\sqrt{N}}\right) - \Phi\left(\frac{a - Nm}{\sigma\sqrt{N}}\right): \qquad (6)$$

Կենտրոնական սահմանային թեորեմը ընդհանրացված և ապացուցված է տարբեր դեպքերի համար, երբ X_n մեծությունները կախյալ են, երբ դրանք բազմաչափ են, և այլն:

Որևէ պարամետր չափելու ժամանակ չափման սխալն արդյունք է մեծ թվով պատահական գումարելիների: Ըստ կենտրոնական սահմանային թեորեմի՝ գումարային սխալն ունի նորմալ բաշխմանը մոտ բաշխում:

Կենտրոնական սահմանային թեորեմը հիմնավորում է նորմալ բաշխման կարևոր դերը տեսական և կիրառական վիճակագրական հետազոտություններում:

Այդ թեորեմի առաջին դրսևորումներից են միևնույն պայմաններում հրաձգության հետևանքով գնդակի՝ նպատակակետից շեղումները: Այստեղ ներկա են մի շարք գործոններ (քամին, զենքի տատանումը և այլն), որոնք պատահականորեն ազդում են հրաձգության արդյունքի վրա: Սակայն այդ գործոնների համեմատական համարժեքության և դրանց մեծ թվի պատճառով շեղումների գումարը կունենա մոտավորապես նորմալ բաշխում:

Տնտեսագիտության մեջ նույնպես որոշ ցուցանիշներ մեծ թվով գործոնների, որոնցից յուրաքանչյուրը փոքր դեր է կատարում, ազդեցության արդյունք են: Ուստի այդպիսի ցուցանիշները ունեն մոտավորապես նորմալ բաշխում:

Սահմանային թեորեմների վրա է հիմնված մի հաշվողական եղանակ, որը կոչվում է **վիճակագրական փորձարկումների** (կամ Մոնտե-Կարլոյի) **եղանակ**: Այն հաճախ կիրառում են ինտեգրալների մոտավոր հաշվարկի համար, եթե ինտեգրալը չի արտահայտվում տարրական ֆունկցիաների միջոցով, բարձր կարգի գծային հանրահաշվական հավասարումների համակարգերի լուծման համար և բազմաթիվ այլ խնդիրներում, երբ սովորական հաշվման մեթոդները հանդիպում են բարդությունների:

Օրինակ 5: Ապահովագրական ընկերությունում ապահովագրված են 10000 ավտոմեքենա: Վթարի հետևանքով մեկ մեքենայի վնասվածք ստանալու հավանականությունը հավասար է 0.006: Դիցուք մեկ տարով մեքենայի ապահովագրման վարձը 240 դրամ է, իսկ վնասվածք ստանալու դեպքում մեքենայի տիրոջը վճարվում է 20000 դրամ: Գտնել հետևյալ պատահույթների հավանականությունները. A = {տարվա ավարտին ապահովագրական ընկերությունը վնաս կկրի}, B_C = {տարվա ավարտին ապահովագրական ընկերությունը կստանա ոչ պակաս քան C դրամ օգուտ}, եթե C = 800000, 1200000, 1600000:

Լուծում: Նշանակենք X_n-ով մեկ մեքենայից ապահովագրական ընկերության որոշակի շահույթ (կամ կորուստ) ստանալը արտահայտող պատահական մեծությունը: Խնդրից երևում է, որ

$$\mathbf{P}\{X_n = 240\} = 1 - 0.006 = 0.994,$$
$$\mathbf{P}\{X_n = -(20000 - 240)\} = 0.006:$$

Հաշվենք՝

$$m = \mathbf{E}X_n = 240 \cdot 0.994 - 19760 \cdot 0.006 = 238.56 - 118.56 = 120,$$
$$\sigma^2 = \mathbf{D}X_n = (240 - 120)^2 \cdot 0.994 + (-19760 - 120)^2 \cdot 0.006 = 14313.6 + 2371286.4 = 2385600,$$
$$\sigma = 1544.54:$$

Հարկավոր է գտնել (համարելով ավտովթարները իրարից անկախ)

$$\mathbf{P}\{A\} = \mathbf{P}\{X(10000) < 0\} = \mathbf{P}\left\{-\infty < \frac{X(10000) - 10000 \cdot 120}{1544.54 \cdot 100} < 0\right\},$$

$$\mathbf{P}\{B_C\} = \mathbf{P}\{X(10000) > C\} = \mathbf{P}\left\{\frac{C - 10000 \cdot 120}{1544.54 \cdot 100} < \frac{X(10000) - 10000 \cdot 120}{1544.54 \cdot 100} < \infty\right\} =$$

$$= 1/2 - \Phi\left(\frac{C - 1200000}{154454}\right)$$

հավանականությունները: Օգտվելով Լինդեբերգի-Լևիի թեորեմից (բանաձև (6)), գտնում ենք՝

$$\mathbf{P}\{A\} = \Phi\left(\frac{-1200000}{154454}\right) + 0.5 = \Phi(7.77) + 0.5 = 0.0000,$$

$$\mathbf{P}\{B_{800000}\} = 0.5 - \Phi\left(\frac{800000 - 1200000}{154454}\right) = 0.5 + \Phi(2.59) = 0.9952,$$

$$\mathbf{P}\{B_{1200000}\} = 0.5, \;\; \mathbf{P}\{B_{1600000}\} = 0.5 - \Phi(2.59) = 0.0048 :$$

Այսպիսով, նշված պայմաններում ապահովագրական ընկերությունը մեկ տարում մեկին մոտ հավանականությամբ կվաստակի ոչ պակաս քան 800000 դրամ:

Օրինակ 6: Դիցուք X_1-ը, X_2-ը,...,X_N-ը $[0,1]$ հատվածի վրա հավասարաչափ բաշխված անկախ պատահական մեծություններ են: $N = 108$ համար գտնել

$$\mathbf{P}\{50 < X(108) < 60\}:$$

Լուծում: Գիտենք (տե՛ս գլուխ 3), որ $\mathbf{E}(X_n) = 1/2$, $\mathbf{D}(X_n) = 1/12$, $n = \overline{1, N}$, հետևաբար, հաշվի առնելով պատահական մեծությունների անկախությունը, ունենք՝

$$\mathbf{E}(X(108)) = 108 \cdot 1/2 = 54, \;\; \mathbf{D}(X(108)) = 108 \cdot 1/12 = 9, \;\; \sigma_{X(108)} = \sqrt{9} = 3 :$$

Ըստ (6) բանաձևի՝

$$\mathbf{P}\{50 < X(108) < 60\} \approx \Phi\left(\frac{60 - 54}{3}\right) - \Phi\left(\frac{50 - 54}{3}\right) = \Phi(2) - \Phi(-4/3) = 0.885 :$$

Հարկ է նշել, որ կենտրոնական սահմանային թեորեմը վիճակագրական ուսումնասիրություններում կիրառելիս անհրաժեշտ է որոշակի զգուշություն: Եթե պատահական մեծությունների գումարի սահմանային բաշխումը որոշ պայմանների դեպքում նորմալ է և կախված չէ գումարելիների բաշխումներից, ապա, զուգամիտության արագությունը զգալիորեն կախված է սկզբնական անդամների բաշխումից: Այսպես, օրինակ, հավասարաչափ բաշխված պատահական մեծությունների գումարի դեպքում 6-10 գումարելիներով կարելի է բավականաչափ մոտենալ նորմալ բաշխմանը, իսկ որևէ այլ բաշխման համար նույնպիսի արդյունքի կարելի է հասնել միայն շատ ավելի մեծ թվով գումարելիների դեպքում:

Նորմալ բաշխումը միակ սահմանային բաշխումը չէ: Օրինակ, Պուասոնի թեորեմը նշում է Պուասոնի բաշխումը որպես սահմանային դիտարկելու դեպք (տե՛ս գլուխ 2):

Գլուխ 5
Պատահական ընթացքներ

Հավանականության տեսության մեզ ամենագրավիչը նրա կիրառությունների բազմազանությունն է:

Մարկ Կաց

Պատահումները միշտ չէ որ նկարագրվում են Գաուսի օրենքով: Հակառակ դեպքում խոշոր երկրաշարժերը կամ բորսայի սնանկացումները համարյա երբեք տեղի չէին ունենա:

Անդրեյ Կոլմոգորով

5.1. Հիմնական գաղափարներ

Մեզ շրջապատող իրականությունում բազմաթիվ երևույթներ պատահական գործոնների ազդեցությամբ ժամանակի ընթացքում փոփոխման են ենթարկվում, փոփոխվում են նաև այդ երևույթների հատկանիշները նկարագրող պատահական մեծությունները և դրանց հավանականությունների բաշխումները:

> Պատահական մեծության ժամանակից կախվածությունը արտահայտող ֆունկցիան կոչվում է **պատահական ընթացք**:

Օդերևութաբանական հատկանիշների մեծ մասի (օրինակ՝ օդի ջերմաստիճանի, մթնոլորտային ճնշման, տեղումների քանակի) վարքը պատահական ընթացքներ են, որոնք տեղի են ունենում այնպիսի պատահական գործոնների ազդեցությամբ, ինչպիսիք են՝ Արևի ակտիվությունը, ամպամածությունը և այլն:

Բազմաթիվ տնտեսական ցուցանիշներ փոփոխվում են պատահական ձևով, օրինակ՝ որևէ ապրանքի պահանջարկը և առաջարկը, բնակչության քանակը երկրում կամ մարզում, արտադրությունում խոտանի տոկոսը, բուհ ավարտող ուսանողների թիվը, տարադրամի փոխարժեքը: Բերենք ևս մի շարք պատահական ընթացքների օրինակներ՝ կապի համակարգում զբաղված գծերի քանակը՝ կախված ժամանակից, մասնիկի բրոունյան շարժումը, էլեկտրացանցում լարվածության տատանումները, նավահանգստում բեռնաթափման սպասող նավերի քանակը:

Մաթեմատիկական տեսակետից պատահական ընթացքը մեկ պարամետրից կախված պատահական մեծությունների ընտանիք է՝ X_t, $t \in T$: T-ն կոչվում է պատահական ընթացքի **պարամետրական բազմություն** և ընդհանուր առմամբ կարող է լինել կամայական: Մենք կդիտարկենք միայն իրական թվերի՝ $\mathcal{R}$ և ամբողջ թվերի՝ $\mathcal{Z}$ բազմությունները կամ նրանց ենթաբազմությունները: Սովորաբար այդ բազմությունները մեկնաբանվում են որպես ժամանակահատված կամ ժամանակի պահերի վերջավոր կամ հաշվելի բազմություն:

> Կամայական պատահական ընթացքը երկու փոփոխականից ֆունկցիա է՝ $X_t(\omega)$, $t \in T$, $\omega \in \Omega$: Պարամետրի յուրաքանչյուր t_0 արժեքի դեպքում $X_{t_0}(\omega)$-ն պատահական մեծություն է, որը կոչվում է **պատահական ընթացքի բաղադրիչ**: Ամեն մի ω_0-ի համար $\varphi(t) = X_t(\omega_0)$-ն t-ից ոչ պատահական ֆունկցիա է, որը կոչվում է **պատահական ընթացքի իրագործում** կամ **նմուշային ֆունկցիա**:

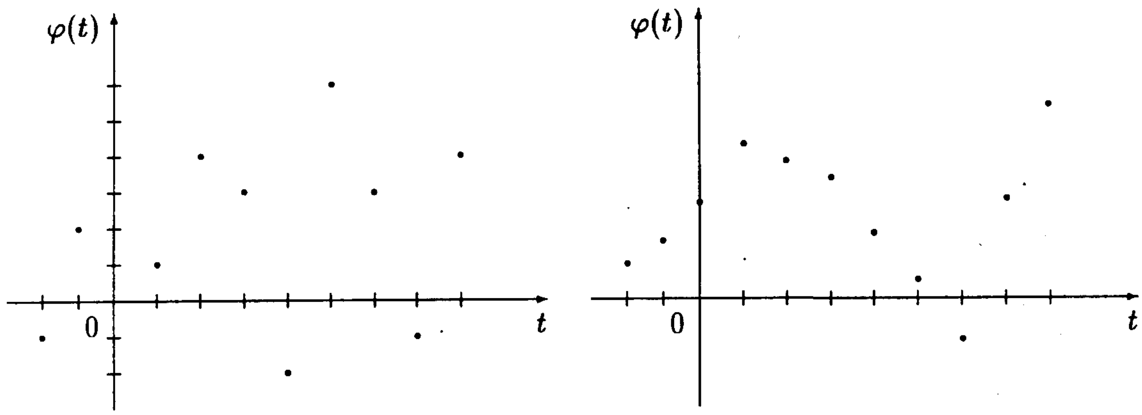
Կախված պարամետրական $\mathcal{T}$ բազմությունից և բաղադրիչների արժեքների (վիճակների) $\mathcal{X}$ տիրույթից, պատահական ընթացքները դասակարգվում են հետևյալ կերպ՝

ընդհատ պարամետրով (ժամանակով) **ընդհատ պատահական ընթացքներ**, օրինակ, ժամանակի որոշակի պահերին փոստարկղից հանված նամակների քանակը,

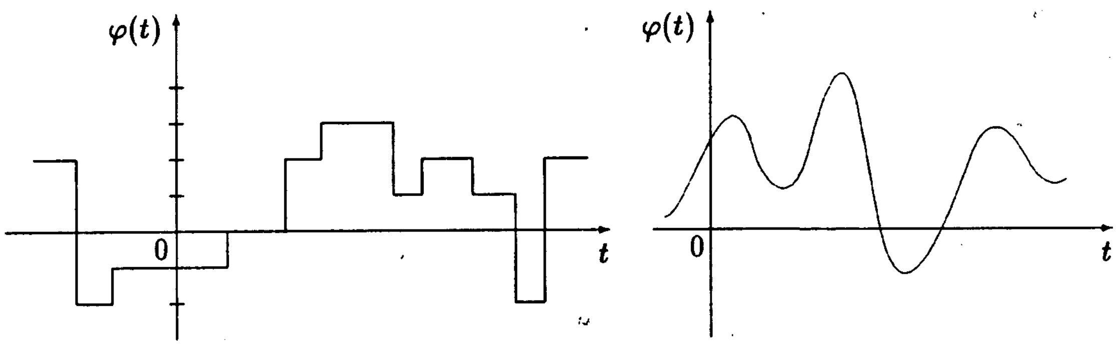
ընդհատ պարամետրով անընդհատ պատահական ընթացքներ, օրինակ, նշված վայրը գնացքի ամենօրյա ժամանելու փաստացի ժամկետը,

անընդհատ պարամետրով ընդհատ պատահական ընթացքներ, օրինակ, բորսայում սակարկությունների ընթացքում մինչ նշված պահը գործարքների քանակը,

անընդհատ պարամետրով անընդհատ պատահական ընթացքներ, օրինակ, էլեկտրական ցանցում հոսանքի լարվածության տատանումները:



Նկար 1: Ընդհատ ժամանակով ընթացքների իրագործումների օրինակներ:



Նկար 2: Անընդհատ ժամանակով ընթացքների իրագործումների օրինակներ:

Դիցուք X_t, $t \in \mathcal{T}$ պատահական ընթացք է: Կամայական $N = 1, 2, \ldots$, և պարամետրի կամայական տարբեր $t_1, t_2, \ldots, t_N \in \mathcal{T}$ արժեքների համար որոշված է **պատահական ընթացքի վերջավորչափանի բաշխման ֆունկցիան**՝

$$F^X_{t_1,t_2,\ldots,t_N}(x_1, x_2, \ldots, x_N) = \mathbf{P}(X_{t_1} < x_1, X_{t_2} < x_2, \ldots, X_{t_N} < x_N),\ x_n \in \mathcal{R},\ n = \overline{1,N}:$$

Պատահական ընթացքի բոլոր վերջավորչափանի բաշխման ֆունկցիաների ընտանիքը՝

$$\{F^X_{t_1,t_2,\ldots,t_N}(x_1, x_2, \ldots, x_N),\ t_1, t_2, \ldots, t_N \in \mathcal{T},\ N = 1, 2, \ldots\},$$

օժտված է հետևյալ կարևոր՝ **համաձայնեցվածության հատկությամբ**, կամայական բնական n, $1 \le n \le N$, թվի համար՝

$$F^X_{t_1,\ldots,t_{n-1},t_n,t_{n+1},\ldots,t_N}(x_1, \ldots, x_{n-1}, \infty, x_{n+1}, \ldots, x_N) =$$
$$= F^X_{t_1,\ldots,t_{n-1}\,t_{n+1},\ldots,t_N}(x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N):$$

Եթե պատահական ընթացքի բաղադրիչները ընդհատ պատահական մեծություններ են, ապա համաձայնեցվածության հատկությունը ընդունում է հետևյալ տեսքը՝

$$\sum_{x_n\in\mathcal{X}} \mathbf{P}\{X_{t_1}=x_1,\ldots,X_{t_{n-1}}=x_{n-1},X_{t_n}=x_n,X_{t_{n+1}}=x_{n+1},\ldots,X_{t_N}=x_N\}=$$
$$=\mathbf{P}\{X_{t_1}=x_1,\ldots,X_{t_{n-1}}=x_{n-1},X_{t_{n+1}}=x_{n+1},\ldots,X_{t_N}=x_N\}:$$

Իսկ եթե պատահական ընթացքի բաղադրիչները անընդհատ պատահական մեծություններ են, ապա՝

$$\int_{\mathcal{X}} f^X_{t_1,\ldots,t_{n-1},t_n,t_{n+1},\ldots,t_N}(x_1,\ldots,x_{n-1},x_n,x_{n+1},\ldots,x_N)\,dx_n=$$
$$=f^X_{t_1,\ldots,t_{n-1},t_{n+1},\ldots,t_N}(x_1,\ldots,x_{n-1},x_{n+1},\ldots,x_N),$$

որտեղ $f^X_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N)$ ֆունկցիան $(X_{t_1},X_{t_2},\ldots,X_{t_N})$ պատահական վեկտորի բաշխման խտությունն է: Այսպիսով,

> Յուրաքանչյուր X_t, $t\in\mathcal{T}$, պատահական ընթացքին համապատասխանում է համաձայնեցված վերջավորչափանի բաշխման ֆունկցիաների ընտանիք:

Տեղի ունի նաև հակառակ փաստը՝

> **Կոլմոգորովի թեորեմը**: Յուրաքանչյուր համաձայնեցված վերջավորչափանի բաշխման ֆունկցիաների ընտանիքի
> $$\{F_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N),\ t_1,t_2,\ldots,t_N\in\mathcal{T},\ N=1,2,\ldots\}$$
> համար կարելի է կառուցել այնպիսի պատահական ընթացք X_t, $t\in\mathcal{T}$, որ
> $$F^X_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N)=F_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N):$$

Այս երկու փաստը թույլ է տալիս չտարբերել պատահական ընթացքի և նրա վերջավորչափանի բաշխման ֆունկցիաների ընտանիքի գաղափարները:

Դիտարկենք պատահական ընթացքների մի քանի օրինակներ:

<u>Օրինակ</u> 1: *Անջատվող փոփոխականներով պատահական ընթացքներ*: Դիցուք $\widetilde{X}=\widetilde{X}(\omega)$, $\omega\in\Omega$, որևէ պատահական մեծություն է, և $\psi(t)$, $t\in\mathcal{T}$, որևէ ոչ պատահական ֆունկցիա է: Հետևյալ արտահայտությունը՝

$$X_t(\omega)=\widetilde{X}(\omega)\psi(t),\ \omega\in\Omega,\ t\in\mathcal{T}, \tag{1}$$

որոշում է պատահական ընթացք, որը կոչվում է **անջատվող փոփոխականներով պատահական ընթացք**: Կախված $\widetilde{X}(\omega)$-ի և $\psi(t)$-ի ընտրությունից, կարելի է ստանալ այդպիսի ընթացքների տարբեր օրինակներ:

ա) Դիցուք $\widetilde{X}(\omega)$-ն ընդունում է -1 և $+1$ արժեքները միևնույն $1/2$ հավանականությամբ, իսկ $\psi(t)=t$: Համապատասխան պատահական ընթացքը ունի միայն երկու իրագործում՝ $\varphi(t)=t$ և $\varphi(t)=-t$ (տես նկար 3):

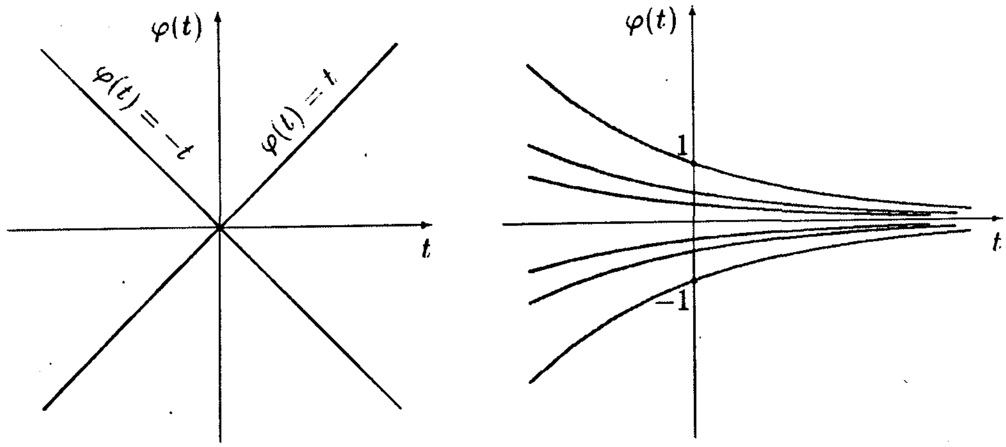
բ) Դիցուք $\widetilde{X}=\widetilde{X}(\omega)$ պատահական մեծությունը հավասարաչափ է բաշխված $[-1,+1]$ միջակայքում, իսկ $\psi(t)=\exp(-t)$, $t\in\mathcal{R}$: Այս դեպքում (1)-ում որոշված ընթացքը ունի անվերջ թվով իրագործումներ: Դրանցից մի քանիսը ներկայացված են նկար 3-ում:

Եթե $\psi(t)$ ֆունկցիան խիստ դրական է՝ $\psi(t)>0$, $t\in\mathcal{R}$, ապա անջատվող փոփոխականներով ընթացքների վերջավորչափանի բաշխման ֆունկցիաները կարելի է գրել ավելի պարզ տեսքով: Իրոք,

$$F^X_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N)=\mathbf{P}\{X_{t_1}<x_1,X_{t_2}<x_2,\ldots,X_{t_N}<x_N\}=$$

$$= \mathbf{P}\{\widetilde{X}\psi(t_1) < x_1, \widetilde{X}\psi(t_2) < x_2, \ldots, \widetilde{X}\psi(t_N) < x_N\} =$$
$$\mathbf{P}\{\widetilde{X} < \frac{x_1}{\psi(t_1)}, \widetilde{X} < \frac{x_2}{\psi(t_2)}, \ldots, \widetilde{X} < \frac{x_N}{\psi(t_N)}\} =$$
$$= \mathbf{P}\{\widetilde{X} < \min_{1\le n\le N} \frac{x_n}{\psi(t_n)}\} = F_{\widetilde{X}}(\min_{1\le n\le N} \frac{x_n}{\psi(t_n)}) :$$

Ինչպես տեսնում ենք, այս դեպքում բոլոր վերջավորչափանի բաշխումները արտահայտվում են $\widetilde{X}$ պատահական մեծության բաշխման ֆունկցիայի միջոցով:



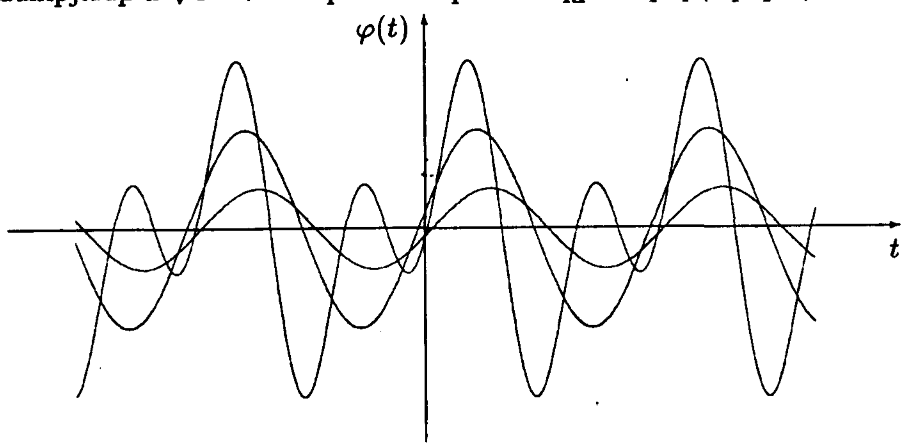
Նկար 3: Անջատվող փոփոխականներով ընթացքների իրագործումները:

Պատահական ընթացքների հետևյալ դասը ունի բազմաթիվ կիրառություններ տնտեսագիտության տարբեր բնագավառներում, երկրաշարժագիտության և օդերևութաբանության մեջ:

<u>Օրինակ</u> 2: *Պատահական տատանումների ընթացք*: Դիցուք

$$X_t = Y \cos \alpha t + Z \sin \alpha t, \quad t \in \mathcal{T},$$

որտեղ (Y, Z)-ը որևէ պատահական վեկտոր է (հաճախ՝ գաուսյան), իսկ α-ն իրական հաստատուն է: Այսպիսի ընթացքի իրագործումները ներդաշնակ տատանումներ են α հաճախականությամբ և $\sqrt{Y^2 + Z^2}$ պատահական ամպլիտուդով (նկար 4):



Նկար 4: Պատահական տատանումների ընթացքի իրագործումները:

<u>Օրինակ</u> 3: *Անկախ արժեքներով պատահական ընթացքներ*: Եթե կամայական N-ի, $N = 1, 2, \ldots$, և կամայական $t_1, t_2, \ldots, t_N \in \mathcal{T}$, համար X_t, $t \in \mathcal{T}$, ($\mathcal{T}$-ն վերջավոր է, կամ հաշվելի), պատահական ընթացքի $X_{t_1}, X_{t_2}, \ldots, X_{t_N}$, բաղադրիչները համախմբությամբ անկախ են, ապա այդպիսի ընթացքը կոչվում է **անկախ արժեքներով պատահական ընթացք**:

Անկախ արժեքներով պատահական ընթացքի վերջավորչափանի բաշխման ֆունկցիաների ընտանիքը լիովին որոշվում է նրա բաղադրիչների մեկչափանի $F_{X_t}(x)$, $t \in \mathcal{T}$ բաշխման ֆունկցիաների միջոցով: Իրոք,

$$F_{t_1,t_2,\dots,t_N}(x_1,x_2,\dots,x_N) = F_{X_{t_1},X_{t_2},\dots,X_{t_N}}(x_1,x_2,\dots,x_N) =$$
$$= F_{X_{t_1}}(x_1)\cdot F_{X_{t_2}(x_2)}\cdots F_{X_{t_N}}(x_N):$$

Ունենալով մեկչափանի $F_n(x), n = 1,2,\dots,$ բաշխման ֆունկցիաների որևէ հաջորդականություն, կարելի է կառուցել վերջավորչափանի բաշխման ֆունկցիաների հետևյալ ընտանիքը՝

$$\{F_{t_1,t_2,\dots,t_N}(x_1,x_2,\dots,x_N) = F_{t_1}(x_1)F_{t_2}(x_2)\cdots F_{t_N}(x_N),$$
$$t_n = 1,2,\dots,\ n=\overline{1,N},\ N=1,2,\dots\}:$$

Հեշտ է տեսնել, որ կառուցված բաշխման ֆունկցիաների ընտանիքը համաձայնեցված է և ըստ Կոլմոգորովի թեորեմի՝ որոշում է անկախ արժեքներով որոշակի պատահական ընթացք:

Նշված դասի պատահական ընթացք է $\Omega = [0,1)$-ի վրա որոշված պատահական մեծությունների հետևյալ հաջորդականությունը՝ $N = 1,2,\dots,$

$$X_N(\omega) = \begin{cases} 1, & \omega \in \left[\dfrac{2k}{2^N}, \dfrac{2k+1}{2^N}\right), \quad k = 0,1,\dots,2^{N-1}-1, \\ 0, & \omega \in \left[\dfrac{2k+1}{2^N}, \dfrac{2(k+1)}{2^N}\right), \quad k = 0,1,\dots,2^{N-1}-1: \end{cases}$$

Այսպիսի պատահական մեծություններ արդեն դիտարկվել են II գլխում, որտեղ ապացուցվել է նրանց անկախությունը:

Պատահական ընթացքների հետևյալ դասն ունի կիրառություններ խաղերի տեսության, տնտեսագիտության, կառավարման տեսության մեջ և այլն:

<u>Օրինակ</u> 4: *Պատահական թափառման ընթացք*: Դիցուք X_n, $n = 1,2,\dots,$ պատահական ընթացք է, որն ունի վերջավոր թվով արժեքներ ընդունող անկախ բաղադրիչներ: Կազմենք նոր պատահական ընթացք՝ $S_N = \sum\limits_{n=1}^{N} X_n$, $N = 1,2,\dots$:

Այդպիսի ընթացքները կոչվում են **պատահական թափառումներ**: Դժվար չէ համոզվել, որ դրանց վերջավորչափանի բաշխումները որոշվում են X_n, $n = 1,2,\dots,$ պատահական ընթացքի մեկչափանի բաշխումների միջոցով: Իսկապես,

$$\mathbf{P}\{S_1 = x_1\} = \mathbf{P}\{X_1 = x_1\},$$
$$\mathbf{P}\{S_1 = x_1,\ S_2 = x_2\} = \mathbf{P}\{X_1 = x_1,\ X_1 + X_2 = x_2\} =$$
$$= \mathbf{P}\{X_1 = x_1,\ X_2 = x_2 - x_1\} = \mathbf{P}\{X_1 = x_1\}\mathbf{P}\{X_2 = x_2 - x_1\},$$

և, ընդհանրապես,

$$\mathbf{P}\{S_1 = x_1,\ S_2 = x_2,\dots,S_N = x_N\} =$$
$$= \mathbf{P}\{X_1 = x_1,\ X_2 = x_2 - x_1,\dots,X_N = x_N - x_{N-1}\} =$$
$$= \mathbf{P}\{X_1 = x_1\}\mathbf{P}\{X_2 = x_2 - x_1\}\cdots\mathbf{P}\{X_N = x_N - x_{N-1}\}:$$

Դիտարկենք պատահական թափառման մի որոշակի դեպք՝ *խաղամոլի սնանկացման մասին խնդիրը*:

Դիցուք խաղասկզբում երկու խաղացող (Ա և Բ) միասին ունեն K դրամական միավոր: Խաղի կանոններն են՝ յուրաքանչյուր հաջող ելքի դեպքում խաղացողն իր խաղընկերոջից շահում է մեկ դրամական միավոր, իսկ անհաջողության դեպքում հակառակորդն է շահում մեկ միավոր: Խաղը շարունակվում է մինչև խաղացողներից մեկի սնանկացումը, այսինքն, մինչև նրա ունեցած գումարը հավասարվի զրոյի:

Պահանջվում է գտնել խաղացողի սնանկացման հավանականությունը:

Թեև տվյալ խնդիրը ձևակերպված է խաղերի տեսության տերմիններով, այն ունի նաև բազմաթիվ կիրառություններ տնտեսագիտության մեջ, օրինակ, ապահովագրական և բանկային ոլորտներում:

Լուծում: Նշանակենք X_n-ով այն պատահական մեծությունը, որը Ա խաղացողի համար n-րդ հերթախաղի բարենպաստ ելքի դեպքում ընդունում է 1 արժեքը, հակառակ դեպքում՝ -1-ը: Ենթադրվում է, որ X_n, $n = 1, 2, \ldots$, անկախ պատահական մեծությունների հաջորդականություն է, ընդ որում $\mathbf{P}\{X_n = 1\} = p$ (Ա խաղացողի շահելու հավանականությունն է), $\mathbf{P}\{X_n = -1\} = q$ (Ա խաղացողի տանուլ տալու հավանականությունն է), p, $q > 0$: Եթե $p > q$, ապա խաղը համարվում է բարենպաստ Ա-ի համար, իսկ եթե $p = q$, ապա խաղը համարվում է արդար (անվնաս): Հասկանալի է, որ $S_N = \sum_{n=1}^{N} X_n$, $N = 1, 2, \ldots$, պատահական թափառման ընթացք է, որի S_N բաղադրիչը ժամանակի N պահին Ա-ի շահույթի (կամ տանուլ տվածի) մեծությունն է, $S_N > 0$ ($S_N < 0$):

Դիցուք Ա-ն սկզբում ուներ k դրամական միավոր, $0 \le k \le K$: Նշանակենք $Q_k^{(N)}$-ով նրա՝ N հերթախաղերից հետո սնանկացման հավանականությունը: Բնական է այդ դեպքում Ա-ի սնանկացման Q_k հավանականությունը վերցնել՝ $Q_k = \lim\limits_{N\to\infty} Q_k^{(N)}$: Տվյալ սահմանը գոյություն ունի, քանի որ $Q_k^{(N)}$ հավանականությունը ըստ N-ի չի նվազում:

Քանի որ Ա-ի դրամագլուխը առաջին հերթախաղից հետո կարող է հավասար լինել $k + 1$, կամ $k - 1$, ապա նա կարող է սնանկանալ երկու դեպքում՝ կամ նա շահում է հերթախաղը և հետագայում սնանկանում՝ ունենալով $k + 1$ միավոր, կամ հերթախաղը տանուլ է տալիս և հետագայում սնանկանում՝ ունենալով $k - 1$ միավոր: Հետևաբար, համաձայն լրիվ հավանականության բանաձևի, կարող ենք գրել

$$Q_k = pQ_{k+1} + qQ_{k-1}: \tag{2}$$

Այստեղ անհրաժեշտ է առանձնացնել $k = 0$ և $k = K$ դեպքերը: Բնական է ընդունել, որ $k = 0$ դեպքում $Q_0 = 1$ (խաղացողը չունի դրամագլուխ, այսինքն նա սնանկացած է ի սկզբանե), իսկ $k = K$ դեպքում ընդունել $Q_K = 0$ (խաղացողը չի կարող սնանկանալ, քանի որ հակառակորդը չունի դրամագլուխ):

Դժվար չէ ստուգել, որ երբ $p \ne q$, ապա (2) հավասարման լուծումը կլինի՝

$$Q_k = \frac{(q/p)^K - (q/p)^k}{(q/p)^K - 1}, \text{ իսկ երբ } p = q = 1/2, \text{ ապա } Q_k = 1 - \frac{k}{K}: \tag{3}$$

Այժմ դիտարկենք խնդրի այն կարևոր դեպքը, երբ Բ-ի դրամագլուխը անսահմանափակ է: Այդ պարագայում Ա-ն, ունենալով k դրամական միավոր, կարող է իր առջև նպատակ դնել՝ կամ սնանկանալ, կամ իր դրամագլուխը հասցնել K միավորի:

Խաղի այդպիսի մեկնաբանությամբ խաղացողի վերջնական շահույթը (նշանակենք այն V_k-ով) պատահական մեծություն է, որի միջին արժեքը հավասար է

$$\mathbf{E}V_k = -kQ_k + (K - k)(1 - Q_k) = K(1 - Q_k) - k:$$

Եթե $p = q = 1/2$, ապա $Q_k = 1 - k/K$, $\mathbf{E}V_k = 0$:

Ստացված բանաձևերից կարելի է անել հետևյալ եզրակացությունները: Եթե խաղացողները նույն հմտության են, այսինքն՝ $p = q = 1/2$, և Ա-ն՝ ունենալով բավականաչափ մեծ k դրամագլուխ, նպատակ է դնում շահել համեմատաբար ոչ մեծ $K - k$ գումար, ապա դրան հասնելու նրա հնարավորությունները բավականաչափ մեծ են: Օրինակ, եթե $k = 9999$, $K = 10000$, ապա համաձայն (3) բանաձևի՝ շահելու հավանականությունը հավասար կլինի $1 - Q_{9999} = 0.9999$:

Խաղի ընթացքում հերթախաղի գումարի (կոպարի) փոփոխությունը նշանակալի ազդեցություն ունի սնանկացման հավանականության վրա: Դիցուք խաղացողների նույն սկզբնական դրամագլուխների դեպքում խաղագումարը երկու անգամ ավելանում է: Դա նշանակում է, որ (3) բանաձևում K-ն պետք է փոխարինել $K/2$-ով, իսկ k-ն՝ $k/2$-ով: Այս դեպքում՝

$$Q_k^* = \frac{(q/p)^{K/2} - (q/p)^{k/2}}{(q/p)^{K/2} - 1} = Q_k \frac{(q/p)^{K/2} + 1}{(q/p)^{K/2} + (q/p)^{k/2}}:$$

Եթե $q > p$ (այսինքն՝ $p < 1/2$) ապա վերջին կոտորակը փոքր է մեկից, և ստանում ենք $Q_k > Q_k^*$, այսինքն՝ խաղացողի սնանկացման հավանականությունը խաղագումարի մեծացման դեպքում փոքրանում է, իսկ հակառակորդի սնանկացման հավանականությունը՝ մեծանում:

Այստեղից հետևում է, որ եթե սկզբնական դրամի չափը բավականին մեծ է, ապա որոշ պայմաններում նպատակահարմար է մասնակցել նույնիսկ ոչ բարենպաստ խաղին: Այսպես, օրինակ, արտակարգ վիճակներից գույքի ապահովագրությունը չնայած կարելի է դիտարկել որպես ոչ բարենպաստ խաղ, հաճախ լինում է հիմնավորված:

5.2. Պատահական ընթացքների բնութագրիչներ

X_t, $t \in \mathcal{T}$, պատահական ընթացքի **սպասելի կամ միջին արժեք** կոչվում է $m(t) = \mathbf{E}X_t$ ոչ պատահական ֆունկցիան, որի արժեքը $t = t_0$ դեպքում հավասար է X_{t_0} պատահական մեծության սպասելիին՝ $m(t_0) = \mathbf{E}X_{t_0}$:

Պատահական ընթացքի սպասելին՝ $m(t)$-ն, այն «միջին» ֆունկցիան է, որի շուրջը «համախմբվում են» ընթացքի իրագործումները:

X_t, $t \in T$, պատահական ընթացքի **ցրվածք** կոչվում է $\sigma^2(t) = \mathbf{D}X_t = \mathbf{E}(X_t - m(t))^2$ ոչ պատահական ֆունկցիան, որի արժեքը $t = t_0$ դեպքում հավասար է X_{t_0} պատահական մեծության ցրվածքին՝ $\sigma^2(t_0) = \mathbf{D}X_{t_0}$: Պատահական ընթացքի ցրվածքը բնութագրում է իրագործումների շեղումը միջինից, այսինքն՝ $m(t)$-ից:

X_t, $t \in \mathcal{T}$, պատահական ընթացքի **համացրվածքային ֆունկցիա** կոչվում է

$$\mathbf{B}(t,s) = \operatorname{cov}(X_t, X_s) = \mathbf{E}(X_t - m(t))(X_s - m(s))$$

ոչ պատահական ֆունկցիան, որի արժեքը $t = t_0$ և $s = s_0$ դեպքում հավասար է X_{t_0} և X_{s_0} պատահական մեծությունների համացրվածքին՝

$$\mathbf{B}(t_0, s_0) = \operatorname{cov}(X_{t_0}, X_{s_0}):$$

Նկատենք, որ եթե $t = s$, ապա $\mathbf{B}(t,s) = \sigma^2(t)$:

X_t, $t \in \mathcal{T}$, պատահական ընթացքի **հարաբերակցության ֆունկցիա** կոչվում է

$$\rho(t,s) = \frac{\mathbf{B}(t,s)}{\sigma(t)\sigma(s)}, \quad \sigma(t) > 0, \ \sigma(s) > 0,$$

ոչ պատահական ֆունկցիան, որի արժեքը $t = t_0$ և $s = s_0$ դեպքում հավասար է X_{t_0} և X_{s_0} պատահական մեծությունների հարաբերակցության գործակցին՝

$$\rho(t_0, s_0) = \frac{\mathbf{E}(X_{t_0} - \mathbf{E}X_{t_0})(X_{s_0} - \mathbf{E}X_{s_0})}{\sqrt{\mathbf{D}X_{t_0}}\sqrt{\mathbf{D}X_{s_0}}}:$$

Հարաբերակցության ֆունկցիան բնութագրում է ժամանակի կամայական t և s պահերին X_t և X_s պատահական մեծությունների կախվածության աստիճանը:

<u>*Օրինակ*</u> 5: Դիցուք $X_t = Y\cos\alpha t + Z\sin\alpha t$, $t \in \mathcal{R}$, պատահական տատանումների ընթացք է, որտեղ Y-ը և Z-ը անկախ, $(0, \sigma^2)$ պարամետրերով բաշխված գաուսյան պատահական մեծություններ են: Գտնենք այդ ընթացքի բնութագրիչները:

Լուծում:

$$m(t) = \mathbf{E}(Y\cos\alpha t + Z\sin\alpha t) = \mathbf{E}Y \cdot \cos\alpha t + \mathbf{E}Z \cdot \sin\alpha t = 0,$$

$$B(t,s) = \mathbf{E}(Y\cos\alpha t + Z\sin\alpha t)(Y\cos\alpha s + Z\sin\alpha s) =$$

$$= \mathbf{E}Y^2\cos\alpha t \cdot \cos\alpha s + \mathbf{E}YZ \cdot (\cos\alpha t \cdot \sin\alpha s + \sin\alpha t \cdot \cos\alpha s) +$$

$$+ \mathbf{E}Z^2\sin\alpha t \cdot \sin\alpha s = \sigma^2(\cos\alpha t \cdot \cos\alpha s + \sin\alpha t \cdot \sin\alpha s) = \sigma^2\cos\alpha(s-t),$$

$$\sigma^2(t) = \sigma^2 \cos\alpha(t-t) = \sigma^2, \quad \rho(s,t) = \frac{\sigma^2\cos\alpha(s-t)}{\sigma^2} = \cos\alpha(s-t):$$

5.3. Պատահական ընթացքների դասերի օրինակներ

Անկախ աճերով պատահական ընթացքներ: X_t, $t \in \mathcal{T}$, պատահական ընթացքը կոչվում է անկախ աճերով ընթացք, եթե կամայական $t_0 < t_1 < t_2 < \ldots < t_N$, $t_n \in \mathcal{T}$, $n = \overline{1,N}$, $N = 1,2,\ldots$, համար $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_N} - X_{t_{N-1}}$ պատահական մեծությունները անկախ են: Այդ ընթացքի վերջավորչափանի բաշխումների որոշման համար բավական է իմանալ X_{t_0} և $X_{t_n} - X_{t_{n-1}}$, $n = \overline{1,N}$, $N = 1,2,\ldots$, աճերի բաշխումները:

Ստացիոնար (նեղ իմաստով) պատահական ընթացքներ: X_t, $t \in \mathcal{T}$, պատահական ընթացքը կոչվում է ստացիոնար (նեղ իմաստով), եթե պարամետրի կամայական $t_1, t_2, \ldots, t_N \in \mathcal{T}$ արժեքների և կամայական h-ի, $h \in \mathcal{T}$, համար †

$$F_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N) = F_{t_1+h,t_2+h,\ldots,t_N+h}(x_1,x_2,\ldots,x_N),$$

այսինքն՝ վերջավորչափանի բաշխումները կախված չեն ժամանակի տեղաշարժից:

Ստացիոնար պատահական ընթացքներին հատկանշական է նրանց վիճակագրական բնութագրիչների ժամանակի ընթացքում անփոփոխությունը՝ հաստատուն լինելը:

Հետևյալ դասի վերջավորչափանի բաշխումները լրիվին որոշվում են $m(t)$ և $\mathbf{B}(t,s)$ ֆունկցիաների միջոցով:

Գաուսյան (նորմալ) պատահական ընթացքներ: X_t, $t \in \mathcal{T}$, պատահական ընթացքը կոչվում է գաուսյան, եթե դրա վերջավորչափանի բաշխումները նորմալ են, այսինքն՝ խտությունները ունեն հետևյալ տեսքը՝

$$f_{t_1,t_2,\ldots,t_N}(x_1,x_2,\ldots,x_N) = \frac{|A|^{1/2}}{(2\pi)^{N/2}}\exp\left\{-\frac{1}{2}\left(\sum_{i,j=1}^{N} a_{ij}(x_i-m_i)(x_j-m_j)\right)\right\},$$

$$A = R^{-1},\ R = \{r_{ij}\}_{i,j=1}^{N},\ r_{ij} = \mathbf{E}X_{t_i}X_{t_j},\ m_i = \mathbf{E}X_{t_i},\ i,j = \overline{1,N},\quad |A| = \mathbf{det}A:$$

Գաուսյան պատահական ընթացքը մեծ թվով փոքր գործոնների ազդեցության տակ գտնվող բնության բազմաթիվ երևույթների մոդելն է: Այդ ընթացքը կիրառական կարևորություն ունի նաև այն պատճառով, որ նրա բոլոր վիճակագրական հատկությունները որոշվում են սպասելիով և հարաբերակցության ֆունկցիայով:

Ստացիոնար (լայն իմաստով) պատահական ընթացքներ: X_t, $t \in \mathcal{T}$, պատահական ընթացքը կոչվում է ստացիոնար (լայն իմաստով), եթե $m(t) = m = \text{const}$, իսկ $\mathbf{B}(t,s) = \mathbf{B}(s-t)$: Եթե նեղ իմաստով ստացիոնար ընթացքի համար գոյություն ունեն $m(t)$-ն և $\mathbf{B}(t,s)$-ը, ապա այն ստացիոնար է նաև լայն իմաստով, հակառակ պնդումը, ընդհանրապես ասած, ճիշտ չէ:

Քանի որ գաուսյան ընթացքների համար վերջավորչափանի բաշխումները լրիվին որոշվում են $m(t)$ և $\mathbf{B}(t,s)$ ֆունկցիաների միջոցով, ապա այդ ընթացքների համար նեղ և լայն իմաստով ստացիոնարության գաղափարները համարժեք են:

Անկախ աճերով պատահական ընթացքների կարևորագույն օրինակներից են վիներյան և պուասոնյան ընթացքները:

Վիներյան պատահական ընթացք (բրոունյան շարժման մոդել): X_t, $t \geq 0$ պատահական ընթացքը կոչվում է վիներյան, եթե

1) $X_0 = 0$,

2) X_t, $t \geq 0$ ընթացքը անկախ աճերով պատահական ընթացք է,

† Այստեղ, իհարկե, h-ը ընտրվում է այնպես, որ $t_1+h, t_2+h, \ldots, t_N+h \in \mathcal{T}$:

3) $X_t - X_s$, $t > s$, աճերը ունեն նորմալ բաշխումներ $\mathbf{E}(X_t - X_s) = 0$ և $\mathbf{D}(X_t - X_s) = \sigma^2(t-s)$, $\sigma^2 > 0$ պարամետրերով՝

$$\mathbf{P}(X_t - X_s < x) = \frac{1}{\sigma\sqrt{2\pi(t-s)}} \int_{-\infty}^{x} e^{-\frac{u^2}{2\sigma^2(t-s)}} du :$$

Օգտվելով աճերի անկախությունից, գրենք վիներյան պրոցեսի վերջավորչափանի խտությունները՝

$$f_{t_1,t_2,\ldots,t_N}(x_1, x_2, \ldots, x_N) = \frac{1}{\sigma\sqrt{2\pi t_1}} e^{-\frac{x_1^2}{2\sigma^2 t_1}} \times \frac{1}{\sigma\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(x_2 - x_1)^2}{2\sigma^2(t_2 - t_1)}} \times$$

$$\times \frac{1}{\sigma\sqrt{2\pi(t_3 - t_2)}} e^{-\frac{(x_3 - x_2)^2}{2\sigma^2(t_3 - t_2)}} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi(t_N - t_{N-1})}} e^{-\frac{(x_N - x_{N-1})^2}{2\sigma^2(t_N - t_{N-1})}} :$$

Վիներյան պատահական ընթացքը նկարագրում է հեղուկի մեջ տեղավորված փոքր մասնիկների շարժումը: Մասնիկների քաոսային անընդհատ շարժումը բնութագրող պատահական ընթացքն առաջին անգամ 1827 թ. դիտել է անգլիացի գիտնական Բրոուն, և այդ պատճառով այն նաև կոչվում է բրոունյան շարժում:

X_t, $t \geq 0$, **պատահական ընթացքը** կոչվում է **պուասոնյան** λ պարամետրով ($\lambda > 0$), եթե.

1) $X_0 = 0$,

2) X_t, $t \geq 0$, ընթացքը անկախ աճերով պատահական ընթացք է,

3) կամայական $s < t$ համար $X_t - X_s$ ունի պուասոնյան բաշխում $\lambda(t-s)$ պարամետրով՝

$$\mathbf{P}\{X_t - X_s = k\} = \frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}, \quad k = 0, 1, 2, \ldots :$$

Օգտվելով աճերի անկախությունից, կստանանք վերջավորչափանի բաշխումները՝

$$\mathbf{P}\{X_{t_1} = k_1, X_{t_2} = k_2, \ldots, X_{t_N} = k_N\} =$$

$$= \frac{(\lambda t_1)^{k_1}}{k_1} e^{-\lambda t_1} \frac{(\lambda(t_2 - t_1))^{k_2 - k_1}}{(k_2 - k_1)!} e^{-\lambda(t_2 - t_1)} \ldots \frac{(\lambda(t_N - t_{N-1}))^{k_N - k_{N-1}}}{(k_N - k_{N-1})!} e^{-\lambda(t_N - t_{N-1})} :$$

Պուասոնյան պատահական ընթացքը հազվադեպ պատահույթները նկարագրող մաթեմատիկական մոդել է: Այս մոդելով են նկարագրվում, օրինակ, տվյալ բնակավայրից ստացված հեռախոսականչերի քանակը, քաղաքի նշված տեղում վթարների քանակը և այլն:

5.4. Մարկովի շղթաներ

Այս ենթաբաժնում ուսումնասիրվող պատահական ընթացքները, որոնք կոչվում են մարկովյան, լայնորեն կիրառվում են գիտության և տեխնիկայի բազմաթիվ բնագավառներում. տնտեսագիտության մեջ, ֆիզիկայում, սոցիոլոգիայում, կենսաբանությունում, ռադիոտեխնիկայում և այլն:

Կդիտարկվեն միայն ընդհատ ժամանակով մարկովյան ընթացքները (Մարկովի շղթաներ), որոնց քաղաքրիչների հնարավոր արժեքների (վիճակների) $\mathcal{X}$ տարածությունը վերջավոր է կամ հաշվելի: Պարզության համար $\mathcal{X}$ բազմության տարրերը հաճախ նշանակում են բնական թվերի միջոցով:

Մենք գիտենք, որ կամայական ընդհատ պարամետրով և հաշվելի $\mathcal{X}$ տարածություն ունեցող X_n, $n = 0, 1, 2, \ldots$, պատահական ընթացքի համար որոշված է նրա վերջավորչափանի բաշխումների

$$\{\mathbf{P}\{X_{t_1} = k_{t_1}, X_{t_2} = k_{t_2}, \ldots, X_{t_N} = k_{t_N}\}\ t_n = 0, 1, \ldots,\ n = \overline{1, N},\ N = 1, 2, \ldots,\}$$

ընտանիքը: Յուրաքանչյուր վերջավորչափանի համատեղ հավանականությունը կարելի է ներկայացնել պայմանական հավանականությունների արտադրյալով՝

$$\mathbf{P}\{X_0 = k_0, X_1 = k_1, \ldots, X_N = k_N\} =$$
$$= \mathbf{P}\{X_0 = k_0\} \cdot \mathbf{P}\{X_1 = k_1 | X_0 = k_0\} \cdot \mathbf{P}\{X_2 = k_2 | X_0 = k_0, X_1 = k_1\} \cdots$$
$$\cdots \mathbf{P}\{X_N = k_N | X_0 = k_0, \ldots, X_{N-1} = k_{N-1}\},$$

որտեղ՝

$$\mathbf{P}\{X_n = k_n | X_0 = k_0, \ldots, X_{n-1} = k_{n-1}\} = \frac{\mathbf{P}\{X_0 = k_0, \ldots, X_n = k_n\}}{\mathbf{P}\{X_0 = k_0, \ldots, X_{n-1} = k_{n-1}\}},\ n = \overline{1, N}:$$

Մարկովի շղթա կանվանենք այնպիսի X_n, $n = 0, 1, 2, \ldots$, պատահական մեծությունների հաջորդականությունը, որը բավարարում է հետևյալ հատկությանը՝

$$\mathbf{P}\{X_n = k | X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i\} =$$
$$= P\{X_n = k | X_{n-1} = i\},\ i, k = 1, 2, \ldots,\ n = 1, 2, \ldots:$$

Այս **հատկությունը**, որը կոչվում է **մարկովյան**, կարելի է հակիրճ բնութագրել հետևյալ կերպ. հայտնի ներկայի պայմանում շղթայի ապագան կախված չէ նրա անցյալից:

$p_{ik}^{(n)} = \mathbf{P}\{X_n = k / X_{n-1} = i\}$ հավանականությունները կոչվում են i վիճակից k վիճակին մեկ քայլով **անցման հավանականություններ**:

$\mathrm{P} = \{p_{ik}^{(n)}\}$ մատրիցը, որը կոչվում է **անցման հավանականությունների մատրից**, ունի հետևյալ հատկությունները՝

$$p_{ik}^{(n)} \geq 0,\ \sum_{k=1}^{\infty} p_{ik}^{(n)} = 1,\quad i, k = 1, 2, \ldots:$$

Այս հատկություններով օժտված մատրիցները կոչվում են **ստոխաստիկ**:

Այսուհետև մենք կդիտարկենք միայն **Մարկովի համասեռ շղթաները**, այսինքն՝ այնպիսի շղթաները, որոնց անցման հավանականությունները կախված չեն n-ից՝

$$p_{ik}^{(n)} = p_{ik},\ i, k = 1, 2, \ldots:$$

Մարկովի շղթաների վերջավորչափանի բաշխումները անցման հավանականությունների միջոցով արտահայտվում են հետևյալ կերպ՝

$$\mathbf{P}\{X_0 = k_0, X_1 = k_1, \ldots, X_N = k_N\} = p_{k_0} \cdot p_{k_0 k_1} \cdot p_{k_1 k_2} \cdots p_{k_{N-1} k_N}: \tag{4}$$

Այստեղ $p_{k_0} = \mathbf{P}\{X_0 = k_0\}$, $k_0 = 1, 2, \ldots$, կոչվում է **Մարկովի շղթայի սկզբնական բաշխում**: Օգտվելով համաձայնեցված բաշխումների մասին Կոլմոգորովի թեորեմից, կարելի է ցույց տալ, որ կամայական ստոխաստիկ մատրիցի և կամայական սկզբնական բաշխման համար կարելի է կառուցել Մարկովի շղթա, որի վերջավորչափանի բաշխումները ունեն (4) տեսքը:

Այժմ բերենք Մարկովի շղթաների որոշ պարզ օրինակներ:

Օրինակ 6: Դիցուք X_n, $n = 0, 1, 2, \ldots$, անկախ, միանման բաշխված պատահական մեծությունների հաջորդականություն է, որոնք ընդունում են վերջավոր թվով արժեքներ հետևյալ հավանականություններով՝

$$\mathbf{P}\{X_n = k\} = p_k,\ k = \overline{0, K}:$$

Այս հաջորդականությունը Մարկովի շղթա է, քանի որ

$$\mathbf{P}\{X_n = k | X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i\} =$$

$$= P\{X_n = k|X_{n-1} = i\} = P\{X_n = k\} = p_k, \; k = \overline{0, K} :$$

Անցման հավանականությունների մատրիցը կազմված է միևնույն տողերից՝

$$\{p_0, p_1, \ldots, p_K\}:$$

<u>*Օրինակ*</u> 7: Դիցուք Y_n, $n = 1, 2, \ldots$, այն ընդհատ, անկախ, միանման բաշխված այն ընդհատ պատահական մեծությունների հաջորդականությունն է, որոնք ընդունում են հաշվելի թվով արժեքներ՝ հետևյալ հավանականություններով՝

$$\mathbf{P}\{Y_n = k\} = p_k, \quad k = 0, 1, 2, \ldots :$$

Կառուցենք X_n, $n = 0, 1, 2, \ldots$, հաջորդականությունը հետևյալ կերպ՝

$$X_0 = 0, \quad X_1 = Y_1, \quad X_2 = Y_1 + Y_2, \quad \ldots, \quad X_n = Y_1 + Y_2 + \ldots + Y_n, \; \ldots :$$

Ցույց տանք, որ այսպես կառուցված ընթացքը Մարկովի շղթա է, և գտնենք նրա անցման հավանականությունների մատրիցը:

Լուծում: Իսկապես՝

$$\mathbf{P}\{X_n = k|X_1 = i_1, X_2 = i_2, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i\} =$$
$$= \mathbf{P}\{Y_1 + Y_2 + \cdots + Y_n = k|Y_1 = i_1, Y_1 + Y_2 = i_2, \ldots, Y_1 + Y_2 + \cdots + Y_{n-1} = i\} =$$
$$= \mathbf{P}\{i + Y_n = k|Y_1 = i_1, Y_1 + Y_2 = i_2, \ldots, Y_1 + Y_2 + \cdots + Y_{n-1} = i\} =$$
$$= \mathbf{P}\{i + Y_n = k|Y_1 + Y_2 + \ldots + Y_{n-1} = i\} = \mathbf{P}\{Y_n = k - i|Y_1 + Y_2 + \ldots + Y_{n-1} = i\} =$$
$$= \mathbf{P}\{Y_n = k - i\} = \begin{cases} p_{k-i}, & \text{եթե } k \geq i, \\ 0, & \text{եթե } k < i : \end{cases}$$

Հետևաբար, X_n, $n = 0, 1, \ldots$ Մարկովի շղթայի անցման հավանականությունների մատրիցը կունենա հետևյալ տեսքը՝

$$\mathbf{P} = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_k & \cdots \\ 0 & p_0 & p_1 & \cdots & p_{k-1} & \cdots \\ 0 & 0 & p_0 & \cdots & p_{k-2} & \cdots \\ . & . & . & \cdots & . & \cdots \end{pmatrix} :$$

Դիտարկենք կարևոր կիրառություններ ունեցող մարկովյան շղթաների կառուցման մի եղանակ:

<u>*Օրինակ*</u> 8: Դիցուք Y_n, $n = 1, 2, \ldots$, անկախ պատահական մեծությունների հաջորդականություն է և $f(u, v)$-ն երկու փոփոխականի իրական ֆունկցիա է: Ցույց տանք, որ $X_n = f(X_{n-1}, Y_n)$, $n = 1, 2, \ldots$, $X_0 = 0$ ռեկուրենտ բանաձևով որոշվող պատահական մեծությունների հաջորդականությունը Մարկովի շղթա է:

Լուծում: Իրոք,

$$\mathbf{P}\{X_n = i_n|X_1 = i_1, X_2 = i_2, \ldots, X_{n-1} = i_{n-1}\} =$$
$$= \mathbf{P}\{f(i_{n-1}, Y_n) = i_n|f(0, Y_1) = i_1, f(i_1, Y_2) = i_2, \ldots, X_{n-1} = i_{n-1}\} =$$
$$= \mathbf{P}\{X_n = i_n|X_{n-1} = i_{n-1}\} = \mathbf{P}\{f(i_{n-1}, Y_n) = i_n|f(i_{n-2}, Y_{n-1}) = i_{n-1}\} =$$
$$= \mathbf{P}\{f(i_{n-1}, Y_n) = i_n\} :$$

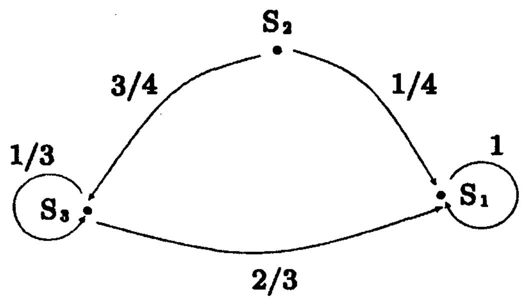
Մարկովի շղթան նկարագրելու համար կարելի է օգտագործել կողմնորոշված գրաֆ, որի գագաթներն են շղթայի վիճակները, իսկ S_i գագաթից S_k գագաթը տանող սլաքը $S_i \xrightarrow{p_{ik}} S_k$ և սլաքի վերևում $p_{ik} > 0$ թիվը հավասար է S_i վիճակից S_k վիճակին անցնելու հավանականությանը: Երբ $p_{ik} = 0$, սլաք չի դրվում:

<u>*Օրինակ*</u> 9: Դիցուք մարկովյան շղթայի հնարավոր վիճակներն են S_1, S_2, S_3, իսկ անցման հավանականությունների P մատրիցն է՝

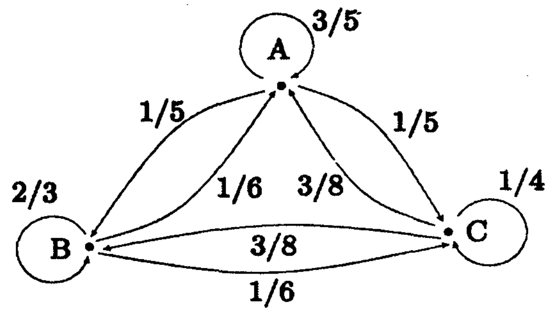
$$\mathrm{P}=\begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 0 & 3/4 \\ 2/3 & 0 & 1/3 \end{pmatrix}:$$

Ինչպիսի՞ կողմնորոշված գրաֆ է նկարագրում մարկովյան շղթան:

Լուծում: Այս շղթային համապատասխանում է նկար 5-ում պատկերված կողմնորոշված գրաֆը:



Նկար 5:



Նկար 6:

Մարկովյան շղթայի S_i վիճակից S_k վիճակին n քայլով անցման հավանականությունները նշանակենք

$$p_{ik}(n)=\mathrm{P}\{X_n=k|X_0=i\},$$

իսկ P_n-ով՝ $p_{ik}(n)$ տարրերով կազմված ստոխաստիկ մատրիցը: Եթե $n=1$, ապա $\mathrm{P}_1=\mathrm{P}$:

Լրիվ հավանականության բանաձևի համաձայն՝ կամայական ոչ բացասական n-ի և m-ի համար ճիշտ է Կոլմոգորով-Չեպմենի հավասարությունը՝

$$p_{ij}(n)=\sum_{r=1}^{K}p_{ir}(m)p_{rj}(n-m):$$

Այս առնչությունը կարելի է ներկայացնել մատրիցային տեսքով՝

$$\mathrm{P}_n=\mathrm{P}_m\times\mathrm{P}_{n-m}:$$

Այստեղից՝

$$\mathrm{P}_2=\mathrm{P}^2,\quad \mathrm{P}_3=\mathrm{P}^3,\ldots,\mathrm{P}_n=\mathrm{P}^n:$$

Մարկովյան շղթաների տեսության մեջ կարևոր դեր ունի վիճակների դասակարգումը: Ասում են, որ S_k վիճակը **հասանելի** է S_i վիճակից, եթե գոյություն ունի այնպիսի $n>0$, որ $p_{ik}(n)>0$: S_i և S_k վիճակները կոչվում են **հաղորդակցվող**, եթե նրանք փոխադարձաբար հասանելի են: S_i վիճակը կոչվում է **ոչ էական**, եթե գոյություն ունի այնպիսի S_k վիճակ, որ S_k-ն հասանելի է S_i-ից, իսկ S_i վիճակը հասանելի չէ S_k վիճակից: Հակառակ դեպքում S_i վիճակը կոչվում է **էական**:

Մարկովյան շղթայի բոլոր էական վիճակների բազմությունը տրոհվում է հաղորդակցվող վիճակներ պարունակող չհատվող դասերի այնպես, որ նույն դասի կամայական երկու վիճակ հաղորդակցվող են, իսկ եթե S_i և S_k վիճակները տարբեր դասերից են, ապա կամայական n-ի համար $p_{ik}(n)=p_{ki}(n)=0$:

Մարկովի շղթան կոչվում է չտրոհվող, եթե նրա բոլոր վիճակները կազմում են հաղորդակցվող վիճակների մեկ դաս:

S_k վիճակը կոչվում է **կլանող**, եթե շղթան, հասնելով այդ վիճակին, նրանից այլևս դուրս չի գալիս, այսինքն՝ $p_{kk}=1$:

Այժմ դիտարկենք մի քանի կարևոր գործնական իրավիճակներ, որոնք կարելի է նկարագրել մարկովյան շղթաների միջոցով: Դա հնարավորություն է տալիս համապատասխան վերլուծության միջոցով հանգել այս իրավիճակներում ծագող մի շարք հարցերի պատասխաններին:

<u>*Օրինակ*</u> 10: *Պահուստների մասին խնդիրը*: Կազմակերպությունը պահանջարկը անընդհատ բավարարելու նպատակով պահեստավորում է ինչ-որ ապրանք: Պահուստի համալրումը կատարվում է ժամանակի $t_0, t_1, t_2, \ldots$ պահերին: (t_{n-1}, t_n) ժամանակահատվածում ստացված ապրանքի պահանջարկների գումարը պատահական մեծություն է՝ Y_n, որի բաշխման օրենքն է՝

$$\mathbf{P}\{Y_n = k\} = p_k, \quad k = 1,2,\ldots, \quad p_k \geq 0, \quad \sum_{k=1}^{\infty} p_k = 1:$$

Ենթադրվում է, որ Y_n, $n = 1,2,\ldots$, պատահական մեծություններն անկախ են: Յուրաքանչյուր ժամանակահատվածի սկզբում պահուստի մակարդակը հաստատուն է: Պահուստը համալրվում է հետևյալ կերպ. եթե առկա ապրանքի քանակը նախօրոք նշված կրիտիկական S_0 մակարդակից ցածր է, ապա կատարվում է անհապաղ համալրում՝ մինչև պահանջվող $S > S_0$ մակարդակը: Եթե առկա ապրանքի քանակը գերազանցում է S_0 կրիտիկական մակարդակը, ապա համալրում չի կատարվում: Նշանակենք X_n-ով մինչ t_n պահը առկա պահուստի մակարդակը: X_n, $n = 0,1,2,\ldots$, ընթացքի վիճակների տարածությունը բաղկացած է պահուստի հնարավոր մակարդակների հետևյալ արժեքներից՝ S, $S-1$, $\ldots$, 1, 0, -1, $-2,\ldots$, որտեղ բացասական արժեքները համապատասխանում են չբավարարված պահանջարկի դեպքերին (այդ պատվերները պետք է անհապաղ կատարվեն պահուստի համալրումից հետո): Ցույց տանք, որ X_n, $n = 0,1,2,\ldots$, պատահական ընթացքը մարկովյան շղթա է, և գտնենք անցման հավանականությունների մատրիցը:

Լուծում: Կարող ենք գրել հետևյալ ռեկուրենտ արտահայտությունը՝

$$X_{n+1} = \begin{cases} X_n - Y_{n+1}, & \text{եթե} \quad S_0 < X_n \leq S, \\ S - Y_{n+1}, & \text{եթե} \quad X_n \leq S_0: \end{cases}$$

Հետևաբար, համաձայն օրինակ 8-ի դիտարկումների, X_n, $n = 0,1,2,\ldots$, ընթացքը մարկովյան շղթա է հետևյալ անցման հավանականություններով՝

եթե $k > S$, ապա $p_{ik} = 0$, եթե $k \leq S$, իսկ $S_0 < i < S$, ապա $p_{ik} = \mathbf{P}\{Y_n = i - k\}$,

եթե $k \leq S$, իսկ $i \leq S_0$, ապա $p_{ik} = \mathbf{P}\{Y_n = S - k\}$:

<u>*Օրինակ*</u> 11: Ա քաղաքի յուրաքանչյուր բնակիչ ունի A, B, C երեք արհեստներից մեկը: A, B, C արհեստներ ունեցող հայրերի երեխաները պահպանում են հայրերի արհեստները, համապատասխանաբար, 3/5, 2/3, 1/4 հավանականությամբ, իսկ եթե չեն պահպանում, ապա հավասար հավանականությամբ ընտրում են մյուս երկու արհեստներից որևէ մեկը: Գտնել հաջորդ սերնդի ընտրած արհեստների հավանականությունների բաշխումը, եթե տվյալ սերնդի համար A արհեստն ունեն բնակիչների 20%-ը, B արհեստը՝ 30%-ը, C արհեստը՝ 50%-ը:

Լուծում: Մարկովյան շղթան ունի A, B, C երեք վիճակ: Հայտնի է սկզբնական բաշխումը՝

$$(p_1^0, p_2^0, p_3^0) = \left(\frac{1}{5}, \frac{3}{10}, \frac{1}{2}\right),$$

և մեկ քայլով անցման հավանականությունների մատրիցը՝

$$\mathbf{P} = \begin{pmatrix} 3/5 & 1/5 & 1/5 \\ 1/6 & 2/3 & 1/6 \\ 3/8 & 3/8 & 1/4 \end{pmatrix}:$$

P մատրիցին համապատասխանում է նկար 6-ում պատկերված կողմնորոշված գրաֆը:
Որոնելի բաշխումը՝ (p_1^1, p_2^1, p_3^1), կգտնենք, օգտվելով լրիվ հավանականության բանաձևից՝

$$p_1^1 = p_1^0 \cdot p_{11} + p_2^0 \cdot p_{21} + p_3^0 \cdot p_{31} = \frac{1}{5}\cdot\frac{3}{5} + \frac{3}{10}\cdot\frac{1}{6} + \frac{1}{2}\cdot\frac{3}{8} = \frac{143}{400},$$

$$p_2^1 = p_1^0 \cdot p_{12} + p_2^0 \cdot p_{22} + p_3^0 \cdot p_{32} = \frac{1}{5}\cdot\frac{1}{5} + \frac{3}{10}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{3}{8} = \frac{171}{400},$$

$$p_3^1 = p_1^0 \cdot p_{13} + p_2^0 \cdot p_{23} + p_3^0 \cdot p_{33} = \frac{1}{5}\cdot\frac{1}{5} + \frac{3}{10}\cdot\frac{1}{6} + \frac{1}{2}\cdot\frac{1}{4} = \frac{86}{400}:$$

Հետևաբար,

$$(p_1^1, p_2^1, p_3^1) = \left(\frac{143}{400}, \frac{171}{400}, \frac{86}{400}\right):$$

<u>Օրինակ</u> 12: *Քոլեջում ուսանողի ուսման ժամկետի մոդելը*: Ենթադրենք, մի ուսանող յուրաքանչյուր տարվա վերջում p հավանականությամբ դուրս է մնում քոլեջից, q հավանականությամբ մնում է նույն կուրսում և r հավանականությամբ փոխադրվում է հաջորդ կուրս: Պարզ է, որ քոլեջում ուսանողի վիճակի փոփոխման ընթացքը մարկովյան շղթա է: Կառուցենք այս մարկովյան շղթայի անցման հավանականությունների մատրիցը:

Լուծում: Դիտարկենք հետևյալ վիճակները՝ S_1-ը՝ ուսանողը դուրս է մնացել, S_2-ը՝ ավարտել է քոլեջը, S_3-ը՝ 4-րդ կուրսեցի է, S_4-ը՝ 3-րդ կուրսեցի է, S_5-ը՝ 2-րդ կուրսեցի է, S_6-ը՝ առաջին կուրսեցի է: Նշենք, որ S_1-ը և S_2-ը կլանող վիճակներ են:

Անցման հավանականությունների P մատրիցը կունենա հետևյալ տեսքը՝

$$\mathrm{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ p & r & q & 0 & 0 & 0 \\ p & 0 & r & q & 0 & 0 \\ p & 0 & 0 & r & q & 0 \\ p & 0 & 0 & 0 & r & q \end{pmatrix}:$$

Մարկովյան շղթաների տեսությունը հնարավորություն է տալիս պատասխանել կարևոր կիրառական նշանակություն ունեցող հետևյալ հարցերին:

1. Եթե առկա պահին շղթան գտնվում է S_i վիճակում, ապա ինչպիսի՞ն է n քայլով S_k, $k = 1, 2, \ldots$ վիճակին հասնելու հավանականությունների բաշխումը: Ինչպե՞ս է կախված այդ բաշխումը S_i վիճակից:

2. Կարելի՞ է արդյոք կանխագուշակել այն միջին ժամանակը, որը շղթան կանցկացնի տվյալ վիճակում:

Այս հարցերին սպառիչ պատասխան կարելի է տալ էրգոդիկ և կլանող մարկովյան շղթաների դեպքում:

Վերջավոր թվով վիճակներ ունեցող մարկովյան շղթան կոչվում է **էրգոդիկ**, եթե կամայական i-ի և k-ի համար՝

$$\lim_{n\to\infty} p_{ik}(n) = p_k > 0,$$

$$\sum_{k=1}^{K} p_k = 1, \quad i = \overline{1, K}:$$

p_k, $k = \overline{1, K}$, թվերը կոչվում են **Մարկովի շղթայի սահմանային հավանականություններ**: Եթե վերջավոր թվով վիճակներ ունեցող մարկովյան շղթայի համար գոյություն ունի այնպիսի n_0, որ P_{n_0} մատրիցի բոլոր տարրերը դրական են, ապա մարկովյան շղթան էրգոդիկ է:

Սահմանային հավանականությունները հետևյալ համակարգի լուծումն են՝

$$p_k = \sum_{i=1}^{K} p_i \cdot p_{ik}(n), \ k = \overline{1,K}, \ \sum_{k=1}^{K} p_k = 1 :$$

Դիցուք շղթան գտնվում է S_i վիճակում: Պատահական մեծությունը, որը հավասար է S_i վիճակից շղթայի S_k վիճակին հասնելու ժամանակի պահերի ընդհանուր քանակին, նշանակենք τ_{ik}-ով: Երգոդիկ շղթայի S_k վիճակում անցկացնելու միջին $\mathrm{E}\tau_{ik}$ ժամանակը մոտավորապես հավասար է np_k՝ $\mathrm{E}\tau_{ik} \sim np_k$, $k = \overline{1,K}$:

Օրինակ 13: Օրինակ 11-ի համար գտնենք սահմանային հավանականությունները որպես հավասարությունների հետևյալ համակարգի լուծում՝

$$\begin{cases} p_1 = p_1 \cdot p_{11} + p_2 \cdot p_{21} + p_3 \cdot p_{31}, \\ p_2 = p_1 \cdot p_{12} + p_2 \cdot p_{22} + p_3 \cdot p_{32}, \\ p_3 = p_1 \cdot p_{13} + p_2 \cdot p_{23} + p_3 \cdot p_{33} \end{cases} \quad \sum_{j=1}^{3} p_j = 1 :$$

Լուծում: Մեր պայմանների դեպքում ունենք՝

$$\begin{cases} p_1 = 3/5 \cdot p_1 + 1/6 \cdot p_2 + 3/8 \cdot p_3, \\ p_2 = 1/5 \cdot p_1 + 2/3 \cdot p_2 + 3/8 \cdot p_3, \\ p_3 = 1/5 \cdot p_1 + 1/6 \cdot p_2 + 1/4 \cdot p_3, \end{cases}$$

երկրորդ հավասարությունից հանելով առաջինը, կստանանք՝ $p_2 - p_1 = -2p_1/5 + p_2/2$, որտեղից $p_2 = 6p_1/5$: Երրորդ հավասարությունից հանելով առաջինը, կստանանք $p_3 - p_1 = -2p_1/5 - p_3/8$, որտեղից $p_3 = 8p_1/15$: Քանի որ $p_1 + p_2 + p_3 = 1$, ունենք $p_1 + 6p_1/5 + 8p_1/15 = 1$: Այստեղից $p_1 = 15/41$, $p_2 = 18/41$, $p_3 = 8/41$: Հետևաբար՝ $(p_1, p_2, p_3) = (15/41, 18/41, 8/41)$:

K վիճակներ ունեցող Մարկովի շղթան կոչվում է **կլանող**, **եթե նրա անցման** հավանականությունների մատրիցն ունի հետևյալ տեսքը՝

$$\begin{array}{cc} & \begin{array}{cc} m & K-m \end{array} \\ \mathrm{P} = & \begin{pmatrix} I_m & . & O \\ . & . & . \\ R & . & Q \end{pmatrix} \begin{array}{c} m \\ \\ K-m \end{array} \end{array}, \ 1 \le m \le K,$$

որտեղ I_m-ն $(m \times m)$-չափանի միավոր մատրից է, O-ն $((K-m) \times m)$-չափանի՝ զրոներից կազմված մատրից է, R-ը՝ $(m \times (K-m))$-չափանի է, իսկ Q-ն $((K-m) \times (K-m))$-չափանի մատրից է:

Նշանակենք b_{ik}-ով շղթայի S_i վիճակից S_k կլանող վիճակին հասնելու հավանականությունը, իսկ B-ով՝ համապատասխան մատրիցը՝ $\mathrm{B} = \{b_{ik}\}$: Տեղի ունի հետևյալ թեորեմը:

> Մարկովի կլանող շղթայի համար $\mathrm{E}\tau_{ik} = T_{ik}$, որտեղ T_{ik}-երը $\mathrm{T} = (I_{K-m} - Q)^{-1}$ մատրիցի տարրերն են,
>
> $$\mathrm{E}\left(\sum_{k=m+1}^{K} \tau_{ik}\right) = \alpha_i,$$
>
> որտեղ α_i-ն T մատրիցի առաջին սյան տարրերն են, իսկ $\mathrm{B} = \mathrm{TR}$:

Օրինակ 14: Կիրառենք այս թեորեմը օրինակ 12-ում դիտարկված մոդելի նկատմամբ: Դիցուք $p = 0.2$, $r = 0.7$ և $q = 0.1$: Գտնենք P, T և B մատրիցները: B մատրիցի տարրերը ուսանողի՝ ուսումը դադարեցնելու կամ ավարտելու հավանականություններն են:

Լուծում:

$$P = \begin{pmatrix} I_2 & . & O \\ . & . & . \\ R & . & Q \end{pmatrix} = \begin{pmatrix} 1 & 0 & . & 0 & 0 & 0 & 0 \\ 0 & 1 & . & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . \\ 0.2 & 0.7 & . & 0.1 & 0 & 0 & 0 \\ 0.2 & 0 & . & 0.7 & 0.1 & 0 & 0 \\ 0.2 & 0 & . & 0 & 0.7 & 0.1 & 0 \\ 0.2 & 0 & . & 0 & 0 & 0.7 & 0.1 \end{pmatrix}:$$

Դժվար չէ համոզվել, որ $T = (I_4 - Q)^{-1}$ և

$$T = \begin{pmatrix} 1.11 & 0 & 0 & 0 \\ 0.86 & 1.11 & 0 & 0 \\ 0.67 & 0.86 & 1.11 & 0 \\ 0.52 & 0.67 & 0.86 & 1.11 \end{pmatrix}:$$

Այստեղից հասկանալի է, որ $\alpha_1 = 1.11$, $\alpha_2 = 0.86$, $\alpha_3 = 0.67$, $\alpha_4 = 0.52$, և

$$B = \begin{array}{c} \text{դադարեցնել ուսումը} \quad \text{ավարտել ուսումը} \\ \begin{pmatrix} 0.22 & 0.78 \\ 0.40 & 0.60 \\ 0.53 & 0.47 \\ 0.63 & 0.37 \end{pmatrix} \end{array} \begin{array}{l} \\ \text{4-րդ կուրսեցի} \\ \text{3-րդ կուրսեցի} \\ \text{2-րդ կուրսեցի} \\ \text{առաջին կուրսեցի:} \end{array}$$

Այստեղից կարելի է ստանալ, օրինակ, հետևյալ եզրակացությունը՝ ուսանողը պետք է փոխադրվի 3-րդ կուրս, որպեսզի նրա քոլեջ ավարտելու հավանականությունը լինի 1/2-ից ոչ պակաս:

Բաժին Բ
Կիրառական և մաթեմատիկական վիճակագրության հիմունքներ

Գլուխ 6
Նմուշահանում: Նկարագրական վիճակագրություն

Մեծ սխալ է դատողությունների մեջ ընկնել՝ փաստերը ձեռք բերելուց առաջ:

Արթուր Կոնան Դոյլ

Մի նկարը տասը հազար բառ արժե:

Ֆրեդերիկ Բարնարդ

Զույգ թվերն, իհարկե, ավելի լավն են, բայց կենտերն ունեն կենտրոն:

Վահան (5 տարեկան)

6.1. Նմուշահանման հիմնական գաղափարները և սկզբունքները

«Վիճակագրություն» բառը հաճախ գործածում են, երբ որևէ երևույթ ուսումնասիրելու համար հավաքում են ստվարաթիվ տվյալներ: Իսկ որպես գիտական առարկա՝ **կիրառական վիճակագրությունը** այնպիսի եղանակների, ընթացակարգերի ու հնարքների համակարգ է, որոնք հնարավորություն են տալիս փորձնական տվյալները հիմնավորված ձևով հավաքել, կազմավորել, ի մի բերել, ներկայացնել և վերլուծել՝ դրանց հիման վրա եզրահանգումներ անելու և որոշումներ ընդունելու նպատակով: **Մաթեմատիկական վիճակագրության** խնդիրն է հիմնավորել, կատարելագործել և զարգացնել այդ եղանակներն ու ընթացակարգերը, ընդլայնել դրանց ներգործության ոլորտը:

Մաթեմատիկական վիճակագրության եղանակների մեծ մասը հենվում է հավանականության տեսության գաղափարների և արդյունքների վրա: Դա թույլ է տալիս, մասնավորապես, գնահատել վիճակագրական նյութի հիման վրա արված եզրակացությունների հուսալիությունը և ճշգրտությունը:

Գրքի այս Բ բաժինը նվիրված է մաթեմատիկական վիճակագրության հիմունքների շարադրմանը, նկատի ունենալով տնտեսագիտական կիրառությունները:

Վիճակագրական հետազոտության սկզբնական փուլի եղանակները կազմում են այսպես կոչված **նկարագրական վիճակագրությունը**: Դրանք ուղղված են հավաքված տվյալները լավագույնս (ակնառու, մատչելի) ներկայացնելու նպատակին՝ աղյուսակների, գծագրերի, ամփոփիչ բնութագրիչների միջոցով:

Ծանոթանանք այդ փուլի հիմնական գաղափարներին և համապատասխան տերմիններին†: Հիշեցնենք, որ վիճակագրությունը ծագել է ժողովրդագրական խնդիրներից,

†Ցավոք, հայերենում վիճակագրական տերմինները դեռ հղկման կարիք ունեն, երբեմն օգտագործվում են բավարար չափով չհիմնավորվածները: Գրքի հավելվածում տրված է հայերեն այն նոր կամ ճշտված տերմինների ցանկը, որոնք ընտրվել են դասավանդման բազմամյա փորձի, մեծ թվով մասնագիտական և թարգմանական բառարանների տեղեկությունների համադրման և բազմաթիվ քննարկումների հիման վրա: Ցանկում տրվում են նաև համապատասխան անգլերեն և ռուսերեն համարժեքները:

և, հետևաբար, հիմնական տերմինները փոխ են առնվել հենց ժողովրդագրությունից: Մենք դա հաշվի ենք առել, միաժամանակ նկատի ունենալով նաև հարյուրամյակների ընթացքում մշակված միջազգային տերմինները:

Սկսենք որոշակի ուսումնասիրման ենթակա բոլոր առարկաների բազմությունից, որի անվանման համար եվրոպական լեզուներում օգտագործում են «բնակչություն» բառի համապատասխան թարգմանությունները (անգլերեն՝ population): Մենք կօգտագործենք հանույթ տերմինը:

> Տվյալ վիճակագրական խնդրում ուսումնասիրման ենթակա, կամ մտովի հնարավոր, բոլոր առարկաների բազմությունը կանվանենք **հանույթ**[†]: Հանույթի տարրը կոչվում է **անհատ**:

Անհատների դիտարկվող հատկությունները կոչվում են **հատկանիշներ**: Դրանք կարող են լինել **որակական** կամ **քանակական**: Քանակական հատկանիշները լինում են **ընդհատ** կամ **անընդհատ**: Հանույթի հատկանիշները կնշանակենք $X, Y, \ldots$ մեծատառերով, իսկ դրանց արժեքները որոշակի անհատների դեպքում, համապատասխանաբար, $x, y, \ldots$ փոքրատառերով:

> Հանույթի այն ենթաբազմությունը, որը վերցվում է քննության համար, կոչվում է **նմուշ**[‡], իսկ նմուշի ստացման գործընթացը՝ **նմուշահանում**: Եթե ուսումնասիրվում են հանույթի բոլոր անհատները, ապա նմուշահանումը կոչվում է **հաշվեհամար** (ժողովրդագրությունում՝ **մարդահամար**):

Նմուշի մեջ ընդգրկված անհատների թիվը կոչում են **նմուշի ծավալ** և նշանակում են N: Նմուշի անհատների X հատկանիշի արժեքների գրանցումը կազմում է

$$\mathbf{x} = (x_1, x_2, \ldots, x_N)$$

վեկտորը, որը հարմար է նույնպես անվանել նմուշ, քանի որ հետագա գործողությունները և եզրակացությունները հիմնվում են հենց $\mathbf{x}$-ի վրա:

Իհարկե, նմուշահանման, առավել ևս հաշվեհամարի ժամանակ, կարող են միաժամանակ գրանցվել մի քանի հատկանիշներ՝ $X_1, X_2, \ldots, X_L$: Այդ դեպքում առաջին դիտարկված անհատի համար գրանցվում է $(x_{11}, \ldots, x_{1L})$ վեկտորը, իսկ ամբողջ նմուշը ներկայացվում է մատրիցով: Մի քանի հատկանիշների ուսումնասիրությանը կանդրադառնանք այս գլխի 6.3 և 6.5 ենթաբաժիններում, իսկ այժմ զբաղվենք մեկ հատկանիշի նկատմամբ տվյալների մշակման հարցերով:

Օրինակ 1: Ժողովրդագրական հետազոտությունների նպատակով անցկացվում են հարցումներ: Եթե մանրամասնորեն ուսումնասիրվում է որևէ երկրի ամբողջ բնակչությունը, ապա, ինչպես ասվեց, գործընթացը կոչվում է մարդահամար (բնակչության հաշվեհամար), հակառակ դեպքում կատարվում է **նմուշային հետազոտություն**: Պարզ է,

---

[†]«Հանույթ»-ը առօրյա իմաստով օգտագործվում է որպես ածական, որի հոմանիշներն են՝ ընդհանուր, բոլոր, ամբողջ, ամեն, բովանդակ, համայն, ողջ բառերը: Մենք կկիրառենք այն որպես գոյական:

[‡]Ռուսերենում «выборка» տերմինը արմատավորվել է, քանի որ չի գտնվել ռուսերենում «брать» բայի գործողության արդյունքը նկարագրող այլ բառ: Անհաջող է նաև այն, որ և՛ գործողությունն է անվանվում выборка, և՛ նրա արդյունքը: Հայերենում որոշ հեղինակներ «выборка»-ն թարգմանելիս «ստեղծել» են «վերցվածք» [1], «ընտրանք» [7] բառերը: Վերջինի իմաստը հակասում է նմուշի պատահական ձևով ստանալու գաղափարին: Մենք ելնում ենք այն սկզբունքից, որ «նմուշ» և «նմուշահանում» բառերը հայերենում կան և լիովին համապատասխանում են անգլերենում, ֆրանսերենում ընդունված տերմիններին (տե՛ս Է. Աղայանի «Արդի հայերենի բացատրական բառարան»-ը):

որ մարդահամարը տալիս է ավելի բազմակողմանի և ճշգրիտ տեղեկություններ, սակայն այն շատ մեծ ծախսերի հետ է կապված և չի կարող հաճախակի անցկացվել: Հարկ է լինում դիմել ընտրանքային ուսումնասիրությունների, որոնք վերջին տարիներին ավելի հաճախ են կատարվում: Այդպիսին են, օրինակ, հասարակական կարծիքի հարցումները, որոնք անցկացվում են նախընտրական փուլերում կամ այլ նպատակներով:

Ամեն մի բնակչի (անհատի) համար դիտարկվում են տարբեր **հատկանիշներ՝ որակական** (սեռը, առողջական վիճակը, մասնագիտությունը և այլն) և **քանակական** (տարիքը, հասակը, կշիռը, աշխատավարձը և այլն):

Նկատենք, որ ընդհատ և անընդհատ տվյալների տարբերությունը վերանում է չափման գործիքների սահմանափակ ճշտության պատճառով: Օրինակ, մարդկանց հասակը չափվում է սանտիմետրի ճշտությամբ, և այդ սկզբունքորեն անընդհատ, մեծության դիտումները իրականում գրանցվում են ընդհատ արժեքներով:

Օրինակ 2: Մեծ թվով էլեկտրական լամպերի բազմությունից (համույթը՝ արտադրված և որևէ ժամանակահատվածում արտադրվելիք լամպերն են) վերցնում են մի քանիսը (նմուշ) և չափում են դրանց աշխատանքի տևողությունը (հետազոտվող հատկանիշ): Քանի որ չափումն ավարտվում է լամպը շարքից դուրս գալիս, հաշվեհամարը սկզբունքորեն հնարավոր չէ, եթե բոլոր լամպերը «ստուգվեն», վաճառելու ոչինչ չի մնա:

Այժմ կարող ենք վերաձևակերպել **մաթեմատիկական վիճակագրության հիմնական խնդիրը**. այն կոչված է մշակել եղանակներ, որոնց միջոցով կարելի է նմուշի տվյալների հիման վրա հնարավորին չափ ճշգրիտ դատողություններ անել դիտարկվող համույթի մասին, մասնավորապես, գնահատել դրա մեկ կամ մի քանի հատկանիշների կարևոր բնութագրիչների արժեքները:

Պետք է նշել, որ հաշվեհամարների (մարդահամարների) տվյալները, լինելով շատ մեծ թվով մասնակիցների աշխատանքի արդյունք, նրանցից ոչ բոլորի բավարար որակավորման կամ պարտաճանաչության պատճառով կարող են պարունակել անճշտություններ: Դա ևս մի փաստարկ է հօգուտ նմուշային ուսումնասիրությունների:

Իհարկե, համույթի մասին նմուշային հետազոտության միջոցով ստացվող տեղեկությունները իրենց հերթին միշտ պարունակում են որոշ սխալներ, քանի որ հենվում են անհատների միայն մի մասի տվյալների վրա: Այստեղից ծագում են նմուշահանման երկու փոխկապված հիմնախնդիրներ: Առաջինը՝ ինչպե՞ս կազմակերպել նմուշահանումը, որպեսզի ստացված տեղեկությունները հնարավորին չափ ճիշտ արտացոլեն ամբողջ համույթի հետազոտվող հատկությունները՝ **նմուշի ներկայացուցչականության խնդիրը**, և երկրորդը՝ ինչպե՞ս մշակել նմուշը՝ նրանից համույթի վերաբերյալ առավել հուսալի տեղեկություններ ստանալու համար: Մենք այս գրքում հակիրճորեն կանդրադառնանք առաջին խնդրին, հետագա շարադրանքը հիմնականում նվիրելով երկրորդին:

Տվյալների տարբեր տեսակներին համապատասխան ընտրվում են դրանց մշակման առանձնահատկությունները: Վիճակագրական հետազոտություններում քանակական տվյալները կարող են լինել չորս տեսակի:

ա) **Անվանական տվյալները** առաջ են գալիս, երբ նշվում (համարակալվում) են անհատների դասերը, հատկությունները: Օրինակ՝ ապրանքների տեսակների կոդերը, ուսանողների ֆակուլտետների համարները և այլն: Այս տվյալները չի կարելի մշակել թվայինների պես, քանի որ *չափի իմաստ չեն կրում*:

բ) **Կարգային** տվյալները ներկայացնում են առարկաների ու երևույթների կարգավորումը՝ ըստ կարևորության, ուժգնության կամ այլ որակի, օրինակ՝ քամու կամ երկրաշարժի ուժգնության աստիճանը (բալը): Այդ տվյալների հետ թվաբանական գործողությունները նույնպես անիմաստ են:

գ) **Միջակայքային** տվյալները արտացոլում են կարգավորում, դասակարգում ըստ որևէ քանակական սանդղակի, օրինակ՝ ջերմաստիճան ըստ Ցելսիուսի: Այստեղ որոշ դեպքերում իմաստ ունեն գումարման և հանման գործողությունները, սակայն բազմապատկումը և բաժանումը՝ ոչ: Չի կարելի ասել, օրինակ, որ 60°-ը երկու անգամ ավելի տաք է, քան 30°-ը, ուրեմն 60/30 հարաբերությունը իմաստ չունի:

դ) **Հարաբերական** տվյալները կարող են մասնակցել հիմնական թվաբանական գործողություններում, ներառյալ անչափում թվով բազմապատկումը և մեկը մյուսի վրա բաժանումը: Ֆինանսական տվյալները, օրինակ՝ եկամուտը, շահույթը, գինը և այլն, այս դասից են: Ֆիզիկական չափումները և ժամանակահատվածները նույնպես հարաբերական տվյալներ են:

Նմուշահանման կազմակերպման հարցերը մանրամասնորեն ուսումնասիրվում են, մասնավորապես, «Վիճակագրություն» դասընթացում: Մենք կսահմանափակվենք մի քանի դիտողություններով:

Տարբերում են նմուշահանման երկու եղանակ՝ **պատահական** և **ոչ պատահական**:

> **Պատահական** կոչվում է այնպիսի **նմուշահանումը**, որի ընթացքում հանույթի բոլոր անհատները հավասար հնարավորություն ունեն ընդգրկվելու նմուշի մեջ՝ փոխադարձ անկախությունն ու պատահականությունն ապահովող որևէ գործընթացի միջոցով:

Հավանականության տեսության տեսակետից այդպիսի նմուշահանումը անկախ փորձերի հաջորդականություն է: Պատահական նմուշահանումը կարող է կատարվել պարզագույն վիճակահանության միջոցով: Օգտագործում են նաև **պատահական թվերի** հատուկ աղյուսակներ՝ հանույթի անհատներին համարակալելուց հետո նմուշի կազմի մեջ վերցնում են այն անհատներին, որոնց համարները իրար ետևից գրված են պատահական թվերի աղյուսակի որևէ մասում:

Ներկայումս պատահական նմուշահանումը **հեշտությամբ** կատարվում է **հաշվարկների** (կոմպյուտերների) օգնությամբ:

Կարևորն այն է, որ նմուշահանման ընթացքում թույլ չտրվեն պատահականության սկզբունքի խախտումներ, որոնց հետևանքով՝ կարող են առաջանալ նմուշահանման անճշտություններ: Սակայն նմուշահանման պատահական լինելու ապահովումն իրականում դժվար է: Օրինակ, ընտրանեկան բյուջեների հետազոտության ժամանակ գործնականում հնարավոր չէ համարակալել երկրի կամ մեծ քաղաքի բոլոր ընտանիքները՝ պատահական նմուշահանում կազմակերպելու համար: Նմանապես, դա հնարավոր չէ իրականացնել հասարակական կարծիքի ուսումնասիրման ժամանակ: Այդ պատճառով կիրառում են **ոչ լրիվ պատահական** նմուշահանում, ձգտելով հնարավորին չափ ապահովել «ներկայացուցչությունը», որպեսզի նմուշում մասնակցեն հանույթի տարբեր որակական շերտերի անհատները: Օրինակ, հարցումների ժամանակ տարբեր խավերի, տարիքային խմբերի, ազգությունների, բնակավայրերի և այլն, անհատների մասնակցությունը: Երբեմն ընտրում են հանույթի որոշ ենթաբազմություններ և դրանցից կատարում պատահական նմուշահանում: Նմուշահանման գործընթացն ուսումնասիրող և մշակող հատուկ գիտությունը քննարկում է նաև ոչ լրիվ պատահական ձևով ստացված նմուշի հետազոտության առանձնահատկությունները: Սակայն մենք հետագա շարադրանքում հիմնականում կսահմանափակվենք լրիվ պատահական նմուշահանման դիտարկումով:

Նմուշ է համարվում նաև պատահական ելքերով փորձի N անգամ իրագործման ելքերի հաջորդականությունը: Օրինակ, եթե ուզում ենք ստուգել մետաղադրամի կանոնավոր լինելը (մետաղադրամը այն «հանույթն» է, որի մասին պետք է դատողություն արվի), ապա կարող ենք N անգամ կատարել նետումներ և արդյունքների հիման վրա պատասխանել մեզ հետաքրքրող հարցերին: Եթե բավականաչափ մեծ N-երի դեպքում զինանշանի հաճախությունը զգալիորեն տարբերվում է 1/2-ից, ապա կարող ենք մերժել մետաղադրամի կանոնավոր լինելու վարկածը:

Այստեղ տեղին է զուգահեռներ անցկացնել հավանականության տեսության և վիճակագրության տերմինների միջև: $(\Omega, \mathcal{F}, \mathbf{P})$ հավանականային տարածության կամ դրան համարժեք պատահական մեծության մոդելներում, որոնք նկարագրում են փորձը, *հավանականությունները համարվում են տրված*: Մաթեմատիկական վիճակագրությունում կատարվում է փորձերի շարքից ստացված նմուշի հետազոտումը, որպեսզի հնարավորին չափ ճշգրտորեն որոշվի հանույթը նկարագրող $(\Omega, \mathcal{F}, \mathbf{P})$ հավանականային տարածության *անհայտ* $\mathbf{P}$ *հավանականությունը*, կամ գնահատվեն դրանով որոշվող բնութագրիչները: Եթե հանույթի անհատների՝ մեզ հետաքրքրող հատկանիշը քանակական է, ապա կարող ենք համարել, որ հանույթը նկարագրվում է պատահական մեծությամբ, որի *բաշխման օրենքը հայտնի չէ*: Այդ տեսակետից՝

փորձարկումների իրագործումից առաջ նմուշը դիտարկվում է որպես N անկախ պատահական վեկտոր՝ $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, իսկ *փորձարկումներից հետո* ստացված $\mathbf{x} = (x_1, x_2, \ldots x_N)$ նմուշը $\mathbf{X}$-ի մի իրականացում է:

Պատահական $\mathbf{X}$ նմուշի ամեն մի բաղադրիչի բաշխման օրենքը համընկնում է հանույթի որոշակի հատկանիշը նկարագրող X պատահական մեծության օրենքի հետ: Այդ օրենքը նախօրոք հայտնի չէ, այն դիտարկվում է վիճակագրական եղանակների հետազոտման, հիմնավորման հարցերում և կոչվում է **տեսական**, որպեսզի տարբերվի փորձնական տվյալների հիման վրա կազմված **վիճակագրական** (*էմպիրիկ*) **բաշխման օրենքից**:

$\mathbf{X}$ **պատահական նմուշի տեսական բաշխման ֆունկցիան**՝ $F_{\mathbf{X}}(x_1, x_2, \ldots, x_N)$-ը, որոշվում է X հայտանիշի $F_X(x)$ **տեսական բաշխման ֆունկցիայի** միջոցով՝

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_N) = \mathbf{P}\{X_1 < x_1, \ldots, X_n < x_N\} = \prod_{n=1}^{N} \mathbf{P}\{X_n < x_n\} = \prod_{n=1}^{N} F_X(x_n):$$

Երբեմն ասում են նաև՝ հանույթը ենթարկվում է $F(x)$ բաշխման օրենքին:

Պատահական նմուշահանումը կարող է կատարվել որպես հանույթից **դարձով նմուշահանում**, որի ժամանակ (տե՛ս գլուխ 1, վարժություն 19) նմուշի մեջ ընդգրկված անհատը նորից մասնակցում է հաջորդ պատահական վիճակահանությանը, բոլոր իրար հաջորդող վիճակահանությունները իրարից անկախ են և կատարվում են միևնույն պայմաններում: Այդ դեպքում, ինչպես ասվեց, $\mathbf{X}$ նմուշի բաղադրիչները իրարից անկախ են և միանման բաշխված:

Անդարձ (կամ **սպառիչ**), **նմուշահանման** դեպքում (տե՛ս գլուխ 1, վարժություն 20) նմուշի մեջ ընդգրկված անհատն այլևս չի մասնակցում հաջորդ վիճակահանություններին: Այդպես է էլեկտրական լամպերին վերաբերող օրինակ 2-ում՝ լամպի ուսումնասիրությունը (աշխատունակության տևողության չափումը) կապված է նրա շարքից դուրս գալու հետ: Եթե հանուրն անվերջ է կամ շատ մեծ թվով անհատներ է պարունակում, կարելի է համարել, որ նորից հաջորդական վիճակահանությունները կատարվում են համարյա անփոփոխ պայմաններում և իրարից անկախ: Սակայն, եթե հանուրը ստվար չէ, ապա ամեն մի հաջորդ վիճակահանությունը իրականացվում է մեկով պակաս թվով տարրեր ունեցող բազմությունից, և X նմուշի բաղադրիչները՝ X_1-ը,…, X_N-ը, չեն լինի միանման բաշխված ու անկախ. այդ հանգամանքը պետք է հաշվի առնվի նմուշի բնութագրիչների հաշվարկման ժամանակ:

Մասնագետը հեշտությամբ կարող է կողմնորոշվել, թե նմուշահանման որ ձևն է ավելի նպատակահարմար տվյալ հետազոտության համար:

X հատկանիշի չափումներից ստացված և գրանցված **տվյալները** կոչվում են **չմշակված**: Առաջին գործողությունը, որը բնական է կատարել՝ այդ թվերի՝ *չնվազման կարգով վերադասավորումն* է: Ստացված վեկտորը կոչվում է **փոփոխման** (վարիացիոն) **շարք**, կամ **կարգավորված նմուշ**, և գրանցվում է հետևյալ կերպ՝

$$\left(x_{(1)}, x_{(2)}, \ldots, x_{(N)}\right):$$

Այդ պարզ գործողությունից հետո ի հայտ են գալիս տվյալների որոշակի հատկություններ. տարբեր արժեքների թիվը, դրանց կրկնվելու հաճախությունները, ամենամեծ՝ $x_{(N)}$ և ամենափոքր՝ $x_{(1)}$ արժեքները, արժեքների փոփոխման միջակայքը՝ $\left(x_{(1)}, x_{(N)}\right)$, և այլն:

> Փոփոխման շարքի տարրերը՝ $x_{(n)}$-երը, կոչվում են **կարգային վիճականիներ**՝
>
> $$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)},$$
>
> իսկ $\left(x_{(1)}, x_{(N)}\right)$ միջակայքի երկարությունը՝ $\left(x_{(N)} - x_{(1)}\right)$-ը, կոչվում է **լայնք**, այն կատարում է նմուշի արժեքների ցրվածության բնութագրիչներից մեկի դերը:

Օրինակ 3: Դիցուք մի մեծ մրցաշարի ընթացքում 30 լավագույն բասկետբոլիստներ հավաքել են այսպիսի միավորներ՝

60, 65, 63, 62, 70, 68, 65, 67, 67, 69, 70, 78, 70, 67, 68,
68, 66, 69, 74, 67, 68, 69, 69, 69, 71, 71, 72, 69, 72, 68 :

Հարկավոր է կազմել փոփոխման շարքը:

Լուծում: Փոփոխման շարքը կլինի՝

60, 62, 63, 65, 65, 66, 67, 67, 67, 67, 68, 68, 68, 68, 68,
69, 69, 69, 69, 69, 69, 70, 70, 70, 71, 71, 72, 72, 74, 78 :

Կարգային վիճականիների գաղափարը լայնորեն օգտագործվում է վիճակագրական տարբեր խնդիրների լուծման, մասնավորապես՝ ոչ պարամետրական գնահատականների և ոչ պարամետրական հայտանիշների կառուցման, իրական համակարգերի և ընթացքների մոդելավորման ժամանակ: Հարկ է նշել, որ ի տարբերություն պատահական նմուշի բաղադրիչների, որպես պատահական մեծություններ՝ *կարգային վիճականիները փոխադարձաբար անկախ չեն* և դրանց մասնակի բաշխման օրենքները միանման չեն, չեն համընկնում հանուրի X հատկանիշի տեսական բաշխման օրենքի հետ, սակայն նրանք կարող են արտահայտվել տեսական $F(x)$ բաշխման ֆունկցիայի միջոցով:

Տվյալների հարմար գրանցման համար հաճախ օգտակար է «ցողուն և տերևներ» կոչվող հնարքը: Դրա կիրառման ժամանակ չափումների տվյալների ընդհանուր մասը կազմում է «ցողունը» (տեղադրվում է ուղղաձիգ գծով), որի վրա «շարվում» են «տերևները»՝ տվյալների փոփոխվող մասը:

Օրինակ 4: Դիցուք փորձնական տվյալներն են՝

0.034	0.034	0.035	0.032	0.033	0.037	0.032	0.035	0.038	0.035
0.032	0.035	0.034	0.030	0.039	0.041	0.035	0.038	0.033	0.034:

Ներկայացնենք դրանք «ցողուն և տերևներ» աղյուսակով:

Լուծում: «Ցողունը» կկազմեն 0.03-ը և 0.04-ը, իսկ «տերևները» կդասավորվեն հետևյալ կերպ՝

```
0.03   1
       –
       222
       33
       4444
       55555
       –
       7
       88
       9
0.04   –
       1  :
```

Ինչպես տեսնում ենք օրինակից, գրառման այս ձևը ոչ միայն խնայում է թվանշանները (օրինակում՝ 80-ի փոխարեն 26), այլև հեշտացնում է հաճախ հանդիպող արժեքների (0.034 և 0.035) բացահայտումը:

6.2. Վիճակագրական բաշխման ֆունկցիա: Հաճախությունների սյունապատկեր

Արդեն նշվել է, որ վիճակագրությունը ծառայում է նմուշից՝ տվյալ կիրառական խնդրի համար օգտակար տեղեկություններ ստանալուն: Այդ նպատակով ստեղծված են տարբեր եղանակներ, գործընթացներ, «գործիքներ», բնութագրիչներ: Դրանց ընկալումն ավելի հեշտ կլինի, եթե նկատի ունենանք, որ հաճախ հանդիպում են հավանականության տեսության գաղափարների և դատողությունների՝ որոշ իմաստով զուգահեռներ. այնտեղ ուսումնասիրվում է պատահական մեծությունը (կամ հավանականային տարածությունը), որի հավանականությունների բաշխումը տրված է, իսկ վիճակագրությունում լուծվում է, կարելի է ասել, *հակադարձ խնդիրը*՝ փորձնական տվյալների, այսինքն՝ նմուշի հիման վրա պետք է եզրակացության գանք հանույթի (կամ նրան համապատասխանող պատահական մեծության) հավանականությունների բաշխման կամ տարբեր բնութագրիչների վերաբերյալ: Նկատի ունենալով այդ տեսակետների «հակադարձ» լինելը, միաժամանակ պետք է գիտակցել դրանց հիմքային նմանությունը:

Ստեղծված են վիճակագրական տվյալների պատկերավոր ներկայացման բազմապիսի ձևեր: Մենք կքննարկենք դրանցից մի քանիսը: Նշենք, որ այդ տեսակետից նույնպես մեծ հնարավորություններ են ընձեռում ժամանակակից հաշվարկները: Վիճակագրության մասնագետը հաճախ պետք է քննարկի տարբեր գծագրերի, գծապատկերների առավելությունները, մինչև որ կգտնի այն, որը լավագույնս է ներկայացնում ստացված տվյալները:

Նմուշի մշակման եղանակները ընդհատ և անընդհատ հատկանիշների դեպքերում որոշ չափով տարբերվում են, ստորև նշված են դրանց առանձնահատկությունները:

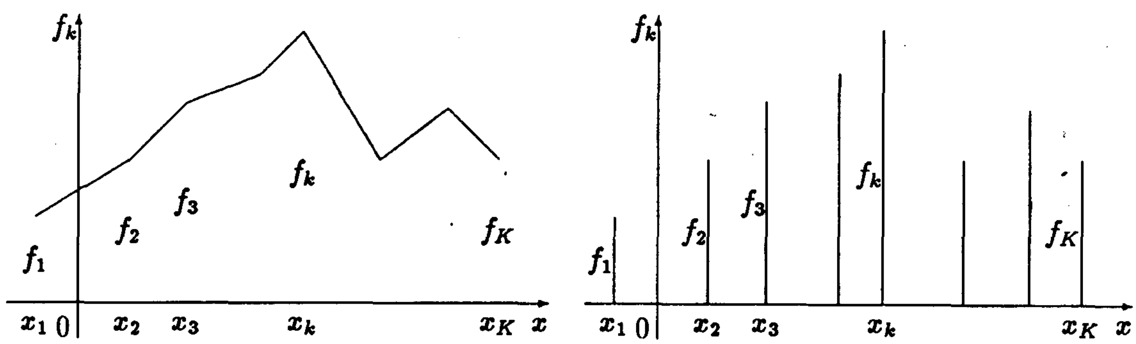
Ընդհատ հատկանիշի դեպքը: Դիցուք հանույթի ուսումնասիրվող X հատկանիշի ընդունած տարբեր արժեքների թիվը մեծ չէ: Երբեմն, դեռ փորձերից առաջ, այդ արժեքների ցուցակը խնդրի բովանդակային իմաստից արդեն հայտնի է: Օրինակ, զառի նետման ժամանակ հնարավոր արժեքները մեկից վեցն են: Իսկ եթե տարբեր արժեքների ցուցակը նախօրոք հայտնի չէ, ապա կարող ենք այն կազմել փորձարկման արդյունքում ստացված կարգավորված նմուշից: Այդ արժեքները կկոչենք **տարբերակներ** և կնշանակենք x_k, $k = \overline{1, K}$, այստեղ K-ն դրանց թիվն է, իսկ k-ն՝ տարբերակի համարը, որը պետք չէ շփոթել դիտման n համարի հետ, $n = \overline{1, N}$: Այնուհետև կազմվում է աղյուսակ (շատ նման ընդհատ պատահական մեծության բաշխման աղյուսակին), որում մասնակցում են աճման կարգով ներկայացված տարբերակները, և դրանց մոտ գրվում է նմուշում ամեն մի տարբերակի կրկնվելու թիվը՝ **բացարձակ հաճախությունը**՝ n_k, կամ **հարաբերական հաճախությունը**՝ f_k՝

X	$x_1,\ x_2,\ \ldots,\ x_K$
n_k	$n_1,\ n_2,\ \ldots,\ n_K$
f_k	$f_1,\ f_2,\ \ldots,\ f_K$

$$f_k = \frac{n_k}{N}, \quad \sum_{k=1}^{K} n_k = N:$$

Այս աղյուսակը (որտեղ բավական է գրել միայն երկրորդ, կամ (ավելի հաճախ) միայն երրորդ տողը) կկոչենք **վիճակագրական բաշխման աղյուսակ**: Այն **վիճակագրական բաշխման օրենքի** ներկայացման ձևերից մեկն է:

Ընդհատ հատկանիշի վիճակագրական բաշխման օրենքը կարելի է պատկերավոր ներկայացնել **հաճախությունների** (բացարձակ կամ հարաբերական) **բազմանկյուն** կոչվող գծապատկերի միջոցով, և **ուղղաձիգ** (կամ **հորիզոնական**) **ձողիկներով** գծապատկերի միջոցով, որոնց կառուցման ընթացքը պարզ է նկար 1-ից:



Նկար 1. Հարաբերական հաճախությունների բազմանկյունը և համապատասխան ուղղաձիգ ձողիկներով գծապատկերը:

Վիճակագրական բաշխման աղյուսակը կառուցելու համար պետք է հաշվել տարբերակների հաճախությունները: Իհարկե, դա դժվար չէ անել փոփոխման շարքը կազմելուց հետո: Սակայն երբեմն հարկավոր է աղյուսակը կազմել, ուղղակի ելնելով չմշակված տվյալներից: Եթե դիտումների թիվը մեծ չէ (օրինակ՝ չի գերազանցում 100-ը), ապա հարմար է օգտվել հետևյալ հնարքից:

Նաշվեփայլիկների (նրբագծերի) **եղանակն** այն է, որ աղյուսակը կազմվում է տվյալների ցուցակը մեկ անգամ հերթով անցնելով, և ամեն **հանդիպած** արժեքի համար աղյուսակում դրվում է մեկ ուղղաձիգ նրբագիծ: Ամեն հինգերորդ նրբագիծը դրվում է թեք, հատելով նախորդ չորսը: Հաշվվում են ամեն տարբերակի մոտ նրբագծերը և հնգյակները, ստացվում են հաճախությունները: Այս գործընթացն ավելի հեշտ է, քան առանձին հաշվելը, թե քանի անգամ է կրկնվել ամեն մի տարբերակը:

<u>Օրինակ</u> 5: Օրինակ 3-ի տվյալներով կառուցենք վիճակագրական բաշխման աղյուսակը:

Լուծում: Կազմենք հաշվեփայտիկներով օժանդակ աղյուսակ, որից հեշտությամբ կստանանք վիճակագրական բաշխման աղյուսակը և $N = \sum_{k=1}^{K} n_k = 30$:

տարբերակներ	հաշվեփայտիկները
60	ǀ
62	ǀ
63	ǀ
64	
65	ǀǀ
66	ǀ
67	ǀǀǀǀ
68	𝍸
69	𝍸ǀ
70	ǀǀǀ
71	ǀǀ
72	ǀǀ
73	
74	ǀ
75	
78	ǀ

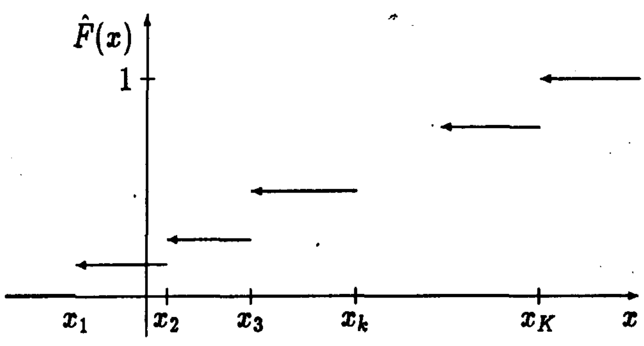
X	n_k
60	1
62	1
63	1
64	0
65	2
66	1
67	4
68	5
69	6
70	3
71	2
72	2
73	0
74	1
75	0
78	1

Ընդհատ հատկանիշի վիճակագրական բաշխման օրենքը ներկայացվում է նաև **վիճակագրական բաշխման ֆունկցիայի** միջոցով, որը տեսական բաշխման ֆունկցիայից տարբերվում է միայն նրանով, որ դրա սահմանման մեջ հավանականությունների փոխարեն մասնակցում են (հարաբերական) հաճախությունները: «Հարաբերական» բառը շատ անգամ չկրկնելու համար մենք այն հաճախ բաց կթողնենք:

Վիճակագրական բաշխման ֆունկցիան՝

$$\hat{F}_X(x) = \sum_{x_k<x} f_k \qquad (1)$$

կպատկերվի ընդհատ պատահական մեծության բաշխման ֆունկցիայի գծապատկերին նմանվող պատկերով (տե՛ս նկար 2):



Նկար 2. Ընդհատ հատկանիշի վիճակագրական բաշխման ֆունկցիան:

Հաճախությունների վիճակագրական կայունության հատկությունը, որը հիմնված է Բեռնուլիի թեորեմի վրա, արդարացնում է $\hat{F}(x)$ ֆունկցիայի օգտագործումը մեծ N – երի դեպքում՝ որպես հատկանիշի անհայտ $F(x)$ տեսական բաշխման ֆունկցիայի մոտարկում:

Դժվար չէ համոզվել, որ վիճակագրական բաշխման ֆունկցիան օժտված է հատկություններով, որոնք, ի դեպ, հավանականությունների բաշխման ֆունկցիայի հատկությունների **հանգունակներն են**.

1. $0 \le \hat{F}_X(x) \le 1$:
2. Եթե $x' \le x''$, ապա $\hat{F}_X(x') \le \hat{F}_X(x'')$:
3. Երբ $x < x_1$, ապա $\hat{F}_X(x) = 0$, երբ $x \ge x_K$, ապա $\hat{F}_X(x) = 1$:

4. $\hat{F}_X(x)$ ֆունկցիան կտոր առ կտոր հաստատուն է, և x_k կետում դրա թռիչքն է՝ $\hat{F}_X(x_k+0) - \hat{F}_X(x_k) = f_k$, $k = \overline{1,K}$:

Սակայն ընդհատ հատկանիշի դեպքում վիճակագրական բաշխման ֆունկցիայի օգտագործումն այնքան էլ արդյունավետ չէ՝ հաճախությունների բեկյալը կամ ձողիկներով գծապատկերն ավելի լավ են ներկայացնում ընդհատ մեծության վիճակագրական բաշխման օրենքը:

Լայնորեն օգտագործվում են նաև տվյալները ներկայացնող գծապատկերների բազմաթիվ այլ ձևեր՝ ժապավենային (bar charts), սեկտորային (pie charts), եռանկյունային, բնակչության բուրգեր (population pyramids՝ ներկայացնում են բնակչության սեռատարիքային բաշխումը) և այլն: Դրանց մեծ մասը հեշտությամբ և արագ կառուցվում է հաշվարային վիճակագրական ծրագրաշարերի օգնությամբ:

Անընդհատ հատկանիշի դեպքը: Երբ ուսումնասիրվող հատկանիշն անընդհատ է, ապա փորձնականորեն ստացված $\mathrm{x} = (x_1, \ldots, x_N)$ նմուշի տվյալները կազմում են հիմնականում չկրկնվող թվերի շարք, որից դժվար է որևէ հատկություն նկատել: Մշակման առաջին քայլը, ինչպես վերը նշվեց, տվյալները փոփոխման շարքի վերածելն է: Սակայն այս դեպքում դրանում նույնպես կլինեն մեծ թվով իրարից տարբերվող արժեքներ:

Երբ նմուշի N ծավալը մեծ է 50-ից, հատկանիշը անընդհատ կամ ընդհատ է, բայց տարբեր հնարավոր արժեքների թիվը մեծ է, ապա հետագա գործողությունները հեշտացնելու նպատակով հարմար է նմուշային տվյալները **ներկայացնել խմբավորված ձևով**: Փոփոխման շարքի առաջին (ամենափոքր) $x_{(1)}$ և վերջին (ամենամեծ) $x_{(N)}$ տարրերով կազմված միջակայքը բաժանում են K հատ հավասար հատվածների: Երբեմն նպատակահարմար է տարբեր երկարության հատվածների բաժանումը, սակայն այդ դեպքի ոչ բարդ առանձնահատկությունների վրա կանգ չենք առնի:

Հատվածների K թիվը պետք է լինի 7-ից 20-ի սահմաններում, K-ի որոշումը կախված է նմուշի N ծավալից: Ընտրության հարցում կողմնորոշվելու համար կարելի է օգտվել **Ստերջեսի մոտավոր բանաձևից**՝

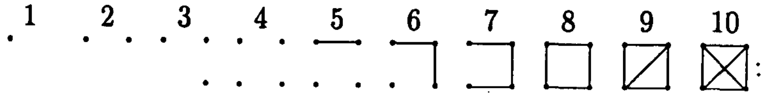
$$K \approx \log_2 N + 1:$$

Ընտրվում են հատվածների բաժանման կետերը՝ $a_0, a_1, \ldots, a_k \ldots a_K$, և հաշվարկվում են այդ հատվածների մեջ ընկած արժեքների քանակները (**բացարձակ հաճախությունները**)՝ $n_1, \ldots, n_k, \ldots, n_K$:

Եթե կան դիտարկված արժեքներ, որոնք հավասար են բաժանման կետերի a_k արժեքներին, ապա պայմանավորվում են դրանք վերագրել, օրինակ, ձախից գտնվող հատվածին: Կամ, որպեսզի բացառվի բաժանման կետերի համընկնումը դիտարկված արժեքների հետ, միջակայքերի բաժանման կետերի համար վերցնում են թվային արժեքներ, որոնք ավելի ստույգ ճշտություն ունեն (ստորակետից հետո՝ մեկ թվանշան ավելի), քան դիտարկվող մեծությունները: Օրինակ, եթե դիտարկվող արժեքները ամբողջ թվեր են, բաժանման c ամբողջ թվի փոխարեն կարելի է վերցնել, օրինակ, c.1 տեսքի թիվ: Իսկ միջակայքի h երկարությունը կարելի է վերցնել

$$h \approx l/K = (x_{(N)} - x_{(1)})/K:$$

Հաճախությունների հաշվարկի համար կարելի է օգտվել հաշվեփայտիկների եղանակից: Իսկ երբ դիտումների թիվն ավելի մեծ է, հարմար է օգտվել **«ծրարների»**՝ [illegible]յալ եղանակից: Միջակայքերի ցուցակը գրվում է սյունակով, հերթականությամբ գննվում են տվյալները և այն միջակայքի աջից, որի մեջ է ընկնում արժեքը, դրվում է կետ կամ գծիկ, այնպես որ հետագայում հաշվարկը հեշտ լինի: Դրվում է մինչև 4 կետ, դրանք կազմում են քառակուսու գագաթները, այնուհետև դրվում է 4 գծիկ՝ քառակուսու կողմերը, և վերջապես՝ 2 անկյունագիծ (ընդամենը՝ 10 տարր): Արդյունքում ամեն մի լրիվ տասնյակին համապատասխանում է մի «ծրար»:



Այնուհետև կազմվում է **վիճակագրական միջակայքային բաշխման աղյուսակը** (երբեմն հարմար է լինում այն տեղադրել ուղղաձիգ, տե՛ս օրինակ 6-ի վերջում):

X	(a_0, a_1),	(a_1, a_2),	$\ldots$,	(a_{k-1}, a_k),	$\ldots$,	(a_{K-1}, a_K)
n_k	n_1	n_2	$\ldots$	n_k	$\ldots$	n_K
f_k	f_1	f_2	$\ldots$	f_k	$\ldots$	f_K :

Օրինակ 6: Մանրամասերի մշակումից հետո չափվել են դրանց տրամագծերը: Արդյունքները ներկայացված են աղյուսակում (միլիմետրերով), հարկավոր է կառուցել միջակայքային բաշխման աղյուսակը:

0.75	0.77	0.77	0.73	0.76	0.74	0.70	0.75	0.71	0.72	0.77	0.79	0.71	0.78	0.73	0.78
0.70	0.73	0.77	0.75	0.74	0.71	0.70	0.78	0.76	0.81	0.69	0.80	0.80	0.77	0.68	0.75
0.74	0.70	0.70	0.74	0.77	0.83	0.76	0.76	0.82	0.77	0.71	0.74	0.77	0.75	0.74	0.77
0.75	0.77	0.72	0.74	0.80	0.75	0.80	0.72	0.78	0.70	0.75	0.78	0.78	0.76	0.77	0.76
0.74	0.74	0.77	0.73	0.74	0.77	0.74	0.75	0.74	0.76	0.76	0.74	0.74	0.74	0.74	0.72
0.76	0.74	0.72	0.80	0.76	0.78	0.73	0.70	0.76	0.76	0.77	0.75	0.78	0.72	0.76	
0.78	0.68	0.75	0.73	0.82	0.73	0.80	0.81	0.71	0.82	0.77	0.80	0.80	0.70	0.70	
0.82	0.72	0.69	0.73	0.76	0.74	0.77	0.72	0.76	0.78	0.78	0.73	0.76	0.80	0.76	
0.72	0.76	0.76	0.70	0.73	0.75	0.77	0.77	0.70	0.81	0.74	0.73	0.77	0.74	0.78	
0.69	0.74	0.71	0.76	0.76	0.77	0.70	0.81	0.74	0.74	0.77	0.75	0.80	0.74	0.76	
0.77	0.77	0.81	0.75	0.78	0.73	0.76	0.76	0.76	0.77	0.76	0.80	0.77	0.74	0.77	
0.72	0.75	0.76	0.77	0.81	0.76	0.76	0.76	0.80	0.74	0.80	0.74	0.73	0.75	0.77	
0.74	0.76	0.77	0.77	0.75	0.76	0.74	0.82	0.76	0.73	0.74	0.75	0.76	0.72	0.78	

Լուծում: Միջակայքերի կառուցման համար որոշենք K-ն, a_0-ն, h-ը: Ըստ բանաձևի՝ $K \approx \log_2 200 + 1 \approx 9$: Քանի որ լայնքը՝ $l = 0.83 - 0.68 = 0.15$, մեկ միջակայքի երկարությունը վերցնենք $h = 0.02$: Ընտրենք a_0-ն: Հաշվենք

$$x_{(1)} - h/2 = 0.68 - 0.01 = 0.67:$$

Որպեսզի բաժանման կետերը չհամընկնեն դիտված արժեքների հետ, վերցնենք $a_0 = 0.669$: Հետևաբար բաժանման կետերը կլինեն՝

$$0.669, 0.689, 0.709, 0.729, 0.749, 0.769, 0.789, 0.809, 0.829, 0.849 :$$

Կազմենք հետևյալ աղյուսակը, կիրառելով «ծրարների» եղանակը, և հաշվենք հաճախությունները:

	տրամագծերի միջակայքերը	հաճախությունների հաշվարկը	բացարձակ հաճախությունը	հարաբերական հաճախությունը
1	0.669 – 0.689		2	0.010
2	0.689 – 0.709		15	0.075
3	0.709 – 0.729		17	0.085
4	0.729 – 0.749		44	0.220
5	0.749 – 0.769		52	0.260
6	0.769 – 0.789		44	0.220
7	0.789 – 0.809		14	0.070
8	0.809 – 0.829		11	0.055
9	0.829 – 0.849		1	0.005
		ընդամենը	200	1

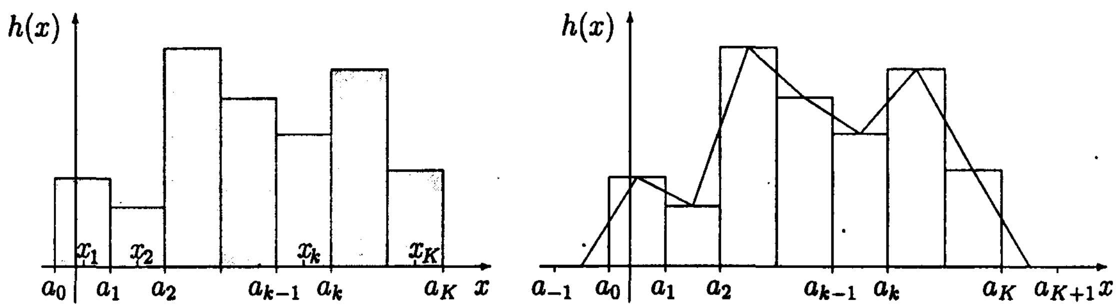
Արդյունքում ստացանք միջակայքային բաշխման աղյուսակը՝

X	f_k
(0.669, 0.689)	0.010
(0.689, 0.709)	0.075
(0.709, 0.729)	0.085
(0.729, 0.749)	0.220
(0.749, 0.769)	0.260
(0.769, 0.789)	0.220
(0.789, 0.809)	0.070
(0.809, 0.829)	0.055
(0.829, 0.849)	0.005 :

Անընդհատ հատկանիշի դեպքում վիճակագրական բաշխման ներկայացման առավել արդյունավետ միջոցներ են **սյունապատկերը (հիստոգրամը)** և **վիճակագրական բաշխման ֆունկցիան**: Սյունապատկերը այն ֆունկցիայի ներկայացումն է, որը հավանականության տեսությունից մեզ ծանոթ անընդհատ պատահական մեծության հավանականությունների բաշխման խտության ֆունկցիայի հանգունակն է, և որը կոչվում է **վիճակագրական խտության ֆունկցիա**:

Վիճակագրական (կամ **նմուշային**) **խտության ֆունկցիան** (նշանակենք այն $h(x)$, $-\infty < x < \infty$) ըստ սահմանման k-րդ միջակայքում, $k = \overline{1, K}$, հաստատուն է և հավասար է այդ միջակայքի վրա կառուցված այն ուղղանկյան բարձրությանը, որի մակերեսը հավասար է միջակայքում դիտված արժեքների f_k հարաբերական հաճախությանը (տե՛ս նկար 3)՝

$$h(x) = \begin{cases} 0, & x < a_0, \quad x \geq a_K, \\ h_k = \dfrac{f_k}{a_k - a_{k-1}}, & a_{k-1} \leq x < a_k, \ k = \overline{1, K} : \end{cases}$$



Նկար 3. Սյունապատկերը և համապատասխան հաճախությունների բազմանկյունը:

Անընդհատ հատկանիշի հաճախությունների բաշխումը կարելի է պատկերել նաև **հաճախությունների բազմանկյան** միջոցով (տե՛ս նկար 3), որը կառուցվում է հետևյալ եղանակով: Սյունապատկերի վրա ավելացվում է միևնույն երկարության երկու միջակայք՝ (a_{-1}, a_0) և (a_K, a_{K+1}), որոնցում հաճախությունը ընդունվում է հավասար 0-ի: Այնուհետև ուղղագիծ հատվածներով հաջորդաբար միացվում են (a_{-1}, a_0) հատվածի միջնակետը $((a_0 + a_1)/2, h_1)$ կետին և այլն, վերջինը կլինի (a_K, a_{K+1}) հատվածի միջնակետը: Կստացվի նկար 3-ում պատկերված բազմանկյունը:

Հաճախությունների բազմանկյունով և x առանցքով սահմանափակված պատկերի մակերեսը և սյունապատկերի գունանշված ընդհանուր մակերեսն **իրար հավասար** են ...սար են մեկի (համոզվե՛ք ինքնուրույն):

Անընդհատ հատկանիշի վիճակագրական բաշխման ֆունկցիան, իհարկե, կարելի է սահմանել և կառուցել համաձայն (1) բանաձևի, բայց կստացվի մեծ թվով փոքր թռիչքներ ունեցող ֆունկցիա, որը կիրառման տեսակետից այնքան էլ հարմար չէ: Հարկավոր է տալ այնպիսի սահմանում, որ անընդհատ հատկանիշի բաշխման ֆունկցիան խզվող չլինի: Ընդունելով $h_X(x)$ ֆունկցիան որպես հաճախությունների բաշխման խտության ֆունկցիա և այն ինտեգրելով՝ կստանանք մի անընդհատ ֆունկցիա, որն անընդհատ հատկանիշի դեպքում բնական է անվանել **հաճախությունների բաշխման ֆունկցիա** կամ **վիճակագրական** (կամ **նմուշային**) **բաշխման ֆունկցիա**:

Դիտողություն: Հարկավոր է գիտակցել, որ $h_X(x)$ ֆունկցիան k-րդ միջակայքում հաստատուն է, և եթե այն ընդունում ենք որպես խտություն, ուրեմն համարում ենք, որ ամեն մի (a_{k-1}, a_k) միջակայքում դիտված արժեքները «հավասարաչափ են բաշխված» (հիշենք, որ հավասարաչափ բաշխման խտության ֆունկցիան հաստատուն է): Այդպիսով, որոշ չափով շեղվում ենք իրականում դիտված տվյալներից, քանի որ դրանց արժեքները նշված հատվածում կարող են գտնվել կամայական դիրքերում, ոչ թե անպայման իրարից հավասար հեռավորության վրա: Սակայն առաջացած շեղումները մեծ չեն, իսկ ստացված բաշխման ֆունկցիան և նրա գծապատկերն ավելի ցայտուն են ներկայացնում դիտարկվող հատկանիշի բաշխման առանձնահատկությունները:

Այսպիսով, անընդհատ հատկանիշի վիճակագրական բաշխման ֆունկցիան՝ $\hat{F}_X(x)$-ը, ստացվում է տեսական $F_X(x)$ բաշխման ֆունկցիայի նման ձևով, միայն թե $f_X(x)$ խտության ֆունկցիային փոխարինում է վիճակագրական խտության ֆունկցիան՝ $h_X(x)$-ը՝

$$\hat{F}_X(x) = \int_{-\infty}^{x} h_X(u)\,du, \quad -\infty < x < \infty :$$

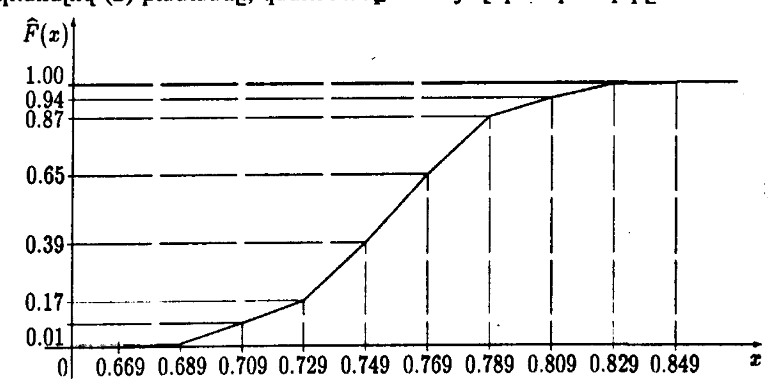
Պարզ է, որ

$$\hat{F}_X(x) = \begin{cases} 0, & x \le a_0, \\ \int_{a_0}^{x} h_X(u)\,du, & a_0 \le x \le a_K, \\ 1, & x \ge a_K : \end{cases} \tag{2}$$

Նկատենք, որ $a_0, a_1, \ldots, a_K$ կետերում $\hat{F}(x)$ ֆունկցիան ճշգրտորեն է համապատասխանում դիտարկված արդյունքներին, այսինքն հավասար է (1) բանաձևով կառուցվածին, իսկ միջակայքերի ներսում այն պարզության և հարմարության համար ընդունվել է ուղղագիծ և անընդհատ:

Օրինակ 7: Օրինակ 6-ում ստացված միջակայքային բաշխման աղյուսակի տվյալներով կառուցենք նմուշային բաշխման ֆունկցիան:

Լուծում: Կիրառելով (2) բանաձևը, կստանանք հետևյալ գծապատկերը՝



Նկար 4. Օրինակ 6-ի տվյալներին համապատասխանող նմուշային բաշխման ֆունկցիան:

Ընդհատ հատկանիշի վիճակագրական բաշխման ֆունկցիայի՝ վերը նշված հատկությունները, բացի չորրորդից, մնում են ուժի մեջ նաև անընդհատ հատկանիշի դեպքում, չորրորդը փոխարինվում է անընդհատության հատկությամբ:

Հարց է ծագում՝ ինչքանո՞վ է տարբերվում $\hat{F}_X(x)$ նմուշային բաշխման ֆունկցիան X հատկանիշի իրական $F_X(x)$ բաշխման ֆունկցիայից: Մեծ N-երի դեպքում $\hat{F}_X(x)$ վիճակագրական ֆունկցիան մոտ է տեսականին: Ավելի ճշգրիտ՝ ամեն մի x կետում $-\infty < x < \infty$, $\hat{F}_X(x)$ նմուշային բաշխման ֆունկցիան նմուշի N ծավալի աճման հետ ձգտում է ըստ հավանականության տեսական $F_X(x)$ բաշխման ֆունկցիային:

Վարկածների ստուգմանը նվիրված գլուխ 9-ում դիտարկվում են տեսական և նմուշային բաշխման ֆունկցիաների համեմատությանը նվիրված ընթացակարգեր, որոնք կոչվում են **համաձայնության հայտանիշներ**:

6.3 Կենտրոնական դիրքի նմուշային բնութագրիչներ

Նմուշային տվյալների թվային բնութագրիչներից առաջին հերթին դիտարկենք կենտրոնական դիրքի բնութագրիչները: Դրանք են. **միջին թվաբանականը** (կրճատ՝ **միջինը**), **կիսողը**, **մոդը** և այլ միջինները: Դրանցից ամեն մեկն ունի իր առավելությունները ու թերությունները և կիրառվում է, նայած թե ինչ նպատակներ ունի հետազոտողը՝ տվյալները մշակելիս: Հիշենք գլուխ 3-ում ուսումնասիրված պատահական մեծության բնութագրիչների սահմանումները: Վիճակագրական բնութագրիչները դրանց հանգունակներն են, սահմանումները շատ նման են: Սկսենք կիսողից:

Նմուշային կիսողը՝ $\hat{x}_{\mathrm{med}}$-ը, այն արժեքն է, որը բաժանում է փոփոխման շարքը երկու խմբի՝ $\hat{x}_{\mathrm{med}}$-ից փոքր և $\hat{x}_{\mathrm{med}}$-ից մեծ, այնպես, որ խմբերում լինեն հավասար քանակով դիտարկված արժեքներ:

Այդ ընդհանուր սահմանումից կիսողը հաճախ միակ ձևով չի որոշվում: Ընդհատ և անընդհատ հատկանիշների համար ընդունված են հաշվարկի բանաձևեր, որոնք տալիս են որոշակի արժեք:

Եթե X հատկանիշը **ընդհատ** է, N-ը նմուշի ծավալն է, իսկ $x_{(l)}$-ը՝ l-րդ կարգային վիճականին, ապա

$$\hat{x}_{\mathrm{med}} = \begin{cases} (x_{(l)} + x_{(l+1)})/2, & N = 2l, \\ x_{(l+1)}, & N = 2l+1 : \end{cases}$$

Կիսողը գտնելու համար օգտակար է տվյալները ներկայացնել «ցողուն և տերևներ» աղյուսակով, ստանալով նաև դրանց կարգավորումը:

Անընդհատ հատկանիշի դեպքում՝ ելնելով խմբավորված տվյալների միջակայքային աղյուսակից, **կիսողը հաշվարկում** են հետևյալ բանաձևով՝

$$\hat{x}_{\mathrm{med}} = a_{l-1} + \frac{1}{n_l}\left(\frac{N}{2} - \sum_{k=1}^{l-1} n_k\right)(a_l - a_{l-1}), \tag{3}$$

որտեղ l-ը այն միջակայքի համարն է, որի համար $\sum_{k=1}^{l-1} n_k < N/2$, և $\sum_{k=1}^{l} n_k > N/2$:

Կիսողը կարելի է մոտավորապես հաշվարկել նմուշային բաշխման ֆունկցիայի՝ $\hat{F}(x)$-ի ըմբռնումների օգնությամբ (տե՛ս նկար 5):

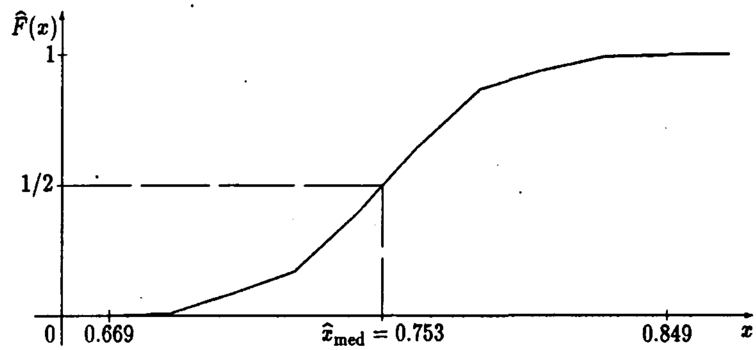
Օրինակ 8: Օրինակ 6-ում դիտարկված տրամագծերի կիսողը գտնենք (3) բանաձևով և նկար 4-ում պատկերված նմուշային բաշխման ֆունկցիայի օգնությամբ, գտնելով այն x արժեքը, որի դեպքում $\hat{F}(x) = 1/2$:

Լուծում: (3) բանաձևը կիրառելու համար նկատենք, որ $l = 5$, ուրեմն

$$\hat{x}_{\text{med}} = 0.749 + \frac{100-78}{52}0.02 = 0.749 + 0.0085 = 0.7575 :$$

Իսկ նմուշային բաշխման ֆունկցիայի գծապատկերից (տե՛ս նկար 5) կգտնենք կիսողի մոտավոր արժեքը՝

$\hat{x}_{\text{med}} \approx 0.753$:



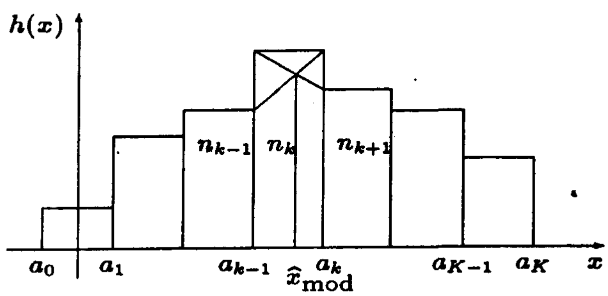
Նկար 5. Կիսողի որոշումը նմուշային բաշխման ֆունկցիայի գծապատկերի օգնությամբ:

Նմուշային մոդը՝ $\hat{x}_{\text{mod}}$, նմուշի արժեքների փոփոխման ամբողջ տիրույթում կամ որևէ ենթատիրույթում ամենահաճախ տարբերակն է: Մոդը կարող է գոյություն չունենալ կամ միակը չլինել, այդ դեպքում բաշխումը կոչվում է **բազմամոդալ**:

Անընդհատ հատկանիշի դեպքում մոդը միջակայքային բաշխման աղյուսակի հիման վրա հաշվարկվում է հետևյալ բանաձևով (a_{l-1}-ը մոդը պարունակող **մոդալ միջակայքի** ձախ տերմինն է)՝

$$\hat{x}_{\text{mod}} = a_{l-1} + \frac{n_l - n_{l-1}}{2n_l - n_{l-1} - n_{l+1}}(a_l - a_{l-1}) : \qquad (4)$$

Կարելի է ապացուցել (խորհուրդ ենք տալիս կատարել ինքնուրույն), որ ստորև բերված նկար 6-ի սյունապատկերի վրա կատարված կառուցումը տալիս է x_{mod}-ի նույն արժեքը, որը ստացվում է (4) բանաձևի հաշվարկով:



Նկար 6. Մոդի որոշումը երկրաչափական եղանակով:

Օրինակ 9: Օրինակ 5-ում ստացված վիճակագրական բաշխման աղյուսակից տեսնում ենք, որ մոդը միակն է և $\hat{x}_{\text{mod}} = 69$: Օրինակ 6-ում անընդհատ հատկանիշի միջակայքային բաշխման ստացված աղյուսակի տվյալներով գտնենք մոդը:

Լուծում: Պարզ է, որ մոդը միակն է: Համաձայն (3) բանաձևի, քանի որ մոդալ միջակայքը հինգերորդն է՝ $l = 5$, ապա

$$\hat{x}_{\text{mod}} = 0.749 + \frac{52-44}{2\cdot 52 - 44 - 44}\cdot 0.02 = 0.749 + \frac{1}{2}\cdot 0.02 = 0.759 :$$

Ինչպես նշեցինք, նախնական ուսումնասիրության նպատակով նմուշը դիտվում է որպես N պատահական մեծությունների վեկտոր $\mathbf{X} = (X_1, \ldots, X_N)$, իսկ փորձարկումներից ստացված նմուշը այդ վեկտորի մի իրականացումն է՝ $\mathbf{x} = (x_1, \ldots, x_N)$: Համապատասխանաբար **նմուշային միջինը** որպես գործողություն նմուշի նկատմամբ նշանակվում է $\overline{\mathbf{X}}$, իսկ որպես այդ գործողության արդյունք թվային տվյալների նկատմամբ նշանակվում է $\overline{\mathbf{x}}$ կամ $\overline{x}$:

$$\overline{\mathbf{X}} = \frac{1}{N}\sum_{n=1}^{N} X_n, \quad \overline{\mathbf{x}} = \overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n : \tag{5}$$

Ընդգծենք, որ $\overline{\mathbf{X}}$-ը *պատահական մեծություն* է, որի մի իրականացումն է $\overline{\mathbf{x}}$-ը:

Եթե N-ը մեծ է, ապա թվային հաշվարկների համար նպատակահարմար է ընդհատ հատկանիշի դեպքում օգտվել վիճակագրական բաշխման աղյուսակից, իսկ անընդհատ դեպքում՝ վիճակագրական միջակայքային բաշխման աղյուսակից: Որպեսզի հնարավոր լինի կիրառել միևնույն բանաձևը, անընդհատ դեպքում միջակայքային բաշխումը հարմար է «դարձնել ընդհատ»՝ k-րդ միջակայքի տարբեր արժեքները փոխարինել n_k անգամ կրկնվող **միջակայքի ներկայացուցիչ** x_k **արժեքով**, որը վերցվում է հավասար միջակայքի ծայրակետերի կիսագումարի (կլորացված) արժեքին (տե՛ս նկար 3): Այժմ

և՛ ընդհատ, և՛ անընդհատ հատկանիշների նմուշային միջինը հաշվարկվում է այսպես՝

$$\overline{\mathbf{x}} = \overline{x} = \frac{1}{N}\sum_{k=1}^{K} n_k x_k = \sum_{k=1}^{K} f_k x_k : \tag{6}$$

Դիտողություն: Իհարկե, ընդհատ դեպքում (5) և (6) բանաձևերը տալիս են նույն արդյունքը: Իսկ անընդհատ դեպքում (6)-ի մեջ ամեն միջակայքի դիտարկված արժեքները փոխարինվել են միջակայքի ներկայացուցիչ x_k արժեքով, առաջացնելով որոշ անճշտություն, սակայն իրական արժեքների մի մասը x_k-ից փոքր է, իսկ մյուսը՝ մեծ, այնպես որ տարբերությունը աննշան պիտի լինի և կարելի է այն անտեսել:

Օրինակ 10: Հաշվել օրինակ 6-ի տվյալների նմուշային միջինը (5) և (6) բանաձևերով:

Լուծում : Եթե կիրառենք (5) բանաձևը, երկար հաշվարկից հետո կստանանք՝

$$(0.75 + \ldots + 0.78)/200 = 0.7538 :$$

Իսկ (6) բանաձևը կիրառելու համար լրացնենք ներկայացուցիչ արժեքները միջակայքային բաշխման աղյուսակում: Միջակայքերի ծայրակետերը մենք վերցրել էինք ոչ թե $0.67, 0.69, \ldots$, այլ $0.669, 0.689, \ldots$: Այժմ վերցնում ենք $x_1 = 0.68$, կլորացնելով 0.679 արժեքը և նման ձևով՝ բոլոր ներկայացուցիչ արժեքները:

X	f_k	x_k
(0.669, 0.689)	0.010	0.68
(0.689, 0.709)	0.075	0.70
(0.709, 0.729)	0.085	0.72
(0.729, 0.749)	0.220	0.74
(0.749, 0.769)	0.260	0.76
(0.769, 0.789)	0.220	0.78
(0.789, 0.809)	0.070	0.80
(0.809, 0.829)	0.055	0.82
(0.829, 0.849)	0.005	0.84 :

Ըստ (6) բանաձևի՝

$$\overline{x} = \frac{1}{200}\sum_{k=1}^{9} n_k x_k = \frac{1}{200}(2 \cdot 0.68 + 15 \cdot 0.70 + 17 \cdot 0.72+$$
$$+44 \cdot 0.74 + 52 \cdot 0.76 + 44 \cdot 0.78 + 14 \cdot 0.80 + 11 \cdot 0.82 + 1 \cdot 0.84) = 0.7578 :$$

Ինչպես տեսնում ենք, արդյունքների տարբերությունը՝ 0.0040, խիստ բավականին փոքր է:

Երբ դիտվում է երկու հատկանիշի՝ X և Y, նկատմամբ ստացված N_1 ծավալի $\mathbf{X}$ նմուշը և N_2 ծավալի $\mathbf{Y}$ նմուշը, դրանց գումարը՝ $\mathbf{X}+\mathbf{Y}$ (կամ $\mathbf{x}+\mathbf{y}$), սահմանվում է որպես համապատասխան բաղադրիչների գումարների վեկտոր՝

$$\mathbf{X}+\mathbf{Y} = (X_1 + Y_1, X_1 + Y_2, \ldots, X_{N_1} + Y_{N_2}),$$
$$\mathbf{x}+\mathbf{y} = (x_1 + y_1, x_1 + y_2, \ldots, x_{N_1} + y_{N_2}) :$$

Նման ձևով սահմանվում է $\mathbf{X}$ նմուշի նկատմամբ գծային ֆունկցիան (α-ն և β-ն հաստատուններ են)

$$\mathbf{Y} = \alpha\mathbf{X} + \beta = (\alpha X_1 + \beta, \ldots, \alpha X_N + \beta) :$$

Միջինի հետևյալ օգտակար հատկությունները բխում են սահմանումից:

1. $\overline{\mathbf{X}+\mathbf{Y}} = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$, $\overline{\mathbf{x}+\mathbf{y}} = \overline{\mathbf{x}} + \overline{\mathbf{y}}$:
2. Եթե $\mathbf{Y} = \alpha\mathbf{X} + \beta$, ապա $\overline{\mathbf{Y}} = \alpha\overline{\mathbf{X}} + \beta$, համապատասխանաբար՝ $\overline{\mathbf{y}} = \alpha\overline{\mathbf{x}} + \beta$:
3. $\min\limits_a \sum\limits_{k=1}^{K} n_k(x_k - a)^2 = \sum\limits_{k=1}^{K} n_k(x_k - \overline{\mathbf{x}})^2$:

Փորձնականորեն նկատվել է, որ միջինը, կիսողը և մոդը բավարարում են հետևյալ առնչությանը:

Եթե նմուշի հաճախությունների բազմանկյունը միամոդալ է և շատ անհամաչափ չէ, ապա

$$\overline{x} - \hat{x}_{\text{mod}} \approx 3(\overline{x} - \hat{x}_{\text{med}}) :$$

Նշենք միջինի և կիսողի առանձնահատկությունները (առավելությունները և թերությունները), որոնք նկատի են առնվում կիրառելիս:

Միջին թվաբանականը քիչ է փոփոխվում նմուշահանման արդյունքների տատանումների հետևանքով, հարմար է համեմատվում այլ միջինների հետ: Սակայն նմուշի առանձին ընկած հեռավոր արժեքները զգալիորեն ազդում են միջինի վրա: Կիսողը ավելի զգայուն է նմուշահանման տատանումների նկատմամբ, իսկ հեռավոր արժեքներից կախվածությունը թույլ է: Սակայն կիսողի նկատմամբ հանրահաշվական գործողությունները հարմար չեն: Կիսողի առավելությունը դրսևորվում է այն դեպքում, երբ միջինը և մոդը գոյություն չունեն (կամ մոդը միակը չէ): Համոզվենք դրանում օրինակի օգնությամբ: Դիցուք 89 փոստային աղավնիներ Ա վայրից միաժամանակ ուղարկվում են Բ վայրը: Դրանց թռիչքի ժամանակների կիսողը որոշվում է, երբ Բ վայրն է հասնում 45-րդ աղավնին, կարիք չկա սպասելու մնացածին (մինչդեռ միջինը հաշվելու համար պետք է սպասել բոլորին, և հնարավոր է նույնիսկ, որ ոչ բոլորը տեղ հասնեն):

Որոշ դեպքերում օգտակար են լինում թվային տվյալների միջին երկրաչափականը, միջին ներդաշնակը, միջին քառակուսայինը: Դրանց սահմանումները շատ նման են գլուխ 3-ի համապատասխան սահմանումներին. նմուշի

միջին քառակուսայինն է՝

$$\hat{x}_Q = \sqrt{\frac{1}{N}\sum_{k=1}^{K} x_k^2 n_k} = \left(\sum_{k=1}^{K} x_k^2 \frac{n_k}{N}\right)^{1/2},$$

միջին ներդաշնակն է՝

$$\hat{x}_H = \frac{1}{\sum_{k=1}^{K} \frac{n_k}{N}\frac{1}{x_k}} = \left(\sum_{k=1}^{K} \frac{n_k}{N} x_k^{-1}\right)^{-1},$$

միջին երկրաչափականն է (միայն դրական X հատկանիշի համար)՝

$$\hat{x}_G = \sqrt[N]{\prod_{k=1}^{K} x_k^{n_k}} = \exp\left\{\frac{1}{N}\sum_{k=1}^{K} n_k \ln x_k\right\}:$$

Կարելի է ապացուցել, որ տեղի ունեն հետևյալ անհավասարությունները, $\hat{x}_H \le \hat{x}_G \le \overline{x} \le \hat{x}_Q$, որտեղ հավասարությունները տեղի ունեն այն և միայն այն դեպքում, երբ բոլոր x_k-երն իրար հավասար են:

Նշված միջինները հետևյալի մասնավոր դեպքերն են:

Դիցուք $\psi(x)$-ը, $x \in \mathcal{R}$, անընդհատ **միընթաց** (մոնոտոն) ֆունկցիա է, որը, հետևաբար, ունի հակադարձ՝ ψ^{-1}: x **նմուշի ψ-միջին** կոչվում է հետևյալ առնչությանը բավարարող $\hat{x}_\psi$ թիվը՝ $$\psi(\hat{x}_\psi) = \frac{1}{N}\sum_{k=1}^{K} n_k \psi(x_k):$$

Եթե X հատկանիշը դրական է, իսկ $\psi(x) = x^r$, որտեղ r-ը ամբողջ թիվ է, ապա ψ-միջինը կոչվում է r-րդ **կարգի միջին**՝

$$(\hat{x}_r)^r = \frac{1}{N}\sum_{k=1}^{K} n_k x_k^r:$$

Երբ $r = -1, 1, 2$, ստանում ենք, համապատասխանաբար՝ $\hat{x}_{-1} = \hat{x}_H$, $\hat{x}_1 = \overline{x}$, $\hat{x}_2 = \hat{x}_Q$:

6.4. Նմուշային մոմենտներ, քանորդիչներ, ցրվածության, անհամաչափության և կուտակվածության բնութագրիչներ

Վիճակագրական տվյալները լինում են այս կամ այն չափով սփռված, դրանց փոփոխման տիրույթը կարող է լինել ավելի կամ պակաս լայն դասավորված: Այդ տեսակետից կան նմուշային տարբեր բնութագրիչներ: Ցրվածության պարզագույն բնութագրիչն է **նմուշի լայնքը**՝ $(x_{(N)} - x_{(1)})$, այսինքն՝ նմուշի մեծագույն և փոքրագույն արժեքների տարբերությունը (սահմանվել է 6.1. ենթաբաժնում):

Պատահական մեծությունների ցրվածության բնութագրիչներին ծանոթ ենք գլուխ 3–ից: Դրանց վիճակագրական հանգունակները սահմանվում են նմուշային մոմենտների օգնությամբ: Սկսենք դրանց ընդհանուր սահմանումից:

x նմուշի՝ x_0-ի **նկատմամբ r-րդ կարգի մոմենտները**՝ ${}_{x_0}\hat{m}_r$, սահմանվում են հետևյալ կերպ՝ $${}_{x_0}\hat{m}_r = \frac{1}{N}\sum_{k=1}^{K} n_k (x_k - x_0)^r, \quad r = 0, 1, 2, \ldots:$$

Կարևոր են հետևյալ դեպքերը:

Երբ $x_0 = 0$, մոմենտները կոչվում են **սկզբնական** և նշանակվում են $\hat{m}_r$: Պարզ է, որ $\hat{m}_0 = 1$, $\hat{m}_1 = \overline{x}$:

Երբ $x_0 = \overline{x}$, մոմենտները կոչվում են կենտրոնական և նշանակվում են՝ $\hat{\mu}_r$: Պարզ է, որ $\hat{\mu}_0 = 1$, $\hat{\mu}_1 = 0$:

Մինչև փորձնական տվյալների ձեռք բերումը, նախնական հետազոտության նպատակով դիտարկում են $\mathbf{X} = (X_1, \ldots, X_N)$ նմուշը (որպես պատահական վեկտոր), և դրա բնութագրիչները, օրինակ, սկզբնական մոմենտները՝

$$\hat{m}_r(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N} X_n^r,$$

և կենտրոնական մոմենտները՝

$$\hat{\mu}_r(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N} (X_n - \overline{\mathbf{X}})^r :$$

Տեղի ունեն գլուխ 3-ում նշված առնչությունների հանգունակները՝

$$\hat{\mu}_2 = \hat{m}_2 - \hat{m}_1^2, \quad \hat{\mu}_3 = \hat{m}_3 - 3\hat{m}_1\hat{m}_2 + 2\hat{m}_1^3,$$
$$\hat{\mu}_4 = \hat{m}_4 - 4\hat{m}_1\hat{m}_3 + 6\hat{m}_1^2\hat{m}_2 - 3\hat{m}_1^4:$$

Ցրվածության բնութագրիչներից առավել հաճախ օգտագործվողներն են **նմուշային** (կամ **վիճակագրական**, կամ **փորձնական** (էմպիրիկ)) **ցրվածքը**՝ $\hat{\mathrm{D}}(\mathrm{x})$ և **նմուշային միջին քառակուսային** **շեղումը**՝ $\hat{\sigma}(\mathrm{x})$-ը՝

$$\hat{\mu}_2 = \hat{\mathrm{D}}(\mathrm{x}) = \frac{1}{N}\sum_{k=1}^{K} n_k(x_k - \overline{x})^2, \quad \hat{\sigma}(\mathrm{x}) = \sqrt{\hat{\mathrm{D}}(\mathrm{x})} : \qquad (7)$$

Ցրվածքի որոշ թերությունն այն է, որ նրա արժեքը արտահայտվում է պատահական մեծության չափի միավորների քառակուսիներով, իսկ նույն միավորներով արտահայտվող ցրվածության բնութագրիչը միջին քառակուսային շեղումն է:

Տեսական ուսումնասիրություններում դիտարկվում է $\mathbf{X}$-ի **նմուշային ցրվածքը**՝

$$\hat{\mathrm{D}}(\mathbf{X}) = \hat{\mu}_2(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{n} (X_n - \overline{\mathbf{X}})^2,$$

այսինքն՝ $\hat{\mathrm{D}}(\mathrm{x})$-ը $\hat{\mathrm{D}}(\mathbf{X})$-ի այն արժեքն է, որը ստացվում է, երբ x-ը $\mathbf{X}$-ի իրականացումն է: Նմանապես $\hat{\sigma}(\mathrm{x})$-ը $\hat{\sigma}(\mathbf{X})$-ի արժեքն է:

Նմուշային ցրվածքի կարևոր հատկություններն են (համեմատության համար տե՛ս գլուխ 3)՝

1. $\hat{\mathrm{D}}(\mathbf{X} + a) = \hat{\mathrm{D}}\mathbf{X}$,
2. $\hat{\mathrm{D}}(a\mathbf{X}) = a^2\hat{\mathrm{D}}(\mathbf{X})$,
3. $\hat{\mathrm{D}}(\mathrm{x}) = \frac{1}{N}\sum_{k=1}^{K} n_k x_k^2 - \left(\frac{1}{N}\sum_{k=1}^{K} n_k x_k\right)^2 :$

Օրինակ 11: Օրինակ 10-ում ստացված աղյուսակում ներկայացուցիչների արժեքներով (7) բանաձևերով հաշվարկել նմուշային ցրվածքը և միջին քառակուսային շեղումը:

Լուծում : Քանի որ այդ ներկայացուցիչների միջինը հավասար է 0.7578, ապա

$$\hat{\mathrm{D}}(\mathrm{x}) = (0.68 - 0.7538)^2 \cdot 0.010 + (0.70 - 0.7538)^2 \cdot 0.075 + \ldots + (0.84 - 0.7538)^2 \cdot 0.005 =$$
$$= 0.000824686 \approx 0.0008,$$

որտեղից՝

$$\hat{\sigma}(\mathbf{x}) = 0.028717346 \approx 0.029:$$

Համախությունների բազմանկյան հաջորդ ամենատարածված բնութագրիչներն են անհամաչափության և կուտակվածության բնութագրիչները: Ավելի հաճախ կիրառվում են Ֆիշերի կողմից առաջարկված երրորդ և չորրորդ կարգի կենտրոնական մոմենտների միջոցով սահմանվող բնութագրիչները: Դրանք են (տե՛ս գլուխ 3) **անհամաչափության և կուտակվածության գործակիցները՝**

$$\hat{\beta}(\mathbf{x}) = \frac{\hat{\mu}_3(\mathbf{x})}{\hat{\sigma}^3(\mathbf{x})}, \quad \hat{\gamma}(\mathbf{x}) = \frac{\hat{\mu}_4(\mathbf{x})}{\hat{\sigma}^4(\mathbf{x})} - 3: \tag{8}$$

Լրիվ պահպանվում են դրանց տեսական հանգունակների հատկությունները, որ նշված են գլուխ 3-ում և ներկայացված են համապատասխան գծապատկերներում:

<u>*Օրինակ*</u> 12: Օրինակ 10-ի աղյուսակի տվյալներով և (8) բանաձևերով հաշվարկել անհամաչափության և կուտակվածության գործակիցները:

Լուծում: Հաշվենք $\hat{\mu}_3$-ը և $\hat{\mu}_4$-ը՝

$$\hat{\mu}_3 = (0.68 - 0.7578)^3 \cdot 0.010 + \ldots + (0.84 - 0.7578)^3 \cdot 0.005 \approx -0.00000134,$$
$$\hat{\mu}_4 = (0.68 - 0.7578)^4 \cdot 0.010 + \ldots + (0.84 - 0.7578)^4 \cdot 0.005 \approx 0.000002835:$$

Հետևաբար՝

$$\hat{\beta}(\mathbf{x}) = -\frac{0.00000134}{0.00002378} = -0.0563, \; \hat{\gamma}(\mathbf{x}) = \frac{0.000002835}{0.000000707} - 3 = 4.001 - 3 = 1.001 > 0:$$

Ստացանք, որ բաշխումը թեքված է դեպի աջ և նորմալ կորից ավելի «սրագագաթ» է:

Նման նպատակների են ծառայում երկրորդ խմբի բնութագրիչները, որ սահմանվում են քանորդչի գաղափարի հիման վրա: Դրանք հարմար է կիրառել հատկապես այն դեպքերում, երբ մոմենտները գոյություն չունեն:

Հիշեցնենք քանորդիչների սահմանումը՝ նմուշային q **քանորդիչն** այն $\hat{x}_q$ արժեքն է, որից փոքր են դիտումների ոչ ավել քան q մասը, իսկ մեծ են՝ $(1-q)$ մասը: Քանորդիչների հաշվարկումը **ընդհատ և անընդհատ հատկանիշների դեպքում** նման է կիսողի հաշվարկմանը:

Քառորդիչները բաժանում են տվյալները չորս մասի, **տասնորդիչները**՝ տաս մասի, **հարյուրորդիչները**՝ հարյուր: Ուրեմն կա երեք քառորդիչ՝ ստորին՝ $\hat{u}_{0.25}$, կիսողը՝ $\hat{u}_{0.5} = \hat{x}_{\text{med}}$ և վերին՝ $\hat{u}_{0.75}$: Հարյուրորդիչներն անվանում են առաջին, երկրորդ և այլն: Կիսողը հավասար է 5-րդ տասնորդչին և 50-րդ հարյուրորդչին:

Լայնքը շատ պարզ սահմանում ունի, սակայն այն հնարավոր չէ հաշվարկել խմբավորված տվյալների դեպքում, բացի դրանից այն հաշվի չի առնում ծայրագույն արժեքների միջև ընկած այլ արժեքները: Ավելի կիրառելի է **միջքառորդչային լայնքը՝** $\hat{u}_{0.75} - \hat{u}_{0.25}$, կամ **միջքառորդչային կիսալայնքը՝** $(\hat{u}_{0.75} - \hat{u}_{0.25})/2$, դրանք հաշվի են առնում նմուշի բոլոր արժեքները (տե՛ս նկար 7):

$\hat{u}_{0.25}$ $\quad$ $\hat{u}_{0.5} = \hat{x}_{\text{med}}$ $\quad$ $\hat{u}_{0.75}$

Նկար 7. Քառորդիչների դասավորությունը:

Երբեմն դիտարկում են **միջհարյուրորդչային հեռավորությունը՝** $\hat{u}_{0.90} - \hat{u}_{0.10}$, համապատասխան $(\hat{u}_{0.10}, \hat{u}_{0.90})$ միջակայքում գտնվում է տվյալների 80%-ը:

Երբեմն օգտակար են. **քանորդչային անհամաչափության գործակիցը՝**

$$\beta_1 = \frac{\hat{u}_{0.75} - \hat{u}_{0.5} - (\hat{u}_{0.5} - \hat{u}_{0.25})}{\hat{u}_{0.75} - \hat{u}_{0.25}} = \frac{\hat{u}_{0.75} - 2\hat{u}_{0.5} + \hat{u}_{0.25}}{\hat{u}_{0.75} - \hat{u}_{0.25}},$$

որը հավասար է 0-ի համաչափ նմուշի դեպքում, դրական է, երբ դրական է Ֆիշերի $\hat{\beta}$-ն, և **կուտակվածության գործակիցը**՝

$$\hat{\gamma}_1 = \frac{\hat{u}_{0.75} - \hat{u}_{0.25}}{2(\hat{u}_{0.9} - \hat{u}_{0.1})},$$

որը նորմալ բաշխված հատկանիշի դեպքում հավասար է 0.263, իսկ եթե փոքր է այդ արժեքից, ապա կուտակվածությունը մեծ է:

Օրինակ 13: Օրինակ 6-ի տվյալներով հաշվարկել $\hat{u}_{0.1}, \hat{u}_{0.25}, \hat{u}_{0.5}, \hat{u}_{0.75}, \hat{u}_{0.9}$ քանորդիչները և այնուհետև գտնել $\hat{\beta}_1$-ը և $\hat{\gamma}_1$-ը:

Լուծում: Կատարենք հաշվարկը (3) բանաձևով $\hat{u}_{0.5} = \hat{x}_{\text{med}}$ համար, այնուհետև նույն ձևով՝ տեղադրելով համապատասխան l-ը և Nq-ն:

$\hat{u}_{0.1} = 0.709 + (20 - 17) \cdot 0.02/17 = 0.71253,$

$\hat{u}_{0.25} = 0.729 + (50 - 34) \cdot 0.02/44 = 0.73627,$

$\hat{u}_{0.5} = 0.749 + (100 - 78) \cdot 0.02/52 = 0.75746,$

$\hat{u}_{0.75} = 0.769 + (150 - 130) \cdot 0.02/44 = 0.77809,$

$\hat{u}_{0.9} = 0.789 + (180 - 174) \cdot 0.02/14 = 0.79757,$

$$\hat{\beta}_1 = \frac{0.77809 - 2 \cdot 0.75746 + 0.73627}{0.77809 - 0.73627} = \frac{0.02063 - 0.02119}{0.04182} = -\frac{0.00056}{0.04182} = -0.01339 < 0,$$

$$\hat{\gamma}_1 = \frac{0.04182}{2(0.79757 - 0.71253)} = \frac{0.04182}{2 \cdot 0,08504} = \frac{0.04182}{0.17008} = 0.24588 < 0.263 :$$

Այսպիսով, արդյունքն այն է, որ տվյալները թեքված են դեպի աջ և ավելի կուտակված են, քան նորմալ բաշխման դեպքում: Տեսնում ենք, որ եզրակացությունը համընկնում է օրինակ 12-ի արդյունքի հետ:

6.5. Երկու հատկանիշի նմուշային նկարագրումը

Նկատենք, որ 6.3. ենթաբաժնում մենք արդեն անդրադարձել ենք երկու նմուշների գումարի միջինի հատկություններին: Այժմ դիտարկենք այն դեպքը, երբ գրանցվում են անհատի երկու կամ մի քանի հատկանիշներ: Եթե միաժամանակ ստացվել են տվյալներ երևույթի երկու հատկանիշների վերաբերյալ, ապա կազմվում է «երկու մուտքով» աղյուսակ (տե՛ս 2.5 ենթաբաժինը և գլուխ 3-ի օրինակ 6-ը): Ենթադրենք, միաժամանակ գրանցվում են որոշակի թվով անձանց հասակը և քաշը կամ մի խումբ ընտանիքների երեխաների թիվը և բնակարանի սենյակների թիվը և այլն: Դիցուք գրանցվող ընդհատ հատկանիշներն են X-ը և Y-ը, իսկ դրանց հնարավոր արժեքներն են, համապատասխանաբար, x_k, $k = \overline{1,K}$, և y_i, $i = \overline{1,I}$, նմուշն է $(\mathbf{X},\mathbf{Y}) = ((X_1,Y_1),\dots,(X_N,Y_N))$, իսկ դիտված արժեքները $(\mathbf{x},\mathbf{y}) = ((x_1,y_1),\dots,(x_N,y_N))$, իսկ n_{ki}-ն N ծավալի նմուշում x_k և y_i հատկանիշներ ունեցող անհատների թիվն է: **Երկու մուտքով** աղյուսակում, որը կոչվում է նաև **զուգակցության աղյուսակ**, ներկայացված են հատկանիշների **համատեղ բացարձակ հաճախությունները**՝

$X \backslash Y$	y_1	y_2	$\dots$	y_i	$\dots$	y_I	
x_1	n_{11}	n_{12}	$\dots$	n_{1i}	$\dots$	n_{1I}	$n_{1.}$
x_2	n_{21}	n_{22}	$\dots$	n_{2i}	$\dots$	n_{2I}	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{ki}	$\dots$	n_{kI}	$n_{k.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_K	n_{K1}	n_{K2}	$\dots$	n_{Ki}	$\dots$	n_{KI}	$n_{K.}$
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.i}$	$\dots$	$n_{.I}$	N

Գումարելով ըստ տողերի, այնուհետև՝ ըստ սյուների, ստանում ենք, համապատասխանաբար, Y և X հատկանիշների **մասնավոր** բացարձակ հաճախությունները՝

$$n_{.k} = \sum_i n_{ik}, \quad k = \overline{1,K}, \quad n_{i.} = \sum_k n_{ik}, \quad i = \overline{1,I}:$$

Բաժանելով աղյուսակի n_{ik} տարրերը $n_{.i}$-ի՝ ստանում ենք x_k արժեքների **պայմանական** հարաբերական **հաճախությունները** ըստ y_i-ի արժեքի՝ $f_{k|i}$, $k = \overline{1,K}$, $i = \overline{1,I}$: Նման ձևով ստանում ենք y_i արժեքի պայմանական հաճախությունն ըստ x_k-ի՝

$$f_{i|k} = \frac{n_{ik}}{n_{.k}}, \; k = \overline{1,K}, \; i = \overline{1,I}:$$

Հասարակական կարծիքն ուսումնասիրող կազմակերպությունը հարցում անցկացնելիս չի սահմանափակվում մեկ հարց տալով: Սովորաբար առաջարկվում են բազմաթիվ հարցեր՝ ուսումնասիրվող երևույթը ավելի լրիվ հասկանալու համար:

<u>*Օրինակ*</u> 14: Ենթադրենք, որ 20 տղամարդ պատասխանել են 2 հարցի՝ նրանք գերադասում են սու՞րճ (Ս), թե՞ թեյ (Թ), և արդյոք նրանք նախընտրում են կոնյա՞կ (Կ), թե՞ գարեջուր (Գ): Դիցուք պատասխաններն են՝ (Ս, Կ) (Ս, Գ) (Թ, Կ) (Ս,Կ) (Ս, Կ) (Թ, Գ) (Ս, Կ) (Թ, Գ) (Թ, Գ) (Ս, Կ) (Ս, Գ) (Ս, Կ) (Ս,Գ) (Թ, Գ) (Ս, Գ) (Ս, Կ) (Ս, Գ) (Թ, Կ) (Ս, Կ) (Թ, Գ) : Ամփոփել արդյունքները զուգակցման աղյուսակում:

Լուծում:

Նախընտրում են	Սուրճ	Թեյ	
Կոնյակ	8	2	10
Գարեջուր	5	5	10
	13	7	20

Երկչափ նմուշի թվային բնութագրիչները հաշվարկվում են երկչափ ընդհատ պատահական վեկտորի համապատասխան բնութագրիչների հաշվարկմանը նման եղանակով: X-ի և Y-ի մասնատեղ բնութագրիչներին ավելանում է նրանց փոխադարձ կապվածությունը բնութագրող համահարաբերակցության նմուշային $\hat{\rho}_{xy}$ գործակիցը:

Այնպես, ինչպես (5) և (6) բանաձևերը հավասարազոր են նմուշի միջինը հաշվարկելու համար (միայն հարկավոր է տարբերել, որ x_n-ով, $n = \overline{1,N}$, նշանակված է n-րդ դիտված արժեքը, իսկ x_k-ով՝ հատկանիշի k-րդ հնարավոր արժեքը, որը հանդիպել է $n_{.k}$ անգամ, $k = \overline{1,K}$), հետևյալ բանաձևերով կարելի է հաշվարկել հարաբերակցության նմուշային գործակիցը՝

$$\hat{\rho}_{XY} = \frac{\sum\limits_n (x_n - \bar{x})(y_n - \bar{y})}{\left(\sum\limits_n (x_n - \bar{x})^2 \sum\limits_n (y_n - \bar{y})^2\right)^{1/2}} =$$

$$= \frac{\sum\limits_n x_n y_n - (\sum\limits_n x_n \sum\limits_n y_n)/N}{[\sum\limits_n x_n^2 - (\sum\limits_n x_n)^2/N]^{1/2}[\sum\limits_n y_n^2 - (\sum\limits_n y_n)^2/N]^{1/2}} = \tag{9}$$

$$= \frac{\sum\limits_{ik} n_{ik}(x_k - \sum\limits_k n_{.k}x_k/N)(y_i - \sum\limits_i n_{i.}y_i/N)}{[\sum\limits_k n_{.k}(x_k - \sum\limits_k n_{.k}x_k/N)^2 \sum\limits_i n_{i.}(y_i - \sum\limits_i n_{i.}y_i/N)^2]^{1/2}}:$$

Երկչափ հանույթի մասին պատկերացում է տալիս **ցրվածության գծապատկերը**, որը ներկայացնում է նմուշի տվյալները հարթության կետերի միջոցով՝ որոշակի դեկարտյան [illegible]առանցքների համակարգում:

Օրինակ 15: Հաշվարկենք աղյուսակում ներկայացված տվյալների մասնատեղ նմուշային միջիններն, ցրվածքները, հարաբերակցության գործակիցը: Կառուցենք ցրվածության գծապատկերը (տես նկար 8):

x	y	x	y	x	y	x	y	x	y
8.35	3.50	10.50	6.00	11.35	9.50	12.15	6.00	12.85	9.50
8.74	1.49	10.75	2.50	11.50	6.00	12.25	8.05	13.15	9.02
9.25	6.40	10.76	5.74	11.50	9.00	12.35	5.01	13.25	6.49
9.50	4.50	11.00	8.50	11.62	8.50	12.50	7.03	13.26	10.50
9.75	5.00	11.00	5.26	11.75	10.00	12.76	7.53	13.40	7.51
10.24	7.00	11.25	8.00	12.00	9.00	12.85	6.01	13.50	10.00
13.65	9.50	14.50	10.00	13.75	8.51	14.75	12.00	14.00	11.00
15.25	12.50	14.23	8.40	16.00	11.50	14.26	10.00	16.00	13.00
14.51	9.50	16.25	12.00						

Լուծում: Պետք է հաշվարկենք x-ը , $\hat{\sigma}^2(\mathrm{x})$-ը, $\overline{\mathrm{y}}$-ը $\hat{\sigma}^2(\mathrm{y})$-ը, $\hat{\rho}_{\mathrm{xy}}$-ը: Սկզբից գտնենք անհրաժեշտ գումարները՝

$$\sum_{n=1}^{42} x_n = 522.23, \quad \sum_{n=1}^{42} y_n = 336.41,$$

$$\sum_{n=1}^{42} x_n^2 = 6652.25, \quad \sum_{n=1}^{42} y_n^2 = 2987.80,$$

$$\sum_{n=1}^{N} x_n y_n = 4358.626 :$$

Այնուհետև բաժանելով 42-ի, կստանանք՝ $\overline{\mathrm{x}} = 12.434$, $\overline{\mathrm{y}} = 8.011$:

Երկրորդ կենտրոնական մոմենտները գտնելու համար հաշվենք՝

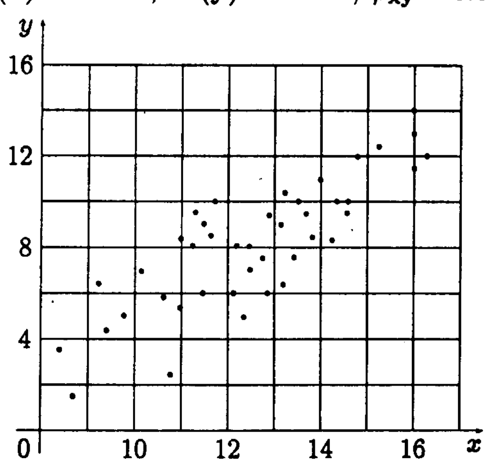
$$\sum_{n=1}^{N} x_n^2 - \sum_{n=1}^{N} x_n^2/N = 6652.25 - 522.23^2/42 \approx 158.8182,$$

$$\sum_{n=1}^{N} y_n^2 - (\sum_{n=1}^{N} y_n)^2/N = 2987.80 - 336.41^2/42 \approx 292.5958,$$

$$\sum_{n=1}^{N} x_n y_n - (\sum_{n=1}^{N} x_n)(\sum_{n=1}^{N} y_n)/N = 4358.626 - (522.23 \times 336.41)/42 \approx 175.1912 :$$

Վերջնականապես, օգտվելով (9) բանաձևերից երկրորդից, բաժանելով 42-ի, կստանանք՝

$$\hat{\sigma}^2(\mathrm{x}) \approx 3.7814, \quad \hat{\sigma}^2(\mathrm{y}) \approx 6.9666, \quad \hat{\rho}_{\mathrm{xy}} \approx 0.813 :$$



Նկար 8: Օրինակ 15-ի տվյալների համապատասխան ցրվածության գծապատկերը:

Գոյություն ունեն երկու և ավելի հատկանիշների համատեղ հաճախությունների բաշխումների պատկերավոր ներկայացման բավականին սրամիտ եղանակներ, որոնք սակայն այստեղ չեն շարադրվի՝ գրքի ծավալի սահմանափակության պատճառով: Ընթերցողը դրանց կարող է ծանոթանալ, օրինակ, Ուպտոնի և Կուկի (Upton, Cook) գրքում (տե՛ս գրականության ցանկը):

Գլուխ 7
Բաշխման բնութագրիչների վիճակագրական գնահատում

Որևէ երևույթի ամեն մի լուրջ վիճակագրական ուսումնասիրություն պահանջում է հատուկ գաղափարական գործիքներ, տեսություն, որոնք հետազոտողին հնարավորություն են տալիս փաստերի անհամար բազմությունից ընտրել այն, որը համապատասխանում է որոշ մոդելի և, հետևաբար, ենթարկվում է համակարգման:

Վասիլի Լեոնտև

Չկա ոչինչ ավելի հեշտ, քան նոր գնահատիչ հորինելը:

Ռոնալդ Ֆիշեր

7.1. Վիճակագրական մոդելների մասին

Առօրյա գործընթացները առաջադրում են անորոշության պայմաններում վճիռներ կայացնելու բազմապիսի իրադրություններ: Քննարկենք, օրինակ, հացթուխի խնդիրը, որը պետք է ամեն օր որոշի թխվելիք հացի քանակը, որպեսզի բավարարի հաճախորդների փոփոխվող պահանջարկը: Նա կարող է համարել, որ ամբողջ չվաճառված հացը կորուստ է և, հետևաբար, պատրաստել միայն ստույգ վաճառվող քանակ: Հակառակը, նա կարող է ցանկանալ բավարարել իր բոլոր հաճախորդներին և արտադրել մի փոքր ավելի՝ գուցե և ունենալով չվաճառված մնացորդի վտանգ: Ո՞րն է նրա համար լավագույն վճիռը: Ինչպե՞ս դասակարգել նման վիճակներում հնարավոր որոշումները: Այդպիսի հարցերի պատասխանները տալիս է **վճիռներ կայացնելու վիճակագրական տեսությունը**: Չանդրադառնալով այդ տեսության ընդհանուր ձևակերպումներին՝ անցնենք վճիռներին նախորդող վիճակագրական խնդիրներին: Ինչպես նշել ենք, ճիշտ վճիռներ կայացնելու համար մաթեմատիկական վիճակագրությունում մշակվում են եղանակներ, որոնք իրականացնում են ուսումնասիրվող հանույթի բնութագրիչների անհայտ արժեքների գնահատումը, վիճակագրական վարկածների ստուգումը, երևույթների տարբեր հատկանիշների միջև կապի հետազոտությունը: Միաժամանակ ուսումնասիրվում են այդ և մի շարք այլ վերլուծական եղանակների ավելի լայն կիրառման հնարավորությունները, սահմանափակումները և այլ առանձնահատկությունները: Գրքի ներկա և հաջորդ գլուխները նվիրված են թվարկված եղանակների մի մասին:

Ընդհանուր առմամբ, գիտական գործընթացները բաղկացած են ուսումնասիրվող տարբեր երևույթների, օբյեկտների **մոդելների** կառուցումից ու դրանց հետազոտությունից: Տեսությունները, փաստորեն, վերացական մոդելներ են, իսկ կիրառական հետազոտություններում կառուցվում են նյութական մոդելներ:

> **Մոդելը** երևույթի կամ ընթացքի՝ ուսումնասիրման տեսակետից հիմնական, էական կողմերը, հատկությունները արտապատկերող հանգույցն է, որն անտեսում է տվյալ տեսակետից ոչ էական դրսևորումները:

Մոդելի ստեղծման գործընթացը կոչվում է **մոդելավորում**: Մոդելը պետք է այնպես հաշվի առնի բոլոր կարևոր օրինաչափությունները և զարգացման պայմանները, շրջապատի երևույթների հետ փոխադարձ կապերը, որպեսզի դրա միջոցով հնարավոր լինի կատարել փորձարկումներ, որոնց նպատակն է մոդելավորման օբյեկտի վարքի որոշումը տարբեր (հաճախ իրականում չդիտարկվող) պայմաններում:

Հիշենք, որպես կարևոր օրինակ, որ պատահական մեծությունը պատահական ելքերով փորձի մաթեմատիկական մոդելն է:

Տնտեսագիտական երևույթներն ու ընթացքները ուսումնասիրվում են տնտեսա-մաթեմատիկական և տնտեսա-վիճակագրական մոդելների կառուցման միջոցով:

Տնտեսա-մաթեմատիկական մոդելն ընդհանրացնում է վերլուծության ենթակա օբյեկտի մասին էական քանակական և որակական տեղեկությունները, այն հաշվողական փորձարկումներ անցկացնելու միջոց է:

Տնտեսա-վիճակագրական մոդելը որոշակի տնտեսական օբյեկտը, ընթացքը կամ երևույթը նկարագրող մաթեմատիկական հարաբերությունների համակարգ է, որի պարամետրերը որոշվում (գնահատվում) են վիճակագրական եղանակներով՝ փաստացի տվյալների հիման վրա: Մոդելի կառուցման ընթացքը բաղկացած է փոխադարձաբար կապված երկու փուլից՝ մոդելի հարաբերությունների ընդհանուր տեսքի ու դրանում մասնակցող փոփոխականների որոշումից և օբյեկտի փորձարկումներից ստացված տվյալների հիման վրա պարամետրերի արժեքների գնահատումից: Տնտեսա-վիճակագրական մոդելի կազմվածքը և տեսքը որոշվում են մոդելավորվող օբյեկտի առանձնահատկություններով, հետազոտողի տեսական պատկերացումներով, հետազոտության նպատակով և տվյալների մշակման եղանակով:

Փորձարկումը N փորձերի իրականացում է, որոնց արդյունքները մաթեմատիկորեն նկարագրվում են X_n, $n = \overline{1,N}$ պատահական մեծություններով: Դիտված N պատահական մեծությունների միահամռությունը կոչում ենք **նմուշ**՝ $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, X_n-երը **նմուշի տարրերն** են, իսկ N-ը՝ **նմուշի ծավալը**: X_n-ի արժեքները $\mathcal{X}$ բազմությունից են: **Փորձարկման որոշակի արդյունքը** նշանակվում է $\mathbf{x} = (x_1, \ldots, x_N)$, և նույնպես կոչվում է նմուշ, իսկ բոլոր հնարավոր արդյունքների բազմությունը՝ $\mathcal{X}^N = \{\mathbf{x}\}$, կոչվում է **նմուշային տարածություն**: Երբ $\mathbf{X}$ նմուշի հավանականությունների բաշխման օրենքը (բաշխման ֆունկցիան՝ $F_{\mathbf{X}}(x_1, \ldots, x_N) = \mathbf{P}(X_1 < x_1, \ldots, X_N < x_N)$), անհայտ է, և հայտնի է միայն հնարավոր բաշխումների ընտանիքը՝ $\mathcal{F} = \{F(x_1, \ldots, x_N)\}$, որին պատկանում է $\mathbf{X}$ նմուշի $F_{\mathbf{X}}$ բաշխումը, ապա ասում են, որ դիտարկվում է $(\mathbf{X}, \mathcal{F})$ **վիճակագրական մոդելը**: Վիճակագրի նպատակն է՝ տրված $\mathcal{F}$ մոդելի սահմաններում ճշտել, երևան բերել իրական $F_{\mathbf{X}}$ բաշխման տարբեր հատկությունները՝ գտնել որոշակի պարամետրի գնահատականները, գուցե և ուղղակի $F_{\mathbf{X}}$-ի հնարավորին չափ ճշգրիտ արտահայտությունը:

Երբ փորձարկումը կազմող հաջորդական փորձերն իրարից անկախ են և անցկացվում են միևնույն պայմաններում, այսինքն՝ նմուշի $X_1, \ldots, X_N$ տարրերը միանման բաշխված, իրարից անկախ պատահական մեծություններ են, ապա մոդելը բնորոշվում է դրանց ընդհանուր միաչափ F_X բաշխումով: Այսպիսով նկարագրել մոդելը նշանակում է տալ հնարավոր բաշխման ֆունկցիաների դասը, որը նշանակվում է $\mathcal{F} = \{F_X\}$: Եթե $\mathcal{F} = \{F_X(x,\theta), \quad \theta \in \Theta\}$, այսինքն թույլատրելի բաշխման ֆունկցիաները կախված են մեկ կամ մի քանի պարամետրերի արժեքների ճշտությամբ, ապա **մոդելը** կոչվում է **պարամետրական**, իսկ Θ բազմությունը՝ **պարամետրական բազմություն**:

Պարամետրական մոդելի դեպքում θ-ից կախված հավանականությունների բաշխումը նշանակում են P_θ, իսկ X-ի որևէ $T(\mathrm{X})$ վիճականու մոմենտները նշանակում են $\mathrm{E}_\theta T(\mathrm{X})$, $\mathrm{D}_\theta T(\mathrm{X})$ և այլն:

Բոլոր վիճակագրական **եղանակները** կարելի է բաժանել երկու մեծ խմբի՝ **պարամետրական և ոչ պարամետրական**: Խնդրի լուծման համար **վիճակագրական պարամետրական եղանակի** օգնությանը դիմում են այն դեպքում, երբ հետազոտվող երևույթը նկարագրող պատահական մեծության վերաբերյալ կան նախնական տեղեկություններ, որոնց հիման վրա կարելի է բնորոշել դրա բաշխման օրենքի տեսակը (երկանդամային, նորմալ, ցուցչային և այլն) և համապատասխան պարամետրերի փոփոխման տիրույթները: Իսկ եթե չկա նախնական պատկերացում երևույթը բնութագրող բաշխման օրենքի և, առավել ևս, պարամետրերի արժեքների մասին, ու առաջադրված խնդրում չի պահանջվում դրանց մասին դատողություններ անել, ապա առաջ են քաշվում վերջին տասնամյակում ակտիվորեն մշակվող **վիճակագրական ոչ պարամետրական եղանակները**: Այս գրքում հիմնականում կներկայացվեն դասական, պարամետրական եղանակները:

Մաթեմատիկական և կիրառական վիճակագրության մի մեծ բաժին են կազմում նմուշային տվյալների հիման վրա պատահական երևույթի կամ դրա մոդելի **պարամետրերի** անհայտ արժեքների **գնահատման եղանակները**:

Գոյություն ունի այդ եղանակների երկու տեսակ՝ կետային և միջակայքային: **Կետային գնահատման եղանակների** նպատակն է գտնել մի թիվ, որը «մոտ» է ուսումնասիրվող բնութագրիչի անհայտ արժեքին: **Միջակայքային գնահատականը** հնարավորին չափ իրար մոտ այն սահմաններն են, որոնց մեջ է գտնվում պարամետրի անհայտ արժեքը՝ նախօրոք տրված մեկին մոտ հավանականությամբ: Այլ կերպ ասած՝ միջակայքի եզրերի հեռավորությունը գնահատման «ճշգրտությունն» է, իսկ նշված հավանականությունը ցույց է տալիս, որ միջակայքը միայն հազվադեպ կարող է չպարունակել պարամետրի արժեքը: Այդ եղանակներին է նվիրված հաջորդ՝ 8-րդ գլուխը, իսկ այս գլխում կուսումնասիրենք կետային գնահատականների կառուցման սկզբունքներն ու եղանակները:

7.2. Գնատուները և դրանց ներկայացվող պահանջները

Վիճակագրության ամենատարածված խնդիրներից մեկը ուսումնասիրվող հանույթի մի որոշակի θ պարամետրի արժեքը որոշելն է: Նախօրոք հայտնի է, որ $\theta \in \Theta \subset \mathcal{R}^K$: Սովորաբար պարամետրի հնարավոր արժեքների Θ բազմությունը վերջավորչափանի է՝ K-ն բնական թիվ է: Սակայն պարամետրական տարածությունը կարող է լինել և անվերջչափանի, օրինակ, եթե որպես θ պարամետր դիտարկենք բաշխման ֆունկցիան, իսկ Θ-ն լինի բոլոր անընդհատ բաշխման ֆունկցիաների դասը:

θ-ի անհայտ արժեքը գնահատելու համար կատարվում են N անկախ փորձեր, ստացվում է $\mathrm{X} = (X_1, \ldots, X_N)$ պատահական նմուշի որևէ $\mathrm{x} = (x_1, \ldots, x_N)$ իրացումը, դրա հետ կատարվում են որոշակի գործողություններ, և արդյունքում ստացված $\hat{\theta}$ արժեքը համարում են θ-ի գնահատական:

Ներմուծենք հետևյալ հիմնական գաղափարները:

Հիշեցնենք, որ նախքան փորձարկման իրականացումը մենք դիտարկում ենք պատահական $\mathrm{X} = (X_1, \cdots, X_N)$ նմուշը՝ որպես N հատ անկախ, միանման բաշխված պատահական մեծությունների վեկտոր: Հետևաբար պատահական վեկտորից $\varphi(\mathrm{X})$ ֆունկցիան նույնպես պատահական մեծություն է, և իմաստ ունի դիտարկել դրա բնութագրիչները, մասնավորապես, սպասելին ու ցրվածքը:

Նմուշի կամայական ֆունկցիա կոչվում է **վիճականի**†, իսկ եթե այդ $\varphi(\mathbf{X})$ ֆունկցիան հատուկ ընտրված է որոշակի պարամետրի գնահատման համար, ապա այն կոչվում է համապատասխան **պարամետրի գնատ**: Տվյալ հանուրը բնութագրող **թվային** θ **պարամետրի գնատ** կոչվում է նմուշից $\hat{\theta}(\mathbf{X})$ ֆունկցիան, որն ընտրված է այնպես, որ նրա արժեքները որոշ իմաստով մոտ լինեն θ-ին: Երբ $\mathbf{X}$ նմուշի փորձնականորեն իրականացած արժեքը x-ն է, ապա $\hat{\theta}(\mathbf{X})$-ը տալիս է մեկ թիվ՝ $\hat{\theta}(\mathrm{x}) = \hat{\theta}$, որը **պարամետրի** տվյալ դեպքում ստացված **գնահատականն** է‡:

Այլ կերպ ասած, գնահատականը գնատի մեկ փորձարկման արդյունքում ստացված որոշակի արժեքն է:

Ինչպե՞ս կառուցել «լավ» գնատուներ: Մինչ այդ հարցին պատասխանելը, նախ պետք է սահմանել «լավ» հատկանիշը:

Գնատուներից պահանջվող կարևորագույն հատկություններն են՝ **անշեղությունը, ունակությունը** և **արդյունավետությունը** (էֆեկտիվությունը), որոնց առկայությունն ապահովում է պարամետրի իրական արժեքին վերը նշված մոտիկությունը:

Գնատուն կոչվում է **անշեղ**, եթե դրա սպասելին հավասար է պարամետրի իրական θ արժեքին՝

$$\mathbf{E}(\hat{\theta}(\mathbf{X})) = \theta : \qquad (1)$$

Եթե գնատուի սպասելին հավասար չէ պարամետրի արժեքին, ապա ասում են, որ **գնատուն շեղ** է: Գնատուի արժեքները՝ նմուշի պատահական լինելու հետևանքով, կարող են քիչ թե շատ տարբերվել պարամետրի θ իրական արժեքից. անշեղությունը նշանակում է, որ այդ պատահական տատանումները տեղի են ունենում պարամետրի իրական արժեքի շուրջը և միջինում հավասար են դրան:

Բնական է նաև պահանջել, որ նմուշի N ծավալի մեծացմանը զուգընթաց գնահատականներն ավելի ու ավելի մոտենան որոնելի θ արժեքին, այլ կերպ ասած՝ տեղեկությունների ավելացման արդյունքում գնատուն պետք է ավելի ճշգրիտ դառնա, ավելի քիչ շեղվի θ-ից:

$\hat{\theta}(\mathbf{X})$ գնատուն կոչվում է **ունակ** (կամ **զուգամետ**), եթե նմուշի N ծավալի մեծացման հետ այն ըստ հավանականության ձգտում է θ պարամետրի որոնելի արժեքին՝

$$\hat{\theta}(\mathbf{X}) \xrightarrow[N\to\infty]{P} \theta :$$

Վերջապես, երրորդ պահանջը վերաբերում է գնատուների ցրվածքի փոքր լինելուն, այսինքն՝ միջին քառակուսային իմաստով փոքր շեղումներ ունենալուն:

Օրինակ, եթե ունենք θ պարամետրի երկու անշեղ գնատու՝ $\hat{\theta}_1 = \phi_1(\mathbf{X})$ և $\hat{\theta}_2 = \phi_2(\mathbf{X})$, ապա բնական է գերադասել ավելի փոքր ցրվածք ունեցողը. այն համարվում է **ավելի արդյունավետ**:

---

† Ռուսերեն՝ «статистика», որը սակայն համանուն է ամբողջ առարկայի անվանմանը:

‡ Ցավոք «գնատու» և «գնահատական» տերմինների համար ռուսերենում հազվադեպ բացառությամբ օգտագործում են միևնույն «оценка» բառը: Հայերենում (ինչպես և եվրոպական մի շարք լեզուներում) [illegible]ինների առկայությունը նպաստում է այդ գաղափարների տարբերակմանը և ճիշտ յուրացմանը:

$\hat{\theta}_1(\mathbf{X})$ գնահատումն կոչվում է **արդյունավետ** (էֆեկտիվ), եթե գնահատումների որոշակի $\hat{\Theta}$ դասում այն ամենաճշգրիտն է՝ ունի փոքրագույն ցրվածք՝
$$\mathbf{D}(\hat{\theta}_1(\mathbf{X})) = \min_{\hat{\theta}(\mathbf{X})\in\hat{\Theta}} \mathbf{D}(\hat{\theta}(\mathbf{X})) :$$

Որոշ դեպքերում θ պարամետրի տվյալ $\hat{\Theta}$ դասի գնահատումների ցրվածքների փոքրագույն արժեքը հնարավոր է իմանալ նախօրոք: Դրա հետ համեմատելով որոշակի գնահատումի ցրվածքի մեծությունը, կարող ենք եզրակացնել՝ այն արդյունավետ է, թե ոչ:

Ունակ լինելու հատկությունը պարտադիր է գնահատումի համար: Եթե գնահատումն շեղ է, ապա հնարավոր է այն ուղղել, որպեսզի դառնա անշեղ: Երբեմն դիտարկվում է հետևյալ՝ ավելի թույլ պահանջը:

Եթե փորձերի N թիվը մեծացնելիս գնահատումի սպասելիի շեղումը պարամետրի իրական արժեքից ձգտում է զրոյի՝
$$\mathbf{E}(\hat{\theta}(\mathbf{X}))\xrightarrow[N\to\infty]{}\theta,$$
ապա գնահատումն կոչվում է **ասիմպտոտորեն անշեղ**:

<u>*Օրինակ*</u> 1: *Նմուշային միջինը որպես սպասելիի գնահատում*: Երբ նմուշը տեսականորեն դիտարկվում է միջ փորձարկումները՝ որպես անկախ միանման բաշխված պատահական մեծությունների վեկտոր՝ $\mathbf{X} = (X_1, \ldots, X_N)$, նմուշային միջինը՝
$$\overline{\mathbf{X}} = \frac{1}{N}\sum_{n=1}^{N} X_n,$$
պատահական մեծություն է: Ըստ մեծ թվերի օրենքի (տե՛ս գլուխ 4) անկախ, միանման բաշխված, ցրվածք ունեցող X_n պատահական մեծությունների միջինը ձգտում է ըստ հավանականության դրանց ընդհանուր միջինին (նշանակենք այն $\mathbf{E}(X_n) = \mathbf{E}(X) = m$)՝
$$\overline{\mathbf{X}}\xrightarrow[N\to\infty]{P} m :$$
Այսպիսով,

եթե X պատահական մեծությունն ունի σ^2 ցրվածք, ապա նմուշային միջինը՝ $\overline{\mathbf{X}}$-ը, տեսական սպասելիի՝ m-ի համար ունակ գնահատում է: $\hat{m}(\mathbf{X}) = \overline{\mathbf{X}}$ գնահատումն նաև անշեղ է և արդյունավետ է գծային անշեղ գնահատումների $\hat{\Theta}$ դասում, ընդ որում դրա ցրվածքը՝ $\mathbf{D}(\hat{m}(\mathbf{X})) = \sigma^2/N$:

Իսկապես՝
$$\mathbf{E}(\hat{m}(\mathbf{X})) = \mathbf{E}(\overline{\mathbf{X}}) = \mathbf{E}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}(X_n) = m :$$
Գծային գնահատումները ունեն հետևյալ տեսքը՝ $\hat{m}(\mathbf{X}) = \sum_n c_n X_n$: Անշեղ գնահատումն պետք է բավարարի (1) պայմանին՝
$$\mathbf{E}(\hat{m}(\mathbf{X})) = \mathbf{E}(\sum_n c_n X_n) = \sum_n c_n \mathbf{E}(X_n) = m\sum_n c_n,$$
որտեղից անհրաժեշտաբար $\sum_n c_n = 1$: Հաշվենք $\hat{m}(\mathbf{X})$-ի ցրվածքը՝
$$\mathbf{D}(\hat{m}(\mathbf{X})) = \mathbf{D}(\sum_n c_n X_n) = \sum_n c_n^2 \mathbf{D}(X_n) = \sigma^2\sum_n c_n^2 :$$
Անշեղ գծային գնահատումների ցրվածքի փոքրագույն արժեքը գտնելու համար հարկավոր է լուծել պայմանական էքստրեմումի հետևյալ խնդիրը՝
$$\min\sum_n c_n^2, \quad \text{երբ} \sum_n c_n = 1 :$$

Այս խնդիրը բերվում է բացարձակ էքստրեմումի հետևյալ խնդրին՝

$$\min S(c_1,\ldots,c_N,\lambda) = \min(\sum_n c_n^2 - \lambda(\sum_n c_n - 1)),$$

որտեղ λ-ն Լագրանժի անորոշ գործակիցն է: Էքստրեմումի որոշման համար S-ի ածանցյալները հավասարեցնենք 0-ի՝

$$\frac{\partial S}{\partial c_n} = 2c_n - \lambda = 0,\ \ n = \overline{1,N},\ \ \frac{\partial S}{\partial \lambda} = \sum_n c_n - 1 = 0:$$

Լուծելով համակարգը, կստանանք՝ $c_n = \lambda/2$, $\lambda/2 = 1/N$, որտեղից $c_n = 1/N$, $n = \overline{1,N}$: Տեղադրելով c_n-երը. գտնում ենք, որ $\hat{m} = \sum_n X_n/N = \mathbf{\overline{X}}$: Նշենք, որ այդ դեպքում $\mathbf{D}(\hat{m})$ ցրվածքը կլինի փոքրագույնը և հավասար $\mathbf{D}(X_n)/N = \sigma^2/N$:

Այս արդյունքից մասնավոր դեպքում ստանում ենք կարևոր և հաճախ օգտագործվող հետևանք:

Բեռնուլիի N փորձերում $\mathcal{A}$ պատահույթի $\hat{p} = n_{\mathcal{A}}/N$ հաճախությունը անհայտ $p = \mathbf{P}(\mathcal{A})$ հավանականության անշեղ, ունակ և արդյունավետ գնատու է:

<u>*Օրինակ*</u> 2: *Անդարձ նմուշի միջինի ցրվածքը*: Անկախ (դարձով) նմուշահանման փոխարեն ժամանակավորապես դիտարկենք K տարրեր պարունակող համույթից անդարձ նմուշահանումը (տե՛ս գլուխ 1, վարժություն 20 և 6.1. ենթաբաժինը): Այդ դեպքում հաջորդական փորձերի արդյունքները՝ X_n, $n = \overline{1,N}$, $N \le K$, չեն լինի անկախ, քանի որ հաջորդ փորձում չեն մասնակցում արդեն հանված և գրանցված տարրերը: Կարելի է ցույց տալ, որ, այնուամենայնիվ, $\hat{m}(\mathbf{X}) = \mathbf{\overline{X}}$ անդարձ նմուշի միջինը m տեսական միջինի անշեղ, ունակ գնատու է, որի ցրվածքը հավասար է՝

$$\mathbf{D}(\mathbf{\overline{X}}) = \frac{\sigma^2}{N}\left(\frac{K-N}{K-1}\right):$$

«Անդարձության գործակիցը»՝ $(K-N)/(K-1)$-ը, փոքր է 1-ից և ձգտում է 0-ի, երբ N-ը մոտենում է K-ին, իսկ երբ $N = K$, ապա ուղղակի $\hat{m} = m$: Սակայն, եթե $K \gg N$, այսինքն K-ն շատ մեծ է N-ից, ապա $K - N \approx K - 1$, և անկախ ու անդարձ նմուշների ցրվածքները կարելի է համարել իրար հավասար. անդարձ և պատահական նմուշահանման տարբերությունը աննշան է դառնում:

<u>*Օրինակ*</u> 3: *Տեսական ցրվածքի կետային գնատուներ*: Բնական է փորձել տիեզերքի տեսական ցրվածքի՝ $\sigma^2 = \mathbf{D}(X)$, գնատուի դերում դիտարկել նմուշային ցրվածքը՝

$$\hat{\sigma}^2(\mathbf{X}) = \hat{\mathbf{D}}(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N}(X_n - \mathbf{\overline{X}})^2:$$

Ցույց տանք, որ $\hat{\sigma}^2$-ն ունակ, բայց շեղ գնատու է: Իսկ ո՞րը կլինի անշեղ:

Լուծում: Գտնենք $\hat{\sigma}^2(\mathbf{X})$-ի սպասելին՝

$$\mathbf{E}(\hat{\sigma}^2(\mathbf{X})) = (1/N)\sum_n(\mathbf{E}(X_n^2) - 2\mathbf{E}(X_n\mathbf{\overline{X}}) + \mathbf{E}(\mathbf{\overline{X}})^2),$$

$$\mathbf{E}(X_n^2) = \mathbf{D}(X_n) + (\mathbf{E}(X_n))^2 = \sigma^2 + m^2,$$

$$\mathbf{E}(X_n\mathbf{\overline{X}}) = \mathbf{E}(X_n\cdot(1/N)\sum_i X_i) = (1/N)(\mathbf{E}(X_n^2) + \sum_{i\neq n}\mathbf{E}(X_n\mathbf{E}X_i)) = \sigma^2/N + m^2,$$

$$\mathbf{E}(\mathbf{\overline{X}}^2) = \mathbf{E}((1/N)\sum_n X_n\cdot(1/N)\sum_i X_i) = \frac{1}{N^2}(\sum_n \mathbf{E}(X_n^2) + \sum_n\sum_{i\neq n}\mathbf{E}(X_n\mathbf{E}X_i)) = \sigma^2/N + m^2:$$

Ուրեմն՝

$$\mathbf{E}(\hat{\sigma}^2(\mathbf{X})) = \sigma^2 - \sigma^2/N = \sigma^2(N-1)/N:$$

Տեսնում ենք, որ $\hat{\sigma}^2(\mathbf{X})$ գնատուն շեղ է, նրա սպասելին ունի $-\sigma^2/N$ շեղում որոնելի σ^2-ուց:

քանի որ $(N-1)/N$ գործակիցը ձգտում է 1-ի, երբ $N \to \infty$, $\hat{\sigma}^2(\mathbf{X})$ գնատուն ասիմպտոտորեն անշեղ է: σ^2-ու անշեղ գնատու է **նմուշային ուղղված ցրվածքը**՝

$$\hat{s}^2(\mathbf{X}) = \frac{N}{N-1}\hat{\sigma}^2(\mathbf{X}) = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \overline{\mathbf{X}})^2 :$$

Իսկապես, $\mathbf{E}\hat{s}^2(\mathbf{X}) = \mathbf{E}\hat{\sigma}^2(\mathbf{X})N/(N-1) = \sigma^2$:

Անցնենք $\hat{s}^2(\mathbf{X})$-ի ունակ լինելու ապացուցմանը: Որպեսզի գնատուն լինի ունակ, բավական է, որ նրա ցրվածքը ձգտի զրոյի, քանի որ Չեբիշևի անհավասարությունից կստանանք, որ, երբ $N \to \infty$, $\hat{s}^2(\mathbf{X}) \to \sigma^2$ ըստ հավանականության: Հաշվենք

$$\mathbf{D}(\hat{s}^2(\mathbf{X})) = \mathbf{E}(\hat{s}^4(\mathbf{X})) - (\mathbf{E}(\hat{s}^2(\mathbf{X})))^2 = \mathbf{E}(\hat{s}^4(\mathbf{X})) - \sigma^4,$$

$$\mathbf{E}(\hat{s}^4(\mathbf{X})) = \mathbf{E}\left[\frac{1}{(N-1)^2}\sum_{i,n}(X_n - \overline{\mathbf{X}})^2(X_i - \overline{\mathbf{X}})^2\right] =$$

$$= \frac{1}{(N-1)^2}\sum_{i,n}\Big[\mathbf{E}(X_n^2X_i^2) - 2\mathbf{E}(X_n^2X_i\overline{\mathbf{X}}) + \mathbf{E}(X_n^2\overline{\mathbf{X}}^2) + \mathbf{E}(X_n^2\overline{\mathbf{X}}^2) - 2\mathbf{E}(X_nX_i^2\overline{\mathbf{X}}) +$$

$$+ 4\mathbf{E}(X_nX_i\overline{\mathbf{X}}^2) - 2\mathbf{E}(X_n\overline{\mathbf{X}}^3) + \mathbf{E}(X_i^2\overline{\mathbf{X}}) - 2\mathbf{E}(X_i\overline{\mathbf{X}}^3) + \mathbf{E}(\overline{\mathbf{X}}^4)\Big] :$$

Քանի որ $\hat{s}^2(\mathbf{X})$-ի սահմանման մեջ մասնակցում են միայն $X_n - \overline{\mathbf{X}}$ տարբերությունները, կարող ենք համարել, որ X_n-երը կենտրոնացված են, ուստի $\mathbf{E}X_n = 0$: Հետևաբար՝

$$\mathbf{E}(X_n^2X_i^2) = \begin{cases} \mathbf{E}(X_n^4) = \mu_4, & n = i, \\ \mathbf{E}(X_n^2X_i^2) = \sigma^4, & n \neq i, \end{cases}$$

$$\mathbf{E}(X_n^2X_i\overline{\mathbf{X}}) = \frac{1}{N}\sum_r \mathbf{E}(X_n^2X_iX_r) = \begin{cases} \mu_4/N, & n = i, \\ \sigma^4/N, & n \neq i, \end{cases}$$

$$\mathbf{E}(X_n^2\overline{\mathbf{X}}^2) = \frac{1}{N^2}\sum_{r,l}\mathbf{E}(X_n^2X_rX_l) = \frac{1}{N^2}\sum_r \mathbf{E}(X_n^2X_r^2) = \frac{\mu_4}{N^2} + \frac{(N-1)}{N^2}\sigma^4,$$

$$\mathbf{E}(X_nX_i^2\overline{\mathbf{X}}) = \begin{cases} \mu_4/N, & n = i, \\ \sigma^4/N, & n \neq i, \end{cases}$$

$$\mathbf{E}(X_nX_i\overline{\mathbf{X}}^2) = \frac{1}{N^2}\sum_l \mathbf{E}(X_nX_iX_rX_l) = \begin{cases} \mu_4/N^2 + 3(N-1)\sigma^4/N^2, & n = i, \\ 2\sigma^4/N^2, & n \neq i, \end{cases}$$

$$\mathbf{E}(X_n\overline{\mathbf{X}}^3) = \frac{1}{N^3}\sum_{r,l,m}\mathbf{E}(X_nX_rX_lX_m) = \frac{1}{N^3}[\mu_4 + 3(N-1)\sigma^4],$$

$$\mathbf{E}(X_i^2\overline{\mathbf{X}}^2) = \mu_4/N^2 + (N-1)\sigma^4/N^2, \quad \mathbf{E}(X_i\overline{\mathbf{X}}^3) = [\mu_4 + 3(N-1)\sigma^4]/N^3,$$

$$\mathbf{E}(\overline{\mathbf{X}}^4) = \frac{1}{N^4}\sum_{r,l,m,s}\mathbf{E}(X_rX_lX_mX_s) = \frac{1}{N^3}[\mu_4 + 3(N-1)\sigma^4]:$$

Տեղադրելով ստացված արտահայտությունները, կստանանք՝

$$\mathbf{E}(\hat{s}^4(\mathbf{X})) = \mu_4/N + \sigma^4(N^2 - 2N + 3)/(N(N-1)),$$

$$\mathbf{D}(\hat{s}^2(\mathbf{X})) = [\mu_4 - (N-3)\sigma^4/(N-1)]/N,$$

որը, իսկապես, երբ $N \to \infty$, ձգտում է զրոյի, եթե μ_4 կենտրոնական մոմենտը գոյություն ունի: Այսպիսով, քանի որ $\hat{\sigma}^2(\mathbf{X})$-ու և $\hat{s}^2(\mathbf{X})$-ու սահմանները, երբ $N \to \infty$, իրար հավասար են,

եթե տեսական չորրորդ կենտրոնական մոմենտը՝ μ_4-ը, գոյություն ունի, ապա $\hat{\sigma}^2(\mathbf{X})$-ը և $\hat{s}^2(\mathbf{X})$-ը տեսական σ^2 ցրվածքի ունակ գնատուներ են:

Երբեմն, երբ բաշխման սպասելիի՝ m-ի արժեքը հայտնի է, հարկ է լինում դիտարկել ցրվածքի նմուշային մի այլ բնութագրիչ՝

$$\hat{s}_0^2(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N}(X_n - m)^2:$$

Դժվար չէ համոզվել, որ $\hat{s}_0^2(\mathbf{X})$-ը ցրվածքի անշեղ և ունակ գնատու է:

<u>Օրինակ</u> 4: *Նմուշային մոմենտները՝ տեսական մոմենտների գնատուներ*: Համաձայն 6.4 ենթաբաժնում տրված սահմանման, նմուշային սկզբնական մոմենտներն են՝

$$\hat{m}_r(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N} X_n^r, \quad r = 1, 2, \ldots:$$

Ցույց տանք, որ

> եթե հատկանիշի m_{2r} տեսական մոմենտը գոյություն ունի, ապա $\hat{m}_r(\mathbf{X})$ նմուշային մոմենտը համապատասխան m_r տեսական մոմենտի անշեղ, ունակ և գծային գնատուների դասում արդյունավետ գնատու է:

Լուծում: Դիտարկենք $Y = X^r$ պատահական մեծությունը: Դրա համար m_r և $\hat{m}_r(\mathbf{X})$ մոմենտները կլինեն առաջին կարգի, համապատասխանաբար, տեսական և նմուշային մոմենտներ: Համաձայն օրինակ 1-ի, եթե $\sigma^2(Y)$ ցրվածքը գոյություն ունի, ապա $\hat{m}_r(\mathbf{X})$ մոմենտը m_r-ի համար կլինի անշեղ, ունակ և գծային (ըստ $Y_k = X_n^r$-երի) գնատուների դասում արդյունավետ գնատու: Մնում է նկատել, որ $\sigma^2(Y) = \mu_{2r}(X) = m_{2r} - m_r^2$:

Քանի որ, ինչպես տեսանք, նմուշային սկզբնական մոմենտները լավ գնատուներ են համապատասխան տեսական մոմենտների համար, իմաստ ունի անհայտ տեսական մոմենտները՝ $m_1, m_2, \ldots$, մեծ N ծավալի նմուշի դեպքում փոխարինել (գնահատել) համապատասխան նմուշայիններով՝ $\hat{m}_1, \hat{m}_2, \ldots$:

Վերադառնանք θ պարամետրի անշեղ գնատուի արդյունավետության և ասիմպտոտական արդյունավետության գաղափարներին:

> N ծավալի նմուշի դեպքում $\hat{\theta}_1(\mathbf{X})$ գնատուի $e(\hat{\theta}_1(\mathbf{X}))$ **արդյունավետություն** կոչվում է անշեղ գնատուների $\hat{\Theta}$ դասում գնատուի ցրվածքի փոքրագույն արժեքի՝ $\min\limits_{\hat{\theta}(\mathbf{X})\in\hat{\Theta}} \mathrm{D}\hat{\theta}(\mathbf{X})$, հարաբերությունը դիտարկվող $\hat{\theta}_1$ գնատուի $\mathrm{D}\hat{\theta}_1(\mathbf{X})$ ցրվածքին:
>
> $$e(\hat{\theta}_1(\mathbf{X})) = \min_{\hat{\theta}(\mathbf{X})\in\hat{\Theta}} \mathrm{D}\hat{\theta}(\mathbf{X})/\mathrm{D}(\hat{\theta}_1(\mathbf{X}))$$

Եթե *անընդհատ* հայտանիշի տեսական բաշխման $f_X(x,\theta)$ խտությունը (կամ *ընդհատ* դեպքում $\mathrm{P}_X(x,\theta)$ հավանականությունը) բավարարում է ըստ θ-ի ածանցելի լինելու, որոշման $\mathcal{D}$ տիրույթի θ-ից անկախ լինելու և էլի մի քանի կանոնավորության (ռեգուլյարության) պայմաններին, ապա

> տեղի ունի **Ռաո-Կրամերի-Ֆրեշեի անհավասարությունը**՝
>
> $$\mathrm{D}[\hat{\theta}(\mathbf{X})] \geq 1/N\mathrm{I}_X(\theta) = \min_{\hat{\theta}(\mathbf{X})\in\hat{\Theta}} \mathrm{D}[\hat{\theta}(\mathbf{X})], \qquad (2)$$

որտեղ

> $$\mathbf{I}_X(\theta) = \mathbf{E}\left(\frac{\partial \ln f_X(X,\theta)}{\partial \theta}\right)^2$$
>
> **ինֆորմացիայի քանակն է ըստ Ֆիշերի**, որը մեկնաբանվում է որպես θ պարամետրի վերաբերյալ մեկ դիտումով ձեռք բերված տեղեկությունների քանակ:

Այստեղից՝

$$e(\hat{\theta}(\mathbf{X})) = (1/\mathbf{D}[\hat{\theta}(\mathbf{X})]) \min_{\hat{\theta}(\mathbf{X})\in\hat{\Theta}} \mathbf{D}\hat{\theta}(\mathbf{X}) = 1/N\mathbf{I}_X(\theta)\mathbf{D}[\hat{\theta}(\mathbf{X})]) \leq 1 :$$

Եթե $e(\hat{\theta}(\mathbf{X})) = 1$, ապա $\hat{\theta}(\mathbf{X})$ **գնահատումն արդյունավետ է**:

> $\hat{\theta}(\mathbf{X})$ **գնահատի ասիմպտոտական արդյունավետություն** կոչվում է
>
> $$e_0(\hat{\theta}(\mathbf{X})) = \lim_{N\to\infty} e(\hat{\theta}(\mathbf{X})) \leq 1 : \qquad (3)$$
>
> Երբ $e_0(\hat{\theta}(\mathbf{X})) = 1$, **գնահատումն ասիմպտոտորեն արդյունավետ է**:

Օրինակ 5: Եթե նորմալ բաշխված՝ $X \sim \mathcal{N}(m,\sigma^2)$ պատահական մեծության m սպասելին հայտնի է, ապա $\hat{s}^2(\mathbf{X})$-ը σ^2 ցրվածքի ասիմպտոտորեն արդյունավետ գնահատ է:

Լուծում: Ինչպես տեսանք օրինակ 3-ում, $\hat{s}^2(\mathbf{X})$-ը σ^2-ու անշեղ գնահատ է կամայական բաշխման դեպքում: Այժմ, հիշելով գլուխ 3-ից, որ նորմալ բաշխման դեպքում $\mu_4 = 3\sigma^4$, գտնենք՝

$$\mathbf{D}\hat{s}^2(\mathbf{X}) = \frac{1}{N}(3\sigma^4 - \frac{N-3}{N-1}\sigma^4) = \frac{2\sigma^4}{N-1}:$$

Հաշվենք Ֆիշերի $\mathbf{I}_X(\sigma^2)$ ինֆորմացիան նորմալ բաշխման դեպքում՝

$$\mathbf{I}_X(\sigma^2) = \mathbf{E}\left(\frac{\partial \ln f_X(x,m,\sigma^2)}{\partial \sigma^2}\right) = \int_{-\infty}^{\infty}\left(\frac{d \ln f_X(x,m,\sigma^2)}{d\sigma^2}\right)^2 f_X(x,m,\sigma^2)dx =$$

$$= \int_{-\infty}^{\infty}\left[\frac{(x-m)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right]^2 f_X(x,m,\sigma^2)dx = \frac{1}{2\sigma^4} :$$

Քանի որ m-ը հայտնի է, օգտվելով Ռաո-Կրամերի-Ֆրեշեի (2) անհավասարությունից, գտնենք $e(\hat{s}^2(\mathbf{X}))$ արդյունավետությունը՝ $e(\hat{s}^2(\mathbf{X})) = (N-1)/N < 1$: Իսկ (3) սահմանման համաձայն

$$e_0(\hat{s}^2(\mathbf{X})) = \lim_{N\to\infty} e(\hat{s}^2(\mathbf{X})) = 1,$$

այսինքն՝ $\hat{s}^2(\mathbf{X})$-ը ասիմպտոտորեն արդյունավետ գնահատ է:

Քննարկենք Ֆիշերի ինֆորմացիայի քանակի որոշ կարևոր հատկությունները: Լրացուցիչ ենթադրենք, որ X պատահական մեծության արժեքների $\mathcal{D}$ տիրույթում $f_X(x,\theta)$ խտությունը $\theta \in \mathcal{R}$ կամայական արժեքի դեպքում չի ընդունում զրո արժեք, կամայական x-ի համար ունի երկրորդ կարգի ածանցյալ ըստ θ-ի, և ինտեգրալի նշանի տակ կարելի է ածանցել ըստ θ-ի խտության ինտեգրալը $\int_{\mathcal{D}_X} f_X(x,\theta)dx$ (որը հավասար է 1-ի): Այդ դեպքում Ֆիշերի ինֆորմացիան կարելի է գրել նաև հետևյալ ձևով՝

$$\mathbf{I}_X(\theta) = \mathbf{E}\left(\frac{f'_X(X,\theta)}{f_X(X,\theta)}\right)^2 :$$

Արված ենթադրություններից հետևում է, որ

$$\mathbf{E}\left(\frac{f'_X(X,\theta)}{f_X(X,\theta)}\right) = 0 :$$

Այստեղից ստացվում է նաև, որ

$$\mathbf{I}_X(\theta) = \mathbf{D}\left(\frac{f'_X(X,\theta)}{f_X(X,\theta)}\right) = -\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln f_X(X,\theta)\right)^2 :$$

Կարևոր է նաև Ֆիշերի ինֆորմացիայի գումարականության (ադիտիվության) հատկությունը՝

Եթե X և Y անկախ պատահական մեծությունների բաշխումները կախված են θ պարամետրից, իսկ որոշման $\mathcal{D}(X)$ և $\mathcal{D}(Y)$ տիրույթները θ-ից կախված չեն, ապա (X,Y) վեկտորի Ֆիշերի $\mathbf{I}_{(X,Y)}(\theta)$ ինֆորմացիան հավասար է՝

$$\mathbf{I}_{(X,Y)}(\theta) = \mathbf{I}_X(\theta) + \mathbf{I}_Y(\theta) :$$

Այստեղից ստացվում է կարևոր հետևանք՝

$\mathbf{X}$ նմուշի Ֆիշերի $\mathbf{I}_\mathbf{X}(\theta)$ ինֆորմացիայի համար՝

$$\mathbf{I}_\mathbf{X}(\theta) = N\mathbf{I}_X(\theta) :$$

Եթե $T(\mathbf{X})$ վիճակագրին $\mathbf{X}$ նմուշի որևէ ֆունկցիա է, ապա միշտ

$$\mathbf{I}_{T(\mathbf{X})}(\theta) \le \mathbf{I}_\mathbf{X}(\theta),$$

այսինքն՝ նմուշի հետ կատարված գործողությունները չեն կարող նոր տեղեկություններ ստեղծել: Սակայն $T(\mathbf{X})$ վիճակագրին **սպառիչ** է, եթե վերջին անհավասարությունում տեղի ունի հավասարություն:

Երբ $\theta \in \mathcal{R}^K$, այսինքն՝ θ-ն վեկտորական պարամետր է, դիտարկվում է **ինֆորմացիայի** համաչափ **մատրիցը**՝

$$\left\{-\mathbf{E}\left(\frac{\partial^2}{\partial\theta_{k_1}\partial\theta_{k_2}}\ln f_X(X,\theta)\right),\ k_1,k_2=\overline{1,K}\right\} :$$

Կարևոր դեր ունի գնատուների ևս մեկ հատկություն, որը կոչվում է **բավարարություն**: Դրա էությունն այն է, որ գնատուն լրիվ օգտագործում է որոնելի պարամետրի մասին նմուշի մեջ պարունակված տեղեկությունները:

$T(\mathbf{X})$ **վիճակագրին** կոչվում է **բավարար**, եթե անհայտ θ պարամետրի մասին հայտնի $\hat{\theta}$-ի դեպքում *նմուշն այլևս լրացուցիչ տեղեկություն տալ չի կարող*:

Բնական է, որ նպատակահարմար է օգտվել բավարար գնատուներից, սակայն դրանց տեսությունը դուրս է մնում մեր ծրագրից: Հետաքրքրասեր ընթերցողին առաջարկում ենք դիմել գրականության ցանկում նշված գրքերին:

7.3. Գնատուների կառուցման ընդհանուր եղանակներ

Վերադառնանք վերը ձևակերպված հարցին՝ ինչպե՞ս գտնել գնատու որևէ պարամետրի արժեքը լավ գնահատելու համար: Գոյություն ունեն այդ նպատակին ծառայող մի քանի ընդհանուր մեթոդներ: Անցնենք դրանց շարադրմանը:

Մոմենտների եղանակը կետային գնատուների որոնման ուղիներից մեկն է: Դիցուք X պատահական մեծության բաշխման օրենքը, կախված K-չափանի $\theta = (\theta_1, ..., \theta_K)$ պարամետրից, տրված է, սակայն պարամետրի արժեքները հայտնի չեն: Այսինքն, հայտնի է $f_X(x,\theta)$ ֆունկցիան, որը X-ի անընդհատ լինելու դեպքում խտության ֆունկցիան է, իսկ երբ X-ը ընդհատ է՝ $\mathbf{P}_X(x,\theta)$ հավանականությունն է: Հարկավոր է գտնել θ պարամետրի՝ $\mathbf{X}=(X_1,\ldots,X_N)$ նմուշի վրա հիմնված $\hat{\theta}(\mathbf{X})$ գնահատումը:

Դիցուք X պատահական մեծությունն ունի K սկզբնական մոմենտներ՝ m_k, $k=\overline{1,K}$: Դրանցից յուրաքանչյուրը կարելի է արտահայտել θ պարամետրի բաղադրիչների միջոցով՝

$$m_k=g_k(\theta)=\begin{cases}\sum\limits_x x^k\mathbf{P}_X(x,\theta), & \text{երբ } X\text{-ը ընդհատ է,}\\ \int\limits_{-\infty}^{\infty} x^k f_X(x,\theta)dx, & \text{երբ } X\text{-ը անընդհատ է:}\end{cases}$$

Ստացվում է K անհայտով K հավասարումների համակարգ՝

$$m_k=g_k(\theta),\ k=\overline{1,K}: \tag{4}$$

Եթե համակարգը հնարավոր է լուծել $\theta_1,\ldots,\theta_K$ պարամետրերի նկատմամբ, ապա կստանանք՝

$$\theta_k=\hat{g}_k(m_1,\ldots,m_K),\ k=\overline{1,K}: \tag{5}$$

Ինչպես գիտենք օրինակ 4-ից, նմուշային մոմենտները տեսական մոմենտների ունակ գնահատումներ են:

Մոմենտների եղանակի գաղափարն այն է, որ (5) համակարգում տեսական մոմենտները փոխարինվում են նմուշայիններով՝ $$\hat{\theta}_k=\hat{g}_k(\hat{m}_1,\ldots,\hat{m}_K),\ k=\overline{1,K}: \quad (6)$$

Նմուշային տվյալներով հաշվելով $\hat{m}_k$ մոմենտների արժեքները (6) հավասարումներից, կստանանք θ պարամետրի բաղադրիչների գնահատականները: Իհարկե, պարտադիր չէ դիտարկել հենց առաջին K մոմենտները, հնարավոր է (4) համակարգում ընդգրկել K հատ կամայական սկզբնական կամ կենտրոնական մոմենտներ, որոնք տվյալ բաշխման դեպքում հնարավորին չափ պարզ ձևով են արտահայտվում $\theta_1,\ldots,\theta_K$ պարամետրերի միջոցով: Վիճակագրական պրակտիկայում, սակայն, չորրորդից բարձր կարգի մոմենտները հազվադեպ են օգտագործվում :

<u>*Օրինակ*</u> 6: X պատահական մեծությունը բաշխված է Պուասոնի օրենքով $\lambda>0$ պարամետրով՝ $X\sim\Pi(\lambda)$: Մոմենտների եղանակով գտնել λ պարամետրի գնահատումը:

Լուծում: ա) Կառուցենք գնահատում, ելնելով սկզբնական առաջին մոմենտից՝ m_1-ից, գիտենալով, որ Պուասոնի բաշխման դեպքում $m_1=\lambda$ (տե՛ս գլուխ 3), ստանում ենք՝

$$\hat{\lambda}=\bar{x}:$$

բ) Հայտնի է, որ Պուասոնի բաշխման ցրվածքը նույնպես հավասար է λ-ի (տե՛ս գլուխ 3), և քանի որ $\sigma^2=m_2-m_1^2$, ապա $m_2=\lambda+\lambda^2$: Այստեղից՝

$$\lambda=\frac{1}{2}(\sqrt{1+4m_2}-1),$$

և ըստ մոմենտների եղանակի կարող ենք վերցնել՝

$$\hat{\lambda}=\frac{1}{2}(\sqrt{1+4\hat{m}_2}-1):$$

Այս օրինակից երևում է, որ մոմենտների եղանակը կարող է մի պարամետրի համար առաջադրել մի քանի տարբեր գնահատումներ: Դրանցից, իհարկե, կարելի է ընտրել լավագույն հատկություններ ունեցողը:

Մոմենտների եղանակը գործնականում շատ հարմար է, քանի որ (6) հավասարման լուծումները ֆունկցիաներ են նմուշային մոմենտներից: Պարզ է, որ նույն սկզբունքը կիրառելի է նաև բազմաչափ պատահական մեծությունների դեպքում:

Նշենք նաև, որ մոմենտների եղանակի առաջադրած գնատումները հաճախ լինում են շեղ և ոչ արդյունավետ: Օրինակ 3-ում մենք տեսանք, որ $\hat{\sigma}^2 = \hat{\mu}_2$ գնատումն σ^2 տեսական ցրվածքի համար շեղ գնատու է, որը, սակայն, ասիմպտոտորեն անշեղ է:

Կարելի է ցույց տալ, որ բավականին ընդհանուր պայմանների դեպքում մոմենտների եղանակով ստացված գնատումն *ունակ* է, *ասիմպտոտորեն անշեղ* է և դրա բաշխման օրենքը *ասիմպտոտորեն նորմալ* է:

Երբեմն մոմենտների եղանակով ստացված գնատումները դիտարկվում են որպես առաջին մոտարկում, որից այլ եղանակներով կարելի է գտնել ավելի արդյունավետ գնատումներ:

Դիտարկենք մոմենտների եղանակի գնատուի ասիմպտոտական արդյունավետությունը: Քանի որ մեծ N-երի դեպքում կարելի է այդ գնատուն համարել անշեղ, ապա հաշվի առնելով Ռաո-Կրամեր-Ֆրեշեի (2) անհավասարությունը, և եթե

$$\mathbf{D}(\hat{\theta}(X)) = C^2(\theta)/N,$$

որտեղ $C^2(\theta)$-ն θ-ից կախված որևէ հաստատուն է, ստանում ենք՝

$$e_0(\hat{\theta}(X)) = \lim_{N\to\infty} e(\hat{\theta}(X)) = \lim_{N\to\infty} \frac{1}{N\mathbf{I}_X(\theta)\mathbf{D}(\hat{\theta}(X))} = \frac{1}{\mathbf{I}_X(\theta)C^2(\theta)}:$$

<u>*Օրինակ*</u> 7: Դիտարկենք $X \sim \mathcal{N}(m,\sigma^2)$ նորմալ բաշխված պատահական մեծությունը, որի m և σ պարամետրերը երկուսն էլ անհայտ են: Գտնենք դրանց գնատումները մոմենտների եղանակի միջոցով:

Լուծում: Գլուխ 3-ում ստացված նորմալ բաշխման մոմենտներն են՝ $m_1 = m$, $m_2 = m^2 + \sigma^2$: Լուծելով այս համակարգը m-ի և σ-ի նկատմամբ, կստանանք՝ $m = m_1$, $\sigma^2 = m_2 - m_1^2$: Հետևաբար, ըստ մոմենտների եղանակի, գնատումները կլինեն՝ $\hat{m} = \hat{m}_1$, $\hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2$:

Դիտարկենք նաև կիրառական

<u>*Օրինակ*</u> 8: Դիցուք մի քաղաքասի պատահականորեն վերցված 10 կրպակների մեկօրյա եկամուտը, որը ենթադրվում է նորմալ բաշխված, կազմել է 32500, 85650, 77520, 112900, 10200, 8250, 25030, 62850, 201700, 35980 դրամ: Գնահատել մոմենտի եղանակի օգնությամբ օրեկան եկամուտի սպասելին և ցրվածքը:

Լուծում: Համաձայն օրինակ 7-ի արդյունքի գտնենք

$$\hat{m} = \hat{m}_1 = (32500 + \cdots + 35980)/10 = 65258,$$

$$\hat{m}_2 = (32500^2 + \cdots + 35980^2)10 = 7387410920,$$

$$\hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2 = 7387410920 - 4258606564 = 3128804356, \qquad \hat{\sigma} \cong 55936,$$

Ստացվեց, որ տվյալ օրինակում ցրվածքը բավականին մեծ է, իսկ միջին քառակուսային շեղումը համեմատելի է միջինի հետ:

<u>*Օրինակ*</u> 9: Փորձարկումը տվել է $\mathbf{X}$ նմուշը: Մոմենտների եղանակով գտնել գամմա օրենքով բաշխված պատահական մեծության՝ $X \sim \Gamma(\alpha,\beta)$, անհայտ α և β պարամետրերի գնատումները:

Լուծում: Ինչպես նախորդ օրինակում, երկու պարամետր գնահատելու համար հարկավոր է երկու հավասարում: Ելնենք սպասելիի և ցրվածքի գնահատականներից՝ նմուշի $\overline{\mathbf{X}}$ միջինը՝ $\mathbf{E}X = m$ սպասելիի համար, $\hat{\sigma}^2(\mathbf{X})$-ը՝ $\mathbf{D}X$ ցրվածքի համար: Գամմա բաշխման դեպքում $m_1 = (\alpha+1)\beta$, $m_2 - m_1^2 = (\alpha+1)\beta^2$: Լուծելով այս համակարգն ըստ մոմենտների եղանակի, կստանանք՝

$$\hat{\alpha} = \frac{\hat{m}_1^2}{\hat{m}_2 - \hat{m}_1} - 1, \ \hat{\beta} = \frac{\hat{m}_2 - \hat{m}_1^2}{\hat{m}_1}:$$

Այժմ ծանոթանանք **առավելագույն ճշմարտանմանության եղանակին**, որի գաղափարը նշել է դեռևս մեծն Կ. Գաուսը, մշակել է Ռ. Ֆիշերը, և որը շատ կարևոր դեր ունի վիճակագրական գնահատման տեսությունում:

Անընդհատ X հատկանիշի տեսական բաշխումը նկարագրվում է $f_X(x,\theta)$ խտության ֆունկցիայով, իսկ ընդհատ X հատկանիշի դեպքում $\mathrm{P}_X(x,\theta)$ հավանականությամբ: $\mathbf{X}$ նմուշը կազմող $X_1,\ldots,X_N$ բաղադրիչները անկախ են և միանման բաշխված, հետևաբար դրանց համատեղ բաշխման խտությունը (ընդհատ դեպքում՝ հավանականությունը) ստացվում է որպես արտադրյալ՝

$$L(x_1,\ldots,x_N,\theta)=\prod_{n=1}^{N} f_{\mathbf{X}}(x_n,\theta):$$

$L(\mathrm{x},\theta)$-ն ֆունկցիա է x նմուշից և կոչվում է **ճշմարտանմանության ֆունկցիա**, քանի որ դրա արժեքի մեծ լինելը վկայում է x նմուշի երևան գալու բարձր հավանականության (ճշմարտանմանության) մասին: Միաժամանակ $L(\mathrm{x},\theta)$-ն θ պարամետրի ֆունկցիա է: Այս տեսակետից,

> երբ տեղի է ունեցել x նմուշը, **բնական է համարել**, որ θ պարամետրի արժեքն այն է, որի դեպքում $L(\mathrm{x},\theta)$-ն մեծագույնն է: Այդ թեզը **առավելագույն ճշմարտանմանության եղանակի սկզբունքն** է, իսկ այդ սկզբունքով գտնված $\hat{\theta}(\mathbf{X})$ վիճակագիրն կլինի θ-ի առավելագույն ճշմարտանմանության եղանակի գնահատումն:

Այլ կերպ ասած, առավելագույն ճշմարտանմանության եղանակով գնահատականը կարելի է գտնել հետևյալ պայմանից՝

$$\hat{\theta}=\arg\max_{\theta} L(\mathbf{x},\theta): \tag{7}$$

Քանի որ $L(\mathbf{x},\theta)$-ն և $\ln L(\mathbf{x},\theta)$-ն մաքսիմումի են հասնում $\hat{\theta}$-ի նույն արժեքի դեպքում, ապա այն հարմար է գտնել էքստրեմումի հետևյալ պայմանից՝

$$\left.\frac{\partial \ln L(\mathbf{x},\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}}=0: \tag{8}$$

Իհարկե, եթե θ պարամետրը K-չափանի է, $K\ge 1$, ապա էքստրեմումի պայմանները տալիս են K հատ հավասարումներ՝

$$\left.\frac{\partial \ln L(\mathbf{x},\theta_1,\ldots,\theta_K)}{\partial\theta_k}\right|_{\theta_k=\hat{\theta}_k}=0,\quad k=\overline{1,K}: \tag{9}$$

Այս համակարգը կազմող հավասարումները կոչվում են **ճշմարտանմանության հավասարումներ**:

Ապա հարկավոր է համոզվել, որ (8)-ից ստացված լուծումը համապատասխանում է մաքսիմումի և ոչ թե մինիմումի կամ շրջման կետի: Եթե (8) կամ (9) հավասարումները Θ տիրույթում լուծում չունեն, ապա մեծագույն արժեքը գտնվում է այդ տիրույթի եզրին:

Հնարավոր է տեսականորեն հիմնավորել, թե ինչու մեծ N-երի դեպքում θ պարամետրի $\hat{\theta}$ գնահատականը մոտ է լինում դրա իրական (մեզ անհայտ) արժեքին:

Նշել ենք, որ (7) սահմանումը հավասարազոր է

$$\hat{\theta}=\arg\max_{\theta}\ln L(\mathbf{x},\theta)$$

պայմանին, որն ըստ $L(\mathbf{x},\theta)$-ի սահմանման համընկնում է

$$\hat{\theta}=\arg\max_{\theta}\frac{1}{N}\sum_{n=1}^{N}\ln f_X(x_n,\theta)$$

դրույթի հետ:

Տեսականորեն (մինչև նմուշահանումը) $\mathbf{X}$ նմուշը N անկախ միանման բաշխված պատահական մեծությունների վեկտոր է, այդպիսին է նաև $\{\ln f_X(X_n,\theta), n=\overline{1,N}\}$ վեկտորը: Իսկ

$\frac{1}{N}\sum\limits_{n=1}^{N}\ln f_X(X_n,\theta)$ պատահական մեծությունը դրանց միջին թվաբանականն է, որը համաձայն մեծ թվերի օրենքի (տե՛ս գլուխ 4), ըստ հավանականության ձգտում է իր սպասելիին, այն է՝

$$\mathbf{E}\ln f_X(X,\theta)=\int f_X(x,\theta_0)\ln f_X(x,\theta)dx:$$

Այսպիսով, (7) սահմանմանը բավարարող $\hat{\theta}$-ն ըստ անընդհատության մոտ կլինի

$$\theta_1=\arg\max_{\theta}\int f_X(x,\theta_0)\ln f_X(x,\theta)dx$$

արժեքին: Մնում է օգտվել «ինֆորմացիայի տեսության հայտնի անհավասարությունից» (տե՛ս գլուխ 3-ի 3.7 ենթաբաժնում Կուլբակի-Լեյբլերի տարամիտության հատկությունը)

$$\int f_X(x,\theta_0)\ln f_X(x,\theta_0)dx\ge\int f_X(x,\theta_0)\ln f_X(x,\theta)dx,$$

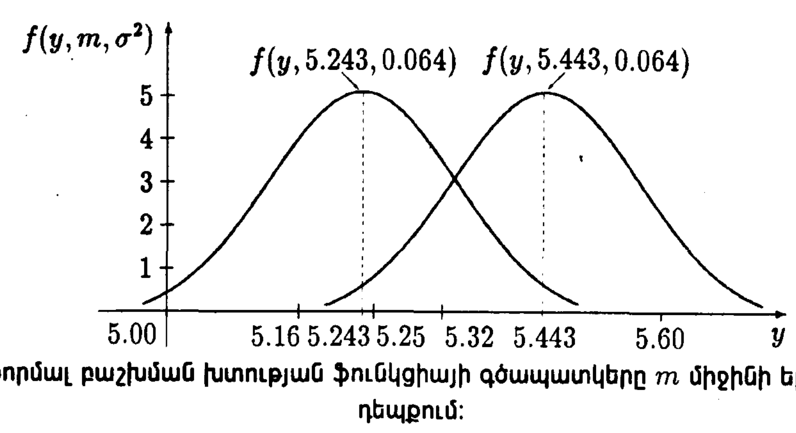
որտեղից հետևում է, որ մաքսիմում տվող արժեքն է $\theta_1=\theta_0$, և (7)-ում որոշվող $\hat{\theta}$-ն ստացվում է որոնելի իրական θ_0 արժեքին մոտ:

Առավելագույն ճշմարտանմանության եղանակի լայն կիրառությունը բացատրվում է դրա շատ օգտակար հատկություններով:

Եթե $f_X(x,\theta)$ խտության ֆունկցիան (կամ ընդհատ դեպքում՝ $\mathbf{P}_X(x,\theta)$ հավանականությունը) բավարարում է ռեգուլյարության բավականին լայն պայմանների, ապա առավելագույն ճշմարտանմանության $\hat{\theta}_N(\mathbf{X})$ գնահատի *բաշխման օրենքը մեծ N-երի դեպքում մոտ է նորմալին*, θ միջինով և $1/(N\mathbf{I}_X(\theta))$ ցրվածքով, որտեղ $\mathbf{I}_X(\theta)$-ն Ֆիշերի ինֆորմացիան է: Այդ $\hat{\theta}_N(\mathbf{X})$ *գնահատն նաև ունակ, ասիմպտոտորեն անշեղ և ասիմպտոտորեն արդյունավետ է*: Բացի դրանից, եթե գոյություն ունի պարամետրի արդյունավետ գնահատ, ապա այն կլինի առավելագույն ճշմարտանմանության հավասարման միակ լուծումը:

Օրինակ 10: Դիցուք X-ը որևէ ֆիրմայի աշխատակիցների աշխատավարձն է: Հայտնի է, որ այն բաշխված է ըստ լոգարիթմորեն նորմալ օրենքի (տե՛ս գլուխ 2), այսինքն՝ $\ln X$ պատահական մեծությունն ունի նորմալ $\mathcal{N}(m,\sigma^2)$ բաշխում: Հարկավոր է գնահատել m պարամետրը: Տրված է մեզ հետաքրքրող աշխատակիցներից պատահականորեն վերցված երեքի աշխատավարձերի նմուշը՝ $x_1=\$190$, $x_2=\$175$, $x_3=\$205$:

Լուծում: Ներկայացնենք $Y\sim\mathcal{N}(m,\sigma^2)$ պատահական մեծության $y_n=\ln x_n$, $n=\overline{1,3}$ արժեքները համապատասխան առանցքի վրա (տե՛ս նկար 1) և աշխատենք գտնել այն $\hat{m}$-ը, որի դեպքում y_1, y_2, y_3 արժեքները կլինեն առավել ճշմարտանման, այսինքն՝ $\varphi(y_n,m,\sigma^2)$ խտությունների արտադրյալը կլինի առավելագույնը: Ունենք՝ $y_1=\ln 190=5.25$, $y_2=\ln 175=5.16$, $y_3=\ln 205=5.32$:



Նկար 1. Նորմալ բաշխման խտության ֆունկցիայի գծապատկերը m միջինի երկու արժեքի դեպքում:

ված նմուշային ցրվածքը՝ $\hat{s}^2(y)=0.0064$: Նկարում պատկերված է $f_Y(y,\hat{m},\hat{s}^2)$ ության ֆունկցիան $\hat{m}=\bar{y}=5.243$ դեպքում, որը համապատասխանում է նշված y_1, y_2, y_3

արժեքների առավել ճշմարտանմանությանը, և $\hat{m} = 5.443$ դեպքում տրված դիտումները ակնհայտորեն ճշմարտանման չեն: Երկու դեպքում էլ $\sigma^2(\mathbf{x})$-ու փոխարեն վերցրել ենք $\hat{s}^2(\mathbf{y})$:

Օրինակ 11: *Պարետոյի բաշխման օրենքի* ($c > 0$, $\alpha > 0$ պարամետրերով) խտության ֆունկցիան է (տե՛ս գլուխ 2)`

$$f_X(x,\alpha,c) = \begin{cases} \alpha c^\alpha / x^{\alpha+1}, & \text{երբ } x > c, \\ 0, & \text{երբ } x \le c : \end{cases}$$

ա. Հարկավոր է առավելագույն ճշմարտանմանության եղանակի օգնությամբ գտնել c պարամետրի գնահատականը, եթե α պարամետրի արժեքը հայտնի է:

բ. Դիցուք c-ն հայտնի է, իսկ α պարամետրի արժեքը՝ անհայտ: Գտնենք դրա գնահատականը՝ $\hat{\alpha}$-ն:

Լուծում: ա. Ճշմարտանմանության ֆունկցիան է՝

$$L(\mathbf{x},\alpha,c) = \begin{cases} \alpha^N c^{N\alpha} / (x_1 x_2 \ldots x_N)^{\alpha+1}, & \text{երբ } x_n > c,\ n = \overline{1,N}, \\ 0, & \text{եթե գոնե մեկ } x_n \le c : \end{cases}$$

Այստեղից հետևում է, որ $L(\mathbf{x},\alpha,c) > 0$ միայն, եթե $c < \min\limits_n x_n$: Բայց c-ի նվազման հետ մեկտեղ c^α-ն նույնպես նվազում է, հետևաբար, քանի որ $c > 0$, $\max\limits_c L(\mathbf{x},\alpha,c)$-ն ստացվում է, երբ $\min x_n = c$: Ուստի գնահատականը՝ $\hat{c}$-ն, կլինի՝

$$\hat{c} = \min x_n :$$

բ. Երբ $x > c$, ապա

$$\ln f_X(x_n,\alpha,c) = \ln\alpha + \alpha \ln c - (\alpha+1)\ln x,$$

ուստի

$$\ln L(\mathbf{x},\alpha,c) = N(\ln\alpha + \alpha\ln c) - (\alpha+1)\sum_n \ln x_n:$$

Այստեղից, ածանցելով ըստ α-ի, ստանում ենք հավասարում α-ի նկատմամբ՝

$$N/\alpha + N\ln c - \sum_n \ln x_n = N/\alpha - \sum_n \ln(x_n/c) = 0:$$

Լուծելով հավասարումը կստանանք՝ $\hat{\alpha} = N / \sum\limits_{n=1}^{N} \ln(x_n/c)$: Քանի որ բոլոր $x_n \ge c$, ապա վերջին հայտարարը դրական է: Որպեսզի համոզվենք, որ ստացված $\hat{\alpha}$-ն տալիս է մաքսիմում, գտնենք երկրորդ ածանցյալը՝

$$\frac{\partial^2 \ln L(\mathbf{x},\alpha,c)}{\partial\alpha^2} = -\frac{N}{\alpha^2} < 0 :$$

Տեսնում ենք, որ կամայական վերջավոր դրական α-ների համար այդ ածանցյալը բացասական է, իսկ դա մաքսիմումի բավարար պայմանն է:

Օրինակ 12: Առավելագույն ճշմարտանմանության եղանակով գտնենք $[a,b]$ հատվածի վրա հավասարաչափ բաշխված հատկանիշի a և b անհայտ պարամետրերը:

Լուծում: Ճշմարտանմանության ֆունկցիան կլինի՝

$$L(\mathbf{x},a,b) = \begin{cases} 1/(b-a)^N, & \text{երբ } x_n \in [a,b],\ n = \overline{1,N}, \\ 0, & \text{եթե գոնե մեկ } x_n \notin [a,b] : \end{cases}$$

Այս ֆունկցիայի նկատմամբ (9) համակարգը խնդրի լուծում չի տա, քանի որ առաջին տիրույթում այն լուծում չունի, իսկ երկրորդի դեպքում անորոշ է: Ուրեմն $\hat{a}$-ն և $\hat{b}$-ն հարկավոր է որոնել a-ի և b-ի թույլատրելի արժեքների տիրույթի՝

$$(a \le x_{(1)}) \cap (b \ge x_{(N)})$$

եզրում, որտեղ (տե՛ս ենթաբաժին 6.1) $x_{(1)}$-ը փոփոխման շարքի փոքրագույն անդամն է (կամ, այլ կերպ ասած, առաջին կարգային վիճականին), իսկ $x_{(N)}$-ը՝ մեծագույն կարգային վիճականին: Այժմ $L(\mathbf{x}, a, b)$-ի մաքսիմումի պայմանը կգրվի այսպես՝

$$1/(\hat{b} - \hat{a})^N = \max_{a \leq x_{(1)}, b \geq x_{(N)}} 1/(b - a)^N :$$

Պարզ է, որ մաքսիմումը ստացվում է, երբ

$$\hat{a} = x_{(1)}, \quad \hat{b} = x_{(N)},$$

ուրեմն հենց դրանք էլ կլինեն որոնելի գնահատականները:

Ըստ **փոքրագույն քառակուսիների եղանակի** գնատուն որոշվում է պարամետրից կախված ուսումնասիրվող ֆունկցիայի և համապատասխան ընտրային տվյալների *տարբերությունների քառակուսիների գումարը նվազագույն դարձնելու պայմանից*:

Դիտարկենք հետևյալ պարզ օրինակը:

<u>*Օրինակ*</u> 13: Դիցուք հարկավոր է գտնել հատկանիշի տեսական m սպասելիի գնատուն:

Լուծում: Ըստ փոքրագույն քառակուսիների եղանակի, գտնենք $u = \sum_n (x_n - \hat{m})^2$ արտահայտության մինիմումը: Էքստրեմումի անհրաժեշտ պայմանից կստանանք՝

$$\frac{du}{d\hat{m}} = -2\sum_n (x_n - \hat{m}) = 0,$$

որտեղից $\hat{m} = (\sum_n x_n)/N$, այսինքն, ըստ փոքրագույն քառակուսիների եղանակի, ընտրային միջինը տեսական միջինի գնատուն է:

Առաջին անգամ այս եղանակը օգտագործվել է 1805 թ. Լեժանդրի աշխատանքում: Հավանականության տեսության տեսակետից դրա հիմնավորումն առաջինը տվել է Գաուսը 1809 և 1821 թթ.: Փոքրագույն քառակուսիների եղանակը ամենալայն տարածումն է գտել վիճակագրական հետազոտություններում՝ շնորհիվ իր երկու կարևոր առավելությունների. այն չի պահանջում հետազոտվող տվյալների բաշխման օրենքի նախնական իմացություն և լավ է մշակված հաշվարկային իրականացման առումով: Բավականին ընդհանուր դեպքերում դրա տված գնատուները լինում են ունակ, ասիմպտոտորեն անշեղ, ասիմպտոտորեն նորմալ և ասիմպտոտորեն արդյունավետ: Մենք այդ եղանակին կանդրադառնանք 11, 12, 14 գլուխներում:

Բացի նշվածներից, գոյություն ունեն մի քանի այլ եղանակներ, օրինակ, **փոքրագույն բացարձակ շեղումների եղանակը**, որի սկզբունքը երևում է անվանումից: Որոշակի իրավիճակներում առավելություն ունի դրանցից որևէ մեկը: Այդ եղանակների կարևոր հատկությունների առկայությունը պարզվում է առավելագույն ճշմարտանմանության եղանակի հետ համեմատության միջոցով:

<u>*Օրինակ*</u> 14: Դիցուք X պատահական մեծությունն ունի $\theta \in \Theta$ միջինով Լապլասի բաշխում, այսինքն՝ դրա խտության ֆունկցիան է՝

$$f_X(x, \theta) = e^{-2|x-\theta|/\sigma}/\sigma:$$

Հարկավոր է գնահատել θ պարամետրի արժեքը N ծավալի $\mathbf{x}$ ընտրանքի հիման վրա:

Լուծում: Ճշմարտանմանության ֆունկցիան կլինի՝

$$\mathrm{L}(\mathbf{x}, \theta) = e^{-2\sum_n |x_n - \theta|/\sigma}/\sigma^N,$$

Ճշմարտանմանության ֆունկցիայի մաքսիմումի պահանջից գնահատականը ստացվում է մաթեմատիկական ծրագրավորման հետևյալ խնդրի լուծումից՝

$$\min \sum_{n=1}^{N} |x_n - \theta|, \quad \theta \in \Theta:$$

Եթե θ-ի վրա սահմանափակումներ չկան, ապա լուծումը ստացվում է, երբ $\hat{\theta}$-ն նմուշային կիսողն է:

Կշռավորված վիճականիների եղանակների հիմնական գաղափարը հետևյալն է՝ $\mathbf{X} = (X_1, \ldots, X_N)$ նմուշի n-րդ դիտման արդյունքին վերագրվում է $w_n \geq 0$ գործակիցը, $\sum_{n=1}^{N} w_n = 1$: Կարելի է, օրինակ, դիտարկել w-կշռավորված մոմենտները՝

$$\hat{m}_r(N, w) = \sum_n w_n X_n^r,$$

որտեղ w-ով նշանակել ենք կշիռների վեկտորը՝ $w = (w_1, \ldots, w_N)$, իսկ որոշակի $\mathbf{x}$ նմուշի դեպքում, իհարկե, $\hat{m}_r$-ը հաշվարկվում է, տեղադրելով X_n-երի փոխարեն x_n-երը:

Երբ աշխատում են միաչափ պատահական մեծության հետ, ապա հաճախ կշիռները ընտրում և վերագրում են ոչ թե հերթական դիտումների արժեքներին, այլ կարգավորված նմուշի տարրերին, այսինքն՝ կարգային $x_{(n)}$ վիճականիներին: Այս եղանակներով հաճախ ստացվում են լավ գնատուներ:

Օրինակ, հավասարաչափ բաշխված հատկանիշի դեպքում $\hat{m}_1 = \overline{\mathbf{X}}$ գնատուի արդյունավետությունը ավելի փոքր է, քան նորմալ բաշխված հատկանիշի դեպքում: Հավասարաչափ բաշխման դեպքում շատ ավելի արդյունավետ է

$$\hat{m}_1 = (x_{(1)} + x_{(N)})/2 \tag{10}$$

գնատուն, որը **կշռավորված գնապուների** դասից է, որտեղ (10) **վիճականիում վերցված** է $w_1 = w_N = 0.5$, իսկ մնացած w_n-երը հավասար են զրոյի:

Նմուշային կիսողը՝ $\hat{x}_{\text{med}}$-ը, որպես տեսական միջինի գնատու, ունի կայունության լավ հատկություններ: Այն նույնպես կշռավորված վիճականի է (ստացվում է, վերցնելով կենտ N-ի դեպքում՝ $w_{(N+1)/2} = 1$, իսկ զույգ N-ի դեպքում՝ $w_{N/2} = w_{(N/2)+1} = 1/2$):

Մի այլ կարևոր գործընթաց կոչվում է **մաղում** (censoring): Այդ հնարքի գաղափարն այն է, որ փոփոխման շարքի «պոչային» անդամներին վերագրվում են զրոյական կշիռներ, իսկ մնացածին՝ հավասար դրականներ: Ընդ որում, այն անդամների ընտրությունը, որոնց կշիռները պետք է լինեն զրոյական, կարող է կատարվել երկու ձևով: Եթե տվյալ $[a, b]$ միջակայքից դուրս եկող արժեքներն են ստանում զրոյական կշիռ, ապա այդ **մաղումը առաջին տիպի է**. այդ դեպքում ոչ զրոյական կշիռներ ստացող դիտումների թիվը պատահական է: Իսկ եթե զրոյական կշիռները վերագրվում են տրված թվով ծայրային անցումներին, օրինակ, դիտումների արժեքների α մասի փոքր ծայրային արժեքներին և β մասի ծայրային մեծ արժեքին, ապա դա **երկրորդ տիպի մաղում** է: Այդ դեպքում պահպանված դիտումների թիվն է $N(1 - \alpha - \beta)$:

Հնարավոր են գնատուներ կառուցելու նաև այլ մոտեցումներ: Եղանակների մի ամբողջ համակարգ կարելի է ստանալ ելնելով հետևյալ նկատառումներից: Որևէ ձևով սահմանենք նմուշային բաշխման $\hat{F}_{\mathbf{X}}(x)$ ֆունկցիայի տեսական $F_X(x, \theta)$ բաշխման ֆունկցիայից շեղման $\mathbf{D}$ չափ, որը կախված է θ-ից և նմուշից՝

$$\mathbf{D} = \mathbf{D}(X, \theta):$$

Այժմ θ պարամետրի գնահատական է համարել այն $\hat{\theta}$ արժեքը, որը նվազեցնում է $\mathbf{D}$ շեղումը: Առավել կարևոր և կիրառվող է Կ. Պիրսոնի կողմից առաջարկված χ^2 չափը:

$$\chi^2 = \sum_{k=1}^{K} \frac{N}{P_k(\theta)}(\frac{n_k}{N} - P_k(\theta))^2 = \sum_{k=1}^{K}(\frac{(n_k - NP_k(\theta))^2}{NP_k(\theta)},$$

որտեղ K-ն պատահական մեծության արժեքների ենթաբազմությունների թիվն է, n_k-ն k-երորդ ենթաբազմությունում ընկած դիտումների թիվն է, $P_k(\theta)$-ն այդ ենթաբազմության մեջ ընկնելու տեսական հավանականությունն է:

Այդ գնատուն բավականին ընդհանուր պայմանների դեպքում ունակ է, ասիմպտոտորեն արդյունավետ և ասիմպտոտորեն նորմալ:

Մեկ այլ եղանակ է ստացվում Ռ. Միզեսի կողմից առաջարկված ω^2 կոչվող չափի դեպքում

$$\mathbf{D}(X,\theta) = \omega^2 = \int_{-\infty}^{\infty} [\hat{F}_{\mathbf{X}}(x) - F_X(x,\theta)]^2 dF_X(x,\theta) :$$

Այս մեծության նվազագույն դարձնելու պայմանից ստացվող գնատուն նույնպես բավականին ընդհանուր պայմանների դեպքում ասիմպտոտորեն նորմալ է, սակայն, ընդհանրապես ասած, ասիմպտոտորեն արդյունավետ չէ:

Քանորդիչների եղանակը օգտագործում է բաշխման քանորդիչների և անհայտ պարամետրերի ֆունկցիոնալ կապը, եթե այն հետազոտողներին հայտնի է:

<u>*Օրինակ*</u> 15: Եթե նմուշը ստացվել է նորմալ $N(m,\sigma^2)$ բաշխում ունեցող պատահական մեծությունից, հարմար է հենվել կիսողի և՛ ստորին, և՛ վերին քառորդիչների վրա: Կանոնական $N(0,1)$ նորմալ բաշխման դեպքում կիսողը հավասար է 0-ի, իսկ ստորին և վերին f քառորդիչները հավասար են համապատասխանաբար $\mp\Phi^{-1}(0.25)$: Քանի որ կանոնական ձևափոխությունն է (տե՛ս գլուխ 3)՝ $Y = (X-m)/\sigma$, ապա X պատահական մեծության քառորդիչները ստացվում են հակադարձ ձևափոխության միջոցով՝

$$u_{0.25} = m - \sigma\Phi^{-1}(0.25) \quad u_{0.75} = m + \sigma\Phi^{-1}(0.25),$$

իսկ կիսողը՝ x_{med}-ը, կլինի հավասար m-ին: Այստեղից գտնում ենք σ-ի համար

$$\sigma = (u_{0.75} - u_{0.25})/2\Phi^{-1}(0.25) :$$

Եթե նմուշի կիսողն է $\hat{x}_{\text{med}}$-ը , իսկ քառորդիչներն են $\hat{u}_{0.25}$-ը և $\hat{u}_{0.75}$-ը, ապա՝ ըստ քանորդիչների եղանակի, m-ի և σ-ի գնատուները կլինեն՝

$$\hat{m} = \hat{x}_{\text{med}}, \quad \hat{\sigma} = (\hat{u}_{0.75} - \hat{u}_{0.25})/2\Phi^{-1}(0.25) :$$

Հարկավոր է գիտակցել, որ առավելագույն ճշմարտանմանության եղանակի տված գնատուները լավն են շատ մեծ N-երի դեպքում: Իսկ փոքր N-երի դեպքում դրանց հետ կարող են մրցել ուրիշ գնատուներ, օրինակ, մոմենտների եղանակի կամ փոքրագույն քառակուսիների եղանակի գնատուները և այլն: Բացի դրանից, եթե բաշխման $f_X(x,\theta)$ խտության ֆունկցիայի ճշգրիտ տեսքը հայտնի չէ, երբ, օրինակ, բաշխումը մի փոքր շեղվում է նորմալ օրենքից, իսկ դա հաճախ է լինում, դրա լավ հատկությունները խախտվում են, և ավելի նպատակահարմար է դիմել «ռոբաստ», կայուն գնատուներ տվող եղանակներին: Դրանք ակտիվորեն մշակվում են վերջին տարիներին, քանի որ թույլ են տալիս կառուցել գնատուներ, որոնք թեկուզ և լավագույնը չեն ենթադրվող բաշխման օրենքի համար, բայց բաշխման օրենքի ենթադրյալից ոչ շատ շեղվելու դեպքում ցուցաբերում են հատկությունների պահպանում և կայունություն: Վերջապես, եթե պարամետրերի K թիվը մեծ է և համեմատելի է նմուշի N ծավալի հետ, ապա առավելագույն ճշմարտանմանության եղանակի գնատուն կարող է նույնիսկ ունակ չլինել:

Գլուխ 8
Միջակայքային գնահատում

Սխալվել կարելի է տարբեր կերպ, մինչդեռ վարվել կարելի է միայն մեկ ուղիով, այդ պատճառով էլ առաջինը հեշտ է, իսկ երկրորդը՝ դժվար, վրիպելը հեշտ է, նպատակին խփելը՝ դժվար:

Արիստոտել

Ես ավելի քիչ կշեղվեմ իրական արժեքից, եթե ես դիտում եմ ավելի շատ անգամ, քան եթե ես դիտում եմ ավելի քիչ անգամ:

Յակոբ Բեռնուլի

8.1. Վստահության հավանականություն: Վստահության միջակայք

Դիցուք $\mathbf{X} = (X_1, \ldots, X_N)$ նմուշը ստացվել է անընդհատ բաշխում ունեցող համույթից, որի $f_X(x,\theta)$ բաշխման խտությունը «պատկանում է Θ պարամետրական ընտանիքին», այսինքն՝ θ պարամետրը Θ բազմությունից է: Պարամետրի θ արժեքն անհայտ է: Նախորդ գլխում մենք ծանոթացանք θ պարամետրի արժեքի գնատուի՝ $\hat{\theta}(\mathbf{X})$ – կառուցման եղանակների հետ: Նման գնահատականները, հիմնված լինելով պատահական $\mathbf{X}$ նմուշի վրա, մոտավոր են նույնիսկ այն դեպքում, երբ օգտագործված գնատուն անշեղ է, ունակ և արդյունավետ: Հարց է ծագում՝ որքա՞ն է շեղվում մոտավոր $\hat{\theta}(\mathbf{X})$ արժեքը իրական θ արժեքից: Գուցե հնարավո՞ր է կառուցել այնպիսի միջակայք՝ $(\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X}))$, որը մեկին մոտ հավանականությամբ ընդգրկի որոնելի θ արժեքը: Այդ դեպքում միջակայքի երկարությունը վերևից կսահմանափակի θ-ի և դրա գնահատականի տարբերությունը: Այս գլխում մենք կտեսնենք, որ հարցի պատասխանը դրական է, և կծանոթանանք համապատասխան եղանակներին: Հարկավոր է ընդգծել, որ $\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X})$ մեծությունները պատահական են, քանի որ կառուցված են պատահական նմուշի հիման վրա, իսկ որոնելի θ արժեքը, իհարկե, պատահական չէ: Կառուցված միջակայքի երկարությունը կախված է նմուշի N ծավալից և դրա աճի հետ, բնականաբար, կարող է փոքրանալ, այսինքն՝ գնահատման ճշտությունը կարող է լավանալ:

Այդպիսի միջակայքը կոչվում է **վստահության միջակայք**, գործընթացը կոչվում է **միջակայքային գնահատում**, $\hat{\theta}_1(\mathbf{X})$-ը և $\hat{\theta}_2(\mathbf{X})$-ը, համապատասխանաբար, **ստորին և վերին վստահության սահմաններն** են, իսկ

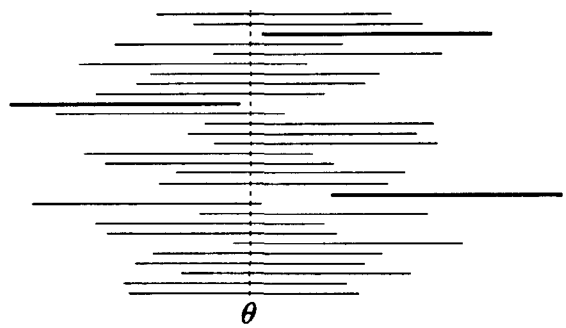
$$\mathbf{P}\{\hat{\theta}_1(\mathbf{X}) < \theta < \hat{\theta}_2(\mathbf{X})\} = 1 - \alpha \tag{1}$$

հավանականությունը՝ **վստահության հավանականությունն** է, որն անվանում են նաև $(1-\alpha)$ **հուսալիություն**, իսկ երբեմն էլ՝ $100(1-p)\%$ վստահության **մակարդակ**: Նշենք, որ α-ն կոչվում է **նշանակալիության մակարդակ**: Այդ տերմինը հաճախ է օգտագործվելու վիճակագրական վարկածների ստուգմանը նվիրված 9-րդ գլխում: α-ի մեծության ընտրությունը կատարվում է, ելնելով լուծվող կիրառական խնդրի պայմաններից՝ սովորաբար $(1-\alpha)$ հուսալիությանը տրվում է $0.90, 0.95, 0.99$ արժեքներից որևէ մեկը: (1) առնչությունը կարդացվում է այսպես՝ «$(\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X}))$ միջակայքը $(1-\alpha)$ հավանականությամբ ծածկում է θ-ն»:

Այժմ ձևակերպենք խիստ սահմանումը:

X պատահական մեծության բաշխման օրենքը բնութագրող θ պարամետրի $(1-\alpha)$ **վստահության հավանականությամբ վստահության միջակայք**, որը ստացված է $\mathbf{x}$ նմուշից, կոչվում է $(\hat{\theta}_1(\mathbf{x}), \hat{\theta}_2(\mathbf{x}))$ միջակայքը, որի ծայրակետերը (1) պայմանին բավարարող $\hat{\theta}_1(\mathbf{X})$, $\hat{\theta}_2(\mathbf{X})$ պատահական մեծությունների իրագործումներն են:

Նմուշի ամեն մի իրականացմանը համապատասխանում է մի որոշակի վստահության միջակայք: (1) պայմանը նշանակում է, որ փորձարկումների երկար շարքում միջինում $(1-\alpha)100\%$ իրականացումների համար վստահության միջակայքը կընդգրկի θ պարամետրի իրական արժեքը: Պատկերավոր դա ներկայացված է նկար 1-ում. θ-ն չընդգրկող միջակայքերը նշված են ավելի հաստ գծով:



Նկար 1. θ պարամետրի վստահության միջակայքի իրականացումները, որոնցից $\alpha \cdot 100\%$ չեն ընդգրկում θ արժեքը:

Ամեն մի որոշակի պարամետրի գնահատման ժամանակ վստահության միջակայքի երկարությունը, նմուշի N ծավալը և վստահության $(1-\alpha)$ հավանականությունը փոխադարձաբար կախված են իրարից: Հանդիպում են երեք տեսակի խնդիրներ.

հայտնի են $(1-\alpha)$-ն և N-ը, հարկավոր է գտնել ամենակարճ վստահության միջակայքը,

տրված են $(1-\alpha)$-ն և $(\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X}))$ միջակայքը, հետևաբար՝ գնահատման ճշտությունը, պահանջվում է գտնել այն փոքրագույն N-ը, որի դեպքում բավարարվում է (1) պայմանը,

տրված են $(\hat{\theta}_1(\mathbf{X}), \hat{\theta}_2(\mathbf{X}))$ միջակայքը և N-ը, պետք է որոշել նվազագույն α-ն:

Այդ խնդիրների լուծման եղանակներին է նվիրված ներկա գլխի շարունակությունը:

Որոշ կիրառական խնդիրների լուծման ժամանակ հարկ է լինում որոշել **ձախակողմյան** կամ **աջակողմյան վստահության միջակայքերը**՝

$$\mathbf{P}\{\theta < \hat{\theta}_2\} = 1-\alpha \quad \text{կամ} \quad \mathbf{P}\{\hat{\theta}_1 < \theta\} = 1-\alpha :$$

Այսպես, օրինակ, եթե θ-ն մետաղի խզման դիմադրությունն է, ապա դիտարկվում է $[\hat{\theta}_1, +\infty)$ միջակայքը, որտեղ $\hat{\theta}_1$-ը փոքրագույն դիմադրությունն է, եթե θ-ն բուժամիջոցում որևէ վտանգավոր քիմիական բաղադրիչի թույլատրելի պարունակությունն է, ապա վերցվում է $(0, \hat{\theta}_2]$ տեսքի միջակայք:

θ պարամետրի վստահության միջակայքը կառուցելու համար օգտագործում են $\hat{\theta}(\mathbf{X})$ կետային գնատուները (տե՛ս գլուխ 7), ընդ որում, նմուշի տրված N ծավալին և վստահության հավանականությանը համապատասխանող փոքրագույն երկարության վստահության միջակայք ստանալու համար հարկավոր է օգտագործել *արդյունավետ* կամ *...ն արդյունավետ գնատուներ*:

Օրինակ 1: Տեսնենք, թե ինչպես՝ N ծավալի $\mathbf{X}$ անկախ նմուշի հիման վրա, կարելի է կառուցել նորմալ բաշխված $X \sim \mathcal{N}(m, \sigma^2)$ պատահական մեծության անհայտ m սպասելիի վստահության միջակայքը: Ենթադրվում է, որ σ պարամետրի արժեքը հայտնի է, և $(1-\alpha)$ վստահությունն ավանականությունը տրված է:

Լուծում: Որպես m-ի գնահատ վերցնենք նմուշի միջինը՝ $\overline{\mathbf{X}}$-ը, որը նույնպես ունի նորմալ բաշխում, սակայն ավելի փոքր ցրվածքով՝ $\overline{\mathbf{X}} \sim \mathcal{N}(m, \sigma^2/N)$: Պետք է գտնենք (1) պայմանին բավարարող $\hat{m}_1(\mathbf{X})$ և $\hat{m}_2(\mathbf{X})$ գնահատումները՝

$$\mathbf{P}\{\hat{m}_1(\mathbf{X}) < m < \hat{m}_2(\mathbf{X})\} = 1 - \alpha :$$

Կազմենք $\hat{m}_1(\mathbf{X}) = \overline{\mathbf{X}} - \varepsilon_1$, $\hat{m}_2(\mathbf{X}) = \overline{\mathbf{X}} + \varepsilon_2$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, և տեղադրենք նախորդ պայմանի մեջ՝

$$\mathbf{P}\{\overline{\mathbf{X}} - \varepsilon_1 < m < \overline{\mathbf{X}} + \varepsilon_2\} = 1 - \alpha :$$

Մնում է որոշել ε_1-ը և ε_2-ը, սակայն երկու անհայտ մեկ պայմանից միակ ձևով չեն որոշվում: Կան այդ պայմանին բավարարող անվերջ թվով զույգեր: Եթե վերցնենք $\varepsilon_1 = \varepsilon_2 = \varepsilon$, ապա միջակայքը կլինի համաչափ $\overline{\mathbf{X}}$-ի նկատմամբ և կորոշվի միակ ձևով: Քանի որ $\overline{\mathbf{X}}$ նորմալ բաշխված պատահական մեծության համար (տե՛ս 2.4 ենթաբաժինը)

$$\mathbf{P}\{a \le \overline{\mathbf{X}} \le b\} = \Phi((b-m)\sqrt{N}/\sigma) - \Phi((a-m)\sqrt{N}/\sigma),$$

ապա

$$\mathbf{P}\{\overline{\mathbf{X}} - \varepsilon \le m \le \overline{\mathbf{X}} + \varepsilon\} = \mathbf{P}\{m - \varepsilon \le \overline{\mathbf{X}} \le m + \varepsilon\} =$$

$$= \mathbf{P}\{-\varepsilon\sqrt{N}/\sigma < (\overline{\mathbf{X}} - m)\sqrt{N}/\sigma) < \varepsilon\sqrt{N}/\sigma\}$$

$$= \Phi(\varepsilon\sqrt{N}/\sigma) - \Phi(-\varepsilon\sqrt{N}/\sigma) = 2\Phi(\varepsilon\sqrt{N}/\sigma) = 1 - \alpha :$$

Գտնելով ε-ը $\Phi(\varepsilon\sqrt{N}/\sigma) = (1-\alpha)/2$ պայմանից, կստանանք վստահության որոնելի միջակայքը: Իհարկե, նմուշի որոշակի $\mathbf{x}$ իրականացման դեպքում վստահության միջակայքը կլինի՝

$$(\overline{\mathbf{x}} - \varepsilon, \ \overline{\mathbf{x}} + \varepsilon) :$$

Նշանակելով u_δ-ով նորմալ բաշխված $U \sim \mathcal{N}(0,1)$ պատահական մեծության δ-քանորդիչը՝ $\mathbf{P}\{U < u_\delta\} = \delta$, կարելի է վերաձևակերպել ստացված արդյունքը, ելնելով $\varepsilon\sqrt{N}/\sigma = -u_{\alpha/2}$ պայմանից և գտնելով $\varepsilon = -u_{\alpha/2}\sigma/\sqrt{N}$:

> Եթե $\mathcal{N}(m, \sigma^2)$ բաշխում ունեցող պատահական մեծության σ միջին քառակուսային շեղումը հայտնի է, ապա անհայտ m սպասելիի վստահության միջակայքն է՝
>
> $$\overline{x} + u_{\alpha/2}\sigma/\sqrt{N} < m < \overline{x} - u_{\alpha/2}\sigma/\sqrt{N} :$$

Նկատենք, որ երբ N-ը մեծ է, այս գնահատականը կարելի է օգտագործել նաև նորմալից տարբեր բաշխում ունեցող հատկանիշների սպասելիի գնահատման համար, քանի որ, համաձայն կենտրոնական սահմանային թեորեմի, $\mathbf{X}$ նմուշի ծավալի՝ N-ի մեծ լինելու դեպքում $(\overline{\mathbf{X}} - m)\sqrt{N}/\sigma$ պատահական մեծությունն ունի մոտավորապես $\mathcal{N}(0,1)$ գաուսյան բաշխում:

Գոյություն ունեն վստահության միջակայքի կառուցման խնդրի լուծման երկու հիմնական մոտեցումներ:

Դրանցից առաջինը, եթե այն հաջողվում է իրականացնել, թույլ է տալիս կառուցել վստահության միջակայքը կամայական վերջավոր (նաև փոքր) ծավալի նմուշի դեպքում: Այդ եղանակն օգտագործում է «նեցուկային» կոչվող ֆունկցիաները:

Ասիմպտոտական կոչվող երկրորդ մոտեցումը ավելի համապարփակ է ուսումնասիրվող պատահական մեծությունների բաշխումների դասերի տեսակետից, սակայն հենվում է կետային գնահատումների ասիմպտոտական հատկությունների վրա և այդ պատճառով կիրառելի է միայն մեծ ծավալի նմուշների դեպքում:

Նշենք, որ եթե X-ն *ընդհատ հատկանիշ* է, ապա, իհարկե, անհայտ պարամետրը միջակայքով գնահատելու խնդիրը նույնպես կարող է լուծվել, միայն հարկավոր է (1) հավասարման հետ առնչվելիս նկատի ունենալ բաշխումների ընդհատությունը:

8.3 ենթաբաժինը նվիրված է վերոհիշյալ առաջին մոտեցմանը, իսկ երկրորդ՝ ասիմպտոտական մոտեցումը կներկայացվի 8.4 ենթաբաժնում:

8.2. Ստյուդենտի, χ^2 և Ֆիշերի բաշխումները

Նորմալ բաշխված համույթից քաղված նմուշի հիման վրա հաշվարկվող վիճակագիրները հաճախ կապված են լինում $\chi^2(N)$, Ստյուդենտի՝ $\mathrm{t}(N)$, Ֆիշերի՝ $\mathcal{F}(N_1, N_2)$ բաշխման օրենքների հետ (տե՛ս գլուխ 2): Այդ բաշխումների քանորդիչները տրված են համապատասխան հավելվածներում: Քանի որ այս և հաջորդ գլուխներում նշված բաշխման օրենքները հաճախ են կիրառվելու, դրանց վերաբերյալ հաղորդենք որոշ լրացուցիչ տեղեկություններ:

N ազատության աստիճաններով $\chi^2(N)$ բաշխում ունեցող պատահական մեծությունը սահմանված է գլուխ 2-ում, տրված է նաև դրա բաշխման խտության ֆունկցիան: Հաճախ, եթե դա թյուրիմացությունների չի բերում, այդ պատահական մեծությունը նույնպես նշանակվում է $\chi^2(N)$: Համապատասխան բաշխման բնութագրիչները տրված են գլուխ 3-ում: Նշենք, որ $\chi^2(N)$ բաշխումը գամմա բաշխման մասնավոր դեպքն է, երբ $\lambda = N/2$, $\alpha = 1/2$: Ի դեպ, **ազատության աստիճան** անվանումը արտացոլում է այն հանգամանքը, որ այդ թիվը ստացվում է որպես χ^2 օրենքով բաշխված պատահական մեծության սահմանմանը մասնակցող պատահական մեծությունների թվի և դրանց անկախ փոփոխությունը սահմանափակող գծային կապերի թվի տարբերություն:

χ^2 բաշխում ունեցող պատահական մեծությունները հաճախ են օգտագործվում վիճակագրական հաշվարկներում, կապված հետևյալ փաստի հետ, որը կձևակերպենք առանց ապացուցման:

> Դիցուք $\mathbf{X} = (X_1, \ldots, X_N)$ նմուշի բաղադրիչներն անկախ են և բաշխված միևնույն նորմալ օրենքով՝ $X_n \sim \mathcal{N}(m, \sigma^2)$, $n = \overline{1,N}$: Այդ դեպքում նմուշի միջինը՝ $\overline{\mathbf{X}}$-ը, և նմուշի ուղղված ցրվածքը՝ $\hat{s}^2(\mathbf{X})$-ը, անկախ պատահական մեծություններ են, ընդ որում, $(N-1)\hat{s}^2(\mathbf{X})/\sigma^2$ վիճակագիրն բաշխված է $\chi^2(N-1)$ բաշխման օրենքով:

Նշենք նաև, որ եթե $\chi^2(N_1)$ և $\chi^2(N_2)$ անկախ պատահական մեծությունները, համապատասխանաբար, բաշխված են N_1 և N_2 ազատության աստիճաններով χ^2 օրենքով, ապա դրանց գումարը բաշխված է $(N_1 + N_2)$ ազատության աստիճաններով χ^2 օրենքով՝

$$\chi^2(N_1) + \chi^2(N_2) = \chi^2(N_1 + N_2)\ :$$

N-ի մեծ արժեքների դեպքում ($N > 30$) $\chi^2(N)$ բաշխումը բավականին մեծ ճշտությամբ մոտարկվում է նորմալ $\mathcal{N}(N, 2N)$ բաշխումով: Այդ հատկությունն օգտագործվում է $\chi^2(N)$ բաշխման $\chi^2_p(N)$ քանորդիչների մոտարկման համար՝ $\mathcal{N}(0,1)$ նորմալ բաշխման u_p քանորդիչներով: Օգտագործում են հետևյալ բանաձևերը՝

$$\chi^2_p(N) \approx (u_p + \sqrt{2N-1})^2/2, \tag{2}$$

$$\chi^2_p(N) \approx N[1 - (2/9N) + u_p\sqrt{2/(9N)}]^3\ : \tag{3}$$

(2) բանաձևը կիրառվում է, երբ $N \geq 30$, $p \geq 0.5$, և տալիս է 1% հարաբերական սխալ, իսկ (2) բանաձևը օգտագործվում է փոքր կարգի քանորդիչների հաշվարկի համար:

Օրինակ 2: Հաշվարկենք $\chi^2_{0.75}(10)$, $\chi^2_{0.95}(40)$, $\chi^2_{0.01}(40)$ քանորդիչները:

Լուծում: Աղյուսակ Ա.3.-ից գտնում ենք $\chi^2_{0.75}(10) = 12.549$: $\chi^2_{0.95}(40)$ քանորդչի հաշվարկման համար օգտագործենք (2) բանաձևը: Քանի որ $u_{0.95} = 1.645$, ապա $\chi^2_{0.95}(40) \approx 1/2 \cdot (1.645 + \sqrt{2 \cdot -1})^2 \approx 55.47$, երկրորդ բանաձևով կստանանք՝ $\chi^2_{0.95}(40) \approx 55.755$, իսկ աղյուսակից՝ $\chi^2_{0.5}(40) = 55.758$:

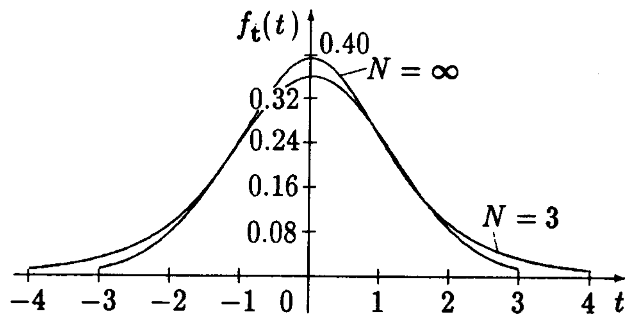
(3) բանաձևից, օգտագործելով $u_{0.01} = -u_{0.99} = -2.326$ արժեքը, կստանանք

$$\chi^2_{0.01}(40) \approx 40(1 - 2/9 \cdot 40 - 2.326\sqrt{2/9 \cdot 40})^3 \approx 22.14:$$

N ազատության աստիճաններով Ստյուդենտի բաշխում (t-բաշխում) ունեցող $\mathrm{t}(N)$ պատահական մեծությունը ստացվում է որպես $U \sim \mathcal{N}(0,1)$ և $\sqrt{\chi^2(N)/N}$ անկախ պատահական մեծությունների հարաբերություն: Հարմար է $\mathrm{t}(N)$-ով նշանակել նաև այդ պատահական մեծության բաշխման օրենքը: Ստյուդենտի բաշխման խտության ֆունկցիան տրված է գլուխ 2-ում: Այն համաչափ է $f_{\mathrm{t}}(t)$ առանցքի նկատմամբ (տե՛ս նկար 2), ուստի $t_p(N)$ քանորդիչները բավարարում են $t_p(N) = -t_{1-p}(N)$ հավասարությանը: Բնութագրիչները տրված են գլուխ 3-ում:

Մեծ N-երի դեպքում ($N > 30$) Ստյուդենտի բաշխման $t_p(N)$ քանորդիչները բավարարում են $t_p(N) \approx u_p$ մոտավոր հավասարությանը: Ավելի ճշգրիտ է հետևյալ բանաձևը՝

$$t_p(N) \approx u_p((1 - 1/(4N))^2 - u_p^2/(2N))^{-1/2} : \tag{4}$$



Նկար 2. Ստյուդենտի և նորմալ բաշխման խտության ֆունկցիաները:

<u>*Օրինակ*</u> 3: Գտնել Ստյուդենտի բաշխման $t_{0.05}(8), t_{0.90}(40)$ քանորդիչները:

Լուծում: Աղյուսակ Հ.3.-ից գտնում ենք $t_{0.95}(8) = 1.86$, $t_{0.05}(8) = -t_{0.95}(8) = -1.86$: Օգտագործելով (4) բանաձևը՝ որոշենք $t_{0.90}(40)$ քանորդիչը: Քանի որ $u_{0.90} = 1.28$ (աղյուսակ Հ. 2.-ից), ապա

$$t_{0.90}(40) \approx 1.28((1 - 1/(4 \times 40))^2 - 1.28^2/(2 \times 40))^{-1/2} \approx 1.297 :$$

Համեմատության համար Հ.3. աղյուսակից գտնենք $t_{0.90}(40)$ քանորդչի ճշգրիտ արժեքը, այն է՝ 1.303:

N_1 և N_2 ազատության աստիճաններով **Ֆիշերի բաշխման օրենքը** բնորոշ է այն $\mathcal{F}(N_1, N_2)$ պատահական մեծությանը, որն ստացվում է որպես $\chi^2(N_1)/N_1$ և $\chi^2(N_2)/N_2$ անկախ պատահական մեծությունների հարաբերություն: Այդ օրենքը նույնպես հարմար է նշանակել $\mathcal{F}(N_1, N_2)$: Դրա բաշխման խտության ֆունկցիան և բնութագրիչները տրված են, համապատասխանաբար, գլուխ 2-ում և 3-ում:

Ֆիշերի բաշխման քանորդիչները կապված են հետևյալ բանաձևով՝

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1) : \tag{5}$$

Նորմալ՝ U, χ^2, Ստյուդենտի՝ t, Ֆիշերի՝ $\mathcal{F}$, բաշխումներ ունեցող պատահական մեծությունների միջև տեղի ունեն հետևյալ առնչությունները՝

$$t^2(N) = \mathcal{F}(1,N),\ \mathcal{F}(N,\infty) = \chi^2(N)/N,\ \chi^2(1) = U^2: \tag{6}$$

Երբ $N_1 \geq 1$, $N_2 \geq 1$, Ֆիշերի բաշխման քանորդիչները կարելի է մոտավորապես հաշվարկել, օգտագործելով հետևյալ բանաձևը՝

$$F_p(N_1,N_2) \approx \frac{N_2}{N_2-2}\sqrt{\frac{2(N_1+N_2-2)}{N_1(N_2-4)}}u_p + \frac{N_2}{N_2-2}: \tag{7}$$

<u>*Օրինակ*</u> 4: Հաշվարկել Ֆիշերի բաշխման $F_{0.01}(3,5), F_{0.90}(4,100)$ և $F_{0.05}(60,120)$ քանորդիչները:

Լուծում: Օգտագործելով (5) հարաբերությունը և Հ.5. աղյուսակը՝ ստանում ենք՝

$$F_{0.01}(3,5) = 1/F_{0.90}(5,3) = 1/28.24 \approx 0.035:$$

Այնուհետև (6) բանաձևից և Հ. 4. աղյուսակից կգտնենք՝

$$F_{0.90}(4,100) \approx \chi^2_{0.9}(4)/4 = 7.78/4 = 1.945:$$

Համաձայն (7) բանաձևի, օգտագործելով $u_{0.05} = -u_{0.95} = -1.645$ արժեքը, կստանանք՝

$$F_{0.05}(60,20) \approx \frac{120}{120-2}\sqrt{\frac{2(60+120-2)}{60(120-4)}}(-1.645) + \frac{120}{120-2} \approx 0.639:$$

8.3. Նեցուկային ֆունկցիա, փոքրագույն երկարության վստահության միջակայք

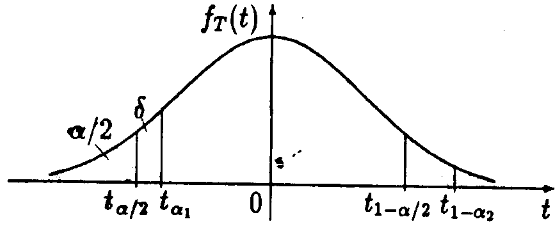
Ձևակերպենք նեցուկային ֆունկցիայի գաղափարը:

Դիցուք հարկավոր է միջակայքով գնահատել θ պարամետրի $g(\theta)$ ֆունկցիան: Եթե $\psi(\mathbf{X}, g(\theta))$ ֆունկցիայի (որն *անընդհատ* և *միընթաց* (մոնոտոն) է ըստ երկրորդ արգումենտի) բաշխման օրենքը *հայտնի է և կախված չէ θ-ից*, ապա այդ **ֆունկցիան** (վիճականին) կոչվում է **նեցուկային** կամ պարզապես $g(\theta)$-ի գնահատման **նեցուկ**:

Այստեղ տեղին է հետևյալ դիտողությունը. կետային գնատուի կառուցման ժամանակ մենք հաճախ կարիք չունենք իմանալու գնատուի բաշխման օրենքը, ընդհակառակը՝ միջակայքային գնահատականը, անհայտ պարամետրի մասին տալով ավելի լրիվ տեղեկություն, հենվում է լրացուցիչ, մասնավորապես գնատուի բաշխման օրենքի իմացության վրա:

Որոշ դեպքերում օգտակար է հետևյալ հանգամանքը:

Դիցուք T-ն պատահական մեծություն է, որի բաշխման խտության ֆունկցիան՝ անկախ θ-ի արժեքից, համաչափ է 0-ի նկատմամբ: Եթե $f_T(t)$ խտության ֆունկցիան անընդհատ է և միամոդալ, ապա տրված α-ի համար $(1-\alpha)$ հավանականություն ունեցող միջակայքի երկարությունը փոքրագույնն է այն դեպքում, երբ միջակայքը *համաչափ* է 0-ի նկատմամբ:



Նկար 3. Միամոդալ անընդհատ բաշխման $t = 0$ ուղղի նկատմամբ համաչափ խտության ֆունկցիա:

Ապացույց: Դիտարկենք $t_{\alpha/2}, t_{1-\alpha/2}, t_{\alpha_1}, t_{1-\alpha_2}$ քանորդիչները (տե՛ս գլուխ 3), որտեղ $\alpha_1 = \alpha/2+\delta$, $\alpha_1+\alpha_2=\alpha$: Պարզ է, որ $-t_{\alpha/2}=t_{1-\alpha/2}$, և $\alpha_1-\alpha/2=\alpha/2-\alpha_2=\delta$: Հարկավոր է ցույց տալ, որ $t_{1-\alpha_2}-t_{\alpha_1}\geq t_{1-\alpha/2}-t_{\alpha/2}$:

Քանի որ

$$\int_{t_{\alpha/2}}^{t_{\alpha_1}} f_T(t)dt = \int_{t_{1-\alpha/2}}^{t_{1-\alpha_2}} f_T(t)dt = \delta,$$

ապա

$$(t_{\alpha_1}-t_{\alpha/2})f_T(t') = (t_{1-\alpha_2}-t_{1-\alpha/2})f_T(t''),$$

որտեղ $t_{\alpha/2} < t' < t_{\alpha_1}$ և $t_{1-\alpha/2} < t'' < t_{1-\alpha_2}$: Քանի որ $f_T(t)$ խտության ֆունկցիան համաչափ է 0-ի նկատմամբ՝ $f_T(t_{\alpha/2}) = f_T(t_{1-\alpha/2})$, և, լինելով միամոդալ, կիսատառանցքներում միընթաց է, ապա

$$f_T(t') > f_T(t_{\alpha/2}) = f_T(t_{1-\alpha/2}) > f_T(t''):$$

Հետևաբար ստանում ենք, որ $t_{1-\alpha_2}-t_{1-\alpha/2}\geq t_{\alpha_1}-t_{\alpha/2}$ կամ $t_{1-\alpha_2}-t_{\alpha_1}\geq t_{1-\alpha/2}-t_{\alpha/2}$:

Նկատենք, որ նշված բաշխման օրենքների թվին են պատկանում մի քանի հաճախ հանդիպող և կարևոր, մասնավորապես նորմալ և Ստյուդենտի բաշխումները: χ_2 և Ստյուդենտի բաշխումների հաճախակի օգտագործումը հետևանք է նորմալ բաշխված հատկանիշի՝ անհայտ պարամետրերի գնահատման համար ստացված նմուշների և դրանցից կազմված վիճականիների բաշխումների հետևյալ հատկությունների:

Եթե N ծավալի $\mathbf{X}$ նմուշի բաղադրիչները անկախ և միանման բաշխված նորմալ պատահական մեծություններ են՝ $X_i \sim \mathcal{N}(m,\sigma^2)$, ապա

$$(\overline{\mathbf{X}}-m)\sqrt{N}/\hat{s}(\mathbf{X}) \sim \mathbf{t}(N-1), \tag{8}$$

$$(N-1)\hat{s}^2(\mathbf{X})/\sigma^2 \sim \chi^2(N-1): \tag{9}$$

$$N\hat{s}_0^2(\mathbf{X})/\sigma^2 \sim \chi^2(N), \tag{10}$$

որտեղ՝ $\hat{s}^2(\mathbf{X}) = (1/(N-1))\sum_n (X_n-\overline{\mathbf{X}})^2$ և $\hat{s}_0^2(\mathbf{X}) = (1/N)\sum_n (X_n-m)^2$:

Օրինակների օգնությամբ տեսնենք, թե ինչպես են կիրառվում նեցուկային ֆունկցիաները:

<u>*Օրինակ*</u> 5: Ելնելով $\mathbf{X}$ նմուշից, գտնել նորմալ բաշխված հատկանիշի m սպասելիի α նշանակալիության մակարդակին համապատասխանող վստահության միջակայքը: Ի տարբերություն օրինակ 1-ի, հատկանիշի σ ցրվածքը հայտնի չէ:

Լուծում: Համաձայն (8)-ի $(\overline{\mathbf{X}}-m)\sqrt{N-1}/\hat{s}$ վիճականին ենթարկվում է $(N-1)$ ազատության աստիճաններով Ստյուդենտի բաշխման օրենքին, այն կարող է կատարել նեցուկային ֆունկցիայի դերը: Վերը ապացուցված հատկության համաձայն, կարող ենք վերցնել 0-ի նկատմամբ համաչափ միջակայք: Այդուսակից որոշում ենք տրված α հավանականությանը համապատասխանող t-ի բաշխման $t_{\alpha/2}(N-1)$ և $t_{1-\alpha/2}(N-1)$ քանորդիչները: Համաչափությունից հետևում է, որ

$$-t_{\alpha/2}(N-1) = t_{1-\alpha/2}(N-1)\geq 0:$$

Ուրեմն

$$\mathbf{P}\{t_{\alpha/2}(N-1) < (\overline{\mathbf{X}}-m)\sqrt{N-1}/\hat{s} < t_{1-\alpha/2}(N-1)\} = 1-\alpha:$$

Այստեղից ստանում ենք վստահության որոնելի միջակայքը (տեղադրելով $\overline{\mathbf{X}}$-ի որոշակի $\overline{\mathbf{x}}$ արժեքը)՝

$$(\overline{\mathbf{x}}+t_{\alpha/2}(N-1)\hat{s}(\mathbf{x})/\sqrt{N-1}, \overline{\mathbf{x}}-t_{\alpha/2}(N-1)\hat{s}(\mathbf{x})/\sqrt{N-1}):$$

Նկատենք, որ, երբ $N-1>30$, Ստյուդենտի բաշխումը մոտ է $\mathcal{N}(0,1)$ նորմալին, հետևաբար կարելի է կիրառել օրինակ 1-ում ստացված բանաձևը, որում սակայն անհայտ σ-ն փոխարինված է նմուշային $\hat{s}(\mathbf{x})$-ով՝

$$\overline{\mathbf{x}} - u_{\alpha/2}\hat{s}(\mathbf{x})/\sqrt{N} < m < \overline{\mathbf{x}} + u_{\alpha/2}\hat{s}(\mathbf{x})/\sqrt{N}:$$

Օրինակ 6: **X** նմուշի հիման վրա, որի բաղադրիչները նորմալ բաշխված $N(m, \sigma^2)$ իրարից անկախ պատահական մեծություններ են, համարելով m-ը հայտնի, կառուցենք σ^2 ցրվածքի $(1-\alpha)$ մակարդակի վստահության միջակայքը:

Լուծում: Ցրվածքի արդյունավետ գնատուն, սպասելի m-ի հայտնի լինելու դեպքում վերը տրված $\hat{s}_0^2(\mathbf{X})$-ն է: Իսկ $\chi^2(N) = N\hat{s}_0^2(\mathbf{X})/\sigma^2$ վիճականին, համաձայն (10)-ի, անկախ m և σ^2 պարամետրերի արժեքներից բաշխված է N ազատության աստիճաններով χ^2 օրենքով և որպես σ^2-ու ֆունկցիա, երբ $\sigma^2 > 0$, անընդհատ է և խիստ միընթաց : Այն կարող է նեցուկ լինել: Ուրեմն

$$\mathbf{P}\{\chi^2_{\alpha/2}(N) < \chi^2(N) < \chi^2_{1-\alpha/2}(N)\} = 1-\alpha$$

պայմանի համաձայն աղյուսակից գտնենք անհրաժեշտ քանորդիչները և կազմենք վստահության միջակայքը, լուծելով

$$\chi^2_{\alpha/2}(N) < N\hat{s}_0^2(\mathbf{X})/\sigma^2 < \chi^2_{1-\alpha/2}(N)$$

անհավասարությունը σ^2-ու նկատմամբ՝

$$N\hat{s}_0^2(\mathbf{X})/\chi^2_{1-\alpha/2}(N) < \sigma^2 < N\hat{s}_0(\mathbf{X})\chi^2_{\alpha/2}(N):$$

Նույն մակարդակով գնահատվում է նաև σ-ն՝

$$\sqrt{N\hat{s}_0^2(\mathbf{X})/\chi^2_{1-\alpha/2}(N)} < \sigma < \sqrt{N\hat{s}_0^2(\mathbf{X})/\chi^2_{\alpha/2}(N)}:$$

Նշենք, որ այս միջակայքերը համաչափ չեն $\hat{s}_0^2$-ու նկատմամբ:

Օրինակ 7: Դիտարկենք նորմալ բաշխում ունեցող հանույթի ցրվածքի միջակայքային գնահատման խնդիրը, երբ սպասելին հայտնի չէ:

Լուծում: σ^2 պարամետրի միջակայքային գնատուի կառուցման համար այժմ կիրառում ենք $(N-1)\hat{s}^2(\mathbf{X})/\sigma^2$ վիճականին, որը համաձայն (9)-ի ենթարկվում է $(N-1)$ ազատության աստիճաններով $\chi^2(N-1)$ բաշխման օրենքին: Աղյուսակից գտնելով $\chi^2_{\alpha/2}(N-1)$ և $\chi^2_{1-\alpha/2}(N-1)$ քանորդիչները, ստանում ենք, որ

$$\mathbf{P}\{\chi^2_{\alpha/2}(N-1) < (N-1)\hat{s}^2(\mathbf{X})/\sigma^2 < \chi^2_{1-\alpha/2}(N-1)\} = 1-\alpha:$$

Լուծելով անհավասարությունները σ^2-ու նկատմամբ, գտնենք որոնելի վստահության միջակայքը՝

$$(N-1)\hat{s}^2(\mathbf{X})/\chi^2_{1-\alpha/2}(N-1), (N-1)\hat{s}^2(\mathbf{X})/\chi^2_{\alpha/2}(N-1),$$

որը $(1-\alpha)$ հավանականությամբ ընդգրկում է անհայտ σ^2 արժեքը (բնականաբար, որոշակի **x** նմուշի դեպքում տեղադրվում է $\hat{s}^2(\mathbf{x})$-ի հաշվարկված արժեքը):

Օրինակ 8: Համատիրությունը ցանկանում է նմուշի հիման վրա որոշ դասի բնակարանների համար 99% վստահությամբ և 100 դրամ ճշտությամբ գնահատել մեկ ամսվա միջին բնակվարձը: Ենթադրելով, որ բնակվարձը բաշխված է նորմալ օրենքով՝ 350 դրամ միջին քառակուսային շեղումով, գտնենք նմուշի նվազագույն ծավալը:

Լուծում : Հարկավոր է գտնել այն N-ը, որի դեպքում

$$\mathbf{P}\{|\overline{\mathbf{X}} - m| < 100\} \geq 0.99:$$

Վերցնելով $1-\alpha = 0.99$, աղյուսակից՝ $\Phi(u_{1-\alpha/2}) = 0.495$ պայմանից, գտնում ենք $u_{1-\alpha/2} = u_{0.995} = 2.6$: Երբ $\varepsilon = 100$, և $\sigma = 350$, ստանում ենք $N = u^2_{0.995}\sigma^2/\varepsilon^2 = 82.81$, ուրեմն, քանի որ N-ը պետք է լինի ամբողջ՝ $N \geq 83$:

Սրբեմն օգտագործվում են **ասիմպտոտորեն նեցուկային ֆունկցիաներ**: Դիցուք $f_X(x,\theta)$-ն X անընդհատ պատահական մեծության խտությունն է: Կարելի է ցույց տալ, որ

$$Z = \frac{\partial \log f_X(X,\theta)}{\partial\theta}$$

պատահական մեծության սպասելին հավասար է զրոյի, իսկ ցրվածքը՝ Ֆիշերի ինֆորմացիային (տե՛ս 7.2 ենթաբաժինը)՝ $\mathbf{E}Z^2 = \mathbf{I}_X(\theta)$:

$\mathbf{X} = (X_1, \ldots, X_N)$ անկախ նմուշի համար, երբ $N \to \infty$, կենտրոնական սահմանային թեորեմից հետևում է, որ

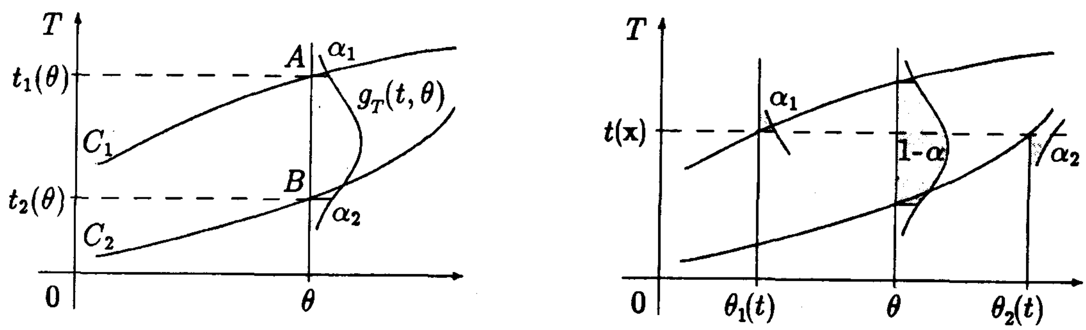
$$\left(\sum_{n=1}^{N} Z_n\right) / \sqrt{N\mathbf{I}_X(\theta)}$$

հարաբերությունը զուգամիտում է կանոնաձև նորմալ պատահական մեծությանը, հետևաբար այն կարելի է համարել **ասիմպտոտորեն նեցուկային**:

Ցույց տանք, ինչպես կարելի է վարվել, երբ $T(\mathbf{X})$-ը θ պարամետրի գնատու է, որի բաշխման խտության ֆունկցիան $g_T(t,\theta)$-ն հայտնի է, բայց, ի տարբերություն նեցուկի, կախված է θ-ից: Տրված θ-ի համար գոյություն ունեն T-ի այնպիսի երկու արժեք՝ $t_1(\theta)$ և $t_2(\theta)$, որ

$$\mathbf{P}\{T(\mathbf{X}) > t_1(\theta)|\theta\} = \alpha_1, \quad \mathbf{P}\{T(\mathbf{X}) < t_2(\theta)|\theta\} = \alpha_2,$$

որտեղ α_1-ը և α_2-ը նախօրոք վերցված հավանականություններ են, և $\alpha_1 + \alpha_2 = \alpha$:



Նկար 4: Վստահության միջակայքի կառուցումը գնատուի խտության ֆունկցիայի կիրառմամբ:

Երբ θ-ն փոփոխվում է Θ-ում, A և B կետերը (տե՛ս նկար 4) տեղափոխվում են C_1 և C_2 կորերով: $\mathbf{x}$ նմուշին համապատասխանող $t(\mathbf{x})$ կետով անցնող հորիզոնական (կետագծային) ուղիղը հատում է C_1 և C_2 կորերը այն կետերում, որոնց $\theta_1(t)$ և $\theta_2(t)$ աբսցիսները կազմում են θ պարամետրի $(1-\alpha)$ վստահության միջակայքը: $\theta_1(t)$ և $\theta_2(t)$ կետերը կարելի է հաշվարկել հետևյալ պայմաններից՝

$$\int_{-\infty}^{t(\mathbf{x})} g_T(t,\theta_1)dt = \alpha_1, \quad \int_{t(\mathbf{x})}^{\infty} g_T(t,\theta_2)dt = \alpha_2 :$$

8.4. Վստահության միջակայքի կառուցման ասիմպտոտական եղանակը

Հայտնի է (տե՛ս գլուխ 7), որ նմուշի N ծավալի մեծ լինելու դեպքում առավելագույն ճշմարտանմանության եղանակի գնատուն ասիմպտոտորեն նորմալ է, հետևաբար, նախորդ ենթաբաժնի օրինակներում դիտարկված եղանակները կիրառելի են դառնում նույնիսկ ոչ նորմալ բաշխված հատկանիշների նկատմամբ:

Դիցուք $\mathbf{x}$-ը դիտարկումներից ստացված նմուշն է, իսկ $t(\mathbf{x})$-ը՝ $\mathbf{x}$-ին համապատասխանող T վիճականու արժեքը: Մենք արդեն դիտարկել ենք նմուշի մեծ ծավալների դեպքում միջակայքային գնահատման օրինակներ: Այսպես, եթե X պատահական մեծությունը տարբերվում է նորմալից, սակայն N-ը մեծ է, և σ^2 ցրվածքը հայտնի է, ապա $(1-\alpha)$ հավանականությամբ տեսական սպասելիի միջակայքային գնահատումը մոտավորապես կիամընկնի օրինակ 1-ում ստացվածի հետ: Օրինակ 6-ում, երբ $N-1>30$, Ստյուդենտի բաշխումը մոտենում է նորմալ բաշխմանը, և m սպասելիի վստահության միջակայքը կարելի է մոտավորապես կառուցել նորմալ բաշխման քանորդիչների միջոցով՝

$$(\bar{\mathbf{x}}+u_{\alpha/2}\hat{s}/\sqrt{N-1}, \bar{\mathbf{x}}-u_{\alpha/2}\hat{s}/\sqrt{N-1}):$$

Օրինակ 9: N անկախ փորձերում A պատահույթը տեղի է ունեցել n_A անգամ: Գտնել մեկ փորձում A պատահույթի տեղի ունենալու p հավանականության վստահության միջակայքը՝ N-ը բավականաչափ մեծ լինելու դեպքում:

Լուծում : Մեկ փորձում A-ի տեղի ունենալու p հավանականության արդյունավետ գնատուն $\hat{p}=n_A/N$ հարաբերական հաճախությունն է: Համաձայն Մուավրի-Լապլասի ինտեգրալային սահմանային թեորեմի (տե՛ս գլուխ 2), n_A/N հաճախությունը ասիմպտոտորեն ունի նորմալ բաշխում, իսկ

$$U=(n_A/n-p)/\sqrt{p(1-p)/N}$$

վիճականին ենթարկվում է $\mathcal{N}(0,1)$ բաշխման օրենքին՝ անկախ p-ից: Ուստի մեծ N-երի դեպքում կունենանք

$$\mathbf{P}\left(\left|\frac{n_A/N-p}{\sqrt{p(1-p)/N}}\right|<u_{1-\alpha/2}\right)\approx 1-\alpha:$$

Այստեղից կստանանք, որ մոտավորապես $1-\alpha$ հավանականությամբ տեղի ունի

$$n_A/N-u_{1-\alpha/2}\sqrt{p(1-p)/N}<p<n_A/N+u_{1-\alpha/2}\sqrt{p(1-p)/N}:$$

Ձախ և աջ մասերում p-ն փոխարինելով իր գնահատականով, կստանանք, որ

> Բեռնուլիի փորձերում A պատահույթի p հավանականության վստահության մոտավոր միջակայքն է՝
>
> $$\hat{p}-u_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N}<p<\hat{p}+u_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N}:$$

Օրինակ 10: Հայտնի է, որ «Պահպանողական» կուսակցության կողմնակիցների քվեների տոկոսը՝ p-ն, մոտավորապես 40% է: Շուկայական ուսումնասիրությունների կազմակերպությունը նպատակ ունի ընտրողների N ծավալի հարցում անցկացնել այնպես, որ պահպանողական կուսակցությունը ընտրողների տոկոսային հարաբերությունը 2%-ից ոչ ավելի տարբերվի բնակչության մեջ համապատասխան տոկոսից:

Ի՞նչ ծավալի նմուշ (100-ի ճշտությամբ) պետք է ունենա կազմակերպությունը:

Լուծում : Պետք է կառուցվի $(\hat{p}-0.02, \hat{p}+0.02)$ տեսքի, օրինակ, 90%-անոց վստահության միջակայք: Ուրեմն N-ը պետք է բավարարի $1.645\sqrt{\hat{p}(1-\hat{p})/N}\approx 0.02$ պայմանին: Լուծելով N-ի նկատմամբ և տեղադրելով $\hat{p}=0.4$, կստանանք՝ $N\approx 1.645^2\times 0.4\times 0.6/0.02^2=1623.6$, այսինքն՝ 1600 ծավալի նմուշը կլինի բավարար:

8.5. Բազմակի պարամետրերի վստահության տիրույթը

Մի քանի պարամետրերի վստահության միջակայքերին անցնելիս ծագում են հետևյալ հարցերը: Հնարավո՞ր է պարամետրերից յուրաքանչյուրի համար ստանալ առանձին վստահության միջակայքեր: Հնարավո՞ր է ստանալ վստահության միջակայքեր պարամետրերի որոշ համակցության համար, օրինակ, դրանց գումարի, տարբերության, հարաբերության և այլն: Հնարավո՞ր է ստանալ մի քանի պարամետրերի վստահության բազմաչափ տիրույթներ:

Ստորև կանդրադառնանք այդ հարցերի մի մասին, որոշ տեղեկություններ կշարադրվեն հաջորդ գլուխներում:

Երբ հարկավոր է գնահատել K հատկանիշներ կամ մի հատկանիշի մի քանի պարամետրեր, $(1-\alpha)$ հավանականությամբ պետք է տեղի ունենան միաժամանակ հետևյալ պատահույթները՝

$$\{\hat{\theta}_{11}(\mathbf{X}) < \theta_1 < \hat{\theta}_{12}(\mathbf{X})\}, \{\hat{\theta}_{21}(\mathbf{X}) < \theta_2 < \hat{\theta}_{22}(\mathbf{X})\}, \ldots \{\hat{\theta}_{K1}(\mathbf{X}) < \theta_K < \hat{\theta}_{K2}(\mathbf{X})\} :$$

Այդ միջակայքերով կազմված տիրույթը կոչվում է $(1-\alpha)$ հուսալիությամբ **վստահության տիրույթ**: Ընդհանուր դեպքում այս խնդիրը բավականին բարդ է: Այն որոշ չափով հեշտանում է, երբ առանձին պարամետրերի հետ կապված պատահույթները *անկախ* են: Ասենք, $K = 2$ դեպքի համար

$$\mathbf{P}\{\hat{\theta}_{11}(\mathbf{X}) < \theta_1 < \hat{\theta}_{12}(\mathbf{X}),\ \hat{\theta}_{21}(\mathbf{X}) < \theta_2 < \hat{\theta}_{22}(\mathbf{X})\} =$$

$$\mathbf{P}\{\hat{\theta}_{11}(\mathbf{X}) < \theta_1 < \hat{\theta}_{12}(\mathbf{X})\}\mathbf{P}\{\hat{\theta}_{21}(\mathbf{X}) < \theta_2 < \hat{\theta}_{22}(\mathbf{X})\} = 1-\alpha : \qquad (11)$$

Անկախության դեպքում վստահության տիրույթի կառուցումը հանգում է K պարամետրերի համար $\sqrt[K]{1-p}$ հուսալիությամբ վստահության միջակայքերի կառուցմանը:

Մենք կանդրադառնանք միայն նորմալ բաշխման m և σ^2 պարամետրերի գնահատմանը, գիտենալով, որ հատկանիշի նորմալ բաշխման դեպքում N ծավալի $\mathbf{X}$ նմուշի համար

$$(\overline{\mathbf{X}} - m)/\sigma\sqrt{N} \sim \mathcal{N}(0,1) \quad \text{և} \quad (N-1)\hat{s}^2(\mathbf{X})/\sigma^2 \sim \chi^2(N-1)$$

պատահական մեծությունները անկախ են:

Օրինակ 11: Նորմալ՝ $\mathcal{N}(m,\sigma^2)$ պատահական մեծության երկու m և σ^2 պարամետրերն էլ անհայտ են, հարկավոր է կառուցել դրանց համատեղ գնահատման $(1-\alpha)$ վստահության տիրույթը, համարելով $N > 30$:

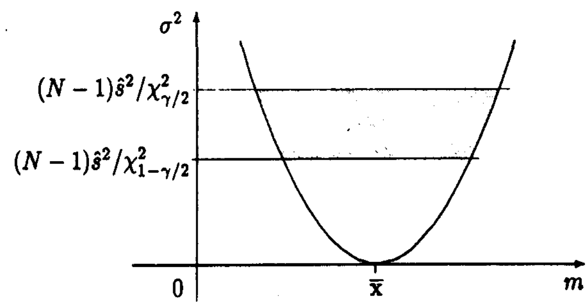
Լուծում : Մենք 5 և 7 օրինակներում դիտարկել ենք այդ պարամետրերի առանձին վստահության միջակայքերի կառուցման խնդիրները: Սակայն այնտեղ կիրառված $(\overline{\mathbf{X}} - m)\sqrt{N-1}/\hat{s}$ և $(N-1)\hat{s}^2(\mathbf{X})/\sigma^2$ վիճականիները անկախ չեն, քանի որ երկուսում էլ մասնակցում է $\hat{s}$-ը: Համաձայն ձևակերպված հատկության, $(\overline{\mathbf{X}} - m)\sqrt{N}/\sigma$ և $(N-1)\hat{s}^2(\mathbf{X})/\sigma^2$ վիճականիները անկախ են: Ելնելով (11) պայմանից՝ կարող ենք պահանջել, որ տեղի ունենան $(1-\beta)$ հավանականությամբ

$$N(\overline{\mathbf{X}} - m)^2/\sigma^2 < u^2_{\beta/2}$$

և $(1-\gamma)$ հավանականությամբ

$$(N-1)\hat{s}^2(\mathbf{X})/\chi^2_{\gamma/2}(N-1) < \sigma^2 < (N-1)\hat{s}^2(\mathbf{X})/\chi^2_{1-\gamma/2}(N-1)$$

անհավասարությունները և, հետևաբար, երկուսը միաժամանակ տեղի կունենան $1-\alpha = (1-\beta)(1-\gamma)$ հավանականությամբ: Որոնելի վստահության տիրույթի տեսքը պատկերված է նկար 5-ում:



Նկար 5: Նորմալ հատկանիշի m և σ^2 պարամետրերի համատեղ վստահության տիրույթի տեսքը:

Օրինակ 12: Դիցուք որոշ անկյան մեծությունը որոշելու համար աստիճաններով ստացվել են 22 չափումների արդյունքները: Միջինը ստացվել է 3.0318, իսկ $\hat{s}^2$-ն՝ 0.0303: Հարկավոր է կառուցել սպասելիի և ցրվածքի վստահության տիրույթը 0.98 վստահության հավանականությամբ:

Լուծում: Քանի որ $\sqrt{0.98} \approx 0.99$, գտնելով $22 - 1 = 21$ ազատության աստիճանի դեպքում՝

$$u_{0.005} = -2.57, \quad \chi^2_{0.005}(21) = 8.034, \quad \chi^2_{0.995}(21) = 41.401,$$

տեղադրենք օրինակում ստացված համապատասխան բանաձևերի մեջ՝

$$3.0318 - 2.57 \cdot \sqrt{0.0303/22} = 2.9364, \quad 3.0318 + 2.57 \cdot \sqrt{0.0303/22} = 3.1272:$$

Այսպիսով, որոնելի տիրույթը կլինի նկար 5-ում պատկերվածը, որտեղ

$$2.9364 < \sigma^2 < 3.1272, \quad 22(3.0318 - m)^2 \leq \sigma^2(2.57)^2 :$$

<u>*Օրինակ*</u> 13: Դիտարկենք երկու նորմալ հատկանիշների՝ $X \sim \mathcal{N}(m_X, \sigma_X^2)$, և $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$, սպասելիների տարբերության՝ $(m_X - m_Y)$-ի վստահության միջակայքի կառուցման խնդիրը, երբ առկա են N_X, N_Y ծավալների իրարից անկախ **X** և **Y** նմուշները, և σ_X-ի և σ_Y-ի արժեքները հայտնի են:

Լուծում: Կազմենք $\overline{\mathbf{X}} - \overline{\mathbf{Y}}$ տարբերությունը և նկատենք, որ այն ունի նորմալ բաշխում

$$\mathcal{N}(m_X - m_Y, \sigma_X^2/N_X + \sigma_X^2/N_Y) :$$

Դրանից հետևում է, որ $(1-\alpha)$ մակարդակի վստահության միջակայքը կլինի (նմուշների միջինների ստացված $\bar{\mathbf{x}}$ և $\bar{\mathbf{y}}$ արժեքների դեպքում)՝

$$\bar{\mathbf{x}} - \bar{\mathbf{y}} + u_{\alpha/2}\sqrt{\sigma_X^2/N_X + \sigma_Y^2/N_Y} < m_x - m_y < \bar{\mathbf{x}} - \bar{\mathbf{y}} - u_{\alpha/2}\sqrt{\sigma_X^2/N_X + \sigma_Y^2/N_Y} :$$

Գլուխ 9
Վիճակագրական վարկածների ստուգում

Եթե հակադիր կարծիքներ չեն արտահայտվել, ապա ինչի՞ց ընտրել լավագույնը:

Հերոդոտոս

Գիտնականը մարդ է, որն ինչ-որ բանում համարյա համոզված է:

Ժյուլ Ռենար

9.1. Վիճակագրական վարկածների դասակարգումը

Վիճակագրական ուսումնասիրության ընթացքում անհրաժեշտ է լինում դատել դիտարկվող ստոխաստիկ մոդելի բնույթի և անհայտ պարամետրերի մասին: Հանուրի վերաբերյալ պատահական նմուշի հիման վրա արվող կամայական դատողություն ենթադրվող է: Հանուրի բաշխման օրենքի կամ որոշակի պարամետրերի հատկությունների վերաբերյալ ենթադրությունները անվանում են **վիճակագրական վարկածներ**:

Հետազոտության նպատակն է լինում որոշել՝ ինչպե՞ս են համաձայնեցվում նմուշային տվյալները առաջադրված վարկածի հետ, կարելի՞ է արդյոք նրանց միջև եղած անհամաձայնությունը վերագրել նմուշահանման ընթացքում առաջացած պատահական տատանումներին, թե՞ վարկածը էապես հակասում է փորձնական տվյալներին և պետք է ժխտվի: Այդ խնդիրը լուծվում է մաթեմատիկական վիճակագրության հատուկ եղանակներով, որոնց հետ ծանոթանալու ենք այս և հաջորդ գլուխներում:

Առաջադրված վարկածի հիմնավորված համադրումը նմուշի հետ կատարվում է այս կամ այն **վիճակագրական հայտանիշի** օգնությամբ: Այդ գործողությունը անվանում են **վարկածի վիճակագրական ստուգում**: Համադրման արդյունքը կարող է լինել բացասական, եթե վիճակագրական տվյալները հակասում են վարկածին և հետևաբար այդ վարկածից պետք է հրաժարվել՝ այն ժխտվում է, կամ ոչ բացասական՝ փորձնական տվյալները չեն հակասում բերված վարկածին, և այդ վարկածը կարելի է ընդունել որպես հնարավոր լուծում: Սակայն հարկավոր է գիտակցել, որ *վարկածի վիճակագրական ստուգման արդյունքը չի նշանակում, որ ենթադրվող պնդումը միակ հարմարն է*:

Վարկածների ստուգման հետ կապված են հետևյալ գաղափարներն ու տերմինները: Ստուգման ենթակա **վարկածը** նշանակում են H_0 և անվանում **հիմնական** (կամ **զրոյական**): Վարկածի բովանդակությունը գրում են H_0-ի աջից երկու կետ դնելուց հետո: Օրինակ, $H_0 : p < p_0$, որտեղ p-ն արտադրանքի խոտանի հավանականությունն է:

Վարկածները լինում են պարզ և բարդ: **Վարկածը պարզ** է, եթե դրա պնդումը միարժեք բնութագրում է օբյեկտը: Օրինակ, դիցուք X հատկանիշը բաշխված է Պուասոնի $\Pi(\theta)$ օրենքով և առաջարկվում է ստուգել $H_0 : \theta = \theta_0$ վարկածը: **Վարկածը** կոչվում է **բարդ**, եթե այն վերաբերում է անհայտ պարամետրի կամ բաշխման՝ որոշակի բազմությանը պատկանելուն: Օրինակ, դիցուք X_n-ը, $n = \overline{1,12}$, տվյալ հիմնարկությունում n-րդ ամսին աշխատանքի ընդունված մասնագետների քանակն է: Ստուգել հետևյալ վարկածը՝ միջին հաշվով ամսական աշխատանքի է ընդունվում հինգից ոչ պակաս աշխատակից: Եթե θ-ով նշանակենք աշխատանքի ընդունված մարդկանց միջին քանակը, ապա վարկածը կգրվի այսպես՝ $H_0 : \theta \geq 5$:

Յուրաքանչյուր զրոյական վարկածին կարող է հակադրվել **երկընտրանքային վարկածը**, որը նշանակում են H_1-ով: Օրինակ, $H_0 : \mathbf{E}X = m_0$ պարզ վարկածը ստուգվում է ընդդեմ $H_1 : \mathbf{E}X = m_1$ պարզ վարկածի կամ հանդեպ $H_1 : \mathbf{E}X > m$ բարդ վարկածի կամ ընդդեմ $H_1 : \mathbf{E}X \neq m$ բարդ երկընտրանքայինի:

Վարկածի վիճակագրական ստուգումը հիմնվում է փորձարկումների արդյունքների վրա և կարող է ընդունվել նաև սխալ որոշում: Վարկածների ստուգման սխալները լինում են երկու սեռի: Եթե ստուգման արդյունքում ժխտվում է H_0 վարկածը, երբ այն իրականում ճիշտ է, ապա ասում են, որ տեղի է ունեցել **առաջին սեռի սխալ**: Եթե H_0 վարկածը ընդունվում է, երբ իրականում ճիշտ է երկընտրանքային H_1-ը, ապա տեղի է ունենում **երկրորդ սեռի սխալ**: Որոշումները ճիշտ են երկու դեպքում (տե՛ս աղյուսակը): Պայմանավորվենք նշանակել H_0-ն ընդունելու հավանականությունը, երբ այն ճիշտ է՝ $\mathbf{P}\{H_0|H_0\} = 1 - \alpha$ և նման ձևով՝ մյուս դեպքերում:

ճիշտ է / ընդունվել է	H_0	H_1
H_0	որոշումը ճիշտ է $\mathbf{P}\{H_0\|H_0\}=1-\alpha$	երկրորդ սեռի սխալ $\mathbf{P}\{H_0\|H_1\}=\beta$
H_1	առաջին սեռի սխալ $\mathbf{P}\{H_1\|H_0\}=\alpha$	որոշումը ճիշտ է $\mathbf{P}\{H_1\|H_1\}=1-\beta$

Սխալների հետևանքները տարբեր են, դրանց կարևորությունը կախված է դիտարկվող խնդրից, վարկածների առարկայական բովանդակությունից: Իհարկե, պետք է ձգտել միաժամանակ փոքրացնել սխալների հավանականությունները: Բայց երբ փոքրացնում ենք առաջին սեռի սխալի հավանականությունը, երկրորդ սեռի սխալի հավանականությունը կարող է մեծանալ: Այդ հանգամանքը մանրամասն կքննարկվի 9.2 ենթաբաժնում:

Վիճակագրական տվյալների մշակման ժամանակ առաջադրված վարկածները ըստ նրանց կիրառական բովանդակության կարելի է բաժանել մի քանի հիմնական դասերի:

1. *Վարկածներ հետազոտվող պատահական մեծության բաշխման օրենքի վերաբերյալ:* Դիտումների արդյունքների մշակման ընթացքում անհրաժեշտ է լինում ընտրել այնպիսի $F_0(x)$ բաշխման ֆունկցիա, որը «լավագույն ձևով» համապատասխանի ուսումնասիրվող X պատահական մեծությանը: Առաջանում է $H_0 : F_X(x) \equiv F(x)$ վարկածի ստուգման խնդիրը: Ընդ որում վարկածը կարող է լինել պարզ ($F(x) \equiv F_0(x)$, որտեղ $F_0(x)$-ը հայտնի ֆունկցիա է) կամ բարդ ($F(x) \equiv F(x, \theta)$, որտեղ θ-ն անհայտ պարամետր է, որը հարկավոր է գնահատել): Պարզ վարկածի դեպքում ստուգումը կատարվում է այսպես կոչված **համաձայնության հայտանիշների** միջոցով:

2. *Վարկածներ հանույթի բնութագրիչների վերաբերյալ:* Հարկ է լինում նմուշային բնութագրիչները (միջինը, ցրվածքը) համեմատել համապատասխան նորմատիվների հետ կամ երկու նմուշների բնութագրիչները համեմատել իրար հետ կամ ստուգել երկու նմուշների միևնույն կամ տարբեր հանույթներին պատկանելու վարկածը:

3. *Վարկածներ վեկտորի բաղադրիչների կախվածության մասին:* Այդպիսի վարկածներ ի հայտ են գալիս բազմաչափ հատկանիշի ուսումնասիրության դեպքում: Դրանք կդիտարկվեն հետագա գլուխներում:

4. *Վարկածներ փորձերի անկախության և պայմանների անփոփոխության մա-սին:* Արդյո՞ք դիտումները կատարվել են անկախ, միևնույն պայմաններում: Այդ հարցերի [illegible] է կախված նմուշի հետագա մշակման արդյունավետ եղանակների ընտրությունը: Համապատասխան վարկածների օրինակներ են՝ $H_0 : \mathbf{E}X_n = m, n = \overline{1, N}$, նորմալ բաշխման դեպքում $H_0 : \rho_{X_n X_{n+1}} = 0, n = \overline{1, N-1}$:

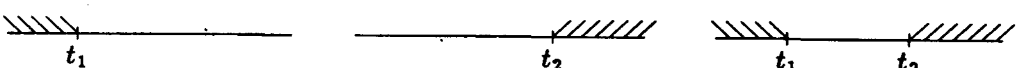
5. *Վարկածներ վիճակագրական տվյալների համասեռության մասին*: Դիցուք ունենք երկու նմուշ՝ **X** և **Y**, որոնք նկարագրում են միևնույն ընթացքը, երևույթը, և այլն, սակայն ստացված են տարբեր պայմաններում: Հարց է ծագում՝ դրանք համապատասխանում են միևնու՞յն բաշխման օրենքին, թե՞ բաշխման օրենքը փոփոխվել է: Այլ կերպ ասած, պետք է ստուգել համասեռության մասին վարկածը՝ $H_0 : F_X(x) \equiv F_Y(x)$: Նման վարկած կարելի է ձևակերպել նաև բնութագրիչների նկատմամբ՝ $H_0 : \mathbf{E}X = \mathbf{E}Y$, կամ $H_0 : \mathbf{D}X = \mathbf{D}Y$:

9.2. Ստուգման հայտանիշները և դրանց բնութագրումը

Ստուգվող վարկածի և նմուշային տվյալներին համատեղելիության մասին եզրակացությունը արվում է որոշակի հայտանիշի օգնությամբ: Ինչպես ասվել է, **X** նմուշի կամայական $T = T(\mathbf{X})$ ֆունկցիան անվանում են **վիճակախի**: T վիճակախու բաշխման օրենքը, սովորաբար, հայտնի է: Օրինակ, $H_0 : \mathbf{E}X = m$ վարկածի ստուգման համար կարելի է ընտրել կամ նմուշային միջինը՝ $T_1(\mathbf{X}) = \overline{\mathbf{X}}$, կամ կենտրոնացված նմուշային միջինը՝ $T_2(\mathbf{X}) = \overline{\mathbf{X}} - m$, կամ կանոնաձև նմուշային միջինը՝ $T_3(\mathbf{X}) = (\overline{\mathbf{X}} - m)/\sigma(\overline{\mathbf{X}})$: Մեծ N-երի դեպքում կարելի է համարել, որ այդ վիճակախիները բաշխված են նորմալ օրենքով, $T_1(\mathbf{X})$-ի դեպքում՝ m կենտրոնով, իսկ $T_2(\mathbf{X})$-ի և $T_3(\mathbf{X})$-ի դեպքերում՝ 0 կենտրոնով:

Հայտանիշը հաճախ հենվում է $T(\mathbf{X})$-ի ընդունած արժեքի վրա: $T(\mathbf{X})$ վիճակախու հնարավոր արժեքների բազմությունից առանձնացնում են այն ենթաբազմությունը, որտեղ H_0 վարկածը ժխտվում է, և անվանում են այն **կրիտիկական տիրույթ**, նրա լրացումը՝ **թույլատրելի տիրույթ**, իսկ երկու տիրույթների սահմանային կետերը՝ **կրիտիկական կետեր**:

Կրիտիկական տիրույթները լինում են **ձախակողմյան**՝ (նկար 1ա), **աջակողմյան**՝ (նկար 1բ), երկկողմանի՝ (նկար 1գ):



Նկար 1ա: Նկար 1բ: Նկար 1գ:

Առաջին սեռի սխալի հավանականության ընտրված α արժեքը անվանում են **նշանակալիության մակարդակ**: Կրիտիկական $\mathcal{T}_\alpha$ տիրույթը որոշում են հետևյալ պայմանից՝

$$\mathbf{P}\{H_1|H_0\} = \mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha \mid H_0\} = \alpha :$$

Եթե $T(\mathbf{X})$ վիճակախու արժեքը պատկանում է $\mathcal{T}_\alpha$-ին, ապա H_0 վարկածը ժխտվում է (նմուշի արդյունքները հակասում են H_0 վարկածին), եթե $T(\mathbf{X}) \notin \mathcal{T}_\alpha$, ապա H_0 վարկածը մնում է ուժի մեջ (կամ ինչպես ասում են, ընդունվում է): Ինչքան էլ փոքր լինի դրական α-ն, $\{T(\mathbf{X}) \in \mathcal{T}_\alpha\}$ պատահույթը քիչ հավանական է, բայց ոչ անհնար: Չի բացառվում, որ ճիշտ H_0 վարկածի դեպքում էլ $T(\mathbf{X})$ վիճակախու արժեքը պատկանի $\mathcal{T}_\alpha$ կրիտիկական տիրույթին, այդ դեպքում հրաժարվելով H_0 վարկածից, կատարում ենք առաջին սեռի սխալ, որի հավանականությունը α-ն է, իսկ H_0-ն հիմնավորված ընդունելու հավանականությունը $(1 - \alpha)$-ն է՝

$$\mathbf{P}\{H_0|H_0\} = \mathbf{P}\{T(\mathbf{X}) \notin \mathcal{T}_\alpha \mid H_0\} = 1 - \mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha \mid H_0\} = 1 - \alpha :$$

α-ն փոքրացնելու հետ մեկտեղ փոքրանում է $\mathcal{T}_\alpha$ կրիտիկական տիրույթը, և, հետևաբար, $T(\mathbf{X})$-ի՝ այդ տիրույթին պատկանելու հավանականությունը դառնում է փոքր նաև այն դեպքում, երբ H_0 վարկածը ճիշտ չէ: Այսինքն, α-ի *փոքրացումը մեծացնում է սխալ H_1 վարկածն ընդունելու β հավանականությունը*՝

$$\mathbf{P}\{T(\mathbf{X}) \notin \mathcal{T}_\alpha \mid H_1\} = \beta = \mathbf{P}\{H_0|H_1\} :$$

Իսկ H_0 վարկածի հիմնավորված ժխտելու հավանականությունը՝

$$\mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha \mid H_1\} = 1 - \mathbf{P}\{T(\mathbf{X}) \notin \mathcal{T}_\alpha \mid H_1\} = 1 - \beta,$$

անվանում են **հայտանիշի հզորություն**:

Հայտանիշը կոչվում է **ունակ**, եթե

$$\lim_{N\to\infty} \mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha \mid H_1\} = 1 :$$

Ունակ հայտանիշը բավականաչափ մեծ N-երի դեպքում մեկին մոտ հավանականությամբ հայտնաբերում է H_0 վարկածի սխալ լինելը:

Որոշ դեպքերում հարմար է դիտարկել այսպես կոչված **պատահականացված** (ռանդոմիզացված) **հայտանիշներ**, երբ համարվում է, որ դիտումների $\mathbf{X}$ նմուշը հակասում է H_0 վարկածին $\varphi(\mathbf{X})$ հավանականությամբ: $\varphi(\mathbf{X})$ ֆունկցիան $(0 \le \varphi(\mathbf{X}) \le 1)$ անվանում են **կրիտիկական ֆունկցիա**: Ոչ պատահականացված T հայտանիշին համապատասխանում է $\mathcal{T}_\alpha$ բազմության հայտիչը՝

$$\varphi(\mathbf{X}) = \begin{cases} 0, & \text{եթե} \quad \mathbf{X} \notin \mathcal{T}_\alpha, \\ 1, & \text{եթե} \quad \mathbf{X} \in \mathcal{T}_\alpha : \end{cases}$$

Այժմ անդրադառնանք մի ավելի նեղ, բայց հաճախ հանդիպող, վարկածների դասին: Երբ հատկանիշի (X պատահական մեծության) ենթադրվող բաշխումների $\mathcal{F}$ բազմությունը տրված է պարամետրական ձևով՝ $\mathcal{F} = \{F(x,\theta),\ \theta \in \Theta\}$, ապա դատողությունները փաստորեն վերաբերում են θ անհայտ պարամետրի արժեքին և անվանվում են **պարամետրական վարկածներ**:

Հիմնական պարամետրական H_0 և H_1 վարկածները ձևակերպվում են հետևյալ կերպ $H_0 :\ \theta \in \Theta_0, \quad H_1 :\ \theta \in \Theta_1\ (\Theta_1 = \Theta - \Theta_0)$:

Տրված α նշանակալիության մակարդակով H_0 վարկածի ստուգման հայտանիշը որոշվում է $T(\mathbf{X})$ վիճականիով և այնպիսի $\mathcal{T}_\alpha$ կրիտիկական տիրույթով, որի դեպքում

$$\sup_{\theta\in\Theta_0} \mathbf{P}_\theta\{T(\mathbf{X}) \in \mathcal{T}_\alpha\} \le \alpha :$$

Սխալների հավանականությունների կախվածությունը θ-ից արտահայտում են

$$\alpha_T(\theta) = \mathbf{P}_\theta\{T(\mathbf{X}) \in \mathcal{T}_\alpha\},\ \theta \in \Theta_0, \quad \beta_T(\theta) = \mathbf{P}_\theta\{T(\mathbf{X}) \notin \mathcal{T}_\alpha\},\ \theta \in \Theta_1,$$

ֆունկցիաները: $(1-\beta_T(\theta))$ ֆունկցիան անվանում են **հայտանիշի հզորության ֆունկցիա**:

Հայտանիշը կոչվում է **անշեղ**, եթե տվյալ α-ի համար

$$\alpha_T(\theta) \le \alpha, \quad \theta \in \Theta_0, \quad \text{և} \quad \alpha_T(\theta) > \alpha, \quad \theta \in \Theta_1 :$$

Պարամետրական հայտանիշը անվանում են **ունակ**, եթե

$$\lim_{N\to\infty} \alpha_T(\theta) = 1, \quad \theta \in \Theta_1 :$$

Դիցուք $T(\mathbf{X})$-ով և $T^*(\mathbf{X})$-ով որոշվում են միևնույն վարկածի α նշանակալիության մակարդակով H_0 ստուգման երկու հայտանիշներ:

Եթե գոնե մեկ θ_1-ի դեպքում $(\theta_1 \in \Theta_1)$ միաժամանակ՝

$$\alpha_{T^*}(\theta_0) \le \alpha_T(\theta_0), \ \theta_0 \in \Theta_0, \quad \alpha_{T^*}(\theta_1) \ge \alpha_T(\theta_1), \quad \theta_1 \in \Theta_1,$$

ապա ասում են, որ $T^*(\mathbf{X})$ **հայտանիշը ավելի հզոր** է, (հետևաբար, այն նախընտրելի է): Եթե նշված անհավասարությունները տեղի ունեն կամայական $T(\mathbf{X})$-ի համար, ապա $T^*(\mathbf{X})$ **հայտանիշը** անվանում են **հավասարաչափ ամենահզոր**: Անշեղ և ամենահզոր **հայտանիշը** անվանում են **լավագույն**:

Եթե երկընտրանքային H_1 վարկածը բարդ է, իսկ $T(\mathbf{X})$ ամենահզոր հայտանիշը մնում է *նույնը բոլոր երկընտրանքային $\theta \in \Theta_1$ վարկածների համար*, ապա **հայտանիշը** θ**-ի նկատմամբ** անվանում են **հավասարաչափ ամենահզոր**:

Որոշ դեպքերում հարկավոր է ընտրել (տվյալ խնդրում) **օպտիմալ հայտանիշը**՝ այն, որի հզորության ֆունկցիան ստանում է մեծագույն արժեքը (Նեյմանի–Պիրսոնի խնդիրը): Օպտիմալ հայտանիշի ընտրությունը հաճախ հնարավոր է կատարել անշեղ հայտանիշների դասի համար:

Հիշեցնենք, որ $L(\mathbf{X},\theta) = \prod_{n=1}^{N} f_X(X_n,\theta)$ վիճականին կոչվում է **ճշմարտանմանության ֆունկցիա** (տե՛ս գլուխ 7): Անվանենք $l(\mathbf{X}) = L(\mathbf{X},\theta_1)/L(\mathbf{X},\theta_0)$ ֆունկցիան **ճշմարտանմանության հարաբերություն** (այն իմաստ ունի, եթե $L(\mathbf{X},\theta_0) \neq 0$):

Օպտիմալ հայտանիշների կառուցման եղանակների մեծամասնությունը հիմնվում է հետևյալ կարևոր արդյունքի վրա, որը կոչվում է **Նեյմանի–Պիրսոնի լեմմա**:

> $H_0 : \theta = \theta_0$ և երկընտրանքային $H_1 : \theta = \theta_1$ պարզ վարկածների ստուգման կամայական α նշանակալիության մակարդակին համապատասխանող ամենահզոր հայտանիշը գոյություն ունի և տրվում է հետևյալ կրիտիկական տիրույթով՝
>
> $$\mathcal{T}^*_\alpha = \{l(\mathbf{X}) : \ l(\mathbf{X}) \geq c_\alpha\} :$$

Երբ պարզ $H_0 : \theta = \theta_0$ վարկածը ստուգվում է ընդդեմ $H_1 : \theta \in \Theta - \{\theta_0\}$ վարկածի, հավասարաչափ ամենահզոր հայտանիշը գոյություն ունի, եթե Նեյմանի–Պիրսոնի լեմմայի մեջ նշված $\mathcal{T}^*_\alpha$ կրիտիկական տիրույթը կախված չէ θ-ի արժեքից:

Վարկածների ստուգման գործընթացը սովորաբար բաղկացած է հետևյալ փուլերից:

1. Ելնելով տնտեսագիտական կամ այլ կիրառական նպատակներից՝ ձևակերպվում է հիմնական H_0 վարկածը: Միաժամանակ ձևակերպվում է երկընտրանքային H_1 վարկածը:

2. Ընտրվում է α նշանակալիության մակարդակը: Սովորաբար α-ին վերագրում են $0.1, 0.05, 0.01, 0.005, 0.001$ արժեքները:

3. Կառուցվում է փորձերի $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ արդյունքներից կախված $T = T(\mathbf{X})$ վիճականին և, եթե հնարավոր է, որոշվում են դրա բաշխման ֆունկցիաները երկու՝ H_0 և H_1 վարկածների պայմաններում:

4. Գտնվում է $\mathcal{T}_\alpha$ կրիտիկական տիրույթը հետևյալ պայմանից՝ $\mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha \mid H_0\} = \alpha$: Նշենք, որ այդ պայմանից $\mathcal{T}_\alpha$-ն որոշվում է ոչ միարժեք, քանի որ t առանցքի վրա կան բազմաթիվ միջակայքեր, որոնց վրա կառուցված կորագիծ սեղանների մակերեսները հավասար են $(1-\alpha)$-ի: Առաջադրվում է նաև հետևյալ պայմանը՝ կրիտիկական $\mathcal{T}_\alpha$ տիրույթը պետք է լինի այնպիսին, որ երկրորդ սեռի սխալի β հավանականությունը լինի փոքրագույնը: Կամայական H_1 երկընտրելի վարկածից կախված կրիտիկական տիրույթը որոշելու համար գտնում են մեկ կրիտիկական կետ, եթե H_1-ը միակողմանի է, և երկու կետ, եթե H_1-ը երկկողմանի է, տե՛ս երեք դեպքերը նկար 2-ում՝

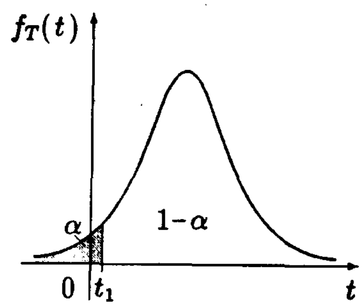
ա) ձախակողմյան կրիտիկական տիրույթ՝ $\mathcal{T}_\alpha = (-\infty, t_1)$, $\mathbf{P}\{T(\mathbf{X}) < t_1\} = \alpha$,
բ) աջակողմյան կրիտիկական տիրույթ՝ $\mathcal{T}_\alpha = (t_2, \infty)$, $\mathbf{P}\{T > t_2\} = \alpha$,
գ) երկկողմանի կրիտիկական տիրույթ՝

$$\mathcal{T}_\alpha = (-\infty, t_1) \cup (t_2, \infty), \ \mathbf{P}\{T < t_1\} = \alpha_1, \ \mathbf{P}\{T > t_2\} = \alpha_2, \ \alpha_1 + \alpha_2 = \alpha,$$

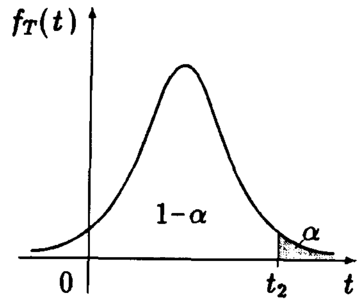
հաճախ գերադասելի է վերցնել $\alpha_1 = \alpha_2 = \alpha/2$:

5. Կատարվում է նմուշահանում, և նրա արդյունքների հիման վրա հաշվվում է $T(\mathbf{X})$-ի արժեքը՝ $\hat{t}$-ն:

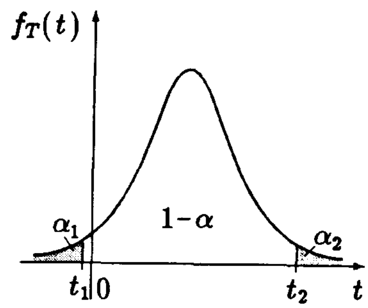
6. Եզրակացություն՝ եթե $\hat{t}$-ն պատկանում է թույլատրելի տիրույթին, ապա H_0-ն մնում է ուժի մեջ՝ փորձի տվյալները չեն հակասում H_0-ին: Հակառակ դեպքում H_0-ն ժխտվում է:



Նկար 2ա:

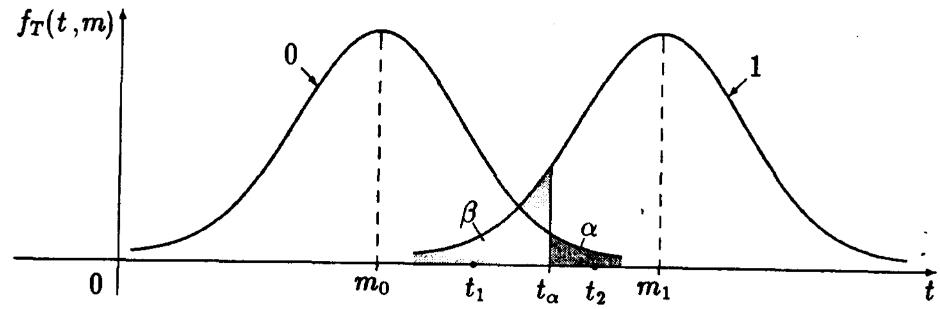


Նկար 2բ:



Նկար 2գ:

Առաջին և երկրորդ սեռի սխալների կապը: Դիցուք $X \sim \mathcal{N}(m,\sigma^2)$ հատկանիշի միջինը անհայտ է, իսկ ցրվածքը՝ σ^2-ին հայտնի է: Ենթադրենք, որ $H_0 : \mathbf{E}X = m_0$ վարկածը ստուգվում է ընդդեմ երկընտրանքային $H_1 : \mathbf{E}X = m_1$ $(m_1 > m_0)$ վարկածի: Ստուգման համար ընտրենք T վիճականին՝ $T(\mathbf{X}) = \overline{\mathbf{X}}$: $\overline{\mathbf{X}}$-ը նորմալ է բաշխված, ընդ որում H_0 վարկածի ճիշտ լինելու դեպքում $\mathbf{E}\overline{\mathbf{X}} = m_0$, $\sigma(\overline{\mathbf{X}}) = \sigma/\sqrt{N}$ (տե՛ս գլուխ 7): Գծապատկեր 3-ի վրա այդ բաշխումը նկարագրվում է «0» խտության կորով:



Նկար 3:

Ըստ ընտրված α-ի գտնում են կրիտիկական t_α կետը՝

$$\mathbf{P}\{T(\mathbf{X}) > t_\alpha \mid H_0\} = \alpha :$$

Եթե H_1 վարկածը ճիշտ է, ապա $T(\mathbf{X}) = \overline{\mathbf{X}} \sim \mathcal{N}(m_1,\sigma^2/N)$, տե՛ս նկար 3-ում «1» խտության կորը: Եթե այդ երկու կորերը համարյա չեն հատվում, ապա $\overline{\mathbf{X}}$ նմուշային արժեքի «0» կորի տակ ընկնելը վկայում է H_0-ի ճիշտ լինելու մասին, իսկ «1» կորի տակ՝ H_1-ի ճիշտ լինելու մասին: Երբ m_0-ն և m_1-ը համեմատաբար մոտ են իրար, ապա կորերը հատվում են զգալի չափով: Եթե $\overline{\mathbf{X}}$-ի նմուշային արժեքը t_2-ն է (t_α-ից աջ), H_0 վարկածը ժխտվում է α առաջին սեռի սխալի հավանականությամբ: Եթե $\overline{\mathbf{X}}$-ի նմուշային արժեքը t_1-ն է (t_α-ից ձախ), ապա H_0-ն ընդունվում է, թեպետ կարող է ճիշտ լինել H_1 վարկածը: Այսպիսով, կարող է կատարվել երկրորդ սեռի սխալ, որի β հավանականությունը կարելի է որոշել «1» խտության կորի օգնությամբ՝

$$\beta = \mathbf{P}\{\overline{\mathbf{X}} \le t_\alpha \mid H_1\} :$$

Գծապատկերից երևում է, որ եթե t_α կետը տանենք աջ, α-ն կփոքրանա, բայց α-ի փոքրանալու հետ մեկտեղ կմեծանա β-ի արժեքը:

α-ն և β-ն միաժամանակ փոքրացնելու միակ հնարավորությունը N-ի մեծացնելն է, քանի որ այդ դեպքում փոքրանում է T(**X**)-ի բաշխման σ^2/N ցրվածքը:

Դիտարկենք մի քանի տեսական օրինակներ:

<u>*Օրինակ*</u> 1: Դիցուք $X \sim \mathcal{N}(m, \sigma^2)$, σ^2-ն հայտնի է:

ա) Օգտագործելով ճշմարտանմանության հարաբերության հայտանիշը, ստուգել $H_0 : m = m_0$ վարկածը ընդդեմ $H_1 : m = m_1$, $(m_1 > m_0)$ վարկածի:

բ) Որքա՞ն մեծ պետք է լինի նմուշի ծավալը, որպեսզի տրված α-ի դեպքում երկրորդ սեռի սխալի հավանականությունը փոքր լինի β-ից:

Լուծում: ա) **X** նմուշի ճշմարտանմանության հարաբերությունն է՝

$$l(\mathbf{X}) = \frac{L(\mathbf{X}, m_1)}{L(\mathbf{X}, m_0)} = \frac{f^{(N)}(\mathbf{X}; m_1, \sigma)}{f^{(N)}(\mathbf{X}; m_0, \sigma)} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}((X_n - m_1)^2 - (X_n - m_0)^2)\right\} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(-2X_n(m_1 - m_0) + (m_1^2 - m_0^2))\right\}:$$

Կրիտիկական տիրույթը որոշվում է $\mathbf{P}\{l(\mathbf{X}) > c_\alpha\} = \alpha$ հավասարումից: Լոգարիթմելով, կստանանք՝

$$-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(-2X_n(m_1 - m_0) + (m_1^2 - m_0^2)) \ge \ln c_\alpha$$

կամ

$$2(m_1 - m_0)\sum_{n=1}^{N} X_n \ge 2\sigma^2 \ln c_\alpha + N(m_1^2 - m_0^2):$$

Ուրեմն՝

$$\overline{\mathbf{X}} \ge \sigma^2 \ln c_\alpha / N(m_1 - m_0) + (m_1 + m_0)/2 = C_\alpha,$$

որտեղից երևում է, որ ճշմարտանմանության հարաբերության հայտանիշը հիմնված է նմուշային միջինի վրա: Աջ մասը պարունակում է անհայտ c_α, այդ պատճառով աջ մասը կնշանակենք C_α-ով, որը նույնպես անհայտ է: Քանի որ $\overline{\mathbf{X}} \sim \mathcal{N}(m_0, \sigma^2/N)$, ապա

$$\mathbf{P}\{\overline{\mathbf{X}} \ge C_\alpha | m = m_0\} = 1/2 - \Phi\left((C_\alpha - m_0)\sqrt{N}/\sigma\right) = \alpha:$$

Եթե նշանակենք $(C_\alpha - m_0)\sqrt{N}/\sigma = z_\alpha$, ապա z_α-ն $\Phi(z) = 1/2 - \alpha$ հավասարման լուծումն է:

$C_\alpha = m_0 + z_\alpha \sigma/\sqrt{N}$ և, հետևաբար, $T_\alpha^* = \{\mathbf{X} : \dfrac{L(\mathbf{X}, m_1)}{L(\mathbf{X}, m_0)} \ge c_\alpha\}$, այսպիսով հայտանիշը որոշված է: H_1 վարկածի դեպքում $\overline{\mathbf{X}} \sim \mathcal{N}(m_1, \sigma^2/N)$, և երկրորդ սեռի սխալի հավանականությունը՝ $\beta = 1/2 + \Phi\left((C_\alpha - m_1)\sqrt{N}/\sigma\right)$, իսկ հզորությունը՝ $1 - \beta = 1/2 - \Phi\left((C_\alpha - m_1)\sqrt{N}/\sigma\right)$: Տեղադրելով C_α-ի արժեքը, կստանանք

$$1 - \beta = 1/2 - \Phi\left(-(m_1 - m_0)\sqrt{N}/\sigma + z_\alpha\right):$$

Քանի որ $(m_1 - m_0)/\sigma$ մեծությունը բնութագրում է H_0 և H_1 վարկածների միջև եղած «հեռավորությունը», ապա նույն հայտանիշի դեպքում H_0-ի և H_1-ի միջև «հեռավորությունը» մեծանալիս մեծանում է հզորությունը: Ստացված հայտանիշը յուրահատուկ է նրանով, որ երբ $m_1 > m_0$, այն կախում չունի m_1-ից: Հետևաբար, միակողմանի հայտանիշը, որը ժխտվում է, երբ $\overline{\mathbf{X}} \ge C_\alpha = m_0 + z_\alpha \sigma/\sqrt{N}$, հավասարաչափ ամենահզորն է (երբ երկընտրանքային H_1 վարկածը նկարագրվում է նորմալ բաշխումով, և $m_1 < m_0$, ապա այն կժխտվի, երբ $\overline{\mathbf{X}} < -C_\alpha$): Ստացված արդյունքը ցույց է տալիս, որ հայտանիշը լավագույնն է նաև բարդ H_1 վարկածի համար, երբ դիտումների միջինը $m \in (m_1, \infty)$, քանի որ $\overline{\mathbf{X}} \ge m_0 + z_\alpha \sigma/\sqrt{N}$ անհավասարությունը m_1 չի պարունակում:

Սակայն նշենք, որ եթե α-ն փոքր է, ապա ճիշտ H_1 վարկածը ժխտելու հավանականությունը կարող է բավականին մեծ լինել:

բ) Եթե կանոնաձև նորմալ բաշխման p-քանորդիչը նշանակենք u_p-ով, ապա, քանի որ $F(-z_\alpha) = \alpha$ և $F\left(z_\alpha - (m_1 - m_0)\sqrt{N}/\sigma\right) = \beta$, կստանանք $-z_\alpha = u_\alpha$ և $z_\alpha - (m_1 - m_0)\sqrt{N}/\sigma = u_\beta$, որտեղից հետևում է, որ, որպեսզի տրված α-ի համար երկրորդ սեռի սխալի հավանականությունը հավասար լինի

β-ին, N-ը պետք է լինի ոչ պակաս, քան՝

$$N_0 = \left[\frac{(u_\alpha + u_\beta)^2\sigma^2}{(m_1 - m_0)^2}\right] + 1 :$$

Օրինակ, եթե $\alpha = \beta = 0.05$ և $(m_1 - m_0)\sigma = 0.1$, ապա կստանանք $N_0 = 1076$:

Օրինակ 2: Դիցուք $X \sim \mathcal{N}(0, \sigma^2)$: $\mathbf{X}$ ընտրքի հիման վրա α նշանակալիության մակարդակով ստուգվում է $H_0 : \sigma = \sigma_0$ վարկածը ընդդեմ $H_1 : \sigma = \sigma_1$, $(\sigma_1 > \sigma_0)$ վարկածի:

Լուծում: Ճշմարտանմանության հարաբերությունն է՝

$$l(\mathbf{X}) = \frac{L(\mathbf{X}, \sigma_1)}{L(\mathbf{X}, \sigma_0)} = \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{n=1}^{N} X_n^2\right\} :$$

H_0 վարկածը ժխտվում է, եթե $l(\mathbf{X}) \geq c_\alpha$, որտեղ c_α-ն որոշվում է $\mathbf{P}\{l(\mathbf{X}) \geq c_\alpha\} = \alpha$ հավասարումից: Լոգարիթմելով $l(\mathbf{X}) \geq c_\alpha$ անհավասարությունը, կստանանք

$$N \ln\frac{\sigma_0}{\sigma_1} + \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_n X_n^2 \geq \ln c_\alpha,$$

որտեղից $\sum\limits_n X_n^2 \geq \dfrac{2(\ln c_\alpha - N\ln(\sigma_0/\sigma_1))}{(1/\sigma_0^2) - (1/\sigma_1^2)} = C_\alpha$: Գտնելով C_α-ն, կգտնենք c_α-ն: Հայտնի է, որ $(\sum\limits_n X_n^2)/\sigma^2 \sim \chi^2(N)$: Կրիտիկական տիրույթը կորոշվի $\mathbf{P}\{(\sum\limits_n X_n^2)/\sigma_0^2 \geq C_\alpha/\sigma_0^2\} = \alpha$ պայմանից, $\chi^2(N)$-բաշխման աղյուսակից, հետևաբար, $C_\alpha = \sigma_0^2\chi_{1-\alpha}^2$:

Երկրորդ սեռի սխալը կլինի՝ $\beta = \mathbf{P}\{(\sum\limits_n X_n^2)/\sigma_1^2 < C_\alpha/\sigma_1^2\}$:

Օրինակ 3: Դիցուք Բեռնուլիի N փորձերում $\mathcal{A}$ պատահույթը տեղի է ունեցել X անգամ: $\mathcal{A}$ պատահույթի մեկ փորձում հանդես գալու p հավանականության նկատմամբ դիտարկենք երկու վարկած՝ $H_0 : p = p_0$, $H_1 : p = p_1$, $0 < p_0 < p_1 < 1$: Գտնել H_0 վարկածի α նշանակալիության մակարդակով ստուգման օպտիմալ հայտանիշը:

Լուծում: Ելնելով հետևյալ անհավասարությունից և Նեյման–Պիրսոնի լեմմայից՝

$$l(X) = \frac{L(X, p_1)}{L(X, p_0)} = \left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]^X \frac{(1-p_1)^N}{(1-p_0)^N} \geq c_\alpha,$$

լոգարիթմելով, կստանանք՝

$$X \ln\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right] \geq \ln c_\alpha + N\ln\left(\frac{1-p_0}{1-p_1}\right),$$

կամ

$$X \geq C_\alpha, \quad C_\alpha = \left[\ln c_\alpha + N\ln\left(\frac{1-p_0}{1-p_1}\right)\right] \Big/ \ln\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right] :$$

Քանի որ X-ը ասիմպտոտորեն նորմալ է բաշխված՝ $X \sim \mathcal{N}(Np, Np(1-p))$, ապա առաջին և երկրորդ սեռի սխալների հավանականությունները կլինեն, համապատասխանաբար՝

$$\alpha = \mathbf{P}\{X \geq C_\alpha | H_0\} = \mathbf{P}\left\{\frac{X - Np_0}{\sqrt{Np_0(1-p_0)}} \geq \frac{C_\alpha - Np_0}{\sqrt{Np_0(1-p_0)}}\right\},$$

$$\beta = \mathbf{P}\{X < C_\alpha | H_1\} = \mathbf{P}\left\{\frac{X - Np_1}{\sqrt{Np_1(1-p_1)}} < \frac{C_\alpha - Np_1}{\sqrt{Np_1(1-p_1)}}\right\} :$$

Նորմալ բաշխման $z_{1-\alpha}$ քանորդչի միջոցով ստանում ենք՝ $C_\alpha = Np_0 + z_{1-\alpha}\sqrt{Np_0(1-p_0)}$:

Օրինակ 4: Դիցուք X_i-ն i-րդ ամսում հիմնարկից ազատված աշխատողների թիվն է, $i = \overline{1,12}$: Պահանջվում է ստուգել հետևյալ վարկածը՝ միջին հաշվով ամսական հիմնարկից ազատվում է մեկ աշխատող: Ենթադրենք, որ տվյալ հիմնարկի համար այդ թիվը հոսունության նորմական մակարդակն է: Կարևոր է կանխել այն պահը, երբ հոսունությունը կավելանա 2.5 անգամ: Բազմաքիվ ստուգումները ցույց են տվել, որ

X_n, $n = \overline{1,N}$ պատահական մեծությունները անկախ են և ունեն Պուասոնի բաշխում θ հոսունության ինտենսիվությամբ: Հիմնական վարկածը՝ $H_0 : \theta = 1$ ստուգվում է ընդդեմ $H_1 : \theta = 2.5$ վարկածի:

Լուծում: Կազմենք ճշմարտանմանության հարաբերությունը՝

$$l(\mathbf{X}) = \frac{L(\mathbf{X},\theta_1)}{L(\mathbf{X},\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^{\sum_n X_n} e^{-(\theta_1-\theta_0)N} :$$

Կրիտիկական տիրույթը որոշվում է $\mathbf{P}\{l(\mathbf{X}) \geq c_\alpha\} = \alpha$ պայմանից: Լոգարիթմելով, կստանանք՝

$$\sum_n X_n \cdot \ln(\theta_1/\theta_0) - (\theta_1 - \theta_0)N \geq \ln c_\alpha,$$

$$\sum_n X_n \geq (\ln c_\alpha + (\theta_1 - \theta_0)N)/\ln(\theta_1/\theta_0) = C_\alpha :$$

Անհայտ c_α-ն կորոշենք α նշանակալիության մակարդակի միջոցով: Քանի որ $X_n \sim \Pi(\theta)$, ապա այդպիսի անկախ N պատահական մեծությունների գումարը՝ $\sum_n X_n \sim \Pi(N\theta)$: Եթե H_0-ն ճիշտ է, $\theta = \theta_0 = 1$ և $N = 12$, ապա

$$\mathbf{P} = \{T \in \mathcal{T}_\alpha \mid H_0\} = \sum_{k \geq C_\alpha} \frac{(N\theta_0)^k}{k!} e^{-N\theta_0} = \sum_{k \geq C_\alpha} \frac{(12)^k}{k!} e^{-12} = \alpha :$$

Վարկածների ստուգման հայտանիշի և վստահության միջակայքի կառուցման կապը: Որոշ դեպքերում հարմար է օգտվել այն հանգամանքից, որ θ պարամետրի նկատմամբ պարզ վարկածի ստուգման խնդիրը սերտորեն կապված է θ **պարամետրի վստահության միջակայքի** կառուցման խնդրի հետ:

Կարելի է ցույց տալ, որ θ-ի համար $(1 - \alpha)$ հավանականությամբ վստահության միջակայքը α նշանակալիության մակարդակով $H_0 : \theta = \theta_0$ վարկածի ստուգման թույլատրելի տիրույթ է:

Ճիշտ է նաև հակառակը՝ եթե որոշենք $H_0 : \theta = \theta_0$ վարկածի α նշանակալիության մակարդակին համապատասխանող ստուգման երկկողմանի կրիտիկական տիրույթը, ապա նրա լրացումը կլինի θ պարամետրի $(1 - \alpha)$ հավանականությամբ վստահության միջակայքը:

Այսպիսով, եթե որևէ վիճակագրական մոդելի համար հայտնի է այդ խնդիրներից մեկի լուծումը, ապա դրա օգնությամբ կարելի է գտնել մյուսի լուծումը, ընդ որում հավասարաչափ ամենահզոր հայտանիշին կհամապատասխանի ամենակարճ վստահության միջակայքը, և հակառակը:

Վարկածների ստուգման հայտանիշների կապը վստահության միջակայքերի կառուցման հետ ցուցադրենք օրինակներով:

<u>*Օրինակ*</u> 5: Դիցուք $X \sim \mathcal{N}(m, \sigma^2)$, m-ը հայտնի է: Պահանջվում է N ծավալի $\mathbf{X}$ նմուշի հիման վրա ստուգել $H_0 : \sigma^2 = \sigma_0^2$ վարկածը ընդդեմ $H_1 : \sigma^2 \neq \sigma_0^2$ վարկածի:

Լուծում: H_1 վարկածի ստուգման համար կունենանք (տե՛ս օրինակ 2) երկկողմանի կրիտիկական տիրույթը՝

$$\sum_n (X_n - m)^2 \leq C_{\alpha/2} \text{ և } \sum_n (X_n - m)^2 \geq C_{1-\alpha/2} :$$

Թույլատրելի բազմությունը կունենա $C_{\alpha/2} < \sum_n (X_n - m)^2 < C_{1-\alpha/2}$ տեսքը: Համեմատելով ստացված անհավասարությունը անհայտ ցրվածքի վստահության միջակայքի հետ (տե՛ս ութերորդ գլխի օրինակ 6-ը)՝

$$N\hat{s}_0^2(\mathbf{X})/\chi^2_{1-\alpha/2}(N) < \sigma_0^2 < N\hat{s}_0^2(\mathbf{X})\chi^2_{\alpha/2}(N),$$

կստանանք՝

$$C_{\alpha/2} = \sigma_0^2 \chi^2_{\alpha/2}(N), \quad C_{1-\alpha/2} = \sigma_0^2 \chi^2_{1-\alpha/2}(N) :$$

Նշենք, որ եթե անհայտ θ պարամետրի վստահության $(\hat{\theta}_1(x), \hat{\theta}_2(x))$ միջակայքը գտնված է առավելագույն ճշմարտանմանության եղանակով, ապա այն տալիս է $H_0 : \theta = \theta_0$ վարկածի ընդդեմ $H_1 : \theta \neq \theta_0$ վարկածի ստուգման ամենալավ հայտանիշը:

Օրինակ 6: Դիտարկենք $\mathcal{N}(m, \sigma^2)$ մոդելը, որում σ^2-ն հայտնի է: Ստուգել $H_0 : m = m_0$ վարկածը ընդդեմ ա) $H_1 : m \neq m_0$ վարկածի և բ) $H_1 : m > m_0$ վարկածի:

Լուծում: Ստուգման օպտիմալ (հավասարաչափ ամենահզոր, անշեղ) $T(\mathbf{X})$ հայտանիշի կրիտիկական տիրույթն է՝

ա) $T(\mathbf{X}) = \sqrt{N}|\overline{\mathbf{X}} - m_0|/\sigma$, $\mathbf{P}\{T(\mathbf{X}) \in \mathcal{T}_\alpha\} = 1 - \alpha$, $\mathbf{P}\{T(\mathbf{X}) \geq z_{1-\alpha/2}\} = \alpha/2$:

Մյուս կողմից, վստահության միջակայքն ունի հետևյալ տեսքը (տե՛ս գլուխ 8, օրինակ 1)՝

$$\{\sqrt{N}|\overline{\mathbf{X}} - m|/\sigma < z_{1-\alpha/2}\} = \{\overline{\mathbf{X}} - \sigma z_{1-\alpha/2}/\sqrt{N} < m < \overline{\mathbf{X}} + \sigma z_{1-\alpha/2}/\sqrt{N}\},$$

$$\Phi(z_{1-\alpha/2}) = (1-\alpha)/2 :$$

Հետևաբար $(\overline{\mathbf{X}} \pm \sigma z_{1-\alpha/2}/\sqrt{N})$ միջակայքը բոլոր $(1-\alpha)$ հավանականությամբ վստահության միջակայքերի միջև ամենակարճն է:

բ) Քանի որ $H_1 : m > m_0$ վարկածի դեպքում հավասարաչափ ամենահզոր հայտանիշը տրվում է $\mathcal{T}_\alpha = \{t : T(\mathbf{X}) \geq z_{1-\alpha}\}$ կրիտիկական տիրույթով, որտեղ $T(\mathbf{X}) = \sqrt{N}(\overline{\mathbf{X}} - m_0)/\sigma$, ապա միակողմանի վստահության միջակայքն է $\{m > \overline{\mathbf{X}} - \sigma z_{1-\alpha}/\sqrt{N}\}$:

9.3. Վարկածներ որոշակի հատկության առկայության հավանականության վերաբերյալ

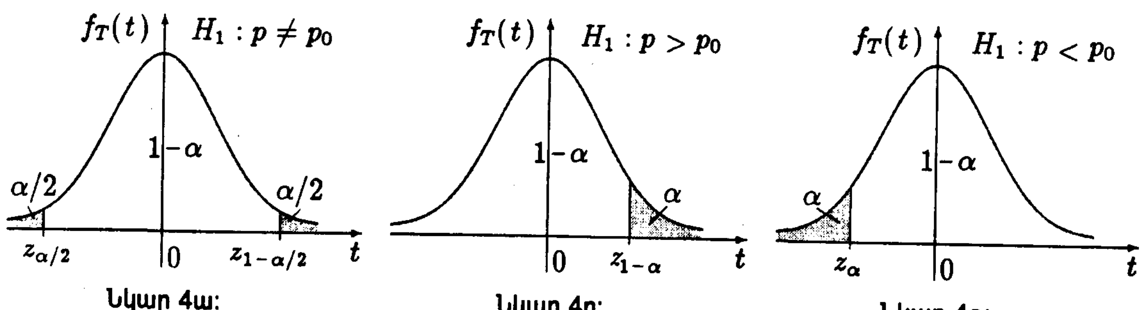
Հատկության առկայության հավանականության համեմատումը նորմատիվի հետ: Դիցուք պետք է ստուգել հանուրի որոշ հատկության առկայության p հավանականության p_0 արժեքին հավասար լինելու վարկածը՝ $H_0 : p = p_0$, ընդդեմ երկընտրանքայինի՝ $H_1 : p \neq p_0$ վարկածի: Հայտանիշի կառուցման համար վերցնենք $T(\mathbf{X}) = n/N$ վիճականին (տե՛ս գլուխ 7), որը p հավանականության ունակ, անշեղ և արդյունավետ գնատու է: Մեծ N-երի դեպքում ասիմպտոտական բաշխումը կլինի նորմալ, հետևաբար, ընտրելով α նշանակալիության մակարդակը, կգտնենք $z_{1-\alpha/2}$-ը (նկար 4ա) հետևյալ հավասարումից՝

$$\mathbf{P}\left\{|n/N - p_0| \leq \sigma(n/N) z_{1-\alpha/2}\right\} = 2\Phi(z_{1-\alpha/2}) = 1 - \alpha :$$

Տեղադրելով $\sigma(n/N) = \sqrt{p_0(1-p_0)/N}$, կստանանք (տե՛ս նաև գլուխ 8, օրինակ 9) p_0-ի վստահության միջակայքը: Այսպիսով, $T(\mathbf{X})$ վիճականու թույլատրելի բազմությունն է՝

$$\left[p_0 - z_{1-\alpha/2}\sqrt{p_0(1-p_0)/N}, p_0 + z_{1-\alpha/2}\sqrt{p_0(1-p_0)/N}\right] :$$

Եթե նմուշային n/N հաճախությունը այդ միջակայքից է, ապա H_0-ն մնում է ուժի մեջ, հակառակ դեպքում ժխտվում է:



Նկար 4ա: Նկար 4բ: Նկար 4գ:

$H_1 : p > p_0$ վարկածի դեպքում (նկար 4բ)՝

$$\mathbf{P}\left\{n/N > p_0 + z_{1-\alpha}\sqrt{p_0(1-p_0)/N}\right\} = 1/2 - \Phi(z_{1-\alpha}) = \alpha :$$

Այսինքն՝ կրիտիկական տիրույթն է՝ $\left[p_0 + z_{1-\alpha}\sqrt{p_0(1-p_0)/N}, +\infty\right)$:

$H_1 : p < p_0$ վարկածի դեպքում (նկար 4գ)՝

$$\mathbf{P}\left\{n/N < p_0 + z_\alpha\sqrt{p_0(1-p_0)/N}\right\} = 1/2 + \Phi(z_\alpha) = \alpha :$$

Կրիտիկական տիրույթն ունի $(-\infty, p_0 + z_\alpha\sqrt{p_0(1-p_0)/N}]$ տեսքը:

Օրինակ 7: 1200 փորձարկումների ընթացքում A պատահույթը տեղի է ունեցել 380 անգամ: Հնարավո՞ր է $\alpha = 0.05$ նշանակալիության մակարդակով պնդել, թե պատահույթի տեղի ունենալու p հավանականությունը հավասար է 0.3:

Լուծում: 1. $H_0 : p = p_0 = 0.3$, $H_1 : p \neq p_0$:

2. $\alpha = 0.05$:

3. $T(\mathbf{X}) = \dfrac{n/N - p_0}{\sigma(n/N)} \sim \mathcal{N}(0,1)$ քանի որ N-ը մեծ է:

4. Կրիտիկական տիրույթը երկկողմանի է՝

$$\mathbf{P}\left\{\left|\frac{n/N - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{N}\right| \le z_{1-\alpha/2}\right\} = 1 - 0.05 = 2\Phi(z_{1-\alpha/2}) - 1 :$$

Աղյուսակից գտնում ենք ՝ $z_{1-\alpha/2} = 1.96$:

5. Գտնում ենք $T(\mathbf{X})$-ի նմուշային արժեքը՝ $\hat{t} = ((380/1200) - 0.3)/\sqrt{0.3 \cdot 0.7/1200} = 1.26$:

6. Քանի որ $-1.96 < 1.26 < 1.96$, $T(\mathbf{X})$-ի թվային արժեքը պատկանում է թույլատրելի բազմությանը, H_0 վարկածը թողնում ենք ուժի մեջ:

Երկու հանույթներում միևնույն հատկանիշի առկայության հավանականությունների համեմատումը: Դիցուք միևնույն հատկանիշի հարաբերական հաճախությունները N_1 և N_2 ծավալներ ունեցող $\mathbf{X}$ և $\mathbf{Y}$ անկախ նմուշներում հավասար են, համապատասխանաբար, n_1/N_1 և n_2/N_2: Ստուգվում է H_0 վարկածը՝ այդ երկու նմուշները վերցված են միևնույն հանույթից, որում հատկանիշի առկայության հավանականությունը հավասար է p-ի (հարաբերական հաճախությունների տարբեր լինելը պատահական նմուշահանման արդյունք է): Նմուշի մեծ և փոքր դեպքերում ստուգումը տարբեր է:

Մեծ նմուշներ: Եթե N_1-ը և N_2-ը մեծ են ($\ge$ 30), նմուշային հարաբերական հաճախության բաշխումը մոտ է նորմալ բաշխմանը, և $H_0 : p_1 = p_2 = p$ վարկածի դեպքում ունենք հետևյալ պարամետրերը՝

$$\mathbf{E}\,(n_1/N_1) = \mathbf{E}\,(n_2/N_2) = p,$$
$$\sigma^2\,(n_1/N_1) = p(1-p)/N_1,\ \sigma^2\,(n_2/N_2) = p(1-p)/N_2 :$$

Քանի որ n_1-ը և n_2-ը անկախ են, անկախ կլինեն նաև n_1/N_1-ն ու n_2/N_2-ը, և $(n_1/N_1 - n_2/N_2)$-ը ևս կունենա մոտավորապես նորմալ բաշխում, որի ցրվածքն է՝

$$\sigma^2\,(n_1/N_1 - n_2/N_2) = \sigma^2\,(n_1/N_1) + \sigma^2\,(n_2/N_2) = p(1-p)/N_1 + p(1-p)/N_2 :$$

Հետևաբար, որպես վիճականի կարելի է վերցնել

$$T(\mathbf{X},\mathbf{Y}) = \frac{(n_1/N_1) - (n_2/N_2)}{\sqrt{p(1-p)(((1/N_1) + (1/N_2))}} \sim \mathcal{N}(0,1) :$$

Հիշելով, որ (N_1+N_2)-ը բավականին մեծ թիվ է, $p(1-p)$-ն կարելի է փոխարինել մոտավոր $(n_1+n_2)/(N_1+N_2)\,[1-(n_1+n_2)/(N_1+N_2)]$ արժեքով:

Քանի որ $H_1 : p_1 \neq p_2$ վարկածը երկկողմանի է, ապա α-ին համապատասխան կգտնենք $z_{1-\alpha/2}$-ը հետևյալ հավասարումից՝

$$\mathbf{P}\{|T(\mathbf{X},\mathbf{Y})| \le z_{1-\alpha/2}\} = 2\Phi(z_{1-\alpha/2}) = 1 - \alpha,$$

և կստանանք կրիտիկական տիրույթը՝

$$T_\alpha = \{(-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)\}:$$

Օրինակ 8: $\alpha = 0.01$ նշանակալիության մակարդակով համեմատել դիմորդների՝ երկու խմբերում նույն առարկայից «գերազանց» ստանալու հավանականությունները, ունենալով հետևյալ տվյալները՝

խմբի համարը	ընտրանքի ծավալը	գերազանցիկների թիվը խմբում
1	$N_1 = 40$	$n_1 = 8$
2	$N_2 = 52$	$n_2 = 10$

Լուծում: 1. $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$, որտեղ p_1-ը p_2-ը , համապատասխանաբար, առաջին և երկրորդ խմբերում «գերազանց» ստանալու հավանականություններն են:

2. $\alpha = 0.01$:

3. $T(\mathbf{X}, \mathbf{Y}) = \dfrac{n_1/N_1 - n_2/N_2}{\sqrt{p(1-p)\left((1/N_1) + (1/N_2)\right)}} \sim \mathcal{N}(0, 1)$:

4. Կրիտիկական տիրույթը՝ T_α-ն, որոշվում է

$$\mathbf{P}\left\{\left|\frac{(n_1/N_1) - (n_2/N_2)}{\sqrt{p(1-p)((1/N_1) + (1/N_2))}}\right| > z_{1-\alpha/2}\right\} = 0.01 = 1 - 2\Phi(z_{1-\alpha/2})$$

հավասարումից, կամ

$$\Phi(z_{1-\alpha/2}) = 0.495, \quad z_{1-\alpha/2} = 2.57, \quad T_\alpha = (-\infty, -2.57) \cup (2.57, \infty):$$

5. Գտնենք T-ի ընտրանքային արժեքը՝

$$\hat{t} = \frac{(8/40) - (10/52)}{\sqrt{18/92\,(1 - (18/92))\,((1/40) + (1/52))}} = 0.092:$$

6. Քանի որ $-2.57 < 0.092 < 2.57$, H_0-ն մնում է ուժի մեջ:

Փոքր ընտրանքներ: Եթե N_1-ը և N_2-ը փոքր են, ապա կիրառվում է հետևյալ հայտանիշը: Յուրաքանչյուր ընտրանքի արդյունքները խմբավորում են ըստ դիտարկվող A հատկանիշի հետևյալ աղյուսակում՝

	փաստացի հաճախություններ		տեսական հաճախություններ	
	A	A'	A	A'
$\mathbf{X}$	n_1	$N_1 - n_1$	$N_1 p$	$(1-p)N_1$
$\mathbf{Y}$	n_2	$N_2 - n_2$	$N_2 p$	$(1-p)N_2$

Եթե ընտրանքները վերցված են միևնույն համույթից, ապա կարելի է որոշել տեսական pN_1, $(1-p)N_1$ և $pN_2, (1-p)N_2$ հաճախությունները, ընդ որում p-ի գնահատականն է $(n_1 + n_2)/(N_1 + N_2)$: Սահմանենք հետևյալ վիճականին՝

$$T(\mathbf{X}, \mathbf{Y}) =$$
$$= \frac{(n_1 - \hat{p}N_1)^2}{\hat{p}N_1} + \frac{(N_1 - n_1 - (1-\hat{p})N_1)^2}{(1-\hat{p})N_1} + \frac{(n_2 - \hat{p}N_2)^2}{\hat{p}N_2} + \frac{(N_2 - n_2 - (1-\hat{p})N_2)^2}{(1-\hat{p})N_2},$$

որն ունի $\chi^2(1)$-բաշխում (չորս տեսական հաճախությունների միջև տեղի ունի երեք անկախ առնչություն, մնում է միայն մեկ ազատության աստիճան): Ընտրելով α նշանակալիության մակարդակը, գտնենք համապատասխան $\chi^2_{1-\alpha}(1)$-ը: Եթե $T(\mathbf{X}, \mathbf{Y})$-ի ընտրանքային արժեքը մեծ է $\chi^2_{1-\alpha}(1)$-ից, ապա H_0 վարկածը ժխտվում է:

Օրինակ 9: $\alpha = 0.05$ նշանակալիության մակարդակով համեմատել երկու բնակելի շենքերում բնակարանների նորոգման անհրաժեշտության հավանականությունները, ըստ աղյուսակում բերված ստուգման արդյունքների՝

ստուգման արդյունքները	առաջին շենք	երկրորդ շենք
նորոգման կարիք կա	$n_1 = 12$	$n_2 = 8$
նորոգման կարիք չկա	$N_1 - n_1 = 10$	$N_2 - n_2 = 10$
բնակարանների թիվը	22	18

Լուծում: 1. Նշանակենք p_1-ով բնակարանի նորոգման անհրաժեշտության հավանականությունը առաջին շենքի համար, p_2-ով՝ երկրորդ շենքի համար: Ստուգվում է $H_0 : p_1 = p_2 = p$ վարկածը ընդդեմ $H_1 : p_1 \neq p_2$ վարկածի: Ենթադրում ենք, որ տեղեկությունները անկախ փորձերի արդյունքներ են: p-ի գնահատականը հավասար է $(n_1 + n_2)/(N_1 + N_2) = 20/40 = 0.5$:

2. $\alpha = 0.05$:

3. Վերցնում ենք $T(\mathbf{X}, \mathbf{Y})$ վիճականին:

4. Աղյուսակից գտնում ենք $\chi^2_{0.95}(1) = 3.84$:

5. Հաշվելով $T(\mathbf{X}, \mathbf{Y})$-ի նմուշային արժեքը, կստանանք՝

$$\frac{(12 - 0.5 \cdot 22)^2}{0.5 \cdot 22} + \frac{(10 - 0.5 \cdot 22)^2}{0.5 \cdot 22} + \frac{(8 - 0.5 \cdot 18)^2}{0.5 \cdot 18} + \frac{(10 - 0.5 \cdot 18)^2}{0.5 \cdot 18} = 0.4 :$$

6. Քանի որ $0.4 < 3.84$, H_0-ն մնում է ուժի մեջ:

9.4. Վարկածներ նորմալ հանուրի սպասելիի վերաբերյալ

Սպասելիի համեմատումը նորմատիվի հետ հայտնի ցրվածքի դեպքում: Ստուգման է ենթարկվում $H_0 : m = m_0$ վարկածը ընդդեմ $H_1 : m \neq m_0$ վարկածի: Ենթադրենք, որ $X \sim \mathcal{N}(m, \sigma^2)$, և σ^2 ցրվածքը հայտնի է: Նման խնդիրներ ծագում են, օրինակ, որևէ միջին ցուցանիշով բնորոշվող արտադրանքի որակը ստուգելիս:

Ստուգման հայտանիշում վերցվում է նորմավորված նմուշային միջինը՝

$$T(\mathbf{X}) = (\overline{\mathbf{X}} - m_0)\sqrt{N}/\sigma, \quad T(\mathbf{X}) \sim \mathcal{N}(0, 1) :$$

Ընտրելով α նշանակալիության մակարդակը, հետևյալ հավասարումից

$$\mathbf{P}\{|T(\mathbf{X})| \leq z_{1-\alpha/2}\} = 2\Phi(z_{1-\alpha/2}) = 1 - \alpha,$$

որոշում են $t_{1-\alpha/2}$ արժեքը և կրիտիկական կետերը՝ $z_{\alpha/2}$, $z_{1-\alpha/2}$:

Եթե $H_0 : m = m_0$ վարկածը ստուգվում է ընդդեմ $H_1 : m > m_0$ (կամ $H_1 : m < m_0$) երկընտրանքային վարկածի, ապա կատարվում է միակողմանի ստուգում: Կրիտիկական տիրույթը և կրիտիկական կետը ստացվում են

$$\alpha = \mathbf{P}\{T(\mathbf{X}) > z_{1-\alpha}\} = 1/2 - \Phi(z_{1-\alpha}) \text{ (կամ } \alpha = \mathbf{P}\{T(\mathbf{X}) < z_\alpha\})$$

հավասարումից: Այդ դեպքում կրիտիկական տիրույթը տրվում է $T(\mathbf{X}) > z_{1-\alpha}$ (կամ $T(\mathbf{X}) < z_\alpha$) անհավասարությամբ:

Հաշվենք հայտանիշի $(1 - \beta)$ հզորությունը պարզ երկընտրանքային $H_1 : m = m_1$ $(m_1 > m_0)$ վարկածի դեպքում: Կառուցենք $T_1(\mathbf{X})$ վիճականին (m_1 կենտրոնի նկատմամբ, քանի որ ըստ H_1 վարկածի $\mathbf{E}X = m_1$)՝

$$T_1(\mathbf{X}) = (\overline{\mathbf{X}} - m_1)\sqrt{N}/\sigma$$

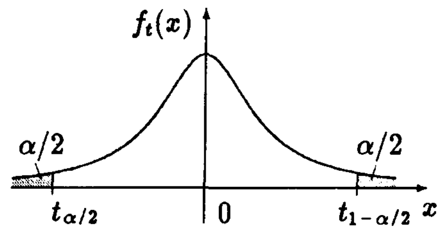
$$1 - \beta = \mathbf{P}\{T_1(\mathbf{X}) > z_\alpha\} = 1/2 + \Phi(z_{1-\alpha}),$$

Ստուգման այս հայտանիշը կիրառելի է նաև այն դեպքում, եթե X-ի բաշխումը նորմալ չէ, սակայն N-ը մեծ է, քանի որ նմուշային միջինը ասիմպտոտորեն նորմալ է բաշխված:

Միջինի համեմատումը տրված արժեքի հետ անհայտ ցրվածքի դեպքում: Ինչպես ցույց ենք տվել գլուխ 8-ում, նորմալ հանուրի դեպքում

$$T(\mathbf{X}) = (\overline{\mathbf{X}} - m_0)\sqrt{N}/\hat{s} \sim \mathcal{T}(N-1):$$

$T(\mathbf{X})$-ը կոչվում է Ստյուդենտի վիճակաձի (նկար 5):



Նկար 5:

Եթե $H_0\ :\ m = m_0$ վարկածը ստուգվում է ընդդեմ $H_1\ :\ m \neq m_0$ վարկածի, ապա կրիտիկական կետերն ստացվում են

$$\mathbf{P}\{t_{\alpha/2} \leq T(\mathbf{X}) \leq t_{1-\alpha/2}\} = 1 - \alpha$$

հավասարումից, հաշվի առնելով, որ Ստյուդենտի բաշխման խտության ֆունկցիան զույգ է, և քանորդիչները կապված են $t_{1-\alpha/2} = -t_{\alpha/2}$ առնչությամբ:

Միակողմանի H_1 վարկածների դեպքում համապատասխան ձևով կառուցվում են միակողմանի կրիտիկական տիրույթները:

Եթե X-ի բաշխումը նորմալ չէ, ապա մեծ N-երի դեպքում Ստյուդենտի վիճակաձին ասիմպտոտորեն բաշխված է $\mathbf{t}(N-1)$ օրենքով, և կարելի է որոշ վերապահումներով կիրառել նույն ընթացակարգը:

Երկու համուրների միջինների համեմատումը հայտնի և հավասար ցրվածքների դեպքում: Դիցուք ունենք $\mathbf{X} = (X_1, \ldots, X_N)$ և $\mathbf{Y} = (Y_1, \ldots, Y_M)$ նմուշները՝ երկու անկախ նորմալ բաշխված համուրներից: Առաջին նմուշի ծավալը N է, իսկ երկրորդինը՝ M: Ենթադրենք, որ նմուշներն ունեն անհայտ m_X և m_Y միջիններ և հայտնի $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ ցրվածքներ: Անհրաժեշտ է ստուգել $H_0\ :\ m_X = m_Y$ վարկածը ընդդեմ $H_1\ :\ m_X \neq m_Y$ վարկածի: H_0 վարկածը բարդ է: Սակայն այն կարող է բերվել պարզ վարկածի, եթե դիտարկենք միջինների $m_X - m_Y$ տարբերությունը: Այդ դեպքում որպես հայտանիշի վիճակաձի բնական է դիտարկել $\overline{\mathbf{X}} - \overline{\mathbf{Y}}$ միջինների տարբերությունը, որը նորմալ է բաշխված, քանի որ նորմալ են և անկախ $\overline{\mathbf{X}}$–ը և $\overline{\mathbf{Y}}$-ը: Հարմար է, որ $\overline{\mathbf{X}} - \overline{\mathbf{Y}}$ տարբերությունը ունենա կանոնական նորմալ բաշխում: Դրա համար հաշվենք $\mathbf{D}(\overline{\mathbf{X}}-\overline{\mathbf{Y}}) = \mathbf{D}\overline{\mathbf{X}}+\mathbf{D}\overline{\mathbf{Y}}$, քանի որ $\overline{\mathbf{X}}$–ը և $\overline{\mathbf{Y}}$–ը անկախ են, $\mathbf{D}(\overline{\mathbf{X}}) = \sigma^2/N$, $\mathbf{D}(\overline{\mathbf{Y}}) = \sigma^2/M$, կստանանք

$$\mathbf{D}(\overline{\mathbf{X}} - \overline{\mathbf{Y}}) = \sigma^2/N + \sigma^2/M = \sigma^2(N+M)/NM :$$

Այսպիսով, ստուգման $T(\mathbf{X}, (\mathbf{Y}))$ վիճակաձին ունի հետևյալ տեսքը՝

$$T(\mathbf{X},\mathbf{Y}) = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\mathbf{D}(\overline{\mathbf{X}} - \overline{\mathbf{Y}})}} = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sigma}\sqrt{\frac{NM}{N+M}} \sim \mathcal{N}(0,1) :$$

Կրիտիկական $\mathcal{T}_\alpha$ տիրույթի տեսքը որոշում է H_1 երկընտրանքային վարկածը:

Եթե $H_1\ :\ m_X \neq m_Y$, ապա $\mathcal{T}_\alpha$-ն երկկողմանի տիրույթ է: Կրիտիկական $z_{\alpha/2}$ և $z_{1-\alpha/2}$ կետերը գտնում ենք $\mathbf{P}\{|T(\mathbf{X},\mathbf{Y})| \leq z_{1-\alpha/2}\} = 1 - \alpha$ հավասարումից: Եթե $t \in (z_{\alpha/2}, z_{1-\alpha/2})$, H_0-ն մնում է ուժի մեջ:

Եթե $H_1 : m_X > m_Y$, կամ $H_1 : m_X < m_Y$, կրիտիկական $\mathcal{T}_\alpha$ տիրույթը որոշվում է համապատասխան ձևով:

Միջինների համեմատումը հայտնի և տարբեր ցրվածքների դեպքում: Եթե նորմալ բաշխում ունեցող երկու համույթներն ունեն m_X և m_Y անհայտ միջիններ և, համապատասխանաբար, հայտնի, բայց տարբեր σ_X, σ_Y ցրվածքներ, ապա $H_0 : m_X = m_Y$ վարկածի ստուգման համար ընտրում են հետևյալ վիճականին, որն ունի նորմալ կանոնաձև բաշխում՝

$$T(\mathbf{X},\mathbf{Y}) = (\overline{\mathbf{X}} - \overline{\mathbf{Y}})/\sqrt{\sigma_X^2/N + \sigma_Y^2/M} \sim \mathcal{N}(0,1):$$

Հետագա քայլերը համընկնում են նախորդ դեպքերի քայլերի հետ:

<u>*Օրինակ*</u> 10: Երկու գործարան արտադրում են նույն ըմպելիքները: Վաճառված արտադրանքի մասին հետևյալ վիճակագրական տվյալների հիման վրա, $\alpha = 0.01$ նշանակալիության մակարդակով ստուգել տարածված կարծիքը՝ Ա ֆիրմայի արտադրանքը ավելի էժան է:

Ա ֆիրման	$N_1 = 80$	$\overline{\mathbf{X}} = 150$ դրամ	$\sigma_1^2 = 10$
Բ ֆիրման	$N_2 = 60$	$\overline{\mathbf{Y}} = 152$ դրամ	$\sigma_2^2 = 8$

Լուծում: 1. Վարկածները նշանակենք $H_0 : m_X - m_Y = 0$, $H_1 : m_X - m_Y < 0$:

2. $\alpha = 0.01$:

3. Որպես վիճականի վերցնենք $T(\mathbf{X},\mathbf{Y}) = (\overline{\mathbf{X}} - \overline{\mathbf{Y}})/\sqrt{\sigma_X^2/N + \sigma_Y^2/M} \sim \mathcal{N}(0,1)$:

4. Գտնենք z_α կրիտիկական կետը՝ $\mathbf{P}\{T(\mathbf{X},\mathbf{Y}) < z_\alpha\} = 1/2 - \Phi(z_{1-\alpha}) = 0.01$, $z_{1-\alpha} = 2.33$, $z_\alpha = -2.33$:

5. Հաշվենք $T(\mathbf{X},\mathbf{Y})$-ի նմուշային արժեքը՝ $\hat{t} = \dfrac{150-152}{\sqrt{10/80 + 8/60}} \approx -\dfrac{2}{\sqrt{0.26}} \approx -3.93$:

6. Քանի որ $\hat{t} < z_\alpha = -2.33$, H_0-ն ժխտվում է, այսինքն ընդունվում է՝ Ա ֆիրմայի արտադրանքը ավելի էժան է, վարկածը:

Միջինների համեմատումը անհայտ, բայց հավասար ցրվածքների դեպքում: Դիցուք $\mathbf{X}$-ը և $\mathbf{Y}$-ը երկու նորմալ բաշխում ունեցող նմուշներ են, որոնց m_X և m_Y միջինները անհայտ են, ինչպես նաև անհայտ են $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ ցրվածքները: Ստուգվում է $H_0 : m_X = m_Y$ վարկածը ընդդեմ $H_1 : m_X \neq m_Y$ վարկածի:

Որպես ստուգման հայտանիշի վիճականի ընտրենք հետևյալը՝

$$T(\mathbf{X},\mathbf{Y}) = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sqrt{\sigma^2(\overline{\mathbf{X}} - \overline{\mathbf{Y}})}} = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sigma}\sqrt{\frac{N+M}{NM}}:$$

Քանի որ σ-ն անհայտ է, վերցնենք նրա անշեղ գնահատումը: Հայտնի է (տե՛ս գլուխ 6), որ σ_X^2-ու և σ_Y^2-ու անշեղ գնահատումներ են, համապատասխանաբար՝

$$\hat{s}^2(\mathbf{X}) = \frac{1}{N-1}\sum_n (X_n - \overline{\mathbf{X}})^2\text{-ը և } \hat{s}^2(\mathbf{Y}) = \frac{1}{M-1}\sum_m (Y_m - \overline{\mathbf{Y}})^2\text{-ը}:$$

Գիտենք նաև, որ $(N-1)\hat{s}^2(\mathbf{X})/\sigma^2 = \sum_n (X_n - \overline{\mathbf{X}})^2/\sigma^2$ վիճականին բաշխված է $\chi^2(N-1)$ օրենքով, $(M-1)\hat{s}^2(\mathbf{Y})/\sigma^2 = \sum_m (Y_m - \overline{\mathbf{Y}})^2/\sigma^2$ վիճականին բաշխված է $\chi^2(M-1)$ օրենքով, ուստի $(1/\sigma^2)(N-1)\hat{s}^2(\mathbf{X}) + (1/\sigma^2)(M-1)\hat{s}^2(\mathbf{Y})$ պատահական մեծությունը բաշխված է $\chi^2(N+M-2)$օրենքով:

Քանի որ $\mathbf{E}\hat{s}^2(\mathbf{X}) = \mathbf{E}\hat{s}^2(\mathbf{Y}) = \sigma^2$, ապա

$$\mathbf{E}\left[\frac{N-1}{N+M-2}\hat{s}^2(\mathbf{X}) + \frac{M-1}{N+M-2}\hat{s}^2(\mathbf{Y})\right] = \frac{N-1}{N+M-2}\sigma^2 + \frac{M-1}{N+M-2}\sigma^2 = \sigma^2,$$

այսինքն, $\dfrac{(N-1)\hat{s}^2(\mathbf{X})+(M-1)\hat{s}^2(\mathbf{Y})}{N+M-2}$ վիճակականին σ^2 ցրվածքի համար անշեղ գնահատում է:

Վերցնենք հետևյալ $T(\mathbf{X},\mathbf{Y})$ վիճակականին, որն անկախ է m_X-ից և m_Y-ից և σ^2-ուց և ունի Ստյուդենտի $T(N+M-2)$-բաշխում՝

$$T(\mathbf{X},\mathbf{Y})=\frac{\overline{\mathbf{X}}-\overline{\mathbf{Y}}}{\sqrt{(N-1)\hat{s}^2(\mathbf{X})+(M-1)\hat{s}^2(\mathbf{Y})}}\sqrt{\frac{MN(M+N-2)}{M+N}}:$$

Ստուգման ենթարկվող $H_0: m_X=m_Y$ վարկածը ժխտվում է, եթե

$$|T(\mathbf{X},\mathbf{Y})|>t_{1-\alpha/2}(N+M-2),$$

որտեղ $t_{1-\alpha/2}(N+M-2)$-ը Ստյուդենտի բաշխման $(1-\alpha/2)$-քանորդիչն է:

Օրինակ 11: Հին եղանակով արտադրանքի մեկ միավորի պատրաստման ժամանակը կազմում էր X րոպե: Նոր եղանակին անցնելուց հետո մեկ շինվածք պատրաստելու համար սկսել են ծախսել Y րոպե: Համարելով X-ի և Y-ի բաշխումը նորմալ m_X և m_Y սպասելիներով, $\alpha=0.1$ նշանակալիության մակարդակով ստուգել նոր եղանակի արդյունավետությունը, ելնելով հետևյալ վիճակագրական տվյալներից՝

հին եղանակ

շինվածքի պատրաստման ժամանակը X_i	23	26	28
արտադրանքի քանակը n_i	30	45	25

, $N=100$,

նոր եղանակ

շինվածքի պատրաստման ժամանակը Y_i	22	24	26	28
արտադրանքի քանակը m_i	40	35	10	5

, $M=90$:

Լուծում: 1. $H_0: m_X=m_Y$ վարկածը ստուգվում է ընդդեմ $H_1: m_X\neq m_Y$ վարկածի:

2. $\alpha=0.1$:

3. Ստուգման հայտանիշը $T(\mathbf{X},\mathbf{Y})$-ն է, որը բաշխված է $T(N+M-2)$ օրենքով՝

$$T(\mathbf{X},\mathbf{Y})=\frac{\overline{\mathbf{X}}-\overline{\mathbf{Y}}}{\sqrt{(N-1)\hat{s}^2(\mathbf{X})+(M-1)\hat{s}^2(\mathbf{Y})}}\sqrt{\frac{MN(M+N-2)}{M+N}}:$$

4. $t_\alpha(N+M-2)$ կրիտիկական արժեքը ստացվում է աղյուսակից՝ $t_{0.1}(188)=3.339$:

5. Հաշվենք $\hat{t}$-ն՝

$$\overline{\mathbf{x}}=25.60,\ \hat{s}^2(\mathbf{x})=3.54,,\ \overline{\mathbf{y}}=23.55,\ \hat{s}^2(\mathbf{y})=3.17,$$

$$\hat{t}(\mathbf{x},\mathbf{y})=\frac{25.60-23.55}{\sqrt{99\cdot3.54+89\cdot3.17}}\cdot\sqrt{\frac{100\cdot90(100+90-2)}{190}}=7.36:$$

6. Քանի որ $\hat{t}(\mathbf{x},\mathbf{y})=7.36>t_\alpha=3.339$, H_0 վարկածը ժխտվում է՝ նոր մեթոդը արդյունավետ է:

9.5. Վարկածներ երկու համույթների ցրվածքների վերաբերյալ

Երկու նորմալ համույթների ցրվածքների համեմատումը: անհայտ միջինների դեպքում: Դիցուք $\mathbf{X}$-ը և $\mathbf{Y}$-ը, համապատասխանաբար, N և M ծավալների երկու նմուշ են, վերցված նորմալ բաշխված և անկախ համույթներից, որոնց m_X, m_Y միջինները, σ_X^2, σ_Y^2 ցրվածքները անհայտ են: Անհրաժեշտ է ստուգել $H_0: \sigma_X^2=\sigma_Y^2$ վարկածը ընդդեմ $H_1: \sigma_X^2\neq\sigma_Y^2$ վարկածի:

Դիտարկենք ստուգման հետևյալ $T(\mathbf{X},\mathbf{Y})$ վիճակականին՝

$$T(\mathbf{X},\mathbf{Y})=\frac{\hat{s}^2(\mathbf{X})}{\hat{s}^2(\mathbf{Y})}=\frac{1}{N-1}\sum_{n=1}^{N}(X_n-\overline{\mathbf{X}})^2\Big/\frac{1}{M-1}\sum_{m=1}^{M}(Y_m-\overline{\mathbf{Y}})^2,$$

որը H_0 վարկածի դեպքում անկախ է նորմալ բաշխման պարամետրերից, ընդ որում $\hat{s}^2(\mathbf{X})/\sigma^2$-ն և $\hat{s}^2(\mathbf{Y})/\sigma^2$-ն ունեն χ^2-բաշխում, համապատասխանաբար, $(N-1)$ և $(M-1)$ ազատության աստիճաններով:

Եթե $H_0 : \sigma_X^2 = \sigma_Y^2$ վարկածը ճիշտ է, այսինքն $\sigma_X^2/\sigma_Y^2 = 1$, ապա գնահատումների $\hat{s}^2(\mathbf{X})/\hat{s}^2(\mathbf{Y})$ հարաբերությունը պետք է մոտ լինի 1-ին: Հարաբերության ո՞ր արժեքի դեպքում այդ եզրակացությունը կլինի բավականին հիմնավորված: Այդ հարցի պատասխանը (H_0 վարկածը ճիշտ լինելու դեպքում) տրվում է

$$\hat{s}^2(\mathbf{X})/\hat{s}^2(\mathbf{Y}) \sim \mathcal{F}(N-1, M-1)$$

վիճականու միջոցով՝ եթե $\hat{s}^2(\mathbf{X})/\hat{s}^2(\mathbf{Y}) < t_1$ կամ $\hat{s}^2(\mathbf{X})/\hat{s}^2(\mathbf{Y}) > t_2$ $(t_1 < 1 < t_2)$, H_0 վարկածը ժխտվում է:

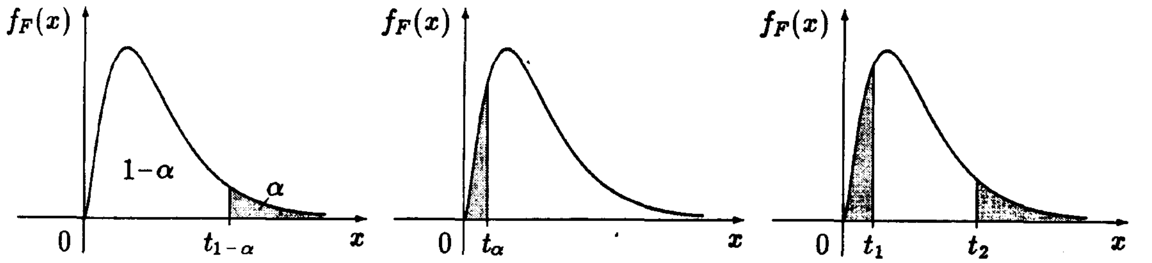
Կրիտիկական կետերն են $F_{\alpha/2}(N-1, M-1)$ և $F_{1-\alpha/2}(N-1, M-1)$ քանորդիչները, որոնք կարելի է գտնել աղյուսակից: Նկատի ունենանք (տե՛ս գլուխ 8), որ ճիշտ է հետևյալ առնչությունը՝ $F_\alpha(N_1, N_2) = 1/F_{1-\alpha}(N_2, N_1)$:

Կառուցենք տրված α նշանակալիության մակարդակին համապատասխանող կրիտիկական տիրույթները H_1 վարկածի բոլոր տարբերակների համար:

1) $H_0 : \sigma_X^2 = \sigma_Y^2$ վարկածը *ստուգվում է ընդդեմ* $H_1 : \sigma_X^2 > \sigma_Y^2$ *վարկածի*: Տվյալ դեպքում կրիտիկական $t_{1-\alpha}$ կետը որոշվում է $\mathbf{P}\{T(\mathbf{X},\mathbf{Y}) > t_{1-\alpha}\} = \alpha$ հավասարումից և կրիտիկական $\mathcal{T}_\alpha$ տիրույթն ունի $(t_{1-\alpha}, \infty)$ տեսքը: Աղյուսակից որոշում ենք $t_{1-\alpha}$ կետը՝

$$t_{1-\alpha} = F_{1-\alpha}(N-1, M-1):$$

Կրիտիկական տիրույթը պատկերված է նկար 6ա-ում:



Նկար 6ա: Նկար 6բ: Նկար 6գ:

Եթե $T(\mathbf{X},\mathbf{Y}) > t_{1-\alpha}$, H_0-ն ժխտվում է:

2) $H_0 : \sigma_X^2 = \sigma_Y^2$ *վարկածը ստուգվում է ընդդեմ* $H_1 : \sigma_X^2 < \sigma_Y^2$ *վարկածի*: Այդ դեպքում կրիտիկական տիրույթը ձախակողմյան է: Կրիտիկական կետը որոշվում է

$$\mathbf{P}\{T(\mathbf{X},\mathbf{Y}) < t_\alpha\} = \alpha$$

պայմանից: Քանի որ $T(\mathbf{X},\mathbf{Y})$-ը բացասական չէ, կրիտիկական տիրույթն ունի $(0, t_\alpha)$ տեսքը: Նշենք, որ կրիտիկական կետը աղյուսակից գտնելու համար օգտվում են $t_\alpha = 1/t_{1-\alpha}$ առնչությունից: Կրիտիկական տիրույթը պատկերված է 6բ նկարում:

Եթե $T(\mathbf{X},\mathbf{Y}) < t_\alpha$, H_0-ն ժխտվում է և ընդունվում է H_1-ը:

3) $H_0 : \sigma_X^2 = \sigma_Y^2$ *վարկածը ստուգվում է ընդդեմ* $H_1 : \sigma_X^2 \neq \sigma_Y^2$ *վարկածի*: Կրիտիկական տիրույթը երկկողմանի է, t_1 և t_2 կրիտիկական կետերը որոշվում են աղյուսակից՝

$$\mathbf{P}\{t_{\alpha/2} \leq T(\mathbf{X},\mathbf{Y}) \leq t_{1-\alpha/2}\} = 1-\alpha$$

պայմանից, դրանք ներկայացված են 6գ գծապատկերում:

Նշենք, որ եթե $H_0 : \sigma_X^2 = \sigma_Y^2$ վարկածը չի ժխտվում, ապա նմուշային $\hat{s}^2(\mathbf{X})$ և $\hat{s}^2(\mathbf{Y})$ արժեքների տարբերությունը համարվում է աննշան և ընդհանուր ցրվածքի արժեքը ընդունում են հավասար $\dfrac{\hat{s}^2(\mathbf{X})(N-1) + \hat{s}^2(\mathbf{Y})(M-1)}{N+M-2}$:

Օրինակ 12: Կարելի՞ է արդյոք $\alpha = 0.1$ նշանակալիության մակարդակով անտեսել օրինակ 11-ում ստացված $\hat{s}^2(\mathbf{X})$ և $\hat{s}^2(\mathbf{Y})$ գնահատումների տարբերությունը:

Լուծում: 1. Ստուգվում է $H_0 : \sigma_X^2 = \sigma_Y^2$ վարկածը ընդդեմ $H_1 : \sigma_X^2 \neq \sigma_Y^2$ վարկածի:

2. $\alpha = 0.1$:

3. Ընտրվում է $T(\mathbf{X},\mathbf{Y}) = \hat{s}^2(\mathbf{X})/\hat{s}^2(\mathbf{Y})$ Ֆիշերի վիճականին:

4. Կրիտիկական կետերն են՝

$$t_{\alpha/2} = F_{\alpha/2}(N-1, M-1) = 0.71 \text{ և } t_{1-\alpha/2} = F_{1-\alpha/2}(N-1, M-1) = 1.41:$$

5. Հաշվում ենք $\hat{t} = 3.54/3.17 = 1.12$:

6. Քանի որ $\hat{t} \in [0.71; 1.41]$, H_0-ն մնում է ուժի մեջ:

Ցրվածքների համեմատումը հայտնի միջինների դեպքում: Դիցուք նորմալ բաշխում ունեցող երկու համույթների միջինները տրված են՝ $\mathbf{EX} = m_X$, $\mathbf{EY} = m_Y$, իսկ σ_X^2 և σ_Y^2 ցրվածքներն անհայտ են: Ստուգվում է $H_0 : \sigma_X^2 = \sigma_Y^2$ վարկածը ընդդեմ $H_1 : \sigma_X^2 \neq \sigma_Y^2$ վարկածի: Անհայտ ցրվածքների գնահատումները՝

$$\hat{s}_0^2(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N}(X_n - m_X)^2, \quad \hat{s}_0^2(\mathbf{Y}) = \frac{1}{M}\sum_{m=1}^{M}(Y_m - m_Y)^2,$$

նմուշային ցրվածքներն են, որտեղ m_X-ը և m_Y-ը հայտնի են: Եթե $H_0 : \sigma_X^2 = \sigma_Y^2 = \sigma^2$ վարկածը ճիշտ է, ապա $\hat{s}_0^2(\mathbf{X})/\sigma^2 \sim \chi^2(N)$ և $\hat{s}_0^2(\mathbf{Y})/\sigma^2 \sim \chi^2(M)$: Ստուգման վիճականին կվերցնենք

$$T(\mathbf{X},\mathbf{Y}) = \hat{s}_0^2(\mathbf{X})/\hat{s}_0^2(\mathbf{Y}),$$

որն ունի Ֆիշերի բաշխում N և M ազատության աստիճաններով, այսինքն՝

$$T(\mathbf{X},\mathbf{Y}) \sim \mathcal{F}(N, M):$$

Կրիտիկական տիրույթը որոշելու ընթացակարգն ընդգրկում է նույն քայլերը, որոնք բերված են նախորդ դեպքում:

9.6. Համաձայնության հայտանիշներ

Համաձայնության հայտանիշները թույլ են տալիս պատասխանել հետևյալ հարցին՝ իրո՞ք փորձնական և տեսական բաշխումների միջև եղած տարբերությունը այնքան փոքր է, որ կարող է համարվել պատահականության հետևանք, թե՞ ոչ:

Հատուկ տեղ են գրավում այն հայտանիշները, որոնցով ստուգվում է համույթի նորմալ բաշխված լինելու վարկածը, քանի որ հաճախ հայտանիշների հիմքում դրվում է նորմալ բաշխման առկայության ենթադրությունը:

Հատկանիշի տեսական բաշխման մասին վարկածի հիմքում կարող են դրվել հատկանիշի մասին տարբեր նախադրյալներ (որոնք, օրինակ, հիմնավորում են բաշխման ասիմպտոտորեն նորմալությունը) կամ դիտվող բաշխման որոշակի հատկություններ (օրինակ, անհամաչափության և կուտակվածության գործակիցների՝ զրոյին հավասար լինելը կարող է ընդունվել որպես բաշխման նորմալ լինելու հայտանիշ, իսկ միջինի և ցրվածքի միմյանց հավասար լինելը՝ Պուասոնի բաշխման):

$H_0 : F_X(x) \equiv F(x)$ վարկածի ստուգման՝ Կոլմոգորովի և χ^2 հայտանիշները:

Կոլմոգորովի հայտանիշը կիրառվում է, երբ $F_X(x)$-ը անընդհատ է: Վիճակագրական հայտանիշը տրվում է

$$D_N(\mathbf{X}) = \sup_{-\infty<x<\infty}\left|\hat{F}_X(x) - F_X(x)\right|$$

վիճականիով, որն արտացոլում է նմուշային $\hat{F}_X(x)$ բաշխման ֆունկցիայի (տե՛ս գլուխ 6, բանաձև (2)) առավելագույն շեղումը տրված $F_X(x)$ վարկածային բաշխման ֆունկցիայից: Տվյալ x-ի համար $\hat{F}_X(x)$ մեծությունը $F_X(x)$-ի համար օպտիմալ գնահատական է և երբ $N \to \infty$, $\hat{F}_X(x) \to F_X(x)$: Այդ պատճառով անվազն մեծ N-երի դեպքում, երբ H_0-ն ճիշտ է, D_N-ի արժեքը էապես չի տարբերվի զրոյից: Ճշգրիտ $\mathrm{P}\{\sqrt{N}D_N(\mathbf{X}) \le \lambda\}$ բաշխումը $N \ge 35$ դեպքում լավ է մոտարկվում Կոլմոգորովի սահմանային բաշխման ֆունկցիայով՝

$$K(\lambda) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2\lambda^2},$$

որի արժեքները տրվում են աղյուսակով: Հայտանիշի կրիտիկական տիրույթը որոշվում է

$$\sqrt{N}D_N(\mathbf{X}) \ge \lambda_\alpha$$

պայմանով, որտեղ $K(\lambda_\alpha) = 1 - \alpha$:

Փոքր N-երի դեպքում ($N < 35$) օգտագործում են հատուկ աղյուսակից վերցված կրիտիկական արժեքները:

Պիրսոնի χ^2 հայտանիշը: Եթե X անընդհատ հատկանիշի նկատմամբ կատարված փորձերի արդյունքներին համապատասխանող միջակայքային բաշխման աղյուսակը կազմված է, ապա քանի որ n_k/N, $k = \overline{1,K}$ հարաբերական հաճախությունները p_k հավանականությունների *ունակ* գնատուներ են, վարկածային p_k արժեքներից շեղումների չափը կարելի է ընտրել տարբեր ձևերով, օրինակ, որպես ֆունկցիաներ $|n_k/N - p_k|$, տարբերություններից: Այդպիսի ֆունկցիաներից ամենակիրառելին Կ. Պիրսոնի կողմից առաջարկված

$$T(\mathbf{X}) = \sum_{k=1}^{K} (n_k - Np_k)^2/Np_k$$

վիճականին է, որտեղ n_k-ն k-րդ՝ $[a_{k-1}, a_k)$ միջակայքին պատկանող տվյալների քանակն է, p_k-ն՝ X պատահական մեծության k-րդ միջակայքին պատկանելու վարկածային հավանականությունը՝

$$p_k = F_X(x_k) - F_X(x_{k-1}),$$

K-ն միջակայքերի քանակն է: Եթե H_0-ն պարզ վարկած է, այսինքն p_k-ները, $k = \overline{1,K}$, որոշվում են միանշանակ, ապա $N \to \infty$ դեպքում համապատասխան հայտանիշը՝ **համաձայնության χ^2 հայտանիշը**, ասիմպտոտորեն տրվում է

$$\{X : T(\mathbf{X}) \ge t_{1-\alpha}\}$$

կրիտիկական տիրույթով: Եթե N-ը մեծ է ($N \ge 50$) և յուրաքանչյուր խմբում տվյալների թիվը անվազն հավասար է 5-ի, այդ վիճականին ենթարկվում է $M = K - 1 - S$ ազատության աստիճաններով χ^2-բաշխմանը, որտեղ S-ը գնահատվող տեսական բաշխման պարամետրերի թիվն է (օրինակ, Պուասոնի բաշխման համար $M = K - 2$, նորմալ բաշխման համար $M = K - 3$): Որոշ դեպքերում հատկանիշի բաշխումը կարող է համեմատվել նախապես տրված բաշխման հետ, կամ մի բաշխման, որի պարամետրերի մի մասի արժեքները տրված են: Այդ դեպքում օգտագործվող կապերի S թիվը պակասում է:

Օրինակ 13: Կատարվում են գործարանում արտադրվող չափիչ գործիքների ստուգումներ՝ դրանց պիտանիությունն ստուգելու նպատակով: Յուրաքանչյուր գործիք ստուգվում է 4 փորձից բաղկացած փորձաշարում: Հետևյալ աղյուսակում բերված է ստուգման արդյունքները 100 գործիքների համար: $\alpha = 0.1$ նշանակալիության մակարդակով ստուգել H_0 վարկածը՝ մեկ փորձաշարում փորձերի X քանակը, որոնց դեպքում գործիքը պիտանի է, ունի երկանդամային բաշխում:

Փորձերի քանակը, որոնցում գործիքը պիտանի է, x_k	0	1	2	3	4
x_k փորձերում պիտանի գործիքների քանակը, n_k	1	1	3	35	60

Լուծում: Եթե H_0 վարկածը ճիշտ է, ապա X-ի հավանականությունների բաշխումը նկարագրվում է հետևյալ բանաձևով՝

$$\mathbf{P}\{X = k\} = C_4^k p^k (1-p)^{4-k}, \quad k = \overline{0,4}:$$

Նկատենք, որ երկանդամային բաշխման p պարամետրը այս խնդրում մեկնաբանվում է որպես կամայական գործիքի՝ կամայական փորձի ժամանակ պիտանի լինելու հավանականություն: Այդ հավանականության լավագույն գնահատումն հարաբերական հաճախությունն է՝

$$\hat{p} = \frac{\text{Փորձերի քանակը, որոնցում պիտանի են 100 գործիքները}}{\text{Փորձերի ընդհանուր քանակը 100 գործիքների համար}} = \frac{\sum_{k=0}^{4} x_k n_k}{4 \cdot 100} = 0.88:$$

Այստեղից, տեղադրելով p-ի փոխարեն նրա գնահատականը, ստանում ենք անհայտ բաշխման (երկանդամային մոդելի) պարամետրական գնահատականը՝

$$p_k = \mathbf{P}\{X = k\} = C_4^k (0.88)^k (0.12)^{4-k}, \; k = \overline{0,4}:$$

Ստուգումը կատարվում է հետևյալ կերպ.

1. H_0: բաշխումը երկանդամային է, H_1: բաշխումը տարբերվում է երկանդամայինից:
2. $\alpha = 0.1$:
3. Քանի որ Պիրսոնի հայտանիշը գործում է, երբ $n_k \geq 5$, անհրաժեշտ է միացնել տվյալների առաջին երեք խմբերը: Ստուգման վիճականին է՝

$$\hat{\chi}^2 = \sum_{k=1}^{3} (n_k - Np_k)^2 / Np_k, \quad N = 100, \quad M = K - 1 - S = 3 - 1 - 1 = 1:$$

Ստանում ենք հետևյալ աղյուսակը՝

փորձարկումների քանակը	գործիքների քանակը n_k	հաճախությունը, n_k/N	p_k	Np_k	$(n_k - Np_k)^2$	$(n_k - Np_k)^2/$ $/Np_k$
0-ից 2-ը	5	0.05	0.07320	7.32	5.382	0.735
3	35	0.35	0.32711	32.711	5.239	0.160
4	60	0.60	0.59969	59.969	0.001	0.000

4. Կրիտիկական կետը գտնում ենք աղյուսակից՝ $\chi^2_{1-\alpha}(3) = 6.64$:
5. Հաշվում ենք $\hat{\chi}^2 = 0.735 + 0.160 + 0.000 = 0.895$:
6. $\hat{\chi}^2 = 0.895 < \chi^2_{1-\alpha}(3) = 6.64$: Ուրեմն, H_0-ն մնում է ուժի մեջ:

Օրինակ 14: Ըստ նշանակալիության $\alpha = 0.025$ մակարդակի ստուգել հանուրի նորմալ բաշխված լինելու մասին վարկածը, եթե հայտնի են տեսական և փորձնական հաճախությունները.

N_{p_k}	6	14	28	18	8	3
n_k	5	10	20	25	14	3

Լուծում: 1. Պահանջվում է ստուգել հետևյալ զրոյական վարկածը՝ $H_0 : X \sim \mathcal{N}(m, \sigma^2)$,

2. $\alpha = 0.025$,

3. Կիրառենք Պիրսոնի χ^2 հայտանիշը,

4. Փորձերի թիվը՝ N-ը, հավասար է 75, այսինքն մեծ է 50-ից, ուրեմն $M = 6 - 1 - 2 = 3$, քանի որ նորմալ բաշխումը բնորոշվում է 2 պարամետրով: Կրիտիկական տիրույթը կլինի՝

$$\mathcal{T}_{0.025} = (\hat{\chi}^2 > \chi_{0.025}(3)):$$

Աղյուսակից գտնում ենք՝ $\chi_{0.025}(3) = 9.4$:

5. Հաշվենք Պիրսոնի վիճականու նմուշային արժեքը՝

$$\hat{\chi}^2 = \frac{(5-6)^2}{6} + \frac{(10-14)^2}{14} + \frac{(20-28)^2}{28} + \frac{(25-18)^2}{18} + \frac{(14-8)^2}{8} + \frac{(3-3)^2}{3} = 10.8175:$$

Քանի որ $\hat{\chi}^2 > \chi_{0.025}(3)$, H_0 վարկածը ժխտվում է, տեսական և փորձնական տվյալների նշանակալի տարամիտումը թույլ չի տալիս ընդունել հանուրի նորմալ բաշխված լինելու մասին վարկածը:

9.7. Երկու հատկանիշների անկախության ստուգումը

Դիցուք, տրված է (X,Y) պատահական մեծության N ծավալի $(\mathbf{X},\mathbf{Y})$ նմուշը: Պահանջվում է ստուգել H_0 վարկածը, ըստ որի X-ը և Y-ը անկախ են:

Ստուգման վիճականին ընտրելու համար կատարենք հետևյալ նշանակումները: Եթե X-ը և Y-ը ընդհատ են և նրանց հնարավոր արժեքներն են, համապատասխանաբար, $x_1, x_2, \ldots, x_K$-ը և $y_1, y_2, \ldots, y_L$-ը, ապա n_{kl}-ը նմուշի այն տարրերի քանակն է, որոնց համար $X = x_k$ և $Y = y_l$, $k = \overline{1,K}$, $l = \overline{1,L}$: Եթե X-ը և Y-ը անընդհատ են, ապա նրանց արժեքների տիրույթը բաժանվում է վերջավոր թվով միջակայքերի, և n_{kl}-ը նմուշի այն տարրերի քանակն է, որոնց համար X-ը ընկնում է k-րդ և Y-ը՝ l-րդ միջակայքը: Այսպիսով, նմուշը կարելի է ներկայացնել $k \times n$ զուգակցությունների աղյուսակի ձևով (տե՛ս գլուխ 6-ը):

Նշանակենք $p_{kl} = \mathbf{P}(X = x_k, Y = y_l)$, $p_{k.} = \mathbf{P}(X = x_k)$, $p_{.l} = \mathbf{P}(Y = y_l)$: Այդ հավանականությունների լավագույն գնահատականներն են, համապատասխանաբար, $\hat{p}_{kl} = n_{kl}/N$, $\hat{p}_{k.} = n_{k.}/N$ և $\hat{p}_{.l} = n_{.l}/N$, որտեղ $n_{k.} = \sum\limits_l n_{kl}/N$, $n_{.l} = \sum\limits_k n_{kl}/N$, $k = \overline{1,K}$, $l = \overline{1,L}$:

Եթե H_0-ն ճիշտ է, ապա $p_{kl} = p_{k.}p_{.l}$, և, համապատասխանաբար, n_{kl}/N և $n_{k.}n_{.l}/N^2$ արժեքները քիչ են տարբերվում:

Ստուգման համար կօգտագործենք հետևյալ ֆունկցիան՝

$$T(\mathbf{x},\mathbf{y}) = \hat{\chi}^2 = \sum_k \sum_l \frac{(n_{kl} - n_{k.}n_{.l}/N)^2}{n_{k.}n_{.l}/N}:$$

Եթե H_0-ն ճիշտ է և $n_{k.}n_{.l}/N \geq 4$, $k = \overline{1,K}$, $l = \overline{1,L}$, ապա $T(\mathbf{X},\mathbf{Y}) \sim \chi^2((K-1)(L-1))$, հետևաբար $\hat{\chi}^2 > \chi^2_{1-\alpha}((K-1)(L-1))$ դեպքում H_0-ն ժխտվում է:

$n_{k.}n_{.l}/N \geq 4$ պայմանն ապահովելու համար անհրաժեշտ է միավորել զուգակցության աղյուսակի համապատասխան սյուները կամ տողերը:

Հաճախ դիտարկվում է 2×2 զուգակցության աղյուսակը, որը համապատասխանում է երկու հակադիր հատկանիշների դեպքին: Անկախության վարկածի ստուգման վիճականին ստանում է հետևյալ տեսքը՝

$$\hat{\chi}^2 = \frac{(n_{11}n_{22} - n_{12}n_{21})^2 N}{(n_{11}+n_{12})(n_{21}+n_{22})(n_{11}+n_{21})(n_{12}+n_{22})} \sim \chi^2(1):$$

Օրինակ 15: Ունենք տվյալների երկու խումբ, որոնք խմբավորված են ըստ երկու հատկանիշների՝ սեռի ($\mathcal{A}$ - տղամարդ, $\mathcal{A}'$ - կին) և ըստ վերաբերմունքի սպորտի նկատմամբ ($\mathcal{B}$ - զբաղվում է սպորտով, $\mathcal{B}'$ - չի զբաղվում): Տվյալները բերված են հետևյալ աղյուսակում՝

	$\mathcal{A}$	$\mathcal{A}'$	Σ
$\mathcal{B}$	74	38	112
$\mathcal{B}'$	30	22	52
Σ	104	60	164

:

Ստուգել հատկանիշների անկախությունը $\alpha = 0.05$ նշանակալիության մակարդակով:

Լուծում: 1. Ստուգում ենք «հատկանիշները անկախ են» H_0 վարկածը ընդդեմ «հատկանիշները կախված են» H_1 վարկածի:

2. $\alpha = 0.05$:

3. Որպես ստուգման հայտանիշ վերցնենք

$$T(\mathbf{X}, \mathbf{Y}) = \frac{(n_{11}n_{22} - n_{12}n_{21})^2 N}{(n_{11} + n_{12})(n_{21} + n_{22})(n_{11} + n_{21})(n_{12} + n_{22})} :$$

4. Աղյուսակից գտնում ենք $\chi^2_{1-\alpha}(1) = 3.84$:

5. Որոշում ենք

$$\hat{\chi}^2 = \frac{(74 \cdot 22 - 38 \cdot 30)^2 \cdot 164}{104 \cdot 60 \cdot 112 \cdot 52} = 1.07 :$$

6. Քանի որ $\hat{\chi}^2 = 1.07 < 3.84$, H_0 վարկածը մնում է ուժի մեջ:

Գլուխ 10
Ցրվածքային վերլուծություն

Ամեն մի բազմազանության մեջ միօրինակություն է թաքնված, ամեն մի փոփոխականության մեջ՝ հաստատունություն:

Շառլ Մոնտեսքյո

10.1. Ներածություն

Ցրվածքային վերլուծությունը մաթեմատիկական վիճակագրության այն եղանակն է, որի նպատակներն են փորձի արդյունքների վրա տարբեր հանգամանքների ազդեցության ի հայտ բերելը և հետագա փորձերի ծրագրումը:

Տարբեր բնագավառներում անհրաժեշտ է լինում համեմատել զանազան նյութեր, մշակման եղանակներ, արտադրության միջոցներ և այլն, պարզելու համար, թե արդյո՞ք դրանք ազդում են որոշակի հատկանիշի վրա, թե՞ ոչ: Օրինակ, ագարակատերը ցանկանում է համեմատել տարբեր հողակտորների բերքատվությունը, բժիշկ-հետազոտողին պետք է տարբերակել մի քանի դեղամիջոցների ազդեցությունը մարդու արյան ճնշման վրա և այլն: Նման խնդիրներում ազդող որակական հատկանիշը կոչվում է **գործոն**, իսկ գործոնի ազդեցությանը ենթարկվող հատկանիշը կոչվում է **արդյունքային**: Թվարկված օրինակներում գործոններ են, համապատասխանաբար, հողակտորները և դեղամիջոցները, որպես արդյունքային հատկանիշ կարող են դիտարկվել, համապատասխանաբար, հողակտորի բերքատվության և արյան ճնշման ցուցանիշները: Գործոնների ազդեցությունն ուսումնասիրելու համար դրանց փորձարկվող տարատեսակները կոչվում են **գործոնների մակարդակներ** կամ **խմբեր**: Հետազոտվող գործոնները նշանակում են A, B և այլն, դրանցից յուրաքանչյուրը կարող է ունենալ վերջավոր թվով մակարդակներ: A գործոնի մակարդակները նշանակում են $A_1, A_2, ..., A_K$, B գործոնի մակարդակները՝ $B_1, B_2, ..., B_L$ և այլն:

Օրինակ, ենթադրենք, որ շինարարական վարչության տնտեսագետը ուսումնասիրում է մեկ հերթափոխում կատարված շինարարական աշխատանքների ծավալի կախվածությունը տարբեր բրիգադների աշխատանքից: Բրիգադների համարներն այստեղ դիտվում են որպես գործոնի մակարդակներ: Արդյունքային հատկանիշն է մեկ հերթափոխում կատարված աշխատանքի ծավալը:

Գործոնի ազդեցությունը հայտնաբերելու համար կատարում են արդյունքային հատկանիշի փորձնական դիտումներ՝ գործոնի բոլոր մակարդակներին համապատասխան: Արդյունքային հատկանիշի արժեքները կարող են կախված լինել ոչ միայն դիտարկվող գործոնից և դրա մակարդակներից, այլ նաև ուրիշ՝ հաշվի չառնված պատահական գործոններից: Ուստի արդյունքային հատկանիշը պատահական մեծություն է, որը նշանակում են Y: Ցրվածքային վերլուծությունը գործիք է՝ ելնելով Y-ի փորձնական դիտումների արդյունքներից, գործոնի մակարդակներին համապատասխանող Y-ի սպասելիների հավասարության վարկածը ստուգելու միջոցով որոշում ընդունելու ազդեցության առկայության վերաբերյալ:

Այս գլխում մենք կծանոթանանք **միագործոն ցրվածքային վերլուծության** մոդելի հետ: Կդիտարկենք **ցրվածքների համասեռության վարկածի** ստուգման եղանակներից մեկը: Այնուհետև, որպես միագործոն ցրվածքային վերլուծության ընդհանրացում, կղիտարկենք **պատահականացված բլոկների եղանակը**, որը հնարավորություն է ընձեռում վարկածներ ստուգելու նաև դիտումների «համասեռության» մասին: Վերջում կդիտարկենք բազմագործոն ցրվածքային վերլուծության մոդելը, պարզության համար սահմանափակվելով երկգործոն դեպքով:

10.2. Միագործոն ցրվածքային վերլուծություն

Հետազոտվում է մեկ որակական գործոնի մակարդակներից արդյունքային հատկանիշի կախվածության առկայությունը (կամ բացակայությունը): Միագործոն ցրվածքային վերլուծության հիմքում դրվում է հետևյալ հավանականային մոդելը՝

$$Y_l = m_l + Z_l, \; l = \overline{1, L},$$

որտեղ Y_l-ը հետազոտվող գործոնի l-րդ մակարդակին համապատասխանող արդյունքային հատկանիշն է՝ բաշխված $\mathcal{N}(0, \sigma_l^2)$, իսկ Z_l-ը՝ արդյունքային հատկանիշի պատահական շեղումը սպասելիից: Ենթադրվում է, որ ցրվածքներն իրար հավասար են՝ $\sigma_l^2 = \sigma^2$, $l = \overline{1, L}$, բայց այդ ընդհանուր σ^2 ցրվածքը հայտնի չէ:

Դիցուք կատարված են Y հատկանիշի l-րդ մակարդակի N_l անկախ դիտումներ՝

$$\mathbf{Y}_l = (Y_{1l}, Y_{2l}, ..., Y_{N_l}), \; l = \overline{1, L}, \; N_1 + N_2 + ... + N_L = N,$$

որոնց արդյունքները կնշանակենք համապատասխան փոքրատառերով՝ y_{nl}: $\mathbf{Y}$-ով կնշանակենք նմուշը, որը բոլոր $\mathbf{Y}_l$ նմուշների միավորումն է:

Արդյունքային հատկանիշի m_l սպասելիները կարելի է ներկայացնել հետևյալ տեսքով՝

$$m_l = m + e_l, \; l = \overline{1, L}, \; \sum_l e_l = 0, \; m = (1/L) \sum_l m_l :$$

Պահանջվում է $\mathbf{Y}_l$ նմուշների հիման վրա ստուգել H_0 վարկածը, ըստ որի արդյունքային հատկանիշի սպասելին կախված չէ գործոնի մակարդակներից, այսինքն՝

$$H_0 : \; m_1 = m_2 = ... = m_L = m:$$

Նշանակենք $\overline{\mathbf{Y}}_l = \sum_{n=1}^{N_l} Y_{nl}/N_l$, $l = \overline{1, L}$, $\overline{\mathbf{Y}} = \sum_l \sum_{n=1}^{N_l} Y_{nl}/N$: Ընդհանուր $\overline{\mathbf{Y}}$ և խմբային $\overline{\mathbf{Y}}_l$ նմուշային միջինները անշեղ և ունակ գնահատներ են, համապատասխանաբար, m և m_l սպասելիների համար: Եթե H_0 վարկածը ճիշտ է, ապա ընդհանուր նմուշային միջինը վիճակագրական իմաստով չի տարբերվի խմբային միջիններից:

H_0 վարկածի ստուգման հայտանիշների կառուցումը սկսենք **նմուշային միջխմբային ցրվածքի** դիտարկումից՝

$$\hat{s}_1^2(\mathbf{Y}) = \frac{1}{L-1} \sum_l N_l (\overline{\mathbf{Y}}_l - \overline{\mathbf{Y}})^2 :$$

Ցույց է տրվում, որ, եթե H_0 վարկածը ճիշտ է, ապա $\hat{s}_1^2(\mathbf{Y})(L-1)/\sigma^2$ վիճականին բաշխված է $\chi^2(L-1)$ օրենքով և $\hat{s}_1^2(\mathbf{Y})$-ը անհայտ σ^2 ցրվածքի անշեղ գնատը է (սակայն, եթե H_0-ն տեղի չունի, ապա $\hat{s}_1^2(\mathbf{Y})$-ը շեղ է):

Առանձին **նմուշային խմբային ցրվածքները**՝

$$\hat{s}^2(\mathbf{Y}_l) = \frac{1}{N_l - 1} \sum_{n=1}^{N_l} (Y_{nl} - \overline{\mathbf{Y}}_l)^2, \; l = \overline{1, L},$$

$\sigma_l^2 = \sigma^2$ ցրվածքների անշեղ գնահատներ են: Հետևաբար, $(N_l - 1)\hat{s}^2(\mathbf{Y}_l)/\sigma_l^2 \sim \chi^2(N_l - 1)$:

Նմուշային ներխմբային (կամ **մնացորդային**) **ցրվածք** կոչվում է

$$\hat{s}_2^2(\mathbf{Y}) = \sum_l (N_l - 1)\hat{s}^2(\mathbf{Y}_l) / \sum_l (N_l - 1)$$

վիճականին: Ակնհայտ է, որ այն ևս σ^2 ցրվածքի անշեղ գնատը է: Բացի դրանից,

$$(N - L)\hat{s}_2^2(\mathbf{Y})/\sigma^2 \sim \chi^2(N - L),$$

քանի որ $\sum_l (N_l - 1) = N - L$:

Դիտարկենք նաև **նմուշային ընդհանուր ցրվածքը**՝

$$\hat{s}^2(\mathbf{Y}) = \frac{1}{N-1}\sum_l \sum_{n=1}^{N_l} (Y_{nl} - \overline{\mathbf{Y}})^2,$$

որը H_0 վարկածի դեպքում նույնպես անհայտ σ^2 ցրվածքի անշեղ գնահատական է: Ապացուցվում է հետևյալ կարևոր առնչությունը՝

$$(N-1)\hat{s}^2(\mathbf{Y}) = (L-1)\hat{s}_1^2(\mathbf{Y}) + (N-L)\hat{s}_2^2(\mathbf{Y}),$$

և, որ $\hat{s}_1^2(\mathbf{Y})$ և $\hat{s}_2^2(\mathbf{Y})$ պատահական մեծությունները անկախ են: Այստեղից հետևում է, որ

$$\hat{F}(\mathbf{Y}) = \hat{s}_1^2(\mathbf{Y})/\hat{s}_2^2(\mathbf{Y}) \sim \mathcal{F}(L-1, N-L):$$

Եթե սպասելիների հավասարության վարկածը տեղի չունի, ապա խմբերի $\overline{\mathbf{Y}}_l$ միջինները զգալի չափով կտարբերվեն միմյանցից: Այս դեպքում $\hat{s}_1^2(\mathbf{Y})$-ը մեծ կլինի, մինչդեռ $\hat{s}_2^2(\mathbf{Y})$-ը, որը հաշվարկվում է ներխմբային նմուշային ցրվածքների միջոցով, բոլորովին չի փոփոխվի: Սա նշանակում է, որ կրիտիկական տիրույթն աջակողմյան է և որոշվում է α նշանակալիության մակարդակին համապատասխանող $F_{1-\alpha}(L-1, N-L)$-քանորդիչի միջոցով: Եթե $\hat{F}(\mathbf{y}) > F_{1-\alpha}(L-1, N-L)$, ապա H_0-ն ժխտվում է:

Ցրվածքային վերլուծության՝ y նմուշի հիման վրա հաշվարկները հարմար է ներկայացնել **միագործոն ցրվածքային վերլուծության աղյուսակով**:

Աղբյուրը	Ազատ. աստիճ.	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
Գործոն	$L-1$	$\hat{s}_1^2(\mathbf{y}) = \sum_l N_l(\overline{\mathbf{y}}_l - \overline{\mathbf{y}})^2/(L-1)$	$\hat{F}(\mathbf{y}) = \hat{s}_1^2(\mathbf{y})/\hat{s}_2^2(\mathbf{y})$
Մնացորդ	$N-L$	$\hat{s}_2^2(\mathbf{y}) = \sum_l \sum_{n=1}^{N_l} (y_{nl} - \overline{\mathbf{y}}_l)^2/(N-L)$	
Ընդհանուրը	$N-1$	$\hat{s}^2(\mathbf{y}) = \sum_l \sum_{n=1}^{N_l} (y_{nl} - \overline{\mathbf{y}})^2/(N-1)$	

:

Օրինակ 1: Արտադրամասում մանրամասերի երկարությունը չափում են երեք տարբեր տեսակի մանրադիտակների (A գործոն) օգնությամբ: Ճարտարագետը պետք է որոշեր՝ արդյո՞ք այդ մանրադիտակները՝ A_1, A_2, A_3, ապահովում են միևնույն արտադրողականությունը: Պատասխանն ստանալու համար նա ընտրեց 15 բանվոր, որոնք ունեն մոտավորապես նույն որակավորումը, տարիքը և աշխատանքի փորձը, պատահական կարգով նրանց բաշխեց երեք մանրադիտակների վրա աշխատելու, յուրաքանչյուր մանրադիտակի վրա՝ 5 հոգի: Այնուհետև նա գրանցեց մանրամասի երկարությունը չափելու համար յուրաքանչյուր բանվորի ծախսած ժամանակը (Y, վայրկյաններով): Տվյալները բերված են աղյուսակում:

	1	2	3	4	5	Միջինը
A_1	25.40	26.31	24.10	23.74	25.10	24.93
A_2	23.40	21.80	23.50	22.75	21.60	22.61
A_3	20.00	22.20	19.75	20.60	20.40	20.59

:

$\alpha = 0.01$ նշանակալիության մակարդակով հարկավոր է ստուգել H_0 վարկածը, ըստ որի երեք մանրադիտակներն ապահովում են միևնույն արտադրողականությունը:

Լուծում: Նախ հաշվենք $\overline{\mathbf{y}}$ ընդհանուր միջինը, $\overline{\mathbf{y}}_l$ խմբային միջինները ($l = \overline{1,3}$), նմուշային միջխմբային $\hat{s}_1^2(\mathbf{y})$ և ներխմբային $\hat{s}_2^2(\mathbf{y})$ ցրվածքները: Աղյուսակի տվյալներով ստանում ենք՝

$$\bar{\mathbf{y}} = (25.40 + 26.31 + ... + 23.40 + ... + 20.40)/15 \approx 22.71,$$
$$\bar{\mathbf{y}}_1 \approx 24.93,\ \bar{\mathbf{y}}_2 \approx 22.61,\ \bar{\mathbf{y}}_3 \approx 20.59,$$
$$\hat{s}_1^2(\mathbf{y}) = 47.164/2 \approx 23.582,\ \hat{s}_2^2(\mathbf{y}) = 11.0532/12 \approx 0.9211:$$

Այս տվյալներով կազմենք միագործոն ցրվածքային վերլուծության աղյուսակը:

Աղբյուրը	Ազատության աստիճանները	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
Գործոն	$3-1=2$	23.5820	25.60
Մնացորդ	$15-3=12$	0.9211	
Ընդհանուրը	$15-1=14$	4.1584	

Ստուգենք H_0 վարկածը: Միագործոն ցրվածքային վերլուծության աղյուսակից գտնում ենք, որ $\hat{F}(\mathbf{y}) \approx 23.582/0.9211 \approx 25.60$, մինչդեռ $F_{0.99}(2,12) = 6.93$, այսինքն՝ H_0-ն ժխտվում է:

10.3. Ցրվածքների համասեռության վարկածի ստուգում

Նախորդ ենթաբաժնում միագործոն ցրվածքային վերլուծության խնդրում մենք ենթադրեցինք, որ դիտման արդյունքների ցրվածքը բոլոր խմբերի համար ունի միևնույն արժեքը՝ σ^2: Սակայն հնարավոր են իրավիճակներ, երբ հետազոտողն այդ ենթադրությունը չի կարող ընդունել առանց էական վերապահումների, ուստի առաջանում է ցրվածքների համասեռության վարկածի ստուգման խնդիրը: Գլուխ 9-ում նման խնդիր լուծված է երկու հանույթների դեպքում: Այս ենթաբաժնում նույն խնդիրը լուծվում է $L \geq 2$ հանույթների դեպքում:

Ցրվածքների համասեռությանը վերաբերող վարկածը հետևյալն է՝

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_L^2 :$$

Դիցուք կատարված են Y_l հանույթի N_l անկախ դիտումներ՝

$$\mathbf{Y}_l = (Y_{1l}, Y_{2l}, ..., Y_{N_l l}),\ l = \overline{1, L} :$$

Նշանակենք՝

$$\hat{s}^2_{\min} = \min(\hat{s}^2(\mathbf{Y}_1), \hat{s}^2(\mathbf{Y}_2), ..., \hat{s}^2(\mathbf{Y}_L)),$$
$$\hat{s}^2_{\max} = \max(\hat{s}^2(\mathbf{Y}_1), \hat{s}^2(\mathbf{Y}_2), ..., \hat{s}^2(\mathbf{Y}_L)),$$

որտեղ $\hat{s}^2(\mathbf{Y}_l)$-ը l-րդ խմբի նմուշային ցրվածքն է, $l = \overline{1, L}$:

H_0 վարկածն ստուգելու համար օգտագործում են հետևյալ վիճակագրին՝

$$\hat{F}_{\max} = \hat{s}^2_{\max}/\hat{s}^2_{\min},$$

որն ունի $\mathcal{F}(L, N')$ բաշխում, որտեղ $N' = [\sum_l N_l/L] - 1 = [N/L] - 1$, իսկ $[a]$-ն a թվի ամբողջ մասն է: Կրիտիկական տիրույթն այստեղ աջակողմյան է, ուստի որպես կրիտիկական կետ ընդունվում է $F_{1-\alpha}(L, N')$-քանորդիչը: Եթե $\hat{F}_{\max} > F_{1-\alpha}(L, N')$, ապա H_0 վարկածը ժխտվում է:

<u>Օրինակ</u> 2: Մանրադիտակների խնդրում $\alpha = 0.05$ համար ստուգել ցրվածքների համասեռության վարկածը:

Լուծում: Օրինակ 1-ի տվյալներով երեք խմբերի համար հաշվարկում ենք՝

$$\hat{s}^2(\mathbf{y}_1) \approx 1.065,\ \hat{s}^2(\mathbf{y}_2) \approx 0.778,\ \hat{s}^2(\mathbf{y}_3) \approx 0.920 :$$

Քանի որ $N_1 = N_2 = N_3 = 5$, $N' = 5$, ստանում ենք՝

$$\hat{F}_{\max} = 1.065/0.778 \approx 1.369 < F_{0.95}(3, 5) = 15.5:$$

Հետևաբար, H_0 վարկածը չի ժխտվում, այսինքն՝ ընդունում ենք, որ երեք ցրվածքների միջև էական տարբերություններ չկան (ցրվածքները համասեռ են): Նկատենք, որ այս արդյունքը թույլ է տալիս օրինակ ...ում ստացված $\hat{s}^2(\mathbf{y}_2) \approx 0.921$ արժեքն ընդունել որպես σ^2 ցրվածքի վիճակագրական գնահատական, որն ստացվել է ազատության 12 աստիճանով: Սակայն որպես σ^2 ցրվածքի գնահատական չի կարելի վերցնել $\hat{s}^2(y_1, y_2, y_3) \approx 4.1584$ արժեքը, որն ստացվում է $N - 1 = 14$ ազատության աստիճանով, քանի որ խմբերի սպասելիների հավասարության մասին վարկածը չի ընդունվել:

10.4. Պատահականացված բլոկների եղանակ

Մենք դիտարկեցինք ցրվածքային վերլուծության միագործոն մոդելը, որի հիման վրա կարելի է ստուգել հետազոտվող գործոնի L խմբերի սպասելիների հավասարության մասին վարկածը: Այդ մոդելն անվանում են նաև լիովին պատահականացված, նկատի ունենալով փորձերի այն ծրագիրը, ըստ որի N փորձերի «դիտորդներն» ընդունվում են համասեռ: Սակայն որոշ դեպքերում դիտորդները կարող են միմյանցից տարբերվել, ուստի հարկավոր է ստուգել արդյունքային հատկանիշի վրա դիտորդների հնարավոր անհամասեռության ազդեցությունը: Նման դեպքերում դիտման արդյունքները խմբավորվում են **բլոկներում**, այնպես որ յուրաքանչյուր բլոկ պարունակի միայն դիտման՝ համասեռ համարվող արդյունքներ, որոնք արդեն պատահական ձևով բաշխվում են ըստ գործոնի մակարդակների (տե՛ս օրինակ 3-ը): Այսպիսով, խնդիր է ծագում՝ վարկածներ ստուգել ոչ միայն խմբերի, այլև սպասելիների՝ ըստ բլոկների հավասարության վերաբերյալ:

Նկատենք, որ պատահականացված բլոկների եղանակը դյուրությամբ կարելի է դիտարկել որպես *երկգործոն* խնդրի լուծման եղանակ, եթե բլոկները դիտարկենք որպես երկրորդ գործոնի մակարդակներ և ենթադրենք, որ այդ գործոնների մակարդակներից յուրաքանչյուր զույգի համար կատարված է ընդամենը մեկ դիտում: Օրինակ, այս գլխի սկզբում դիտարկված՝ շինարարական բրիգադների օրինակում կարող է խնդիր առաջանալ, թե արդյո՞ք հերթափոխի ժամը (ցերեկային, երեկոյան) որևէ ազդեցություն ունի շինարարության ծավալների վրա, այսինքն, կարելի է հերթափոխի ժամը դիտարկել որպես արդյունքային հատկանիշի վրա ազդող հնարավոր գործոն:

<u>*Օրինակ*</u> 3: Ագարակատերն ուսումնասիրում է մշակաբույսերի մի քանի տեսակների բերքատվությունը: Եթե բոլոր հողամասերն ըստ բերքատվության համարյա միանման են, ապա մշակաբույսերի բաշխումն ըստ հողամասերի պետք է կատարվի լիովին պատահական ծրագրով: Սակայն հաճախ հողամասերն էապես տարբերվում են մեկը մյուսից, իսկ դա փորձնական տվյալների մեջ լրացուցիչ ցրվածություն է առաջացնում: Տարատեսակության ազդեցությունը վերացնելու նպատակով փորձի համար նախատեսված մակերեսը բաժանում են մասերի (բլոկների) այնպես, որ յուրաքանչյուր բլոկի սահմաններում հողի որակը լինի համարյա համասեռ (սակայն այդ տեսակետից բլոկների միջև կարող է լինել էական տարբերություն): Այնուհետև բլոկներից յուրաքանչյուրը բաժանում են փորձադաշտերի՝ ըստ մշակաբույսերի քանակի: Տեսակների բաշխումն ըստ

Հողակտորների բլոկները, k	Բույս I A_1	Բույս II A_2	Բույս III A_3	Բույս IV A_4	Ընդամենը
1	70	61	82	74	287
2	77	75	88	76	316
3	76	67	90	80	313
4	80	63	96	76	315
5	84	66	92	84	326
6	78	68	98	86	330
Ընդամենը	465	400	546	476	1887
Միջինը	77.5	66.67	91	79.33	78.625

փործադաշտերի կատարվում է պատահականորեն:

Դիցուք՝ դիտարկվում են չորս տեսակի մշակաբույսեր, որոնք պետք է բաշխվեն 24 հողակտորի վրա (յուրաքանչյուր բլոկում կա չորս հողակտոր): Բերքատվության (Y) տվյալները պատահականորեն վերցված յուրաքանչյուր հողակտորի վրա և յուրաքանչյուր բույսի համար (պայմանական միավորներով) բերված են աղյուսակում, ընդ որում 0-ն համապատասխանում է ամենացածր, իսկ 100-ը՝ ամենաբարձր բերքատվությանը:

Մշակաբույսերի բերքատվության խնդրում ցրվածքային վերլուծությունը կշարունակենք օրինակ 4-ում:

Իսկ այժմ նշանակենք K-ով բլոկների քանակը, L-ով՝ խմբերի քանակը:

Պատահականացված բլոկների եղանակը հենվում է հետևյալ մոդելի վրա՝

$$Y_{kl} = m_{kl} + Z_{kl}, \ k = \overline{1,K}, \ l = \overline{1,L},$$

որտեղ m_{kl}-ը k-րդ բլոկին և l-րդ խմբին համապատասխանող Y_{kl} արդյունքային հատկանիշի սպասելին է, Z_{kl}-երը հատկանիշի պատահական շեղումներն են սպասելիից, որոնք անկախ են և բաշխված են $\mathcal{N}(0,\sigma^2)$ օրենքով: Ինչպես միագործոնի դեպքում, այստեղ էլ կարելի է սպասելիներն արտահայտել հետազոտվող գործոնի և բլոկների միջինների միջոցով՝ $m_{kl} = m+m_{k.}+m_{.l}$, $k = \overline{1,K}$, $l = \overline{1,L}$, որտեղ $m = (1/KL)\sum\limits_{k,l} m_{kl}$, $m_{k.} = (1/L)\sum\limits_{l} m_{kl} - m$, $m_{.l} = (1/K)\sum\limits_{k} m_{kl} - m$, $\sum\limits_{k} m_{k.} = \sum\limits_{l} m_{l.} = 0$: $m_{k.}$-ն և $m_{.l}$-ն արտահայտում են, համապատասխանաբար, k-րդ բլոկի և l-րդ խմբի ազդեցության չափը Y-ի վրա: $m_{k.} = const$, $k = \overline{1,K}$, նշանակում է, որ բլոկները «համասեռ» են: Նմանապես, $m_{.l} = const$, $l = \overline{1,L}$ նշանակում է, որ դիտարկվող գործոնը Y-ի վրա չի ազդում:

Նշանակենք

$$\mathbf{Y} = (1/KL)\sum_{l}\sum_{k} Y_{kl}, \ \mathbf{Y}_{.l} = (1/K)\sum_{k} Y_{kl}, \ \mathbf{Y}_{k.} = (1/L)\sum_{l} Y_{kl},$$

Հիշյալ ազդեցությունների առկայությունը հայտնաբերելու համար օգտագործում են **նմուշային միջխմբային ցրվածքը՝**

$$\hat{s}^2_{.1}(\mathbf{Y}) = \frac{1}{L-1}\sum_{l}\sum_{k=1}^{K}(\mathbf{Y}_{.l} - \mathbf{Y})^2,$$

նմուշային միջբլոկային ցրվածքը՝

$$\hat{s}^2_{1.}(\mathbf{Y}) = \frac{1}{K-1}\sum_{l}\sum_{k=1}^{K}(\mathbf{Y}_{k.} - \mathbf{Y})^2,$$

նմուշային մնացորդային ցրվածքը՝

$$\hat{s}^2_{2}(\mathbf{Y}) = \frac{1}{(L-1)(K-1)}\sum_{l}\sum_{k=1}^{K}(Y_{kl} - \mathbf{Y}_{.l} - \mathbf{Y}_{k.} + \mathbf{Y})^2,$$

նմուշային ընդհանուր ցրվածքը՝

$$\hat{s}^2(\mathbf{Y}) = \frac{1}{KL-1}\sum_{l}\sum_{k}(Y_{kl} - \mathbf{Y})^2 :$$

Կարելի է ցույց տալ, որ

$$(KL-1)\hat{s}^2(\mathbf{Y}) = (L-1)\hat{s}^2_{.1}(\mathbf{Y}) + (K-1)\hat{s}^2_{1.}(\mathbf{Y}) + (L-1)(K-1)\hat{s}^2_{2}(\mathbf{Y}) : \quad (2)$$

Զրոյական վարկածը, ըստ որի խմբերի սպասելիներն իրար հավասար են (խմբային գործոնը բացակայում է), ունի հետևյալ տեսքը՝

$$H_0^{(1)} : \ m_{.1} = m_{.2} = ... = m_{.L}:$$

Հաշվի առնելով, որ (2)-ի աջ մասի գումարելիները H_0 վարկածի դեպքում անկախ պատահական մեծություններ են, կարող ենք կազմել համապատասխան F-հարաբերությունը՝ $\hat{F}(\mathbf{Y}) = \hat{s}^2_{.1}(\mathbf{Y})/\hat{s}^2_{2}(\mathbf{Y})$ վիճականին, որը բաշխված է $\mathcal{F}(L-1,(K-1)(L-1))$ օրենքով: Եթե $\hat{F}(\mathbf{y}) > F_{1-\alpha}(L-1,(K-1)(L-1))$, ապա $H_0^{(1)}$ վարկածը ժխտվում է:

Զրոյական վարկածը, ըստ որի բլոկների սպասելիները իրար հավասար են (բլոկները «համասեռ» են), ունի հետևյալ տեսքը՝

$$H_0^{(2)}: m_{1.} = m_{2.} = ... = m_{K.}:$$

Այս դեպքում $\hat{F}(\mathbf{Y}) = \hat{s}^2_{1.}(\mathbf{Y})/\hat{s}^2_2(\mathbf{Y})$ վիճակագիրը բաշխված է $\mathcal{F}(K-1,(K-1)(L-1))$ օրենքով: Եթե $\hat{F}(\mathbf{y}) > F_{1-\alpha}(K-1,(K-1)(L-1))$, ապա $H_0^{(2)}$ վարկածը ժխտվում է:

Պատահականացված բլոկների ցրվածքային վերլուծության տվյալները ընդունված է միավորել աղյուսակում՝

Աղբյուրը	Ազատ. աստիճ.	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
Գործոն	$L-1$	$\hat{s}^2_{.1}(\mathbf{y}) = \sum\limits_l \sum\limits_k (\overline{\mathbf{y}}_{.l} - \overline{\mathbf{y}})^2/(L-1)$	$\hat{F}(\mathbf{y}) = \hat{s}^2_{.1}(\mathbf{y})/\hat{s}^2_2(\mathbf{y})$
Բլոկ	$K-1$	$\hat{s}^2_{1.}(\mathbf{y}) = \sum\limits_l \sum\limits_k (\overline{\mathbf{y}}_{k.} - \overline{\mathbf{y}})^2/(K-1)$	$\hat{F}(\mathbf{y}) = \hat{s}^2_{1.}(\mathbf{y})/\hat{s}^2_2(\mathbf{y})$
Մնացորդ	$(K-1)\times$ $\times(L-1)$	$\hat{s}^2_2(\mathbf{y}) = \sum\limits_l \sum\limits_k (y_{kl} - \overline{\mathbf{y}}_{.l} - \overline{\mathbf{y}}_{k.} + \overline{\mathbf{y}})^2/(K-1)(L-1)$	
Ընդհանուր	$KL-1$	$\hat{s}^2(\mathbf{y}) = \sum\limits_l \sum\limits_k (y_{kl} - \overline{\mathbf{y}})^2/(KL-1)$	

:

Օրինակ 4: Նախորդ օրինակի տվյալներով կատարել ցրվածքային վերլուծություն:

Լուծում: Ունենք՝ $K = 6$, $L = 4$, $\overline{\mathbf{y}}_{.1} = 465$, $\overline{\mathbf{y}}_{.2} = 400$, $\overline{\mathbf{y}}_{.3} = 546$, $\overline{\mathbf{y}}_{.4} = 476$, $\overline{\mathbf{y}}_{1.} = 287$, $\overline{\mathbf{y}}_{2.} = 316$, $\overline{\mathbf{y}}_{3.} = 313$, $\overline{\mathbf{y}}_{4.} = 315$, $\overline{\mathbf{y}}_{5.} = 326$, $\overline{\mathbf{y}}_{6.} = 330$:

Հետևաբար, կատարելով հաշվարկները, կստանանք՝

$\hat{s}^2_{.1}(\mathbf{y}) = 1787.46/3 \approx 595.820$, $\hat{s}^2_{1.}(\mathbf{y}) = 283.38/5 \approx 56.676$, $\hat{s}^2_2(\mathbf{y}) = 224.79/15 \approx 14.986$:

Պատահականացված բլոկների ցրվածքային վերլուծության համապատասխան աղյուսակն է՝

Աղբյուրը	Ազատության աստիճանները	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
Գործոն	$4-1=3$	$\hat{s}^2_{.1}(\mathbf{y}) \approx 595.820$	$\hat{F}(\mathbf{y}) \approx 39.76$
Բլոկ	$6-1=5$	$\hat{s}^2_{1.}(\mathbf{y}) \approx 56.676$	$\hat{F}(\mathbf{y}) \approx 3.782$
Մնացորդ	$(6-1)(4-1)=15$	$\hat{s}^2_2(\mathbf{y}) \approx 14.986$	
Ընդհանուրը	$(6)(4)-1=23$	$\hat{s}^2(\mathbf{y}) \approx 99.81$	

:

Օգտվելով ստացված տվյալներից՝ $\alpha = 0.05$ դեպքում գտնում ենք, որ խմբային գործոնի ազդեցության բացակայության մասին $H_0^{(1)}$ վարկածի համար $\hat{F}(\mathbf{y}) \approx 39.76 > F_{0.95}(3,15) = 3.29$, այսինքն՝ այն ժխտվում է, որտեղից եզրակացնում ենք, որ դիտարկվող մշակաբույսերի միջին բերքատվությունները էապես տարբեր են:

Բլոկների պատահականացման արդյունավետությունն ստուգելու համար դիմենք $H_0^{(2)}$ վարկածին, ըստ որի՝ բլոկների միջև նշանակալի տարբերություն չկա: Քանի որ այս դեպքում $\hat{F}(\mathbf{y}) \approx 3.78 > F_{0.95}(5,15) = 2.90$, ապա $H_0^{(2)}$ վարկածը ժխտվում է, հետևաբար, հողակտորների բլոկների միջև տարբերությունը նշանակալի է:

10.5. Երկգործոն ցրվածքային վերլուծություն

Այս ենթաբաժնում մենք կընդհանրացնենք նշված մոդելները այն իրավիճակների համար, երբ դիտարկվում են արդյունքային հատկանիշի վրա ազդող երկու գործոն (A և B): Ընդ որում մենք կդիտարկենք այն պարզ դեպքը, երբ A և B գործոններին համապատասխանող նմուշները ունեն միևնույն ծավալը (տարբեր ծավալների դեպքը դիտարկվում է առանց դժվարության): Մեզ հետաքրքրում են հետևյալ հարցերը, թե արդյո՞ք դիտման արդյունքների ցրվածության վրա ազդեցություն ունի A կամ B գործոնը, արդյո՞ք երկու գործոնները ազդում են իրարից անկախ:

Նախ դիտարկենք այն դեպքը, երբ երկու գործոնների մակարդակներից յուրաքանչյուր զույգի համար կատարված է մեկ դիտում: Հեշտ է նկատել, որ այս դեպքում մենք կարող ենք լիովին կիրառել պատահականացված բլոկների եղանակը, երկրորդ գործոնի մակարդակները դիտարկելով որպես բլոկներ: Ուստի համապատասխան երկգործոն ցրվածքային վերլուծության աղյուսակը համընկնում է պատահականացված բլոկների ցրվածքային վերլուծության աղյուսակի հետ:

Փոքր-ինչ ավելի բարդ է այն դեպքը, երբ երկու գործոնների մակարդակների յուրաքանչյուր զույգի համար կատարված են մեկից ավելի դիտումներ: Նախ դիտարկենք օրինակ, որը պարզաբանում է երկգործոն (ինչպես նաև բազմագործոն) ցրվածքային վերլուծության մոդելի կառուցվածքը:

<u>Օրինակ</u> 5: Կատարվել է հետազոտություն, որի նպատակն է պարզել ալկոհոլի ոչ մեծ քանակության ազդեցությունը կոմպյուտերային օպերատորների աշխատունակության վրա: Դիտարկվել է երկու գործոն՝ օպերատորի օգտագործած ալկոհոլի քանակությունը (A գործոն) և կատարած աշխատանքի բարդությունը (B գործոն): Փորձի համար ընտրվել են 12 կոմպյուտերային օպերատորներ՝ մոտավորապես նույն աշխատանքային կարողություններով: Նրանք պատահականորեն բաժանվել են երեք խմբի, յուրաքանչյուր խմբում՝ չորսական: Խմբերի անդամներին առաջարկվել է ընդունել ալկոհոլի միևնույն չափը, ընդ որում առաջին խմբում՝ փոքր (A_1), երկրորդ խմբում՝ միջին (A_2), իսկ երրորդ խմբում՝ մեծ (A_3) չափը: Այնուհետև յուրաքանչյուրին հանձնարարվել է տպագրել մեկական էջ պատահական եղանակով ընտրված տեխնիկական (B_1) կամ ոչ տեխնիկական (B_2) բովանդակությամբ տեքստ (ինչը, ըստ հետազոտողի մտահղացման, արտահայտում է տեքստի բարդությունը): Օպերատորների թույլ տված սխալների քանակը (Y) ներկայացված է աղյուսակում:

Ընդունած ալկոհոլի չափը A	Տեխնիկական տեքստ (B_1)	Ոչ տեխնիկական տեքստ (B_2)
փոքր (A_1)	5, 3	0, 2
միջին (A_2)	12, 14	3, 6
մեծ (A_3)	18, 21	10, 7

Աղյուսակի տվյալների թռուցիկ վերլուծությունն իսկ ցույց է տալիս, որ ալկոհոլային գործոնն ազդում է օպերատորի աշխատունակության վրա: Երևում է նաև տպագրվող տեքստի բարդության գործոնի ազդեցությունը: Սակայն անհրաժեշտ է ստուգել այն վարկածը, ըստ որի գործոնները (միասին, թե առանձին-առանձին) ազդեցություն չունեն օպերատորի աշխատունակության վրա:

Այս տվյալների երկգործոն ցրվածքային վերլուծությունը կատարված է օրինակ 6-ում:

Նշանակենք K-ով A գործոնի մակարդակների քանակը, L-ով՝ B գործոնի մակարդակների քանակը, $N_{kl} = N$-ով՝ A և B գործոնների մակարդակներից յուրաքանչյուր զույգի համար կատարած դիտումների քանակը, Y_{kln}-ով՝ A գործոնի k-րդ և B գործոնի l-րդ մակարդակների դեպքում n-րդ դիտման արդյունքը: Երկգործոն ցրվածքային վերլուծության մոդելն ունի հետևյալ տեսքը՝

$$Y_{kln} = m_{kl} + Z_{kln}, \; k = \overline{1,K}, \; l = \overline{1,L}, \; n = \overline{1,N},$$

որտեղ m_{kl}-ը A գործոնի k-րդ մակարդակին և B գործոնի l-րդ մակարդակին համապատասխանող Y_{kln} արդյունքային հատկանիշի սպասելին է, Z_{kln}-երը հատկանիշի անկախ պատահական շեղումներն են սպասելիից, որոնք բաշխված են $\mathcal{N}(0, \sigma^2)$ օրենքով:

Սպասելիները արտահայտվում են գործոնների միջինների միջոցով՝

$$m_{kl} = m + m_{k..} + m_{.l.} + m_{kl.},$$

որտեղ

$$m = (1/KL)\sum_k\sum_l m_{kl},\ m_{k..} = (1/L)\sum_l m_{kl} - m,\ m_{.l.} = (1/K)\sum_k m_{kl} - m,$$

$$m_{kl.} = m + m_{kl} - (1/K)\sum_k m_{kl} - (1/L)\sum_l m_{kl},\ \sum_k m_{k..} = \sum_l m_{.l.} = \sum_k\sum_l m_{kl.} = 0:$$

Երկգործոն մոդելում $m_{k..}$-ն արտահայտում է Y-ի վրա A գործոնի k-րդ մակարդակի ազդեցության չափը, $m_{.l.}$-ը՝ B գործոնի l-րդ մակարդակի, իսկ $m_{kl.}$-ը՝ A գործոնի k-րդ ու B գործոնի l-րդ մակարդակների համատեղ ազդեցության չափերը Y-ի վրա:

Դիտարկենք հետևյալ նմուշային միջինները՝

$$\mathbf{\overline{Y}}_{kl.} = \sum_n Y_{kln}/N,\ \mathbf{\overline{Y}}_{k..} = \sum_l\sum_n Y_{kln}/LN,\ \mathbf{\overline{Y}}_{.l.} = \sum_k\sum_n Y_{kln}/KN, \mathbf{\overline{Y}} = \sum_k\sum_l\sum_n Y_{kln}/KLN,$$

և հետևյալ նմուշային ցրվածքները՝

ընդհանուր նմուշային ցրվածքը՝

$$\hat{s}^2(\mathbf{Y}) = \sum_n\sum_k\sum_l(Y_{kln} - \mathbf{\overline{Y}})^2/(KLN-1),$$

A **գործոնին համապատասխանող նմուշային ցրվածքը՝**

$$\hat{s}^2_A(\mathbf{Y}) = LN\sum_k(\mathbf{\overline{Y}}_{k..} - \mathbf{\overline{Y}})^2/(K-1),$$

B **գործոնին համապատասխանող նմուշային ցրվածքը՝**

$$\hat{s}^2_B = KN\sum_l(\mathbf{\overline{Y}}_{.l.} - \mathbf{\overline{Y}})^2/(L-1),$$

A և *B* **գործոնների համատեղ ազդեցությանը համապատասխանող նմուշային ցրվածքը՝**

$$\hat{s}^2_{AB}(\mathbf{Y}) = N\sum_k\sum_l(\mathbf{\overline{Y}}_{kl.} - \mathbf{\overline{Y}}_{k..} - \mathbf{\overline{Y}}_{.l.} + \mathbf{\overline{Y}})^2/(K-1)(L-1),$$

և **մնացորդային նմուշային ցրվածքը՝**

$$\hat{s}^2_2(\mathbf{Y}) = \sum_n\sum_k\sum_l(Y_{kln} - \mathbf{\overline{Y}}_{kl.})^2/KL(N-1):$$

Ցույց է տրվում, որ

$$(KLN-1)\hat{s}^2(\mathbf{Y}) =$$
$$= (K-1)\hat{s}^2_A(\mathbf{Y}) + (L-1)\hat{s}^2_B(\mathbf{Y}) + (K-1)(L-1)\hat{s}^2_{AB}(\mathbf{Y}) + KL(N-1)\hat{s}^2_2(\mathbf{Y}):$$

Երկգործոն ցրվածքային վերլուծության մոդելը հանգում է, ըստ էության, երկընտրանքային վարկածների երեք տարբեր համակարգերի ստուգմանը: Առաջին համակարգը վերաբերում է A գործոնի ազդեցությանը: Այս դեպքում զրոյական վարկածն է՝

$$H_0^{(1)} : m_{1..} = m_{2..} = = ... = m_{K..} = 0:$$

Զրոյական վարկածն ստուգելու համար ճևավորենք F-հարաբերությունը, ելնելով տվյալների նորմալ բաշխվածության մասին ենթադրությունից և գործոնների մակարդակներից յուրաքանչյուր զույգի համար կատարված դիտումների անկախությունից: Եթե $H_0^{(1)}$ վարկածը ճիշտ է, ապա

$$\hat{F}_A(\mathbf{Y}) = \hat{s}^2_A(\mathbf{Y})/\hat{s}^2_2(\mathbf{Y}) \sim \mathcal{F}(K-1, KL(N-1)),$$

հետևաբար, եթե $\hat{F}_A(\mathbf{y}) > F_{1-\alpha}(K-1, KL(N-1))$, ապա $H_0^{(1)}$ վարկածը ժխտվում է:

Երկրորդ համակարգը վերաբերում է B գործոնի ազդեցությանը: Զրոյական վարկածը այս դեպքում հետևյալն է՝

$$H_0^{(2)} : m_{.1.} = m_{.2.} = ... = m_{.L.}:$$

Եթե $H_0^{(2)}$ վարկածը ճիշտ է, ապա

$$\hat{F}_B(\mathbf{Y}) = \hat{s}^2_B(\mathbf{Y})/\hat{s}^2_2(\mathbf{Y}) \sim \mathcal{F}(L-1, KL(N-1)),$$

հետևաբար, $\hat{F}_B(\mathbf{y}) > F_{1-\alpha}(L-1, KL(N-1))$ դեպքում $H_0^{(2)}$ վարկածը ժխտվում է:

Երրորդ համակարգը վերաբերում է A և B գործոնների համատեղ ազդեցությանը: Այս դեպքում զրոյական վարկածը հետևյալն է՝

$$H_0^{(3)} : m_{kl.} = 0, \ k = \overline{1,K}, \ l = \overline{1,L}:$$

$H_0^{(3)}$-ը նշանակում է, որ A և B գործոնները փոխկապակցված ազդեցություն չունեն արդյունքային հատկանիշի վրա: Եթե $H_0^{(3)}$ վարկածը ճիշտ է, ապա

$$\hat{F}_{AB}(\mathbf{Y}) = \hat{s}^2_{AB}(\mathbf{Y})/\hat{s}^2_2(\mathbf{Y}) \sim \mathcal{F}((K-1)(L-1), KL(N-1)) :$$

Եթե $\hat{F}_{AB}(\mathbf{y}) > F_{1-\alpha}((K-1)(L-1), KL(N-1))$, ապա $H_0^{(3)}$ վարկածը ժխտվում է:

Երկգործոն ցրվածքային վերլուծության տվյալները միավորում են հետևյալ աղյուսակում՝

Աղբյուրը	Ազատության աստիճանները	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
A	$K-1$	$\hat{s}^2_A(\mathbf{y})$	$\hat{F}_A(\mathbf{y}) = \hat{s}^2_A(\mathbf{y})/\hat{s}^2_2(\mathbf{y})$
B	$L-1$	$\hat{s}^2_B(\mathbf{y})$	$\hat{F}_B(\mathbf{y}) = \hat{s}^2_B(\mathbf{y})/\hat{s}^2_2(\mathbf{y})$
AB	$(K-1)(L-1)$	$\hat{s}^2_{AB}(\mathbf{y})$	$\hat{F}_{AB}(\mathbf{y}) = \hat{s}^2_{AB}(\mathbf{y})/\hat{s}^2_2(\mathbf{y})$
Մնացորդը	$KL(N-1)$	$\hat{s}^2_2(\mathbf{y})$	
Ընդհանուրը	$LKN-1$		

:

Օրինակ 6: Կատարել օրինակ 5-ում դիտարկված՝ օպերատորների աշխատունակության տվյալների ցրվածքային վերլուծությունը:

Լուծում: Նախ հաշվենք նմուշային միջինները ու ցրվածքները՝

$K = 3, \ L = 2, \ N = 2,$

$\overline{\mathbf{y}}_{1..} = 10, \overline{\mathbf{y}}_{2..} = 35, \overline{\mathbf{y}}_{3..} = 56, \overline{\mathbf{y}}_{.1.} = 73, \overline{\mathbf{y}}_{.2.} = 28, \overline{\mathbf{y}}_{11.} = 8, \overline{\mathbf{y}}_{12.} = 2, \overline{\mathbf{y}}_{21.} = 26, \overline{\mathbf{y}}_{22.} = 9,$ $\overline{\mathbf{y}}_{31.} = 39, \overline{\mathbf{y}}_{32.} = 17, \ \hat{s}^2_A(\mathbf{y}) \approx 132.5833, \ \hat{s}^2_B(\mathbf{y}) \approx 168.7497, \ \hat{s}^2_{AB}(\mathbf{y}) \approx 17.25016, \ \hat{s}^2_2(\mathbf{y}) \approx 3.25 :$

Իսկ հիմա ստուգենք A, B և AB գործոնների ազդեցության նշանակալիության վարկածները $\alpha = 0.01$ համար: Կատարելով անհրաժեշտ հաշվումները, կստանանք՝

$\hat{F}_A(\mathbf{y}) = 132.5833/3.25 \approx 40.79 > F_{0.99}(2,6) = 10.92,$

$\hat{F}_B(\mathbf{y}) = 168.7497/3.25 \approx 51.92 > F_{0.99}(1,6) = 13.75,$

$\hat{F}_{AB}(\mathbf{y}) = 16.7375/3.25 \approx 5.15 < F_{0.99}(2,6) = 10.92 :$

Այս օրինակի ցրվածքային վերլուծության թվային արդյունքները ներկայացված են աղյուսակում:

Աղբյուրը	Ազատության աստիճանները	Միջին քառակուսին	$\hat{F}(\mathbf{y})$
A (ալկոհոլ)	2	$\hat{s}^2_A(\mathbf{y}) \approx 132.5833$	$\hat{F}_A(\mathbf{y}) \approx 40.79$
B (տեքստ)	1	$\hat{s}^2_B(\mathbf{y}) \approx 168.7497$	$\hat{F}_B(\mathbf{y}) \approx 51.92$
AB	2	$\hat{s}^2_{AB}(\mathbf{y}) \approx 16.7375$	$\hat{F}_{AB}(\mathbf{y}) \approx 5.15$
Մնացորդ	6	$\hat{s}^2_2(\mathbf{y}) \approx 3.25$	
Ընդամենը	11		

:

Ինչպես տեսնում ենք, A (ալկոհոլ) և B (տեքստի բարդություն) գործոնների ազդեցությունը օպերատորի աշխատունակության վրա զգալի է, և դրանցից յուրաքանչյուրը գործում է մյուսից անկախ, այսինքն A և B գործոնների փոխկապակցված ազդեցությունը բացակայում է:

Գլուխ 11
Զույգային գծային ռեգրեսիա և հարաբերակցություն

Կանխագուշակումը՝ լայն առումով, յուրաքանչյուր գիտության գործնական նպատակն է:

Հարալդ Կրամեր

11.1. Գաղափար ռեգրեսիայի և հարաբերակցության մասին

Մաթեմատիկական վիճակագրության՝ **ռեգրեսիային վերլուծություն** կոչվող բաժինը նվիրված է վիճակագրական տվյալների օգնությամբ մեծությունների միջև կախվածության ուսումնասիրության եղանակներին:

Դիցուք, հետազոտվում է երկու հատկանիշների՝ X-ի և Y-ի փոխկապվածությունը: Պայմանավորվենք X-ով նշանակել անկախ փոփոխականը, և Y-ով՝ նրանից կախված փոփոխականը:

Եթե Y մեծության կախումը X-ից ֆունկցիոնալ է՝ X-ի յուրաքանչյուր արժեքին համապատասխանում է Y մեծության որոշակի արժեք, գրում են՝ $Y = \phi(X)$: Ֆունկցիոնալ կախումների հանդիպում ենք, օրինակ, ֆիզիկայում: Նկատենք, որ երբ X-ը պատահական մեծություն է, ապա նրանից ֆունկցիոնալ կախման մեջ գտնվող Y մեծությունը նույնպես պատահական է: Նման պատահական մեծությունների օրինակներ դիտարկված են 2-րդ գլխում:

Սակայն հաճախ հանդիպում են **հավանականային (ստոխաստիկ) կախվածության** դեպքեր, երբ X-ը պատահական կամ ոչ պատահական մեծություն է, իսկ Y-ը՝ պատահական մեծություն, որը կախված է X-ից: Ունենալով $\mathbf{E}(Y|X)$-ի պայմանական բաշխման ֆունկցիան, կարելի է X-ի կամայական x արժեքի համար մոտավորապես կանխագուշակել Y-ի արժեքները, օրինակ, վստահության միջակայքի ճշտությամբ: Ռեգրեսիային վերլուծությունը հիմնականում զբաղվում է $(\mathbf{X}, \mathbf{Y})$ երկչափ նմուշի հիման վրա Y-ի պայմանական սպասելիի և այլ բնութագրիչների գնահատման խնդիրներով:

Տնտեսագիտական խնդիրներում, որպես կանոն, հետազոտվող մեծությունների համատեղ բաշխման մասին բավարար չափով նախնական տեղեկություններ չեն լինում, հետազոտողը ստիպված հավանականային կախումները մոտարկում է տարբեր մոդելներով, որոնք հիմնավորվում և ճշգրտվում են վիճակագրական տվյալներով:

Նշենք, որ հավանականային կախում մենք արդեն դիտարկել ենք ցրվածքային վերլուծության խնդիրներում: Օրինակ, միագործոն ցրվածքային վերլուծության խնդրում X փոփոխականի դերում է A գործոնը, որի $l = 1, 2, ..., L$ մակարդակները կարելի է համարել X-ի արժեքներ, և դրանց համապատասխանում են Y_l պատահական մեծությունները:

> Y**-ի ռեգրեսիայի ֆունկցիա** X**-ի նկատմամբ** կոչվում է $\mathbf{E}(Y|X = x) = \mathbf{E}(Y|x)$ պայմանական սպասելին, դիտարկված որպես x-ի ֆունկցիա:

Եթե x-ի փոփոխումից $\mathbf{E}(Y|X = x)$-ը փոփոխվում է, ապա ասում են, որ Y-ի և X-ի միջև կա **հարաբերակցային կախվածություն**: Հակառակ դեպքում ասում են, որ Y-ի և X-ի միջև հարաբերակցային կապ չկա:

Սովորաբար ռեգրեսիայի ֆունկցիան օգտագործվում է կանխագուշակման նպատակով, ուստի մենք կօգտագործենք նաև $g(x) = \mathbf{E}(Y|X = x)$ նշանակումը, այսինքն կընդունենք, որ $X = x$ արժեքի համար (որը կարող է նաև նմուշում դիտարկված չլինել) $y = g(x)$ թիվը Y-ի համապատասխան կանխագուշակված միջին արժեքն է: Համապատասխան ենթատեքստում $g(x)$ ֆունկցիան կանվանենք նաև **կանխագուշակիչ**:

Ռեգրեսիայի վերլուծությունը զբաղվում է հետևյալ երեք խնդիրներով.

ա) ռեգրեսիայի մոդելի ընտրություն, որը կատարվում է ռեգրեսիայի ֆունկցիայի վերաբերյալ որոշակի նախնական տեղեկությունների և ենթադրությունների հիման վրա,

բ) ընտրված մոդելը բնութագրող անհայտ պարամետրերի գնահատում,

գ) ռեգրեսիայի մոդելի վերաբերյալ կիրառական խնդրից ծագող վարկածների ստուգում:

Այս գլխում կծանոթանանք **զույգային գծային ռեգրեսիայի** մոդելին: Այնուհետև կսահմանենք **հարաբերակցության** գաղափարը և դրա հետ սերտորեն կապված մի շարք բնութագրիչներ, որոնք առնչվում են ռեգրեսիային վերլուծությանը, դրանք են՝ **մասնակի հարաբերակցության գործակիցը**, **հարաբերակցային քանորդը** և, որպես հարաբերակցության գործակցի առանձնահատուկ տարբերակ, **Սպիրմենի հարաբերակցության կարգակարգային գործակիցը**: Հաջորդ գլխում կդիտարկենք **բազմաչափ**, **«գծայնացված»** և **կորագիծ ռեգրեսիայի** մոդելները:

11.2. Զույգային գծային ռեգրեսիա

Նախ պարզաբանենք այն բնական հարցը, թե ինչպես է ծագել «ռեգրեսիա» տերմինը, որը բառացիորեն նշանակում է «հետադիմություն»: Այն ներմուծել է անգլիացի հոգեբան և մարդաբան Ֆ. Հալտոնը մարդու հասակի՝ ժառանգաբար փոխանցվելու հատկությունն ուսումնասիրելիս: Հայրերի և որդիների հասակն արտահայտող վիճակագրական տվյալները վերլուծելիս նա խմբավորել է որդիներին ըստ հայրերի, որոնց հասակը բոլոր հայրերի միջին հասակից տարբերվում է x դյույմով: Պարզվել է՝ այդ ձևով խմբավորված որդիների միջին հասակը բոլոր որդիների միջին հասակից տարբերվում է x դյույմից պակաս չափով: Այս երևույթը նա անվանել է «հետադիմություն (ռեգրեսիա) դեպի միջին վիճակը»:

Այս խնդրում դիտարկվող երևույթի «հետադիմություն» բնորոշումը բնական է: Սակայն գոյություն ունեն կախումներ, որոնք բնական կլինեին անվանել «առաջադիմություն», այդ պատճառով կախումներ ուսումնասիրելու վիճակագրական եղանակը կիրառվում է որպես «ռեգրեսիային վերլուծություն»: Մենք էլ կօգտագործենք միջազգային «ռեգրեսիա», «ռեգրեսիայի ֆունկցիա» տերմինները:

Դիցուք, տրված է (X, Y) երկչափ պատահական մեծության նմուշը՝

$$(\mathbf{X}, \mathbf{Y}) = \{(X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)\}:$$

Մեր նպատակն է այս տվյալների հիման վրա ստանալ $\mathbf{E}(Y|X = x)$ ռեգրեսիայի ֆունկցիայի վիճակագրական հանգունակը:

Հաճախ օգտակար է մինչև համապատասխան հաշվարկներ կատարելը վիճակագրական տվյալները պատկերել **ցրվածության գծապատկերի** ձևով (տե՛ս գլուխ 6): Աչքի անցկացնելով այդ պատկերները, հետազոտողը նկատում է X-ից Y-ի կախման հիմնական միտումները և կարող է ընտրել ռեգրեսիայի ֆունկցիայի համապատասխան դասը:

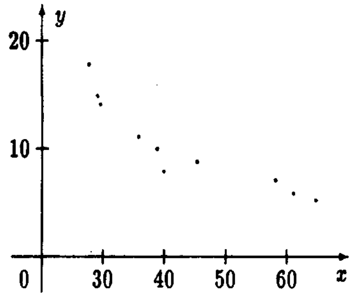
Օրինակ 1: Հանրախանութի պատասխանատու աշխատողը նկատել է, որ խանութի աշխատակիցների կարգապահությունը կախված է նրանց տարիքից: Ուսումնասիրելով պատահականորեն նշված 10 աշխատակիցների տարիքը և մեկ տարվա ընթացքում նրանց՝ աշխատանքից բացակայելու օրերի քանակը, նա ստացել է հետևյալ աղյուսակը՝

Դիտման համարը, n	1	2	3	4	5	6	7	8	9	10
Տարիքը, X (տարի)	27	61	37	23	46	58	29	36	64	40
Օրերի քանակը, Y (օր)	15	6	10	18	9	7	14	11	5	8

:

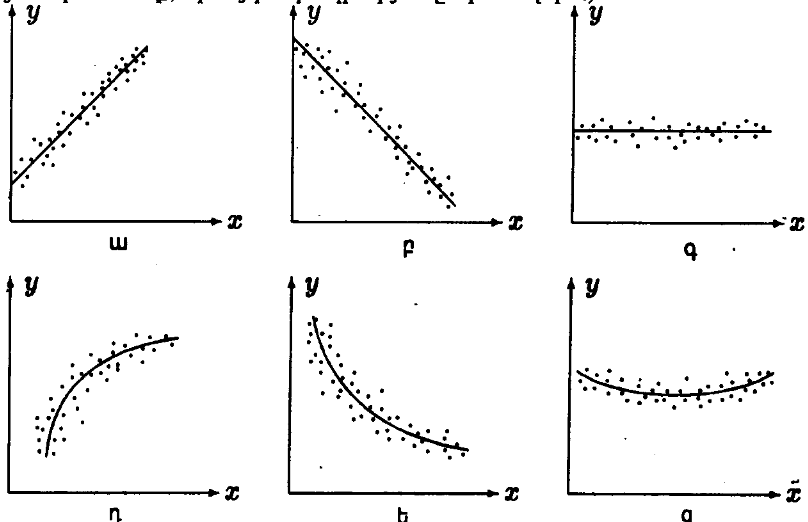
Կառուցել բերված տվյալների ցրվածության գծապատկերը:

Լուծում: Որպես X անկախ փոփոխական վերցնենք աշխատակցի տարիքը, իսկ որպես կախյալ Y՝ աշխատանքի չներկայանալու օրերի քանակը:



Նկար 1: Բացակայությունների մասին օրինակի ցրվածության գծապատկերը:

Նկարից կարելի է առաջադրել այս խնդրում ռեգրեսիայի ֆունկցիայի՝ գծային լինելու վարկածը (մենք հետագայում կտեսնենք, որ այդ ենթադրությունը հիմնավոր է):



Նկար 2: Հավանականային կախումների որոշ տեսակները:

Ցրվածության գծապատկերի վրա ընդունված է պատկերել նաև X-ից Y-ի հավանականային կախումն արտահայտող ռեգրեսիայի ֆունկցիայի գծապատկերը, որը որոշում են վիճակագրական տվյալների օգնությամբ: Նկար 2-ում պատկերված են հավանականային կախումների ցրվածության գծապատկերների և ռեգրեսիայի ֆունկցիաների պարզագույն օրինակներ. աճող գծային կախում (ա), նվազող գծային կախում (բ), Y-ը անկախ է X-ից (գ), ոչ գծային աճող կախում (դ), ոչ գծային նվազող կախում (ե), գոգավոր կախում (զ):

Ռեգրեսիային վերլուծությունում ընդունված է սահմանափակվել երևույթների այնպիսի մոդելներով, որոնցում տարանջատում են ուսումնասիրվող երևույթի ոչ պատահական միտումը (ռեգրեսիայի ֆունկցիայի ձևով) և պատահականությունն արտահայտող բաղադրիչը (գումարվող պատահական մեծության ձևով):

Դիտարկենք ռեգրեսիայի ֆունկցիայի պարզագույն դեպքը, երբ այն գծային է՝

$$y(x) = \mathrm{E}(Y|X = x) = a_0 + a_1 x :$$

a_0 և a_1 պարամետրերը կոչվում են **ռեգրեսիայի գործակիցներ**:

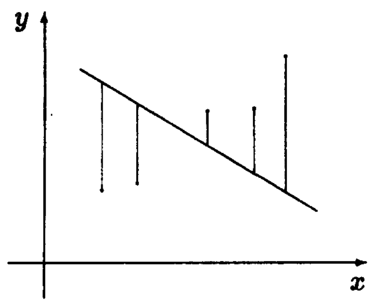
Ռեգրեսիայի գծային մոդելն է՝

$$Y = a_0 + a_1 X + Z,\ n = \overline{1, N} :$$

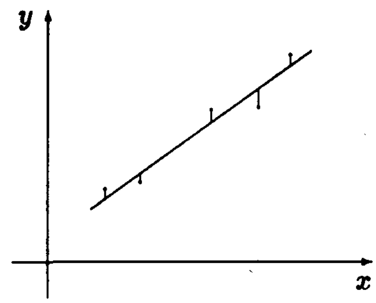
Այստեղ $Z \sim \mathcal{N}(0,\sigma^2)$, որտեղ σ^2 ցրվածքն անհայտ է: Անհայտ են նաև a_0 և a_1 պարամետրերը, որոնց նմուշային գնահատականները կնշանակենք $\hat{a}_0$ և $\hat{a}_1$:

$\hat{y} = \hat{a}_0 + \hat{a}_1 x$ առնչությունը կոչվում է **նմուշային գծային ռեգրեսիայի ֆունկցիա**:

$\hat{a}_0$ և $\hat{a}_1$ գործակիցները հնարավոր է ստանալ տարբեր գնատումների օգնությամբ, որոնք կարելի է իրար հետ համեմատել ըստ ցրվածության գծապատկերի կետերի և տվյալ եղանակով ստացված ուղղի փոխադարձ դասավորության: Որքան ավելի սերտորեն են դասավորված կետերը ուղղի նկատմամբ, այնքան ավելի գերադասելի է տվյալ եղանակը: Այստեղից հանգում ենք նմուշային ռեգրեսիայի ֆունկցիայի որոշման մի եղանակի, որի օգնությամբ որոշված ուղիղը լավագույնն է, այսինքն Y-ի դիտված և կանխագուշակված արժեքները նվազագույն չափով են հեռացված միմյանցից (տե՛ս նկարներ 3-ը և 4-ը): Այդ մոտեցումը հայտնի է որպես **փոքրագույն քառակուսիների եղանակ** (որի սկզբունքը նշված է նաև գլուխ 7-ում):



Նկար 3: Ուղիղը ճիշտ չի ընտրված:



Նկար 4: Ուղիղը ճիշտ է ընտրված:

Դիցուք տրված է (X,Y) երկչափ պատահական վեկտորի փորձնական $(\mathbf{X},\mathbf{Y})$ նմուշը: Պահանջվում է գտնել այնպիսի $\hat{y} = \hat{a}_0 + \hat{a}_1 x$ ուղիղ, որի համար

$$\sum_n (y_n - \hat{y}_n)^2 = \min_{\{a_0,a_1\}} \sum_n [y_n - (a_0 + a_1 x_n)]^2, \tag{1}$$

որտեղ $\hat{y}_n = \hat{a}_0 + \hat{a}_1 x_n$:

Փնտրվող ուղիղը գտնելու նպատակով հաշվում են (1)-ի աջ կողմում գրված գումարի մասնակի ածանցյալները ըստ $\hat{a}_0$ և $\hat{a}_1$ գործակիցների և հավասարեցնում 0-ի: Ստացվում է, այսպես կոչված, **նորմալ հավասարումների համակարգը**՝

$$\overline{\mathbf{y}} = \hat{a}_0 + \hat{a}_1 \overline{\mathbf{x}},$$

$$\sum_n x_n y_n = \hat{a}_0 \sum_n x_n + \hat{a}_1 \sum_n x_n^2 :$$

Լուծելով այն, կստանանք՝

$$\hat{a}_0 = \overline{\mathbf{y}} - \hat{a}_1 \overline{\mathbf{x}}, \quad \hat{a}_1 = (\sum_n x_n y_n - N \overline{\mathbf{x}}\,\overline{\mathbf{y}})/(\sum_n x_n^2 - N \overline{\mathbf{x}}^2) : \tag{2}$$

Այսպիսով, փոքրագույն քառակուսիների եղանակով ստացվում են ռեգրեսիայի գործակիցների $\hat{a}_0 = \hat{a}_0(\mathbf{x},\mathbf{y})$ և $\hat{a}_1 = \hat{a}_1(\mathbf{x},\mathbf{y})$ գնահատականները, և, հետևաբար, նմուշային գծային ռեգրեսիայի ֆունկցիան՝ $\hat{y} = \overline{\mathbf{y}} + \hat{a}_1(x - \overline{\mathbf{x}})$:

Իսկ ինչպե՞ս «չափել» գնահատված նմուշային ռեգրեսիայի ֆունկցիայի համապատասխանությունը նմուշային տվյալներին, ինչպե՞ս որոշել դրա կանխագուշակիչ կարողությունը:

Նշանակենք

$$S^2(\mathbf{x},\mathbf{y}) = \sum_n (y_n - \overline{\mathbf{y}})^2 = \sum_n y_n^2 - (\sum_n y_n)^2/N :$$

Նույն միջինի նկատմամբ Y-ի կանխագուշակված արժեքների ցրվածությունը բնորոշվում է $S_1^2(\mathbf{x},\mathbf{y}) = \sum_n(\hat{y}_n - \overline{\mathbf{y}})^2$ գումարով:

Փոխարինելով Y-ի դիտված՝ y_n արժեքը կանխագուշակված $\hat{y}_n$-ով, թույլ ենք տալիս $y_n - \hat{y}_n$ սխալ (որը կոչվում է **մնացորդ**): Ուստի $S_2^2(\mathbf{x},\mathbf{y}) = \sum_n(y_n - \hat{y}_n)^2$ մեծությունը բնորոշում է կանխագուշակման գումարային քառակուսային սխալը, որը, ինչպես երևում է (1) արտահայտությունից, նվազագույնն է ռեգրեսիայի բոլոր հնարավոր ուղիղների բազմության վրա:

Կարելի է ցույց տալ, որ $S^2(\mathbf{x},\mathbf{y}) = S_1^2(\mathbf{x},\mathbf{y}) + S_2^2(\mathbf{x},\mathbf{y})$, որտեղ

$$S_1^2(\mathbf{x},\mathbf{y}) = \hat{a}_0\sum_n y_n + \hat{a}_1\sum_n x_ny_n - \left(\sum_n y_n\right)^2/N,$$

$$S_2^2(\mathbf{x},\mathbf{y}) = \sum_n y_n^2 - \hat{a}_0\sum_n y_n - \hat{a}_1\sum_n x_ny_n :$$

Հետևյալ հարաբերությունը կոչվում է **որոշակիության գործակից**՝

$$d_{YX}^2(\mathbf{x},\mathbf{y}) = \sum_n(\hat{y}_n - \overline{\mathbf{y}})^2/\sum_n(y_n - \overline{\mathbf{y}})^2 = S_1^2(\mathbf{x},\mathbf{y})/S^2(\mathbf{x},\mathbf{y}) :$$

Այն ցույց է տալիս, թե Y-ի ընդհանուր ցրվածության որ մասն է պայմանավորված ռեգրեսիային կախվածությամբ: Եթե $S_1^2(\mathbf{x},\mathbf{y}) = 0$, այսինքն՝ $\hat{y}_n = \overline{\mathbf{y}}$, $n = \overline{1,N}$, և, հետևաբար, ենթադրյալ ռեգրեսիայի ֆունկցիան կանխագուշակիչ կարողություն չունի, ապա $d_{YX}^2(\mathbf{x},\mathbf{y}) = 0$: Եթե $S_2^2(\mathbf{x},\mathbf{y}) = 0$, այսինքն՝ $\hat{y}_n = y_n$, $n = \overline{1,N}$, և, հետևաբար, կանխագուշակման սխալ չկա, ապա $d_{YX}^2(\mathbf{x},\mathbf{y}) = 1$: Երբեմն $d_{YX}^2(\mathbf{x},\mathbf{y})$-ի փոխարեն դիտարկում են $100d_{YX}^2(\mathbf{x},\mathbf{y})$-ը, արտահայտելով այն տոկոսներով:

Ռեգրեսիայի միջին քառակուսային սխալ անվանում են հետևյալ մեծությունը՝

$$\hat{s}_{YX}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_n(y_n - \hat{y}_n)^2/(N-2)} :$$

Այն արտահայտում է Y-ի դիտման արդյունքների ցրվածության մեծությունը փոքրագույն քառակուսիների եղանակով որոշված ռեգրեսիայի ուղղի նկատմամբ: Մենք հետագայում կտեսնենք, որ $\hat{s}_{YX}^2(\mathbf{x},\mathbf{y})$-ն անհայտ σ^2 ցրվածքի անշեղ գնահատական է:

<u>Օրինակ</u> 2: Պաղպաղակ վաճառողին անհրաժեշտ է իմանալ, թե ինչպիսի կախում կա պաղպաղակի վաճառքի և օդի օրվա ջերմաստիճանի միջև: Նա, օրեցօր գրանցելով օդի միջին ջերմաստիճանը ($X°C$) և օրական եկամուտը (հազար դրամ, Y), ստացել է աղյուսակում բերված տվյալները (տե՛ս առաջին երկու սյունակները):

x_n	y_n	x_n^2	y_n^2	x_ny_n	$\hat{y}$	$y_n - \hat{y}_n$
17	15.2	289	231.04	258.4	13.88	+1.32
21	16.8	441	282.24	352.8	17.78	−0.98
23	18.0	529	324.00	414.0	19.72	−1.72
24	20.5	576	420.25	492.0	20.70	−0.20
27	23.6	729	556.96	637.2	23.62	−0.02
28	22.5	784	506.25	630.0	24.59	−2.09
29	26.8	841	718.24	777.2	25.56	+1.24
31	29.0	961	841.00	899.0	27.51	+1.49
32	31.4	1024	985.96	1004.8	28.49	+2.91
33	30.6	1089	936.36	1009.8	29.46	+1.14
35	29.2	1225	368.64	672.0	31.41	−2.21
37	34.0	1369	1156.00	1258.0	33.35	+0.65
38	32.8	1444	1075.84	1246.4	34.33	−1.53
375	330.4	11301	8886.78	10001.6	330.4	0.00

Գտնենք պաղպաղակի վաճառքից ստացած եկամուտի գծային ռեգրեսիան օդի ջերմաստիճանի նկատմամբ, որոշակիության գործակիցը և ռեգրեսիայի միջին քառակուսային սխալը:

Լուծում: Համապատասխան միջանկյալ արդյունքները բերված են աղյուսակում: Տեղադրելով աղյուսակի վերջին տողի տվյալները $\hat{a}_0$-ի և $\hat{a}_1$-ի (2) բանաձևերի մեջ, կստանանք՝

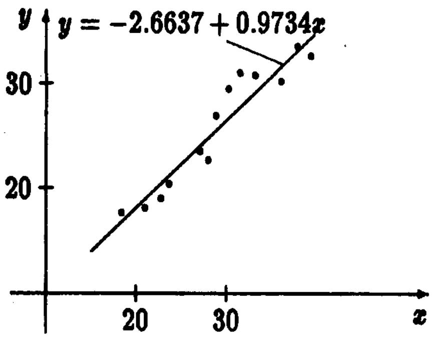
$$\hat{a}_1 = \frac{(13)(10001.6) - (375)(330.4)}{(13)(11301) - (375)^2} \approx \frac{6120.8}{6288} \approx 0.9734,$$

$$\bar{y} = 330.4/13 \approx 25.415, \quad \bar{x} = 375/13 \approx 28.846, \ \hat{a}_0 = 25.415 - (0.9734)(28.846) \approx -2.6637 :$$

Այսպիսով, $\hat{y} = -2.6637 + 0.9734x$:

Նկարում պատկերված են պաղպաղակի վաճառքի տվյալների ցրվածության և նմուշային ռեգրեսիայի ուղղի գծապատկերները:

Այստեղ ուղղի թեքության գործակիցը դրական է, ուստի օդի ջերմաստիճանի 1^0C-ով աճելու դեպքում ($\Delta x = 1$) կարելի է սպասել օրական եկամուտի աճ՝ 0.97 հազար դրամի չափով ($\Delta y \approx 0.97$):



Նմուշային ռեգրեսիայի ֆունկցիան կարելի է օգտագործել կանխագուշակման նպատակով, տեղադրելով x-ի պահանջվող արժեքը: Այսպես, եթե դիտարկվող խնդրում կրպակի տնօրենին հետաքրքրում է, թե որքան կլինի իր եկամուտը, երբ օդի ջերմաստիճանը լինի 30^0C, նա պետք է կատարի հետևյալ հաշվարկը՝ $\hat{y} = -2.6637 + (0.9734)(30) \approx 26.54$:

Ունեցած բոլոր տվյալներից ստացված 26.54 հազար դրամ կանխագուշակումն ավելի հիմնավորված է, քան հարևան 29^o և 31^o դեպքում դիտված եկամուտների միջինը՝ 27.9 հազար դրամը:

Աղյուսակում բերված են նաև $\hat{y}_n$-ի հաշվարկված արժեքները, ինչպես նաև դրանց և Y-ի դիտված արժեքների տարբերությունները:

Հաշվենք $S^2(\mathbf{x},\mathbf{y})$, $S_1^2(\mathbf{x},\mathbf{y})$, $S_2^2(\mathbf{x},\mathbf{y})$ մեծությունների նմուշային արժեքները՝

$S^2(\mathbf{x},\mathbf{y}) = 8886.78 - (330.4)^2/13 \approx 489.537,$

$S_1^2(\mathbf{x},\mathbf{y}) = (-2.6637)(330.4) + (0.9734)(10001.6) - (330.4)^2/13 \approx 458.227,$

$S_2^2(\mathbf{x},\mathbf{y}) = 8886.78 - (-2.6637)(330.4) - (0.9734)(10001.6) \approx 31.310,$

Որոշակիության գործակիցը հավասար է

$$d_{YX}^2(\mathbf{x},\mathbf{y}) = 458.227/489.537 \approx 0.936 :$$

Այսպիսով, կարելի է համարել, որ պաղպաղակի վաճառքից ստացված եկամուտի ցրվածքի 93.6 տոկոսը բացատրվում է ջերմաստիճանի փոփոխությամբ՝ ըստ դիտման օրերի, և միայն 6.4 տոկոսն է, որ կապված է ռեգրեսիայի մոդելում հաշվի չառնված այլ գործոնների հետ:

Ռեգրեսիայի միջին քառակուսային սխալն է $\hat{s}_{YX}(\mathbf{x},\mathbf{y}) = \sqrt{31.310/(13-2)} \approx 1.687$: Մենք այն կօգտագործենք հաջորդ օրինակներում՝ ռեգրեսիայի գործակիցների վստահության միջակայքերը հաշվարկելու համար:

Մինչ այժմ մենք որևէ ենթադրություն չենք արել X և Y պատահական մեծությունների բաշխման ֆունկցիաների վերաբերյալ, այսինքն՝ բոլոր բերված արդյունքները ճիշտ են (X, Y) երկչափ պատահական վեկտորի կամայական համատեղ բաշխման դեպքում: Իսկ այժմ ընդունենք, որ Y-ի պայմանական բաշխումը, երբ $X = x$, նորմալ է՝ $\mathcal{N}(a_0 + a_1 x, \sigma^2)$: Այդ դեպքում ռեգրեսիայի ֆունկցիան և նրա հետ առնչվող մեծություններն օժտված են հետևյալ հատկություններով:

1. $\mathbf{E}(S_2^2(\mathbf{X},\mathbf{Y})) = (N-2)\sigma^2$, ուստի

$$\hat{s}^2_{YX}(\mathbf{x},\mathbf{y}) = \frac{1}{N-2}\sum_n (y_n - \hat{a}_0 - \hat{a}_1 x_n)^2$$

σ^2 ցրվածքի անշեղ գնահատական է: Այն անվանում են նաև **մնացորդային ցրվածք** և նշանակում $\hat{\sigma}^2_{YX}(\mathbf{x},\mathbf{y})$, շեշտելով վերջինիս առնչությունը ռեգրեսիայի վերլուծության հետ:

Նորմալ բաշխվածության պայմանից բխում է, որ $\hat{\sigma}^2_{YX}(\mathbf{X},\mathbf{Y})/\sigma^2 \sim \chi^2(N-2)$:

Եթե $a_1 = 0$, ապա $\mathbf{E}(S_1^2(\mathbf{X},\mathbf{Y})) = \sigma^2$ և $S_1(\mathbf{X},\mathbf{Y})^2/\sigma^2 \sim \chi^2(1)$, ընդ որում $S_1^2(\mathbf{X},\mathbf{Y})$-ն և $S_2^2(\mathbf{X},\mathbf{Y})$-ն անկախ պատահական մեծություններ են: Այսպիսով, $S_1^2(\mathbf{x},\mathbf{y})$-ը նույնպես σ^2 ցրվածքի անշեղ գնահատական է (ազատության 1 աստիճանով)՝ $\hat{\sigma}^2 = S_1^2(\mathbf{x},\mathbf{y})$:

2. $\hat{a}_0(\mathbf{X},\mathbf{Y})$ և $\hat{a}_1(\mathbf{X},\mathbf{Y})$ պատահական մեծությունները բաշխված են համատեղ նորմալ օրենքով և a_0, a_1 գործակիցների զուգամետ և անշեղ գնատուներն են, որոնք անկախ են $S_1^2(\mathbf{X},\mathbf{Y})$ և $S_2^2(\mathbf{X},\mathbf{Y})$ պատահական մեծություններից:

3. Նմուշային ռեգրեսիայի $\hat{a}_0(\mathbf{X},\mathbf{Y})$ գործակցի ցրվածքը կլինի՝

$$\hat{\sigma}^2_{a_0} = \sigma^2 \sum_n x_n^2 / [N \sum_n (x_n - \overline{\mathbf{x}})^2],$$

իսկ a_0 գործակցի վստահության միջակայքի ծայրակետերն են՝

$$\hat{a}_0(\mathbf{x},\mathbf{y}) \pm t_{1-\alpha/2}\hat{\sigma}_{YX}(\mathbf{x},\mathbf{y})\{\sum_n x_n^2/[N\sum_n (x_n - \overline{\mathbf{x}})^2]\}^{1/2},$$

որտեղ $t_{1-\alpha/2}$-ը Ստյուդենտի $\mathbf{t}(N-2)$ բաշխման $(1-\alpha/2)$-քանորդիչն է:

4. Նմուշային ռեգրեսիայի $\hat{a}_1(\mathbf{X},\mathbf{Y})$ գործակցի ցրվածքը կլինի՝

$$\hat{\sigma}^2_{a_1} = \sigma^2 / \sum_n (x_n - \overline{\mathbf{x}})^2 :$$

Հետևաբար a_1-ի վստահության միջակայքը որոշվում է հետևյալ ծայրակետերով՝

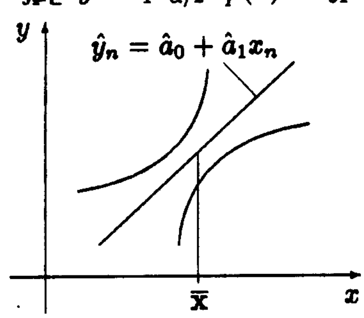
$$\hat{a}_1(\mathbf{x},\mathbf{y}) \pm t_{1-\alpha/2}\hat{\sigma}_{YX}(\mathbf{x},\mathbf{y})/[\sum_n x_n^2 - (1/N)(\sum_n x_n)^2]^{1/2} :$$

5. Y-ի կանխագուշակված (միջին) արժեքը տրված $X = x$-ի համար որոշվում է նմուշային ռեգրեսիայի ֆունկցիայի օգնությամբ՝ $\hat{y} = \hat{a}_0 + \hat{a}_1 x$ կամ՝ $\hat{y} = \overline{\mathbf{y}} + \hat{a}_1(x - \overline{\mathbf{x}})$:

$\hat{Y}$-ի միջին քառակուսային շեղումը տրվում է հետևյալ առնչությամբ՝

$$\hat{\sigma}_{\hat{Y}}(x) = \hat{\sigma}_{YX}(\mathbf{x},\mathbf{y})\left[1/N + (x-\overline{\mathbf{x}})^2/\sum_n (x_n - \overline{\mathbf{x}})^2\right]^{1/2},$$

իսկ $\hat{y}$-ի վստահության միջակայքը՝ $\hat{y} \pm t_{1-\alpha/2}\hat{\sigma}_{\hat{Y}}(x)$ ծայրակետերով:



Այստեղ պետք է նկատել, որ $\hat{Y}$-ի գնահատման ճշտությունը ամենամեծն է (այսինքն՝ $\hat{a}_1(\mathbf{X},\mathbf{Y})$-ի ցրվածքը ամենափոքրն է), երբ $x = \overline{\mathbf{x}}$, իսկ երբ x արժեքը հեռու է $\overline{\mathbf{x}}$-ից, ճշտությունը փոքր է: Այս փաստը բերում է հետևյալ եզրակացության՝ ռեգրեսիայի ֆունկցիան կանխագուշակման նպատակով օգտագործելիս պետք է դիտարկել x-ի այն արժեքները, որոնք բավականաչափ մոտ են $\overline{\mathbf{x}}$ միջին արժեքին: Այն խնդիրներում, որտեղ անհրաժեշտ է կանխագուշակել Y-ի արժեքները, որոնք համապատասխանում են հետազոտվող տիրույթի եզրակետերի մոտակայքում կամ տիրույթից դուրս գտնվող x-ի արժեքներին, ռեգրեսիայի վերլուծության եղանակները վատ են աշխատում:

Նկարում պատկերված է ռեգրեսիայի ուղիղը, իսկ նրա երկու կողմում՝ $\hat{y}$-ի վստահության միջակայքը x-ից կախված կորերի ձևով: Այդ կորերը, ինչպես երևում է ցրվածքի արտահայտությունից, հիպերբոլներ են:

$\hat{Y}$-ի միջին քառակուսային շեղման համար բերված արտահայտությունը վերաբերում է տվյալ x-ի համար Y պատահական մեծության կանխագուշակված արժեքին: Քանի որ Y-ի պայմանական ցրվածքը հավասար է σ^2, ապա Y-ի անկախ դիտման արդյունքի կանխագուշակված մեծությունը նախկինի նման կորոշվի $\hat{y}$-ով, բայց նրա ցրվածքն արդեն հավասար կլինի

$$\hat{\sigma}^2_Y(x) = \sigma^2\left[1 + 1/N + (x-\overline{\mathbf{x}})^2/\sum_n(x_n-\overline{\mathbf{x}})^2\right] :$$

Այստեղ նույնպես σ^2-ի փոխարեն կարելի է տեղադրել նրա նմուշային գնահատականը՝ $\hat{\sigma}^2_{YX}$-ը, ուստի Y-ի դիտման նոր արդյունքի համար կստանանք հետևյալ վստահության միջակայքը՝

$$\hat{y} \pm t_{1-\alpha/2}\hat{\sigma}_{YX}(\mathbf{X},\mathbf{Y})\left[1 + 1/N + (x-\overline{\mathbf{x}})^2/\sum_n(x_n-\overline{\mathbf{x}})^2\right]^{1/2},$$

որտեղ $t_{1-\alpha/2}$-ը $\mathrm{t}(N-2)$ բաշխման $(1-\alpha/2)$-քանորդիչն է:

Օրինակ 3: Օգտագործելով օրինակ 2-ի արդյունքները, հաշվենք ռեգրեսիայի a_0 և a_1 գործակիցների 95-տոկոսանոց վստահության միջակայքերը, պաղպաղակի վաճառքից սպասվող եկամուտի և դիտման արդյունքի 95-տոկոսանոց վստահության միջակայքերը, երբ օդի ջերմաստիճանը հավասար է 29^0C:

Լուծում: Աղյուսակից $\mathrm{t}(11)$-բաշխման համար ունենք $t_{0.975} = 2.201$: Կատարելով պահանջվող գործողությունները, a_0 գործակցի վստահության միջակայքի ծայրակետերի համար կստանանք՝

$$-2.6637 \pm (2.201)\left\{\frac{11301}{13[(11301)-(375)^2/13]}\right\}^{1/2}(1.687) \approx -2.6637 \pm 4.9778,$$

այսինքն՝ a_0 գործակցի 95-տոկոսանոց վստահության միջակայքն է՝ $-7.6415 \le a_0 \le 2.3141$:

Նման ձևով a_1 գործակցի վստահության միջակայքի ծայրակետերի համար կստանանք՝

$$0.9734 \pm (2.201)(1.687)/\left[11301-(375)^2/13\right]^{1/2} \approx 0.9734 \pm 0.1688,$$

այսինքն՝ a_1 գործակցի 95-տոկոսանոց վստահության միջակայքն է՝ $0.8046 \le a_1 \le 1.1422$:

Համապատասխանաբար, $x = 29$ արժեքի համար կստանանք՝

$$\hat{\sigma}_{\hat{Y}} = (1.687)\left\{1/13 + (29-28.846)^2/[11301-(375)^2/13]\right\}^{1/2} \approx 0.468,$$

$$-2.6637 + (0.9734)(29) \pm (2.201)(0.468) \approx 25.565 \pm 1.030 :$$

Հետևաբար, պաղպաղակի վաճառքից սպասվող եկամուտի վստահության միջակայքն է՝ $(24.54, 26.60)$: Իսկ երբ օդի ջերմաստիճանը հավասար է 20^0C, ստանում ենք $\hat{\sigma}_{\hat{Y}} \approx 0.824$, $15.0 \le \hat{Y} \le 18.6$: Y-ի դիտման անկախ արդյունքի վստահության միջակայքը $x = 29$ համար՝ $20.82 \le Y \le 30.31$: Ինչպես տեսնում ենք, Y-ի դիտման արդյունքը տրված x-ի համար կանխագուշակվում է ավելի մեծ սխալով, քան Y-ի սպասվող արժեքը:

6. Փոքրագույն քառակուսիների եղանակով $\hat{a}_1$-ի (2)-ում ստացված արտահայտությունը կարելի է ձևափոխել հետևյալ տեսքի՝

$$\hat{a}_1 = \frac{\sum_n (x_n - \overline{\mathbf{x}})(y_n - \overline{\mathbf{y}})}{\sum_n (x_n - \overline{\mathbf{x}})^2} = \frac{\sum_n (x_n - \overline{\mathbf{x}})(y_n - \overline{\mathbf{y}})}{\sqrt{\sum_n (x_n - \overline{\mathbf{x}})^2 \sum_n (y_n - \overline{\mathbf{y}})^2}} \sqrt{\frac{\sum_n (y_n - \overline{\mathbf{y}})^2}{\sum_n (x_n - \overline{\mathbf{x}})^2}} :$$

Այնուհետև, $\hat{a}_1$-ը կարելի է ներկայացնել հետևյալ ավելի պարզ տեսքով՝

$$\hat{a}_1 = \hat{\rho}_{XY}(\mathbf{x}, \mathbf{y}) \hat{s}(\mathbf{y}) / \hat{s}(\mathbf{x}),$$

որտեղ

$$\hat{\rho}_{XY}(\mathbf{x}, \mathbf{y}) = \sum_n (x_n - \overline{\mathbf{x}})(y_n - \overline{\mathbf{y}}) / \sqrt{\sum_n (x_n - \overline{\mathbf{x}})^2 \sum_n (y_n - \overline{\mathbf{y}})^2},$$

իսկ $\hat{s}(\mathbf{x})$-ը և $\hat{s}(\mathbf{y})$-ը, համապատասխանաբար, X-ի և Y-ի նմուշային միջին քառակուսային շեղումներն են:

$\hat{\rho}_{XY}(\mathbf{x}, \mathbf{y})$-ը կոչվում է X-ի և Y-ի միջև **նմուշային հարաբերակցության գործակից:**

Որոշակիացման և նմուշային հարաբերակցության գործակիցները կապված են հետևյալ առնչությամբ՝ $d^2_{YX}(\mathbf{x}, \mathbf{y}) = \hat{\rho}^2_{XY}(\mathbf{x}, \mathbf{y})$:

Այսպիսով, ռեգրեսիայի և հարաբերակցության գաղափարները սերտորեն կապված են միմյանց հետ: Գրականության մեջ ռեգրեսիային վերլուծությունը երբեմն անվանում են հարաբերակցային-ռեգրեսիային վերլուծություն:

Դիտարկենք ռեգրեսիային վերլուծության երրորդ՝ ռեգրեսիային մոդելին վերաբերող, վիճակագրական վարկածներ ստուգելու խնդիրը:

Վարկածները a_0 գործակցի մասին ունեն հետևյալ տեսքը՝ $H_0 : a_0 = a_0'$, $H_1 : a_0 \neq a_0'$, որտեղ a_0'-ն նախօրոք տրված թիվ է:

Դիտարկենք հետևյալ վիճականիշը՝

$$\hat{t}(\mathbf{X}, \mathbf{Y}) = (\hat{a}_0(\mathbf{X}, \mathbf{Y}) - a_0')[N \sum_n (X_n - \overline{\mathbf{X}})^2 / \sum_n X_n^2]^{1/2} / \hat{\sigma}_{YX}(\mathbf{X}, \mathbf{Y}) :$$

Հենվելով երկրորդ և երրորդ հատկությունների և գլուխ 2-ի արդյունքների վրա, կարելի է ցույց տալ, որ եթե H_0-ն ճիշտ է, ապա $\hat{t}(\mathbf{X}, \mathbf{Y}) \sim \mathrm{t}(N-2)$: Ուստի, եթե $|\hat{t}(\mathbf{x}, \mathbf{y})| > t_{1-\alpha/2}$, ապա H_0 վարկածը ժխտվում է: Նշանակալիության մակարդակն ընտրված է α:

a_1 գործակցի մասին վարկածներն ունեն այսպիսի տեսք՝ $H_0 : a_1 = a_1'$, $H_1 : a_1 \neq a_1'$, որտեղ a_1'-ը նախօրոք տրված թիվ է: Վերջինիս արժեքը թելադրվում է լուծվող կիրառական խնդրի պահանջներով: Օրինակ, հետազոտողին կարող է հետաքրքրել, թե a_1-ի համար տվյալ փորձաշարում ստացված նմուշային արժեքը համապատասխանու՞մ է արդյոք մի այլ փորձաշարում կամ որևէ այլ եղանակով ստացված a_1' արժեքին: Ավելի հաճախ հետազոտողն ուզում է իմանալ, թե a_1-ի համար ստացված նմուշային արժեքը (որը կարող է և հավասար չլինել 0-ի) իրո՞ք վկայում է այն մասին, որ Y-ի և X-ի միջև կա գծային հավանականային կախում (այսինքն՝ որ $a_1 \neq 0$): Այս դեպքում վարկածները ձևակերպելիս ընդունում են $H_0 : a_1 = 0$, ուստի H_0 վարկածն ստանում է այն իմաստը, որ գծային ռեգրեսիան x-ից կախված չէ:

H_0 վարկածի ստուգման համար օգտվում են

$$\hat{t}(\mathbf{X}, \mathbf{Y}) = (\hat{a}_1(\mathbf{X}, \mathbf{Y}) - a_1')[\sum_n (X_n - \overline{\mathbf{X}})^2]^{1/2} / \hat{\sigma}_{YX}(\mathbf{X}, \mathbf{Y})$$

վիճականից, որը H_0-ն ճիշտ լինելու դեպքում բաշխված է $\mathrm{t}(N-2)$ օրենքով: Եթե պարզվում է, որ $\left|\hat{t}(\mathbf{x}, \mathbf{y})\right| > t_{1-\alpha/2}(N-2)$, ապա H_0 վարկածը ժխտվում է:

Կարելի է ցույց տալ, որ այս խնդրի լուծումը տրվում է նաև

$$\hat{F}(\mathbf{X},\mathbf{Y}) = (N-2)S_1^2(\mathbf{X},\mathbf{Y})/S_2^2(\mathbf{X},\mathbf{Y})$$

վիճականու օգնությամբ, որը H_0-ն ճիշտ լինելու դեպքում ունի $\mathcal{F}(1, N-2)$-բաշխում: Հետևաբար, եթե $\hat{F}(\mathbf{x},\mathbf{y}) > F_{1-\alpha}(1, N-2)$, ապա H_0 վարկածը ժխտվում է:

Առաջին հայացքից թվում է, թե $\hat{t}(\mathbf{X},\mathbf{Y})$ և $\hat{F}(\mathbf{X},\mathbf{Y})$ պատահական մեծությունները, որոնք բաշխված են, համապատասխանաբար, t և $\mathcal{F}$ օրենքներով, էապես տարբերվում են: Սակայն, ինչպես տեսնում ենք, դրանք տալիս են լուծվող խնդրի միևնույն պատասխանը: Բանն այն է, որ դիտարկվող դեպքում տեղի ունեն հետևյալ հավասարությունները՝

$$\hat{F}(\mathbf{X},\mathbf{Y}) = \hat{a}_1^2 \sum_n (X_n - \overline{\mathbf{X}})^2/\hat{\sigma}_{YX}^2 = \hat{t}^2 \quad \text{և} \quad F_{1-\alpha}(1,\nu) = t_{1-\alpha/2}^2(\nu):$$

Այս դիտողությունը կարևոր է, քանի որ հաջորդ գլխում դիտարկվում են վարկածներ բազմաչափ գծային ռեգրեսիայի գործակիցների նշանակալիության մասին, որոնք ստուգվում են F-հարաբերության օգնությամբ:

Օրինակ 4: Օրինակներ 2-ի և 3-ի տվյալներով $\alpha = 0.05$ համար ստուգենք վարկած ռեգրեսիայի բացակայության մասին:

Լուծում: $\mathcal{F}(1, 11)$ բաշխման համար ունենք

$$\hat{F}(\mathbf{x},\mathbf{y}) = (11)(458.227)/31.310 \approx 161.0 > F_{0.95}(1, 11) = 4.84,$$

իսկ $\mathbf{t}(11)$ բաշխման համար՝

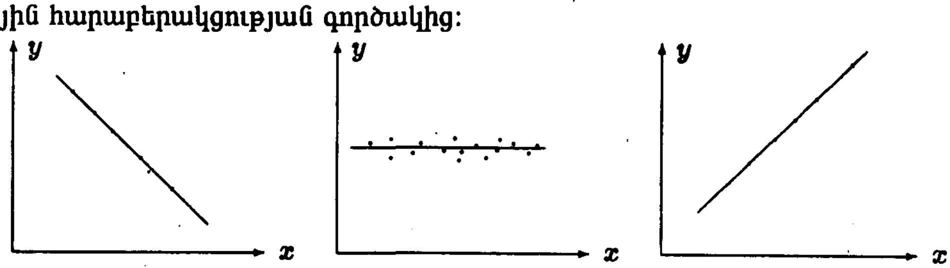
$$\hat{t}(\mathbf{x},\mathbf{y}) = [(0.9734)(11301 - (375)^2/13]/1.687 \approx 12.69 > t_{0.975}(11) = 2.2,$$

ուստի H_0 վարկածը ժխտվում է:

11.3. Զույգային հարաբերակցություն քանակական մեծությունների միջև

Նախորդ ենթաբաժնում նշվեց, որ ռեգրեսիային վերլուծությունում լուծվում է անկախ X փոփոխականի արժեքներով կախյալ Y փոփոխականի արժեքները կանխագուշակելու խնդիրը: Այնտեղ բերված բանաձևերը, օրինակները և գծապատկերները ցույց են տալիս, որ այդ խնդրի լուծման հաջողությունը էապես կախված է այն բանից, թե դիտման արդյունքները որքան մոտ են դասավորված ռեգրեսիայի ուղղի շուրջը: Մենք տեսանք, որ որոշակիության գործակիցը բնութագրում է դիտման արդյունքների և ռեգրեսիայի ուղղի փոխադարձ դասավորության մերձության աստիճանը և թվային մեծությամբ հավասար է հարաբերակցության (կոռելյացիայի) գործակցի քառակուսուն: Հետևաբար, **հարաբերակցության գործակիցը**, որի ընտրանքային համարժեքը բերված է նախորդ ենթաբաժնում, «երկու փոփոխականների գծային կախվածության աստիճանը չափող» մեծություն է:

Եթե ռեգրեսիային վերլուծությունն արդեն կատարված է, հայտնի են ռեգրեսիայի գործակիցները և որոշակիության գործակիցը, ապա ընտրանքային հարաբերակցության գործակցի թվային արժեքը որոշելու համար անհրաժեշտ է հաշվել $\sqrt{\hat{d}_{YX}^2(\mathbf{X},\mathbf{Y})}$-ը և՝ դրան վերագրելով ռեգրեսիայի $\hat{a}_1$ գործակցի նշանը, ստացված մեծությունն, ընդունել որպես ընտրանքային հարաբերակցության գործակից:



Նկարում պատկերված է երկու փոփոխականների գծային կախման (հարաբերակցության) տարբեր աստիճաններն արտահայտող երեք դեպք: ա) առավելագույն ացասական հարաբերակցություն ($\rho_{XY} = -1$), բ) հարաբերակցության բացակայություն

$(\rho_{XY} = 0)$ և գ) առավելագույն դրական հարաբերակցություն $(\rho_{XY} = +1)$:

Այն դեպքերում, երբ հասկանալի է, թե հարաբերակցության գործակիցը որ պատահական մեծություններին է վերաբերում, այն կնշանակենք ուղղակի ρ-ով:

Նմուշային հարաբերակցության գործակիցը որոշվում է հետևյալ բանաձևով՝

$$\hat{\rho}_{YX}(\mathbf{x},\mathbf{y}) = \sum_n (x_n - \overline{\mathbf{x}})(y_n - \overline{\mathbf{y}}) \Big/ \sqrt{\sum_n (x_n - \overline{\mathbf{x}})^2 \sum_n (y_n - \overline{\mathbf{y}})^2} =$$

$$= \frac{\overline{(x - \overline{\mathbf{x}})(y - \overline{\mathbf{y}})}}{\sqrt{\overline{(x - \overline{\mathbf{x}})^2}}\sqrt{\overline{(y - \overline{\mathbf{y}})^2}}} = \frac{\sum_n x_n y_n - (1/N)(\sum_n x_n)(\sum_n y_n)}{\sqrt{\sum_n x_n^2 - (1/N)(\sum_n x_n)^2}\sqrt{\sum_n y_n^2 - (1/N)(\sum_n y_n)^2}} : \qquad (3)$$

Օրինակ 5: Հաշվենք պաղպաղակի վաճառքից ստացված եկամուտի և օդի ջերմաստիճանի միջև նմուշային հարաբերակցության գործակիցը՝ օգտվելով օրինակ 2-ի համար հաշվարկված մեծություններից:

Լուծում: Քանի որ $d^2_{YX}(\mathbf{x},\mathbf{y}) \approx 0.936$, իսկ $\hat{a}_1 < 0$, ստանում ենք $\hat{\rho} = -\sqrt{0.936} \approx -0.967$:

Օրինակ 6: Հաշվենք աշխատակիցների բացակայություններին (Y) և նրանց տարիքի (X) միջև նմուշային հարաբերակցության գործակիցը՝ ընդունելով այդ փոփոխականները որպես պատահական մեծություններ:

Լուծում: Օգտվելով օրինակ 1-ի աղյուսակի տվյալներից՝ ստանում ենք

$$\sum_n x_n = 421,\ \sum_n x_n^2 = 19661,\ \sum_n y_n = 103, \sum_n y_n^2 = 1221,\ \sum_n x_n y_n = 3817,$$

հետևաբար,

$$\hat{\rho}(\mathbf{x},\mathbf{y}) = \frac{3817 - (421)(103)/10}{\sqrt{19661 - (421)^2/10}\sqrt{1221 - (103)^2/10}} \approx -0.933 :$$

Ինչպես տեսնում ենք, այս երկու օրինակներում էլ նմուշային հարաբերակցության գործակիցը բացասական է և բավականին մոտ է իր փոքրագույն արժեքին՝ -1-ին:

Նախորդ ենթաբաժնում ստացված $\hat{a}_1 = \hat{\rho}\ \hat{s}_Y/\hat{s}_X$ բանաձևը հնարավորություն է տալիս նույն վիճականին օգտագործել ինչպես ռեգրեսիայի վերլուծության, այնպես էլ հարաբերակցության բացակայության մասին վարկածներ ստուգելու համար:

Հիմնական վարկածները, որոնք անհրաժեշտ է լինում ստուգել, ունեն հետևյալ տեսքը՝

$$H_0 : \ \rho = \rho_0, \quad H_1 : \ \rho \neq \rho_0,$$

որտեղ ρ-ն տեսական հարաբերակցության գործակիցն է, ρ_0-ն նախօրոք ընտրված արժեք է: Երբ $\rho_0 = 0$, H_0 վարկածը նշանակում է, որ հարաբերակցություն չկա:

Այս վարկածները ստուգելու համար անհրաժեշտ է որոշել նմուշային հարաբերակցության գործակցի՝ որպես պատահական մեծության, բաշխման ֆունկցիան:

Դիտարկենք այն դեպքը, երբ (X, Y) զույգն ունի համատեղ նորմալ բաշխում և $\rho \neq 0$: N-ի փոքր արժեքների դեպքում $\hat{\rho}$-ի բաշխումն ունի բավականին բարդ տեսք և պիտանի չէ կոնկրետ հաշվարկների համար: N-ի մեծ արժեքների դեպքում կարելի է օգտվել այն հանգամանքից, որ $\hat{\rho}$-ի բաշխումը շատ մոտ է $\mathcal{N}(\rho, (1-\rho^2)/\sqrt{N})$ նորմալ բաշխմանը: Այստեղ պետք է հաշվի առնել նաև այն հանգամանքը, որ երբ ρ-ի բացարձակ արժեքը մոտ է 1-ին, նշված մոտարկումը կարելի է հուսալի համարել միայն $N > 500$ դեպքում:

$\rho = 0$ դեպքում $\hat{\rho}$-ի բաշխման ֆունկցիան ընդունում է ավելի պարզ տեսք, մասնավորապես, կարելի է ցույց տալ, որ

$$\hat{t}(\mathbf{X},\mathbf{Y}) = \hat{\rho}\sqrt{(N-2)/(1-\hat{\rho}^2)} \sim \mathbf{t}(N-2) :$$

Հենվելով այս փաստի վրա, կարելի է ստուգել հարաբերակցային կապի բացակայության մասին վարկածը, այն է՝ H_0 : $\rho = 0$, ընդդեմ H_1 : $\rho \neq 0$ վարկածի: Եթե հաշվարկների արդյունքում պարզվի, որ

$$\left|\hat{t}(\mathbf{x},\mathbf{y})\right| = |\hat{\rho}|\,\sqrt{N-2}/\sqrt{1-\hat{\rho}^2} > t_{1-\alpha/2}(N-2), \tag{4}$$

որտեղ $t_{1-\alpha/2}(N-2)$-ը Ստյուդենտի $\mathbf{t}(N-2)$ բաշխման $(1-\alpha/2)$-քանորդիչն է, ապա H_0 վարկածը կժխտվի:

Օրինակ 7: Օրինակ 5-ի տվյալների համար ստուգենք զրոյական վարկածը, ըստ որի պաղպաղակի վաճառքի և օդի ջերմաստիճանի միջև հարաբերակցություն չկա:

Լուծում: Կատարելով համապատասխան հաշվարկները, կստանանք

$$\hat{t}(\mathbf{x},\mathbf{y}) = |-0.967|\,\sqrt{11}/\sqrt{1-(-0.967)^2} \approx 12.59 :$$

Քանի որ $\hat{t}(\mathbf{x},\mathbf{y}) \approx 12.59 > t_{0.975}(11) = 2.20$, ապա H_0 վարկածը ժխտվում է:

Օրինակ 8: Օրինակ 6-ի տվյալների համար ստուգենք զրոյական վարկածը, ըստ որի աշխատակիցների բացակայությունների և տարիքի միջև հարաբերակցություն չկա, ընդունելով, որ (X,Y)-ը ունի համատեղ նորմալ բաշխում:

Լուծում: Կստանանք

$$\hat{t}(\mathbf{x},\mathbf{y}) = |-0.933|\,\sqrt{8}/\sqrt{1-(-0.933)^2} \approx 7.33 > t_{0.975}(8) = 2.31,$$

հետևաբար, հարաբերակցության բացակայության մասին վարկածն այստեղ նույնպես ժխտվում է:

Ընդհանուր դեպքում ($\rho \neq 0$) վարկածներ ստուգելիս հարմար է օգտվել 1921 թ. Ռ. Ֆիշերի կողմից առաջարկված՝ $\hat{\rho}$ պատահական մեծությունը *նորմալացնող ձևափոխությունից*, այն է՝

$$\hat{z} = \frac{1}{2}\ln\frac{1+\hat{\rho}}{1-\hat{\rho}} :$$

Ֆիշերի ձևափոխությունը կոչվում է նորմալացնող, որովհետև $\hat{z} = \hat{z}(\mathbf{X},\mathbf{Y})$ պատահական մեծության բաշխումը, արդեն $N = 20$ դեպքում, բավականին մոտ է նորմալ բաշխմանը հետևյալ պարամետրերով՝

$$\mathbf{E}(\hat{z}) = \frac{1}{2}\ln\frac{1+\rho}{1-\rho},\quad \mathbf{D}(\hat{z}) = \frac{1}{N-3} :$$

Այս փաստը հնարավորություն է տալիս հաշվել հարաբերակցության գործակցի վստահության միջակայքը՝ $\hat{\rho}_1 = \mathbf{th}\hat{z}_1 \le \rho \le \hat{\rho}_2 = \mathbf{th}\hat{z}_2$, որտեղ $\mathbf{th}z = (e^z - e^{-z})/(e^z + e^{-z})$-ը հիպերբոլական տանգենսի ֆունկցիան է,

$$\hat{z}_{1,2} = \frac{1}{2}\ln\frac{1+\hat{\rho}}{1-\hat{\rho}} \mp \frac{z_{1-\alpha/2}}{\sqrt{N-3}},$$

իսկ $z_{1-\alpha/2}$-ը $\mathcal{N}(0,1)$ նորմալ բաշխման $(1-\alpha/2)$-քանորդիչն է:

Օրինակ 9: Օգտագործելով Ֆիշերի ձևափոխությունը, որոշենք օրինակ 6-ում հաշվարկված՝ աշխատակիցների բացակայությունների և տարիքի հարաբերակցության գործակցի վստահության միջակայքը $1-\alpha = 0.95$ նշանակալիության մակարդակի վրա:

Լուծում: Կատարելով համապատասխան հաշվարկները $\hat{\rho} = -0.933$ համար, ստանում ենք՝

$$z_{1,2} \approx -1.681 \pm 0.693,\ \hat{\rho}_1 \approx -0.983,\ \hat{\rho}_2 \approx -0.757,$$

հետևաբար $[-0.983,\ -0.757]$ միջակայքը 0.95 հավանականությամբ «ծածկում է» հարաբերակցության գործակցի իրական արժեքը:

11.4. Պայմանական կամ մասնակի հարաբերակցություն

Դիցուք, դիտարկվում են $X_1, ..., X_M$ պատահական մեծությունները և հարաբերակցության գործակիցները դրանց միջև: Նշանակենք $R = \|\rho_{mm'}\|$, $m, m' = 1, ..., M$, որտեղ $\rho_{mm'}$-ը X_m և X'_m մեծությունների միջև հարաբերակցության գործակիցն է: R մատրիցը կոչվում է **հարաբերակցային մատրից**:

> X_1 և X_2 պատահական մեծությունների **պայմանական** կամ **մասնակի հարաբերակցության գործակից**, երբ մնացած $M-2$ մեծությունները սևեռված են, կոչվում է հետևյալ մեծությունը՝
> $$\rho_{12/34...M} = -R_{12}/(R_{11}R_{22})^{1/2}, \qquad (5)$$

որտեղ R_{mq}-ն $|R| = \det R$ որոշիչում ρ_{mq} տարրի հանրահաշվական լրացումն է: Մնացած տարրերի համար մասնակի հարաբերակցությունը որոշվում է նույն կերպ: Մասնավոր դեպքում, երբ $M = 3$, ստանում ենք

$$\rho_{12/3} = (\rho_{12} - \rho_{13}\rho_{23})/[(1-\rho_{12}^2)(1-\rho_{23}^2)]^{1/2}:$$

Նմուշային մասնակի հարաբերակցության գործակիցը ստացվում է, տեղադրելով (5) բանաձևի մեջ համապատասխան հարաբերակցության գործակիցների նմուշային արժեքները: Վարկածների ստուգումը և վստահության միջակայքերի կառուցումը կատարվում է ճիշտ այնպես, ինչպես սովորական հարաբերակցության դեպքում, ամենուրեք N-ը փոխարինելով $N-M$-ով, որտեղ M-ը սևեռված փոփոխականների քանակն է:

Մասնակի հարաբերակցության գործակիցն օգտագործում են այն դեպքերում, երբ հիմքեր կան ենթադրելու, որ հետազոտվող երկու փոփոխականները հարաբերակցված են մեկ այլ փոփոխականի հետ: Նման դեպքերում հաշվում են թե՛ սովորական, թե՛ մասնակի հարաբերակցության գործակիցները, որոնք հետո համեմատվում են իրար հետ: Եթե երկու մեծությունների միջև մասնակի հարաբերակցության գործակիցը սևեռված երրորդ մեծության դեպքում ավելի փոքր է, քան սովորական հարաբերակցության գործակիցը, ապա նշանակում է, որ վերջինս մասամբ պայմանավորված է այդ երրորդ մեծության ազդեցությամբ: Եթե մասնակի հարաբերակցության գործակիցն ավելի մեծ է, ապա երրորդ մեծությունը թուլացրել է դրանց կապը կամ «քողարկել» է այն: Սա է պատճառը, որ մասնակի հարաբերակցությունը երբեմն անվանում են «մաքրված կապ»:

Դիտարկենք մասնակի հարաբերակցության գործակցի օգտագործման մի հետաքրքիր օրինակ, որը պատկանում է Հուկերին (Journ. Roy. Stat. Soc., 1907, v. 70, 1):

Օրինակ 10: Խոտի կերակրատեսակների բերքատվությունն ուսումնասիրելու համար նա դիտարկել է 20 տարվա տվյալներ, որոնք բնութագրում են բերքատվությունը ($X_{0,n}$), գարնանային տեղումների քանակը ($X_{1,n}$) և գարնանը կուտակված «ակտիվ» ջերմությունը (5.5^0-ից բարձր) ($X_{2,n}$), $n = \overline{1,20}$: Տվյալները վերցվել են Անգլիայի այն շրջաններից, որոնք բնակլիմայական պայմաններով նման են, ինչը, հեղինակի մտահղացմամբ, երաշխավորում է տվյալների համասեռությունը: Հարաբերակցության գործակիցների համար ստացվել են հետևյալ գնահատականները՝ $\hat{\rho}_{01} \approx 0.80$, $\hat{\rho}_{02} \approx -0.40$, $\hat{\rho}_{12} \approx -0.56$: Ստուգելով $H_0 : \rho = 0$ վարկածը $\alpha = 0.1$ համար, ստանում ենք, որ (4) անհավասարության ձախ մասը՝ $|\hat{t}|$-ն, $\hat{\rho}_{01}$, $\hat{\rho}_{02}$ և $\hat{\rho}_{12}$ հարաբերակցության գործակիցների համար ընդունում է, համապատասխանաբար, 5.66, 1.85 և 2.87 արժեքները, մինչդեռ աջ մասը՝ $t_{1-\alpha/2}(18) = 1.73$, ուրեմն՝ հարաբերակցության երեք գործակիցներն էլ նշանակալի են:

Այս օրինակում հիմնական հարցը վերաբերում է եղանակային պայմանների ազդեցությանը խոտի բերքատվության վրա: Հարաբերակցությունների վերլուծությունից պարզվում է, որ բերքատվությունը սերտ կապի մեջ է տեղումների քանակի հետ՝ դրական հարաբերակցության գործակցով, իսկ կուտակված ջերմության հետ՝ բացասական հարաբերակցության գործակցով:

Իրո՞ք ջերմաստիճանի բարձրացումը բացասաբար է անդրադառնում բերքատվության վրա, թե՞ այստեղ միջանցիկ ազդեցություն ունի տեղումների քանակությունը: Կարելի է մտածել, որ ջերմաստիճանը նվազում է տեղումների հետևանքով, իսկ ջերմության տված օգուտն էապես փոքրանում է չոր եղանակի պատճառած վնասի հետևանքով: Այս երկու հնարավոր (իրարամերժ) բացատրություններից մեկը հիմնավորված համարելու համար հաշվենք մասնակի հարաբերակցության գործակիցները և ստուգենք $\hat{\rho}_{02/1}$ հարաբերակցության գործակցի նշանակալիությունը:

Լուծում: Կատարելով հաշվարկները, ստանում ենք

$$\hat{\rho}_{01/2} \approx 0.759,\ \hat{\rho}_{02/1} \approx 0.097,\ \hat{\rho}_{12/0} \approx -0.436:$$

Այս դեպքում (երբ ազատության աստիճանների քանակը պակասում է 1-ով և հավասարվում 17-ի) $\hat{t}(\mathbf{x},\mathbf{y}) \approx 0.40$, իսկ $t_{1-\alpha/2} = 1.80$, հետևաբար, $\hat{\rho}_{02/1}$-ի համար ստացված արժեքը նշանակալի չէ:

Ինչպես տեսնում ենք, բերքատվության ցուցանիշի և տեղումների քանակի հարաբերակցության գործակիցը համարյա նույնն է՝ անկախ ջերմության քանակից: Սակայն կուտակված ջերմության հետ բերքատվության պայմանական հարաբերակցության գործակիցը, երբ տեղումների քանակը սևեռված է, նախկինի համեմատ էապես նվազում է: Սա նշանակում է, որ, իսկապես, $\hat{\rho}_{02} \approx -0.40$ բացասական արժեքը տեղումների բարդ գործոնի ազդեցության հետևանք է:

11.5. Տարակարգային հարաբերակցություն

Որևէ հատկանիշով դասակարգված առարկաների հաջորդականությունը կոչվում է **կարգավորված**: Այդպիսի դասակարգման գործընթացը կոչվում է **կարգավորում**:

Կարգավորման օրինակ է X պատահական մեծության $x_1, x_2, ..., x_N$ նմուշի փոփոխման շարքը $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(N)}$ (տե՛ս գլուխ 6): Այստեղ նմուշի x_n տարրին համապատասխանության է դրվում i_n բնական թիվը, որն արտահայտում է փոփոխման շարքում այդ անդամի զբաղեցրած տեղի համարը և կոչվում է նմուշի n-րդ անդամի **տարակարգ**: Այսպիսով, $x_1, x_2, ..., x_N$ նմուշին համապատասխանում է $i_1, i_2, ..., i_N$ **տարակարգային հաջորդականությունը**, որը $1, 2, ..., N$ բնական թվերի տեղափոխություն է:

Տարակարգային հաջորդականություններ կարելի է դիտարկել նաև այն խնդիրներում, որտեղ հետազոտվող առարկաների հատկությունները հնարավոր չէ նկարագրել պատահական մեծությունների օգնությամբ, սակայն այդ առարկաները հնարավոր է կարգավորել, դասավորելով դրանք տարածության կամ ժամանակի մեջ որոշակի հերթականությամբ, ըստ հետազոտվող հատկության դրսևորման աստիճանի: Այստեղ նույնպես առարկաների հատկանիշի հաջորդականությունը փոխարինվում է դրանց տարակարգերի հաջորդականությամբ:

Երբեմն օգտագործում են նաև կոտորակային տարակարգեր: Այն դեպքերում, երբ երկու կամ ավելի առարկաներ ունեն հետազոտվող հատկանիշի միևնույն չափը, ընդունված է այդ առարկաներին վերագրել միևնույն տարակարգը, որը հավասար է դրանց զբաղեցրած տեղերի համարների միջին թվաբանականին: Այս ձևով ստացված տարակարգերը կոչվում են **միավորված** կամ **կապված տարակարգեր**:

Այս ենթաբաժնում մենք կնշանակենք $X_1, X_2 ..., X_N$ և $Y_1, Y_2 ..., Y_N$ երկու հետազոտվող հատկություններն արտահայտող (պատահական) տարակարգային հաջորդականությունները: Պահանջվում է այդ հաջորդականությունների հիման վրա ստուգել հետազոտվող հատկությունների միջև հարաբերակցության բացակայության մասին վարկածը:

Խնդրի էությունը պարզաբանելու համար ենթադրենք, որ միավորված տարակարգեր չկան: $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$ հաջորդականությունն արտագրենք այնպիսի հերթականությամբ, որպեսզի $x_1, x_2, ..., x_N$ թվերը դասավորվեն աճման կարգով: Կստանանք

$$(1, y_{n1}), (2, y_{n2}, ..., (N, y_{nN}) :$$

Հասկանալի է, որ եթե դիտարկվող հատկությունները լինեն անկախ, ապա y-ների բոլոր հնարավոր հաջորդականությունները կունենան միևնույն հավանականությունը՝ $1/N!$: Առաջադրվել են տարակարգային հարաբերակցական կապի բացակայության վարկածներ ստուգելու մի քանի վիճականիներ, որոնք օգտագործում են նշված հատկությունը: Դրանցից առավել տարածված ու հայտնի վիճականին է Սպիրմենի տարակարգային հարաբերակցության գործակիցը:

Սպիրմենի տարակարգային հարաբերակցության գործակիցն առաջադրվել է 1904 թ.՝ հոգեբանության բնագավառի որոշ խնդիրներ լուծելու նպատակով: Այն մինչև օրս շարունակում է հաջողությամբ կիրառվել՝ երկու փոփոխականների միջև հավանականային կապի աստիճանը որոշելու համար:

Սպիրմենի հարաբերակցության գործակիցը թվային արտահայտությամբ հավասար է սովորական նմուշային հարաբերակցության գործակցին, եթե $X_1, X_2..., X_N$ և $Y_1, Y_2..., Y_N$ պատահական մեծությունները տարակարգեր են:

Ընդհանուր դեպքում, երբ տարակարգային հաջորդականությունները պարունակում են կապված տարակարգեր, Սպիրմենի հարաբերակցության գործակիցը հաշվում են հետևյալ բանաձևով՝

$$\hat{r}_S(\mathbf{X}, \mathbf{Y}) = \frac{(1/6)(N^3 - N) - \sum_n (X_n - Y_n)^2 - T_X - T_Y}{\sqrt{[(1/6)(N^3 - N) - 2T_X][(1/6)(N^3 - N) - 2T_Y]}}, \tag{6}$$

որտեղ

$$T_X = \frac{1}{12}\sum_{t=1}^{M_X}[(N_t^X)^3 - N_t^X], \quad T_Y = \frac{1}{12}\sum_{t=1}^{M_Y}[(N_t^Y)^3 - N_t^Y],$$

և, համապատասխանաբար, M_X-ը և M_Y-ը, X և Y փոփոխականների միավորված տարակարգերի խմբերի քանակներն են, N_t^X-ը և N_t^Y-ը՝ t-րդ խմբի մեջ մտնող միավորված տարակարգերի քանակները:

Եթե միավորված տարակարգեր չկան, ապա $T_X = T_Y = 0$, և

$$\hat{r}_S(\mathbf{X}, \mathbf{Y}) = 1 - \frac{6}{N^3 - N}\sum_n (X_n - Y_n)^2 :$$

Դժվար չէ ցույց տալ, որ այս դեպքում $\hat{r}_S(\mathbf{X}, \mathbf{Y}) = \hat{\rho}_{YX}(\mathbf{X}, \mathbf{Y})$, որտեղ $\hat{\rho}(\mathbf{X}, \mathbf{Y})$-ը որոշվում է (3) բանաձևով, եթե այնտեղ տեղադրվում են տարակարգերը: Այս փաստը որոշ չափով պարզեցնում է ինչպես Սպիրմենի տարակարգային հարաբերակցության գործակցի մեկնաբանությունը՝ որպես հավանականային կապի չափի, այնպես էլ հաշվարկները, երբ օգտագործվում են սովորական հարաբերակցության գործակիցը հաշվող հաշվարային ծրագրերը: Տարբերություն կա միայն վարկածների ստուգման խնդրում:

Եթե երկու տարակարգային հաջորդականությունները համընկնում են՝ $x_n = y_n$, $n = \overline{1, N}$, ապա $\hat{r}_S = 1$: Հակառակը, երբ տարակարգերը հակադիր են՝ $x_n = N - y_n + 1$, ապա $\hat{r}_S = -1$: Մնացած դեպքերում $|\hat{r}_S| < 1$:

Տարակարգային հարաբերակցության գործակցի վերաբերյալ $H_0 : r_S(\mathbf{X}, \mathbf{Y}) = 0$ վարկածն ստուգելու համար օգտվում են նույն վիճականիից, որն օգտագործվում է սովորական հարաբերակցության գործակցի համար (տե՛ս (4)-ը)՝

$$|\hat{t}(\mathbf{X},\mathbf{Y})| = |\hat{r}_S(\mathbf{X},\mathbf{Y})|\sqrt{N-2}/\sqrt{1-\hat{r}_S^2(\mathbf{X},\mathbf{Y})} \sim \mathbf{t}(N-2) :$$

Հետևաբար, եթե $|\hat{t}(\mathbf{x},\mathbf{y})| > t_{1-\alpha/2}(N-2)$, որտեղ $t_{1-\alpha/2}(N-2)$-ը Ստյուդենտի $\mathbf{t}(N-2)$ բաշխման $(1-\alpha/2)$-քանորդիչն է, ապա H_0 վարկածը ժխտվում է:

Կա ավելի պարզ վիճականի՝ $\hat{z}(\mathbf{X},\mathbf{Y}) = \hat{r}_S(\mathbf{X},\mathbf{Y})\sqrt{N-1}$, որի բաշխումը N-ի մեծ արժեքների դեպքում շատ մոտ է $\mathcal{N}(0,1)$ նորմալ բաշխմանը: Եթե $|\hat{z}(\mathbf{x},\mathbf{y})| > z_{1-\alpha/2}$, որտեղ $z_{1-\alpha/2}$-ը $\mathcal{N}(0,1)$ բաշխման $(1-\alpha/2)$-քանորդիչն է, ապա H_0-ն ժխտվում է:

Օրինակ 11: Գրասենյակի ղեկավարի թափուր պաշտոնը զբաղեցնելու համար ներկայացված է 7 հայտ: Երկու պատասխանատու աշխատողի հանձնարարել են իրարից անկախ ստուգել բոլոր թեկնածուների կարողությունները՝ այդ պաշտոնում աշխատելու համար, և դասավորել թեկնածուներին ըստ կարողությունների դրսևորման աստիճանի՝ 1 տարակարգը վերագրելով լավագույն թեկնածուին: Ստացված տարակարգերը բերված են աղյուսակում՝

n	1	2	3	4	5	6	7
X_n	7	2	6	5	4	1	3
Y_n	7	3	5	6	4	2	1

:

Պահանջվում է որոշել երկու ստուգողների գնահատականների համընկնելու աստիճանը, հենվելով Սպիրմենի տարակարգային հարաբերակցության գործակցի արժեքի վրա: Եթե այն նշանակալի է (այսինքն՝ արժեքը բավականաչափ մեծ է), ապա կարելի է համարել, որ թեկնածուների կարողությունները ճիշտ են գնահատվել, և, միավորելով ստացված գնահատականները, կարելի է ընտրել լավագույն թեկնածուին:

Լուծում: Քանի որ $\sum\limits_n (x_n - y_n)^2 = 8$, ապա (6) բանաձևից ստանում ենք հետևյալ արդյունքը՝

$$\hat{r}_S(\mathbf{x},\mathbf{y}) = 1-(6)(8)/(7^3-7) \approx 0.857 :$$

Ստուգենք դիտարկվող փոփոխականների անկախ լինելու H_0 վարկածը: $\hat{z}(\mathbf{X},\mathbf{Y})$ վիճականու համար ստանում ենք $\hat{z}(\mathbf{x},\mathbf{y}) = 0.857\sqrt{6} \approx 2.10$, իսկ $\alpha = 0.05$ համար $z_{0.975} = 1.96$, հետևաբար H_0 վարկածը ժխտվում է: Եթե կիրառենք $\hat{t}(\mathbf{X},\mathbf{Y})$-ը, ապա կստանանք $\hat{t}(\mathbf{x},\mathbf{y}) \approx 3.72$, մինչդեռ $t_{0.975}(5) = 2.57$, հետևաբար, ինչպես և նախորդ դեպքում, H_0 վարկածը ժխտվում է: Նկատենք, որ այս արդյունքի հիման վրա ներկայացված թեկնածուներից ընտրվում է վեցերորդը:

Օրինակ 12: Աղյուսակում բերված են ԱՄՆ մեկ դոլարի և 1000 ՌԴ ռուբլու փոխար-

n	ԱՄՆ դոլար	Տարա-կարգը	ՌԴ ռուբլի	Տարա-կարգը	n	ԱՄՆ դոլար	Տարա-կարգը	ՌԴ ռուբլի	Տարա-կարգը
	I	II	III	IV		I	II	III	IV
1	450.8	1	79.80	1	11	512.7	20	87.55	20
2	462.1	2	81.25	2	12	508.2	19	86.70	19
3	468.2	3	82.10	4	13	504.9	17	85.90	17
4	468.9	4	81.80	3	14	502.7	16	85.45	16
5	484.1	7	84.05	9	15	502.6	15	85.30	15
6	478.2	5	82.60	5	16	501.1	13.5	84.90	12.5
7	481.1	6	82.65	6	17	501.1	13.5	84.90	12.5
8	486.1	8	83.30	7	18	501.0	11.5	84.90	12.5
9	489.7	9	83.75	8	19	501.0	11.5	84.90	12.5
10	505.8	18	86.40	18	20	499.9	10	84.55	10

ժեքները դրամի նկատմամբ 1997 թ. հունվարի 15-ից նոյեմբերի 1-ը ներառյալ (ընդամենը 20 տվյալ): Փոփոխականների նշանակումները չբարդացնելու համար համարակալված են աղյուսակի սյուները, ընդ որում I սյունակում բերված են ամերիկյան մեկ դոլարի փոփոխականների մեծությունները (դրամով), II սյունակում՝ վերջինիս տարակարգերը, որտեղ 1 տարակարգը համապատասխանում է դոլարի փոխարժեքի նվազագույն արժեքին, իսկ 20 տարակարգը՝ առավելագույն արժեքին: III և IV սյունակներում բերված են համապատասխան մեծությունները 1000 ռուսական ռուբլու համար:

Որոշենք աղյուսակում ներկայացված թվային շարքի տարակարգային հարաբերակցության գործակիցը և ստուգենք դրա նշանակալիությունը $\alpha = 0.01$ համար:

Լուծում: Աղյուսակից տեսնում ենք, որ դոլարի տարակարգային հաջորդականությունը պարունակում է միավորված (կապված) տարակարգերի երկու, իսկ ռուսական ռուբլունը՝ մեկ խումբ: Այսինքն,
$M_X = 2, \quad M_Y = 1, \quad T_X = (2^3 - 2 + 2^3 - 2)/12 = 1, \quad T_Y = (4^3 - 4)/12 = 5$:

Տեղադրելով այս տվյալները (6) բանաձևի մեջ, կստանանք $\hat{r}_S(\mathbf{x}, \mathbf{y}) \approx 0.991$: Նշենք, որ եթե հաշվարկը կատարենք, տեղադրելով $T_X = T_Y = 0$, կստանանք նույն թիվը՝ 0.991 (երրորդ նիշի ճշտությամբ):

Ստուգենք H_0 վարկածը, կիրառելով $\hat{z}(\mathbf{X}, \mathbf{Y})$ վիճականին: Կատարելով հաշվարկները, կստանանք $\hat{z}(\mathbf{x}, \mathbf{y}) \approx 4.32$, մինչդեռ $z_{0.995} = 2.58$: Հետևաբար, H_0 վարկածը չի ընդունվում, այսինքն՝ ամերիկյան դոլարի և ռուսական ռուբլու՝ դրամով արտահայտված փոխարժեքները հարաբերակցված են: Սպիրմենի հարաբերակցության գործակիցը նշանակալի է:

Որոշ խնդիրներում անհրաժեշտ է լինում ստուգել համանականային կախումների առկայությունը և «չափել» դրանց աստիճանը միաժամանակ մի քանի տարակարգային փոփոխականների միջև: Այս նպատակով Կենդալը առաջարկել է Սպիրմենի տարակարգային հարաբերակցության գործակցն ընդհանրացնող մի վիճականի՝ $\hat{W}(M)$, որը կոչվում է **համաձայնեցվածության (կամ կոնկորդացիայի) գործակից** և որոշվում է հետևյալ բանաձևով՝

$$\hat{W}(M) = \frac{12}{M^2(N^3 - N)} \sum_{n=1}^{N} \sum_{m=1}^{M} [X_n^m - M(N+1)/2]^2, \tag{7}$$

որտեղ M-ը դիտարկվող տարակարգային փոփոխականների քանակն է, N-ը՝ նմուշի ծավալը, X_n^m-ը՝ m-րդ փոփոխականի n-րդ տարակարգը:

Համաձայնեցվածության $\hat{W}(M)$ գործակիցն օժտված է հետևյալ հատկություններով.

ա. $0 \leq \hat{W}(M) \leq 1$:

բ. $\hat{W}(M) = 1$ այն և միայն այն դեպքում, երբ բոլոր M տարակարգային հաջորդականությունները համընկնում են:

գ. Դիցուք՝ $\hat{r}(M)$-ն M փոփոխականների բոլոր հնարավոր $M(M-1)/2$ զույգերի միջև հաշվարկված՝ Սպիրմենի տարակարգային հարաբերակցության գործակիցների միջին թվաբանական արժեքն է, այդ դեպքում՝

$$\hat{r}(M) = [M\hat{W}(M) - 1]/(M - 1) :$$

$M = 2$ դեպքում ստանում ենք $\hat{W}(2) = [\hat{r}(2) + 1]/2$, մյուս կողմից՝ քանի որ $\hat{r}(2) = \hat{r}_S(\mathbf{x}, \mathbf{y})$, երկու փոփոխականների համաձայնեցվածության գործակիցը գծային կախման մեջ է այդ փոփոխականների համար հաշվարկված՝ Սպիրմենի տարակարգային հարաբերակցության գործակցից:

Այն փաստը, որ $\hat{W}(M)$-ի համար բացասական արժեքներ նախատեսված չեն, բացատրվում է հետևյալ հանգամանքով: Ի տարբերություն զույգային կապերի, $M \geq 3$ տարակարգային փոփոխականի դեպքում *համաձայնեցվածության* և *ոչ համաձայնեցվածության* հականիշ գաղափարները կորցնում են նախկին սիմետրիան (զրոյի նկատմամբ): Դիտարկվող տարակարգային հաջորդականությունները կարող են լրիվ համընկնել, բայց չեն կարող լրիվ չհամընկնել այն իմաստով, որը տեղի ունի երկու փոփոխականի դեպքում:

Ավելացնենք նաև, որ (7) բանաձևը վերաբերում է միավորված տարակարգերի բացակայության դեպքին: Եթե կան միավորված տարակարգեր, ապա օգտվում են հետևյալ բանաձևից՝

$$\hat{W}(M) = \frac{\sum_n \sum_m [X_n^m - M(N+1)/2]^2}{(1/12)M^2(N^3 - N) - M\sum_m T_{X^m}} :$$

Դիտարկենք այն դեպքը, երբ համաձայնեցվածության գործակցի ենթադրյալ տեսական արժեքը համապատասխանում է դիտարկվող փոփոխականների միջև տարակարգային որևէ կապի բացակայության դեպքին: M-ի և N-ի փոքր արժեքների համար ($2 \leq M \leq 20$, $3 \leq N \leq 7$) ստացված են հատուկ աղյուսակներ, որոնք հնարավորություն են տալիս ստուգելու դիտարկվող փոփոխականների միջև տարակարգային հարաբերակցային կապի (կամ համաձայնեցվածության) բացակայության մասին H_0 վարկածը: $N > 7$ դեպքում կարելի է օգտվել այն հանգամանքից, որ

$$\hat{\chi}^2(\mathbf{X}) = M(N-1)\hat{W}(M) \sim \chi^2(N-1) :$$

Եթե

$$\hat{\chi}^2(\mathbf{x}) > \chi^2_{1-\alpha}(N-1),$$

ապա փոփոխականների միջև տարակարգային կապի բացակայության մասին վարկածը ժխտվում է:

Օրինակ 13: «Ֆայնենշլ թայմս» թերթը հրապարակում է աշխարհի 500 առաջատար

	Ընկերության անվանումը	Շուկ. կապ.	Շրջ. կապ.	Մաքուր եկամ.	ROCE	Օգտ. կապ.	Ծառ. քանակը
n		X_1	X_2	X_3	X_4	X_5	X_6
1	General Electric	222.7	78.5	7.3	8.7	83.4	239.0
2	Microsoft Corp	159.7	11.4	3.5	31.9	10.8	22.2
3	Exxon Corp	158.0	116.7	7.5	14.	52.7	79.0
4	Coca - Cola	151.3	18.5	3.5	48.	7.3	26.0
5	Intel Corp	150.8	20.8	5.2	28.9	17.9	48.5
6	Merek	120.8	19.8	3.9	25.1	15.4	49.1
7	IBM Corp	104.1	75.9	5.4	17.2	31.5	240.6
8	Philip Morris	100.7	54.5	6.3	23.5	26.8	154.0
9	Procter & Gamble	93.3	35.8	3.4	20.5	16.2	106.0
10	Wal-Mart Stores	82.5	104.9	3.1	10.8	28.2	728.0
11	Bristol Myerss Squibb	82.5	15.1	2.9	37.8	7.5	51.2
12	Johnson & Johnson	76.9	21.6	2.9	23.6	12.2	89.3
13	Plizer Inc	75.5	11.3	1.9	25.1	7.7	47.0
14	Hewlett-Packard Co	72.3	38.4	2.6	16.1	16.0	112.0
15	AT & T Corp	71.9	52.2	5.6	19.9	28.2	130.4
16	Du Pont De Nemours	69.6	38.5	3.6	22.1	16.4	97.0
17	Lilly (Eli)	67.3	7.4	1.5	17.6	8.6	29.2
18	Bell Atlantic Corp	62.5	13.1	1.7	12.9	13.5	29.2
19	Pepsico	61.5	31.6	1.2	7.6	15.1	486.0
20	Mobil Corp	58.2	71.1	3.0	12.3	23.6	43.0
21	Compaq Computer Corp	56.6	18.1	1.3	20.4	6.4	18.9
22	SBC Communications	56.1	13.9	2.1	17.0	12.3	61.5
23	Berkshire Hathaway	55.2	10.5	2.5	9.7	25.7	34.5
24	Boeing	54.4	22.7	1.1	7.3	14.9	143.0
25	Chevron Corp	54.4	37.6	2.6	13.3	19.6	40.8

ընկերությունների տնտեսական ցուցանիշները, որոնք վերջին տարիներին վերահրատարակում է ռուսական «Իզվեստիա» թերթը: Աղյուսակում բերված են ԱՄՆ-ի 25 խոշորագույն ընկերությունների 1997 թ. հիմնական տնտեսական ցուցանիշները, որոնք վերցված են նշված թերթի 1998 թ. փետրվարի 17-ի համարից: Ընկերությունները դասավորված են ըստ շուկայական կապիտալի մեծության (այսինքն՝ այն արժեքի, որն ստացվում է թողարկված արժեթղթերի քանակի և դրանց շուկայական գնի արտադրյալով):

Աղյուսակում բերված են հետևյալ ցուցանիշները.
շուկայական կապիտալը՝ X_1, մլրդ դոլար, դրամաշրջանառությունը՝ X_2, մլրդ դոլար, մաքուր եկամուտը՝ X_3, մլրդ դոլար, շահույթի ցուցանիշը ներդրված կապիտալի համար (ROCE)՝ X_4, օգտագործվող կապիտալը՝ X_5, մլրդ դոլար, ծառայողների քանակը՝ X_6, հազ. մարդ:

Կարելի՞ է արդյոք ընդունել այն տեսակետը, որ նշված ցուցանիշների արժեքները «համաձայնեցված են», այսինքն, դրանցից որևէ մեկի մեծացման (փոքրացման) դեպքում մնացած հինգ ցուցանիշները նույնպես մեծանալու (փոքրանալու) միտում են դրսևորում: Այս հարցին պատասխանելու համար հաշվենք $\hat{W}(6)$ համաձայնեցվածության գործակիցը $N = 25$ համար, այնուհետև ստուգենք H_0 վարկածը:

Լուծում: Հաջորդ աղյուսակում բերված են նշված վեց ցուցանիշների տարակարգերը, ընդ որում ցուցանիշի նվազագույն արժեքին վերագրված է 1 տարակարգը, իսկ առավելագույնին՝ 25: Այստեղ միացված տարակարգեր չկան, որովհետև աղյուսակում բերված թվերը կլորացված են բնագրի համեմատ, որտեղ համընկնող արժեքներ չկան:

n	X_1	X_2	X_3	X_4	X_5	X_6	$\sum_p x_n^p$	$(\sum_p x_n^p - 78)^2$
1	25	23	24	3	25	22	122	1936
2	24	4	16	23	6	2	75	9
3	23	25	25	9	24	14	120	1764
4	22	9	17	25	2	4	79	1
5	21	11	20	22	16	9	99	441
6	20	10	19	21	12	10	92	169
7	19	22	21	12	23	23	120	1764
8	18	20	23	18	20	21	114	1296
9	17	15	15	16	14	16	93	225
10	16	24	14	5	22	25	106	784
11	15	7	11	24	3	10	71	49
12	14	12	12	19	7	15	79	1
13	13	3	6	20	4	8	54	576
14	12	17	9	10	13	18	79	1
15	11	19	22	14	21	19	106	784
16	10	18	18	17	15	17	95	289
17	9	1	4	13	5	3	35	1849
18	8	5	5	7	9	13	47	961
19	7	14	2	2	11	24	60	324
20	6	21	13	6	18	7	71	49
21	5	8	3	15	1	1	33	2025
22	4	6	7	11	8	12	48	900
23	3	2	8	4	19	5	41	1369
24	2	13	1	1	10	20	47	961
25	1	16	10	8	17	6	58	400
Ընդամենը								18927

Քանի որ $N = 25, M = 6$, կստանանք $M(N+1)/2 = 78$, և

$$\hat{W}(6) = \frac{12}{6^2(25^3 - 25)} 18927 \approx 0.4044 :$$

Այնուհետև ստանում ենք՝

$$\hat{\chi}^2(\mathbf{x}) = M(N-1)\hat{W}(6) = 6(25-1)0.4044 \approx 58.2,$$

մինչդեռ $\chi^2_{0.95}(24) = 35.2$: Հետևաբար դիտարկվող վեց փոփոխականների անկախության մասին վարկածը չի ընդունվում:

Օրինակ 14: Նախորդ օրինակի արդյունքը կարող էր ստացվել այն պատճառով, որ բոլոր փոփոխականները կախված են շուկայական կապիտալի մեծությունից (X_1): Լուծենք նույն խնդիրը վերջին հինգ փոփոխականների համար, առանց հաշվի առնելու շուկայական կապիտալի մեծությունը:

Լուծում: Կատարելով նույն հաշվարկը նշված հինգ փոփոխականների համար, կստանանք հետևյալ արդյունքը՝

$$M(N+1)/2 = 65, \quad \hat{W}(5) \approx 0.395, \quad \hat{\chi}^2(\mathbf{x}) = M(N-1)\hat{W}(5) \approx 47.4:$$

Քանի որ $\chi^2_{0.95}(24) = 35.2$, դիտարկվող հինգ փոփոխականների անկախության մասին վարկածը ժխտվում է: Նկատենք, որ X_1 փոփոխականի արտաքսման հետևանքով $\hat{\chi}^2$ մեծությունն էական փոփոխություն չկրեց:

11.6. Հարաբերակցային քանորդ

Y պատահական մեծության ցրվածքի արտահայտության մեջ կատարելով որոշ տարրական ձևափոխություններ և, հաշվի առնելով պայմանական ցրվածքի հատկությունները կստանանք՝

$$\mathbf{D}(Y) = \sigma_Y^2 = \mathbf{E}[Y - \mathbf{E}(Y)]^2 = \mathbf{E}[Y - \mathbf{E}(Y|X) + \mathbf{E}(Y|X) - \mathbf{E}(Y)]^2 = \mathbf{E}[Y - \mathbf{E}(Y|X)]^2 +$$

$$+\mathbf{E}[\mathbf{E}(Y|X) - \mathbf{E}(Y)]^2 = \mathbf{E}[Y - \mathbf{E}(Y|X)]^2 + \mathbf{D}[\mathbf{E}(Y|X)] = \sigma_{YX}^2 + \mathbf{D}[\mathbf{E}(Y|X)] :$$

Ինչպես տեսնում ենք, վերջին արտահայտության առաջին գումարելին Y-ի պայմանական ցրվածքն է ըստ $X = x$ պայմանի և մեկնաբանվում է որպես Y-ի **ցրվածք ռեգրեսիայի ֆունկցիայի նկատմամբ**: Իսկ երկրորդ գումարելին մեկնաբանվում է որպես **ռեգրեսիայի ֆունկցիայի ցրվածք**:

Y-ի **հարաբերակցային քանորդ** X-ի նկատմամբ կոչվում է հետևյալ հարաբերությունը՝

$$r_{YX} = \sqrt{\mathbf{D}[\mathbf{E}(Y|X=x)]}/\sigma_Y = \sqrt{1 - \sigma_{YX}^2/\sigma_Y^2} : \qquad (8)$$

Հարաբերակցային քանորդն արտահայտում է X-ից Y-ի կախման չափը և օգտագործվում է որպես հավանականային կախումների հայտանիշ: Այդ կապակցությամբ հարաբերակցային քանորդի քառակուսին կոչվում է **որոշակիության գլխավոր գործակից**:

Հարաբերակցային քանորդի հատկությունները մասամբ նման են հարաբերակցության գործակցի հատկություններին:

1. $0 \le r_{YX}^2 \le 1$, ընդ որում $r_{YX}^2 = 1$, միայն եթե $\mathbf{E}[Y - \mathbf{E}(Y|X=x)]^2 = 0$, այսինքն՝ երբ $\mathbf{P}\{Y = \mathbf{E}(Y|X=x)\} = 1$: Սա նշանակում է, որ Y-ի դիտման արդյունքները գտնվում են ռեգրեսիայի կորի վրա, այսինքն՝ Y-ը ֆունկցիոնալ կախում ունի X-ից (նկատենք, որ այս հատկությունը կախված չէ ռեգրեսիայի ֆունկցիայի տեսքից):

2. ի տարբերություն հարաբերակցության գործակցի, հարաբերակցային քանորդը հաաչափ է փոփոխականների նկատմամբ՝ $r_{YX}^2 \neq r_{XY}^2$:

3. Եթե Y-ի ռեգրեսիան X-ի վրա գծային է, ապա $r^2_{YX} = \rho^2_{YX}$, մնացած դեպքերում $r^2_{YX} \geq \rho^2_{YX}$: Ընդհանրապես, $r^2_{YX} - \rho^2_{YX}$ տարբերությունը ծառայում է որպես ոչ գծային կախման ցուցիչ: Եթե $r^2_{YX} = \rho^2_{YX}$, ապա պետք է եզրակացնել, որ ռեգրեսիայի ֆունկցիան գծային է, իսկ եթե $r^2_{YX} > \rho^2_{YX}$, ապա ռեգրեսիայի ֆունկցիան գծային չէ:

Հետևյալ օրինակն ընդգծում է այդ հանգամանքը և ցուցադրում հարաբերակցությյան գործակցի ու հարաբերակցային քանորդի տարբերությունը:

Դիցուք, $Y = X^2$, իսկ X-ը ընդունում է $-1, 0, +1$ արժեքները, յուրաքանչյուրը՝ 1/3 հավանականությամբ: Այս դեպքում $\rho = 0$, $r^2_{YX} = 1$, $r_{XY} = 0$: Օրինակը ցույց է տալիս, որ ֆունկցիոնալ կախման դեպքում հարաբերակցային քանորդը հավասար է 1-ի, մինչդեռ հարաբերակցության գործակիցը կարող է հավասարվել 0-ի:

4. Եթե $\mathrm{E}(Y|X = x)$-ը x-ից կախված չէ, ապա $r_{YX} = 0$: Սա տեղի ունի, մասնավորապես, այն դեպքում, երբ X և Y պատահական մեծությունները անկախ են:

Նմուշային հարաբերակցային քանորդը որոշելու համար անհրաժեշտ է (8)-ում տեղադրել σ^2_{YX} և σ^2_Y ցրվածքների նմուշային գնահատականները:

Խմբավորենք երկչափ նմուշը հետևյալ եղանակով: Բաժանենք X-ի արժեքների տիրույթը K միջակայքերի այնպես, որ k-րդ միջակայքին համապատասխանի $y_1, y_2, \ldots, y_N$ հաջորդականության մի քանի՝ N_k տարր: Նշանակենք y_{kn}-ով Y-ի այն արժեքները, որոնց x_n-երը ընկնում են k-րդ միջակայքը, $k = \overline{1, K}$, $n = \overline{1, N_k}$, և դիցուք՝ $\mathbf{\bar{y}}_k = \sum\limits_{n=1}^{N_k} y_{kn}/N_k$: Այդ դեպքում կարող ենք գրել

$$\mathbf{\bar{y}} = \frac{1}{N}\sum_k N_k \mathbf{\bar{y}}_k, \; N = \sum_k N_k, \; \hat{\sigma}^2_y = \frac{1}{N}\sum_k \sum_{n=1}^{N_k} (y_{kn} - \mathbf{\bar{y}})^2, \; \hat{\sigma}^2_{YX} = \sum_k N_k(\mathbf{\bar{y}}_k - \mathbf{\bar{y}})^2 :$$

Տեղադրելով բերված նմուշային գնահատականները (8)-ում, կստանանք նմուշային հարաբերակցային քանորդը՝

$$\hat{r}^2_{YX}(\mathbf{x}, \mathbf{y}) = \frac{\sum\limits_k N_k(\mathbf{\bar{y}}_k - \mathbf{\bar{y}})^2}{\sum\limits_k \sum\limits_{n=1}^{N_k} (y_{kn} - \mathbf{\bar{y}})^2} = \frac{\sum\limits_k N_k \mathbf{\bar{y}}^2_k - N\mathbf{\bar{y}}^2}{\sum\limits_k \sum\limits_{n=1}^{N_k} y^2_{kn} - N\mathbf{\bar{y}}^2} : \tag{9}$$

Հարաբերակցային քանորդի մասին զրոյական վարկածն է՝ $H_0 : \; r_{YX} = 0$: Եթե այն ճիշտ է, ապա

$$\hat{F}(\mathbf{X}, \mathbf{Y}) = \frac{(K-1)\hat{r}^2_{YX}}{(N-K)(1 - \hat{r}^2_{YX})} \sim \mathcal{F}(K-1, N-K),$$

ուստի եթե $\hat{F}(\mathbf{x}, \mathbf{y}) > F_{1-\alpha}(K-1, N-K)$, որտեղ $F_{1-\alpha}(K-1, N-K)$-ն $\mathcal{F}(K-1, N-K)$ բաշխման $(1-\alpha)$-քանորդիչն է, ապա H_0 վարկածը ժխտվում է:

<u>Օրինակ</u> 15: Ագարակատերը հետաքրքրվում է պարարտանյութի օգնությամբ բերքատվության բարձրացման հարցով: Նա ընտրում է նույն որակի և չափի 10 հողակտոր և աճեցնում միևնույն տեսակի և քանակի լոլիկի սածիլներ, օգտագործելով, սակայն, տարբեր քանակությամբ պարարտանյութ: Պարարտանյութի քանակությունը (X) և դրան համապատասխանող բերքի ցուցանիշը (Y, պայմանական միավորներով) տրված են աղյուսակում՝

X	0	0	10	10	20	20	30	30	40	40
Y	2.7	3.6	5.0	6.4	8.2	10.4	11.3	12.7	13.6	15.4

:

Հաշվելով $\hat{r}_{YX}(\mathbf{X},\mathbf{Y})$ ընտրանքային հարաբերակցային քանորդը, ստուգել պարարտանյութի քանակից լոլիկի բերքատվության կախվածության նշանակալիությունը $\alpha = 0.05$ համար:

Լուծում: Ընդունենք $K = 5$ և խմբավորենք տվյալներն ըստ պարարտանյութի քանակի: Բոլոր խմբերի համար ստանում ենք $N_k = 2$, $k = \overline{1,5}$: Համապատասխան միջանկյալ արդյունքները բերված են աղյուսակում:

Խմբի համարը՝ k	1	2	3	4	5	
$\overline{\mathbf{x}}_k$	0	10	20	30	40	Գումարը
$\sum_{n=1}^{N_k} y_{nk}$	6.3	11.4	18.6	24	29	89.3
$\overline{\mathbf{y}}_k$	3.15	5.7	9.3	12	14.5	
$\overline{\mathbf{y}}_k^2$	9.9225	32.49	86.49	144	210.25	966.305
$\sum_{n=1}^{N_k} y_{nk}^2$	20.25	65.96	175.4	288.98	422.12	972.710

:

Աղյուսակից ստանում ենք, որ $\overline{\mathbf{y}} = 8.93$: Հետևաբար, ըստ (9)-ի կարող ենք գրել

$$\hat{r}_{YX}^2(\mathbf{x},\mathbf{y}) = \frac{966.305 - (10)(8.93)^2}{972.710 - (10)(8.93)^2} = \frac{168.856}{175.261} \approx 0.9635, \quad \hat{r}_{YX}(\mathbf{x},\mathbf{y}) \approx 0.982 :$$

Հարաբերակցային քանորդի նշանակալիությունը ստուգելու համար հաշվենք համապատասխան F-հարաբերությունը: Կստանանք

$$\hat{F}(\mathbf{x},\mathbf{y}) = \frac{(5-1)(0.9635)}{(10-5)(1-0.9635)} \approx 21.12,$$

մինչդեռ, $F_{0.95}(4,5) = 5.19$: H_0 վարկածը ժխտվում է, այսինքն՝ բերքի ցուցանիշը կախված է պարարտանյութի քանակից:

Նկատենք, որ այս օրինակում ընտրանքային հարաբերակցության գործակիցը՝ $\hat{\rho}_{XY}(\mathbf{x},\mathbf{y}) \approx 0.980$, այսինքն՝ այն մոտավորապես հավասար է ընտրանքային հարաբերակցային քանորդին: Ըստ հարաբերակցային քանորդի 3-րդ հատկության, կարող ենք ընդունել, որ այս օրինակում ռեգրեսիայի ֆունկցիան գծային է:

Գլուխ 12
Կորագիծ և բազմաչափ գծային ռեգրեսիա

Բավական չէ փորձի արդյունքները գրանցել, դրանք պետք է կշռադատել ու համեմատել, քննել ու զտել:

Միշել Մոնտեն

12.1. Հիմնական գաղափարներ

Նախորդ գլխում մենք մանրամասնորեն վերլուծեցինք ռեգրեսիայի պարզագույն խնդիրը, որն ուսումնասիրում է Y պատահական մեծության հավանականային կախվածությունը միաչափ X փոփոխականից՝ արտահայտված $\mathrm{E}(Y|X=x)=g(x,a)=a_0+a_1x$ գծային ռեգրեսիայի ֆունկցիայով: Սակայն կան բազմաթիվ խնդիրներ, որոնց լուծման համար միաչափ գծային մոդելը պիտանի չէ, օրինակ, այն դեպքերում, երբ կախումն արտահայտվում է ցուցչային կամ եռանկյունաչափական ֆունկցիաների կամ դրանց գումարի միջոցով, ինչպես նաև այն դեպքերում, երբ Y-ը հավանականային կախման մեջ է մի քանի փոփոխականներից:

Այս գլխում մենք կուսումնասիրենք **ոչ գծային** և **բազմաչափ ռեգրեսիայի մոդելներ,** որոնց արտահայտության մեջ բացահայտ ձևով մասնակցում են M-չափանի x վեկտորը և Q-չափանի a պարամետրը: $g(x,a)$ ռեգրեսիայի ֆունկցիայի արգումենտներն են

$$x=(x_1,x_2,...,x_M),\ a=(a_0,a_1,...,a_Q),$$

իսկ ռեգրեսիայի մոդելն է՝

$$Y=g(x,a)+Z,$$

որտեղ X-ը անկախ, իսկ Y-ը՝ կախյալ փոփոխականներն են, $Z\sim\mathcal{N}(0,\sigma^2)$:

Երբեմն $g(x,a)$ ֆունկցիայի բացահայտ տեսքը նախօրոք հայտնի է, այդ դեպքում ռեգրեսիայի ֆունկցիայի որոշման խնդիրը հանգում է a պարամետրի գնահատմանը:

> $g(x,a)$ **ռեգրեսիայի ֆունկցիան** կոչվում է **գծային**, եթե
>
> $$g(x,a)=a_0+\sum_{m=1}^{M}a_mx_m, \qquad (1)$$
>
> $a_0,a_1,...,a_M$ պարամետրերը կոչվում են ռեգրեսիայի գործակիցներ: Եթե $g(x,a)$ կախումը x-ից գծային չէ, **ռեգրեսիայի ֆունկցիան** կոչվում է **կորագիծ**:

Պարզության համար մենք կանդրադառնանք միայն այն կարևոր դեպքին, երբ $g(x,a)$ ֆունկցիան ըստ x փոփոխականի գծային չէ, իսկ ըստ a վեկտորի գծային է, այսինքն ունի հետևյալ տեսքը՝

$$g(x,a)=a_0+\sum_{m}a_mg_m(x): \qquad (2)$$

Այս տեսակի ֆունկցիաներ առաջանում են, մասնավորապես, երբ $\mathrm{E}(Y|X=x)$ պայմանական սպասելիի բացահայտ տեսքը հայտնի չէ, սակայն տվյալ խնդրի համար ընդունելի ճշտությամբ այն կարելի է մոտարկել (2) տեսքի ֆունկցիայով: Նշանակելով $x'_m=g_m(x)$՝ (2)-ը կգրվի (1) տեսքով, այդ պատճառով (2) տեսքի ռեգրեսիայի ֆունկցիան կոչվում է **գծայնացված**:

Հաճախ ռեգրեսիայի մոդելների տեսակներն անվանում են ըստ $g(x,a)$ ֆունկցիայի դասերի: Օրինակ, երբ $g(x,a)$-ն բազմանդամ է x-ից, ապա համապատասխան **ռեգրեսիայի ֆունկցիան** կոչվում է **բազմանդամային**, երբ այն եռանկյունաչափական բազմանդամ է՝ **եռանկյունաչափական**, երբ $\{g_m(x),\ m=\overline{1,M}\}$-ը համակարգը օրթոգոնալ է, ապա՝ **օրթոգոնալ** և այլն:

12.2. Փոքրագույն քառակուսիների եղանակը բազմաչափ դեպքում

Այս ենթաբաժնում կծանոթանանք փոքրագույն քառակուսիների եղանակի ընդհանրացմանը բազմաչափ դեպքի համար, $M > 1$:

Դիցուք տրված են $(M+1)$-չափանի (X,Y) պատահական վեկտորի N ծավալի $(\mathbf{X},\mathbf{Y})$ նմուշը, որտեղ $X_n = (X_{n1}, X_{n2}, \ldots, X_{nM})$, և $g(x,a)$ ֆունկցիան:

Նշանակենք $\hat{y} = g(x,\hat{a})$, $\hat{y}_n = g(x_n,\hat{a})$, $W(a) = \sum\limits_n [y_n - g(x_n,a)]^2$: Ըստ փոքրագույն քառակուսիների եղանակի պահանջվում է գտնել այնպիսի $\hat{a}$ վեկտոր, որ $W(\hat{a})$ քառակուսիների գումարը լինի փոքրագույնը՝

$$\min_a W(a) = \sum_n [y_n - g(x_n,\hat{a})]^2 : \tag{3}$$

Սահմանափակվենք (2) տեսքի գծայնացված ֆունկցիաների դեպքով, նշանակելով $1 \equiv g_0(x)$, կստանանք

$$\min_a W(a) = \sum_n [y_n - \sum_{m=0}^{M} \hat{a}_m g_m(x_n)]^2 :$$

Ածանցելով $W(a)$ ֆունկցիան ըստ a_m-երի և ածանցյալները հավասարեցնելով 0-ի կստանանք՝

$$\frac{\partial W}{\partial \hat{a}_{m'}} = -2[\sum_n y_n g_{m'}(x_n) - \sum_n \sum_{m=0}^{M} \hat{a}_m g_m(x_n) g_{m'}(x_n)] = -2[b_{m'} - \sum_{m=0}^{M} \hat{a}_m c_{mm'}] = 0,$$

որտեղ

$$m' = \overline{0,M},\ b_{m'} = \sum_n y_n g_{m'}(x_n);\ c_{mm'} = \sum_n g_m(x_n) g_{m'}(x_n), \tag{4}$$

կամ

$$\sum_m c_{mm'} \hat{a}_m = b_{m'},\quad m' = \overline{0,M} : \tag{5}$$

Հավասարումների (5) **համակարգը** անվանում են **նորմալ**:

Այսպիսով, բազմաչափ (2) տեսքի գծայնացված դեպքում փոքրագույն քառակուսիների եղանակը հանգում է հավասարումների նորմալ համակարգի լուծմանը:

12.3. Բազմաչափ գծային ռեգրեսիա

Բազմաչափ գծային ռեգրեսիայի մոդելն է՝

$$y = a_0 + \sum_m a_m x_m + Z,$$

որտեղ $Z \sim \mathcal{N}(0,\sigma^2)$, իսկ **նմուշային բազմաչափ ռեգրեսիայի ֆունկցիան** է՝

$$\hat{y} = \hat{a}_0 + \sum_m \hat{a}_m x_m,$$

որտեղ $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M$ հաստատունները ռեգրեսիայի նմուշային գործակիցներն են:

Այդ գործակիցները գտնելու համար օգտվենք (5) համակարգից: Տեղադրելով $g_0(x_n) = 1$ $x_n) = x_{nm}$ ($m > 0$ դեպքում) (4)-ում, ստանում ենք՝

$$b_m = \sum_n y_n x_{nm}; \ c_{mm'} = \sum_n x_{nm} x_{nm'} : \qquad (6)$$

Այնուհետև, տեղադրելով (6)-ը նորմալ հավասարումների (5) համակարգում, կստանանք՝

$$\begin{aligned}
&\hat{a}_0 N + \hat{a}_1 \sum_n x_{n1} + \hat{a}_2 \sum_n x_{n2} + \ldots + \hat{a}_M \sum_n x_{nM} = \sum_n y_n, \\
&\hat{a}_0 \sum_n x_{n1} + \hat{a}_1 \sum_n x_{n1}^2 + \hat{a}_2 \sum_n x_{n1} x_{n2} + \ldots + \hat{a}_M \sum_n x_{n1} x_{nM} = \sum_n x_{n1} y_n, \\
&\hat{a}_0 \sum_n x_{n2} + \hat{a}_1 \sum_n x_{n2} x_{n1} + \hat{a}_2 \sum_n x_{n2}^2 + \ldots + \hat{a}_M \sum_n x_{n2} x_{nM} = \sum_n x_{n2} y_n, \\
&\qquad\qquad \cdots\cdots\cdots\cdots \qquad\qquad (7) \\
&\hat{a}_0 \sum_n x_{nM} + \hat{a}_1 \sum_n x_{nM} x_{n1} + \hat{a}_2 \sum_n x_{nM} x_{n2} + \ldots + \hat{a}_M \sum_n x_{nM}^2 = \sum_n x_{nM} y_n :
\end{aligned}$$

Լուծելով ստացված համակարգը, կստանանք $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M$ նմուշային բազմաչափ ռեգրեսիայի գործակիցները: Համակարգի լուծման եղանակների վրա կանգ չենք առնի: Սակայն, քանի որ երկչափ դեպքը ($M = 2$) հաճախ է հանդիպում, ստորև կստանանք հավասարումների նորմալ համակարգի լուծման բացահայտ տեսքը: $M > 2$ դեպքում նպատակահարմար է օգտվել համապատասխան հաշվարկային ծրագրերից:

Օրինակ 1: Լուծել (7) համակարգը $M = 2$ համար:

Լուծում: Ստանում ենք՝

$$\begin{aligned}
&\hat{a}_0 N + \hat{a}_1 \sum_n x_{n1} + \hat{a}_2 \sum_n x_{n2} = \sum_n y_n, \\
&\hat{a}_0 \sum_n x_{n1} + \hat{a}_1 \sum_n x_{n1}^2 + \hat{a}_2 \sum_n x_{n1} x_{n2} = \sum_n x_{n1} y_n, \\
&\hat{a}_0 \sum_n x_{n2} + \hat{a}_1 \sum_n x_{n2} x_{n1} + \hat{a}_2 \sum_n x_{n2}^2 = \sum_n x_{n2} y_n :
\end{aligned}$$

Արտաքսելով $\hat{a}_0$-ն վերջին երկու հավասարումներից՝ կստանանք համակարգի վերջնական տեսքը՝

$$\hat{a}_1 d_{11} + \hat{a}_2 d_{12} = e_1, \ \hat{a}_1 d_{21} + \hat{a}_2 d_{22} = e_2, \ \hat{a}_0 = \overline{\mathbf{y}} - \hat{a}_1 \overline{\mathbf{x}}_1 - \hat{a}_2 \overline{\mathbf{x}}_2,$$

որտեղ

$$\begin{aligned}
&d_{11} = \sum_n x_{n1}^2 - (1/N)(\sum_n x_{n1})^2, \ d_{22} = \sum_n x_{n2}^2 - (1/N)(\sum_n x_{n2})^2, \\
&d_{12} = d_{21} = \sum_n x_{n1} x_{n2} - (1/N)(\sum_n x_{n1})(\sum_n x_{n2}), \\
&e_1 = \sum_n y_n x_{n1} - (1/N)(\sum_n x_{n1})(\sum_n y_n), \ e_2 = \sum_n y_n x_{n2} - (1/N)(\sum_n x_{n2})(\sum_n y_n) :
\end{aligned}$$

Այսպիսով, (7) համակարգի լուծումը $M = 2$ դեպքում տրվում է հետևյալ բանաձևերով՝

$$\hat{a}_0 = \overline{\mathbf{y}} - \hat{a}_1 \overline{\mathbf{x}}_1 - \hat{a}_2 \overline{\mathbf{x}}_2, \ \hat{a}_1 = \frac{e_1 d_{22} - e_2 d_{12}}{d_{11} d_{22} - d_{12}^2}, \ \hat{a}_2 = \frac{e_2 d_{11} - e_1 d_{12}}{d_{11} d_{22} - d_{12}^2} : \qquad (8)$$

Ինչպես և միաչափ ռեգրեսիայի դեպքում, Y-ի դիտումների արդյունքների՝ նմուշային միջինից շեղումների ընդհանուր քառակուսիների գումարը՝ $S^2(\mathbf{x}, \mathbf{y})$-ը, տրոհվում է երկու բաղադրիչների՝ $S^2(\mathbf{x}, \mathbf{y}) = S_1^2(\mathbf{x}, \mathbf{y}) + S_2^2(\mathbf{x}, \mathbf{y})$, որտեղ

$$\begin{aligned}
&S^2(\mathbf{x}, \mathbf{y}) = \sum_n (y_n - \overline{\mathbf{y}})^2 = \sum_n y_n^2 - (1/N)(\sum_n y_n)^2, \\
&S_1^2(\mathbf{x}, \mathbf{y}) = \sum_n (\hat{y}_n - \overline{\mathbf{y}})^2 = \\
&= \hat{a}_0 \sum_n y_n + \hat{a}_1 \sum_n x_{n1} y_n + \hat{a}_2 \sum_n x_{n2} y_n + \ldots + \hat{a}_M \sum_n x_{nM} y_n - (1/N)(\sum_n y_n)^2, \qquad (9) \\
&S_2^2(\mathbf{x}, \mathbf{y}) = \sum_n (y_n - \hat{y}_n)^2 = \\
&\qquad = \sum_n y_n^2 - \hat{a}_0 \sum_n y_n - \hat{a}_1 \sum_n x_{n1} y_n - \hat{a}_2 \sum_n x_{n2} y_n - \ldots - \hat{a}_M \sum_n x_{nM} y_n :
\end{aligned}$$

Միաչափ դեպքի նմանությամբ **որոշակիացման գործակից** է կոչվում

$$d_{YX}^2(\mathbf{x}, \mathbf{y}) = S_1^2(\mathbf{x}, \mathbf{y}) / S^2(\mathbf{x}, \mathbf{y})$$

հարաբերությունը:

Օրինակ 2: Որոշել աշխարհի խոշորագույն ընկերությունների մաքուր եկամուտի (գլուխ 11-ի աղյուսակում՝ X_3) գծային ռեգրեսիան շահույթի ցուցանիշի ($ROCE, X_4$) և ծառայողների քանակի (X_6) նկատմամբ, օգտագործելով աղյուսակի տվյալները և (8) բանաձևերը:

Լուծում: Մաքուր եկամուտը նշանակենք Y-ով, շահույթի ցուցանիշը՝ X_1-ով, իսկ ծառայողների քանակը՝ X_2-ով: Կստանանք՝

$$\sum_n x_{n1} = 491.5,\ \ \sum_n x_{n2} = 3138.8,\ \ \sum_n y_n = 85.45,\ \ \sum_n y_n^2 \approx 372.7055\ \ \sum_n x_{n1}^2 \approx 11923.23,$$

$$\sum_n x_{n2}^2 \approx 1014140, \sum_n x_{n1}x_{n2} \approx 46244.67,\ \ \sum_n x_{n1}y_n \approx 1720.216,\ \ \sum_n x_{n2}y_n \approx 11018.18:$$

Տեղադրելով այս արժեքները (8) բանաձևերի մեջ՝ կստանանք նորմալ համակարգի գործակիցների արժեքները, ինչպես նաև լուծումը՝

$$d_{11} \approx 2260.341,\ \ d_{12} \approx -15464.14,\ \ d_{22} \approx 620057.4,\ \ e_1 \approx 40.26904,\ \ e_2 \approx 289.7617,$$

$$\hat{a}_0 \approx 2.782,\ \ \hat{a}_1 \approx 0.02533,\ \ \hat{a}_2 \approx 0.0011:$$

Այսպիսով, մաքուր եկամուտի բազմակի գծային ռեգրեսիան նշված երկու փոփոխականների նկատմամբ ունի հետևյալ տեսքը՝

$$\hat{y}_n = 2.782 + 0.02533\,x_{1n} + 0.0011\,x_{2n}:$$

$\hat{a}_0$ գործակիցը հավասար է ռեգրեսիայի ուղղի՝ y-ների առանցքի հետ հատման կետի օրդինատին, իսկ $\hat{a}_1$ գործակիցը ցույց է տալիս, թե որքանով է փոփոխվում մաքուր եկամուտի սպասելին շահույթի ցուցիչի միավոր փոփոխության դեպքում, երբ ծառայողների քանակը հաստատուն է: Նման մեկնաբանություն կարելի է անել նաև $\hat{a}_2$ գործակցի նկատմամբ: Մասնավորապես, շահույթի ցուցիչի հաստատուն արժեքի դեպքում ծառայողների քանակի միավոր փոփոխությունը (այսինքն՝ 1000 մարդով փոփոխվելը) փոփոխում է մաքուր եկամուտը 1.1 մլն դոլարով (նույն նշանով): Սա նշանակում է, որ մեկ մարդուն ընկնող եկամուտի մասնաբաժինը մոտավորապես 1100 դոլար է:

Ինչպես ընդունեցինք, երբ $X = x$, Y-ի պայմանական բաշխումն է՝ $\mathcal{N}(a_0+a_1x_1+a_2x_2+ \ldots + a_Mx_M, \sigma^2)$: Ուստի ռեգրեսիայի ֆունկցիան և նրա հետ առնչվող մեծությունները օժտված են մի շարք հատկություններով, որոնք նման են նախորդ գլխում դիտարկված՝ միաչափ ռեգրեսիայի հատկություններին: Զևակերպենք դրանցից կարևորագույնները:

1. $\mathbf{E}(S_2^2(\mathbf{X},\mathbf{Y})) = (N-M-1)\sigma^2$, ուստի

$$\hat{s}_{YX}^2(\mathbf{X},\mathbf{Y}) = S_2^2(\mathbf{X},\mathbf{Y})/(N-M-1)$$

σ^2 ցրվածքի անշեղ գնահատու է: Դրա արժեքները անվանում են նաև **մնացորդային ցրվածք** և նշանակում $\hat{\sigma}_{YX}^2(\mathbf{x},\mathbf{y})$:

Եթե $a_1 = a_2 = \ldots = a_M = 0$, ապա $\mathbf{E}(S_1^2(\mathbf{X},\mathbf{Y})) = M\sigma^2$, իսկ $S_1^2(\mathbf{X},\mathbf{Y})/\sigma^2$ պատահական մեծությունը բաշխված է $\chi^2(M)$ օրենքով, ընդ որում $S_1^2(\mathbf{X},\mathbf{Y})$-ը և $S_2^2(\mathbf{X},\mathbf{Y})$-ը անկախ են:

2. Նմուշային ռեգրեսիայի $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M$ գործակիցները բաշխված են համատեղ նորմալ օրենքով, համապատասխան ռեգրեսիայի գործակիցների զուգամետ և անշեղ գնատուներ են, որոնք անկախ են $S_1^2(\mathbf{X},\mathbf{Y})$-ից և $S_2^2(\mathbf{X},\mathbf{Y})$-ից:

Ռեգրեսիայի $a_0, a_1, \ldots, a_M$ գործակիցների մասին վարկածները կարելի է դասակարգել երկու տեսակի.

ա) վարկած այն մասին, որ բոլոր գործակիցները, բացի a_0-ից, հավասար են զրոյի՝

$$H_0:\ a_1 = a_2 = \ldots = a_M = 0,\ H_1 = \overline{H_0}:$$

H_0-ն մեկնաբանվում է որպես բազմաչափ ռեգրեսիայի բացակայություն, այսինքն՝ երբ տեղի ունի H_0 վարկածը, $\mathbf{E}(Y|\mathbf{x}) = a_0$ (ռեգրեսիայի ֆունկցիան $\mathbf{x}$-ից կախված չէ): Այդ դեպքում $\bar{\mathbf{y}}$-ը ընդունվում է որպես Y-ի լավագույն կանխագուշակված արժեք:

H_0 վարկածն ստուգելու համար օգտվում ենք այն հանգամանքից, որ

$$\hat{F}(\mathbf{X},\mathbf{Y}) = \frac{S_1^2(\mathbf{X},\mathbf{Y})/M}{S_2^2(\mathbf{X},\mathbf{Y})/(N-M-1)} \sim \mathcal{F}(M, N-M-1):$$

Ուստի, եթե $\hat{F}(\mathbf{x},\mathbf{y}) > F_{1-\alpha}(M, N-M-1)$ Ֆիշերի բաշխման $(1-\alpha)$-քանորդից, ապա H_0 վարկածը ժխտվում է:

բ) վարկած այն մասին, որ ռեգրեսիայի գործակիցների որոշ ենթաբազմություն՝ $a_{m_1}, a_{m_2}, ..., a_{m_L}$, $1 \le L < M$, հավասար է զրոյի՝

$$H_0: \; a_{m_1} = a_{m_2} = ... = a_{m_L} = 0, \; H_1 = \overline{H_0}:$$

Այս դեպքը մեկնաբանվում է որպես վարկած այն մասին, որ որոշ սևեռված L հատ փոփոխականներ կանխագուշակիչ նշանակություն չունեն (հետևաբար դրանք նմուշային ռեգրեսիայի ֆունկցիայի սխալի վրա չեն ազդում):

H_0 վարկածը ստուգելու համար օգտվում են

$$\hat{F}(\mathbf{X},\mathbf{Y}) = \frac{S_{2,M-L}^2(\mathbf{X},\mathbf{Y}) - S_{2,M}^2(\mathbf{X},\mathbf{Y})}{S_{2,M}^2(\mathbf{X},\mathbf{Y})} \frac{N-M-1}{L} \sim \mathcal{F}(L, N-M-1) \qquad (10)$$

վիճակաnիg, որտեղ $S_{2,M}^2(\mathbf{X},\mathbf{Y})$-ը և $S_{2,M-L}^2(\mathbf{X},\mathbf{Y})$-ը համապատասխանաբար, հավասար են $S_2^2(\mathbf{X},\mathbf{Y}) = \sum_n (\hat{Y}_n - Y_n)^2$ մեծությանը, երբ ռեգրեսիայի ֆունկցիան պարունակում է բոլոր M կամ միայն նշված $M-L$ փոփոխականները: Եթե

$$\hat{F}(\mathbf{x},\mathbf{y}) > F_{1-\alpha}(L, N-M-1),$$

ապա H_0 վարկածը ժխտվում է:

Առանձին հետաքրքրություն է ներկայացնում այն կարևոր մասնավոր դեպքը, երբ $L=1$, այսինքն, ստուգվում է ռեգրեսիայի որևէ գործակցի նշանակալիությունը (այն ավելի մանրամասն կդիտարկենք 12.5 ենթաբաժնում):

Օրինակ 3: Նախորդ օրինակի տվյալներով $\alpha = 0.05$ համար ստուգենք H_0 վարկածը, ըստ որի մաքուր եկամուտը կախված չէ շահույթի ցուցանիշից և ծառայողների թվաքանակից (ռեգրեսիայի բոլոր գործակիցները, բացի a_0-ից, հավասար են զրոյի):

Լուծում: Օգտվելով (9) բանաձևերից՝ ստանում ենք

$S^2(\mathbf{x},\mathbf{y}) \approx 375.7055 - 85.45^2/25 \approx 80.6374,$

$S_1^2(\mathbf{x},\mathbf{y}) \approx (2.782)(491.5) + (0.02533)(1720.216) + (0.0011)(11018.18) - 85.45^2/25 \approx 1.3387,$

$S_2^2(\mathbf{x},\mathbf{y}) = 375.7055 - (2.782)(491.5) - (0.02533)(1720.216) - (0.0011)(11018.18) \approx 79.2987,$

$\hat{F}(\mathbf{x},\mathbf{y}) = (1.3387)(25-2-1)/(79.2987)(2) \approx 0.19 < F_{1-\alpha}(2,22) = 3.44,$

ուստի H_0 վարկածը ընդունելի է: Հետևաբար, հիմք չունենք պնդելու, որ աշխարհի 25 խոշոր ընկերությունների մաքուր եկամուտը կախված է շահույթի ցուցիչից և ընկերության աշխատակիցների քանակից:

Բազմաչափ ռեգրեսիաների վերլուծության արդյունքները կարելի է մեկնաբանել որպես միագործոն ցրվածքային վերլուծություն, միավորելով համապատասխան արդյունքները հետևյալ աղյուսակում:

Աղբյուրը	Ազատության աստիճաններ	Միջին քառակուսին	$\hat{F}(\mathbf{x},\mathbf{y})$
Ռեգրեսիա	M	$\hat{s}_1^2(\mathbf{x},\mathbf{y}) = \sum_n (\hat{y}_n - \overline{\mathbf{y}})^2/M$	$\hat{F} = \frac{\hat{s}_1^2(\mathbf{x},\mathbf{y})}{\hat{s}_2^2(\mathbf{x},\mathbf{y})}$:
Մնացորդը	$N-M-1$	$\hat{s}_2^2(\mathbf{x},\mathbf{y}) = \sum_n (y_n - \hat{y})^2/(N-M-1)$	
Ընդհանուրը	$N-1$	$\hat{s}^2(\mathbf{y}) = \sum_n (y_n - \overline{\mathbf{y}})^2/(N-1)$	

12.4. Բազմակի հարաբերակցության գործակից

Դիցուք՝ Y-ը և $X_1, X_2, ..., X_M$-ը պատահական մեծություններ են: Նորից դիտարկենք նմուշային բազմաչափ գծային ռեգրեսիայի ֆունկցիան՝ $\hat{y} = \hat{a}_0 + \sum_m \hat{a}_m x_m$:

> Y-ի և $X = (X_1, X_2, ..., X_M)$ պատահական վեկտորի **բազմակի հարաբերակցության գործակից** $\rho_{Y(X_1,X_2,...,X_M)}$ կոչվում է Y և $\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + ... + \hat{a}_M X_M$ պատահական մեծությունների հարաբերակցության գործակիցը:

Նշանակենք $Y = Z_0$; $X_m = Z_m$, $m = \overline{1, M}$ և դիտարկենք $(M+1)$-չափանի $(Z_0, Z_1, ..., Z_M)$ պատահական վեկտորը: Նշանակենք $\rho_{mm'}$-ով Z_m և Z'_m պատահական մեծությունների հարաբերակցության գործակիցը, $R = \|\rho_{mm'}\|$-ով՝ $Z_0, Z_1, ..., Z_M$ պատահական մեծությունների հարաբերակցային մատրիցը $(m, m' = \overline{0, M})$ և, պարզության համար, $\rho_{0(12...M)}$-ով՝ բազմակի հարաբերակցության գործակիցը: Կարելի է ապացուցել, որ բազմակի հարաբերակցության գործակիցն արտահայտվում է R մատրիցի բոլոր տարրերի միջոցով, հետևյալ բանաձևով

$$\rho^2_{0(12...M)} = 1 - |R| \,/\, |R_{00}|\,,$$

որտեղ $|R_{00}|$-ն R մատրիցի ρ_{00} տարրի հանրահաշվական լրացումն է:

Բազմակի հարաբերակցության գործակիցն արտահայտվում է նաև մասնակի հարաբերակցության գործակիցների միջոցով.

$$\rho^2_{0(12...M)} = 1 - (1 - \rho^2_{01})(1 - \rho^2_{02/1})(1 - \rho^2_{03/12})...(1 - \rho^2_{0M/12...M-1}) : \qquad (11)$$

Ելնելով բազմակի հարաբերակցության գործակցի սահմանումից՝ կարող ենք նկատել, որ այն Y-ի և X պատահական վեկտորի համաքական գծային կախման չափ է: Ցույց է տրվում, որ դրա մեծությունը հավասար է Y և $\sum_m a_m X_m$ պատահական մեծությունների հարաբերակցության գործակցի մեծագույն արժեքին՝ $a_1, a_2, ..., a_M$ գործակիցների արժեքների բազմության վրա:

Թվարկենք բազմակի հարաբերակցության գործակցի հիմնական հատկությունները:

ա) $0 \leq \rho_{0(12...M)} \leq 1$:

բ) $\rho_{0(12...M)} \geq |\rho_{0m}|$, $m = \overline{1, M}$, այսինքն՝ բազմակի հարաբերակցության գործակիցը փոքր չէ Y-ի և կամայական X_m-ի հարաբերակցության գործակցի բացարձակ արժեքից:

գ) $\rho_{0(12...M)} \geq \rho_{0m(I_m)}$, որտեղ $m = \overline{1, M}$, իսկ I_m-ը $\{1, 2, ..., M\}$ համարների կամայական ենթաբազմություն է:

դ) $\rho_{0(12...M)} = 0$ պայմանը նշանակում է, որ բոլոր սովորական և մասնակի հարաբերակցության գործակիցները, որոնք ներկայացված են (11)-ում, հավասար են զրոյի:

ե) Եթե որևէ մասնակի հարաբերակցության գործակից հավասար է 1 -ի, բազմակի հարաբերակցության գործակիցն ընդունում է իր առավելագույն 1 արժեքը:

զ) յուրաքանչյուր նոր (կանխազուշակող) X_m փոփոխականի ավելացումը, անկախ ավելացման հերթականությունից, չի նվազեցնում բազմակի հարաբերակցության գործակցի մեծությունը, այսինքն՝

$$\rho_{0(1)} \leq \rho_{0(12)} \leq \rho_{0(123)} \leq ... \leq \rho_{0(12...M)}:$$

Բազմակի հարաբերակցության գործակցի այս հատկություններն օգնում են ընտրություն կատարել բազմաչափ ռեգրեսիաների վերլուծության ժամանակ օգտագործվող տարբեր գծային մոդելների միջև, որոնք տարբերվում են ոչ միայն փոփոխականների քանակով, այլև դրանց համակցությամբ:

Նմուշային բազմակի հարաբերակցության գործակիցը ստացվում է նրա սահմանմանը մասնակցող մեծությունների փոխարեն տեղադրելով համապատասխան նմուշային արժեքները: Սակայն, դա հարմար է կատարել հաշվարկի օգնությամբ: Իսկ եթե փոքրաքույն քառակուսիների եղանակով կատարվել է բազմաչափ գծային ռեգրեսիային վերլուծություն, ապա կարելի է օգտվել ավելի հարմար բանաձևից՝

$$\hat{\rho}^2_{0(12\ldots M)} = 1 - \frac{(N-1)\sum\limits_n (y_n - \hat{a}_0 - \sum\limits_m \hat{a}_m x_{nm})^2}{(N-M-1)\sum\limits_n (y_n - \overline{\mathrm{y}})^2} :$$

Բազմակի հարաբերակցության գործակցի նշանակալիության մասին վարկածների զույգը հետևյալն է՝

$$H_0 : \rho_{0(12\ldots M)} = 0, \quad H_1 : \rho_{0(12\ldots M)} > 0 :$$

H_0 վարկածը նշանակում է, որ պատահական մեծությունների միջև գծային (հարաբերակցային) կախում չկա:

Եթե $Z_0, Z_1, \ldots, Z_M$ պատահական մեծություններն ունեն համատեղ նորմալ բաշխում, և H_0 վարկածը ճիշտ է, ապա

$$\hat{F}(\mathbf{X},\mathbf{Y}) = \frac{\hat{\rho}^2_{0(12\ldots M)}}{1-\hat{\rho}^2_{0(12\ldots M)}} \frac{N-M-1}{M} \sim \mathcal{F}(M, N-M-1) :$$

Ուստի, եթե $\hat{F}(\mathbf{x},\mathbf{y}) > F_{1-\alpha}(M, N-M-1)$, ապա H_0 վարկածը ժխտվում է:

Օրինակ 4: Հաշվենք բերքատվության ցուցանիշի և մնացած փոփոխականների բազմակի հարաբերակցության գործակիցը, ելնելով 11-րդ գլխի օրինակ 10-ում բերված նմուշային մասնակի հարաբերակցության գործակիցների արժեքներից:

Լուծում: Ընդունենք, որ Z_0-ն բերքատվության ցուցանիշն է, իսկ Z_1 և Z_2-ը՝ համապատասխանաբար գարնանային տեղումների և կուտակված ջերմության քանակները: Տեղադրելով (11)-ում $\hat{\rho}_{01} = 0.80$ և $\rho_{02/1} = 0.097$ արժեքները՝ կստանանք

$$\hat{\rho}^2_{0(12)} = 1 - (1 - \hat{\rho}^2_{01})(1 - \hat{\rho}^2_{02/1}) \approx 0.643,$$

որտեղից $\hat{\rho}_{0(12)} \approx 0.802$: Ստուգենք բերքատվության ցուցանիշի և գարնանային տեղումների ու կուտակված ջերմության քանակի միջև բազմակի հարաբերակցային կապի բացակայության մասին $H_0 : \rho_{0(12)} = 0$ վարկածը: Հաշվենք $F(\mathbf{X},\mathbf{Y})$-ի արժեքը, տեղադրելով

$$\hat{\rho}^2_{0(12)} = 0.643, \; M = 2, \; N = 20 :$$

Կստանանք $\hat{F}(\mathbf{x},\mathbf{y}) \approx 15.4$, մինչդեռ $F_{0.95}(2,17) = 3.59$, այսինքն H_0 վարկածը ժխտվում է:

Նկատենք, որ $\hat{\rho}_{0(12)} = 0.802$ արժեքը շատ մոտ է ρ_{01} հարաբերակցության գործակցի արժեքին (0.80), այսինքն, երկրորդ փոփոխականի (ջերմության քանակի) ավելացումը ռեգրեսիայի մոդելում լրացուցիչ կանխագուշակիչ նշանակություն չունի: Այս եզրակացությունը, իհարկե, մոտավոր է, ուստի նման դեպքերում անհրաժեշտ է ստուգել բ) կետում դիտարկված վարկածը $L = 1$ համար, ինչպես դա նկարագրված է վերևում:

12.5. Գծայնացված բազմաչափ ռեգրեսիա

Նորից դիտարկենք այն դեպքը, երբ X-ը միաչափ է: Ինչպես նշվեց 12.1 ենթաբաժնում, երբ ռեգրեսիայի ֆունկցիայի կախվածությունը x-ից գծային չէ, սակայն a վեկտորից գծային է՝

$$\mathbf{E}(Y|X=x) = g(x,a) = a_0 + \sum_m a_m g_m(x),$$

ապա, նշանակելով $x_m = g_m(x)$, այն ձևափոխում ենք

$$\mathbf{E}(Y|x) = g(x,a) = a_0 + \sum_{m=1}^{M} a_m x_m$$

տեսքի, և խնդիրը հանգում է բազմաչափ գծային ռեգրեսիաների վերլուծությանը: Օրինակ, երբ $g_m(x) = x_m = x^m$, ստանում ենք **բազմանդամային ռեգրեսիայի ֆունկցիա**՝

$$y = a_0 + \sum_m a_m x^m : \tag{12}$$

$M = 2$ դեպքում (12)-ը ստանում է հետևյալ տեսքը՝ $y = a_0 + a_1x + a_2x^2$, և կոչվում է **պարաբոլային ռեգրեսիայի ֆունկցիա**: Կիրառվում է նաև **պարաբոլային ռեգրեսիայի կենտրոնարկված** տարբերակը՝

$$y = a_0' + a_1'(x - x_0) + a_2'(x - x_0)^2,$$

ընդ որում x_0-ի փոխարեն հաճախ տեղադրում են $\overline{x}$ նմուշային միջինը:

Այսպիսով, գծայնացված ռեգրեսիայի ֆունկցիաների որոշման խնդիրը ըստ էության բազմաչափ գծային ռեգրեսիաների վերլուծության խնդիր է:

Օրինակ 5: Փոքրագույն քառակուսիների եղանակով ստանալ պարաբոլային ռեգրեսիայի գործակիցները:

Լուծում: Փոքրագույն քառակուսիների եղանակով ռեգրեսիայի a_0, a_1 և a_2 գործակիցները որոշվում են հավասարումների (8) համակարգի օգնությամբ, որը դիտարկվող դեպքում ընդունում է հետևյալ տեսքը՝

$$d_{11} = \sum_n x_n^2 - (1/N)(\sum_n x_n)^2,\ d_{12} = d_{21} = \sum_n x_n^3 - (1/N)(\sum_n x_n)(\sum_n x_n^2)^2,$$

$$d_{22} = \sum_n x_n^4 - (1/N)(\sum_n x_n^2)^2,\ e_1 = \sum_n y_n x_n - (1/N)(\sum_n x_n)(\sum_n y_n),$$

$$e_2 = \sum_n y_n x_n^2 - (1/N)(\sum_n x_n^2)(\sum_n y_n) :$$

Օրինակ 6: Ծխախոտի որոշակի տեսակը վաճառվում է մեկ անձին պատկանող 15 տարբեր կրպակներում: Կրպակների տերը յուրաքանչյուր կրպակի համար սահմանում էր անհատական գներ (X) և գրանցում մեկ օրվա ընթացքում վաճառված ծխախոտատուփերի քանակը (Y): Արդյունքները բերված են աղյուսակում: Պահանջվում է այդ տվյալներով կատարել ծխախոտատուփի գնի վրա վաճառքի ծավալի ռեգրեսիայի վերլուծությունը, օգտագործելով պարաբոլային ռեգրեսիայի կենտրոնարկված տարբերակը:

n	x_n	y_n	n	x_n	y_n	n	x_n	y_n
1	79	142	6	99	91	11	119	77
2	79	151	7	99	100	12	119	86
3	79	163	8	99	107	13	119	95
4	79	168	9	99	115	14	119	100
5	79	176	10	99	126	15	119	106

:

Լուծում: Կատարելով համապատասխան հաշվարկները՝ ստանում ենք $\overline{x} = 99$ արժեքի նկատմամբ կենտրոնարկված նմուշային պարաբոլային ռեգրեսիայի ֆունկցիան՝

$$\hat{y} = 107.8 - 1.68(x - 99) + 0.0465(x - 99)^2:$$

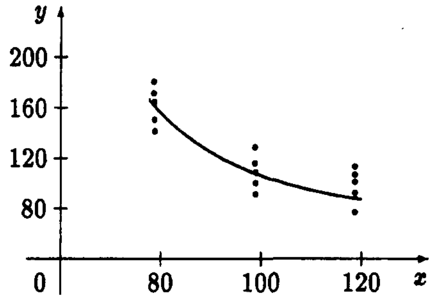
Նկարում բերված են դիտարկվող խնդրի ցրվածության տրամագիրը և պարաբոլի գծապատկերը, որոնք թույլ են տալիս վերլուծել ծխախոտի վաճառքի ծավալների և նրա գնի հարաբերակցությունը.

ա) ռեգրեսիայի ֆունկցիան նվազող է, այսինքն, գներն աճելիս վաճառքի ծավալը նվազում է (դա միանգամայն բնական է):

բ) ծխախոտը միջին գնով ($x = \overline{x} = 99$ դրամ) վաճառելու դեպքում վաճառքի սպասվող քանակը հավասար է $\hat{a}_0' \approx 108$ տուփի (այստեղ երևում է կենտրոնարկված ռեգրեսիայի առավելությունը ոչ կենտրոնարկվածի նկատմամբ, քանի որ $\hat{a}_0'$ գործակիցն անմիջականորեն ցույց է տալիս Y-ի սպասվող արժեքը X-ի միջին արժեքի համար): Այս դեպքում կրպակի սպասվող հասույթը հավասար կլինի $99 \times 108 = 10692$ դրամի:

գ) ծխախոտը $x = 79$ դրամով վաճառելիս վաճառքի սպասվող քանակը հավասար է 160 տուփի, իսկ սպասվող հասույթը՝ $79 \times 160 = 12640$ դրամի:

դ) ծխախոտը $x = 119$ դրամով վաճառելիս վաճառքի սպասվող քանակը հավասար է 93 տուփի, իսկ սպասվող հասույթը՝ $119 \times 93 = 11067$ դրամի:



Ինչպես երևում է այս արդյունքներից՝ ավելի ձեռնտու է ծխախոտը վաճառել 79, քան 99 կամ 119 դրամով:

Գծայնացված բազմաչափ ռեգրեսիայի մոդելի համար վարկածների ստուգումը կատարվում է նույն եղանակներով, ինչպես գծային դեպքում: Սակայն ռեգրեսիայի ֆունկցիայի գծայնացման առնչությամբ առաջանում են նոր խնդիրներ, որոնց լուծումը պահանջում է հատուկ մոտեցումներ:

Մասնավորապես, դիտարկենք բազմանդամային ռեգրեսիայի խնդիրը, երբ անհայտ տեսքի կանխագուշակիչը մոտարկվում է բազմանդամով: Ընդհանուր դեպքում կարող են անհայտ լինել ոչ միայն $a_0, a_1, ..., a_M$ գործակիցները, այլև նրանց քանակը՝ M-ը: Օրինակ, պարաբոլային ռեգրեսիայի դիտարկված խնդրում (տե՛ս օրինակ 6-ը) առաջանում է հետևյալ հարցը. հիմնավորվա՞ծ է արդյոք պարաբոլային մոդելի կիրառությունը, եթե կարելի է օգտվել ավելի պարզ տեսքի գծային մոդելից: Սա նշանակում է, որ մենք պետք է ընտրություն կատարենք գծային և պարաբոլային ռեգրեսիայի ֆունկցիաների միջև: Այսինքն՝ հարց է առաջանում, թե, առհասարակ, ինչպես **ընտրել բազմանդամի կարգը**, որպեսզի այն ոչ մեծ լինելով հանդերձ ապահովի ստացված արդյունքների վստահելիության հնարավորին չափ բարձր մակարդակ:

Այս խնդիրը լուծելու համար նախ ստուգում են հետևյալ վարկածը՝ $H_0 : a_2 = 0$, $H_1 : a_2 \neq 0$: Եթե H_0 վարկածը ժխտվում է, ապա ստուգում են նոր վարկած՝ $H_0 : a_3 = 0$, $H_1 : a_3 \neq 0$, ինչի արդյունքում ընտրություն են կատարում երկրորդ և երրորդ կարգի բազմանդամների միջև: Ընդհանրացնելով ասվածը, $M > 2$ դեպքում կարելի է ընտրություն կատարել $M-1$ և M աստիճանի բազմանդամային ռեգրեսիայի ֆունկցիաների միջև, ստուգելով հետևյալ վարկածը՝ $H_0 : a_M = 0$, $H_1 : a_M \neq 0$:

$H_0 : a_M = 0$ վարկածն ստուգելու համար օգտվում են (10) վիճականուց $L = 1$ դեպքում՝

$$\hat{F}(\mathbf{X}, \mathbf{Y}) = \frac{S^2_{2,M-1}(\mathbf{X}, \mathbf{Y}) - S^2_{2,M}(\mathbf{X}, \mathbf{Y})}{S^2_{2,M}(\mathbf{X}, \mathbf{Y})}(N - M - 1),$$

որտեղ

$$S^2_{2,M}(\mathbf{X}, \mathbf{Y}) = \sum_n (Y_n - \hat{Y}_n)^2 = \sum_n (Y_n - \hat{a}_0 - \sum_{m=1}^{M} \hat{a}_m X_n^m)^2 :$$

H_0 վարկածի ճիշտ լինելու դեպքում $\hat{F}(\mathbf{X}, \mathbf{Y}) \sim \mathcal{F}(1, N - M - 1)$: Հետևաբար, եթե $\hat{F}(\mathbf{x}, \mathbf{y}) > F_{1-\alpha}(1, N - M - 1)$, ապա H_0 վարկածը ժխտվում է:

<u>Օրինակ</u> 7: Ծխախոտի վաճառքի խնդրում համեմատել գծային և պարաբոլային ռեգրեսիայի ֆունկցիաները և հիմնավորել պարաբոլային ֆունկցիայի ընտրությունը: Ընդունել $\alpha = 0.05$:

Լուծում: Նախ ստանանք ցրվածքային վերլուծության աղյուսակը գծային և պարաբոլային ռեգրեսիաների համար:

Աղբյուրը	Ազատության աստիճանները	Միջին քառակուսին	$\hat{F}(\mathbf{x},\mathbf{y})$
Ռեգրեսիա (գծային)	1	$\hat{s}_1^2 = 11289.6$	$\hat{F} = 46.73$
Մնացորդ	$15-1-1=13$	$\hat{s}_2^2 = 241.6$	
Ռեգրեսիա (պարաբ.)	2	$\hat{s}_1^2 = 6221.4$	$\hat{F} = 37.57$
Մնացորդ	$15-2-1=12$	$\hat{s}_2^2 = 165.6$	
Ընդհանուրը	$15-1=14$		

Ինչպես տեսնում ենք աղյուսակից, գծային ռեգրեսիայի դեպքում ստացվում է $\hat{F}(\mathbf{x},\mathbf{y}) = 46.73$, մինչդեռ $F_{1-\alpha}(1,13) = 4.67$, այսինքն $H_0 : a_1 = 0$ վարկածը ժխտվում է: Նույն պատկերն է ստացվում նաև պարաբոլային ռեգրեսիայի դեպքում, քանի որ $\hat{F}(\mathbf{x},\mathbf{y}) = 37.57$, մինչդեռ $F_{1-\alpha}(2,12) = 3.89$: Սակայն մեզ հետաքրքրում է նաև այն հարցը, թե որքանով է արդարացված պարաբոլային ռեգրեսիայի ընտրությունը, քանի որ գծային ռեգրեսիան նույնպես նշանակալի է: Ուստի անհրաժեշտ է ստուգել $H_0 : a_2 = 0$ վարկածը: Ստանում ենք

$$\hat{F}(\mathbf{x},\mathbf{y}) = \frac{3140.8 - 1987.6}{1987.6}(15-2-1) \approx 6.96,$$

մինչդեռ $F_{1-\alpha}(1,12) = 4.75$: Այսպիսով, պարաբոլային ռեգրեսիայի ընտրությունն արդարացվում է, որովհետև պարաբոլի քառակուսային անդամի գործակիցը նշանակալի է: Հետևաբար, ընդունվում են նաև մեր եզրակացությունները, որոնք հիմնված են պարաբոլային ռեգրեսիայի ֆունկցիայի վերլուծության վրա:

Գլուխ 13
Ժամանակային շարքեր

Երկու մեծագույն բռնակալներ կան աշխարհում՝ պատահականությունն ու ժամանակը:

Յոհան Հերդեր

13.1. Հիմնական գաղափարներ

> **Ժամանակային շարք** կոչվում է որևէ ընթացքի ընդհատ պահերին կատարված դիտումների արդյունքների հաջորդականությունը:

Տնտեսական և գործարարական երևույթները սովորաբար ուսումնասիրվում են տարեկան, եռամսյակային, ամսական, շաբաթական և օրական, երբեմն էլ ոչ հավասար ժամանակահատվածներում կատարվող դիտումների հիման վրա: Ժամանակային շարքերի ուսումնասիրության նպատակն է, ելնելով տվյալ ժամանակահատվածում կատարված դիտումների արդյունքներից, հնարավորություն ընձեռել կանխագուշակումներ անելու ժամանակի հետագա պահերի կամ հատվածների համար:

11-րդ գլխում մենք տեսանք, որ ռեգրեսիայի ֆունկցիայի օգնությամբ կատարված կանխագուշակման սխալը կախված է X անկախ փոփոխականի ընդունած x արժեքից, ընդ որում միջին արժեքից էապես տարբերվող x-երի համար այդ սխալը էապես մեծ է: Հետևաբար առաջին հայացքից կարող է թվալ, թե ռեգրեսիայի վերլուծությունը ժամանակային շարքերի ուսումնասիրման համար պիտանի չէ: Սակայն պետք է հաշվի առնել երկու լրացուցիչ գործոնների առկայությունը, որոնք թույլ են տալիս ժամանակային շարքերի վերլուծության խնդիրներում հաջողությամբ կիրառել ռեգրեսիայի վերլուծության եղանակները: Առաջին գործոնը վերաբերում է տնտեսական երևույթների և ընթացքների կայունությանը, այն է՝ դիտարկվող ժամանակահատվածներում դրանք պահպանում են իրենց հատկությունները և օրինաչափությունները (հակառակ դեպքում առհասարակ հնարավոր չէր լինի որևէ կանխագուշակում անել): Երկրորդ գործոնը կապված է ուսումնասիրվող երևույթի՝ ժամանակից կախվածության կառուցվածքին: Հաճախ ժամանակային կախվածությունը ներկայանում է որպես ոչ պատահական (հիմնականում՝ միընթաց կամ տատանողական) և պատահական ընթացքների համակցություն:

> Պարզագույն դեպքում ժամանակային շարքը հետևյալ ընթացքի՝
>
> $$Y_t = f(t) + Z_t,\ t \in [0, T), \tag{1}$$
>
> $t_n,\ n = \overline{0, N-1}$, պահերին Y_t-ի դիտման արդյունքների հաջորդականությունն է՝
>
> $$\{Y_n,\ n = \overline{0, N-1}\},\ Y_0 = Y_{t_0},\ Y_1 = Y_{t_1},\ \dots,\ Y_{N-1} = Y_{t_{N-1}},$$
>
> որտեղ $f(t)$-ն ընթացքի ոչ պատահական բաղադրիչն է, որը կոչվում է **միտում**, իսկ Z_t-ն՝ պատահական բաղադրիչը:

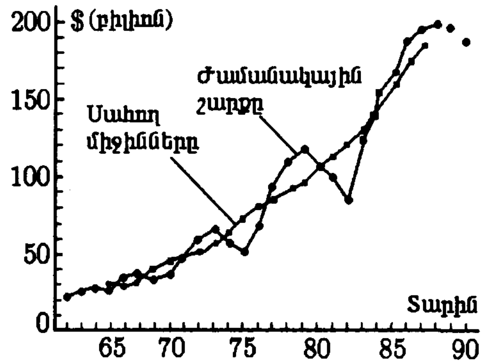
N թիվը կոչվում է ժամանակային շարքի երկարություն: Տնտեսագիտական հետազոտությունները հաճախ կատարվում են ըստ ժամանակի հավասարաչափ սանդղակի, երբ

$$t_n - t_{n-1} = t_{n-1} - t_{n-2} = \dots = t_2 - t_1,$$

և ենթադրվում է, որ $\mathbf{E}(Z_t) = 0$:

Տնտեսական բնագավառում ոչ պատահական բաղադրիչը, օրինակ, արտադրանքի ծավալն է, որը պայմանավորված է տնտեսական աճի ընդհանուր միտումով, գիտատեխնիկական առաջընթացով և ֆինանսական միջոցների ներդրումով: Դրա վրա, բացի տնտեսական գործոններից, կարող են կանխագուշակելի ազդեցություն ունենալ նաև այլ երկարատև գործոններ, ինչպիսիք են, օրինակ, բնակլիմայական գործոնները (որպես օրինակ նշենք Արևի՝ 11,2 տարի պարբերությամբ ակտիվության հայտնի ազդեցությունը մշակաբույսերի բերքատվության վրա):

<u>*Օրինակ*</u> 1: ԱՄՆ-ում 1961-1990 թվականներին բնակարանների մասնավոր շինարարության տարեկան ծավալները (բիլիոն դոլարներով) պատկերված են նկարում:



Նկարի վրա պարզորոշ նկատվում են տատանողական բնույթի ելևէջները, միընթաց աճող միտումը, ինչպես նաև պատահական շեղումները:

Ժամանակային շարքի վերլուծության հիմնական խնդիրն է՝ շարքի $\{y_n,\ n = \overline{0, N-1}\}$ տվյալների հիման վրա տարանջատել ոչ պատահական և պատահական բաղադրիչները և գնահատել դրանց բնութագրիչները:

Այդ խնդրի լուծումը հնարավորություն է տալիս գնահատել պատահական ընթացքի և նրա բաղադրիչների արժեքները ժամանակի անցյալ և ապագա պահերի համար:

Համեմատելով (1) ներկայացումը 12-րդ գլխում դիտարկված ռեգրեսիայի մոդելների հետ՝ տեսնում ենք, որ դրանց միջև սկզբունքային տարբերություն չկա: Այդ մոդելը կարելի է դիտել որպես ռեգրեսիայի մոդել, որտեղ ռեգրեսիայի ֆունկցիայի արգումենտը ժամանակն է, հետևաբար ժամանակային շարքերի վերլուծության համար կիրառելի են նախորդ գլուխներում դիտարկված մոտեցումներն ու հաշվարկային եղանակները:

$f(t)$ **միտումը** կոչվում է **գծային**, եթե $f(t) = a_0 + a_1 t$, **գծայնացված** կամ **բազմաչափ գծային**, եթե

$$f(t) = a_0 + \sum_m a_m g_m(t),\ t \in [0, T),$$

բազմանդամային, եթե

$$f(t) = a_0 + \sum_m a_m t^m,\ t \in [0, T),$$

պարբերական, եթե որևէ $T_0 \le T$ ամբողջ թվի համար $f(t + T_0) = f(t),\ t \in [0, T - T_0)$:

Գծային, գծայնացված և բազմանդամային միտումների վերլուծությունը կատարվում է նախորդ գլուխներում դիտարկված ռեգրեսիայի վերլուծության եղանակներով: Պարբերական միտումի վերլուծությունը կատարվում է եռանկյունաչափական բազմանդամների կիրառմամբ, դրան կանդրադառնանք հաջորդ ենթաբաժնում: Այնուհետև կծանոթանանք ժամանակային շարքերի ողորկացման եղանակների հետ:

13.2. Եռանկյունաչափական ռեգրեսիա

Դիցուք $t_n = nT/N$, $n = \overline{0, N-1}$: Հարմար է նաև համարել, որ շարքի N երկարությունը զույգ թիվ է:

Եթե խնդրի բնույթը թելադրում է T պարբերությամբ $f(t)$ ֆունկցիայի ընտրություն, ապա այն ներկայացնում են եռանկյունաչափական բազմանդամի տեսքով՝

$$f(t) = a_0 + \sum_{m=1}^{M/2-1} [a_{2m-1} \cos(m\frac{2\pi}{T}t) + a_{2m} \sin(m\frac{2\pi}{T}t)],$$

որտեղ M-ը զույգ թիվ է, ընդ որում $4 \leq M \leq N+2$: Եթե ֆունկցիան հաստատուն է, ապա $f(t) = a_0$ և եռանկյունաչափական գումարելիները բացակայում են: M թիվը ընտրվում է եռանկյունաչափական բազմանդամի անդամների քանակը (կարգը) որոշելու համար: Բազմանդամի կարգը մեկով ավելացնելու համար անհրաժեշտ է M-ի արժեքը մեծացնել 2-ով:

Դիտման t_n պահերի համար կարող ենք գրել

$$f(t_n) = a_0 + \sum_{m=1}^{M/2-1} [a_{2m-1} \cos(2\pi mn/N) + a_{2m} \sin(2\pi mn/N)],\ n = \overline{0, N-1}: \quad (2)$$

Պահանջվում է դիտումների $y_0, y_1, ..., y_{N-1}$ արդյունքներով գնահատել $a_0, a_1, ..., a_{M-2}$ գործակիցները: Դրանք կարելի է գնահատել, լուծելով բազմաչափ գծային ռեգրեսիայի խնդիր՝ օգտվելով 12-րդ գլխում բերված բանաձևերից: Սակայն դժվար չէ համոզվել, որ նորմալ հավասարումների համակարգն այս դեպքում ավելի պարզ տեսք ունի, քանի որ

$$\cos(2\pi t/T),\ \sin(2\pi t/T),\ \cos(4\pi t/T),\ \sin(4\pi t/T),\ ...$$
$$...,\ \cos((M/2-1)2\pi t/T),\ \sin((M/2-1)2\pi t/T),$$

ֆունկցիաների համակարգը օրթոգոնալ է $t_n = nT/N$, $n = \overline{0, N-1}$ կետերի նկատմամբ, այսինքն $m \neq m'$, $m, m' = \overline{0, M/2-1}$ համար

$$\sum_n \cos(2\pi mn/N) \sin(2\pi m'n/N) = 0,$$

$$\sum_n \cos(2\pi mn/N) \cos(2\pi m'n/N) = 0,$$

$$\sum_n \sin(2\pi mn/N) \sin(2\pi m'n/N) = 0:$$

Ուստի նորմալ հավասարումների համակարգի մատրիցն անկյունագծային է:

Նմուշային ռեգրեսիայի գործակիցները կստացվեն հետևյալ բանաձևերով՝

$$\hat{a}_0 = (1/N) \sum_n y_n,\ \hat{a}_{2m-1} = (2/N) \sum_n y_n \cos(2\pi mn/N),$$

$$\hat{a}_{2m} = (2/N) \sum_{n=0}^{N-1} y_n \sin(2\pi mn/N),\ m = 1, 2, ..., M/2-1:$$

Օրթոգոնալության հատկության շնորհիվ բավականին պարզ տեսք ունի նաև մնացորդային ցրվածքը՝

$$\hat{\sigma}^2_{M-2} = S^2_{M-2}/(N-M+1),$$

որտեղ

$$S^2_{M-2} = \sum_n (y_n - \hat{y}_n)^2 = \sum_n (y_n - \hat{a}_0)^2 - (N/2) \sum_{m=1}^{M/2-1} (\hat{a}^2_{2m-1} + \hat{a}^2_{2m}): \quad (3)$$

Այն ցույց է տալիս, թե ինչպես է նվազում քառակուսիների գումարը, կախված M-ի արժեքից:

Ռեգրեսիայի ֆունկցիայի վերջին երկու՝ a_{M-3} և a_{M-2} գործակիցների նշանակալիության մասին վարկածն ստուգելու համար 12-րդ գլխի (10) բանաձևի օգնությամբ և օրթոգոնալության շնորհիվ կկիրառվի հետևյալ պարզեցված վիճականին՝

$$\hat{F}(\mathbf{X},\mathbf{Y}) = \frac{N(\hat{a}_{M-3}^2 + \hat{a}_{M-2}^2)}{2S_{M-2}^2}\frac{N-M+1}{2}, \tag{4}$$

որն ունի $\mathcal{F}(2, N-M+1)$ բաշխում: Հետևաբար, եթե $\hat{F}(\mathbf{x},\mathbf{y}) > F_{1-\alpha}(2, N-M+1)$, ապա H_0 վարկածը ժխտվում է:

Ցույց է տրվում, որ գոյություն ունի M-ի այնպիսի արժեք ($M_{max} = N+2$), որի դեպքում $S_{M-2} = 0$: Սակայն այդ դեպքում ստացված ռեգրեսիայի ֆունկցիան անցնում է ժամանակային շարքի բոլոր (t_n, Y_n) կետերով, հարմարվելով ժամանակային շարքի պատահական բաղադրիչի փոփոխություններին, և սոսկ սահմանափակ տեղեկություններ է տալիս շարքի միտումի մասին: Ուստի հարց է առաջանում, թե ինչպե՞ս ընտրել եռանկյունաչափական բազմանդամի կարգը, որպեսզի այն հնարավորին չափ բարձր նշակալիություն ունենա, սակայն շատ բարդ չլինի հետագա վերլուծության համար:

Այս հարցի լուծումը հետևյալն է: Անհրաժեշտ է հաջորդաբար ավելացնել (2) մոտարկող եռանկյունաչափական բազմանդամի կարգը մեկ միավորով (այսինքն՝ M-ի արժեքը՝ 2-ով) ու (4)-ի օգնությամբ ստուգել ավելացված անդամների նշանակալիության մասին վարկածը և շարունակել այնքան, քանի դեռ նշված վարկածը ժխտվում է, այսինքն քանի դեռ ավելացված գործակիցները նշանակալի են: Նման մոտեցում մենք արդեն կիրառել ենք գլուխ 12-ում՝ բազմանդամային ռեգրեսիայի կարգը որոշելիս:

<u>*Օրինակ 2:*</u> «Ջեներալ Մոտորս» ընկերության արտադրանքի իրացման ծավալները 1970-1989 թթ. բերված են աղյուսակում: Կատարենք ժամանակային շարքի վերլուծություն $\alpha = 0.05$ համար:

Տարին	Արտադրանքի ծավալը (մլն միավոր)	$\hat{y}_n$ ($M=8$)	Սահող միջիններ ($L=7$)	Ցուցչ. ողորկացում ($\beta = 0.25$)
1970	5.3	7.12	–	5.3
1971	7.8	7.57	–	5.9
1972	7.8	7.74	–	6.4
1973	8.7	7.53	7.4	7.0
1974	6.7	7.26	7.9	6.9
1975	6.6	7.42	8.1	6.8
1976	8.6	8.13	8.3	7.3
1977	9.1	9.00	8.1	7.7
1978	9.5	9.35	8.1	8.2
1979	9.0	8.82	8.0	8.4
1980	7.1	7.66	7.9	8.1
1981	6.8	6.63	7.8	7.7
1982	6.2	6.44	7.8	7.4
1983	7.8	7.23	7.7	7.5
1984	8.3	8.43	7.8	7.7
1985	9.3	9.20	8.0	8.1
1986	8.6	9.07	8.3	8.2
1987	7.8	8.21	–	8.1
1988	8.1	7.29	–	8.1
1989	7.9	6.89	–	8.1

Լուծում: Նկարում պատկերված է աղյուսակի տվյալներն արտահայտող բեկյալը: Քանի որ պարզորոշ երևում է տվյալների փոփոխականության պարբերական բնույթը, ժամանակային շարքի վերլուծությունը կատարենք, վերցնելով որպես միտումի մոդել (2)-ով որոշվող եռանկյունաչափական բազմանդամը և գնահատելով բազմանդամի կարգը: Այս օրինակի համար $N = 20$, ուստի $M \le 22$: Ընդունենք $M = 10$ և կատարենք համապատասխան հաշվարկները: Նմուշային ռեգրեսիայի գործակիցների բանաձևերում տեղադրելով $2\pi mn/N = \pi mn/10$, $m = 1,2,3,4$ և $n = 0,1,...,19$ ստանում ենք՝

$\hat{a}_0 = (5.3 + 7.8 + ... + 8.1 + 7.9)/12 = 7.85,$

$\hat{a}_1 = [(5.3)(1) + (7.8)(0.9510566) + ... + (8.1)(0.809017) + (7.9)(0.9510566)]/6 = -0.2462842,$

$\hat{a}_2 = [(5.3)(0) + (7.8)(0.309017) + ... + (8.1)(-0.5877854) + (7.9)(-0.309017)]/6 = -0.00295665,$

$\hat{a}_3 = [(5.3)(1) + (7.8)(0.809017) + ... + (8.1)(0.309017) + (7.9)(0.809017)]/6 = -0.4622542,$

$\hat{a}_4 = [(5.3)(0) + (7.8)(0.5877853) + ... + (8.1)(-0.9510566) + (7.9)(-0.5877852)]/6 = -0.6449275,$

$\hat{a}_5 = [(5.3)(1) + (7.8)(0.5877852) + ... + (8.1)(-0.3090177) + (7.9)(0.587785)]/6 = -0.0245746,$

$\hat{a}_6 = [(5.3)(0) + (7.8)(0.809017) + ... + (8.1)(-0.9510566) + (7.9)(-0.809017)]/6 = -0.8872406,$

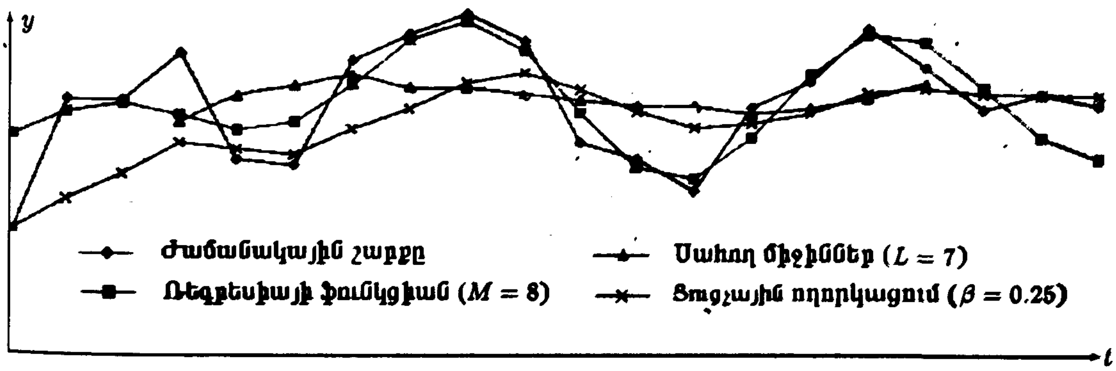
$\hat{a}_7 = [(5.3)(1) + (7.8)(0.309017) + ... + (8.1)(-0.809017) + (7.9)(0.309017)]/6 = -0.4601723,$

$\hat{a}_8 = [(5.3)(0) + (7.8)(0.9510566) + ... + (8.1)(-0.587785) + (7.9)(-0.9510566)]/6 = -0.1976028$:

Աղյուսակում բերված են $S^2_{M-2}(\mathbf{x},\mathbf{y})$, $\hat{\sigma}^2_{M-2}(\mathbf{x},\mathbf{y})$, $\hat{F}(\mathbf{x},\mathbf{y})$ գնահատականները, որոնք համապատասխանում են $M = 4,6,8,10$ արժեքներին, ինչպես նաև $\mathcal{F}(2, N - M + 1)$ բաշխման $(1-\alpha)$-քանորդիչները:

Վիճակագրական ցուցանիշը	$M = 4$	$M = 6$	$M = 8$	$M = 10$
$S^2_{M-2}(\mathbf{x},\mathbf{y})$	23.003	16.707	8.829	6.321
$N - M + 1$	17	15	13	11
$\hat{\sigma}^2_{M-2}(\mathbf{x},\mathbf{y})$	1.353	1.114	0.679	0.575
$\hat{F}(\mathbf{x},\mathbf{y})$	0.22	2.83	5.80	2.18
$F_{1-\alpha}(2, N - M - 1)$	3.59	3.68	3.81	3.98

Աղյուսակի տվյալներից հետևում է, որ ռեգրեսիայի գործակիցները նշանակալի են միայն $M = 8$ համար, ուստի հենց այդ արժեքն է ընդունվում որպես հետազոտվող ժամանակային շարքի միտումը նկարագրող եռանկյունաչափական բազմանդամի կարգի լավագույն գնահատական: $M = 8$-ին համապատասխանող ռեգրեսիայի ֆունկցիայի արժեքները բերված են նախորդ աղյուսակում: Դրանք պատկերված են նաև նկարում:



13.3. Ժամանակային շարքի ողորկացումը

Հետազոտվող ժամանակային շարքի միտումը հայտնաբերելու ամենատարածված և պարզ եղանակներից է շարքի **ողորկացումը**: Ողորկացման էությունն այն է, որ շարքի փաստացի տվյալները փոխարինվում են հաշվարկային մեծություններով, որոնք ավելի պակաս չափով են փոփոխական (ավելի «ողորկ» են), քան նախնական տվյալները: Փոփոխականության պակասեցումը նպաստում է շարքի միտումի ավելի հստակ դրսևորմանը:

Ստորև դիտարկենք ժամանակային շարքի ողորկացման երեք եղանակ՝ սահող միջինների, կշռավորված միջինների և **ցուցչային**:

Սահող միջինների եղանակը: Դիցուք, տրված է Y_t, $t = \overline{0, T-1}$ ժամանակային շարքը: Սահող միջինների եղանակով ողորկացման էությունն այն է, որ Y_t-ն փոխարինվում է նոր՝ $\mathbf{\overline{Y}}_t$ ժամանակային շարքով, ըստ հետևյալ կանոնի: Ընտրենք $L = 2M+1$ $(M \geq 1)$ կենտ թիվը, որտեղ $L \leq T$: Շարքի հաջորդական L անդամների համար որոշենք դրանց միջինը: Առաջին L անդամների միջինը նշանակենք $\mathbf{\overline{Y}}_M$-ով, $Y_2, ..., Y_{L+1}$ անդամների միջինը՝ $\mathbf{\overline{Y}}_{M+1}$-ով և այլն: Այսինքն՝

$$\mathbf{\overline{Y}}_t = \frac{1}{L}\sum_{m=-M}^{M} Y_{t+m},\ M \leq t \leq T-M-1 : \tag{5}$$

Այստեղ L-ը կոչվում է **ողորկացման միջակայք**, որը կարծես թե, սահում է ժամանակային շարքի վրայով՝ մեկական քայլով: Այստեղից էլ առաջացել է եղանակի անվանումը:

Ողորկացման արդյունքում ստացված (5) հաջորդականությունը կոչվում է **առաջին կարգի ողորկացված** ժամանակային շարք: Եթե սահող միջինների եղանակը կիրառենք (5)-ով որոշվող $\mathbf{\overline{Y}}_t$ ժամանակային շարքի համար, ապա այդ ձևով ստացված հաջորդականությունը կկոչվի **երկրորդ կարգի ողորկացված** ժամանակային շարք և այլն:

Սահող միջինների եղանակն ունի երկու թերություն: Առաջինը՝ այդ եղանակն օգտագործելու դեպքում հետազոտվող ժամանակային շարքի $2M$ անդամ դուրս են մնում վերլուծությունից: Դրանք են՝ $Y_0, Y_1, ..., Y_{M-1}$ և $Y_{T-M}, Y_{T-M+1}, ..., Y_{T-1}$ անդամները, որոնք անմիջականորեն չեն փոխարինվում որևէ այլ թվերով: Նկատենք, սակայն, որ T-ի մեծ արժեքների դեպքում այս հանգամանքն այնքան էլ էական չէ:

Երկրորդը՝ ողորկացման միջակայքի ընտրությունը զգալի չափով կամայական է, քանի որ չկան հստակ կանոններ դրանց ընտրության համար: Որքան ավելի մեծ է ողորկացման միջակայքը, այնքան ավելի շատ են մարում շարքի տատանումները, ուստի հայտնաբերված միտումը լինում է ավելի սահուն և ողորկ: Մասնավորապես, եթե շարքն ունի պարբերական բաղադրիչ, ապա վերջինս լրիվ վերանում է, եթե ողորկացման միջակայքը հավասար է այդ բաղադրիչի պարբերությանը կամ նրա պատիկն է: Ողորկացված $\mathbf{\overline{Y}}_t$ շարքի անդամի ցրվածքը հավասար է σ^2/L, որտեղ σ^2-ն ողորկացման միջակայքի անդամների ցրվածքն է: Այսինքն, պատահական բաղադրիչի ցրվածքը նույնպես փոքրանում է: Թվում է, թե որքան մեծ է ողորկացման միջակայքը, այնքան լավ, սակայն այս դեպքում միտումները դրսևորվում են միայն ամենաընդհանուր տեսքով, ուստի ժամանակային շարքի՝ տնտեսական վերլուծության տեսակետից անհրաժեշտ որոշ մանրամասներ կարող են կորչել: Այս պատճառով սովորական ողորկացման եղանակն անհրաժեշտ է կիրառել՝ ելնելով լուծվող խնդրի պահանջներից:

Ավելի նուրբ է, այսպես կոչված **կշռավորված միջինների եղանակը**, ըստ որի Y_t ժամանակային շարքը փոխարինվում է

$$\mathbf{\overline{Y}}_t = \sum_{m=-M}^{M} a_m Y_{t+m},\ \sum_{m=-M}^{M} a_m = 1$$

շարքով, որտեղ a_m գործակիցները կոչվում են կշիռներ:

Կշիռների որոշման համար օգտագործվում են տարբեր սկզբունքներ: Դիտարկենք դրանցից երկուսը: Առաջինը հենվում է $(\frac{1}{2}+\frac{1}{2})^{2M}$ երկանդամի վերածման գործակիցների համակարգի վրա: Այդ դեպքում կշիռները ունեն հետևյալ տեսքը՝

$$a_m = \frac{2M!}{(M+m)!(M-m)!2^{2M}}, \quad m = \overline{-M, M}:$$

Օրինակ, $M = 1, 2, 3$ համար ստանում ենք

$M = 1,\ \{a_m\} = \{1/4,\ 1/2,\ 1/4\}$,

$M = 2,\ \{a_m\} = \{1/16,\ 1/4,\ 3/8,\ 1/4,\ 1/16,\}$

$M = 3,\ \{a_m\} = \{1/64,\ 3/32,\ 15/64,\ 5/16,\ 15/64,\ 3/32,\ 1/64\}$, և այլն:

Երկրորդ սկզբունքը հենվում է $2M + 1$ երկարությամբ ողորկացման միջակայքի տվյալներով նմուշային բազմանդամային ռեգրեսիայի ֆունկցիան որոշելու վրա, որի արդյունքում միջակայքի $(M + 1)$-րդ (կենտրոնական) կետին վերագրվում է այդ կետում ռեգրեսիայի ֆունկցիայի օգնությամբ կանխագուշակված արժեքը: Ողորկացման միջակայքը «սահեցնում» են ժամանակային շարքի վրայով, ամեն անգամ որոշելով նմուշային ռեգրեսիայի ֆունկցիան: Կարելի է ցույց տալ, որ ռեգրեսիայի ֆունկցիան առաջին կարգի բազմանդամով մոտարկելու եղանակը բերում է արդեն դիտարկված՝ սահող միջինների եղանակին: Իսկ երկրորդ կարգի բազմանդամով մոտարկելիս ստացվում է a_m կշիռների հետևյալ համակարգը (եթե միայն ժամանակային շարքը ստացվել է ժամանակի հավասար միջակայքերով դիտումներ կատարելիս)՝

$M = 2,\ \{a_m\} = (1/35)\{-3, 12, 17, 12, -3\}$,

$M = 3,\ \{a_m\} = (1/21)\{-2, 3, 6, 7, 6, 3, -2\}$,

$M = 4,\ \{a_m\} = (1/231)\{-21, 14, 39, 54, 59, 54, 39, 14, -21\}$

և այլն:

Վերջում նշենք, որ սահող միջինների եղանակով ողորկացումը պարզ գործողություն է, որն ունի հստակ բովանդակություն: Սակայն այս գործողությունը ժամանակային շարքն ավելի շատ է ձևափոխում, քան կարող է թվալ առաջին հայացքից: Եթե մինչև ողորկացումը ժամանակային շարքի անդամներն անկախ էին, ապա ողորկացումից հետո հաջորդական անդամների միջև առաջանում է որոշակի հավանականային կախվածություն:

Օրինակ 3: Աղյուսակում ներկայացված «Ջեներալ Մոտորս» ընկերության արտադրանքի իրացման ծավալների վերաբերյալ տվյալներով կատարել ժամանակային շարքի ողորկացում սահող միջինների եղանակով:

Լուծում: Ողորկացման միջակայքի համար ընդունենք $M = 3$ $(L = 7)$ և կատարենք ողորկացում $t = 3, 4, ..., 16$ համար՝

$$\overline{y}_3 = (5.3 + 7.8 + 7.8 + 8.7 + 6.7 + 6.6 + 8.6)/7 = 51.5/7 = 7.36,$$

......

$$\overline{y}_{16} = (7.8 + 8.3 + 9.3 + 8.6 + 7.8 + 8.1 + 7.9)/7 = 57.8/7 = 8.26,$$

Օրինակ 2-ի աղյուսակում բերված է այս եղանակով ողորկացված շարքը $L = 7$ համար: Ինչպես տեսնում ենք, ստացված շարքը իսկապես բավականին «ողորկ» է, այնտեղից անհետացել են ինչպես պատահական, այնպես էլ պարբերական բաղադրիչները:

Ցուցչային ողորկացում: Այս եղանակով ողորկացման դեպքում գործակիցներն ընտրվում են այնպես, որ դիտման արդյունքները մասնակցում են ողորկացմանը՝ ցուցչային օրենքով նվազող կշիռներով:

Ցուցչային ողորկացման մոդելն ունի հետևյալ տեսքը՝

$$\overline{Y}_0 = Y_0,\ \overline{Y}_t = \beta Y_t + (1-\beta)\overline{Y}_{t-1},\ t = 1, 2, ..., T-1,$$

որտեղ $0 \leq \beta \leq 1$ և, ինչպես նախկինում,

$$Y_t = f(t) + Z_t,\ \mathbf{E}(Z_t) = 0,\ \mathbf{D}(Z_t) = \sigma^2,$$

Որպեսզի երևա կշիռների աճման ցուցչային հատկությունը, հետևենք ողորկացման ընթացքին $t = 0, 1, ..., T-1$ պահերին: Կստանանք

$$\overline{\mathbf{Y}}_1 = \beta Y_1 + (1-\beta)\overline{\mathbf{Y}}_0 = \beta Y_1 + (1-\beta)Y_0,$$

$$\overline{\mathbf{Y}}_2 = \beta Y_2 + (1-\beta)\overline{\mathbf{Y}}_1 = \beta Y_2 + \beta(1-\beta)Y_1 + (1-\beta)^2 Y_0,$$

$$\cdots\cdots \qquad (6)$$

$$\overline{\mathbf{Y}}_t = \beta Y_t + (1-\beta)\overline{\mathbf{Y}}_{t-1} = \beta Y_t + \beta(1-\beta)Y_{t-1} + \beta(1-\beta)^2 Y_{t-2} + ... + (1-\beta)^t Y_0 :$$

Ինչպես տեսնում ենք, որքան t-ն մեծանում է, այնքան t-ի փոքր արժեքներին համապատասխանող դիտման արդյունքներն ավելի փոքր կշռով են մասնակցում $\overline{\mathbf{Y}}_t$ ողորկացված ժամանակային շարքի ձևավորմանը: t-ի մեծ արժեքների դեպքում ողորկացված շարքի ցրվածքը հավասար է

$$\overline{\sigma}^2 = \frac{\beta}{2-\beta}\sigma^2 \leq \sigma^2 :$$

Քանի որ $0 \leq \beta \leq 1$, ապա դժվար չէ համոզվել, որ ողորկացված ժամանակային շարքի ($\overline{Y}_t$) ցրվածքը չի գերազանցում ելակետային ժամանակային շարքի (Y_t) ցրվածքը, այսինքն`

$$\overline{\sigma}^2 \leq \sigma^2 :$$

Որքան ավելի փոքր է β-ն, այնքան ավելի մեծ է ցուցչային ողորկացման ազդեցությունը:

<u>*Օրինակ 4:*</u> Ցուցչային ողորկացման եղանակով կատարենք օրինակ 2-ի աղյուսակի տվյալների վերլուծությունը, ընդունելով $\beta = 0.25$:

Լուծում: Կիրառելով (6)-ը, կստանանք

$\overline{y}_0 = 5.3,$

$\overline{y}_1 = (0.25)(7.8) + (1-0.25)(5.3) = 5.9,$

$\overline{y}_2 = (0.25)(7.8) + (1-0.25)(5.9) = 6.4,$

....

$\overline{y}_{18} = (0.25)(8.1) + (1-0.25)(8.1) = 8.1,$

$\overline{y}_{19} = (0.25)(8.1) + (1-0.25)(8.1) = 8.1 :$

Ցուցչային եղանակով ողորկացված ժամանակային շարքի վերլուծության տվյալները բերված են օրինակ 2-ի աղյուսակում:

Գլուխ 14
Տնտեսաչափության տարրեր

Տնտեսաչափությունը տնտեսագիտության դերի մասին որոշակի հայեցակարգ է: Այն ապահովում է մաթեմատիկական տնտեսագիտության՝ մոդելներ կառուցելիս թվային արդյունքներ ստանալու, և ընդգրկում է տնտեսական տվյալներին փորձնական հիմնավորում ստանալու մաթեմատիկական վիճակագրության կիրառությունները:

Պոլ Սամուելսոն

14.1. Տնտեսաչափության էությունն ու խնդիրները

Նախևառաջ ընթերցողին ծանոթացնենք տնտեսագիտության բնագավառում Նոբելյան մրցանակակիրների կողմից տնտեսաչափության բնութագրումներին: Դրանցից առաջինը մենք օգտագործել ենք որպես բնաբան: Իսկ հաջորդ ձևակերպումը Ս. Գոլդբերգինն է.

> «Տնտեսաչափությունը կարելի է սահմանել որպես սոցիալական գիտություն, որտեղ տնտեսագիտության, մաթեմատիկայի և վիճակագրության եզրահանգումների միաձուլումը կիրառվում է տնտեսական օրինաչափությունների վերլուծության համար»:

Իսկ Ս. Այվազյանի և Վ. Մխիթարյանի՝ 1998 թ. հրատարակված մեծածավալ գրքի «Տնտեսաչափության հիմունքներ» բաժնում (որը կազմում է 365 էջ) կարդում ենք.

> «**Տնտեսաչափությունը** գիտության ինքնուրույն տնտեսա-մաթեմատիկական ճյուղ է, որը թույլ է տալիս տնտեսագիտական դրույթների և տնտեսական վիճակագրության սկզբնանյութի հիման վրա, օգտագործելով անհրաժեշտ մաթեմատիկական-վիճակագրական գործիքներ, տնտեսագիտական տեսությամբ պայմանավորված ընդհանուր (որակական) օրինաչափություններին տալ որոշակի քանակական արտահայտություն»:

Այնուհետև, հարմար ենք գտնում մեջբերել այդ գրքից տնտեսաչափության հիմնական խնդիրների սեղմ և դիպուկ նկարագրությունը.

«Տնտեսաչափության ձևակերպված սահմանումից հետևում է, որ դրա գլխավոր նպատակը տնտեսագիտական և սոցիոլոգիական կիրառություններն են, հատկապես վերլուծվող ցուցանիշների միջև գոյություն ունեցող որոշակի քանակական փոխադարձ կապերի մոդելային նկարագրությունը:

Տնտեսաչափության օգնությամբ լուծվող խնդիրների ամբողջ բազմազանությունը հարմար է դասակարգել ըստ երեք հատկանիշի. ըստ վերջնական կիրառական նպատակների, ըստ ստորակարգության մակարդակի և ըստ վերլուծվող տնտեսական համակարգի տեսակի:

Ըստ վերջնական կիրառական նպատակների կարելի է առանձնացնել վերլուծման ենթարկված համակարգի զարգացումը և վիճակը բնութագրող տնտեսական ու սոցիալական ցուցանիշների *կանխագուշակումը* և դրա սոցիալ-տնտեսական զարգացման հնարավոր տարբերակների *նմանակումը*, երբ արտադրանքի սպառման, սոցիալական և ֆինանսական քաղաքականության և այլնի բնութագրիչների միջև վիճակագրորեն հայտնաբերված փոխադարձ կապերը օգտագործվում են հետազոտելու, թե ինչպես են այս կամ այն ղեկավարվող արտադրության և բաշխման գործոնների հնարավոր (պլանավորվող) փոփոխությունները ազդում մեզ հետաքրքրող «ելքային» բնութագրիչների վրա:

Ըստ վերլուծման առարկա՝ տնտեսական համակարգի ստորակարգության մակարդակի, առանձնացվում են *մակրոմակարդակը* (այսինքն՝ ամբողջ երկրի), *մեզոմակարդակը* (ռեգիոններ, ճյուղեր, միավորումներ) և *միկրոմակարդակը* (ընտանիքներ, ձեռնարկություններ, ֆիրմաներ):

Որոշ դեպքերում հարկավոր է որոշել տնտեսաչափական մոդելավորման բնագավառը՝ հետազոտությունը կարող է կենտրոնացվել շուկայի հիմնահարցերի, ներդրումների, ֆինանսական կամ սոցիալական քաղաքականության, գնակազմության, բաշխման հարաբերությունների, պահանջարկի ու սպառման կամ հիմնահարցերի որոշակի համալիրի վրա: Սակայն որքան տնտեսաչափական հետազոտությունը հիմնահարցերի ընդգրկման տեսակետից հավակնոտ է, այնքան պակաս են այն արդյունավետ անցկացնելու հնարավորությունները»:

Տնտեսաչափությունը մշակում և կիրառում է վիճակագրական եղանակներ, որոնց միջոցով փորձնական տվյալների վերլուծության հիման վրա գնահատվում են տնտեսական երևույթների միջև առկա քանակական փոխկապվածությունները:

Այս գլխում նպատակ է դրվում ներկայացնելու տնտեսաչափության էությունը և դրա առանձնահատկությունները, ինչպես նաև որոշ տնտեսաչափական մոդելների օրինակներ:

Տնտեսաչափական վերլուծության անկյունաքարերն են՝

-տնտեսագիտական վարկածների և մոդելների ձևակերպումը,

-ռեգրեսիայի մոդելի կառուցումը,

-ներգրավված փոփոխականների, պարամետրերի, պատահական շեղումների և բուն հավասարումների վիճակագրական տեստավորումը,

-մոդելից ստացված արդյունքների վստահելիության և իրական պատկերին համապատասխանության ստուգումը:

Տնտեսաչափության առանցքային խնդիրներից է նմուշի հիման վրա կառուցված ռեգրեսիային հավասարման (հավասարումների համակարգի) կամ ժամանակային շարքերի հետազոտման միջոցով այս կամ այն տնտեսագիտական վարկածի ստուգումը:

Կարևոր է, որ ստացված արդյունքները բավարարեն հետևյալ երեք պայմաններին: Առաջինը. պետք է համոզվել, որ ուսումնասիրվող երևույթի առաջադրված տնտեսագիտական մոդելը լավագույն ձևով է համապատասխանում նմուշին: Երկրորդը՝ ցանկալի է, որ մոդելի բոլոր պարամետրերի վերաբերյալ վարկածների ստուգման արդյունքներն ունենան հնարավորին չափ բարձր նշանակալիություն: Եթե այդ երկու պայմանները բավարարված են, անցնում ենք տնտեսաչափական վերլուծության հաջորդ փուլին՝ կանխագուշակմանը, որի արդյունքները պետք է բավարարեն երրորդ պայմանին՝ լինեն իմաստալից:

Օրինակ, միկրոտնտեսագիտությունից հայտնի է, որ եթե որևէ ապրանքի գինը աճում է, ապա այլ պայմանների հաստատուն լինելու դեպքում այդ ապրանքի նկատմամբ պահանջարկը պետք է կրճատվի: Հարց է առաջանում՝ ինչքա՞ն կնվազի պահանջարկը: Պարզ է, որ եթե մենք ունենանք մաթեմատիկական բանաձև, որը տվյալ ապրանքի իրացման քանակը կապում է նրա գնի հետ, ապա կկարողանանք համապատասխան հաշվարկներ կատարել:

Տնտեսաչափությունն օգտվում է մաթեմատիկական վիճակագրության վերլուծական եղանակներից: Տնտեսաչափական մոդելներում հաճախ օգտագործվում են նաև այնպիսի տեղեկություններ, որոնք հնարավոր չէ ստանալ փորձնական եղանակներով: Օրինակ, հնարավոր չէ ամբողջ ազգային տնտեսության համար կատարել փորձեր տարբեր հարկային դրույքաչափերի տեսանկյունից և ստուգել, թե ինչ արդյունքի կարելի է հասնել այս կամ այն տարբերակի դեպքում: Տնտեսագիտական հետազոտություններ կատարելիս հաճախ հնարավոր չէ իրականացնել վերահսկելի փորձեր տնտեսությունում տեղի ունեցող գործընթացների վերաբերյալ:

Տնտեսաչափությունը օգտվում է նաև այնպիսի թվային տվյալներից, որոնց նկատմամբ նախապես հայտնի է, որ դրանք պարունակում են չափման որոշակի անճշտություններ: Ավելին, տնտեսաչափությունն ունի հնարավորություն գնահատելու և գիտականորեն շրջանցելու այդպիսի տվյալների թերությունները:

Տնտեսագետները և գործարարները իրենց աշխատանքում հայտնվում են այնպիսի իրավիճակներում, երբ անհրաժեշտ է լինում կանխագուշակել, օրինակ, ինչ-որ ապրանքի վաճառքի քանակը կամ շուկայում դրամի պահանջարկը կամ բանկերի շահաբաժինը և այլն: Նման իրավիճակներում տնտեսաչափությունը կարող է անփոխարինելի դեր կատարել տնտեսական տարբեր իրավիճակների մոդելավորման, կանխագուշակման, հիմնավորված և շահավետ որոշումներ կայացնելու գործում:

Ուսումնասիրվող տնտեսական խնդրի (երևույթի) տնտեսագիտական նկարագրությունը մենք կանվանենք **տնտեսագիտական մոդել**:

Տնտեսագիտական մոդելը (տե՛ս նաև գլուխ 8) հնարավորություն է տալիս առանձնացնել երևույթի էական գործոնները, ուշադրություն չդարձնել երկրորդական (մասնակի) հանգամանքների վրա և ի հայտ բերել ուսումնասիրվող օբյեկտի հիմնական հատկությունները:

Ինչպե՞ս է ստեղծվում տնտեսագիտական մոդելը: Առաջին հերթին պետք է հստակ ձևակերպել հետազոտվող խնդրի բովանդակությունը և տեսնել, թե ինչպիսի կապեր է ակնկալում տնտեսագիտությունը դիտարկվող փոփոխականների միջև: Այնուհետև ուսումնասիրվող երևույթի վերաբերյալ ձևակերպվում և հիմնավորվում է վարկածը:

Տեսությունում տրվում է ապրանքի գնի և իրացման ծավալի փոխկապվածության միտումը այլ ապրանքների գների և սպառողների եկամուտների հաստատուն մնալու պայմաններում: Սակայն ցանկալի է գիտենալ այդ կապի ֆունկցիոնալ տեսքը, օրինակ, անկման օրինաչափությունը գծային ֆունկցիա՞ է, թե՞ ոչ:

Օգտագործվող փոփոխականը կարող է լինել **բացատրվող** կամ **բացատրող** (կախյալ և անկախ փոփոխականների նմանությամբ), ինչպես նաև **էնդոգեն** կամ **էկզոգեն**, **կանխորոշված** կամ **կեղծ** (ֆիկտիվ):

Էնդոգեն փոփոխականը բնութագրում է ուսումնասիրվող տնտեսական երևույթի (համակարգի) արդյունավետությունը: Էնդոգեն փոփոխականները փոխկապված են: Դրանց արժեքները որոշվում են մոդելի «ներսում», տնտեսական համակարգի գործունեության արդյունքում և կախված են մեծ թվով այլ փոփոխականների արժեքներից (բացատրող, կանխորոշված և այլն): Այդ պատճառով էնդոգեն փոփոխականները, ըստ էության, պատահական բնույթ ունեն և, հետևաբար, կարող են որոշվել միայն որոշակի ճշտությամբ:

Էկզոգեն փոփոխականների արժեքները տրվում են մոդելից «դուրս», դրանք նախապես հայտնի են, և մոդելը չի բացատրում, թե դրանք ինչպես են ստացվում: Ենթադրվում է, որ էկզոգեն փոփոխականները հարաբերակցված չեն պատահական շեղումների հետ: Դրանք անվանվում են նաև **բացատրող** փոփոխականներ կամ ռեգրեսորներ, որոնց միջոցով բացատրվում է էնդոգեն փոփոխականների վարքագիծը:

Կանխորոշված կոչվում են այն էնդոգեն **փոփոխականները**, որոնք մասնակցում են մոդելում նաև անցած ժամանակաշրջանի տարբեր պահերին, և որոնց ազդեցությունը բնութագրվում է ինչպես ժամանակի ընթացիկ պահերին, այնպես էլ որոշ ուշացումով, որն անվանում են **լագ**: Դա է պատճառը, որ այդպիսի փոփոխականները կոչվում են **լագային**: Լագային փոփոխական կարող է լինել նաև էկզոգեն փոփոխականը:

Կեղծ են կոչվում այն փոփոխականները, որոնք բնութագրում են գործոնի որակական (ոչ քանակական) կողմը: Օրինակ, ծառայողի օտար լեզվի իմացությունը, սեռը և այլն:

Մոդելի մանրամասներ ասելով հասկանում ենք հետևյալ հարցերը՝

ա. ինչպիսի՞ բացատրող փոփոխականներ են ներգրավված,

բ. ո՞ր պայմաններին են բավարարում փոփոխականները և պատահական շեղումները,

գ. ո՞ր մաթեմատիկական **ֆունկցիաներն են օգտագործվում ռեգրեսիայի**ն **մոդելը կառուցելու** համար:

Եթե ընտրված ֆունկցիան կամ փոփոխականներից որևէ մեկը կամ էլ բացատրող փոփոխականների վերաբերյալ ենթադրությունները չեն համապատասխանում իրականությանը, ապա ասում ենք, որ թույլ է տրված **մոդելի մանրամասների սխալ**: Հնարավոր են մոդելի մանրամասների երեք տեսակի սխալներ: Դրանք վերաբերում են ռեգրեսիայի մոդելի ֆունկցիոնալ տեսքին և էական (կարևոր) բացատրող փոփոխականի բացթողմանը կամ ոչ էական (կարևոր) փոփոխականի ներգրավմանը մոդելի մեջ:

Հաջորդ հիմնախնդիրը, որին մենք հանդիպում ենք թե՛ տնտեսական, թե՛ սոցիալական իրավիճակներ հետազոտելիս, այդ իրավիճակների բնութագրիչների չափման և գնահատման դժվարություններն են: Օրինակ, եթե մենք գործ ունենք ընտանեկան բյուջեի հետազոտման խնդրի հետ, ապա ընթերցողը կընդունի, որ դա դժվար խնդիր է: Ենթադրենք մի պահ, որ մենք կարողացել ենք սահմանել, թե ինչ ենք հասկանում «ընտանեկան եկամուտ» ասելով: Հարց է առաջանում, թե ինչպես չափել այն, որովհետև, ի տարբերություն, ասենք գործվածք արտադրող ձեռնարկության, որի արտադրանքը կարելի է չափել, օրինակ, մետրերով, «ընտանեկան բյուջեի» պարագայում մեզ պետք է տեղեկություններ հավաքել ընտանիքի բոլոր անդամների եկամուտների մասին: Մենք պետք է ունենանք այդ խնդրի լուծման որոշակի մոտեցում, որը կարող է պարունակել նաև ոչ ճշգրիտ չափումներ: Հասկանալի է, որ հնարավոր չէ բոլոր ընտանիքների համար կատարել հաշվարկներ միևնույն ճշտությամբ: Այդպիսի հաշվարկներ կատարելիս, սովորաբար հետազոտողները որպես նմուշ վերցնում են ընտանիքների մի որոշակի բազմություն: Այստեղ մի շատ կարևոր հարց է ծագում՝ որքա՞նով է այդ նմուշը ներկայացուցչական, օրինակ, ամբողջ երկրի համար:

Այժմ ենթադրենք, որ մենք կարողացել ենք հաղթահարել տվյալների հավաքման հետ առնչվող դժվարությունները և ցանկանում ենք օգտագործել տեսական վարկածները քանակական եզրակացությունների համար: Տեսությունը մեզ հուշում է, որ, օրինակ, ընտանեկան ծախսերը կախված են այն եկամուտից, որը մնում է հարկերի վճարումից հետո, ընտանիքի առկա հարստությունից և, բնական է, նախորդ ժամանակահատվածում կատարած սպառման կառուցվածքից: Տեսությունը նաև հուշում է, որ սպառողական ծախսերը աճում են եկամուտի աճման հետ, բայց գոյություն ունի «հագեցման» կետ, որից հետո աճը խիստ դանդաղում է, չգերազանցելով որոշակի քանակ:

Տնտեսաչափությամբ զբաղվող մասնագետն ինքը պետք է որոշի մոդելի փոփոխականների միջև եղած ֆունկցիոնալ կապերը: Այստեղ կարևորվում են նաև առկա տեղեկությունները, հետազոտողի փորձառությունը և հմտությունը:

Եթե ընտրված է գծային մոդել, օրինակ $Y = \alpha_0 + \alpha_1 X$, որտեղ Y-ը ընտանեկան ծախսերն են, X-ը՝ եկամուտը, իսկ α_0-ն ու α_1-ը՝ մոդելի պարամետրը, ապա հաշվի առնելով, որ նմուշը պարունակում է ոչ լրիվ ճշգրիտ տեղեկություններ, տնտեսաչափական մոդելը պետք է նախատեսի նաև պատահական սխալի գոյությունը: Վերջնական տնտեսաչափական մոդելը կունենա հետևյալ տեսքը՝

$$Y = a_0 + a_1 X + U,$$

որտեղ U-ն պատահական մեծություն է և դասական ռեգրեսիայի դեպքում բավարարում է հետևյալ պայմաններին՝

$$\mathbf{E}(U) = 0,\ \mathbf{E}(U^2) = \sigma^2,\ \mathbf{E}(U_t, U_s) = 0 \ \ (\text{երբ } t \neq s)$$

(տե՛ս ստորև ներկայացվող Գաուսի-Մարկովի թեորեմի 3-րդ պայմանը):

Հարց է ծագում, իսկ ճիշտ չի՞ լինի արդյոք հետազոտել մեկ այլ ֆունկցիոնալ կապ՝

$$Y = a_0 + a_1 X + a_2 P + U,$$

որտեղ P-ն գնի ընդհանրացված ցուցիչն է տվյալ ժամանակահատվածի համար, և, ընդհանրապես, ինչու՞ չնայել այդ առնչությունները ժամանակի մեջ և հաշվի չառնել նաև անցյալ տարվա եկամուտը (t-ով նշելով ընթացիկ տարին)՝

$$Y_t = a_0 + a_1 X_t + a_2 P_t + a_3 X_{t-1} + U$$

և այլն:

Տնտեսաչափական մոդելի կիրառելիությունը և օգտակարությունը կախված է դրա իրական տնտեսական գործընթացների (երևույթների) արտապատկերման ճշգրտության աստիճանից:

Այս դրույթները և տնտեսաչափական մոդելավորման գործընթացը ցուցադրելու նպատակով դիտարկենք *խոշորացման (ագրեգացման) և ապախոշորացման (դեզագրեգացման) խնդիրը:* Պարզության համար դիտարկենք սպառողական ֆունկցիայի մոդելի հետևյալ տարբերակը՝

$$C_t = \alpha + \beta Y_t + U_t$$

որտեղ C_t-ն և Y_t-ն ներկայացնում են, համապատասխանաբար, սպառման և եկամուտի խոշորացված ցուցանիշները t տարում, իսկ U_t-ն պատահական մեծություն է: Ցույց տանք, որ այդ երկու փոփոխականների (C-ն՝ էնդոգեն, Y-ը՝ էկզոգեն) միջև կապը վերը բերված տեսքով լիարժեք չի ներկայացնում տեսական դրույթները: Օրինակ, պարզ է, որ, եթե սպառողական ապրանքների վրա ծախսերը կրճատվեն և այլ Z_t ծախսեր չկատարվեն, ապա եկամուտը պետք է նույնպես կրճատվի: Հետևաբար պետք է՝

$$Y_t = C_t + Z_t :$$

Դա նշանակում է, որ Y_t-ն նոր դերում է, որպես էնդոգեն, այլ ոչ թե էկզոգեն փոփոխական:

Այժմ տեսնենք՝ հնարավո՞ր է արդյոք Z_t-ն դիտարկել որպես էկզոգեն: Դա հնարավոր է, եթե այն կախված չլինի U_t-ից, Y_t-ից և C_t-ից: Բայց այստեղ դժվար է ընդունել այն վարկածը, որ միջոցները, որոնք չեն ծախսվում սպառողական ապրանքների վրա, կապված չեն այն փոփոխականների հետ, որոնցով որոշվում է սպառման մեծությունը: Պարզ է, որ Z_t-ն պետք է պարունակի I_t ներդրումները, որոնք հավանաբար կախված են U_t-ից: Հաշվի առնելով, որ U_t-ն պետք է արտացոլի գների ազդեցությունը (ներառյալ տոկոսադրույքը), ինչպես նաև կուտակած հարստության չափը և նրա բաշխումը: Բնական է ենթադրել, որ Z_t փոփոխականը բաղկացած է երկու բաղադրիչներից՝ I_t ներդրումներից և G_t կառավարման ծախսերից:

Այսպիսով, ստանում ենք

$$Y_t = C_t + I_t + G_t :$$

Ակնհայտ է նաև, որ I_t ներդրումները, լինելով եկամուտի մասը, պետք է ապահովեն հիմնական միջոցների այնպիսի կուտակում, որն իր հերթին ապահովելու է Y_t վերջնական արդյունքը և r_t տոկոսադրույքի՝ փոխառու միջոցների վճարումները: Հետևաբար՝

$$I_t = f(Y_t, r_t):$$

Դժվար թե տոկոսադրույքը կարելի է համարել էկզոգեն փոփոխական, այն ավելի շուտ կախված է M_t դրամական մնացորդների իրական մեծությունից: Տրամաբանական կլիներ M_t-ն համարել էկզոգեն (անկախ) փոփոխական: Այդ դեպքում կունենանք

$$M_t = g(Y_t, r_t)$$

Ամփոփելով տեսնում ենք, որ մեր սկզբնական մոդելը պարունակում է չորս հավասարում, որոնք համատեղ թույլ են տալիս որոշել Y_t, C_t, I_t և r_t փոփոխականների արժեքները, և այդ հավասարումներում օգտագործվում են միայն երկու էկզոգեն փոփոխականներ՝ M_t-ն և G_t-ն: Բոլոր փոփոխական մեծությունները հաշվարկվում են հաստատուն գների պայմաններում: Բայց այդ ենթադրությունը իրական չէ, և գների մակարդակը պետք է լինի էնդոգեն: Դա հնարավոր է, եթե ներմուծենք ևս երեք հավասարում՝ արտադրական ֆունկցիան և առնչություններ, որոնք հնարավորություն են ստեղծում ընդգրկելու տնտեսաչափական մոդելում ևս երկու էնդոգեն փոփոխականներ՝ աշխատավարձի դրույքաչափը ընթացիկ գներով՝ W_t-ն, և զբաղվածության մակարդակը՝ L_t-ն:

Տնտեսության քեյնզյան մոդելում ենթադրվում է արտադրական հզորությունների ավելցուկի առկայություն, և թողարկվող արտադրանքի ծավալը կարելի է դիտարկել որպես ֆունկցիա ամբողջական սպառումից, այլ ոչ թե արտադրական հզորություններից: (Սրանով է, որ դիտարկվող մոդելը տարբերվում է նեոդասական պատկերացումից, որի դեպքում ամբողջական արտադրանքի մեծությունը որոշվում է առկա ռեսուրսներով և տեխնիկական առաջընթացի մակարդակով): Արտադրական ֆունկցիայի միջոցով կարելի է որոշել աշխատանքի ամբողջական պահանջարկը՝ որպես ֆունկցիա տրված արտադրանքի թողարկման Y_t և K_t կուտակած հիմնական միջոցների ծավալներից, ինչպես նաև տրված (մեզ հայտնի) արտադրողականությունից և հիմնական միջոցներից (որոնք պարզության համար կարելի է համարել հաստատուն)՝

$$Y_t = F(L_t, K_t) :$$

Ուսումնասիրվող մոդելում ենթադրվում է նաև, որ տնտեսությունը գտնվում է կատարյալ մրցակցության պայմաններում և աշխատանքի սահմանային արտադրողականության և իրական P_t աշխատավարձի միջև տեղի ունի

$$\frac{\partial Y_t}{\partial L} = \frac{W_t}{P_t}$$

առնչությունը:

Վերջինս կարելի է օգտագործել գների մակարդակի որոշման համար, եթե հնարավոր լինի բացատրել նոմինալ աշխատավարձի չափերը (դասական տեսությունում այդ առնչությունը օգտագործվում է աշխատանքի պահանջարկի որոշման համար):

Դիցուք դիտարկվող t տարում այդ մակարդակը հաստատուն է, այդ դեպքում ավելացնելով ևս երկու հավասարում՝

$$P_t = P(W_t, L_t, Y_t), \quad \text{և} \quad W_t = W_0,$$

մենք կստանանք հավասարումների համակարգի վերջնական տեսքը, և հնարավոր կլինի որոշել G_t, M_t և W_0 էկզոգեն փոփոխականներից կախված յոթ փոխկապված փոփոխականների արժեքները՝ $Y_t, C_t, I_t, r_t, L_t, P_t$ և W_t:

Վերը բերված հավասարումների համակարգը, իհարկե, ավելի բովանդակալից է, քան սպառողական ֆունկցիան: Նկարագրված մոտեցումը հնարավորություն է տալիս կառուցելու տնտեսաչափական մի մոդել, որը կարելի է օգտագործել կանխագուշակման համար և որն ավելի լավ է նկարագրում իրականությունը

$$C_t = \alpha_0 + \alpha_1 Y_t + U_{1t},$$
$$I_t = \beta_0 + \beta_2 \dot{Y}_t + \beta_2 r_t + \beta_3 K_{t-1} + U_{2t},$$
$$M_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 r_t + U_{3t},$$
$$Y_t = \delta_0 L_t^{\delta_1} K_t^{\delta_2} U_{4t},$$
$$P_t = \dot{P}_t = \Theta_0 + \Theta_1 \dot{W}_t + U_{5t},$$
$$\dot{W}_t = \varphi_0 + \varphi_1 \dot{P}_1 + \varphi_2 U_t^{-1} + \varphi_3 \dot{U}_t + U_{6t},$$
$$Y_t = C_t + I_t + G_t,$$
$$U_t = N_t - L_t,$$
$$K_t = K_{t-1} + (I_t - \beta_3 K_{t-1}) = (1 - \beta_3) K_{t-1} + I_t:$$

Այստեղ նոր փոփոխականներն են. N_t-ն՝ աշխատանքի քանակը, U_t-ն՝ գործազրկության մակարդակը, K_{t-1}-ը՝ նախորդ տարում կուտակված հիմնական միջոցները:

Սակայն այս մոդելը նույնպես զերծ չէ թերություններից, օրինակ, այն չի արտացոլում արտաքին առևտուրը:

Տնտեսաչափական մոդելի մանրամասների վերաբերյալ այս դատողությունները կարելի է շարունակել, բայց մենք այսքանով կսահմանափակվենք: Պետք է նշել, որ երբ նույն խնդիրը հետազոտում են երկու տարբեր մասնագետներ, ապա, եթե նույնիսկ նրանք ունենեն նույն տեսական կարողությունները, միևնույն է՝ նրանց վերջնական արդյունքները իրարից տարբերվելու են, և ոչ թե այն պատճառով, որ կարող են տարբեր լինել հավաքված տեղեկությունները՝ նմուշը, տնտեսաչափական մոդելը, և այլն, այլ նաև այն պատճառով, որ տնտեսաչափությունը սոսկ մաթեմատիկական մոդելների և տնտեսագիտության սոցիալ-տնտեսական իրավիճակների նկարագրում չէ:

14.2. Տնտեսաչափական հետազոտությունների առանձնահատկությունները

Ռեգրեսիային հավասարման պարամետրերի գնահատումը տնտեսաչափական մոդելի կառուցման կարևոր փուլերից մեկն է:

Բազմաչափ ռեգրեսիայի տեսական հավասարումն ունի հետևյալ տեսքը՝

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_MX_M + U :$$

Փոքրագույն քառակուսիների եղանակի օգնությամբ ստանում ենք պարամետրերի գնահատականները և նմուշային ռեգրեսիայի հավասարման տեսքը (տե՛ս գլուխ 12)՝

$$\hat{Y} = \hat{b}_0 + \hat{b}_1X_1 + \hat{b}_2X_2 + \ldots + \hat{b}_MX_M :$$

Սովորաբար զուգահեռ կատարվում է բացատրող փոփոխականների և նրանց միջև գոյություն ունեցող կապերի ընտրությունը, կարևորելով գնահատված կախվածությունների բովանդակային իմաստը:

Հիշեցնենք, որ պարզ ռեգրեսիայի դեպքում որևէ ռեգրեսիայի գործակցի նշանակալիության ստուգումն իրականացվում է Ստյուդենտի $t(N-2)$ վիճականու միջոցով, իսկ բազմաչափ ռեգրեսիայի դեպքում՝ և $t(N-M-1)$ վիճականու միջոցով, որտեղ M-ը բացատրող փոփոխականների քանակն է:

Ռեգրեսիայի հավասարման ճշգրտությունը ստուգելու համար օգտագործում են R^2_{YX} **որոշակիացման (դետերմինացիայի) գործակիցը**՝

$$R^2_{YX} = \sum_{n=1}^{N}(\hat{Y}_n - \bar{Y})^2 / \sum_{n=1}^{N}(Y_n - \bar{Y})^2,$$

Հայտնի է, որ R^2_{YX} գործակիցը հավասար է բազմակի հարաբերակցության գործակցի քառակուսուն և բնութագրում է նմուշային տվյալների ցրվածության այն մասը (բաժինը), որը կապված է տվյալ ռեգրեսիային կախվածության հետ: Որոշակիացման գործակցի նշանակալիությունն ստուգելու վիճականին է Ֆիշերի վիճականին՝

$$F(\mathbf{X}, \mathbf{Y}) = [R^2_{YX}/(1 - R^2_{YX})] \cdot [(N - M - 1)/M],$$

որն ունի $\mathcal{F}(M, N - M - 1)$ բաշխում: Մասնավոր դեպքում, երբ $M = 1$, Ֆիշերի վիճականին հավասար է Ստյուդենտի վիճականու քառակուսուն՝ $F = \mathbf{t}^2$:

$F(\mathbf{X}, \mathbf{Y})$ վիճականին նաև կիրառվում է, երբ K քանակությամբ լրացուցիչ փոփոխականներ են ներմուծվում (կամ հանվում) ռեգրեսիայի հավասարում (հավասարումից):

F վիճականին օգտագործվում է ոչ միայն այն ժամանակ, երբ ստուգվում է գծային ռեգրեսիայի բոլոր գործակիցների՝ զրոյի հավասար լինելու վարկածը, այլ նաև դրանց միայն մի մասի զրոյի հավասար լինելու վարկածը: Ենթադրենք՝ հետազոտության սկզբում N դիտումների արդյունքներով քննվում է

$$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_MX_M$$

հավասարումը M բացատրող փոփոխականների համար, և որոշակիացման գործակիցն է R_1^2: Դիցուք, տնտեսագիտական վերլուծությամբ գալիս ենք եզրակացության, որ, օրինակ, վերջին K բացատրող փոփոխականները կարելի է անտեսել: Այս դեպքում նույն դիտումների արդյունքների հիման վրա գնահատվում է

$$Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_{M-K}X_{M-K}$$

հավասարումը, և որոշակիացման գործակիցը ընդունում է R_2^2 արժեքը: Պարզ է, որ $R_2^2 \leq R_1^2$: Ռեգրեսիայի նոր հավասարման համար, հեռացված բացատրող փոփոխականների զրոյի հավասար լինելու վարկածը ստուգելու նպատակով, հաշվում ենք

$$F(\mathbf{X}, \mathbf{Y}) = (R_1^2 - R_2^2)/(1 - R_1^2) \cdot (N - M - 1)/K$$

վիճականին, որն ունի Ֆիշերի $\mathcal{F}(K, N - M - 1)$ բաշխում (տե՛ս նաև գլուխ 12-ը): $F(\mathbf{X}, \mathbf{Y})$ վիճականին օգտագործվում է նաև ռեգրեսիայի հավասարման մեջ նոր (K հատ) բացատրող փոփոխականներ ընդգրկելու դեպքում:

Այդ դեպքում նույն վարկածը ստուգվում է

$$F(\mathbf{X},\mathbf{Y}) = (R_2^2 - R_1^2)/(1 - R_2^2) \cdot (N - M - K - 1)/K$$

վիճականու նկատմամբ, որն ունի $F(\mathbf{X},\mathbf{Y})(K, N-M-K-1)$ բաշխում: Եթե զրոյական վարկածը ժխտվում է, ապա նոր ներմուծված փոփոխականները չեն բացատրում կախյալ փոփոխականի տվյալների ցրվածության էական մասը, այսինքն դրանց ավելացնելը ռեգրեսիայի հավասարման մեջ չի հիմնավորվում նմուշային տվյալներով: Հասկանալի է, որ նոր փոփոխականների ընդգրկումը ռեգրեսիայի հավասարման մեջ հարմար է կատարել մեկ առ մեկ:

Այժմ անցնենք Գաուսի-Մարկովի թեորեմի ներկայացմանը: Երբ այս թեորեմի պայմանները չեն բավարարվում, հատկապես երրորդը, ապա առաջանում են որոշակի դժվարություններ և առանձնահատկություններ, որոնք, տնտեսաչափական հետազոտություններ կատարելիս, անհրաժեշտ է հաղթահարել կամ շրջանցել՝ փոքրագույն քառակուսիների եղանակի լիարժեք կիրառելիության ապահովելու համար:

Ենթադրենք տրված է N ծավալի երկչափ նմուշ՝ (X_n, Y_n), $n = \overline{1,N}$, և տեղի ունեն հետևյալ պայմանները՝

1. $Y_n = a_0 + a_1 X_n + U_n$, $n = \overline{1,N}$,

2. X_n-երը ոչ պատահական մեծություններ են և $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ վեկտորը կոլինեար չէ $e = (1, 1, \ldots, 1)$ N-չափանի վեկտորին:

3ա. $\mathbf{E}(U_n) = 0$, $\mathbf{E}(U_n^2) = \mathbf{D}(U_n) = \sigma^2$ (σ^2-ն n-ից կախված չէ):

3բ. $\mathbf{E}(U_{n_1}, U_{n_2}) = 0$, երբ $n_1 \neq n_2$:

3բ$'$. $\mathbf{E}(X_n, U_n) = 0$, $n = \overline{1,N}$,

3գ. $U_n \sim N(0, \sigma^2)$, $n = \overline{1,N}$:

Գաուսի-Մարկովի թեորեմը պնդում է, որ

> $1 - 3$ պայմանների փոքրագույն քառակուսիների եղանակով ստացված
> $$\hat{a}_1(\mathbf{X},\mathbf{Y}) = \frac{N\sum X_nY_n - \sum X_n \sum Y_n}{N\sum X_n^2 - (\sum X_n)^2} = \frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\mathbf{D}(\mathbf{X})},$$
> և
> $$\hat{a}_0(\mathbf{X},\mathbf{Y}) = \sum Y_n/N - \hat{a}_1(\mathbf{X},\mathbf{Y}) \sum X_n/N$$
> գնահատումներն անշեղ գծային գնահատումների դասում ունեն նվազագույն ցրվածք (այսինքն՝ արդյունավետ են):

Նշենք, որ Գաուսի-Մարկովի թեորեմը ճիշտ է նաև բազմաչափ դեպքում:

Տարացրիվություն: Դիտարկենք այն դեպքը, երբ Գաուսի-Մարկովի թեորեմի 3բ պայմանը տեղի չունի:

> Եթե U_n պատահական սխալները բավարարում են 3ա և 3բ պայմաններին, ապա ասում են, որ տեղի ունի **համացրիվություն**: Իսկ երբ սխալները հարաբերակցված չեն, բայց ցրվածքը կախված է դիտարկման համարից՝ $\mathbf{E}(U_n^2) = \mathbf{D}(U_n) = \sigma_n^2$, ասում ենք, որ տեղի ունի **տարացրիվություն**:

Վերջին հատկությունը տեղի է ունենում, երբ ուսումնասիրվող տնտեսական օբյեկտները *համասեռ* չեն: Օրինակ, եթե նմուշի մեջ ընդգրկված են ինչպես մեծ, այնպես էլ փոքր ձեռնարկությունների ցուցանիշները:

Քննարկենք այն դեպքը, երբ ռեգրեսիայի հավասարմանը համապատասխանող Ω **համացրվածքային** մատրիցն անկյունագծային է և հայտնի է ճնուշային **համացրվածքային** մատրիցը: Հաճախ, ելնելով խնդրի բովանդակությունից, կարելի է ընդունել, որ $\sigma_n^2 = \omega_n \cdot \sigma^2$, որտեղ σ^2-ն անհայտ է, իսկ ω_n գործակիցները հայտնի են, ընդ որում $\sum\limits_{n=1}^{N} \omega_n = N$: Երբ $\omega_n = 1$, ունենք դասական ռեգրեսիայի դեպքը, Գաուսի-Մարկովի պայմանը բավարարված է: Եթե $\omega_n \neq 1$, ապա տարացրիվության երևույթը կարելի է չեզոքացնել, դիտարկելով ռեգրեսիայի նոր հավասարում, նշանակելով $Y_n' = Y_n/\omega_n$, $U_n' = U_n/\omega_n$՝

$$Y_n' = a_0' + a_1'X + U_n':$$

Այստեղ արդեն տեսական ցրվածքը կլինի՝

$$\mathbf{E}(U_n'^2) = \mathbf{E}(U_n^2/\omega_n^2) = (1/\omega_n^2)\mathbf{E}(U_n^2) = (1/\omega_n^2) \cdot \omega_n^2\sigma_n^2 = \sigma^2,$$

այսինքն՝ ստանում ենք համացրիվության դեպքը և նոր ռեգրեսիոն մոդելի համար բավարարված է Գաուսի-Մարկովի 3ք պայմանը:

Փոփոխականներին կշիռներ վերագրելու եղանակը կիրառելի է նաև բազմաչափ ռեգրեսիայի դեպքում: Ապացուցումը թողնում ենք ընթերցողին:

14.3. Բազմակոլինեարություն

Տնտեսաչափական վերլուծության ընթացքում հանդիպում են որոշակի իրավիճակներ, երբ բացատրող փոփոխականները գտնվում են միմյանց հետ խիստ հարաբերակցման կապի մեջ: Այդ դեպքում հարաբերակցության մասնակի գործակիցները խիստ զգայուն են չափումների սխալների նկատմամբ և որպես հետևանք, հանգեցնում են ռեգրեսիայի ոչ հուսալի գնահատմանը:

Բազմակոլինեարության առկայությունը բացահայտվում է, երբ, օրինակ, ռեգրեսիայի գործակիցների ցրվածքը ընդունում է բավականին մեծ արժեքներ:

Բազմակոլինեարության երևույթը բնորոշ է ժամանակային շարքերին, երբ երկու և ավելին անկախ փոփոխականներ ունեն բացահայտ ժամանակային միտում: Զարգացվել է մի եղանակ (կոնֆլյուենտային վերլուծություն), որի էությունն է՝ բոլոր այն փոփոխականները, որոնք կարևորվում են տվյալ մոդելի համար, հաջորդաբար դիտարկվում են որպես կախյալ փոփոխական և կառուցվում է ռեգրեսիայի հավասարում, բայց և մնացած փոփոխականների և և դրանց բոլոր հնարավոր ենթաբազմությունների միջև: Հաշվարկվում է բոլոր մասնակի ռեգրեսիայի գործակիցները և դետերմինացիայի գործակիցները և դրանց վերլուծությամբ որոշվում է որ փոփոխականն է օգտակար, ավելորդ կամ խանգարող դիտարկվող տնտեսաչափական մոդելի համար: Փոփոխականը օգտակար է, եթե նրա առկայությունը շոշափելի մեծացնում է R^2-ին, եթե դա այդպես չէ և նրա ներգրավումը չի ազդում հավասարման գործակիցների վրա, ապա այդ փոփոխականը ավելորդ է: Իսկ, եթե փոփոխականի ներգրավումը հիմնարար փոխում է ռեգրեսիայի գործակիցների գնահատումները և չի լավացնում կախյալ փոփոխականի արժեքների մոտեցումը նշուշի արժեքներին, ապա այն խանգարող է:

Ստորև կբերենք գործնականում կիրառելի եղանակ, որը հնարավորություն է տալիս ձևափոխել $X_1, X_2, \ldots, X_N$ բազմակոլինեար փոփոխականները $Z_1, Z_2, \ldots, Z_N$ ոչ կոլինեար մեծությունների պահպանելով դրանց տնտեսական բովանդակությունը:

Դիցուք հայտնի է նմուշի $X_1, X_2, \ldots, X_N$ փոփոխականների նմուշային համահարաբերակցային մատրիցը և $s_n^2 = 1$ և $\mathbf{E}X_n = 0$ $n = \overline{1, N}$: Ընդունենք նաև, որ հարաբերակցության գործակիցները $\rho_{n_1n_2} \neq 0$ և $\rho'_{n_1n_2} \neq 0$ ($\rho'_{n_1n_2}$ սահմանումը բերված է ստորև):

1. $Z_1 = X_1$,

2. $\hat{X}_2 = \rho_{12}X_1,\ Z_2 = X_2 - \hat{X}_2 = X_2 - \rho_{12}X_1 = \begin{vmatrix} 1 & \rho_{12} \\ X_1 & X_2 \end{vmatrix}$,

3. $\hat{X}_3 = \rho_{13}X_1 + \rho'_{23}Z_2/s_2^2$, որտեղ

$$\rho'_{23} = \frac{\operatorname{cov}(Z_2, X_3)}{s_2 s_3} = \frac{\operatorname{cov}(X_2 - \rho_{12}X_1, X_3)}{s_2 \cdot 1} = \frac{\rho_{23} - \rho_{12}\cdot\rho_{13}}{\sqrt{1-\rho_{12}^2}},$$

$$Z_3 = X_3 - \hat{X}_3 = X_3 - \frac{\rho_{23} - \rho_{12}\rho_{13}}{s_2^2}(X_2 - \rho_{12}X_1) =$$

$$= X_3 - \frac{\rho_{12} - \rho_{12}\rho_{13}}{1-\rho_{12}^2}(X_2 - \rho_{12}X_1) =$$

$$= K_2\begin{vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ X_1 & X_2 & X_3 \end{vmatrix}, \text{ որտեղ } K_2 = \begin{vmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{vmatrix}^{-1},$$

և այլն մինչև N-րդ քայլ: Եղանակը համընկնում է Գրամի որոշիչների միջոցով փոփոխականների օրթոգոնալացման եղանակի հետ: Ստացված նոր $Z_1, Z_2, \ldots, Z_N$ փոփոխականները օրթոգոնալ են միմյանց և կրում են $X_1, X_2, \ldots, X_N$ փոփոխականների նույն տնտեսագիտական բովանդակությունը և կոլինեար չեն:

14.4. Տնտեսաչափական մոդելների կիրառման օրինակներ

Այս ենթաբաժնում կներկայացնենք տնտեսաչափական վերլուծության օրինակներ: Դրանցից առաջինը պարզաբանում է կառավարության բյուջետային քաղաքականության հնարավոր ազդեցությունը Հայաստանի տնտեսության վրա, իսկ երկրորդը վերաբերում է Մեծ Բրիտանիայի տնային տնտեսությունների եկամուտներին և խնայողություններին:

<u>*Օրինակ*</u> 1: Նկարագրենք 1997 թ. հունվար-հունիս ժամանակահատվածում Հայաստանի բնակչության սպառման մակարդակը հետևյալ բանաձևի միջոցով՝

$$C = 80 + 0.8(Y - T),$$

որտեղ C-ն սպառումն է, Y-ը համախառն ներքին արդյունքն է, T-ն հարկերից ստացված գումարն է, $(Y-T)$-ն՝ տնօրինվող եկամուտը: Այստեղ ընդունված է, որ կարճաժամկետ սպառման հակումը լրացուցիչ 100 դրամ եկամուտի համար 80 դրամ է: Ընդունենք, որ տնային տնտեսությունների և ձեռնարկությունների կողմից I ներդրումները կազմում են 10 մլրդ դրամ: Քեյնզյան տեսության համաձայն, ներդրումները նշված ժամանակաշրջանում նվազ զգայուն են կոնյունկտուրային: Այդ գումարը վտանգված չէ՝ արտադրության մակարդակի ոչ էական փոփոխություններով: Անհրաժեշտ է ուսումնասիրել, թե ինչպիսի՞ ազդեցություն կունենա կառավարության բյուջետային վարվելակերպը, որը բնութագրվում է T եկամուտի մակարդակով կամ t հարկադրույքով՝ $T = tY$, և արտադրության վրա ծախսերի մակարդակով:

Լուծում: *Նախնական իրավիճակ*: Ենթադրենք հարկադրույքի սկզբնական t չափը՝ 20% և արտադրության զարգացման հեռանկարը թույլ է տալիս կանխատեսել համախառն ներքին արդյունքի՝ $Y = 600$ մլրդ դրամ, մակարդակը: Հետևաբար կունենանք՝

$$Y = 600 \text{ (մլրդ դրամ)},\ T = \frac{20 \times 600}{100} = 120 \text{ (մլրդ դրամ)},$$

$$C = 80 + 0.8(600 - 120) = 464 \text{ (մլրդ դրամ)},\ I = 10 \text{ (մլրդ դրամ)}:$$

Ընդհանուր հավասարակշռությունից կարելի է որոշել՝ կառավարության ծախսերի G մակարդակը՝

$$Y = C + I + G,$$

$$G = 600 - 464 - 10 = 126 \text{ (մլրդ դրամ)},$$

և բյուջեի D բացը (դեֆիցիտը)՝

$$D = G - T = 126 - 120 = 6 \text{ (մլրդ դրամ)} :$$

Ենթադրենք բյուջեի բացի այդպիսի մակարդակը Ազգային ժողովի կողմից ընդունելի չէ, հետևաբար, կառավարությունը պետք է փոփոխության ենթարկի իր բյուջետային քաղաքականությունը, որպեսզի հավասարակշռի բյուջեն, հնարավորինս պահպանելով արտադրության մակարդակը՝ $Y = 600$ մլրդ դրամ ցուցանիշը: Դրա համար կառավարությանը սպասարկող տնտեսագիտական գործակալությունը պետք է կատարի որոշակի հաշվարկներ: Պետք է հետազոտել՝ կարելի՞ է արդյոք ճշգրտորեն հասնել կառավարության ծախսերի կրճատումով, թե՞ հակառակը, պետք է միաժամանակ ավելացնել եկամուտները և ծախսերը: Տեսնենք, թե ինչ մակրոտնտեսական իրավիճակի կհանգեցնի, եթե կառավարությունն ընդունի ծախսերի կրճատման ռազմավարությունը: Ենթադրենք որոշվել է կրճատել ծախսերը 10 մլրդ դրամով, միաժամանակ պահպանելով 20%-անոց հարկադրույքը: Այդ դեպքում կունենանք՝

$$Y = C + I + G,$$

$$Y = 80 + 0.8(Y - T) + 10 + 116 = 80 + 0.8(Y - 0.2Y) + 10 + 116 = 206 + 0.64Y,$$

$$Y = 572 \text{ (մլրդ դրամ)}, \ T = \frac{20 \times 572}{100} = 114.4 \text{ (մլրդ դրամ)},$$

$$G = 116 \text{ (մլրդ դրամ)}, \ D = G - T = 116 - 114.4 = 1.6 \text{ (մլրդ դրամ)} :$$

Ի՞նչ եզրահանգման եկանք. կառավարությունը, կրճատելով իր ծախսերը 10 մլրդ դրամով, բյուջեի 6 մլրդ դրամ բացը նվազեցրեց մինչև 1.6 մլրդ դրամի, բայց միաժամանակ եկամուտները կրճատվեցին $600 - 572 = 28$ մլրդ դրամով, որը կազմում է համախառն ազգային արդյունքի մոտ 3.4%-ը, իսկ դա բերում է, որպես հետևանք, գործազրկության աճի: Եզրակացությունն այսպիսին է՝ միայն ծախսերի կրճատումը բավարար լուծում չէ, և այդ ռազմավարությունը մերժելի է:

Այժմ ուսումնասիրենք կառավարության երկրորդ ռազմավարությունը՝ ավելացնել ծախսերը և միաժամանակ եկամուտները: Կրկնելով դատողությունները, եզրակացնում ենք, որ եթե 1 մլրդ դրամով մեծացնենք ծախսերը, պահպանելով արտադրության $Y = 600$ մլրդ դրամ մակարդակն անփոփոխ, ապա եկամուտը կաճի $1 : 0.8 = 1.25$ մլրդ դրամով և, որպեսզի նվազեցնենք 6 մլրդ բացը, անհրաժեշտ է ավելացնել ծախսերը՝

$$\frac{c \times \Delta G}{1 - c} = \frac{0.8 \times 6}{0.2} = 24 \text{ (մլրդ դրամ)}:$$

Համաձայն հաշվարկի, 24 մլրդ դրամ լրացուցիչ ծախսերը կբերեն $24 \times 1.25 = 30$ մլրդ դրամ եկամուտ, հետևաբար կառավարությունը կունենա՝ $G = 126 + 24 = 150$ (մլրդ դրամ), $T = 120 + 30 = 150$ (մլրդ դրամ), $C = 80 + 0.8(600 - 150) = 440$(մլրդ դրամ):

Տնտեսական հավասարակշռությունը վերականգնվեց՝ $Y = C + G + I$, այսինքն $600 = 440 + 150 + 10$, և նոր հարկի դրույքաչափը հավասար է $t = 150/600 = 0.25$ կամ 25%:

Ինչպես տեսնում ենք, երկրորդ ռազմավարությունը՝ ավելացնել ծախսերը և եկամուտները, նպաստեց բյուջեի բացի վերացմանը և բերեց տնտեսական հավասարակշռության: Հետևաբար, անհրաժեշտ է հարկադրույքը դարձնել 25% և ավելացնել պետության ծախսերը 24 մլրդ դրամով: Այս արդյունքը կարելի էր ստանալ նաև այլ եղանակով՝ $Y = C + G + I$, $Y = 90 + 0.8(1 - t)Y + G$ և, որովհետև պետք է ունենանք բյուջետային հաշվեկշիռ, ապա $G = T = t - Y$, հետևաբար

$$Y = 90 + 0.8(1 - t)Y + tY,$$

$$(1 - t)Y - 0.8(1 - t)Y = 90, \quad 0.2(1 - t)Y = 90,$$

$$t = 1 - \frac{90}{0.2 \times 600} = 0.25\% :$$

Ենթադրենք այժմ, որ կառավարությունը ընդունել է երկրորդ ռազմավարությունը՝ հարկադրույքը 25%, $Y = 600$ (մլրդ դրամ), $C = 440$ (մլրդ դրամ), $I = 10$ (մլրդ դրամ), $G = T = 150$ (մլրդ դրամ) և բյուջեն հավասարակշռված է: Ընդունենք փարկած՝ կառավարությունը եկել է այն եզրակացության, որ հնարավոր է արտադրության անկում և գործազրկություն: Ի՞նչ պետք է անի կառավարությունը, որպեսզի կանխարգելի գործազրկության սպառնալիքը: Պարզ է, որ պետք է խթանել արտադրությունը, ստեղծել նոր աշխատատեղեր: Միաժամանակ անհրաժեշտ է խթանել սպառումը, որպեսզի այն կլանի ստեղծված հավելյալ արտադրանքը:

Այդպիսով գալիս ենք հետևյալին՝ պետք է ստեղծել բյուջեի բաց, նպատակ ունենալով ստանալ արտադրության մակարդակի ավելացում՝ սկզբնական 600 մլրդ դրամի նկատմամբ: Թե ինչպիսին կլինեն մակրոցուցանիշների արժեքները, երևում է հետևյալ հաշվարկներից: Ընդունելով, որ բյուջեի բացը հավասար է 10 մլրդ դրամի: $G - T = 10$ (մլրդ դրամ), $Y = C + I + G$: Անհրաժեշտ ΔY աճը պետք է բավարարի $\Delta Y = \Delta C + \Delta G = 0.8(\Delta Y - \Delta T) + 10 + \Delta T$, բայց $\Delta T = t\Delta Y$, ապա $\Delta Y = 0.8(\Delta Y - t\Delta Y + 10 + t\Delta Y)$, $(1 - t)\Delta Y - 0.8(1 - t)\Delta Y = 10$ (մլրդ դրամ), $0.2(1 - t)\Delta Y = 10$ (մլրդ դրամ), $(1 - t)\Delta Y = 50$ (մլրդ դրամ): Եթե $t = 25\%$, ապա

$\Delta Y = 66.6$ մլրդ դրամ $= 10.1\%Y$, $\Delta T = 16.65$ (մլրդ դրամ), $\Delta G = 26.65$ (մլրդ դրամ) :

Ամփոփելով հաշվարկները, գալիս ենք հետևյալ եզրահանգման՝ 10 մլրդ դրամ բյուջեի բացը հնարավոր չէ ստանալ միայն ծախսերը 10 մլրդ դրամ ավելացնելով, որովհետև արտադրության ընդլայնումը կբերի հավելյալ ֆիսկալ եկամուտների ստեղծմանը, որի հարկման պատճառով բյուջեի բացը կնվազի: 10 մլրդ դրամ բյուջեի բացը կարելի է ստեղծել, եթե ծախսերը ավելացնենք 26.65 մլրդ դրամով, և այդ դեպքում 16.65 մլրդ դրամը կլինի բյուջեի լրացուցիչ եկամուտը:

<u>Օրինակ</u> 2: Դիտարկենք Մեծ Բրիտանիայի 1946–1963թթ. եկամուտներին (X) և խնայողություններին (Y) վերաբերող տվյալները, որոնք ներկայացված են աղյուսակում, որտեղ x_n-երը և y_n-երը 1946–1963 թթ. տվյալներն են միլիարդ ֆունտ ստեռլինգով.

տարի	եկամուտ x_n	խնայողություն y_n	${x_n}^2$	y_n^2	$x_n y_n$
1946	8.8	0.36	77.44	0.1296	3.168
1947	9.4	0.21	88.36	0.0441	3.168
1948	10.0	0.08	100	0.0064	0.8
1949	10.6	0.20	112.36	0.0400	2.12
1950	11.0	0.10	121	0.01	1.1
1951	11.9	0.12	141.61	0.0144	1.428
1952	12.7	0.41	161.29	0.1681	5.207
1953	13.5	0.50	182.25	0.25	6.75
1954	14.3	0.43	204.49	0.18.49	6.149
1955	15.5	0.59	240.25	0.3481	9.145
1956	16.7	0.90	278.89	0.81	15.03
1957	17.7	0.95	313.29	0.9025	16.815
1958	18.6	0.82	345.96	0.6724	15.252
1959	19.7	1.04	388.09	1.0815	20.488
1960	21.1	1.53	445.21	2.3409	32.283
1961	22.8	1.94	519.84	3.7636	39.9
1962	23.9	1.75	571.21	3.0625	41.825
1963	25.2	1.99	635.04	3.9601	50.148
$\sum$	283.4	13.92	4926.58	17.7892	273.914

:

Լուծում: Y և X փոփոխականների միջև պետք է լինի որոշակի հավանականային կապ՝ եկամուտի աճի դեպքում բնակչության խնայողությունները նույնպես աճում են:

Ենթադրենք, որ այդ կապը գծային է, հետևաբար տնտեսաչափական մոդելը կունենա հետևյալ տեսքը՝

$$Y_n = a_0 + a_1 X_n + U_n:$$

Գտնենք նմուշային ռեգրեսիայի գործակիցները՝ $\hat{a}_0$-ն և $\hat{a}_1$-ը: Համաձայն փոքրագույն քառակուսիների եղանակի՝

$$\mathbf{\overline{x}} = 15.74, \mathbf{\overline{y}} = 0.773,$$

$$\sum_n x_n^2 - (\sum_n x_n)^2/N = 4926.58 - (283.4)^2/18 = 464.60,$$

$$\sum_n x_n y_n - (\sum_n x_n)^2(\sum_n y_n)^2/N = 273.914 - 283.4 \times 13.92/18 = 54.751,$$

$$\sum_n y_n^2 - (\sum_n y_n)^2/N = 17.7892 - (13.92)^2/18 = 7.0244,$$

$$\hat{a}_1 = [\sum_n x_n y_n - (\sum_n x_n)^2(\sum_n y_n)^2/N]/[\sum_n x_n^2 - (\sum_n)^2/N] = 54.751/464.60 = 0.1178,$$

$$\hat{a}_0 = 0.773 - 0.1178 \times 15.74 = -1.081 :$$

Հետևաբար, տնտեսաչափական տեսական մոդելը կընդունի հետևյալ տեսքը՝

$$y = -1.081 + 0.1178x:$$

Որոշակիացման գործակիցը կլինի՝ $R^2_{YX} = 0.918$: Այն ցույց է տալիս, որ գծային կապը բավական լավ է արտահայտում որոնելի կախվածությունը, սակայն չպետք է մոռանալ, որ եկամուտների և խնայողությունների աճը տեղի է ունեցել ինչպես ընդհանուր տնտեսության զարգացման, այնպես էլ դիտարկվող ժամանակաշրջանում գների աճի շնորհիվ:

Այժմ որոշենք ռեգրեսիայի միջին քառակուսային սխալը և a_1 գործակցի 95% վստահության միջակայքը՝

$$U_n = Y_n - \hat{Y}_n, \ \sum_n U_n^2 = 0.5747, \ s^2 = \sum_n U_n^2/(N-2) = 0.03592,$$

$$s_b = \sqrt{s^2/(\sum_n x_n^2 - (\sum_n x_n)^2/N)} = 0.0088:$$

Հետևաբար՝

$$0.1178 - 2.12 \times 0.0088 \le a_1 \le 0.1178 + 2.12 \times 0.0088, \ 0.0991 \le a_1 \le 0.1365 :$$

$H_0: \ a_1 = 0$ վարկածն ստուգելու համար հաշվենք Ֆիշերի վիճականին՝

$$F(\mathbf{x}, \mathbf{y}) = 180 > F_{0.95}(1, 16) = 4.49 :$$

Այսպիսով, մենք համոզվում ենք, որ խնայողությունների և եկամուտի միջև $Y = a_0 + a_1X$ գծային կախման առկայության մասին վարկածը ընդունելի է:

Գլուխ 15
Տվյալների վերլուծության ծրագրաշարեր

Ուսումը ոչ միայն լույս է, այլ նաև ազատություն է:

Իվան Տուրգենև

Մենք պետք է ի վիճակի լինենք կարգավորելու տվյալների զանգվածներն ավելի շուտ, քան դրա կհասնենք խառը:

Ջոն Գրիֆիթս

15.1. Վիճակագրական ծրագրաշարերի տեսակները

Ժամանակակից հաշվարկային միջոցների և դրանց հնարավորությունների բազմազանության բուռն կատարելագործումը հնարավոր է դարձնում տարբեր գործնական և հետազոտական աշխատանքներում տվյալների վերլուծության գործում վիճակագրական ծրագրաշարերի լայն կիրառումը:

Տվյալների վերլուծության միջոցների մեծ պահանջարկը զարգացրել է վիճակագրական ծրագրաշարերի շուկան: Արևմուտքում տարածում գտած ծրագրաշարերի թիվը հասնում է հազարի: Լինելով տարբեր իրենց ծավալով և ընտրված եղանակների որակով, հնարավոր կիրառության ոլորտներով և գնով՝ դրանք բավարարում են մարդկային գործունեության տարբեր ոլորտներում տվյալների մշակման զանազան պահանջները: Լայն տարածում են գտել STATGRAPHICS, SYSTAT, BMDP, SPSS, SAS, CSS, STATISTICA և նորագույն S-PLUS ծրագրաշարերը:

Սույն գլխում նկարագրվում են ամենահայտնի (առաջին տասնյակից) STATGRAPHICS, STATISTICA, SPSS և S-PLUS ծրագրաշարերի հնարավորությունները, դիտարկվում են դրանց աշխատանքի օրինակները, քննարկվում են մի քանի այլ ծրագրաշարերի հնարավորությունները:

Ժամանակակից մաթեմատիկական և վիճակագրական ծրագրաշարերը դասակարգվում են 4 խմբի.

- ընդհանուր նշանակության,
- մասնագիտացված, մեթոդաուղղված,
- առարկայա- (կամ խնդրա-) ուղղված,
- ուսումնական:

Մասնագիտացված ծրագրերը սովորաբար պարունակում են եղանակներ վիճակագրության մի քանի բաժիններից: Ամենից հաճախ հանդիպում են ժամանակային շարքերի վերլուծման (օրինակ МЕЗОЗАВР, TREND), ռեգրեսիոն վերլուծության ծրագրերը: Այդ ծրագրերը պարունակում են իրենց բնագավառների տարածում գտած եղանակների մի որոշակի խումբ, երբեմն էլ ներառում են յուրօրինակ եղանակներ:

Առարկայաուղղված ծրագրաշարերը նախատեսված են որոշակի կիրառական ոլորտի խնդիրների լուծման համար, օրինակ՝ արտադրանքի որակի ապահովման, ապահովագրական խնդիրների հաշվարկի, բանկային գործընթացների:

Հատուկ հետաքրքրություն են ներկայացնում միավորված մեթոդաուղղված կամ ընդհանուր նշանակության ծրագրաշարերը: Այս ծրագրաշարերի համապարփակվածությունը հնարավորություն է տալիս դրանք կիրառել և՛ մշակման սկզբնական փուլում, և՛ վիճակագրական պարզագույն եղանակների շրջանակներից դուրս լինելու դեպքերում, ինչպես նաև վիճակագրության հիմունքների ուսուցման գործընթացում:

STATGRAPHICS, STATISTICA և S-PLUS համակարգերը ընդհանուր նշանակության վիճակագրական ծրագրաշարեր են, որոնց մասին հետագայում կխոսվի ավելի մանրամասն:

STATGRAPHICS-ի առանձնահատկություններից հատկապես պետք է նշել լավ մտածված աշխատանքային դաշտը: Տարածված են այս համակարգի 7-րդ տարբերակի երեք տարատեսակներ, որոնցից երրորդը (STATGRAPHICS Plus for Windows) գործում է Windows միջավայրում:

STATISTICA բազմակողմանի վիճակագրական ծրագրաշարը արտադրվել է ամերիկյան StatSoft ընկերության կողմից: Այն ստեղծվել է 90-ական թվականների սկզբներին՝ անմիջապես Windows միջավայրի համար: Ծրագրաշարում իրենց արտացոլումն են գտել տեսական և կիրառական վիճակագրության վերջին նորույթներից շատերը:

Ընդհանուր նշանակության ծրագրաշարերից են նաև SYSTAT, SPSS և SAS համակարգերը:

SYSTAT-ը բազմակողմանի վիճակագրական ծրագրաշար է: Այն հարուստ է վիճակագրական եղանակներով և առանձնանում է հիանալի գրաֆիկական հնարավորություններով:

SPSS-ը մոդուլային ծրագիր է: Համակարգը օգտագործողին տալիս է տվյալների փոփոխման հնարավորություններ, ընդգրկված են նկարագրական վիճակագրությունը, ցրվածքային վերլուծությունը, հարաբերակցային վերլուծությունը, ֆայլերի հետ աշխատելու, գծապատկերներ և գրաֆիկներ կառուցելու և հաշվետվությունների պատրաստման միջոցներ: Ծրագրաշարի հավելյալ բաժինները ներառում են աղյուսակների վերլուծության և կառուցման, ժամանակային շարքերի վերլուծության, դասակարգման, խորացված և ընդլայնված վիճակագրական վերլուծության եղանակներ և այլն:

Ամերիկյան Visual Numerics հիմնարկության նոր մշակումներից լայն ճանաչում են ստացել IMSL, PV WAVE 6.2 և մի քանի ուրիշ համակարգեր:

СТАТ-ДИАЛОГ կենտրոնի (Մոսկվա) կողմից մշակված ծրագրաշարերից է ընդհանուր նշանակության STADIA ծրագրաշարը, որը պարունակում է վիճակագրական տվյալների վերլուծության եղանակների հարուստ ընտրանի և միևնույն ժամանակ մատչելի է մասնագետների, գործավարների և ուսանողների լայն շրջաններին: Ծրագրաշարի վերջին տարբերակն ունի գրաֆիկական բազմազան հնարավորություններ:

Հաջորդ երեք ծրագրաշարերը նույնպես ստեղծվել են СТАТ-ДИАЛОГ մասնագիտացված կենտրոնում:

МЕЗОЗАВР ծրագրաշարը նախատեսված է ժամանակային շարքերի վերլուծության համար և պարունակում է ողորկացման, սեզոնային տատանումների առանձնացման, սպեկտրային վերլուծության, մասնակի զտման եղանակներ, ինչպես նաև միտումի գծային և ոչ գծային մոդելներ, Բոքսի-Ջենկինսի մոդելներ և այլն:

КЛАС-МАСТЕР ծրագրաշարը նախատեսված է քանակական, որակական և տրամաբանական («այո-ոչ» տեսակի) տվյալների դասակարգման խնդիրների լուծման համար:

САНИ ծրագրաշարը նախատեսված է բազմաբնույթ, այդ թվում ոչ թվային տվյալների վերլուծության և ներկայացման համար: Այն հնարավորություն է տալիս տվյալները ներկայացնել մատչելի տեսքով, կառուցել խմբավորումներ, ստուգել անկախության վարկածները և այլն:

BMDP-ն մասնագիտացված, եղանակաուղղված ծրագրաշար է: Այն իրագործում է հատուկ բնագավառների և հատկապես բժշկակենսաբանական հետազոտությունների ...եսիոն կախվածությունների կառուցում և վերլուծություն:

GLIM ծրագրաշարը կարելի է դասել խնդրաուղղված համակարգերի շարքին: Նրանում առավել ամբողջականորեն իրագործված են կյանքի տևողության վերլուծության եղանակները: Դրա անհրաժեշտությունն առաջանում է արդյունաբերության մեջ՝ հուսալիության փորձարկման ժամանակ, մարքեթինգային, ապահովագրական և բժշկական հետազոտություններում:

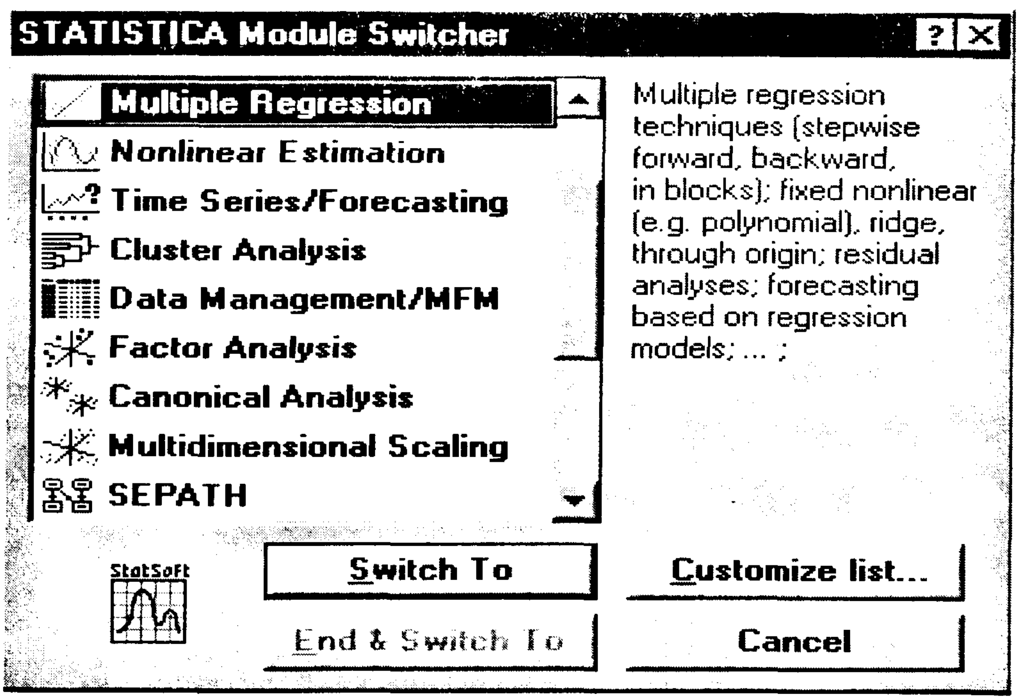
S-PLUS 2000 նորագույն վիճակագրական ծրագրաշարն ունի վերոհիշյալ ծրագրաշարերի գրեթե բոլոր առավելությունները և միևնույն ժամանակ համալրված է հաշվարկային միջոցների ժամանակակից հնարավորություններով: Այն բազմակողմանի ծրագիր է՝ նախատեսված տվյալների վերլուծության և մատչելի ներկայացման համար: Microsoft Office ծրագրաշարին համանման արտաքին ձևավորումը, աշխատանքային դաշտը և գործիքների ցանկը կարելի է վերափոխել ըստ աշխատողի ճաշակի և հարմարության:

15.2. STATISTICA ծրագրաշարը

Ընդհանուր նշանակության ժամանակակից վիճակագրական համակարգերը կազմված են տարբեր բաժիններից: Յուրաքանչյուր բաժին միմյանց հետ տրամաբանորեն կապված վիճակագրական եղանակները միավորող բարձր մակարդակի լեզվով գրված ծրագիր է: Այն ապահովում է տվյալների որոշակի մշակումը՝ առանց այլ բաժինների դիմելու: Հնարավորություն կա նաև աշխատել միաժամանակ մի քանի բաժինների հետ:

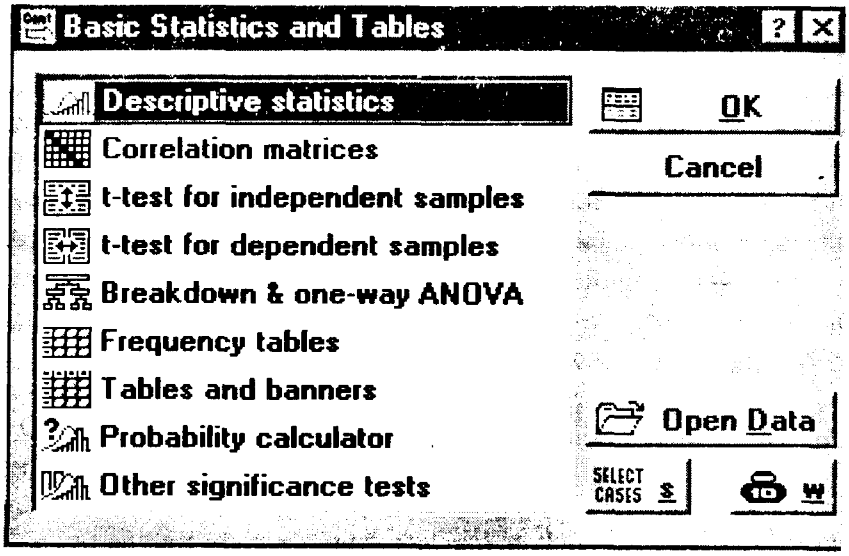
Վերոհիշյալ երկու ծրագրաշարերն էլ պարունակում են ինչպես զանազան վիճակագրական եղանակներ, որոնք թույլ են տալիս լուծել գործնականում կամայական խնդիր, այնպես էլ գրաֆիկների և հաշվետվությունների պատրաստման տարբեր համակարգեր:

Բացի այդ STATISTICA-ում կա ավտոմատորեն հաշվետվություններ կազմելու հնարավորություն: Ծրագրաշարը ներառում է SCL (Statistica Command Language) ներկառուցված լեզուն, որի օգնությամբ տվյալների մշակումը կարելի է կատարել փաթեթային եղանակով և STATISTICA BASIC լեզուն, որն աշխատողին հնարավորություն է տալիս գրել մասնավոր ծրագրեր:



Նկար 1: STATISTICA ծրագրաշարի բաժինների մեկնարկիչը:

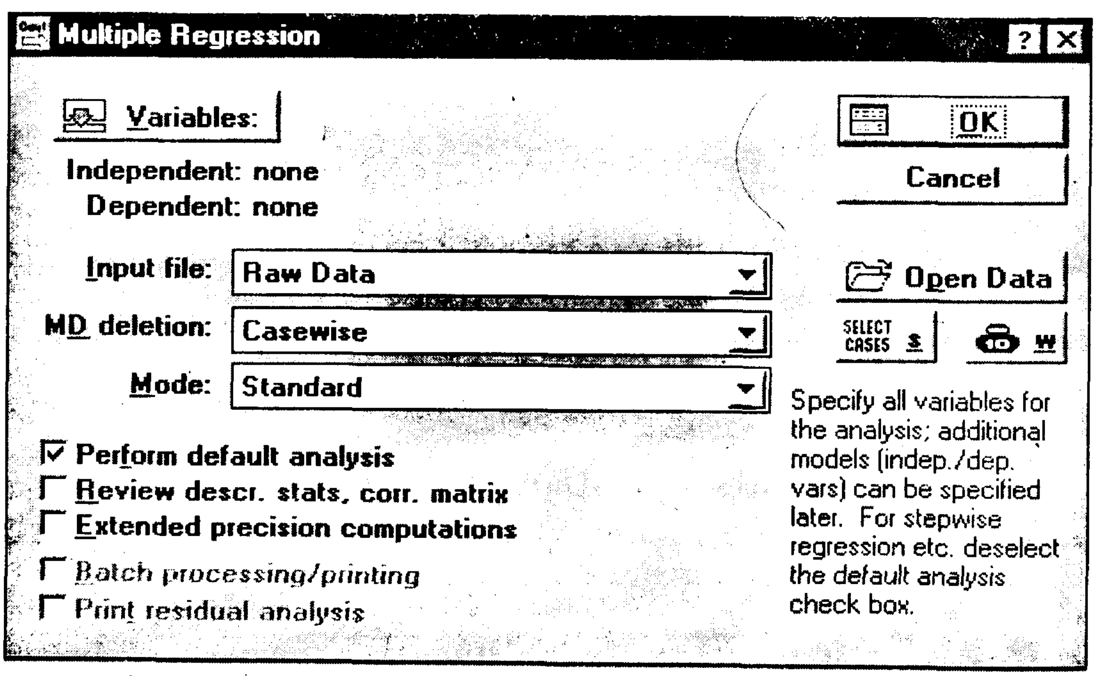
STATISTICA ծրագրաշարի գործարկումից հետո էկրանին հայտնվում է բաժինների մեկնարկիչը (Module Switcher), որի միջոցով կարելի է ընտրել անհրաժեշտ բաժինը:



Նկար 2: «Հիմնական վիճականիներ/աղյուսակներ» բաժնի աշխատանքային պատուհանը:

Համակարգում աշխատանքը սովորաբար սկսվում է «Հիմնական վիճականիներ/աղյուսակներ» բաժնից (Basic Statistics/Tables): Այն նախատեսված է տվյալների նախնական մշակման, հետազոտական վերլուծության, տվյալները տարբեր ձևերով խմբավորելու, այդ խմբերը ցուցադրելու և դրանց միջև փոխադարձ կապն ապահովելու համար:

Բոլոր նկարագրող բնութագրիչները, ինչպիսիք են՝ միջինը, միջին քառակուսային շեղումը, կիսողը, մոդը, կուտակվածության և անհամաչափության գործակիցները, կարող են հաշվարկվել ըստ մեկ կամ մի քանի փոփոխականների խմբավորված տվյալների համար, օրինակ՝ ըստ սեռի և տարիքի:

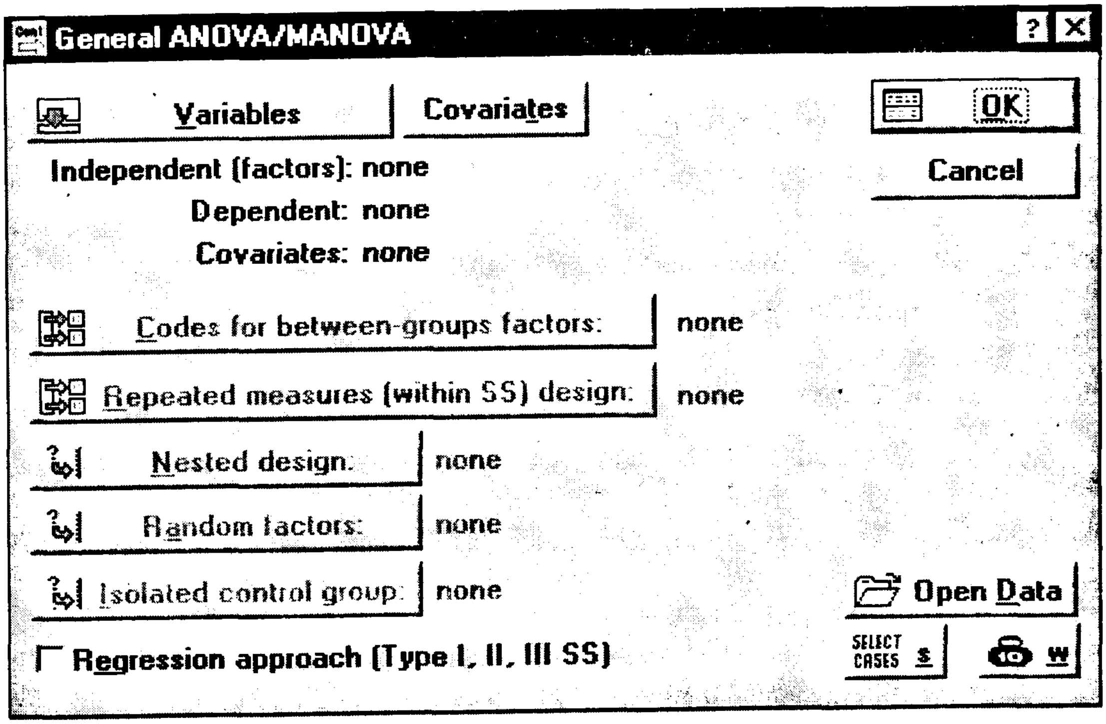


Նկար 3: «Բազմակի ռեգրեսիա» բաժնի աշխատանքային պատուհանը:

Այդ բաժնի «հարաբերակցություն» մասը հնարավորություն է տալիս պատկերացում կազմել փոփոխականների միջև կախվածության մասին: Հարաբերակցային մատրիցը դիտողական դարձնելու համար օգտագործվում են զանազան գրաֆիկական եղանակներ (օրինակ՝ շրջանակային հարթ գծապատկերներ և այլն):

«Հիմնական վիճականիներ/աղյուսակներ» բաժնի սկզբնական պատուհանում առկա է «հավանականային հաշվարկիչ» (Probably Calculator) ենթածրագիրը: Այն սովորաբար փոխարինում է դասագրքերում և մաթեմատիկական վիճակագրության տեղեկատուներում ընդգրկված աղյուսակներին:

«Բազմակի ռեգրեսիա» (Multiple-Regression) բաժինը հնարավորություն է տալիս կառուցել բազմակի փոփոխականների կախվածության մոդելներ և գնահատել դրանց համապատասխանությունը: Բաժինն ունի բազմակի գծային և ոչ գծային ռեգրեսիայի մոդելների հարուստ խումբ (բազմանդամային, ցուցչային, լոգարիթմական և այլն):



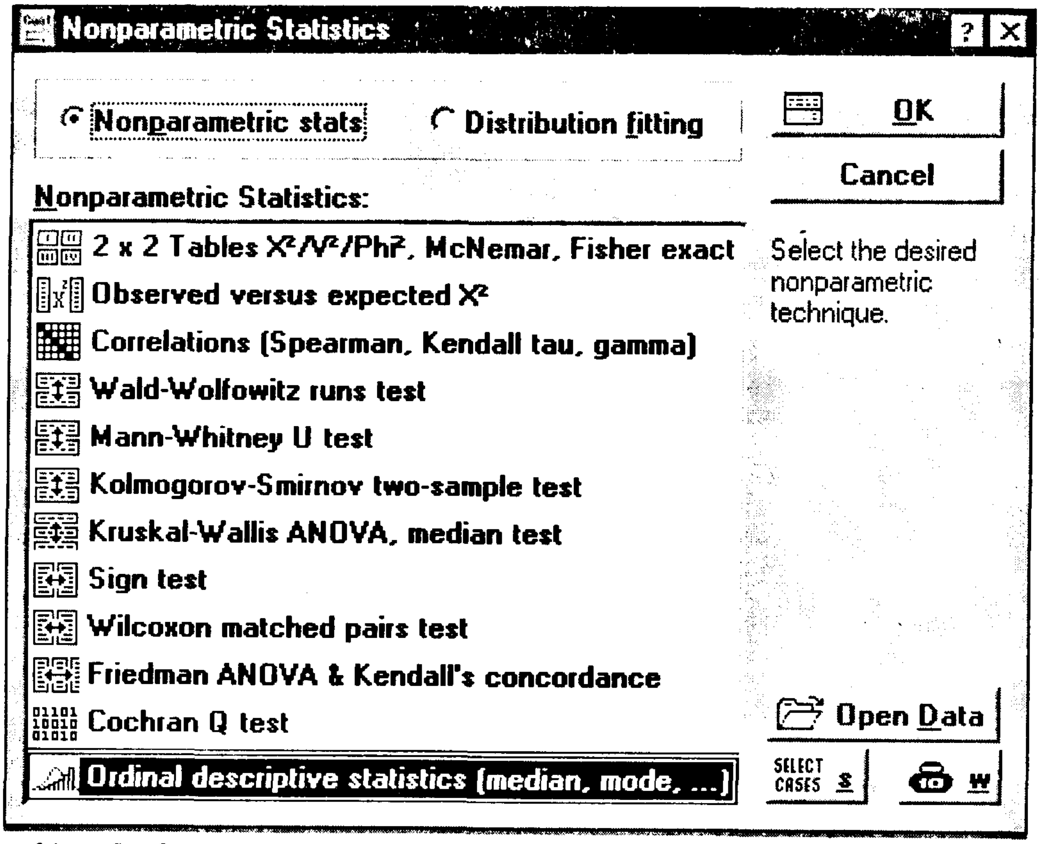
Նկար 4: «Ցրվածքային վերլուծություն» բաժնի աշխատանքային պատուհանը:

«Ցրվածքային վերլուծություն» (ANOVA/MANOVA) բաժինը միաչափ և բազմաչափ, ցրվածքային և հարաբերակցային վերլուծության ընդհանուր խումբ է: Բաժնում առկա են ցրվածքային վերլուծության հիմնական վարկածների ստուգման բազմաթիվ վիճակագրական ընթացակարգեր:

«Ոչ պարամետրական վիճակագրություն/Բաշխում» (Nonparametrics/Distribution) բաժինը պարունակում է համաձայնության հայտանիշների ընդգրկուն խումբ, մասնավորապես՝ Կոլմոգորով-Սմիրնովի հայտանիշը և տարբեր տարակարգային հայտանիշներ:

«Ժամանակային շարքերի վերլուծության / կանխագուշակման» (Time Series / Forecasting) բաժինը բաղկացած է ժամանակային շարքերի գրաֆիկական պատկերման, հարթեցնող և մոդելավորող ձևափոխությունների մի քանի ընդհանուր ծրագրերից:

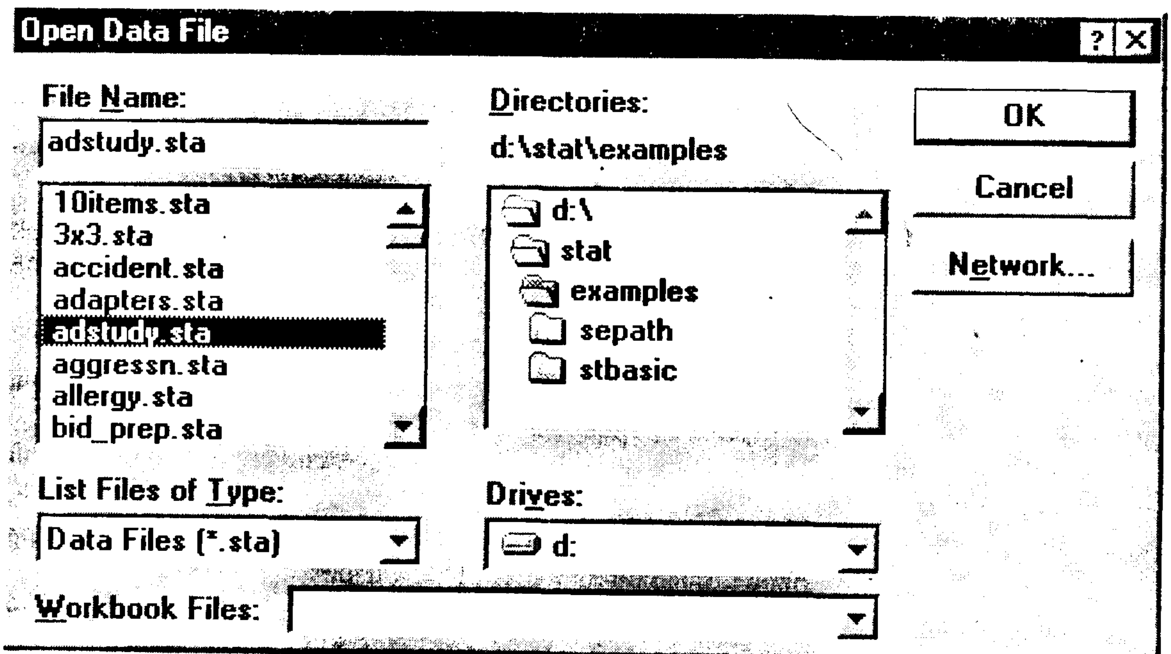
Տվյալների ներածումը STATISTICA համակարգում կազմակերպվում է էլեկտրոնային աղյուսակների տեսքով (Spreadsheet): Դրանք կարող են պարունակել ինչպես թվային, այնպես էլ տեքստային տեղեկություններ: Էլեկտրոնային աղյուսակներում տվյալները կարող են լինել տարբեր տեսքի, օրինակ՝ օրացուցային, ժամային, դրամական և այլն: Տվյալները կարելի է ներմուծել նաև անմիջապես ստեղնաշարից: Նոր տվյալները հաշվարկվում են էլեկտրոնային աղյուսակում տրված բանաձևերի օգնությամբ:



Նկար 5: «Ոչ պարամետրական վիճակագրություն/Բաշխում» բաժնի պատուհանը:

Տվյալների ֆայլը բացելու համար հարկավոր է գործարկել File հրամանացանկի «Բացել տվյալները» (Open Data) հրամանը (տե՛ս նկար 6):

Հնարավոր է օգտագործել նաև ուրիշ հավելվածներում պատրաստված տվյալները՝ կիրառելով պատճենման գործողությունը, ինչպես նաև՝ DDE դինամիկ կապը STATISTICA -ի և Windows-ի այլ ծրագրերի միջև:



6: adstudy.sta ֆայլի բացումը Open Data File պատուհանից STATISTICA ծրագրաշարում:

•

STATISTICA ծրագրաշարում նախնական տվյալների պատկերավոր ներկայացման, հետազոտական վերլուծության համար առկա են բազմատեսակ գծապատկերներ ստեղծելու հարյուրավոր միջոցներ: Ծրագրաշարը թույլ է տալիս տարբեր կոորդինատային համակարգերում կառուցել մեծ քանակության երկչափ և եռաչափ գծապատկերներ, ինչպես նաև վիճակագրական մասնագիտացված գրաֆիկներ, այունապատկերներ և այլն:

Համակարգի գրաֆիկական միջոցները մատչելի են կամայական բաժնում և վիճակագրական վերլուծության կամայական փուլում: Դրանք կարելի է օգտագործել հետևյալ նպատակներով.

- թվային և տեքստային արժեքների ցուցադրում անմիջապես ծրագրաշարի նախնական տվյալների էլեկտրոնային աղյուսակից կամ վերլուծության արդյունքների աղյուսակից (Scrollsheet):

- վիճակագրական վերլուծության արդյունքների արտածում գծապատկերների հաջորդականության տեսքով: Վիճակագրական բոլոր եղանակների պատուհաններում հնարավորություն է տրվում կառուցել պահանջվող վերլուծության գծապատկերը:

STATISTICA համակարգում ներառված են տվյալների գրաֆիկական վերլուծության համար անհրաժեշտ ինտերակտիվ գործիքներ: Դրանց օգնությամբ կարելի է կառուցված գծապատկերի վրա նշել որոշ կետեր և շարունակել վիճակագրական վերլուծությունն առանց այդ կետերի համապատասխան արժեքների: Նշված կետերի համապատասխան թվային արժեքները կարող են գրանցվել գրաֆիկի հետ կապված և *գրաֆիկի տվյալների խմբագրի* կողմից դիտարկված հատուկ էլեկտրոնային աղյուսակում:

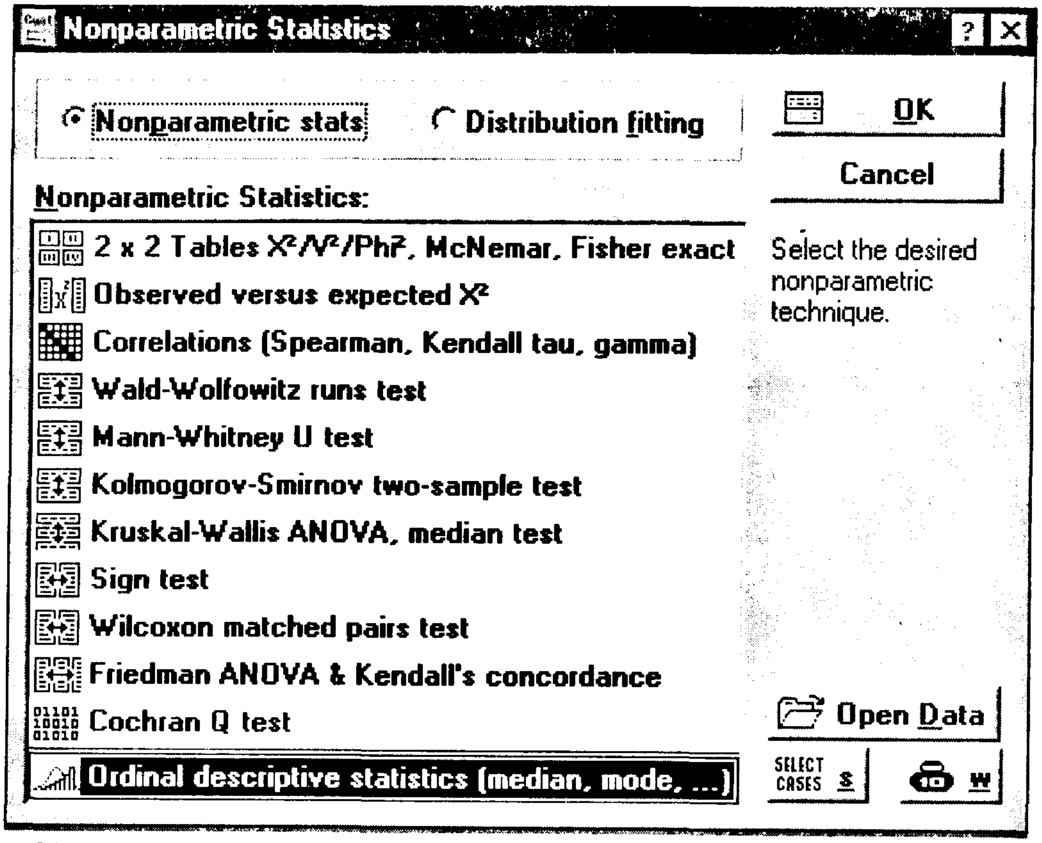
Ծրագրաշարն ունի հարմար գործիքներ՝ մի քանի գրաֆիկներ և փաստաթղթեր մի պատուհանում տեղադրելու համար: Այդ գործիքների օգնությամբ կարելի է հեշտությամբ միավորել բարդ գրաֆիկական, տեքստային և թվային տեղեկությունները: Դեռ ավելին, STATISTICA-ն ամբողջությամբ օժանդակում է OLE-Օբյեկտների կապակցման և ներդրման տեխնոլոգիային, որի օգնությամբ հնարավոր է տարբեր ծրագրերից փաստաթղթեր կապակցել և ներդնել STATISTICA-ի գրաֆիկական փաստաթղթերում և հակառակը՝ STATISTICA-ի գծապատկերները տեղադրել Windows-ի այլ ծրագրերում:

STATISTICA համակարգի գրաֆիկական հնարավորությունների մանրակրկիտ բացատրությունը տրված է [ ] գրքի գլուխ 5-ում: Պարզագույն գրաֆիկների կառուցման օրինակները տրված են հաջորդ ենթաբաժնում:

STATISTICA *ծրագրաշարի օգտագործման օրինակ*: Դիտարկենք STATISTICA ծրագրաշարի «Հավանականային հաշվարկիչը» (Probability calculator), որը գտնվում է «Հիմնական վիճականիներ/աղյուսակներ» (Basic Statistics/Tables) սկզբնական պատուհանում:

Հաշվարկիչն աշխատեցնելու արդյունքում բացվում է «Հավանականային բաշխման հաշվարկիչ» (Probability Distribution Calculator) պատուհանը, որի ձախ կողմում գտնվող ցուցակից կարելի է ընտրել հաճախակի հանդիպող բաշխումներ՝ Բետա, Կոշու, χ^2, նորմալ, լոգարիթմորեն նորմալ, Ստյուդենտի և այլն: Պատուհանում տրվում են ընտրված բաշխման պարամետրերը: Օրինակ նորմալ բաշխման համար՝ միջինը և միջին քառակուսային շեղումը, χ^2-ու համար՝ ազատության աստիճանը և այլն: p տողում տրվում է հավանականությունը, Z տողում հաշվվում է ընտրված բաշխման համապատասխան p-քանորդիչը: Նույնը կարելի է կատարել հակադարձ կարգով. Z տրված արժեքի միջոցով գտնել p համապատասխան հավանականությունը:

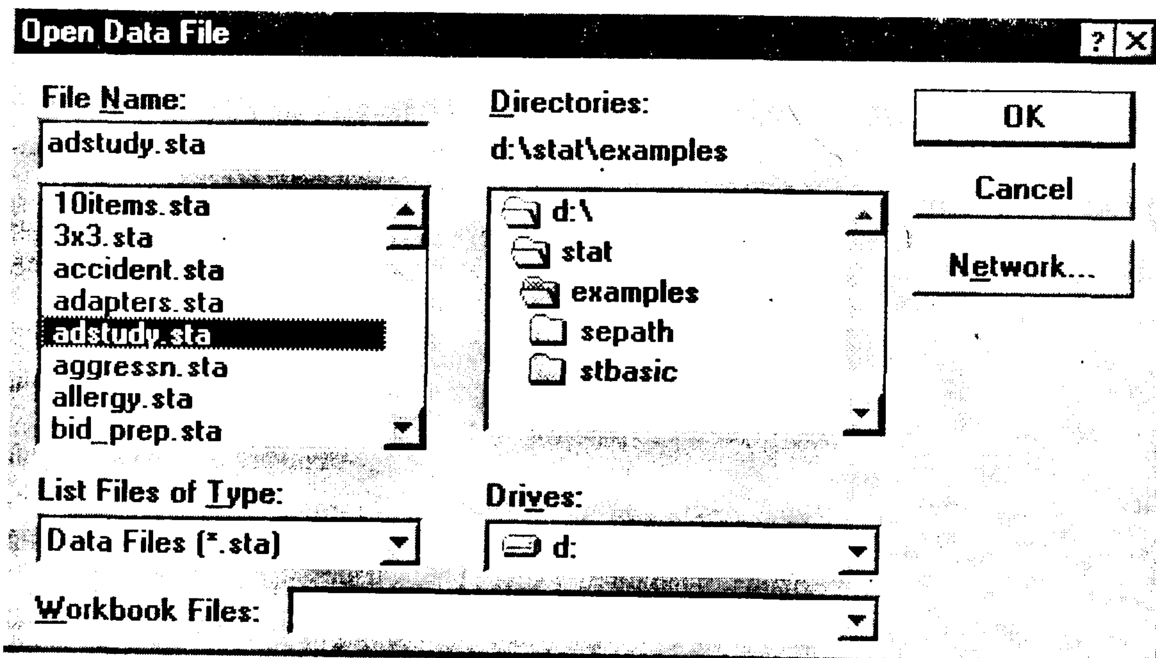
Հաշվարկիչը հնարավորություն է տալիս հաշվել բաշխման հակադարձ ֆունկցիան



Նկար 5: «Ոչ պարամետրական վիճակագրություն/Բաշխում» բաժնի պատուհանը:

Տվյալների ֆայլը բացելու համար հարկավոր է գործարկել File հրամանացանկի «Բացել տվյալները» (Open Data) հրամանը (տե՛ս նկար 6):

Հնարավոր է օգտագործել նաև ուրիշ հավելվածներում պատրաստված տվյալները՝ կիրառելով պատճենման գործողությունը, ինչպես նաև՝ DDE դինամիկ կապը STATISTICA -ի և Windows-ի այլ ծրագրերի միջև:



Նկար 6: adstudy.sta ֆայլի բացումը Open Data File պատուհանից STATISTICA ծրագրաշարում:

•

STATISTICA ծրագրաշարում նախնական տվյալների պատկերավոր ներկայացման, հետազոտական վերլուծության համար առկա են բազմատեսակ գծապատկերներ ստեղծելու հարյուրավոր միջոցներ: Ծրագրաշարը թույլ է տալիս տարբեր կոորդինատային համակարգերում կառուցել մեծ քանակության երկչափ և եռաչափ գծապատկերներ, ինչպես նաև վիճակագրական մասնագիտացված գրաֆիկներ, սյունապատկերներ և այլն:

Համակարգի գրաֆիկական միջոցները մատչելի են կամայական բաժնում և վիճակագրական վերլուծության կամայական փուլում: Դրանք կարելի է օգտագործել հետևյալ նպատակներով.

- թվային և տեքստային արժեքների ցուցադրում անմիջապես ծրագրաշարի նախնական տվյալների էլեկտրոնային աղյուսակից կամ վերլուծության արդյունքների աղյուսակից (Scrollsheet):

- վիճակագրական վերլուծության արդյունքների արտածում գծապատկերների հաջորդականության տեսքով: Վիճակագրական բոլոր եղանակների պատուհաններում հնարավորություն է տրվում կառուցել պահանջվող վերլուծության գծապատկերը:

STATISTICA համակարգում ներառված են տվյալների գրաֆիկական վերլուծության համար անհրաժեշտ ինտերակտիվ գործիքներ: Դրանց օգնությամբ կարելի է կառուցված գծապատկերի վրա նշել որոշ կետեր և շարունակել վիճակագրական վերլուծությունն առանց այդ կետերի համապատասխան արժեքների: Նշված կետերի համապատասխան թվային արժեքները կարող են գրանցվել գրաֆիկի հետ կապված և *գրաֆիկի տվյալների խմբագրի* կողմից դիտարկված հատուկ էլեկտրոնային աղյուսակում:

Ծրագրաշարն ունի հարմար գործիքներ՝ մի քանի գրաֆիկներ և փաստաթղթեր մի պատուհանում տեղադրելու համար: Այդ գործիքների օգնությամբ կարելի է հեշտությամբ միավորել բարդ գրաֆիկական, տեքստային և թվային տեղեկությունները: Դեռ ավելին, STATISTICA-ն ամբողջությամբ օժանդակում է OLE-Օբյեկտների կապակցման և ներդրման տեխնոլոգիային, որի օգնությամբ հնարավոր է տարբեր ծրագրերից փաստաթղթեր կապակցել և ներդնել STATISTICA-ի գրաֆիկական փաստաթղթերում և հակառակը՝ STATISTICA-ի գծապատկերները տեղադրել Windows-ի այլ ծրագրերում:

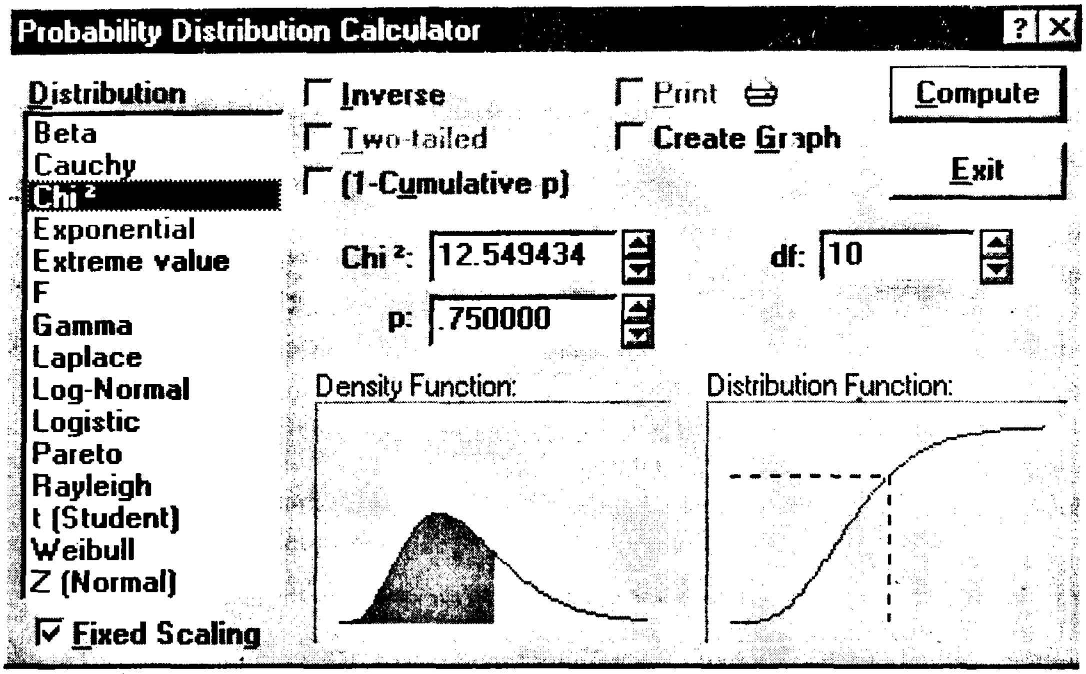
STATISTICA համակարգի գրաֆիկական հնարավորությունների մանրակրկիտ բացատրությունը տրված է [ ] գրքի գլուխ 5-ում: Պարզագույն գրաֆիկների կառուցման օրինակները տրված են հաջորդ ենթաբաժնում:

STATISTICA *ծրագրաշարի օգտագործման օրինակ*: Դիտարկենք STATISTICA ծրագրաշարի «Հավանականային հաշվարկիչը» (Probability calculator), որը գտնվում է «Հիմնական վիճականիներ/աղյուսակներ» (Basic Statistics/Tables) սկզբնական պատուհանում:

Հաշվարկիչն աշխատեցնելու արդյունքում բացվում է «Հավանականային բաշխման հաշվարկիչ» (Probability Distribution Calculator) պատուհանը, որի ձախ կողմում գտնվող ցուցակից կարելի է ընտրել հաճախակի հանդիպող բաշխումներ՝ Բետա, Կոշու, χ^2, նորմալ, լոգարիթմորեն նորմալ, Ստյուդենտի և այլն: Պատուհանում տրվում են ընտրված բաշխման պարամետրերը: Օրինակ նորմալ բաշխման համար՝ միջինը և միջին քառակուսային շեղումը, χ^2-ու համար՝ ազատության աստիճանը և այլն: p տողում տրվում է հավանականությունը, Z տողում հաշվվում է ընտրված բաշխման համապատասխան p-քանորդիչը: Նույնը կարելի է կատարել հակադարձ կարգով. Z տրված արժեքի միջոցով գտնել p համապատասխան հավանականությունը:

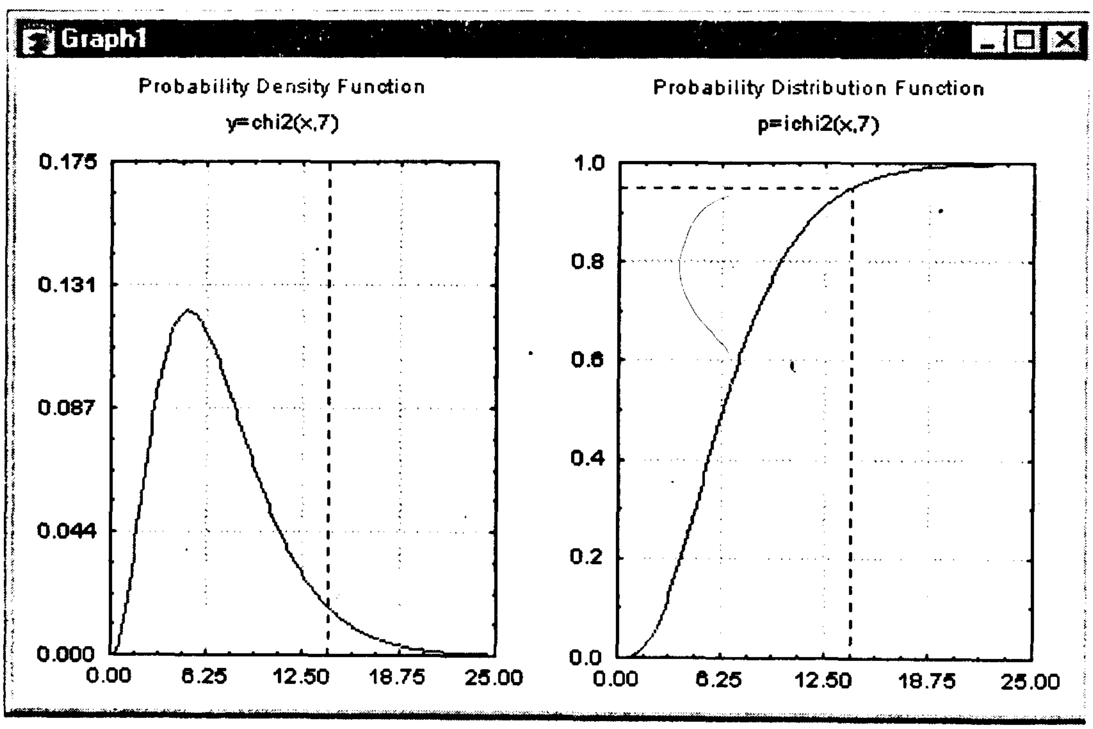
Հաշվարկիչը հնարավորություն է տալիս հաշվել բաշխման հակադարձ ֆունկցիան

(Inverse), ստեղծել գծապատկերը (Create graph) և տպել (Print) արդյունքները:



Նկար 7: «Հավանականային բաշխման հաշվարկիչի» նախնական պատուհանը:

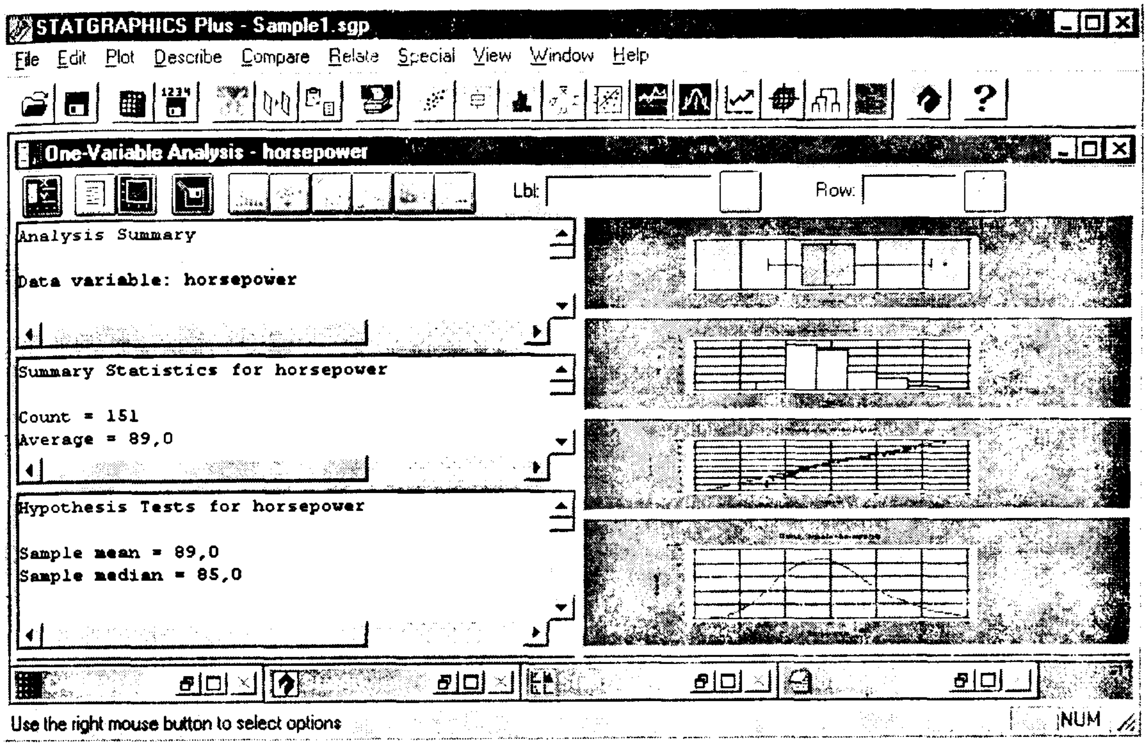
Օրինակ ընտրենք բաշխումների ցանկից χ^2 բաշխումը, *df* տողում ներածենք ազատության աստիճանը 7, վերցնենք $p = 0.95$: Հաշվման արդյունքում (Compute) կստվի 7 ազատության աստիճանով χ^2 բաշխման 0.95 քանորդիչը. ChI : 14.068419



Նկար 8: 7 ազատության աստիճանով χ^2 խտության և բաշխման ֆունկցիաների գծապատկերները:

15.3. STATGRAPHICS ծրագրաշարը

STATGRAPHICS ծրագրաշարի հիմքում ընկած է լայն տարածում գտած հրամանացանկերի համակարգը: Ծրագրաշարի հիմնական աշխատանքային դաշտի վերին մասում գտնվում են հաճախակի օգտագործվող հրամանների սեղմակները: Դրանց օգնությամբ հեշտորեն կարելի է օգտագործել ծրագրի հիմնական հնարավորություններն ու միջոցները:



Նկար 9: STATGRAPHICS ծրագրաշարի աշխատանքային դաշտը:

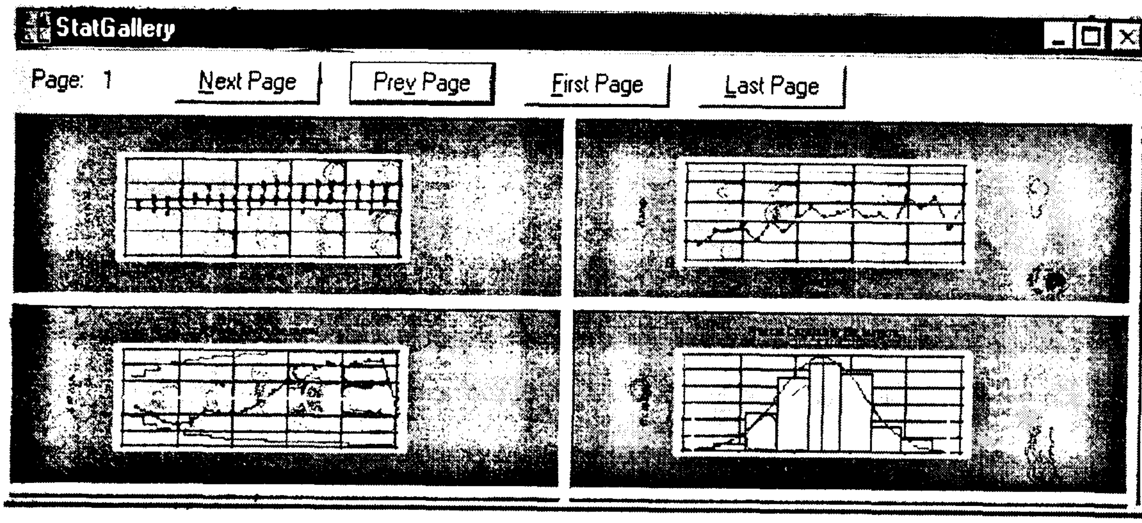
Ծրագրաշարն ունի հարուստ գրաֆիկական հնարավորություններ: Նախատեսված են ցրվածքի երկչափ և եռաչափ, մեկ կամ մի քանի փոփոխականների ֆունկցիոնալ, սյունակային, շրջանային և մատրիցային գրաֆիկներ: Բացի դրանից, ծրագրաշարն իրականացնում է բաշխման ֆունկցիաների և դրանց ածանցյալների գծապատկերների հարթեցում և կառուցում հավանականային ընտանիքների լայն դասի համար:

Ի տարբերություն STATISTICA համակարգի, տվյալների մուտքագրումը STATGRAPHICS-ի գրեթե բոլոր բաժիններում իրագործվում է ընտրված եղանակի տվյալների դաշտերի լրացման միջոցով: Դա հաճախ պահանջում է վիճակագրական հատկությունների և ծրագրաշարում ընդունված կանոնների լավ իմացություն:

Մուտքագրման դաշտերում կարող են տրվել ինչպես տվյալների անմիջական արժեքները (ոչ մեծ օբյեկտների դեպքում), այնպես էլ դրանք ներկայացնող փոփոխականների անուններն ու արտահայտություններ:

STATGRAPHICS ծրագրաշարի հավանականային հաշվարկիչի դերը կատարում են «Բաշխման ֆունկցիաներ» (DISTRIBUTION FUNCTIONS) բաժնի «Բաշխման պատկերում» (Distribution Plotting) և «Կրիտիկական արժեքներ» (Critical Values) ենթածրագրերը:

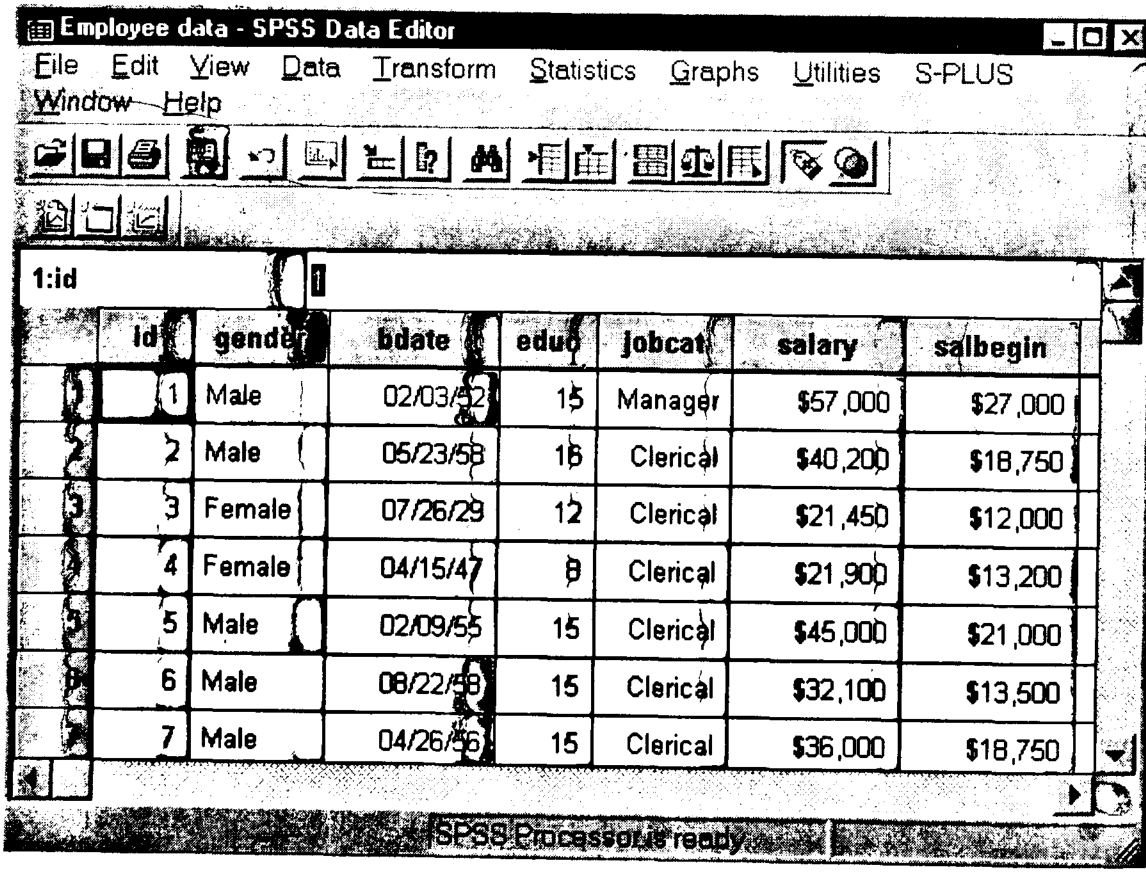
STATGRAPHICS Plus 3 տարբերակը հնարավորություն է տալիս ստեղծել բազմաթիվ StatGallery էջեր, որոնցից յուրաքանչյուրը իր հերթին կարող է ունենալ մինչև 9 ենթապատուհան: Դրանք օգտագործվում են տարբեր տեղերում գտնվող տեքստային և գրաֆիկական վախանակները միաժամանակ դիտելու կամ տպելու համար: Մկնիկի աջ կոճակի սեղմամբ հայտնվող հրամանացանկի միջոցով կարելի է ձևափոխել, տեղափոխել, տպել կամ ջնջել պատուհաններում գտնվող տեղեկությունները:



Նկար 10: STATGRAPHICS ծրագրաշարի StatGallery էջը:

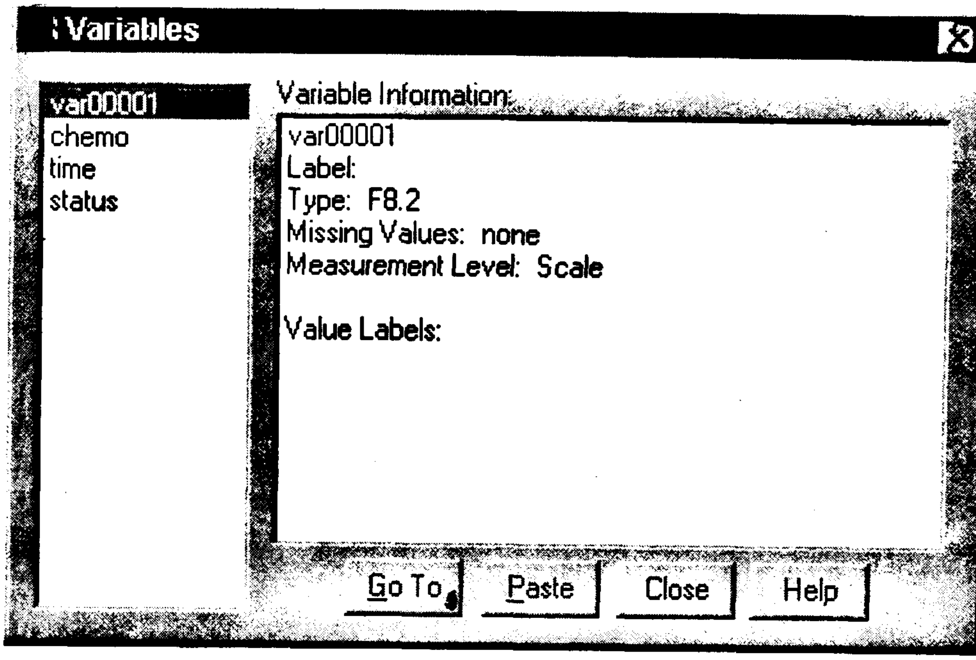
15.4. SPSS ծրագրաշարի բովանդակությունը

SPSS 8.0 ծրագրաշարում տեղ գտած գրաֆիկական հնարավորությունները հեշտացնում և մատչելի են դարձնում աշխատանքը տվյալների հետ:



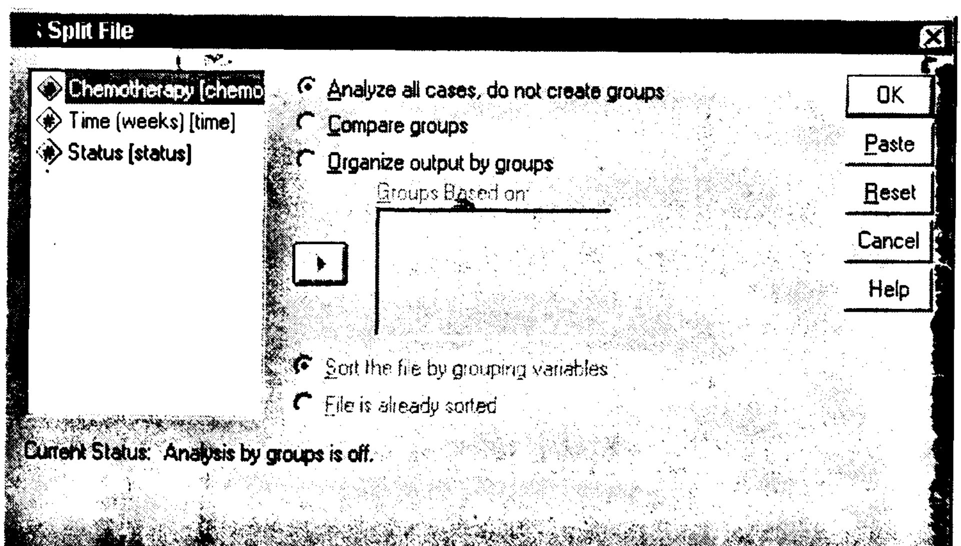
Նկար 11: SPSS ծրագրաշարի գլխավոր աշխատանքային դաշտը:

Ծրագրաշարի հրամանացանկի կարևորագույն հրամանները տեղ են գտել նաև գործիքային տողերում: Բացի մի շարք հասարակ հրամաններից, ինչպիսիք են՝ ֆայլի բացումը (Open File), գրանցումը (Save File), տպումը (Print), սխալի ուղղումը (Undo) և արժեքի փնտրումը (Search for Data), այստեղ ներառված են նաև SPSS-ին հատուկ որոշ հրամաններ: Դրանց օգնությամբ կարելի է ստանալ պահանջվող փոփոխականի սահմանումը, տվյալների տիպը, փոփոխականի պիտակը, արժեքների պիտակները և այլն (Variables) (տե՛ս նկար 10):



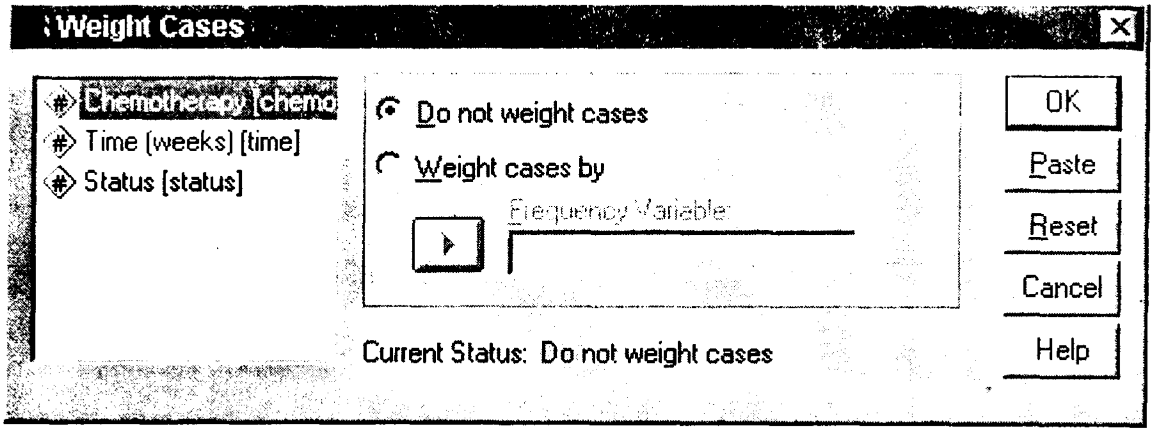
Նկար 12: Variables հրամանի պատուհանը:

Ինչպես նաև տվյալների ֆայլը բաժանել տարբեր խմբերի՝ մեկ կամ մի խումբ փոփոխականների համար վերլուծություն կատարելու նպատակով (Split File) (տե՛ս նկար 11):



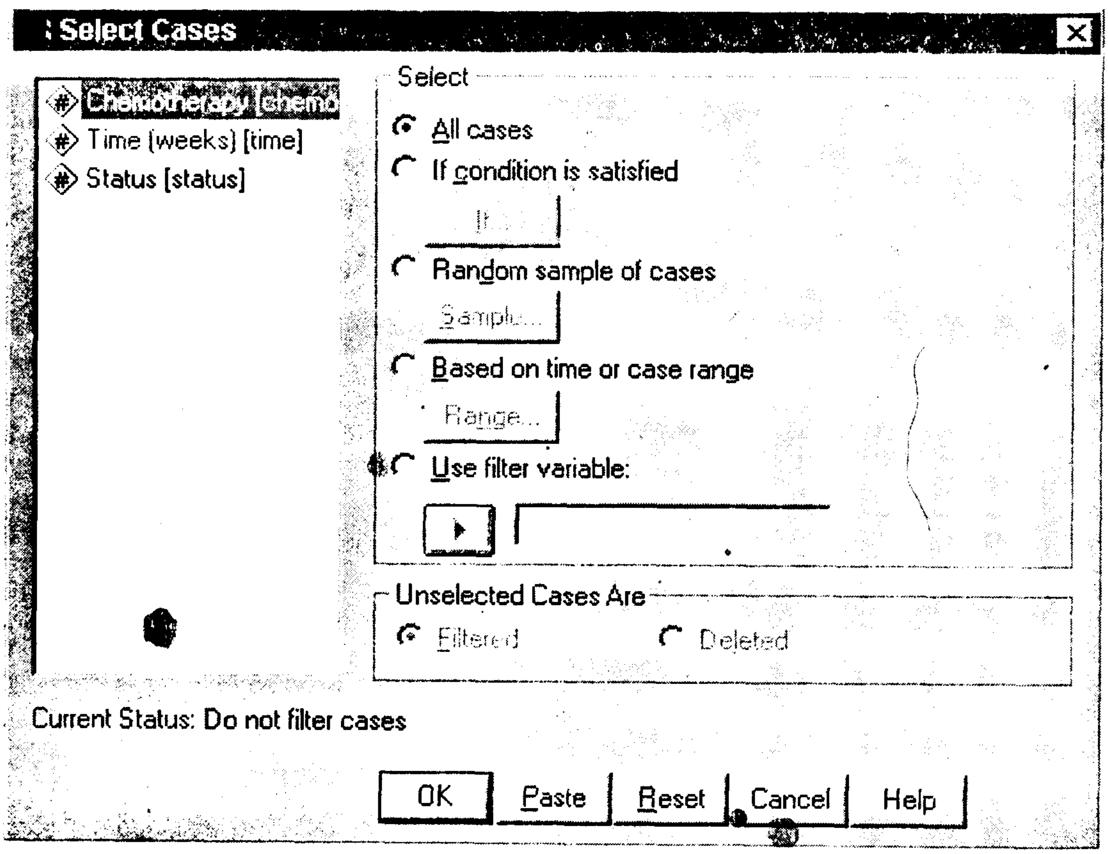
Նկար 13: «Ֆայլի բաժանում» հրամանի պատուհանը:

Գործիքային տողից «Պատահարների կշիռ» (Weight Cases) հրամանը վերլուծության նպատակով պատահարներին վերագրում է տարբեր կշիռներ: Կշռվող փոփոխականի արժեքը պետք է ցույց տա տվյալների ֆայլի յուրաքանչյուր պատահարով ներկայացվող դիտումների քանակը: Զրոյական, բացասական կամ անհայտ արժեքով պատահարները վերլուծության մեջ բացառվում են: Կոտորակային արժեքները թույլատրելի են:



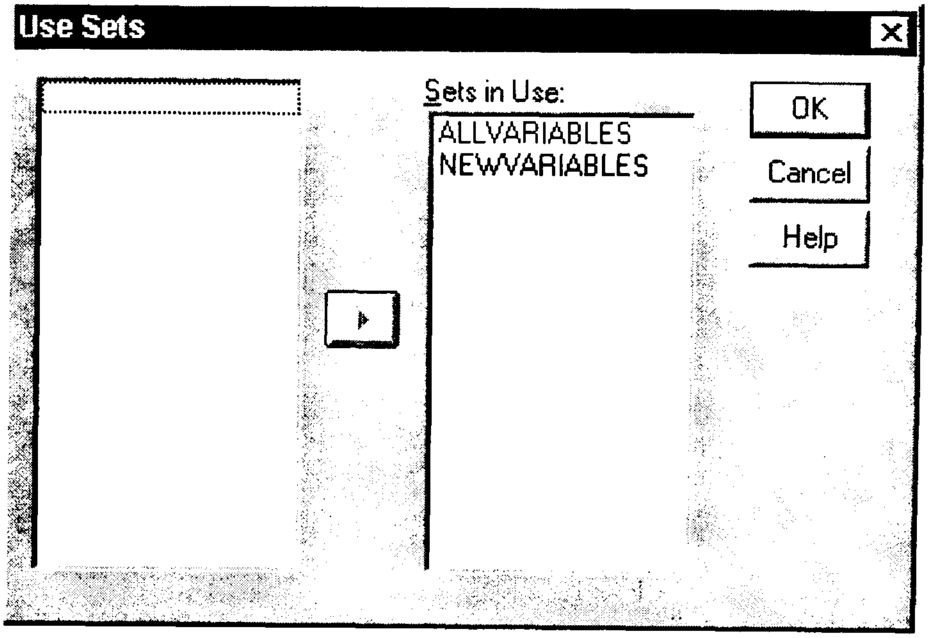
Նկար 14: «Պատահարների կշիռ» հրամանի պատուհանը:

Իսկ «Ընտրել պատահար» (Select Cases) հրամանն առաջարկում է պատահարների ենթախմբերի տարբեր եղանակների ընտրություն: Հնարավորություն կա ընտրել կամայական պատահարների օրինակներ: Ենթախումբ սահմանելիս՝ կարելի է սահմանել փոփոխականի արժեքները և միջակայքը, ժամանակային միջակայքը, տողի համարը, հանրահաշվական, տրամաբանական տարբեր արտահայտություններ և ֆունկցիաներ:



Նկար 15: «Ընտրել պատահար» հրամանի պատուհանը:

Ծրագրաշարում հնարավորություն կա նաև սահմանափակել երկխոսության պատուհանում ցուցադրվող փոփոխականների բազմությունը (Use Sets): Փոքր բազմությունը հեշտացնում է պահանջվող փոփոխականը գտնելը, ինչպես նաև մեծացնում է հաշվարկների արագությունը:



Նկար 16: Use Sets հրամանի պատուհանը:

15.5. S-PLUS 2000 ծրագրաշարը

S-PLUS համակողմանի ծրագրաշարի հրամանացանկերը, պատուհանները և գործիքային տողերը տվյալների հետ արհեստավարժորեն աշխատելու, գծապատկերներ և գրաֆիկներ ստեղծելու և S-PLUS-ի ենթածրագրեր գրելու լայն հնարավորություններ են ընձեռում: Հրամանացանկերին կարելի է դիմել ինչպես մկնիկի, այնպես էլ ստեղնաշարի օգնությամբ: Երկխոսության պատուհանները կարելի է բացել կամ հրամանացանկի տարբերակները ընտրելու, կամ էլ օբյեկտի վրա մկնիկի ձախ ստեղնի կրկնակի կանչի միջոցով:

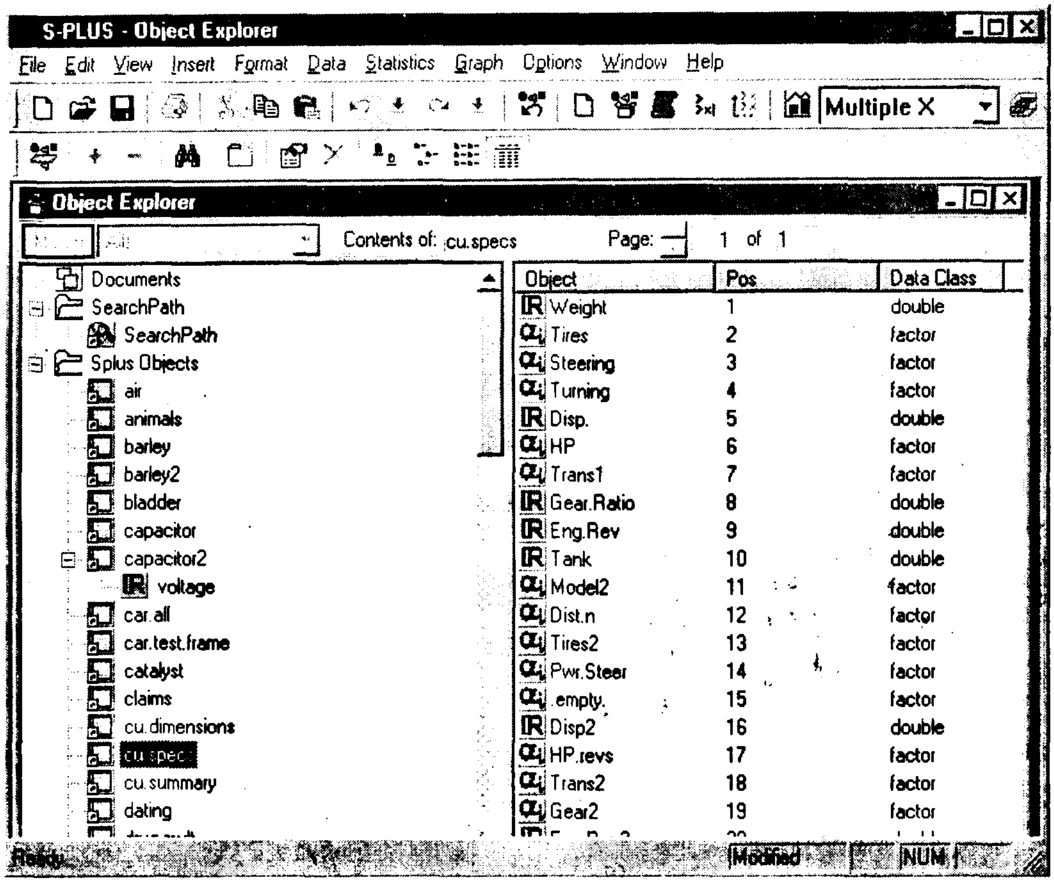
S-PLUS-ում կարելի է աշխատել միաժամանակ մի քանի պատուհանների հետ: Դա հեշտացնում է տվյալների խմբագրումը, ենթածրագրեր աշխատեցնելը և գծապատկերներ կառուցելը: Ծրագրաշարի պատուհաններն իրենց կառուցվածքով նման են Windows-ի սովորական պատուհաններին:

Գլխավոր գործիքային տողերը տասն են: Գործիքային տողերը պարունակում են հրամանների արագ օգտագործման սեղմակներ: Դրանք կարելի է օգտագործել ինչպես ֆայլային այնպիսի հասարակ գործողությունների համար, ինչպիսիք են՝ ֆայլի բացումը և պահպանումը, այնպես էլ միայն S-PLUS-ին հատուկ գործողությունների համար: Կամայական գործիքային տող կարելի է տեղաշարժել բաց պատուհանի երկայնքով:

Ծրագիրն ունի 6 տեսակի պատուհաններ. «Օբյեկտներ հետազոտող» (Object explorer), «Տվյալների պատուհան» (Data window), «Գրաֆիկական շերտ» (Graph sheet), «Հրամանների պատուհան» (Commands window), «Ենթածրագրային պատուհան» (Script window) և «Հաշվետվության պատուհան» (Report window): Այս պատուհանները հնարավորություն են տալիս հեշտորեն կազմակերպել աշխատանքը միաժամանակ տվյալների, ենթածրագրերի և գծապատկերների հետ:

S-PLUS-ի հրամանացանկերը փոփոխվում են աշխատանքային պատուհանին համապատասխան: Օրինակ, եթե ակտիվ է «Տվյալներ» (Data) պատուհանը, ապա հրամանացանկերը ցուցադրում են միայն այդ պատուհանում պիտանի հրամանները: Այդ դեպքում գրաֆիկական և ծրագրավորման հնարավորությունները կամ բացակայում են, կամ էլ պասիվ են:

Ծրագրաշարում աշխատողի սահմանած պատուհանները կարող են կապվել ծրագրի գործողությունների հետ: Գործողությունները կարող են ցուցադրվել «Օբյեկտներ հետազոտող» պատուհանում և հրավիրվել մկնիկի կրկնակի կանչով:



Նկար 17: «Օբյեկտներ հետազոտող» պատուհանը:

Հնարավորություն կա տվյալները մշակել հատուկ պատուհանում ցուցադրվող տեղեկությունների սյունակների տեսքով: Կարելի է բացել միաժամանակ մի քանի, բնույթով տարբեր տվյալներ պարունակող պատուհաններ: «Տվյալների պատուհանը» մեծ հնարավորություններ է ընձեռում տվյալների մշակման և փոխակերպման համար:

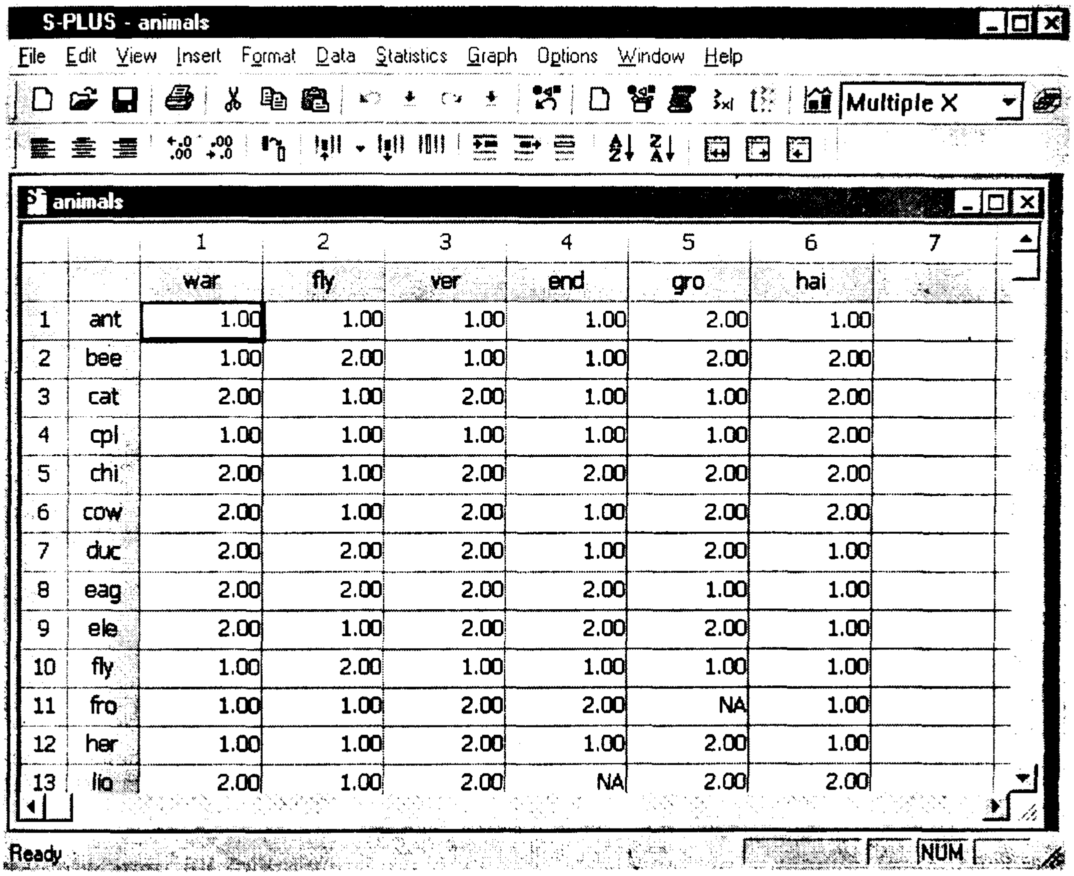
Տվյալների բազմությունը կարող է պարունակել շատ ավելի մեծաքանակ տեղեկություններ, քան հնարավոր է միանգամից ցուցադրել էկրանին: «Տվյալների պատուհանի» միջոցով կարելի է տեղեկությունները դիտել մաս-մաս: Պատուհանը հնարավոր է կրկնապատկել՝ տվյալների երկու տարբեր հատվածները միաժամանակ ուսումնասիրելու համար:

Վերլուծության նպատակով տվյալներ ներածելու պարզագույն եղանակներից է պատրաստի ֆայլի ներածումը: S-PLUS-ը հնարավորություն է տալիս նաև տվյալները և գծապատկերները արտածել տարբեր տեսակի ֆայլերով՝ տպելու և այլ ծրագրերում օգտագործելու համար: Բացի ASCII և FASCII ֆայլերից, S-PLUS հնարավոր է ներածել հետևյալ տեսակի ֆայլերը. Microsoft Excel (.XLS), Quattro Pro Worksheets (.WQ1), Paradox Databases (.DB), Lotus Worksheets (.WKS), dBase (.DBF), FoxPro, Systat, SigmaPlot (.JNB), SPSS (.SAV), SAS (.SD2), SAS Transport (version 6.x. TPT), Microsoft Access (.MDB), Matlab (.MAT), SPSS Export (.POR), S-PLUS Transport (.SDD), STATA (.DTA), Gauss (.DAT):

Ծրագիրն առարկայառուղղված է: «Օբյեկտ հետազոտողը» (Object Explorer) պարզ, բայց հզոր միջավայր է, որը հնարավորություն է տալիս նշել, դիտել, ստեղծել և մշակել օբյեկտները: Այն նման է Windows Explorer-ին, որը ծառանման կառույցով ներկայացնում է ամբողջ հաշվարը՝ ցուցադրելով ֆայլերը, էկրանը, ղեկավարման վահանակը և այլն: «Օբյեկտ հետազոտողը» համանման ծառով ներկայացնում է օբյեկտները:

S-PLUS ծրագրաշարն ունի նաև կապակցման միջոցներ, այնպես որ գրաֆիկական շերտերում կարելի է օգտագործել այլ ծրագրերով ստեղծված տվյալներ կամ օբյեկտներ: Հնարավոր է ծրագրի որևէ կետ կապել մեկ այլ ծրագրի հետ: Այդպիսի կետերը, համապատասխան տվյալները փոփոխելիս, ինքնաբերաբար վերափոխվում են:

Կարելի է այլ ծրագրերից S-PLUS-ում ներգրավել տարբեր գրաֆիկներ (օրինակ Excel-ից): Ներդրված օբյեկտը վերածվում է S-PLUS-ի գրաֆիկական շերտի մասի:



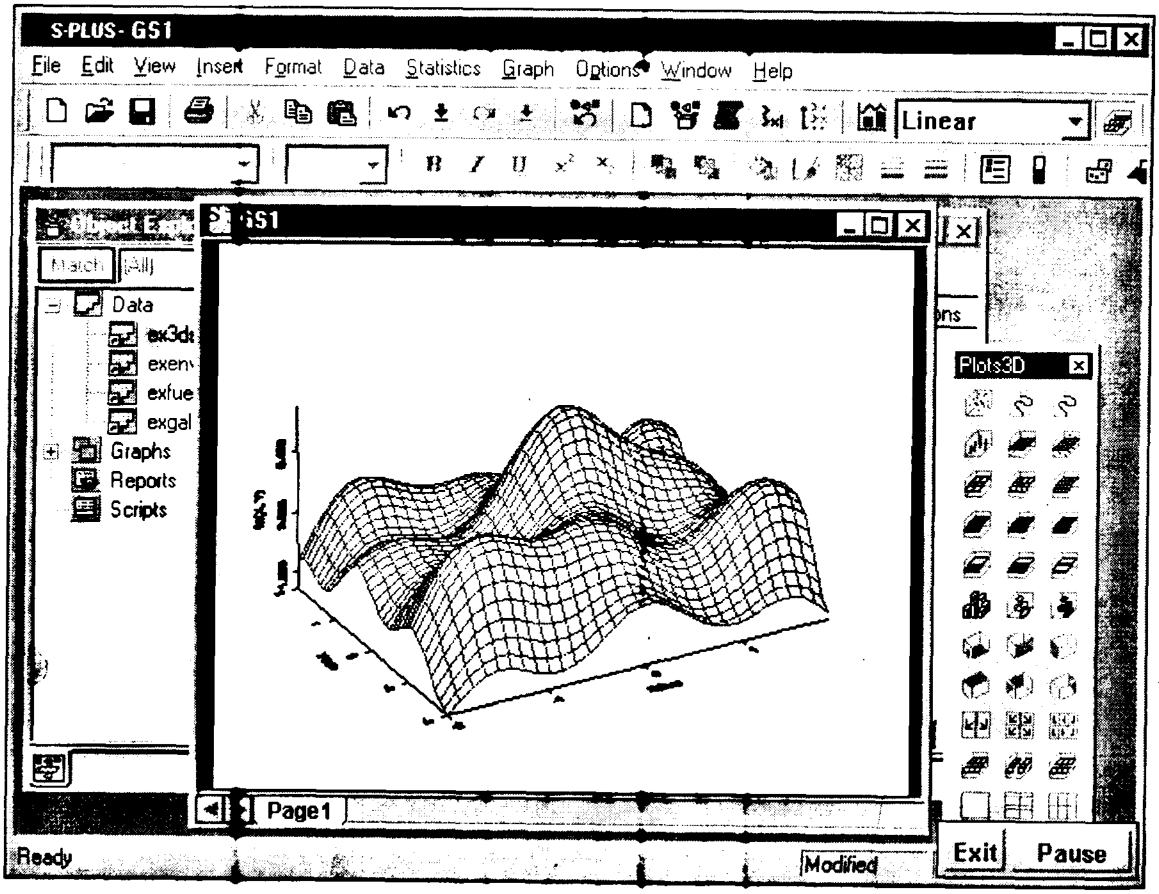
		1	2	3	4	5	6	7
		war	fly	ver	end	gro	hai	
1	ant	1.00	1.00	1.00	1.00	2.00	1.00	
2	bee	1.00	2.00	1.00	1.00	2.00	2.00	
3	cat	2.00	1.00	2.00	1.00	1.00	2.00	
4	cpl	1.00	1.00	1.00	1.00	1.00	2.00	
5	chi	2.00	1.00	2.00	2.00	2.00	2.00	
6	cow	2.00	1.00	2.00	1.00	2.00	2.00	
7	duc	2.00	2.00	2.00	1.00	2.00	1.00	
8	eag	2.00	2.00	2.00	2.00	1.00	1.00	
9	ele	2.00	1.00	2.00	2.00	2.00	1.00	
10	fly	1.00	2.00	1.00	1.00	1.00	1.00	
11	fro	1.00	1.00	2.00	2.00	NA	1.00	
12	her	1.00	1.00	2.00	1.00	2.00	1.00	
13	lio	2.00	1.00	2.00	NA	2.00	2.00	

Նկար 18: «Տվյալների պատուհանը»:

Գծապատկերներ կարելի է կառուցել «Տվյալների պատուհանի» միջոցով կամ մի շարք այլ ծրագրաշարերից (ներառյալ Excel, Lotus, dBase, SAS, SPSS և տեքստային (ASCII) ֆայլերը) ներածված տվյալների միջոցով: Գծապատկերները պահվում են «գրաֆիկական շերտերում» (Graph sheet): «Գրաֆիկական շերտում» հնարավոր է տեղադրել մեկ կամ մի քանի գծապատկերներ, ինչպես նաև կարելի է աշխատել միաժամանակ մի քանի «գրաֆիկական շերտերի» հետ: Գրաֆիկական շերտը պահպանելու դեպքում նրանում գտնվող բոլոր գծապատկերները պահպանվում են .SGR ընդլայնումով ֆայլի մեջ: Գծապատկերի յուրահատկությունները, ինչպես օրինակ, տեսակը, տվյալները, ձևավորումը և այլն, նույնպես պահպանվում են: Սակայն իրական տվյալները գծապատկերի հետ չեն պահվում, քանի որ շատ հաճախ միևնույն տվյալներով ստեղծվում են մի քանի գծապատկերներ:

«Գրաֆիկական շերտերը» իրենց հերթին կարելի է արտածել տարբեր տեսակի ֆայլերով, ներառյալ Windows Metafile (.WMF), Windows Bitmap (.BMP) և Tagged Image File (.TIFF) տեսակի ֆայլերը: Այս արտածված ֆայլերն արդեն կարելի է ներմուծել այլ ծրագրեր: Սակայն հետագայում մշակելու համար անհրաժեշտ է «գրաֆիկական շերտը» պահպանել որպես S-PLUS-ի գծապատկեր:

S-PLUS-ն ունի գրաֆիկական ձևավորման մեծ հնարավորություններ, ներառյալ կոորդինատային առանցքների, կետերի, վերնագրերի և բացատրությունների ձևավորումը: Այդ հնարավորություններից կարելի է օգտվել գծապատկերը ստեղծելուց անմիջապես հետո:



Նկար 19: «Գրաֆիկական շերտը»:

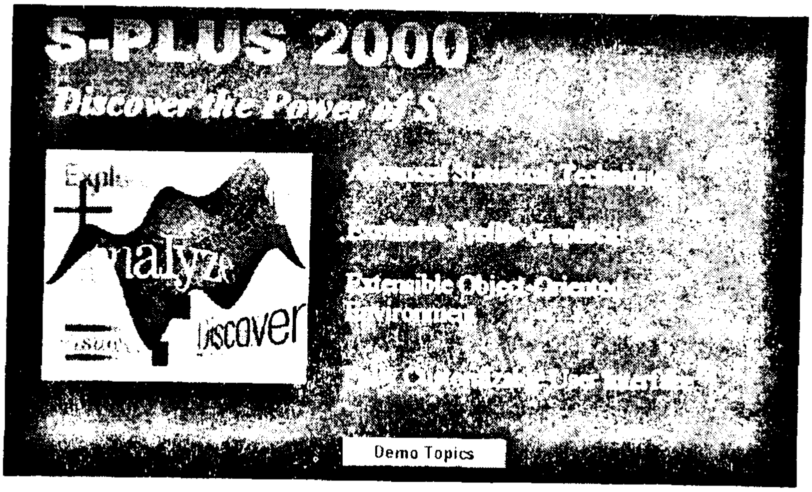
«Գրաֆիկական շերտում» գտնվող բոլոր օբյեկտները համարվում են գրաֆիկական օբյեկտներ: Դրանց հետ աշխատելիս՝ հնարավոր է փոխել ձևը, չափը և գույնը, վերադասավորել, պատճենել և տեղափոխել:

Օբյեկտը գծապատկերին ավելացնելուց անմիջապես հետո կարելի է այն տեղափոխել, պատճենել, չափը փոխել կամ փոփոխել պարամետրերը: Օբյեկտները նշելու և ձևավորելու նպատակով կարելի է օգտագործել «Օբյեկտներ հետազոտողը»: Այն նպատակահարմար է ընթացիկ բոլոր օբյեկտները հաջորդականորեն դիտելու համար: Այն նաև հարմար է այնպիսի օբյեկտների մշակման ժամանակ, երբ ցուցադրման այլ ձևերով տվյալ օբյեկտը դժվարությամբ է նշվում (օրինակ՝ մասամբ ծածկված օբյեկտները):

Վիճակագրական գործողությունների հարուստ ընտրանի է ընդգրկված գլխավոր հրամանացանկի «Տվյալներ» (Data) և «Վիճակագրություն» (Statistics) ենթաբաժիններում:

«Տվյալներ» հրամանացանկը պարունակում է տվյալների աղյուսակացման, բաշխման ֆունկցիաների հաշվարկման և պատահական օրինակներ ու թվեր ստեղծելու հրամաններ: «Վիճակագրություն» հրամանացանկը պարունակում է համառոտ տվյալների ստեղծման, վարկածների ստուգման և այլ վիճակագրական մոդելներ:

«Ենթածրագրային պատուհանում» (Script window) կարելի է գրել ենթածրագրեր՝ տվյալների վերլուծության և գծապատկերների ստեղծման շատ գործողություններ ավտոմատացնելու համար: Յուրաքանչյուր «Ենթածրագրային պատուհան» ունի արտածման վահանակ, որը ցույց է տալիս աշխատող ենթածրագրի արտաբերած տեղեկությունները և ծրագրային վահանակ, որը նախատեսված է ենթածրագրի հրամանները մուտքագրելու համար: Ենթածրագրերը հնարավորություն են տալիս, օգտագործելով S-PLUS-ի ծրագրավորման լեզուն, գրել հրամաններ՝ տվյալների ներածման, արտածման կամ փոխակերպման համար, վերլուծությունների, ինչպես նաև գծապատկերներ ստեղծելու, ձևավորելու կամ տպելու համար: Ենթածրագրերը կարելի է աշխատեցնել S-PLUS-ից կամ որևէ այլ ծրագրից:



Նկար 20: S-PLUS ծրագրաշարի ուսումնական ենթաբաժինը:

«Ենթածրագրային պատուհանը» «Հրամանային պատուհանի» (Commands window) երկընտրանքն է: «Հրամանային պատուհանը» ինտերակտիվ է, և մուտքագրված հրամանները թարգմանչի օգնությամբ անմիջապես գնահատվում են յուրաքանչյուր հրամանի տակ ցուցադրվող արդյունքով: Իր հերթին «Ենթածրագրային պատուհանը» հնարավորություն է տալիս մուտքագրել հրամաններ և ֆունկցիաներ, սակայն գնահատում է դրանք միայն պահանջի դեպքում: Ենթածրագիրը կարելի է աշխատեցնել «Run» սեղմակի օգնությամբ: Ենթածրագրի որևէ հատվածի նշված լինելու դեպքում կատարվում է միայն այդ հատվածը: Արդյունքը ցուցադրվում է արտածման վահանակում, այլ ոչ թե յուրաքանչյուր հրամանի տակ:

«Հրամանային պատուհանը» նախընտրելի է տվյալների ինտերակտիվ վերլուծության ժամանակ, մինչդեռ «Ենթածրագրային պատուհանը» օգտակար է բարդ ֆունկցիաներ գրելիս:

S-PLUS ծրագրաշարի առավելություններից մեկն էլ մատչելի ուսուցանողական ենթաբաժինն է (DEMO), համալրված մուլտիմեդիայի ժամանակակից հնարավորություններով և հրապույրներով: Այն օգնում է հեշտորեն յուրացնել ծրագրաշարի հիմունքները և արագորեն ծանոթանալ հրամանացանկերից և գործիքներից օգտվելու կանոններին: Առաջարկվող վարժությունները ու ցուցադրական օրինակները ուղղեկցվում են հաճելի երաժշտությամբ: Նկար 20-ում պատկերված է ուսումնական ենթաբաժնի պատուհանը:

Գրականության ցանկ

Համբարձումյան Գ. Հ. Հավանականությունների տեսություն, III հրատ., Երեվան, Լույս, 1977:

Համբարձումյան Գ. Հ. Պատահական պրոցեսներ, Երևան, Լույս, 1974:

Համբարձումյան Գ. Հ., Հարությունյան Ե. Ա., Պետրոսյան Ն. Հ., Սիմոնյան Ա. Խ. Հավանականությունների տեսության խնդրագիրք, Մաս առաջին, Երևան, ԵՊՀ, 1973:

Հարությունյան Ե. Ա., Հարությունյան Մ. Ե. Ինֆորմացիայի տեսություն, Երևան, ԵՊՀ, 1987:

Ղազանչյան Տ. Պ., Մեսրոպյան Ն. Խ. Հավանականությունների տեսության խնդրագիրք, Մաս երկրորդ, Երևան, Փյունիկ, 1997:

Ջանփոլադյան Բ. Քոմփյութերային կիրառական տեխնոլոգիաներ, Երևան, 1999:

Մեսրոպյան Ն. Խ., Ղազանչյան Տ. Պ. Հավանականությունների տեսության խնդրագիրք, Մաս առաջին, ԵՊՀ, 1986:

Պողոսյան Ա., Դավթյան Ա. Հավանականությունների տեսության և մաթեմատիկական վիճակագրության խնդիրների լուծման ձեռնարկ, Երևան, Տնտեսագետ, 1997:

Alreck P. L., Settle R. B. The Survey Research Handbook, Irwin, 1985.

Anton H., Kolman B., Averbach B. Mahtematics with Applications for the Management, Life and Social Sciences, New York, HBJ, 1988.

Berenson M. L., Levine D. M. Basic Business Statistics, Annotated Instructors Edition, New Jersey, Prentice Hall Englewood Cliffs, 1992.

Blahut R. E. Principles and Practice of Information Theory, New York, Addison – Wesley, 1987.

Cover T. M., Thomas J. A. Elements of information theory, New York, Wiley, 1991.

Edmondson A., Druce D. Advanced Biology Statistics, Oxford, Oxford University Press, 1996.

Hanke E., Reitsch G. Understanding Business Statistics, Boston, Irwin, 1991.

Hall O. P. Jr., Adelman H. V. Business Statistics. Text, Cases, Software, Boston, Irwin, 1991.

Lapin L. Statistics for Modern Business Decisions, Harcourt Brace Joanovich, 1990

Mann P. S. Statistics for Business and Economics. Annotated Instructors Edition, New York, Wiley, 1995.

Mc Clave T., Benson G., Sincich T. Statistics for Business and Economics, Prentice Hall, 1998.

Qittard Pinon F. Lingnelet Cours Statistiques, Tome 1, Calcul des Probabilites, Tome 2, Decision, Estimation, Test, Alger, Office des Publications Universitaires, 1983

Upton G., Cook I. Understanding Statistics, Oxford, Oxford University Press, 1996.

Webster A. L. Applied Statistics for Business and Economics, Instructors Edition. Chicago, Irwin, 1995.

Агапов Г. И. Задачник по теории вероятностей, Москва, Высшая школа, 1986.

Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений, Москва, Статистика, 1974.

Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных, Москва, Финансы и статистика, 1983.

Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей, Москва, Финансы и статистика, 1985.

Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности, Москва, Финансы и статистика, 1989.

Айвазян С. А., Степанов В. С. Инструменты статистического анализа данных, Мир ПК, N8, 1997, стр. 32–41.

Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрии, Москва, Юнити, 1998.

Большев Л. Н., Смирнов Н. В. Таблицы математической статистики, Москва, Наука, 1983.

Боровиков В. П., Боровиков И. П. Статистика. Статистический анализ и обработка данных в среде Windows, Москва, Филин, 1998.

Боровиков В. П. Популярное введение в программу STATISTIKA, Москва, Компьютер пресс, 1998.

Бочаров П. П., Печинкин А. В. Теория вероятностей. Математическая статистика, Москва, Гардарика, 1998.

Булдык Г. М. Теория вероятностей и математическая статистика, Минск, Вышэйша школа, 1989.

Венецкий И. Г., Кильдишев Г. С. Теория вероятностей и математическая статистика, Москва, Статистика, 1975.

Винн Р., Холден К. Введение в прикладной эконометрический анализ, Москва, Финансы и статистика, 1981.

Гликман Н. Эконометрический анализ региональных систем, Москва, Прогресс, 1980.

Гмурман В. Е. Теория вероятностей и математическая статистика, Москва, Высшая школа, 1972.

Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике, Москва, Высшая школа, 1979.

Доугерти К. Введение в эконометрику, Москва, Инфра–М, 1997.

Ефимова М. Р., Петрова Е. В., Румянцев В. Н. Общая теория статистики, Москва, Инфра–М, 1998.

Ивченко Г. И., Медведев Ю. И. Математическая статистика, Москва, Высшая школа, 1984.

Кади Дж. Количественные методы в экономике, Москва, Прогресс, 1977.

Калинина В. Н., Панкин В. Ф. Математическая статистика, Москва, Высшая школа, 1998.

Карасев А. И. Теория вероятностей и математическая статистика, Москва, Статистика, 1979.

Кейн Э. Экономическая статистика и эконометрия. Вып. 1, 2, Москва, Статистика, 1977.

Колемаев В. А., Калинина В. Н. Теория вероятностей и математическая статистика, Москва, Инфра–М, 1997.

Колемаев В. А., Староверов О. В. Турундаевский В. Б. Теория вероятностей и математическая статистика, Москва, Высшая школа, 1991.

Колмогоров А. Н., Журбенко И. Г., Прохоров А. В. Введение в теорию вероятностей, Москва, Наука, 1982.

Магнус Я. В., Катышев П. К., Пересецкий А. П. Эконометрика. Начальный курс, Москва, Дело, 1997.

Маленво Э. Статистические методы эконометрии, Вып. 1, 2, Москва, Статистика, 1975.

Манзон Б. Статистика 5.1 — программа для ученых и предпринимателей, Компьютер Пресс, N8, 1997, стр. 97—103.

Мешалкин Л. Д. Сборник задач по теории вероятностей, Москва, МГУ, 1963.

Михок Г., Урсяну В. Выборочный метод и статистическое оценивание, Москва, Финансы и статистика, 1982.

Мостеллер Ф., Рурке Р, Томас Дж. Вероятность, Москва, Мир, 1969.

Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике, Серия "Библиотека иностранных книг для экономистов и статистиков", Москва, Финансы и статистика, 1982.

Нейман Ю. Вводный курс теории вероятностей и математической статистики, Москва, Наука, 1968.

Ниворожкина Л. И., Морозова З. А. Основы статистики с элементами теории вероятностей для экономистов. Руководство для решения задач, Ростов—на—Дону, Феникс, 1999.

Реньи А. Трилогия о математике. Диалоги о математике. Письма о вероятности. Дневник. Записки студента по теории информации, Москва, Мир, 1980.

Розанов Ю. А. Лекции по теории вероятностей, Москва, Наука, 1986.

Сборник задач по теории вероятностей, математической статистике, и теории случайных функций. Под редакцией А. А. Свешникова, Москва, Наука, 1970.

Севастьянов Б. А. Курс теории вероятностей и математической статистики, Москва, Наука, 1982.

Севастьянов Б. А., Чистяков В. П., Зубков А. М. Сборник задач по теории вероятностей, Москва, Наука, 1980.

Секей Г. Парадоксы в теории вероятностей и математической статистике, Москва, Мир, 1990.

Сильвестров Д. С. Программное обеспечение прикладной статистики, Москва, Финансы и статистика, 1988.

Справочник по прикладной статистике, Тома 1 и 2. Под редакцией Ллойда Э., Ледермана У., Тюрина Ю.Н., Москва, Финансы и статистика, 1990.

Статистический словарь. Редактор Королев М. А., Москва, Финансы и статистика, 1989.

Тутубалин В. Н. Теория вероятностей, Москва, МГУ, 1972.

Тюрин Ю. Н., Макаров А. А. Статистический анализ на компьютере, Москва, Инфра—М, 1998.

Четыркин Е. М., Калихман И. Л. Вероятность и статистика, Москва, Финансы и статистика, 1982.

Чисар И., Кернер Я. Теория информации, Москва, Мир, 1985.

Феллер В. Введение в теорию вероятностей и ее приложения, В 2—х томах, Москва, Мир, 1984.

Ширяев А. Н. Вероятность, Издание второе, Москва, Наука, 1989.

Яглом А. М., Яглом И. М. Вероятность и информация, Издание третье, Москва, Наука, 1973.

Աղյուսակներ

Ա.1. Պուասոնի բաշխման գումարային հավանականությունների աղյուսակ՝

$$\mathbf{P}\{X \geq x\} = \sum_{k=x}^{\infty} \frac{\lambda^k}{k!} \exp^{-\lambda}:$$

x \ λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	0.09516	0.18127	0.25918	0.32968	0.39347	0.45119	0.50341
2	0.00468	0.01752	0.03694	0.06155	0.09020	0.12190	0.15580
3	0.00016	0.00115	0.00360	0.00793	0.01439	0.02312	0.03414
4	0	0.00006	0.00027	0.00078	0.00175	0.00336	0.00575
5		0	0.00002	0.00006	0.00017	0.00039	0.00079
6					0.00001	0.00004	0.00009
7							0.00001

x \ λ	0.8	0.9	1.0	2.0	4.0	6.0	8.0
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	0.55067	0.59343	0.63212	0.86466	0.98168	0.99752	0.99968
2	0.19121	0.22752	0.26424	0.59399	0.90842	0.98265	0.99698
3	0.04742	0.06286	0.08030	0.32332	0.76190	0.93803	0.98625
4	0.00908	0.01346	0.01899	0.14288	0.56653	0.84880	0.95762
5	0.00141	0.00234	0.00368	0.05226	0.37116	0.71494	0.90037
6	0.00018	0.00034	0.00059	0.01656	0.21487	0.55432	0.80976
7	0.00002	0.00004	0.00008	0.00453	0.11067	0.89370	0.68663
8		0.00001	0.00002	0.00110	0.05113	0.25602	0.54704
9				0.00024	0.02136	0.15276	0.40745
10				0.0005	0.00823	0.08392	0.28338
11				0.0001	0.00284	0.04262	0.18411
12					0.00092	0.02009	0.11192
13					0.00027	0.00883	0.06380
14					0.00008	0.00363	0.03418
15					0.00002	0.00140	0.01726
16					0.00001	0.00051	0.00823
17						0.00017	0.00372
18						0.00006	0.00159
19						0.00002	0.00065
20						0.00001	0.00025
21							0.00009
22							0.00003
23							0.00001
24							0

Ա.2. Նորմալ բաշխման խտության ֆունկցիայի՝ $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$,

և Լապլասի ֆունկցիայի՝ $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-\frac{x^2}{2}) dx$, աղյուսակ:

x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$
0.00	0.3989	0.0000	1.25	0.1826	0.3944	2.50	0.0175	0.4938
0.05	0.3984	0.0199	1.30	0.1714	0.4032	2.55	0.0154	0.4946
0.10	0.3970	0.0398	1.35	0.1604	0.4115	2.60	0.0136	0.4953
0.15	0.3945	0.0596	1.40	0.1497	0.4192	2.65	0.0119	0.4960
0.20	0.3910	0.0793	1.45	0.1394	0.4265	2.70	0.0104	0.4965
0.25	0.3867	0.0987	1.50	0.1295	0.4332	2.75	0.0091	0.4970
0.30	0.3814	0.1179	1.55	0.1200	0.4394	2.80	0.0079	0.4974
0.35	0.3752	0.1368	1.60	0.1109	0.4452	2.85	0.0069	0.4978
0.40	0.3683	0.1554	1.65	0.1023	0.4505	2.90	0.0060	0.4981
0.45	0.3605	0.1736	1.70	0.0940	0.4554	2.95	0.0051	0.4984
0.50	0.3521	0.1915	1.75	0.0863	0.4599	3.00	0.0044	0.4986
0.55	0.3429	0.2088	1.80	0.0790	0.4641	3.05	0.0038	0.4989
0.60	0.3332	0.2257	1.85	0.0721	0.4678	3.10	0.0033	0.4990
0.65	0.3230	0.2422	1.90	0.0656	0.4713	3.15	0.0028	0.4992
0.70	0.3123	0.2580	1.95	0.0596	0.4744	3.20	0.0024	0.4993
0.75	0.3011	0.2734	2.00	0.0540	0.4773	3.25	0.0020	0.4994
0.80	0.2897	0.2881	2.05	0.0488	0.4798	3.30	0.0017	0.4995
0.85	0.2780	0.3023	2.10	0.0440	0.4821	3.35	0.0015	0.4996
0.90	0.2661	0.3159	2.15	0.0396	0.4842	3.40	0.0012	0.4997
0.95	0.2541	0.3289	2.20	0.0355	0.4861	3.45	0.0010	0.4997
1.00	0.2420	0.3413	2.25	0.0317	0.4878	3.50	0.0009	0.4998
1.05	0.2299	0.3531	2.30	0.0283	0.4893	3.55	0.0007	0.4998
1.10	0.2179	0.3643	2.35	0.0252	0.4906	3.60	0.0006	0.4998
1.15	0.2059	0.3749	2.40	0.0224	0.4918	3.70	0.0004	0.4999
1.20	0.1942	0.3849	2.45	0.0198	0.4929	3.80	0.0003	0.4999

Ա.3. $\chi^2(N)$-բաշխման $p = \mathbf{P}\{\chi^2(N) < \chi_p^2(N)\}$ հավանականությանը համապատասխանող $\chi^2(N)$ քանորդիչների աղյուսակ:

N \ p	0.01	0.05	0.10	0.90	0.95	0.99
1	0.0002	0.004	0.02	2.71	3.84	6.64
2	0.02	0.10	0.21	4.61	5.99	9.21
3	0.12	0.35	0.58	6.25	7.82	11.34
4	0.30	0.71	1.06	7.78	9.49	13.28
5	0.55	1.15	1.61	9.24	11.07	15.09
6	0.87	1.64	2.20	10.65	12.59	16.81
7	1.24	2.17	2.83	12.02	14.06	18.48
8	1.65	2.73	3.49	13.36	15.51	20.09
9	2.09	3.33	4.17	14.68	16.92	21.67
10	2.56	3.94	4.87	15.99	18.31	23.21
11	3.05	4.58	5.58	17.28	19.68	24.72
12	3.57	5.23	6.30	18.55	21.03	26.22
13	4.11	5.89	7.04	19.81	22.36	27.68
14	4.66	6.57	7.79	21.06	23.69	29.14
15	5.23	7.26	8.55	22.31	25.00	30.58
16	5.81	7.96	9.31	23.54	26.30	32.00
17	6.41	8.67	10.09	24.77	27.59	33.41
18	7.02	9.39	10.86	25.99	28.87	34.81
19	7.63	10.12	11.65	27.20	30.14	36.19
20	8.26	10.85	12.44	28.41	31.41	37.57
21	8.90	11.59	13.24	29.62	32.67	38.93
22	9.54	12.34	14.04	30.81	33.92	40.29
23	10.20	13.09	14.85	32.01	35.17	41.64
24	10.86	13.85	15.66	13.19	36.42	43.98
25	11.52	14.61	16.47	34.38	37.65	44.31
26	12.20	15.37	17.29	35.56	38.89	45.64
27	12.88	16.15	18.11	36.74	40.11	46.96
28	13.56	16.93	18.94	37.92	41.34	48.28
29	14.26	17.71	19.77	39.09	42.56	49.59
30	14.95	18.49	20.60	40.26	43.77	50.89

Ա.4. Ստյուդենտի $\mathrm{t}(N)$-բաշխման $p = \mathbf{P}\{\mathrm{t}(N) < t_p(N)\}$ հավանականությանը համապատասխանող $t_p(N)$ քանորդիչների աղյուսակ:

N \ p	0.1	0.05	0.75	0.01	0.005
1	6.31	12.71	31.82	63.66	127.32
2	2.92	4.30	6.96	9.92	14.09
3	2.35	3.18	4.54	5.84	7.45
4	2.13	2.78	3.75	4.60	5.60
5	2.01	2.57	3.36	4.03	4.77
6	1.94	2.45	3.14	3.71	4.32
7	1.89	2.36	3.00	3.50	4.03
8	1.86	2.31	2.90	3.36	3.83
9	1.83	2.26	2.82	3.25	3.69
10	1.81	2.23	2.76	3.17	3.58
11	1.80	2.20	2.72	3.11	3.50
12	1.78	2.18	2.68	3.05	3.43
13	1.77	2.16	2.65	3.01	3.37
14	1.76	2.14	2.62	2.98	3.33
15	1.75	2.13	2.60	2.95	3.29
16	1.75	2.12	2.58	2.92	3.25
17	1.74	2.11	2.57	2.90	3.22
18	1.73	2.10	2.55	2.88	3.20
19	1.73	2.09	2.54	2.86	3.17
20	1.72	2.09	2.53	2.85	3.15
25	1.71	2.06	2.49	2.79	3.08
30	1.70	2.04	2.46	2.75	3.03
40	1.68	2.02	2.42	2.70	2.97
60	1.67	2.00	2.39	2.66	2.91
120	1.66	1.98	2.36	2.62	2.86
∞	1.64	1.96	2.33	2.58	2.81

Ա.5.1. Ֆիշերի $\mathcal{F}(N_1, N_2)$ բաշխման $p = \mathbf{P}\{\mathcal{F}(N_1, N_2 < F_p(N_1, N_2))\}$ հավանականությանը համապատասխանող $F_p(N_1, N_2)$ քանորդիչների աղյուսակ:

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1):$$

$$p = 0.90:$$

N_2 \ N_1	1	2	3	4	5	6	8	12	24	120
1	39.9	49.5	53.6	55.8	57.2	58.2	59.4	60.7	62.0	63.1
2	8.53	9.00	9.16	9.24	9.29	9.33	9.37	9.4	9.45	9.48
3	5.54	5.46	5.39	5.34	5.31	5.28	5.25	5.22	5.18	5.14
4	4.54	4.32	4.19	4.11	4.05	4.01	3.95	3.90	3.83	3.78
5	4.06	3.78	3.62	3.52	3.45	3.40	3.34	3.27	3.19	3.12
6	3.78	3.46	3.29	3.18	3.11	3.05	2.98	2.90	2.82	2.74
7	3.59	3.26	3.07	2.96	2.88	2.83	2.75	2.67	2.58	2.49
8	3.46	3.11	2.92	2.81	2.73	2.67	2.59	2.50	2.40	2.32
9	3.36	3.01	2.81	2.69	2.61	2.55	2.47	2.38	2.28	2.18
10	3.29	2.92	2.73	2.61	2.52	2.46	2.38	2.28	2.18	2.08
11	3.23	2.86	2.66	2.54	2.45	2.39	2.30	2.21	2.10	2.00
12	3.18	2.81	2.61	2.48	2.39	2.33	2.24	2.15	2.04	1.93
13	3.14	2.76	2.56	2.43	2.35	2.28	2.20	2.10	1.98	1.88
14	3.10	2.73	2.52	2.39	2.31	2.24	2.15	2.05	1.94	1.83
15	3.07	2.70	2.49	2.36	2.27	2.21	2.12	2.02	1.90	1.79
16	3.05	2.67	2.46	2.33	2.24	2.18	2.09	1.99	1.87	1.75
17	3.03	2.64	2.44	2.31	2.22	2.15	2.06	1.96	1.84	1.72
18	3.01	2.62	2.42	2.29	2.20	2.13	2.04	1.93	1.81	1.69
19	2.99	2.61	2.40	2.27	2.18	2.11	2.02	1.91	1.79	1.67
20	2.97	2.59	2.38	2.25	2.16	2.09	2.00	1.89	1.77	1.64
25	2.92	2.53	2.32	2.18	2.09	2.02	1.93	1.82	1.69	1.56
30	2.88	2.49	2.28	2.14	2.05	1.98	1.88	1.77	1.64	1.50
40	2.84	2.44	2.23	2.09	2.00	1.93	1.83	1.71	1.57	1.42
60	2.79	2.39	2.18	2.04	1.95	1.87	1.77	1.66	1.51	1.35
120	2.75	2.35	2.13	1.99	1.90	1.82	1.72	1.60	1.45	1.26
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.67	1.55	1.38	1.17

Ա.5.2. Ֆիշերի $\mathcal{F}(N_1, N_2)$ բաշխման $p = \mathbf{P}\{\mathcal{F}(N_1, N_2 < F_p(N_1, N_2))\}$ հավանականությանը համապատասխանող $F_p(N_1, N_2)$ քանորդիչների աղյուսակ:

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1):$$

$$p = 0.95:$$

N_2 \ N_1	1	2	3	4	5	6	8	12	30	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.1	254
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.5
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2541
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.69	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

Ա.5.3. Ֆիշերի $\mathcal{F}(N_1, N_2)$ բաշխման $p = \mathrm{P}\{\mathcal{F}(N_1, N_2 < F_p(N_1, N_2))\}$ հավանականությանը համապատասխանող $F_p(N_1, N_2)$ քանորդիչների աղյուսակ:

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1):$$

$$p = 0.975:$$

N_2 \ N_1	1	2	3	4	5	6	8	12	24	120
1	648	800	864	900	922	937	957	977	997	1014
2	38.5	39.0	39.2	39.3	39.3	39.3	39.4	39.4	39.5	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.5	14.3	14.1	14.0
4	12.2	10.7	9.98	9.60	9.36	9.20	8.98	8.75	8.51	8.31
5	10.0	8.43	7.76	7.39	7.15	6.98	6.76	6.52	6.28	6.07
6	8.81	7.26	6.60	6.23	5.99	5.82	5.60	5.37	5.12	4.90
7	8.07	6.54	5.89	5.52	5.29	5.12	4.90	4.67	4.42	4.20
8	7.57	6.06	5.42	5.05	4.82	4.65	4.43	4.20	3.95	3.73
9	7.21	5.71	5.08	4.72	4.48	4.32	4.10	3.87	3.61	3.39
10	6.94	5.46	4.83	4.47	4.24	4.07	3.85	3.62	3.37	3.14
11	6.72	5.26	4.63	4.28	4.04	3.88	3.66	3.43	3.17	2.94
12	6.55	5.10	4.47	4.12	3.89	3.73	3.51	3.28	3.02	2.79
13	6.41	4.97	4.35	4.00	3.77	3.60	3.39	3.15	2.89	2.66
14	6.30	4.86	4.24	3.89	3.66	3.50	3.29	3.05	2.79	2.55
15	6.20	4.77	4.15	3.80	3.58	3.41	3.20	2.96	2.70	
16	6.12	4.69	4.08	3.73	3.50	3.34	3.12	2.89	2.63	2.46
17	6.04	4.62	4.01	3.66	3.44	3.28	3.06	2.82	2.56	2.38
18	5.98	4.56	3.95	3.61	3.38	3.22	3.01	2.77	2.50	2.32
19	5.92	4.51	3.90	3.56	3.33	3.17	2.96	2.72	2.45	2.26
20	5.87	4.46	3.86	3.51	3.29	3.13	2.91	2.68	2.41	2.20
25	5.69	4.29	3.69	3.35	3.13	2.97	2.75	2.51	2.24	1.98
30	5.57	4.18	3.59	3.25	3.03	2.87	2.65	2.41	2.14	1.87
40	5.42	4.05	3.46	3.13	2.90	2.74	2.53	2.29	2.01	1.72
60	5.29	3.93	3.34	3.01	2.79	2.63	2.41	2.17	1.88	1.58
120	5.15	3.80	3.23	2.89	2.67	2.52	2.30	2.05	1.76	1.43
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.19	1.94	1.64	1.27

Ա.5.4. Ֆիշերի $\mathcal{F}(N_1, N_2)$ բաշխման $p = \mathbf{P}\{\mathcal{F}(N_1, N_2 < F_p(N_1, N_2))\}$ հավանականությանը համապատասխանող $F_p(N_1, N_2)$ քանորդիչների աղյուսակ:

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1):$$

$p = 0.99$:

N_2 \ N_1	1	2	3	4	5	6	8	12	24	120
1	4052	5000	5403	5625	5764	5859	5982	6106	6235	6339
2	98.5	99.00	99.2	99.25	99.3	99.3	99.4	99.4	99.5	99.5
3	34.1	30.82	29.5	28.7	28.2	27.9	27.5	27.1	26.6	26.2
4	21.2	18.00	16.7	16.0	15.5	15.2	14.8	14.4	13.9	13.6
5	16.3	13.27	12.1	11.4	11.0	10.7	10.3	9.89	9.47	9.11
6	13.75	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.97
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.74
8	11.3	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.95
9	10.6	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.40
10	10.0	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	4.00
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.69
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.45
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	3.96	3.59	3.25
14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.80	3.43	3.09
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.96
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.84
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.46	3.08	2.75
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.66
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.58
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.52
25	7.77	5.57	4.68	4.18	3.85	3.63	3.22	2.99	2.62	2.27
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.11
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.92
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.73
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.53
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.32

Ա.5.5. Ֆիշերի $\mathcal{F}(N_1, N_2)$ բաշխման $p = \mathbf{P}\{\mathcal{F}(N_1, N_2 < F_p(N_1, N_2))\}$ հավանականությանը համապատասխանող $F_p(N_1, N_2)$ քանորդիչների աղյուսակ:

$$F_{1-p}(N_1, N_2) = 1/F_p(N_2, N_1):$$

$$p = 0.995:$$

N_2 \ N_1	1	2	3	4	5	6	8	12	24	120
1	16211	20000	21615	22500	23056	23437	23925	24426	24940	25359
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.13	43.49	42.62	41.99
4	31.33	26.28	24.26	23.15	22.46	21.97	21.35	20.70	20.03	19.47
5	22.78	18.31	16.53	15.56	14.94	14.51	13.96	13.38	12.78	12.27
6	18.63	14.54	12.92	12.03	11.46	11.07	10.57	10.03	9.47	9.00
7	16.24	12.40	10.88	10.05	9.52	9.16	8.68	8.18	7.65	7.19
8	14.69	11.04	9.60	8.81	8.30	7.95	7.50	7.01	6.50	6.06
9	13.61	10.11	8.72	7.96	7.47	7.13	6.69	6.23	5.73	5.30
10	12.83	9.43	8.08	7.34	6.87	6.54	6.12	5.66	5.17	4.75
11	12.23	8.91	7.60	6.88	6.42	6.10	5.68	5.24	4.76	4.34
12	11.75	8.51	7.23	6.52	6.07	5.76	5.35	4.91	4.43	4.01
13	11.37	8.19	6.93	6.23	5.79	5.48	5.08	4.64	4.17	3.76
14	11.06	7.92	6.68	6.00	5.56	5.26	4.86	4.43	3.96	3.55
15	10.80	7.70	6.48	5.80	5.37	5.07	4.67	4.25	3.79	3.37
16	10.58	7.51	6.30	5.64	5.21	4.91	4.52	4.10	3.64	3.22
17	10.38	7.35	6.16	5.50	5.07	4.78	4.39	3.97	3.51	3.10
18	10.22	7.21	6.03	5.37	4.96	4.66	4.28	3.86	3.40	2.99
19	10.07	7.09	5.93	5.27	4.85	4.56	4.18	3.76	3.31	2.89
20	9.94	6.99	5.82	5.17	4.76	4.47	4.09	3.68	3.22	2.81
25	9.48	6.60	5.46	4.84	4.43	4.15	3.78	3.37	2.92	2.50
30	9.18	6.35	5.24	4.62	4.23	3.95	3.58	3.18	2.73	2.30
40	8.83	6.07	4.98	4.37	3.99	3.71	3.35	2.95	2.50	2.06
60	8.49	5.79	4.73	4.14	3.76	3.49	3.13	2.74	2.29	1.83
120	8.18	5.54	4.50	3.92	3.55	3.28	2.93	2.54	2.09	1.61
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.74	2.36	1.90	1.36

Ա.6. Ֆիշերի հակադարձ ձևափոխությունը՝ z-ի արժեքներին համապատասխանող ρ-ի արժեքները:

z	0	1	2	3	4	5	6	7	8	9
0.0	0000	0100	0200	0300	0400	0500	0599	0699	0798	0898
1	0997	1096	1194	1293	1391	1489	1586	1684	1781	1877
2	1974	2070	2165	2260	2355	2449	2543	2636	2729	2821
3	2913	3004	3095	3185	3275	3364	3452	3540	3627	3714
4	3800	3885	3969	4053	4136	4219	4301	4382	4452	4542
5	4621	4699	4777	4854	4930	5005	5080	5154	5227	5299
6	5370	5441	5511	5580	5649	5717	5784	5850	5915	5980
7	6044	6107	6169	6231	6291	6351	6411	6469	6527	6584
8	6640	6696	6751	6805	6858	6911	6963	7014	7064	7114
9	7163	7211	7259	7306	7352	7398	7443	7487	7531	7574
1.0	7616	7658	7699	7739	7779	7818	7857	7895	7932	7969
1	8005	8041	8076	8110	8144	8178	8210	8243	8275	8306
2	8337	8367	8397	8426	8455	8483	8511	8538	8565	8591
3	8617	8643	8668	8692	8717	8741	8764	8787	8810	8832
4	8854	8875	8896	8917	8937	8957	8977	8996	9015	9033
5	9051	9069	9087	9104	9121	9138	9154	9170	9186	9201
6	9217	9232	9246	9261	9275	9289	9302	9316	9329	9341
7	9354	9366	9379	9391	9402	9414	9425	9436	9447	9358
8	9468	9478	9488	9498	9508	9518	9527	9536	9545	9554
9	9562	9571	9579	9587	9595	9603	9611	9618	9626	9633
2.0	9640	9647	9654	9661	9668	9674	9680	9686	9693	9699
1	9704	9710	9716	9722	9727	9732	9738	9743	9748	9753
2	9757	9762	9767	9771	9776	9780	9785	9789	9793	9797
3	9801	9805	9809	9812	9816	9820	9823	9827	9830	9834
4	9837	9840	9843	9846	9849	9852	9855	9858	9861	9864
5	9866	9869	9871	9874	9876	9879	9881	9884	9886	9888
6	9890	9892	9894	9897	9899	9901	9903	9904	9906	9908
7	9910	9912	9914	9915	9917	9919	9920	9922	9923	9925
8	9926	9928	9929	9931	9932	9933	9935	9936	9937	9938
9	9940	9941	9942	9943	9944	9945	9946	9947	9948	9949

ՀԱՎԱՆԱԿԱՆՈՒԹՅԱՆ ՏԵՍՈՒԹՅԱՆ, ՄԱԹԵՄԱՏԻԿԱԿԱՆ ՎԻՃԱԿԱԳՐՈՒԹՅԱՆ և ՈՐՈՇ ԸՆԴՀԱՆՈՒՐ ՏԵՐՄԻՆՆԵՐ

տերմինը և նրա ածանցյալները	բացատրություն	անգլերեն տերմինը	ռուսերեն տերմինը
անշեղություն, չշեղված լինելը	վիճակագրական գնահատման չշեղված լինելու հատկություն (սխալ է՝ «անշեղելի»)	unbiasedness	несмещенность
անհատ, հանուրի տարր	տիեզերքի տարր	element of a population, population member	индивид(уум), элемент совокупности
անհամաչափություն		asymmetry	асимметрия
անչափում մեծություն	«չեզոք», չափողականություն չունեցող	dimensionless, nondimensional variable	безразмерная величина
աստիճանավորում		gradation	градация
աստիճանանիշ		grad	градация
անփոփոխ		invariant	инвариантный
անփոփոխություն		invariance	инвариантность
բաշխման անհա-մաչափության, թեքվածության գործակից		coefficient of asymmetry, of ditribution, coefficient of skewness	коэффициент асимметрии распределения, коэффициент скошенности
անցումային վիճակ	Մարկովի շղթայի վիճակների դասակարգման տերմին	transient state	переходное состояние
բաղադրույթ		superposition	суперпозиция
հավանականության բաշխման օրենք (հակիրճ՝ հավանակա-նության բաշխում, օրենք, բաշխում)		probability distribution, law	закон распре-деления, рас-пределение вероятностей
համատեղ բաշխում		joint distribution, simultaneous distribution	совместное распределение
մասնատեղ, լուսանցքային բաշխում		marginal distribution	маргинальное распределение
նմուշային բաշխում		sampling distribution	выборочное распределение
նմուշային բաշխման ֆունկցիա		sample distribution function	выборочная функция распределения
հավանականության մասնատեղ (լուսանցքային) խտություն		marginal probability density	маргинальная плотность вероятности
երկանդամային օրենք, երկանդամային բաշխում		binomial distribution	биномиальное распределение
նորմալ բաշխում, Գաուսի օրենք, գաուսյան բաշխում		normal distribution, Gauss distribution	нормальное, (гауссовское) распределение, распределение Гаусса

կանոնական նորմալ բաշխում		standard normal distribution	стандартное нормальное распределение
գնատու	ֆունկցիա փորձի տվյալներից (նմուշից), որը տալիս է որոշակի հատկանիշի գնահատումը	estimator, estimating function	оценка (оценочная функция)
անշեղ գնատու	սխալ է՝ «անշեղելի»	unbiased estimator	несмещенная оценка
շեղ գնատու		biased estimator	смещенная оценка
ունակ գնատու		consistent estimator	состоятельная оценка
ուղղված գնատու		corrected estimator	исправленная оценка
գնահատական	գնատուի տված արժեքը որոշակի փորձնական տվյալների դեպքում	estimate	оценка (значение оценочной функции)
բավարար գնատու, գնահատական		sufficient exhaustive, estimator, estimate	достаточная оценка
կետային գնահատական (գնատու)		point estimate (estimator)	точечная оценка
միջակայքային գնահատական		interval estimate	интервальная оценка
առավել(ագույն) ճշմարտանմանության գնատու		maximum likelihood estimator	оценка максимального правдоподобия
դասակարգում, խմբավորում	օբյեկտների տարանջատում (դասակարգում, խմբավորում) փորձնական տվյալների հիման վրա	grouping, classification	классификация, группировка
եղանակ, մեթոդ		method	метод
մոմենտների եղանակ		method of moments	метод моментов
լայնք		range	размах
ծրագրաշար		package	пакет программ
կանոնական (ստանդարտ) շեղում		standard deviation	стандартное, среднеквадратичное отклонение
կանոնական		standard	стандарт
կանոնականում		standardization	стандартизация
կենտրոնացված մեծություն		centred variable	центрированная величина
կիսող		median	медиана
կորուստների ֆունկցիա		risk function, loss function	функция риска, функция потерь
կշռավորված, կշռյալ		weighed, weighted	взвешенный
համույթ		population, universe, parent population	генеральная совокупность
հավանականության տեսություն	գերադասելի է, քան «հավանականությունների» տեսություն	probability theory	теория вероятностей
հիստոգրամ, սյունապատկեր		histogram	гистограмма

հարյուրորդիչ		centile	центиль
համակցություն		combination	комбинация
համակցական		combinatorial	комбинаторный
համակցաբանություն		combinatorics	комбинаторика
համացրվածք		covariance	ковариация
համացրվածքային վերլուծություն		analysis of covariance	ковариацион-ный анализ
համացրվածքային մատրից		covariance matrix	ковариационная матрица
համացրիվություն	ցրվածքների հավասարություն	homoscedasticity	гомоскедастич-ность
(համա)**հարաբերակ-ցություն**, կոռելյացիա		correlation	корреляция
չհարաբերակցված, չկոռելյացված		uncorrelated	некоррелиро-ванный
հարաբերակցության գործակից		correlation coefficient	коэффициент корреляции
հարաբերակցային բանորդ		correlation ratio	корреляционное отношение
կարգային հարաբերակ-ցություն		rank correlation	ранговая корреляция
Սպիրմենի կարգային հարաբերակցության գործակից		Spirmen rank correlation coefficient	коэффициент ранговой корреляции Спирмена
համախություն	անհաչող է` «համախականություն»	frequency	частота, частость
հարաբերական համախություն		relative frequency	относительная частота, частость
բացարձակ համախություն		absolute frequency	абсолютная частота, частость
(բացա)**հայտիչ**, ինդիկատոր, (բազմու-թյան) հայտիչ ֆունկցիա	ֆունկցիա, որը տվյալ բազմու-թյան կետերում հավասար է 1, դրսում` 0	indicator, indicator function	индикатор, ха-рактеристичес-кая функция множества
հաշվիչ		computer	компьютер
տվյալների ներածում հաշվիչ		data input to computer	ввод данных в компьютер
տվյալների արտածում հաշվիչից		data output from computer	вывод данных из компьютера
հատկանիշ		character	признак
արդյունքային հատկանիշ		resulting property	результативный (результирую-щий) признак
որակական հատկանիշ		qualitative character	качестенный признак
քանակական հատկանիշ		quantitative	количественный признак
ճշմարտանմանություն		likelihood	правдоподобие

մաթեմատիկական մակածություն		induction mathematical	математичекая индукция
միտում		trend	тренд
մոդ		mode, modal value	мода
նմուշ	«որևէ բանի փոքր քանակություն, որը վերցվում է ուսումնասիրության համար» (Է.Աղայան, Բաց. բառարան), վիճակագրությունում՝ փորձնական տվյալների վեկտորը: Կիրառվող «ընտրանք» տերմինը սխալ է, անհաջող թարգմանություն է ռուսերենից, քանի որ ռուսերենում «выбрать» բառի «վերցնել», «հանել» իմաստն է օգտագործված	sample	выборка, выборочная совокупность
անկախ նմուշներ		independent samples	независимые выборки
մաղված նմուշ, տվյալներ		censored, trimmed sample, data	цензурированная выборка, данные
նմուշահանում	նմուշի ստացման գործողությունը	sampling	выборка, выборочный метод,
նմուշահանում անդարձ, սպառիչ		sampling without replacement	бесповторный отбор, выборка без возвращения
նմուշահանում դարձով		sampling with replacement	выборка с возвращением
ենթանմուշ		subsample	выборка из выборки
ողորկացում		smoothing	сглаживание
պատահական ընթացքների տեսություն		theory of random processes	теория случайных процессов
պատահական մեծություն		random variable	случайная величина
կենտրոնացված պատահական մեծություն		centered random variable	центрированная случайная величина
նորմավորված պատահական մեծություն		normalized random variable	нормированная случайная величина
կանոնաձև պատահական մեծություն	կենտրոնացված և նորմավորված պատահական մեծություն	standardized random variable	стандартная случайная величина
պատահականացում		randomization	рандомизация
սպասելի, միջին արժեք		expectation, expected value, mathematical expectation	математическое ожидание, среднее значение
պայմանական սպասելի		conditional expectation	условное математическое ожидание
վարկած		hypotesis	гипотеза

վարկածն ընդունել, ճիշտ համարել		accept hypotesis	принять гипотезу
վերադարձելի վիճակ		recurrent state	возвратное состояние
վարկածը չընդունել, ժխտել		reject, discard hypotesis	отвергнуть гипотезу
վիճակագրություն		statistics	статистика
նկարագրական վիճակագրություն		descriptive statistics	описательная статистика
ժողովրդագրական վիճակագրություն		demographic statistics, population statistics	демографическая статистика, статистика народонаселения
տնտեսական վիճակագրություն		economic statistics	экономическая статистика
վիճականի,	նմուշի բնութագրիչ, ֆունկցիա փորձի տվյալներից	statistic, sample characteristic	статистика, функция от выборки
բավարար վիճականի	«բավարար» լինելու հատկությամբ օժտված վիճականի	sufficient statistic, exhaustive statistic	достаточная статистика
կարգային վիճականի		rank statistic	ранговая статистика
տասնորդիչ		decile	дециль
կուտակվածության գործակից		coefficient of kurtosis, excess, flatness	коэффициент крутости, эксцесс, куртозис
ցրվածության բնութագրիչներ		measures of dispersion	характеристика вариации, рассеяния
վստահության հավանականություն	սխալ է՝ «վստահելի»	confidence, fiducial probability, confidence level	доверительная вероятность
վստահության միջակայք		confidence interval, region	доверительный интервал
տեստ, հայտանիշ		criterion, test	критерий, тест
կարգային տեստ, հայտանիշ		rank test	ранговый критерий
տեստում, ստուգում		testing	проверка, тестирование
դիրքի հայտանիշ		test of location	критерий смещения
համաձայնության հայտանիշ		goodness-of-fit test	критерий согласия
ամենահզոր հայտանիշ		most powerful test	наиболее мощный критерий
կարգերի գումարի հայտանիշ		rank sum test	критерий суммы рангов
խի-քառակուսի տեստ, հայտանիշ		chi-squared test	критерий хи-квадрат
տարակարգ, կարգ		rank	ранг
տարակարգերի վեկտոր		rank vector	вектор рангов
կարգային, տարակարգային		rank	ранговый

անտեսաչափություն		econometrics	эконометрика
տարացրիվություն	ցրվածքների անհամասեռություն	heteroscedasticity	гетероскедастичность
զուգակցության աղյուսակ		contingency table	таблица сопряженности
համադրման բանաձև		compozition formula	формула композиции
ցրվածք		variance	дисперсия
նմուշային ցրվածք		sample variance	выборочная дисперсия
ցրվածքային վերլուծություն		analysis of variance, ANOVA	дисперсионный анализ
ունակություն, զուգամիտություն (ըստ հավանականության)		consistency	состоятельность
զուգամիտություն ըստ օրենքի		convergence in distribution	сходимость по распределению
չզուգամիտող		inconsistent	несостоятельный
փորձ		experiment, trial	эксперимент, опыт
փոփոխման շարք		variational series	вариационный ряд
փոփոխման գործակից		variation coefficient	коэффициент вариации
քառորդիչ		quartile	квартиль
քանորդիչ		quantile	квантиль
A-ի հավանականությունը	անհաջող է ասել «այն բանի հավանականությունը, որ A»	probability of A	вероятность того, что A
A-ի ըստ B-ի պայմանական հավանականությունը		conditional probability of A upon B	условная вероятность A при условии B

ՕԳՏԱԳՈՐԾՎԱԾ ԲԱՌԱՐԱՆՆԵՐԻ ՑԱՆԿ

Աղայան Է. Արդի հայերենի բացատրական բառարան, Հ. 1, 2, Երևան, Հայաստան, 1976:

Անգլերեն-հայերեն բառարան, խմբ. Ասմանգուլյանի Հ. Ա. և Հովհաննիսյան Մ. Ի., Հայաստան, 1984:

Գրիգորյան Ս. Ա. Մաթեմատիկական տերմինների ռուս-հայերեն բառարան, Երևան, Լույս, 1996:

Ժամանակակից հայոց լեզվի բացատրական բառարան, Հ. 1-4, Երևան, ՀՍՍՀ ԳԱ, 1969-1980:

Ղարիբյան Ա. Ս., Տեր-Մինասյան Է. Գ., Գևորգյան Մ. Ս. Հայ-ռուսերեն բառարան, Երևան, Հայաստանի պետական հրատարակչություն, 1960:

Ռադիոէլեկտրոնիկայի ռուս-հայերեն բառարան, ընդհ. խմբագիր Գասպարյան Ռ. Բ. Երևան, Լույս, 1982:

Ռուս-հայերեն բառարան չորս հատորով, Երևան, ՀՍՍՀ ԳԱ, 1954-1958:

Սուքիասյան Ա. Մ. Հայոց լեզվի հոմանիշների բառարան, Երևան, ՀՍՍՀ ԳԱ, 1967:

Տերմինաբանական և ուղղագրական տեղեկատու, Երևան, Երևանի համալսարան, 1984:

Տոնյան Ա. Հ., Տոնյան Վ. Ա. Մաթեմատիկական տերմինների բառարան անգլերեն, ռուսերեն, հայերեն, գերմաներեն, ֆրանսերեն լեզուներով, Երևան, ՀՍՍՀ ԳԱ, 1965:

Hebak P., Hustopecky J. Sestijazycny slovnik terminu z matematicke statistiky, Praha, 1979.

Lohwater A. J. Russian-English dictionary of themathematical sciences, Second Edition, American Mathematical Society, Providence, 1990.

Rudler G., Anderson N. French-English, English-French dictionary, London, Collins, 1975.

Англо-русский словарь математических терминов, под редакцией Александрова П. С., Издание второе, Москва, ИЛ, 1994.

Боровков К. А. Англо-русский, русско-английский словарь по теории вероятностей, статистике и комбинаторике, Москва, ТВП, SIAM, Philadelphia, 1994.

Немецко-русский математический словарь, под редакцией Погребысского И. Б., Москва, Советская энциклопедия, 1968.

Французско-русский математический словарь, под редакцией Розова Н. Х., Москва, Советская энциклопедия, 1970.

Այբբենական բառացույց

Եվգենի Հարությունյան, Տաթևիկ Ղազանչյան,
Նաիրա Մեսրոպյան, Դավիթ Աստրյան, Մարիամ Հարությունյան,
Մելս Սահակյան, Հարություն Շահումյան

Հավանականություն և կիրառական վիճակագրություն: Տնտեսագետների ու գործարարների համար

Խմբագիրներ՝
Ռուբեն Համբարձումյան, Բորիս Նահապետյան,
Եվգենի Հարությունյան, Մամիկոն Գինովյան,
Ալեքսան Սիմոնյան, Դավիթ Աստրյան,
Աշոտ Հարությունյան

Պատվեր 8 Տպաքանակ 1500

Տպագր. 37.25 մամուլ: Չափսը՝ 60×84 1/8 : Թուղթ՝ տպագրական 1:
Տպագրված է ՀՀ ԳԱԱ Հայոց ցեղասպանության թանգարան-ինստիտուտի տպարանում:
Հասցեն՝ Երևան-1, Արովյան 15: