АНДРЕЙ НИКОЛАЕВИЧ
КОЛМОГОРОВ
60-е годы

А. Н. КОЛМОГОРОВ

ТЕОРИЯ ИНФОРМАЦИИ И ТЕОРИЯ АЛГОРИТМОВ

Ответственный редактор
академик
Ю. В. ПРОХОРОВ

МОСКВА
«НАУКА»
1987

УДК 519.72 + 510.5

Колмогоров А. Н. **Теория информации и теория алгоритмов.**— М.: Наука, 1987.—304 с.

В настоящую третью книгу избранных трудов академика А. Н. Колмогорова включены работы по теории информации и теории алгоритмов и их приложениям к различным областям знания. Комментарии специалистов дают представление о развитии работ А. Н. Колмогорова и современном состоянии рассматриваемых в них проблем.

Составитель

А. Н. ШИРЯЕВ

Рецензенты:

С. И. АДЯН, Ю. И. МАНИН

Андрей Николаевич Колмогоров

ТЕОРИЯ ИНФОРМАЦИИ И ТЕОРИЯ АЛГОРИТМОВ

Утверждено к печати Отделением математики Академии наук СССР

Редактор В. И. Битюцков. Редактор издательства Н. Н. Левнова.
Художник Л. С. Эрман. Художественный редактор С. А. Литвак.
Технический редактор З. Б. Павлюк. Корректоры Л. В. Лукичева, И. А. Талалай

ИБ № 35678

Сдано в набор 29.09.86. Подписано к печати 24.02.87. Т-16114. Формат 60×90 1/16
Бумага книжно-журнальная. Гарнитура обыкновенная новая. Печать высокая
Усл. печ. л. 19,12. Усл. кр. отт. 20,12. Уч.-изд. л. 18,9. Тираж 5700 экз. Тип. зак. 3258

Цена 1 р. 60 к.

Ордена Трудового Красного Знамени издательство «Наука»
117864, ГСП-7, Москва, В-485, Профсоюзная ул., 90

2-я типография издательства «Наука» 121099, Москва, Г-99, Шубинский пер., 6

К $\frac{1501000000-085}{042(02)-87}$ 114-87-II

ОТ РЕДАКЦИИ

В соответствии с постановлением Президиума АН СССР первая книга избранных трудов академика А. Н. Колмогорова «Математика и механика» (М.: Наука) вышла в свет в 1985 г., вторая книга «Теория вероятностей и математическая статистика» (М.: Наука) — в 1986 г. В настоящую третью книгу включены труды А. Н. Колмогорова по теории информации и теории алгоритмов. Кроме того, в книгу помещены найденные недавно в архивах автора его студенческий «Доклад математическому кружку о квадрильяже» и вторая часть его работы «Об операциях над множествами» (ее первая часть опубликована в первой книге избранных трудов).

Подготовка третьей книги избранных трудов А. Н. Колмогорова осуществлена Ю. В. Прохоровым и А. Н. Ширяевым.

ПРИВЕТСТВИЕ А. Н. КОЛМОГОРОВУ ОТ МОСКОВСКОГО МАТЕМАТИЧЕСКОГО ОБЩЕСТВА*

Дорогой Андрей Николаевич!

Правление Московского математического общества, члены Общества сердечно поздравляют Вас с 80-летним юбилеем. Вся Ваша научная жизнь самым тесным образом связана с Московским математическим обществом. Вы впервые выступили на заседании Общества 8 октября 1922 г. с «Примером ряда Фурье—Лебега, расходящегося почти всюду», сделавшего Вас, тогда 19-летнего студента, сразу известным всему математическому миру. Начиная с этого момента, Вы выступали на заседаниях Общества с большими докладами на самые разнообразные темы 98 раз.

Список Ваших выступлений поражает своей широтой. Достаточно перечислить лишь названия некоторых Ваших докладов:

«О всюду расходящемся ряде Фурье» (16.11.1926);

«Об одной общей схеме теории вероятностей» (20.03.1928);

«Новая интерпретация интуиционистской логики» (18.12.1928);

«О геометрических идеях Plücker'a и Klein'a» (11.12.1932);

«Цепи Маркова и обратимость законов природы» (5.01.1935);

«Статистическая теория кристаллизации застывших металлов» (16.02.1937);

«Современные вопросы теоретико-множественной геометрии» (22.12.1938);

«Стационарные последовательности элементов гильбертова пространства» (4.06.1939);

«О двух видах аксиоматического метода» (18.03 и 1.04.1941);

«О мерах, инвариантных по отношению к группе преобразований» (16.04.1941);

«Унитарные представления унитарных групп» (2.02.1944);

«Математическая теория турбулентности» (3.10.1944);

«Строение полных метрических алгебр Буля» (9.12.1947);

«Наилучшее приближение комплексных функций» (24.02.1948);

«О некоторых математических задачах, связанных с контролем производства» (9.01.1951);

«Двузначные функции двузначных переменных и применение их к релейно-контактным схемам» (27.11.1951);

«О спектрах динамических систем на торе» (30.09.1952);

«О почти периодических движениях твердого тела вокруг неподвижной точки» (26.05.1953);

«Оценки минимального числа элементов ε-сетей в различных функциональных классах и их применение к вопросу о представи-

---

* УМН, 1983, т. 38, вып. 4, с. 3—5.

мости функций нескольких переменных суперпозициями функций меньшего числа переменных» (27.04.1954);

«О некоторых асимптотических характеристиках вполне ограниченного метрического пространства» (5.06.1956);

«Равномерные предельные теоремы для сумм независимых слагаемых» (18.12.1956);

«Малые знаменатели в задачах механики и анализа» (13.01. 1956);

«Что такое информация» (4.06.1961);

«Самоконструирующиеся аппараты» (21.11.1961);

«Вычислимые функции и основания теории информации и теории вероятностей» (19.11.1963);

«Эксперимент и математическая теория в изучении турбулентности» (18.05.1965);

«Статистическая гидродинамика океана» (24.02.1970);

«Сложность задания и сложность построения математических объектов» (23.11.1971);

«О статистических решениях уравнений Навье—Стокса» (18.01. 1978).

Вы не раз выступали с обзорными докладами, которые послужили источником идей дальнейших исследований:

«Меры и распределения вероятностей в функциональных пространствах» (30.11.1948);

«Решенные и нерешенные задачи, связанные с 13-й проблемой Гильберта» (17.05.1960);

«Математические методы исследования русского стиха» (27.12. 1960);

«О равномерных предельных теоремах для сумм независимых переменных» (26.03.1963);

«Сложность алгоритмов и объективное определение случайности» (16.04.1974).

Андрей Николаевич, Вы были инициатором и докладчиком многочисленных заседаний Общества, посвященных обсуждению проектов статей, подготовленных для «Большой Советской Энциклопедии». В 1938 г. (10 марта) было проведено заседание, посвященное обсуждению Вашей знаменитой энциклопедической статьи «Математика». И еще несколько раз Вы выступали с проектами Ваших статей для «Большой Советской Энциклопедии»: «Аксиома», «Бесконечно малая» (18 и 19.10.1949) и др.

С большим интересом проходили заседания Общества с Вашими докладами:

«Развитие в СССР математических методов познания природы» (10.11.1937);

«О критике Остроградского на работы Лобачевского» (29.09. 1948);

«Развитие математики в советскую эпоху» (23.12.1949);

«Элементы математики Николая Бурбаки» (29.05.1956);

«О некоторых чертах современного этапа развития математики» (21.04.1959);

«Из опыта работы» (25.04.1963);

«О Международном математическом конгрессе в Амстердаме» (28.09.1954);

«О научной командировке во Францию, ГДР и Польшу» (28.02. 1956).

Вы всегда придавали большое значение обсуждению широкой математической общественностью вопросов школьного образования.

22 и 28 ноября 1937 г. было проведено заседание Общества, посвященное обсуждению составленного Вами и П. С. Александровым плана нового учебника элементарной алгебры.

Вами сделаны следующие доклады:

«О проекте программы по математике для средней школы» (3.03.1946);

«О проекте программ для средней школы» (17.02.1948);

«О факультативных занятиях по математике в средней школе в 1967/68 учебных годах» (14.02.1967).

Дорогой Андрей Николаевич, Вы — член Математического общества с 1 февраля 1930 г., почетный член — с 28 апреля 1953 г., сейчас Вы — президент нашего Общества, и можно смело сказать, что вместе с П. С. Александровым Вы определили лицо современного Московского математического общества.

Ваш юбилей — это праздник для всех нас. Примите, дорогой Андрей Николаевич, наши чувства глубокого к Вам уважения и преклонения перед Вами.

Доброго Вам здоровья и радостей!

Правление Московского математического общества

АНДРЕЙ НИКОЛАЕВИЧ КОЛМОГОРОВ*
(К восьмидесятилетию со дня рождения)

Н. Н. Боголюбов, Б. В. Гнеденко, С. Л. Соболев

25 апреля 1983 г. исполнилось восемьдесят лет со дня рождения Андрея Николаевича Колмогорова. В приветствии Отделения математики АН СССР, Московского математического общества и редакционной коллегии журнала «Успехи математических наук» к семидесятипятилетнему юбилею Андрея Николаевича было сказано: «Вы внесли выдающийся вклад в развитие современной математики и ее приложений. Ваши фундаментальные исследования определили лицо многих областей математики XX в. Теория тригонометрических рядов, теория меры, теория множеств, теория интеграла, конструктивная логика, топология, теория приближений, теория вероятностей, теория случайных процессов, теория информации, математическая статистика, динамические системы, конечные автоматы, теория алгоритмов, математическая лингвистика, теория турбулентности, небесная механика, дифференциальные уравнения, 13-я проблема Гильберта, теория стрельбы, приложение математики к проблемам биологии, геологии, кристаллизации металлов — вот неполный перечень областей математики, ее приложений, которые Вы обогатили своими глубокими идеями. Вы всегда придавали большое значение приложениям математики и ее связям с другими науками. Исключительная широта Ваших научных интересов, дар педагога всегда привлекали к Вам талантливую молодежь. Вами создана всемирно известная научная школа. Представители этой школы успешно работают в самых различных областях математики и других естественных наук» [1].

В приветствии к семидесятилетнему юбилею Андрея Николаевича от Президиума АН СССР, в частности, отмечалось: «Большой вклад в распространение математических знаний Вы внесли, возглавляя Отдел математики в «Большой Советской Энциклопедии». Вы создали большую научную школу, из которой вышли многие крупные ученые. Вы принимаете активное участие в деятельности Отделения математики Академии наук СССР. Общеизвестны Ваши заслуги в деле развития высшего образования в нашей стране...» [2].

А. Н. Колмогоров родился в Тамбове. Его отец был агрономом. Мать Андрея Николаевича скончалась при его рождении. Заботы

---

* УМН, 1983, т. 38, вып. 4, с. 11—23.
[1] УМН, 1978, т. 33, вып. 2, с. 212—213.
[2] УМН, 1973, т. 28, вып. 5, с. 3.

по уходу за ребенком взяла на себя сестра матери Вера Яковлевна, независимая женщина, обладавшая высокими общественными идеалами. Эти идеалы она передала племяннику, воспитав в нем чувство ответственности, самостоятельность суждений, нетерпимость к безделию и плохо выполненному поручению, стремление понимать, а не только запоминать. Он рано начал работать и перед поступлением в Московский университет некоторое время служил проводником на железной дороге.

Осенью 1920 г. Андрей Николаевич поступил в Московский университет, но определил себя как математика не сразу. Долгое время наряду с математикой он увлекался русской историей, принимал активное участие в семинаре профессора С. В. Бахрушина. Им было выполнено серьезное научное исследование по писцовым книгам XV—XVI вв. о земельных отношениях в древнем Новгороде. Известно, что Андрей Николаевич высказывал в 20-х годах гипотезу о путях заселения верховьев Пинеги и что эта гипотеза Колмогорова нашла подтверждение в ходе организованной в эти места экспедиции П. С. Кузнецова.

Интересы А. Н. Колмогорова в области гуманитарных наук с новой силой и новыми возможностями вспыхнули в 60-е годы, когда Андрей Николаевич уделил много сил и внимания вопросам стихосложения и применению к изучению стиха математических методов; он писал статьи, привлекал к работе молодежь, участвовал в научных конференциях и выступал с докладами на эти темы. Систематическая работа началась в 1960 г., когда А. Н. Колмогоров осуществил свой давний замысел и развернул на кафедре теории вероятностей и в только что им созданной проблемной статистической лаборатории исследования по математическим методам в языкознании. Эти исследования сразу вызвали широкий интерес и доверие среди специалистов в области языка и литературы. И уже в 1961 г. на IV Всесоюзном математическом съезде в Ленинграде и затем на совещании в Горьком, посвященном применению математических методов в изучении языка художественных произведений, молодые сотрудники Андрея Николаевича докладывали первые результаты по вероятностным методам стиховедения и определению энтропии русской речи путем вероятностного эксперимента.

Но вернемся к юношеским годам Андрея Николаевича. Большое влияние на формирование А. Н. Колмогорова как математика оказал семинар В. В. Степанова по тригонометрическим рядам. На семинаре перед участниками возникали нерешенные вопросы, ответы на которые казались насущными.

В 1922 г., после того как А. Н. Колмогоровым была выполнена его первая самостоятельная работа (о порядке величины коэффициентов Фурье), он стал учеником Н. Н. Лузина.

С участия в семинаре В. В. Степанова и последовавших затем занятий под руководством Н. Н. Лузина начался первый творческий период в жизни Колмогорова-математика. Этот период был характерен тем, что постановки задач и направленность мысли Колмогорова находились в русле основных устремлений Московской математической школы того времени.

В частности, под влиянием работ М. Я. Суслина, Н. Н. Лузина и П. С. Александрова у А. Н. Колмогорова возникло желание разработать общую теорию операций над множествами. Уже весной 1922 г. А. Н. Колмогоров завершил большое исследование по теории операций над множествами. Им был введен очень широкий класс операций над множествами — δs-операции. По не зависящим от автора причинам эта работа была опубликована лишь в 1928 г. Она привлекла внимание исследователей, и заложенная в ней идея развивалась в дальнейшем рядом ученых — Л. В. Канторовичем, А. А. Ляпуновым и другими.

В июне 1922 г. А. Н. Колмогоров построил пример ряда Фурье—Лебега, расходящегося почти всюду, вслед за тем — ряда Фурье—Лебега, расходящегося в каждой точке. Оба эти примера были для специалистов полной неожиданностью и произвели огромное впечатление. В теорию функций действительного переменного вошел крупный исследователь. Интерес к теории рядов Фурье и теории ортогональных функций Андрей Николаевич сохранил на всю жизнь, время от времени возвращаясь к задачам из этой области исследований и передавая молодежи ряд вопросов для разработки.

В теории функций А. Н. Колмогорова интересовали все основные проблемы — вопросы дифференцирования, интегрирования, теории меры, теории приближения функций и т. д. В каждый из изученных им вопросов он вносил элемент существенной новизны.

К 1924 г. относится возникновение интереса у А. Н. Колмогорова к теории вероятностей — к той области науки, в которой его авторитет особенно велик. Первый шаг в этой новой для него области исследований был сделан совместно с А. Я. Хинчиным. Им удалось найти необходимые и достаточные условия сходимости рядов, члены которых являются взаимно независимыми случайными величинами. Значение этой работы не исчерпывается тем, что в ней было получено полное решение важной задачи; в ней были созданы основы метода, который затем многократно и успешно использовался для решения весьма разнообразных вопросов. В 1928 г. А. Н. Колмогорову удалось указать необходимые и достаточные условия для закона больших чисел — задача, в решение которой значительный вклад внесли П. Л. Чебышев и А. А. Марков (старший). Через год появилась большая работа, в которой был доказан закон повторного логарифма для сумм независимых случайных величин при весьма широких условиях, наложенных на слагаемые. За несколько лет перед этим закон повторного логарифма был открыт А. Я. Хинчиным сначала для схемы Бернулли, а затем перенесен на схему Пуассона. В том же году был получен замечательный результат — широкие условия применимости усиленного закона больших чисел. Несколько позднее эти условия послужили А. Н. Колмогорову основой для получения необходимого и достаточного условия для усиленного закона больших чисел в случае одинаково распределенных независимых слагаемых. Это условие исключительно просто — существование конечного математического ожидания у слагаемых.

Вновь следует подчеркнуть, что во всех названных работах важны не только окончательные результаты, но и методы, которые впоследствии вошли в арсенал, используемый современной теорией вероятностей. В частности, было доказано знаменитое «неравенство Колмогорова», обобщающее классическое «неравенство Чебышева».

Специального упоминания заслуживает небольшая заметка Андрея Николаевича «Общая теория меры и исчисление вероятностей», в которой был предложен первый набросок аксиоматики теории вероятностей, основанный на базе теории меры и теории функций действительного переменного. Сам по себе этот подход не был абсолютно нов (поскольку был намечен Э. Борелем и А. Ломницким), но именно у А. Н. Колмогорова он получил завершение в виде получивших всеобщее признание простых и четких математических формулировок. Э. Борель еще в 1909 г. высказал общие соображения о важности теории меры для построения фундамента теории вероятностей. В 1923 г. эти общие идеи получили развитие в большой статье А. Ломницкого, а в 1929 г. стали предметом изучения А. Н. Колмогорова. Через четыре года в 1933 г. первичная идея Э. Бореля приняла окончательную форму в классической монографии А. Н. Колмогорова «Основные понятия теории вероятностей», изданной на немецком языке в издательстве Шпрингера. Через четыре года она появилась в русском переводе и вскоре же стала библиографической редкостью (в 1974 г. вышло второе русское издание). Эта монография не только определила новый этап в развитии теории вероятностей как математической науки, но и дала необходимые основополагающие данные для построения теории случайных процессов.

В 1931 г. вышла замечательная работа А. Н. Колмогорова «Аналитические методы в теории вероятностей». В ней были заложены основы современной теории марковских случайных процессов и вскрыты глубокие связи теории вероятностей с теорией дифференциальных уравнений, как обыкновенных, так и в частных производных второго порядка. Отрывочные результаты, связывающие задачи теории цепей Маркова и марковские блуждания на прямой с уравнениями параболического типа, были получены давно еще Лапласом, а позднее Фоккером и Планком. Но это были лишь примеры. Настоящую теорию марковских процессов, или, как говорил А. Н. Колмогоров, процессов без последействия, заложил Андрей Николаевич в этой работе. В ней были получены как прямые, так и обратные уравнения. Появилась новая область математики с многочисленными выходами в практику. Она была подхвачена представителями физики и биологии, химии и инженерного дела и быстро превратилась в одно из самых мощных математических орудий современного естествознания.

Большое впечатление произвела статья «Математика», написанная А. Н. Колмогоровым для 2-го издания «Большой Советской Энциклопедии». В этой своей работе А. Н. Колмогоров в сжатой форме и на принципиальной основе проследил историческое развитие математики, указал узловые моменты этого развития и предложил для него оригинальную схему периодизации. Жаль, что эта статья не была у нас издана отдельной книгой (вероятно, это еще не поздно сделать; статьи А. Н. Колмогорова для БСЭ вообще заслуживают специального внимания и издания). Но и без этого статья «Математика» оказала серьезное влияние на направление исследований в области истории математики. И не только в нашей стране. Кстати сказать, в ГДР отдельное издание этой работы было осуществлено.

Мысль А. Н. Колмогорова работала в самых разнообразных направлениях, выдвигая новые проблемы и давая решение принципиальных вопросов. За десятилетие, непосредственно предшествующее Великой Отечественной войне, им было опубликовано более шестидесяти работ, которые относились к теории вероятностей, проективной геометрии, математической статистике, теории функций действительного переменного, топологии, математической логике, математической биологии, философии и истории математики. Среди работ этого десятилетия были такие, которые закладывали фундамент новых направлений математической мысли в теории вероятностей, в топологии, в теории приближения функций.

Именно в этот период А. Н. Колмогоров ввел в топологию понятие верхнего граничного, или ∇-оператора, одновременно с американским математиком Александером и независимо от него. С помощью этого оператора он построил теорию когомологических групп (как тогда говорили, ∇-групп) первоначально для комплексов, а затем для любых бикомпактных пространств. Понятие когомологической группы оказалось весьма удобным и очень сильным средством для исследования многочисленных вопросов топологии, в том числе связанных с изучением непрерывных отображений. На этой базе им было построено понятие когомологического кольца — одного из важных топологических понятий. Необходимо указать также исключительно общую формулировку закона двойственности, относящегося к замкнутым множествам, расположенным в любых локально бикомпактных вполне регулярных топологических пространствах, удовлетворяющих лишь условию ацикличности в тех размерностях, о которых идет речь в самой формулировке результата.

Тогда же им был построен пример открытого отображения компакта на компакт большей размерности.

В теории вероятностей за этот же срок была создана монография «Основные понятия теории вероятностей», дано представление безгранично делимых распределений, найдено предельное распределение для максимума уклонения эмпирического распределения от истинного (критерий Колмогорова), построена теория цепей Маркова со счетным множеством состояний, найдена связь между геометрией гильбертова пространства и рядом задач теории стационарных последовательностей.

В конце 30-х годов внимание А. Н. Колмогорова все более и более стала привлекать механика турбулентности, т. е. закономерности тех часто встречающихся на практике течений жидкости и газа, которые сопровождаются беспорядочными пульсациями скорости, давления и других гидродинамических величин. Корректное математическое описание таких течений неизбежно должно быть статистическим и опираться на общее понятие случайной функции, как это было разъяснено еще в 1939 г. в первой заметке М. Д. Миллионщикова — первого представителя колмогоровской школы в теории турбулентности. Строгий статистический подход систематически использовался и самим А. Н. Колмогоровым; именно в работах А. Н. Колмогорова и его учеников теория турбулентности приобрела четкое математическое оформление в виде прикладной главы теории меры в функциональных пространствах. Однако основные работы Андрея Николаевича по механике турбулентности, появившиеся в 1941 г., имели совсем не формально-математический, а отчетливо физический характер. В основе этих работ лежало глубокое проникновение в самую суть процессов, характерное для той крайне сложной нелинейной физической системы с очень большим числом степеней свободы, какой является развитое турбулентное течение; именно глубокая физическая интуиция помогла А. Н. Колмогорову выявить в этих процессах наличие очень тонкого «локального самоподобия» (понятие, позже сыгравшее большую роль во многих разделах теоретической физики) и вывести отсюда фундаментальные количественные соотношения, имеющие характер новых законов природы. Среди таких законов — знаменитый колмогоровский «закон двух третей», обладающий, как это и свойственно фундаментальным законам природы, величественной простотой: во всяком развитом турбулентном течении средний квадрат разности скоростей в двух точках, находящихся на (не слишком малом и не слишком большом) расстоянии r, пропорционален $r^{2/3}$.

Ко времени появления первых работ А. Н. Колмогорова экспериментальные данные, позволяющие проверить предсказанные им количественные законы, отсутствовали; позже, однако, эти законы многократно сопоставлялись с данными измерений как в природных средах (атмосфера, океан), так и на разнообразных лабораторных установках и всегда оказывались выполняющимися с высокой степенью точности. (А. Н. Колмогоров делал и качественные предсказания; так, он предсказал подтвержденную экспериментом слоистую структуру океана: эффект, известный под названием «блины».) В 1961—1962 гг. А. Н. Колмогоров еще раз возвратился к результатам своих старых исследований по механике турбулентности и показал, что, отказавшись от одного принимавшегося им ранее предположения, кажущегося вполне естественным, можно получить небольшие уточнения открытых им раньше законов, весьма трудно обнаруживаемые на опыте из-за своей малости, но тем не менее в самое последнее время также надежно подтвержденные рядом экспериментов.

Возвращение к прежней тематике характерно для творчества А. Н. Колмогорова в целом: он может лишь открывать для себя новые и новые области, но не может покидать их окончательно. Эта черта особенно заметна в послевоенный период, когда Андрей Николаевич снова возвращается и к турбулентности, и к функциям действительного переменного (в связи с 13-й проблемой Гильберта), и к логическим основам математики (в применении к геометрии, теории вероятностей, теории информации). При том что Андрей Николаевич занимается необычайно широким спектром тем (здесь и классическая механика, и эргодическая теория, и теория функций, и теория информации, и теория алгоритмов), этот спектр непрерывен, его темы, на первый взгляд отдаленные, оказываются связанными между собой совершенно неожиданными, найденными А. Н. Колмогоровым связями. Число работ по каждой теме обычно невелико, но все они фундаментальны — и не только по силе результата, но и по влиянию на все дальнейшее развитие соответствующей области науки.

Для иллюстрации сказанного достаточно обратиться к работам Андрея Николаевича по теории динамических систем, в которых сказалось все своеобразие его дарования. Этих работ немного, но они значительны тем, что в них открыты новые стороны теории, на которых выросли целые большие направления. Работы эти распадаются на два цикла, один из которых произошел из задач классической механики, а другой — из проблематики теории информации.

Работы первого цикла посвящены общей теории гамильтоновых систем и относятся к краткому двухлетнему периоду 1953—1954 гг. По итогам этих работ А. Н. Колмогоров сделал большой обзорный доклад на Международном математическом конгрессе в Амстердаме в 1954 г. Задача об эволюции орбит в проблеме трех и более тел восходит к Ньютону и Лапласу; в случае малых масс планет эта задача является частным случаем задачи о поведении квазипериодических движений гамильтоновых систем при малом изменении функции Гамильтона. Именно эту последнюю задачу Пуанкаре назвал «основной проблемой динамики». В работах

А. Н. Колмогорова эта проблема оказалась решенной для большинства начальных условий в случае общего положения. Приложение теории Колмогорова к различным конкретным задачам дало возможность впоследствии решить множество проблем, десятилетиями ждавших своего решения. Например, из теоремы А. Н. Колмогорова вытекает устойчивость движения астероида в плоской ограниченной круговой задаче трех тел и устойчивость быстрого вращения тяжелого несимметричного твердого тела. Теорема Колмогорова применяется также при изучении магнитных силовых линий, что имеет большое значение в физике плазмы.

Исключительно плодотворным оказался сам метод получения этих результатов. Впоследствии этот метод был усовершенствован учеником Андрея Николаевича В. И. Арнольдом и Ю. Мозером и ныне вошел в обиход под названием теории КАМ (т. е. теории Колмогорова—Арнольда—Мозера).

Работы А. Н. Колмогорова и В. И. Арнольда по теории возмущений гамильтоновых систем были отмечены Ленинской премией 1965 г.

Второй цикл работ по динамическим системам заключался в применении к эргодической теории этих систем идей теории информации.

К теории информации интересы Андрея Николаевича обращаются с 1955 г. Для распространения и популяризации в нашей стране этой теории он сделал исключительно много. В июне 1956 г. он (совместно с И. М. Гельфандом и А. М. Ягломом) выступает на эту тему на третьем Всесоюзном математическом съезде в Москве, а в октябре 1956 г. делает доклад перед Общим собранием Академии наук СССР. Напряженное сосредоточение мысли на идеях шенноновской теории информации приводит А. Н. Колмогорова к совершенно неожиданному и смелому синтезу этих идей сперва с развитыми им же в 30-е годы идеями теории приближения, а затем с идеями теории алгоритмов. Об алгоритмической теории информации мы скажем позже, а сейчас отметим плодотворность введения информационных (а именно энтропийных) характеристик в изучение и метрических пространств, и динамических систем.

В работах 1955—1956 гг. А. Н. Колмогоров вводит понятие ε-энтропии множества в метрическом пространстве и тем самым получает средство оценки «метрической массивности» функциональных классов и пространств. Используя это понятие, Андрей Николаевич дал энтропийную трактовку замечательных результатов А. Г. Витушкина о непредставимости функций n переменных гладкости r с помощью суперпозиции функций m переменных гладкости l, если $n/r > m/l$. Дело в том, что ε-энтропия какого-либо класса функций — это, грубо говоря, количество информации, позволяющее указать функцию этого класса с точностью ε. Как показал А. Н. Колмогоров, ε-энтропия естественных классов r

раз дифференцируемых функций n переменных растет как $\varepsilon^{-n/r}$ (необходимое количество информации тем больше, чем больше независимых переменных, и тем меньше, чем выше гладкость, причем существенно именно отношение размерности к гладкости). Отсюда сразу следует теорема Витушкина: ведь количество информации, необходимое для задания суперпозиции, не может превосходить (по порядку величины) количество информации, достаточное для задания составляющих суперпозицию функций меньшего числа переменных.

Эти исследования побудили А. Н. Колмогорова обратиться непосредственно к 13-й проблеме Гильберта, состоящей в том, чтобы доказать для некоторой непрерывной функции трех переменных невозможность ее представления в виде суперпозиции непрерывных функций двух переменных. А. Н. Колмогоров пришел в 1956 г. к крайне неожиданному результату: каждая непрерывная функция любого числа переменных представима в виде суперпозиции непрерывных функций трех переменных. Сама же проблема Гильберта оказалась редуцированной к некоторой задаче о представлении функций, заданных на универсальных деревьях трехмерного пространства. Эта последняя задача была решена В. И. Арнольдом в работе, выполненной под руководством А. Н. Колмогорова в 1957 г., причем решена с получением ответа, опровергающего гипотезу Гильберта: каждая непрерывная функция трех переменных оказалась представима как суперпозиция непрерывных функций двух переменных. В том же году А. Н. Колмогоров сделал решающий шаг, показав, что непрерывная функция любого числа переменных представима в виде суперпозиции непрерывных функций одного переменного и сложения.

Очень большой резонанс имело введение энтропийных характеристик в теорию динамических систем. Такое введение было осуществлено в небольшой заметке Андрея Николаевича «Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега» — заметке, открывшей второй цикл работ по динамическим системам и сыгравшей исключительную роль в развитии эргодической теории динамических систем. Дело в том, что существовавшие к тому времени метрические (т. е. не зависящие ни от каких структур, кроме меры) инварианты (спектр, перемешивание) не позволяли различить метрически даже такие системы, как простейшие (т. е. с двумя исходами) автоморфизмы Бернулли с различными вероятностями p. Заметка А. Н. Колмогорова решила этот классический вопрос: системы с разными p оказались, как и следовало ожидать, неизоморфными. Успех был обеспечен введением совершенно нового метрического инварианта, навеянного теорией информации,— энтропии динамической системы. Эта заметка открыла длинный ряд исследований у нас в стране и за рубежом, завершившийся в последние годы теоремами Орнстейна, утверждающими, что для достаточно быстро перемешивающихся динамических систем инвариант Колмогорова полностью определяет системы с точностью до метрического изоморфизма (так что, например, автоморфизмы Бернулли с равными энтропиями метрически изоморфны). Другое важное понятие, введенное в той же работе А. Н. Колмогорова,— понятие квазирегулярной или, как теперь принято говорить, K-системы («K» — в честь А. Н. Колмогорова). Это понятие играет очень важную роль при анализе классических динамических систем с сильными свойствами стохастичности. Появление таких свойств у динамических систем объясняется их внутренней динамической неустойчивостью, и принятый путь анализа этих свойств состоит в обнаружении того, что рассматриваемая система является K-системой. Нет сомнений, что широта проникновения идей эргодической теории в задачи физики, биологии, химии имеет одним из своих источников именно эту работу А. Н. Колмогорова.

В 1958—1959 гг. Андрей Николаевич объединил в одном семинаре исследования по эргодической теории динамических систем и по гидродинамической неустойчивости. Программа этого семинара предусматривала построение, хотя бы в простейших модельных примерах, эргодической теории движений жидкости, устанавливающейся после потери устойчивости ламинарным течением. Это программа применения идей эргодической теории к явлениям типа турбулентности оказала большое влияние на все последующие работы в этом направлении.

Если в 50-е годы усилия А. Н. Колмогорова направлены на использование понятий теории информации в других областях математики, то в 60-е годы он предпринимает реконструкцию самой теории информации — реконструкцию на алгоритмической основе. В двух основополагающих статьях, опубликованных в 1965 и 1969 гг. в журнале «Проблемы передачи информации», Андрей Николаевич создает новую область математики — алгоритмическую теорию информации. Центральным в этой теории является понятие сложности конечного объекта при фиксированном (алгоритмическом) способе его описания; эта сложность определяется вполне естественно — как минимальный объем описания. Теорема Колмогорова устанавливает, что среди всевозможных алгоритмических способов описания существуют оптимальные — те, для которых сложности описываемых объектов оказываются сравнительно небольшими; хотя оптимальный способ и неединствен, для заданных двух оптимальных способов соответствующие им сложности отличаются не более чем на аддитивную константу. Как всегда у А. Н. Колмогорова, новые понятия оказывались одновременно естественными, неожиданными и простыми. В рамках этих идей стало возможным, в частности, определить понятие индивидуальной случайной последовательности (что недоступно классической теории вероятностей): случайной следует признать всякую последовательность, у которой сложность (при оптимальном способе описания) ее начального отрезка растет достаточно быстро с увеличением длины отрезка.

С 1980 г. Андрей Николаевич заведует кафедрой математической логики механико-математического факультета МГУ. Математическая логика (в широком смысле, включающем и теорию алгоритмов, и основания математики) — это ранняя и поздняя его любовь. Еще в 1925 г. А. Н. Колмогоров опубликовал в «Математическом сборнике» работу о законе исключенного третьего, навсегда вошедшую в золотой фонд работ по математической логике. Это была первая отечественная публикация по математической логике, содержащая математические результаты (к тому же очень значительные), и первое в мире систематическое исследование интуиционистской логики (логики, не признающей закона исключенного третьего). В этой работе, в частности, впервые появились так называемые погружающие операции, посредством которых одни логические исчисления погружаются в другие. С помощью одной из таких операций (исторически первой, известной теперь под названием «операции Колмогорова») в рассматриваемой работе было осуществлено погружение классической логики в интуиционистскую, чем было доказано, что применение закона исключенного третьего само по себе не может привести к противоречию. А в 1932 г. Андрей Николаевич опубликовал вторую работу по интуиционистской логике, в которой для этой логики была впервые предложена семантика, свободная от философских установок интуиционизма. Именно эта работа дала возможность трактовать интуиционистскую логику как конструктивную логику.

В 1952 г. А. Н. Колмогоров предлагает наиболее общее определение конструктивного объекта и наиболее общее определение алгоритма. Значимость этих определений раскрывается в полной мере лишь сейчас. В 1954 г. он формулирует начальные понятия теории нумераций.

В 1972 г. по инициативе Андрея Николаевича на механико-математическом факультете МГУ впервые вводится обязательный курс математической логики. А. Н. Колмогоров создает действующую и поныне программу и первым читает этот курс.

И здесь мы подходим к одной из самых характерных черт биографии А. Н. Колмогорова: всю жизнь Андрей Николаевич не только развивал науку, но и учил молодежь.

Педагогическая деятельность А. Н. Колмогорова началась в 1922 г., когда он стал учителем опытно-показательной школы Наркомпроса РСФСР. В ней он проработал до 1925 г. Эта деятельность его увлекала, и он навсегда остался в душе и в жизни учителем и воспитателем. В 1931 г. Андрей Николаевич стал профессором МГУ, а с 1933 по 1939 г. был также директором НИИ математики Московского государственного университета. Много сил и инициативы вложил он в то, чтобы Институт математики университета стал одним из ведущих научных учреждений страны. Особое внимание при этом он уделял подготовке молодых ученых, старательно и с большим успехом отыскивая талантливых, увлеченных наукой молодых людей среди студентов и стажеров. Одним из учеников А. Н. Колмогорова стал пришедший на его семинар М. Д. Миллионщиков (впоследствии вице-президент АН СССР), ранее занимавшийся непосредственно практической работой. Он проходил далее аспирантуру под руководством Андрея Николаевича, а в 1938 г. поступил по его рекомендации в Институт теоретической геофизики АН СССР, возглавлявшийся в то время О. Ю. Шмидтом.

Один из авторов вспоминает семинар по теории вероятностей 1934 г., которым руководили А. Н. Колмогоров и А. Я. Хинчин. Характер проведения семинара, на котором остро обсуждались доклады, выдвигались гипотезы, вносились предложения, предлагались иные, чем у докладчика, решения, производил огромное впечатление.

Аспирантами Андрея Николаевича в ту пору были А. И. Мальцев, С. М. Никольский, Б. В. Гнеденко, И. М. Гельфанд, Г. М. Бавли, И. Я. Верченко. Их специальностями были соответственно математическая логика, функциональный анализ, теория вероятностей, снова функциональный анализ, теория вероятностей, теория функций.

Андрей Николаевич стремился создать аспирантский коллектив, находящийся в постоянном научном возбуждении, постоянном искании. И это ему удалось. Как директор НИИ математики, он систематически встречался со всеми аспирантами независимо от того, являлись ли они его учениками или нет, и выяснял, над чем работает аспирант, какие проблемы решает. И если оказывалось, что таких проблем у него нет, то предлагал подумать над той или иной задачей, которую тут же формулировал. Позднее он интересовался, как продвигается решение, и аспирант вовлекался в активную исследовательскую работу. Время аспирантуры для всех учеников А. Н. Колмогорова на всю жизнь останется незабываемым периодом в жизни, вхождения в научные искания, размышлений о назначении науки, формирования веры в неисчерпаемость творческих сил человека. Одновременно Андрей Николаевич стремился развивать и общекультурные интересы, прежде всего к изобразительному искусству, архитектуре и литературе, а также к физической культуре (сам он был превосходным лыжником и пловцом, неутомимым ходоком).

Неповторимая личность Андрея Николаевича, его многогранная эрудиция, разнообразие его научных устремлений, глубина его проникновения в тему — все это служило для его учеников хотя и недосягаемым, но чрезвычайно влекущим примером.

Совершенно незабываемы загородные прогулки, которые организовывал Андрей Николаевич для аспирантов и студентов, приглашая на них по три—пять человек, вызвавших его интерес. Эти прогулки были заполнены беседами о проблемах математики и ее применений, об архитектурных достопримечательностях мест, по которым пролегал маршрут, о выдающихся событиях культурной жизни. Но в первую очередь, конечно, были беседы о проблемах, которые предназначались для самостоятельной работы собеседников. Андрей Николаевич умеет так поставить задачу, что она заинтересовывает собеседника, становится близкой ему, вызывает прилив желания решить ее. Поражает то умение, которое проявляет Андрей Николаевич, подбирая задачу по силам молодого человека и в то же время требуя от него полного напряжения сил. Вопросы Андрея Николаевича о том, что удалось сделать, заставляли постоянно держать себя в напряжении и думать направленно. Ведь так неловко при встречах с ним было признаться, что ничего не удалось сделать, что ничего не можешь сказать нового. В этой индивидуальной работе проявлялся своеобразный и огромный педагогический талант Андрея Николаевича — воспитателя творческой молодежи. Он заставлял работать, не принуждая; прививал веру в силу разума и в необходимость научного поиска. Во время прогулки каждый из участников по нескольку раз беседовал с Андреем Николаевичем, а он менял тему разговора, держа в руках все его нити. Многие прогулки заканчивались в Комаровке, где Андрей Николаевич и Павел Сергеевич Александров угощали все общество обедом. Участники прогулки, студенты и аспиранты, усталые и переполненные впечатлениями от бесед и от полученных математических идей, возвращались пешком на станцию, чтобы уехать в Москву. Переоценить воспитательное значение этих прогулок невозможно.

Перечисленные выше лица еще оставались аспирантами, а Андрей Николаевич уже искал среди студентов способных молодых людей, готовых отдать свои силы науке. Так появились Г. Е. Шилов, М. К. Фаге, а также два аспиранта из Саратова А. М. Обухов и В. Н. Засухин (последний, защитив в 1941 г. кандидатскую диссертацию, погиб на фронте). Активное привлечение к работе учеников не прекращалось и во время войны: в 1942 г. в аспирантуру МГУ к А. Н. Колмогорову поступил А. С. Монин (вскоре после этого ушедший на фронт и вернувшийся в аспирантуру только в 1946 г.), а в 1943 г. в аспирантуру МИАН — А. М. Яглом.

После войны Андрея Николаевича окружили новые молодые математики, которых он ввел в науку. Среди них Б. А. Севастьянов, С. X. Сираждинов, М. С. Пинскер, Ю. В. Прохоров, Г. И. Баренблатт, Л. Н. Большев, Р. Л. Добрушин, Ю. Т. Медведев, В. С. Михалевич, В. А. Успенский, А. А. Боровков, В. М. Золотарев, В. М. Алексеев, Ю. К. Беляев, Л. Д. Мешалкин, В. Д. Ерохин, Ю. А. Розанов, Я. Г. Синай, В. М. Тихомиров, А. Н. Ширяев, В. И. Арнольд, Л. А. Бассалыго, Ю. П. Офман. В более позднее время к этому списку учеников А. Н. Колмогорова добавились А. В. Прохоров, М. В. Козлов, И. Г. Журбенко, А. М. Абрамов, А. В. Булинский, а также ряд зарубежных математиков, в том числе известный шведский специалист П. Мартин-Лёф.

Некоторые из учеников Андрея Николаевича были впоследствии избраны в Академию наук СССР и Академии наук союзных республик. Это действительные члены АН СССР А. И. Мальцев (алгебра, математическая логика), М. Д. Миллионщиков (механика, прикладная физика), С. М. Никольский (теория функций), А. М. Обухов (физика атмосферы), Ю. В. Прохоров (теория вероятностей); члены-корреспонденты АН СССР Л. Н. Большев (математическая статистика), А. А. Боровков (теория вероятностей, математическая статистика), И. М. Гельфанд (функциональный анализ), А. С. Монин (океанология), действительные члены АН УССР Б. В. Гнеденко (теория вероятностей, история математики), В. С. Михалевич (кибернетика), действительный член АН УзССР С. Х. Сираждинов (теория вероятностей) [3]. Какое разнообразие специальностей!

Последние приблизительно двадцать лет Андрей Николаевич в значительной мере посвятил проблемам средней школы. И, как во всем, он отдает этим исключительно важным для общества проблемам всего себя.

Начнем с организации в Москве школы-интерната № 18, или, как принято говорить, колмогоровской школы. В первые же годы ее существования он не только читал лекции и проводил упражнения для школьников, но и писал для учащихся конспекты своих лекций. Помимо обязательных занятий, он выступал перед учащимися с лекциями о музыке, искусстве, литературе.

То, что идея школы-интерната верна, подтверждает вся ее деятельность. Достаточно сказать, что среди учащихся школы 97—98% успешно поступили в лучшие вузы страны; около 400 ее выпускников завершили или завершают обучение в аспирантуре Московского университета; свыше 250 ее питомцев защитили кандидатские диссертации. А сейчас ряд ее воспитанников защитили и докторские диссертации. А ведь школа невелика — в ней всего 360 учащихся. Следует добавить, что учащиеся этой школы систематически занимают призовые места на всесоюзных и международных математических олимпиадах. Несомненно, что своими успехами школа в значительной мере обязана Андрею Николаевичу, его беззаветной преданности делу воспитания молодежи и развития ее способностей.

В 1964 г. А. Н. Колмогоров возглавил математическую секцию Комиссии АН СССР и АПН СССР по определению содержания

---

[3] В 1984 г. действительными членами АН СССР избраны И. М. Гельфанд и В. С. Михалевич, членами-корреспондентами АН СССР — В. И. Арнольд и Б. А. Севастьянов.— *Примеч. ред.*

среднего образования. Разработанная этой Комиссией в 1965 — 1968 гг. новая программа для 6—8 и 9—10 классов была одобрена Отделением математики АН СССР и до сих пор является, несмотря на многие изменения, основой для дальнейшего совершенствования математического образования и базой для написания учебников. Сам Андрей Николаевич принял непосредственное участие в написании новых учебников для массовой средней школы. По написанному под его руководством учебному пособию «Алгебра и начала анализа, 9—10» занимаются во всех общеобразовательных школах Советского Союза. Пособие «Геометрия, 6—8» встретило достаточно резкую критику. В настоящее время оно заменяется другими учебниками. Но при создании этих новых учебников их идейное содержание в значительной мере формируется под воздействием концепций, заложенных в учебном пособии, написанном под руководством А. Н. Колмогорова.

Вскоре после избрания в 1939 г. действительным членом АН СССР Андрей Николаевич был избран академиком-секретарем физико-математического Отделения. Делам Отделения он отдался с присущей ему страстью. Огромная работа была осуществлена Андреем Николаевичем на посту заведующего редакцией «Издательства иностранной литературы» и редактора отдела математики в «Большой Советской Энциклопедии». На написанных им и вышедших под его редакцией энциклопедических статьях по математике еще многие годы будут учиться энциклопедическому делу последующие авторы.

С 1 декабря 1964 г. по 13 декабря 1966 г. и с 13 ноября 1973 г. А. Н. Колмогоров — президент Московского математического общества; с 1946 по 1954 г. и с 1983 г.— главный редактор «Успехов математических наук».

Но центром деятельности А. Н. Колмогорова всегда был Московский университет, в котором он работает непрерывно с 1925 г. после окончания физико-математического факультета. В университете он заведовал: с 1938 по 1966 г. основанной им кафедрой теории вероятностей, с 1966 по 1976 г.— основанной им межфакультетской лабораторией статистических методов, с 1976 по 1980 г.— основанной им кафедрой математической статистики, с 1980 г. по настоящее время — кафедрой математической логики. С 1951 по 1953 г. А. Н. Колмогоров — директор Института математики и механики МГУ, а с 1954 по 1956 г. и с 1978 г. по настоящее время А. Н. Колмогоров заведует отделением математики механико-математического факультета МГУ; на этом посту, как и ранее (возглавляя НИИ математики МГУ и Институт механики и математики МГУ), он сделал чрезвычайно много для улучшения подготовки аспирантов математических специальностей. С 1954 по 1958 г. А. Н. Колмогоров — декан механико-математического факультета МГУ.

Особого внимания заслуживает деятельность Андрея Николаевича по организации в Московском университете небольшого, но крайне необходимого подразделения — статистической лаборатории. Мысль о создании статистической лаборатории при кафедре теории вероятностей возникла у Андрея Николаевича в начале 1960 г. Он приступил к осуществлению этой идеи с того, что организовал для сотрудников кафедры рабочий семинар по решению прикладных статистических задач, на котором сам прочитал ряд вводных лекций с демонстрацией примеров. При лаборатории удалось создать сектор вычислительных работ, оснащенный ЭВМ. Без хорошо организованной библиотеки, которая была бы снабжена основными журналами и систематически пополнялась монографиями, невозможно плодотворно вести научные исследования. Для организации специализированной библиотеки статистической лаборатории Андрей Николаевич не только добился отпуска средств, но и передал значительную часть полученной им международной бальцановской премии на приобретение литературы.

Статистическая лаборатория — не единственное подразделение Московского университета, да и не только университета, созданное по инициативе Андрея Николаевича. Так по его предложению в 1946 г. в Институте теоретической геофизики АН СССР была создана маленькая лаборатория атмосферной турбулентности, из которой впоследствии выросла значительная часть Института физики атмосферы.

Забота о библиотеках, об издании математической литературы, о подготовке специалистов всех уровней, о создании новых кафедр и лабораторий и о многом другом — все это проявление присущего Андрею Николаевичу чувства личной ответственности за судьбы науки и просвещения. Это чувство заставляло его принимать активное участие в формировании учебных планов и программ на всех ступенях математического образования. То же чувство побудило Андрея Николаевича решительно поддержать в свое время молодую кибернетику. Поддержка эта была существенной для признания научного статуса кибернетики не только в нашей стране, но и за рубежом. В приветствии к 80-летию А. Н. Колмогорова от Института кибернетики им. В. М. Глушкова АН УССР говорится: «В нашем институте всегда помнят о Вашем выдающемся вкладе в становление советской кибернетики. Ваши теории и научные идеи воплощены во многих кибернетических разработках. Многие из советских кибернетиков — Ваши ученики либо ученики Ваших учеников». Как известно, кибернетика прокладывала себе путь в борьбе с непризнанием. Подобной же была судьба генетики — и тут А. Н. Колмогоров также смело выступил в поддержку научной истины. В статье, опубликованной в 1940 г. в разделе «Генетика» Докладов АН СССР, он показал, что материал, собранный последователями академика Т. Д. Лысенко, служит, вопреки мнению этих последователей, новым подтверждением законов Менделя.

Чувство личной ответственности — один из компонентов высокой нравственной позиции А. Н. Колмогорова в науке. Другими компонентами этой его позиции являются объективность, отсутствие завышенной оценки своих учеников, готовность поддержать любые существенные достижения, скромность в вопросах приоритета и авторства (фактически Андрей Николаевич соавтор многих работ своих учеников), отзывчивость и готовность помочь.

Научные заслуги Андрея Николаевича высоко оценены в нашей стране и за ее пределами. Он награжден семью орденами Ленина, ему присвоено звание Героя Социалистического Труда, он — лауреат Ленинской и Государственной премий.

Свыше двадцати научных организаций избрали Андрея Николаевича своим членом. Он является, в частности, членом Королевской Нидерландской академии наук (1963), Лондонского королевского общества (1964), Национальной академии наук США (1967), Парижской академии наук (1968), Румынской академии наук (1965), Академии естествоиспытателей «Леопольдина» (1959), Американской академии наук и искусств в Бостоне (1959). Он почетный доктор Парижского, Стокгольмского и Варшавского университетов; избран почетным членом Московского, Индийского, Калькуттского математических обществ, Лондонского королевского статистического общества, Международного статистического института, Американского общества метеорологов. В 1963 г. ему присуждена Международная бальцановская премия. (Одновременно с А. Н. Колмогоровым, но по другим разделам премию фонда Бальцана получили папа Иоанн XXIII, историк С. Морисон, биолог К. Фриш, композитор П. Хиндемит.)

Андрей Николаевич Колмогоров занимает уникальное место в современной математике, да и в мировой науке в целом. По широте и разнообразию своих научных занятий он напоминает классиков естествознания прошлых веков.

В каждой из областей, которыми занимался Андрей Николаевич, он достигал ее вершин. Его работы по праву можно назвать классическими; их влияние с ходом времени не только не убывает, но и постоянно усиливается.

Для научного облика А. Н. Колмогорова характерно редкое соединение в одном лице черт абстрактного математика и естествоиспытателя, теоретика и практика; от аксиоматики теории вероятностей он непосредственно переходит к статистическому контролю массовой продукции, от теоретической гидромеханики к личному участию в океанографических экспедициях.

Вся жизнь Андрея Николаевича Колмогорова — это беспримерный подвиг во имя науки. А. Н. Колмогоров навсегда вошел в число великих русских ученых.

1

О ПОНЯТИИ АЛГОРИТМА *

Мы отправляемся от следующих наглядных представлений об алгоритмах:

1) Алгоритм Г, примененный ко всякому «условию» («начальному состоянию») A из некоторого множества $\mathfrak{G}$ (Г) («области применимости» алгоритма Г), дает «решение» («заключительное состояние») B.

2) Алгоритмический процесс расчленяется на отдельные шаги заранее ограниченной сложности; каждый шаг состоит в «непосредственной переработке» возникшего к этому шагу состояния S в состояние $S^* = \Omega_\Gamma(S)$.

3) Процесс переработки $A^0 = A$ в $A^1 = \Omega_\Gamma(A^0)$, A^1 в $A^2 = \Omega_\Gamma(A^1)$, A^2 в $A^3 = \Omega_\Gamma(A^2)$ и т. д. продолжается до тех пор, пока либо не произойдет безрезультатная остановка (если оператор Ω_Γ не определен для получившегося состояния), либо не появится сигнал о получении «решения». При этом не исключается возможность неограниченного продолжения процесса (если никогда не появится сигнал о решении).

4) Непосредственная переработка S в $S^* = \Omega_\Gamma(S)$ производится лишь на основании информации о виде заранее ограниченной «активной части» состояния S и затрагивает лишь эту активную часть.

Предполагается строгое определение, отображающее эти наглядные представления в точных математических терминах.

Состояния изображаются топологическими комплексами специального вида. Активная часть состояния образуется из всех элементов комплекса, удаленных на расстояние не более заданного от некоторой начальной вершины. Оператор Ω_Γ задается конечным набором правил. Каждое из этих правил имеет вид $U_i \to W_i$; его применение состоит в том, что если активная часть комплекса S есть U_i, то, чтобы получить $S^* = \Omega_\Gamma(S)$, следует U_i заменить на W_i, оставив неизменным $S \setminus U_i$.

Устанавливается эквивалентность между понятием алгоритма (в смысле предложенного определения) и понятием частично рекурсивной функции.

---

* УМН, 1953, т. 8, вып. 4, с. 175—176.

2
К ОБЩЕМУ ОПРЕДЕЛЕНИЮ КОЛИЧЕСТВА ИНФОРМАЦИИ *

Совместно с И. М. Гельфандом и А. М. Ягломом

Предлагаемые далее определения (2), (4) и свойства I—IV величины $I(\xi, \eta)$ были даны в докладе А. Н. Колмогорова на совещании по теории вероятностей и математической статистике (Ленинград, июнь 1954 г.). Теорема 4 и теорема 5 о полунепрерывности $I(\xi, \eta)$ при слабой сходимости распределений были найдены И. М. Гельфандом и А. М. Ягломом. После этого А. Н. Колмогоровым была предложена окончательная редакция заметки, в которой подчеркивается, что по существу переход от конечного случая к общему и вычисление и оценка количества информации при помощи предельных переходов совершенно тривиальны, если вести изложение в терминах нормированных булевых алгебр. Не представляло бы, конечно, большого труда сформулировать и более общие принципы предельного перехода не по $n \to \infty$, а по той или иной частичной упорядоченности.

§ 1. В этом параграфе система $\mathfrak{S}$ всех «случайных событий» $A, B, C, \ldots$ предполагается булевой алгеброй с операциями образования дополнительного элемента A', соединения $A \cup B$ и умножения AB, единичным элементом ε и нулевым элементом N. Рассматриваемые распределения вероятностей $P(A)$ являются определенными на $\mathfrak{S}$ неотрицательными функциями, удовлетворяющими условию аддитивности $P(A \cup B) = P(A) + P(B)$, если $AB = N$, и нормировки $P(\varepsilon) = 1$.

Классическое понятие «испытания» естественно идентифицировать с понятием подалгебры алгебры $\mathfrak{S}$ (наглядно: подалгебра $\mathfrak{A}$ состоит из всех событий, исход которых становится известным после совершения данного испытания). Если подалгебры $\mathfrak{A}$ и $\mathfrak{B}$ конечны, то «количество информации, содержащейся в результатах испытания $\mathfrak{A}$ относительно результатов испытания $\mathfrak{B}$», мы определяем формулой Шеннона

$$I(\mathfrak{A}, \mathfrak{B}) = \sum_{i, j} P(A_i B_j) \log \frac{P(A_i B_j)}{P(A_i) P(B_j)}. \tag{1}$$

В общем случае естественно положить

$$I(\mathfrak{A}, \mathfrak{B}) = \sup_{\mathfrak{A}_1 \subseteq \mathfrak{A},\ \mathfrak{B}_1 \subseteq \mathfrak{B}} I(\mathfrak{A}_1, \mathfrak{B}_1), \tag{2}$$

где верхняя грань берется по всем конечным подалгебрам $\mathfrak{A}_1 \subseteq \mathfrak{A}$ и $\mathfrak{B}_1 \subseteq \mathfrak{B}$. Символически можно записать определение (2)

* ДАН СССР, 1956, т. 111, № 4, с. 745—748.

в виде

$$I(\mathfrak{A}, \mathfrak{B}) = \int\limits_{\mathfrak{A}} \int\limits_{\mathfrak{B}} P(d\mathfrak{A}\, d\mathfrak{B}) \log \frac{P(d\mathfrak{A}\, d\mathfrak{B})}{P(d\mathfrak{A})\, P(d\mathfrak{B})} \,. \tag{3}$$

В случае конечных $\mathfrak{A}$ и $\mathfrak{B}$ количество информации $I(\mathfrak{A}, \mathfrak{B})$ действительно и неотрицательно. В общем случае, кроме действительных неотрицательных значений I, может появиться еще только значение $I = +\infty$. Сохраняются следующие хорошо известные для конечного случая свойства величины I ($[\mathfrak{M}]$ обозначает наименьшую подалгебру алгебры $\mathfrak{S}$, содержащую множество $\mathfrak{M}$):

1. $I(\mathfrak{A}, \mathfrak{B}) = I(\mathfrak{B}, \mathfrak{A})$.

2. $I(\mathfrak{A}, \mathfrak{B}) = 0$ в том и только в том случае, если системы событий $\mathfrak{A}$ и $\mathfrak{B}$ независимы.

3. Если $[\mathfrak{A}_1 \cup \mathfrak{B}_1]$ и $[\mathfrak{A}_2 \cup \mathfrak{B}_2]$ независимы, то

$$I([\mathfrak{A}_1 \cup \mathfrak{A}_2],\ [\mathfrak{B}_1 \cup \mathfrak{B}_2]) = I(\mathfrak{A}_1, \mathfrak{B}_1) + I(\mathfrak{A}_2, \mathfrak{B}_2).$$

4. Если $\mathfrak{A}_1 \subseteq \mathfrak{A}$, то $I(\mathfrak{A}_1, \mathfrak{B}) \leqslant I(\mathfrak{A}, \mathfrak{B})$.

Следующие теоремы почти очевидны, но в применениях к конкретным случаям доставляют полезные методы вычисления и оценки количества информации при помощи предельных переходов.

Теорема 1. *Если алгебра $\mathfrak{A}_1 \subseteq \mathfrak{A}$ всюду плотна на алгебре $\mathfrak{A}$ в смысле метрики*

$$\rho(A, B) = P(AB' \cup A'B),$$

то

$$I(\mathfrak{A}_1, \mathfrak{A}) = I(\mathfrak{A}, \mathfrak{B}).$$

Теорема 2. *Если*

$$\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \ldots \subseteq \mathfrak{A}_n \subseteq \ldots,$$

то для алгебры $\mathfrak{A} = \bigcup\limits_n \mathfrak{A}_n$ имеет место равенство

$$I(\mathfrak{A}, \mathfrak{B}) = \lim_{n\to\infty} I(\mathfrak{A}_n, \mathfrak{B}).$$

Теорема 3. *Если последовательность распределений $P^{(n)}$ сходится на $[\mathfrak{A} \cup \mathfrak{B}]$ к распределению P, т. е. если $\lim\limits_{n\to\infty} P^{(n)}(C) = P(C)$ для $C \in [\mathfrak{A} \cup \mathfrak{B}]$, то для определенных в соответствии с (1) и (2) через распределения $P^{(n)}$ и P количеств информации $I^{(n)}(\mathfrak{A}, \mathfrak{B})$ и $I(\mathfrak{A}, \mathfrak{B})$ имеет место неравенство*

$$\liminf_{n\to\infty} I^{(n)}(\mathfrak{A}, \mathfrak{B}) \geqslant I(\mathfrak{A}, \mathfrak{B}).$$

§ 2. Будем теперь считать, что основная булева алгебра $\mathfrak{S}$ является σ-алгеброй, а все рассматриваемые распределения $P(A)$ будем предполагать σ-аддитивными. Будем обозначать одной буквой X «измеримое пространство», т. е. совокупность множества X и σ-алгебры S_X некоторых его подмножеств с единицей X (в некоторое отклонение от терминологии [1], где S_X может быть σ-кольцом). «Случайным элементом» ξ пространства X называется гомоморфное отображение

$$\xi^*(A) = B$$

булевой алгебры S_X в булеву алгебру $\mathfrak{S}$. Наглядно $\xi^*(A)$ обозначает «событие, заключающееся в том, что $\xi \in A$».

При отображении ξ^* алгебра S_X переходит в подалгебру алгебры $\mathfrak{S}$, которую мы обозначим

$$\mathfrak{S}_\xi = \xi^*(S_X).$$

Естественно положить по определению

$$I(\xi, \eta) = I(\mathfrak{S}_\xi, \mathfrak{S}_\eta). \tag{4}$$

Обычным образом в соответствии с изложением в [1] определим измеримое пространство $X \times Y$. Условие

$$(\xi, \eta)^*(A \times B) = \xi^*(A)\,\eta^*(B)$$

однозначно определяет гомоморфизм $(\xi, \eta)^*$ булевой σ-алгебры $S_{X\times Y}$ в $\mathfrak{S}$. Этот гомоморфизм служит определением пары (ξ, η) случайных объектов ξ и η в качестве нового случайного объекта — случайного элемента пространства $X \times Y$. Формула

$$\mathcal{P}_\xi(A) = P(\xi^*(A))$$

определяет меру в пространстве X, которая называется распределением случайного объекта ξ. В соответствии с [1] определим, наконец, в пространстве $X \times Y$ меру

$$\Pi = \mathcal{P}_\xi \times \mathcal{P}_\eta.$$

Применяя к заданным на $X \times Y$ двум мерам Π и $\mathcal{P}(\xi, \eta)$ теорему Радона—Никодима, напишем равенство

$$\mathcal{P}_{\xi,\eta}(C) = \iint\limits_C a(x, y)\, d\mathcal{P}_\xi\, d\mathcal{P}_\eta + S(C),$$

где мера S сингулярна относительно меры Π.

Т е о р е м а 4. *Количество информации* $I(\xi, \eta)$ *может быть конечно лишь в случае* $S(C) = 0$.

В этом случае

$$I(\xi, \eta) = \int\limits_X \int\limits_Y a(x, y) \log a(x, y)\, d\mathcal{P}_\xi\, d\mathcal{P}_\eta =$$

$$= \int\limits_{X\times Y} \log a(x, y)\, d\mathcal{P}(\xi, \eta). \tag{5}$$

Заметим еще, что для любого случайного элемента ξ пространства X и измеримого по Борелю отображения $y = f(x)$ пространства X в пространство Y случайный элемент $\eta = f(\xi)$ пространства Y определяется отображением

$$\eta^*(A) = \xi^* f^{-1}(A)$$

алгебры S_Y в алгебру $\mathfrak{S}$. Определяя обычным образом понятие математического ожидания M, можно формулу (5) записать в виде

$$I(\xi, \eta) = \mathsf{M} \log a(\xi, \eta). \tag{6}$$

Из свойств 1—4 величины $I(\mathfrak{A}, \mathfrak{B})$ вытекают следующие свойства величины $I(\xi, \eta)$:

I. $I(\xi, \eta) = I(\eta, \xi)$.

II. $I(\xi, \eta) = 0$ в том и только в том случае, если ξ и η независимы.

III. Если пары (ξ_1, η_1) и (ξ_2, η_2) независимы, то

$$I((\xi_1, \xi_2), (\eta_1, \eta_2)) = I(\xi_1, \eta_1) + I(\xi_2, \eta_2).$$

IV. Если ξ_1 есть функция от ξ (в указанном выше смысле), то

$$I(\xi_1, \eta) \leqslant I(\xi, \eta).$$

Легко подобно паре (ξ, η) определить последовательность $\xi = (\xi_1, \xi_2, \ldots, \xi_n, \ldots)$ в качестве случайного элемента пространства $X = \mathop{\times}\limits_{n=1}^{\infty} X_n$. Из теорем 1 и 2 легко выводится, что при этом всегда

$$I(\xi, \eta) = \lim_{n\to\infty} I((\xi_1, \xi_2, \ldots, \xi_n), \eta).$$

Укажем в заключение еще на одно применение теоремы 3. Пусть X и Y — полные метрические пространства. Если за S_X и S_Y принять алгебры их борелевских множеств, то пространства X и Y становятся в то же время измеримыми пространствами. В этой обстановке имеет место следующая

Теорема 5. *Если для случайных элементов $\xi \in X$, $\eta \in Y$ распределения $\mathcal{P}^{(n)}(\xi, \eta)$ слабо сходятся к распределению $\mathcal{P}(\xi, \eta)$, то для соответствующих количеств информации имеет место неравенство*

$$\liminf_{n\to\infty} I^{(n)}(\xi, \eta) \geqslant I(\xi, \eta).$$

Для доказательства достаточно применить теорему 3 к алгебрам $\mathfrak{A}$ и $\mathfrak{B}$, состоящим соответственно из событий $\xi \in A$, $\eta \in B$, где A и B суть множества непрерывности распределений $\mathcal{P}_\xi$ и соответственно $\mathcal{P}_\eta$ (т. е. множества, для границ которых соответственно $\mathscr{P}_\xi = 0$ или $\mathscr{P}_\eta = 0$), и заметить, что алгебра $[\mathfrak{A} \bigcup \mathfrak{B}]$ всюду плотна на $\mathfrak{S}_{X \times Y}$ в метрике, порождаемой предельным распределением P.

8 октября 1956 г.

ЛИТЕРАТУРА

1. *Халмош П. Р.* Теория меры. М.: Изд-во иностр. лит., 1953.

3

ТЕОРИЯ ПЕРЕДАЧИ ИНФОРМАЦИИ *

I. ВОЗНИКНОВЕНИЕ ТЕОРИИ И ЕЕ СОДЕРЖАНИЕ

Выступая второй раз перед Общим собранием нашей Академии, мне хочется начать с замечания, что тема моего сегодняшнего сообщения тесно связана с темой доклада «Статистическая теория колебаний с непрерывным спектром», который я делал на Общем собрании Академии в 1947 г. [1] В самом деле, наиболее богатая содержанием и наиболее увлекательная с чисто математической стороны часть теории информации — теория стационарно работающих каналов связи, передающих непрерывные сообщения,— не могла бы быть создана без заранее разработанной теории стационарных случайных процессов и, в частности, без спектральной теории таких процессов.

Интерес к разнообразным задачам передачи и хранения информации имеет большую давность. Давно, по существу, возникали вопросы об оценке «количества» информации. Вопрос о возможности введения универсальной числовой меры для количества информации особенно важен в случаях необходимости преобразования информации одного рода в информацию качественно другого рода. Типичной задачей на преобразование информации

---

* В кн.: Сессия Академии наук СССР по научным проблемам автоматизации производства, 15—20 окт. 1956 г.: Пленар. заседания. М.: Изд-во АН СССР, 1957, с. 66—99. В конце статьи приводится выступление А. Н. Колмогорова при обсуждении доклада.

[1] См. [1]. Более подробное изложение математической стороны вопросов, которым был посвящен этот мой доклад в 1947 г., выделено в статью [2]. Сейчас аналогичную роль играет глава II настоящей работы, воспроизводящая с небольшими изменениями доклад, прочитанный Б. В. Гнеденко от моего имени на симпозиуме по теории информации Американского института радиоинженеров (Массачусетсский технологический институт, сентябрь 1956 г.) и основанный во многих частях на докладе, прочитанном мною совместно с И. М. Гельфандом и А. М. Ягломом на Всесоюзном математическом съезде в июне 1956 г.

является задача табулирования непрерывных функций от непрерывно меняющихся аргументов. Если, например, функцию $f(x, y)$ от двух аргументов

$$0 \leqslant x \leqslant 1, \qquad 0 \leqslant y \leqslant 1$$

необходимо задать с точностью до ε и известно, что ее приращение Δf не превосходит по абсолютной величине ε, когда

$$|\Delta x| \leqslant \varepsilon_x, \qquad |\Delta y| \leqslant \varepsilon_y,$$

то заведомо достаточно табулировать функцию f с шагом ε_x по x и ε_y по y, т. е. задать приблизительно

$$N \sim 1/\varepsilon_x\varepsilon_y$$

ее значений. Если допустить, что $|f| \leqslant 1$, то для фиксации одного значения функции достаточно приблизительно

$$K \sim \log_{10}(1/\varepsilon)$$

десятичных знаков. Таким образом, если следовать намеченному сейчас плану, всего в таблицу придется ввести приблизительно

$$NK \sim (1/\varepsilon_x\varepsilon_y)\log_{10}(1/\varepsilon)$$

десятичных знаков (не считая записи значений аргументов x и y, которая стандартна для всех функций f, подчиненных сделанным ограничениям). Общеизвестно, что такой примитивный подход к делу, особенно в случае функций нескольких переменных, привел бы к таблицам огромного объема и практически неосуществимым. На самом деле при табулировании достаточно «гладких» функций выбирают шаг таблицы по независимым переменным значительно большим, а для нахождения значений функции при промежуточных значениях аргумента прибегают к интерполированию с помощью разностей до какого-либо порядка (в случае таблиц функций нескольких переменных часто до четвертого или шестого порядка). Кроме того, часто первые знаки $f(x, y)$, повторяющиеся для близких значений аргументов, не выписывают в каждой клеточке таблицы, соответствующей определенной паре значений (x, y), а выписывают, например, лишь в начале строки, состоящей из многих таких клеточек.

При всей общеизвестности и давнем происхождении описанных приемов построения таблиц соответствующие общие теоретические исследования, которые должны выяснить минимальное количество информации, необходимое для фиксации с заданной точностью ε произвольной функции f, подчиненной лишь тем или иным общим условиям, находятся еще в начальной стадии. Например, только недавно в моей заметке [3] была явно дана формула [2]

$$H_\varepsilon (F^n_{p,\alpha}) \asymp (1/\varepsilon)^{n/(p+\alpha)}, \tag{1}$$

указывающая порядок роста при $\varepsilon \to 0$ количества информации, необходимого для фиксации заданной в ограниченной области функции n переменных с ограниченными производными первых p порядков и производными порядка p, удовлетворяющими условию Липшица степени α.

Вполне понятно, что вопросы преобразования и хранения информации имеют решающее значение при разработке конструкций и способов употребления современных вычислительных машин. Здесь «объем памяти» запоминающего устройства характеризуется числом удерживаемых этим устройством двоичных знаков (0 или 1). Способ измерения количества информации при помощи сравнения любой информации с информацией в виде некоторого числа двоичных знаков стал сейчас стандартным в теории информации.

Переходя к задачам из теории «каналов связи» в той ее форме, которая связана с работой телеграфных или телефонных линий и подобных им устройств, начнем с очень простой задачи о передаче последовательности десятичных знаков

$$a_1, a_2, \ldots, a_n, \ldots \tag{2}$$

при помощи последовательности двоичных знаков

$$b_1, b_2, \ldots, b_n, \ldots \tag{3}$$

Допустим, что последовательность (2) возникает во времени по одному знаку в единицу времени. Спрашивается, с какой скоростью надо передавать двоичные знаки последовательности (3), если желать, чтобы восстановление по последовательности (3) исходной последовательности (2) осуществлялось без неограниченно возрастающего со временем запоздания. Если изображать каждый знак a_n отдельно двоичными знаками, то на один знак a_n приходится тратить четыре двоичных знака (так как $2^3 = 8 < 10$ и только $2^4 = 16 \geqslant 10$). Если передавать каждую пару (a_{2n-1}, a_{2n}) знаков последовательности (2) группой двоичных знаков, то на нее придется тратить семь двоичных знаков (так как $2^6 = 64 < 100$ и только $2^7 = 128 \geqslant 100$), для передачи трех десятичных знаков достаточно тратить десять двоичных знаков (так как $2^9 < 1000$, но $2^{10} = 1024 \geqslant 1000$) и т. д. Соответствующие этим системам передачи скорости образования знаков последовательности (3), т. е. соответствующее число двоичных знаков, которые необходимо в

[2] Формулу эту можно назвать формулой А. Г. Витушкина, так как в его работе [4] по другому поводу выяснена роль показателя $n/(p+\alpha)$ и по существу содержится половина доказательства формулы (1), а именно доказательство того факта, что порядок $H_\varepsilon (F^n_{p,\alpha})$ не может быть меньше указываемого формулой (1).

среднем передавать за единицу времени, равны

$$\nu_1 = 4; \qquad \nu_2 = {}^7/_2 = 3{,}5; \qquad \nu_3 = {}^{10}/_3 = 3{,}33\ldots$$

Легко доказать, что возникающий таким образом ряд чисел сходится к пределу

$$\nu = \log_2 10 = 3{,}32\ldots$$

Именно это ν и является нижней гранью тех скоростей создания знаков последовательности (3), при которых можно без систематически возрастающего запоздания передавать последовательность (2).

Уже на этом нарочито наивном примере можно наблюдать многие типичные явления, открытые современной теорией передачи информации в применении к значительно более общим видам каналов связи.

1. Скорость создания информации, несмотря на разнородность ее качества, допускает разумное количественное измерение. В частности, мерой скорости создания информации при фиксировании в единицу времени в среднем ν знаков, каждый из которых может принимать k значений, разумно считать число

$$\overline{H} = \nu \log_2 k.$$

(В нашем случае

$$\overline{H} = \log_2 10 = 3{,}32\ldots$$

для последовательности (2) и

$$\overline{H}' = \nu \log_2 2 = \nu$$

для последовательности (3) при создании ν знаков в единицу времени.)

Для дальнейшего заметим, что в случае точно (без ошибок) работающих каналов связи скорость выдаваемой ими информации разумно называть их «пропускной способностью». В нашем случае пропускная способность последовательности (3), рассматриваемой как канал связи, равна $\overline{C} = \overline{H}' = \nu$. В общем случае пропускная способность определяется несколько иначе, что будет позднее объяснено. После этого замечания сформулируем второй общий принцип.

2. Для возможности передачи возникающей со скоростью $\overline{H}$ информации через канал связи с пропускной способностью $\overline{C}$ без систематически возрастающего запоздания необходимо условие

$$\overline{H} \leqslant \overline{C}.$$

Наиболее существен, однако, третий принцип, содержание которого в приводимой сейчас формулировке еще очень расплывчато, но в котором и заключается, пожалуй, наиболее существенное новое, внесенное в теорию передачи информации по каналам связи работами Шеннона.

3. Даже при значительной качественной разнородности информации, подлежащей передаче, и той информации, которую способен непосредственно воспринимать и выдавать канал связи, в случае

$$\overline{H} < \overline{C}$$

в принципе возможно производить передачу без систематически возрастающего запаздывания, если подлежащая передаче информация перед поступлением в канал связи надлежащим образом «кодируется», а после выхода из канала «декодируется». Однако, вообще говоря, при приближении $\overline{H}$ к $\overline{C}$ способы кодирования и декодирования неизбежно усложняются и возникает все большее и большее (при $\overline{H} \to \overline{C}$) запаздывание передачи.

Заметим, что запаздывание, о котором идет речь в принципе 3, возрастает при $\overline{H} \to \overline{C}$, но при постоянном источнике информации с $\overline{H} < \overline{C}$ может быть оценено некоторой величиной τ, которая сохраняется при неограниченном продолжении работы канала.

В виде следующего примера рассмотрим передачу текста, состоящего из букв русского алфавита. Так как различных букв в нашем алфавите 33, то можно формально образовать

$$N = 33^n$$

различных «текстов» длины n (т. е. последовательностей из n букв). Количество информации, содержащейся в указании одного определенного такого текста, равно по предыдущему

$$I = n \log 33$$

(здесь и всюду далее мы употребляем логарифмы при основании 2, не оговаривая этого специально). Существование стенографии показывает, однако, что реальные языковые тексты можно передавать значительно более кратким образом. Это вполне понятно: число N^* «осмысленных» текстов из n букв, конечно, несравненно меньше N. Поэтому в принципе возможна система записи («идеальная» стенография), которая требовала бы для фиксации любого из этих N^* осмысленных текстов лишь

$$I^* \sim \log N^*$$

двоичных знаков. Полное использование этой принципиальной возможности «сжатия текста», конечно, лежит за пределами осуществимого при помощи каких-либо систем стенографирования или кодирования текста, следующих достаточно простым формальным правилам: закономерности построения реального языкового текста вряд ли когда-либо будут полностью формализованы.

Мы обратим сейчас внимание лишь на одну особенность языковых текстов: различные буквы встречаются в них с различной частотой. Если фиксировать числа

$$n_1, n_2, \ldots, n_k$$

вхождений в текст длины n каждой из букв

$$a_1, a_2, \ldots, a_k$$

(естественно, что при этом

$$n_1 + n_2 + \ldots + n_k = n),$$

то число текстов длины n сократится до

$$N' = \frac{n!}{n_1!\, n_2! \ldots n_k!}.$$

Пользуясь так называемой формулой Стирлинга, которая нужна здесь даже лишь в ослабленной форме:

$$\log(n!) \sim n \log n,$$

легко подсчитать, что при больших n_i

$$I' = \log N' \sim -n \sum_i p_i \log p_i,$$

где

$$p_i = n_i/n$$

— частоты появления отдельных букв. Полученный результат можно выразить так: при употреблении отдельных букв с частотами p_i количество информации, передаваемой «на одну букву текста», равно

$$H = -\sum_i p_i \log p_i. \tag{4}$$

В случае равных частот

$$p_1 = p_2 = \ldots = p_k = 1/k$$

мы вновь получаем

$$H = \log k;$$

при любых других частотах p_i

$$H < \log k.$$

Легко указать приемы кодирования, позволяющие передавать произвольный текст, в котором буквы $a_1, a_2, \ldots, a_k$ встречаются с частотами $p_1, p_2, \ldots, p_k$, затрачивая в среднем (при передаче достаточно длинных текстов) на одну букву число двоичных знаков, сколь угодно мало отличающееся от H, определенного по формуле (4). Это еще одно проявление третьего из сформулированных выше принципов. Мы оставим, однако, теперь рассмотрение элементарных примеров и перейдем к элементам общей теории.

Формула (4) допускает, кроме чисто статистической (в которой p_i являются частотами), вероятностную интерпретацию. Пусть совсем общим образом дано распределение вероятностей

$$\mathsf{P}\{\xi = x_i\} = p_i \tag{5}$$

случайного объекта ξ. Если дано только это распределение, то в ответе на вопрос о том, какому из возможных значений

$$x_1, x_2, \ldots, x_k$$

равно в действительности ξ, остается некоторая неопределенность. Сообщая, что ξ равно такому-то определенному x_i, мы устраняем эту неопределенность, т. е. сообщаем некоторую дополнительную информацию об объекте ξ. Мерой неопределенности распределения (5) является его *энтропия*, выражающаяся уже знакомой нам формулой (4). Эта же формула выражает и *количество информации*, необходимое для устранения неопределенности в задании ξ лишь распределением (5), т. е. информации, содержащейся в указании точного значения ξ.

Пусть теперь, далее, дано совместное распределение вероятностей

$$\mathsf{P}\{\xi = x_i, \eta = y_j\} = p_{ij}$$

двух случайных объектов ξ и η. Если $\eta = y_j$ задано, то ξ получает условное распределение

$$\mathsf{P}\{\xi = x_i \mid \eta = y_j\} = p_{i\mid j}.$$

Количество информации, содержащееся в указании точного значения ξ, если значение $\eta = y_j$ уже известно, равно

$$H(\xi \mid \eta = y_j) = -\sum_i p_{i\mid j} \log p_{i\mid j},$$

в среднем же оно равно

$$\mathsf{M}H(\xi \mid \eta) = -\sum_j \mathsf{P}\{\eta = y_j\} \sum_i p_{i\mid j} \log p_{i\mid j}.$$

Естественно считать, что разность

$$I(\eta, \xi) = H(\xi) - \mathsf{M}H(\xi \mid \eta) \tag{6}$$

есть то количество информации относительно ξ, которое уже содержится в задании η. Легко подсчитать, что $I(\eta, \xi)$ можно записать в виде

$$I(\eta, \xi) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{\mathsf{P}\{\xi = x_i\}\, \mathsf{P}\{\eta = y_j\}}, \tag{7}$$

где ξ и η играют совершенно одинаковую роль: количество информации в η относительно ξ и количество информации в ξ относительно η численно равны друг другу.

Легко проверить, что всегда

$$I(\xi, \xi) = H(\xi). \tag{8}$$

Теперь без труда определяются скорость передачи информации через канал связи, работающий с ошибками, и пропускная способность таких каналов (пропускная способность определяется по-прежнему как верхняя грань возможной скорости передачи информации). Приведу еще один элементарный пример. Пусть «на входе» в канал подаются знаки η, равные 0 или 1, а на выходе получаются соответствующие знаки η', причем

$$\mathsf{P}\{\eta' = 0\} = 1 - \Delta, \qquad \mathsf{P}\{\eta' = 1\} = \Delta \quad \text{в случае} \quad \eta = 0,$$

$$\mathsf{P}\{\eta' = 0\} = \Delta, \qquad \mathsf{P}\{\eta' = 1\} = 1 - \Delta \quad \text{в случае} \quad \eta = 1.$$

Легко подсчитать, что верхняя грань

$$C = \sup I(\eta, \eta') = 1 + [\Delta \log \Delta + (1 - \Delta) \log (1 - \Delta)] \tag{9}$$

достигается в случае

$$\mathsf{P}\{\eta = 0\} = \mathsf{P}\{\eta = 1\} = {}^1/_2.$$

По формуле (9) количество информации $I(\eta, \eta')$, передаваемое нашим каналом на одну букву, равно заведомо нулю, если $\Delta = {}^1/_2$. Это вполне понятно, так как с вероятностной точки зрения в этом случае знаки η и η' независимы. Максимальное значение пропускной способности $C = 1$ получается при $\Delta = 0$ и при $\Delta = 1$. В случае $\Delta = 0$ канал работает без ошибок: с вероятностью единица $\eta = \eta'$. Хотя во втором случае с вероятностью единица $\eta \neq \eta'$, η легко восстанавливается по η' и в этом случае: надо только вместо 0 читать 1 и наоборот. Во всех остальных случаях

$$0 < C < 1.$$

Когда Шеннон в конце 40-х годов предложил измерять количество информации формулами (4), (6), (7) и обнаружил, что и для каналов связи, работающих с ошибками, при довольно широких предположениях относительно источника информации и строения канала сохраняются сформулированные выше второй и третий принципы, то этим и были созданы основы современной теории информации в виде дисциплины, способной к систематическому развитию.

Во всяком крупном открытии имеются элементы неожиданности. Этим крупное открытие и отличается от постепенно накапливаемых результатов текущей научной работы. Подчеркну здесь, в чем я усматриваю то качественно новое и неожиданное, что содержится уже в описанной элементарной части теории информации.

1. По первоначальному замыслу «информация» не есть скалярная величина. Различные виды информации могут быть чрезвычайно разнообразны. Можно было заранее ожидать, что можно предложить те или иные способы измерять «количество» информации, но было неясно, существует ли среди этих способов какой-либо один, имеющий принципиальное преимущество перед другими, а главное было совершенно неясно, можно ли качественно различные информации, для которых та или иная условная количественная мера одинакова, действительно считать эквивалентными в смысле трудности их передачи через каналы связи или их хранения в запоминающих устройствах.

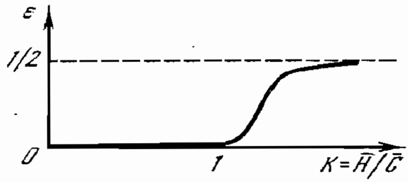


Рис. 1

Оказалось, что такая «правильная» по преимуществу мера количества информации существует и позволяет решать до конца широкий класс задач, в которых априори независимость решения от более тонких качественных особенностей информации была совершенно неясна.

2. В случае каналов связи (или запоминающих устройств), работающих с ошибками, можно было опасаться, что достижение передачи с достаточно малой вероятностью ошибки ε связано с очень большим уменьшением скорости передачи. Вопреки этим опасениям оказалось, что при условии $\overline{H} < \overline{C}$ сразу становится возможной передача информации со сколь угодно малой вероятностью ошибки. Для пояснения этого обстоятельства укажу на следующее. Пусть дан канал связи с пропускной способностью C. Рассмотрим задачу передачи по нему двоичных знаков, возникающих в числе H за единицу времени.

Тогда достижимый с любым приближением минимум вероятности ошибки в передаче отдельного знака является универсальной функцией ε (K) отношения $K = \overline{H}/\overline{C}$ (изображенной на рис. 1). При $K \leqslant 1$ она равна нулю, а при дальнейшем возрастании K начинает возрастать и при $K \to \infty$ стремится к половине (аналитически при $K > 1$ функция ε (K) определяется из уравнения

$$1/K = 1 + \varepsilon \log \varepsilon + (1 - \varepsilon) \log (1 - \varepsilon), \quad \varepsilon < {}^1/_2.$$

Таким образом, от значений, точно равных нулю при $K \leqslant 1$, наша функция при $K > 1$ сразу переходит к возрастанию.

Правда, как уже отмечалось ранее, при $\overline{H}$, близких к $\overline{C}$, достижение малой вероятности ошибки требует часто очень сложных правил кодирования информации и приводит к большому запаздыванию ее выдачи на выходе канала. Но именно эта большая сила сложных способов кодирования в смысле устранения ошибок и является принципиально важным выводом, к возможным применениям которого мы еще вернемся.

Обоснование предложенных Шенноном способов измерения количества информации, т. е. формул (4) и (7), можно мыслить двояко. Прежде всего можно выдвинуть те или иные естественные аксиоматические требования, которым мера количества информации должна удовлетворять. В применении к величине $H(\xi)$, т. е. в соответствии с формулой (8), к количеству информации, содержащейся в указании случайного объекта ξ относительно самого этого объекта ξ, образцово ясное изложение такого подхода к задаче обоснования теории информации дано в работе А. Я. Хинчина [5]. Представляло бы некоторый интерес дать столь же естественную аксиоматику непосредственно более общего понятия количества информации $I(\xi, \eta)$ в одном случайном объекте относительно другого случайного объекта. Дело в том (ср. главу II), что в непрерывном случае обычно $H(\xi)$ и $H(\eta)$, взятые отдельно, бесконечны, и $I(\xi, \eta)$ не может вычисляться по формуле (6).

В качестве итога исследований аксиоматического направления можно считать выясненным, что никакой другой столь же естественной скалярной сводной характеристики информации, содержащейся в одном случайном объекте ξ относительно другого случайного объекта η, кроме $I(\xi, \eta)$, существовать не может. Однако так как «информация» по своей природе не обязана быть (и в действительности не является!) скалярной величиной, то никакие аксиоматические исследования указанного направления не могут ответить на вопрос о том, сколь полно характеризует величина $I(\xi, \eta)$ интересующую нас «информацию». Некоторый ответ на этот вопрос, как уже можно было понять из предшествующего изложения, дают теоремы, обосновывающие второй и третий сформулированные выше общие принципы теории передачи информации. В силу этих теорем в весьма широком классе случаев приспособленность канала связи или запоминающего устройства к передаче или хранению информации с достаточной полнотой характеризуется одним-единственным числом C (при расчете на единицу времени $\overline{C}$), а приспособленность самой информации к передаче или хранению — одним-единственным числом H (или при расчете на единицу времени $\overline{H}$). Качественное своеобразие информации оказывается при этом несущественным. Трудно переоценить значение такого рода результатов, если учесть широту применения в современной технике таких преобразований, как преобразование изображений, передаваемых в телевидении, в электрические колебания и обратно (в терминологии теории передачи информации это частный случай «кодирования» и «декодирования») и т. п. Вместе с тем надо понимать, что при всей увлекательности идей теории информации подобное стирание качественных особенностей информации имеет место только с известным приближением и при определенных условиях. Эти условия в хорошо исследованных случаях, грубо говоря, состоят в накоплении большого количества однородной информации, в действительной осуществимости сложных методов кодирования и в безвредности возникающего при этом запаздывания выдачи информации. Кроме того, следует ясно представлять себе, что точное математическое обоснование реальности этих представлений пока выполнено лишь при довольно ограниченных предположениях. При необычайном богатстве идей, данных в работах самого Шеннона, изложение в них обычно крайне туманно. Лишь позднее в ряде работ чистых математиков для случая стационарно работающих каналов, передающих дискретные сигналы, «теоремы Шеннона» были доказаны безукоризненно и в достаточно общих предположениях. Наиболее законченный характер из работ этого направления имеет работа А. Я. Хинчина [6].

С значительно опережающим фундаментальные математические исследования развитием теории информации, осуществляемым исследователями более прикладного направления, советский читатель может с достаточной полнотой познакомиться по книгам [7—9] (в [7] включен перевод основной работы Шеннона [10]).

Существенные элементы теории информации для непрерывного случая, казалось бы более трудного, возникли до Шеннона. Логарифмическая мера количества информации, вполне аналогичная шенноновскому выражению $H(\xi)$, лежит в основе асимптотических методов статистики, разработанных Фишером еще в 1921—1925 гг. (см. [11, гл. 32—33])[3].

Еще ближе непосредственно к работам Шеннона стоят результаты В. А. Котельникова [12], полученные еще в 1933 г. Здесь была сформулирована фундаментальная идея спектральной теории передачи информации при помощи непрерывных сигналов, о которой я говорю подробнее в главе II[4].

II. ПРИНЦИПЫ ПОСТРОЕНИЯ ТЕОРИИ В СЛУЧАЕ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

Роль энтропии случайного объекта ξ, способного принимать значения $x_1, x_2, \ldots, x_n$ с вероятностями $p_1, p_2, \ldots, p_n$,

$$H(\xi) = -\sum_k p_k \log p_k$$

---

[3] Общеизвестно и то обстоятельство, что выражение $H(\xi)$ формально тождественно с выражением энтропии в физике. Это совпадение я считаю вполне достаточным для оправдания наименования величины $H(\xi)$ и в теории информации «энтропией»: такие математические аналогии следует всегда подчеркивать, так как сосредоточение на них внимания содействует прогрессу науки. Но было бы преувеличением считать, что физические теории, связанные с понятием энтропии, уже содержат в себе в готовом виде элементы теории информации: первое использование выражений типа энтропии в качестве меры количества информации, которое мне известно, находится в только что упомянутых работах Фишера.

[4] Глава II настоящей работы содержит больше математических подробностей, чем устный доклад.

в теории информации и теории передачи сообщений при помощи дискретных сигналов можно считать достаточно выясненной. Далее я настаиваю на той идее, что основным понятием, допускающим обобщение на совершенно произвольные непрерывные сообщения и сигналы, является не непосредственно понятие энтропии, а понятие количества информации $I(\xi, \eta)$ в случайном объекте ξ относительно объекта η. В дискретном случае эта величина корректно вычисляется по известной формуле Шеннона [5]

$$I(\xi, \eta) = H(\eta) - \mathsf{M}H(\eta \mid \xi).$$

Для конечномерных распределений, обладающих плотностью, величина $I(\xi, \eta)$ по Шеннону определяется аналогичной формулой

$$I(\xi, \eta) = h(\eta) - \mathsf{M}h(\eta \mid \xi),$$

где $h(\eta)$ есть «дифференциальная энтропия»

$$h(\eta) = -\int p_\eta(y) \log p_\eta(y)\, dy,$$

а $h(\eta \mid \xi)$ — аналогичным образом определяемая условная дифференциальная энтропия. Общеизвестно, что величина $h(\xi)$ не имеет непосредственного реального истолкования и даже неинвариантна по отношению к преобразованиям координат в пространстве $x_1, \ldots, x_n$. Для бесконечномерных распределений аналога выражения $h(\xi)$, вообще говоря, не существует.

В собственном смысле слова энтропия объекта ξ с непрерывным распределением всегда бесконечна. Если непрерывные сигналы не могут тем не менее служить для передачи неограниченно большой информации, то только потому, что они всегда наблюдаются с ограниченной точностью. Естественно поэтому, задав точность наблюдения ε, определить соответствующую ей «ε-энтропию» $H_\varepsilon(\xi)$ объекта ξ. Это и было сделано Шенноном под названием «скорости создания сообщений». Хотя выбор для этой величины нового названия и не меняет существа дела, я решаюсь предложить такое переименование, подчеркивающее более широкий интерес понятия и его глубокую аналогию с обыкновенной точной энтропией. Укажу заранее на то, что, как указывается в § 3, для ε-энтропии сохраняется теорема об экстремальной роли нормального распределения (как в конечномерном, так и в бесконечномерном случае). Далее, в § 1, 2 я даю, не претендуя на безусловную новизну, абстрактную формулировку определения и основных свойств величины $I(\xi, \eta)$ и обзор основной проблематики теории передачи сообщений Шеннона. В § 3—6 излагаются некоторые конкретные результаты, полученные советскими математиками в самое последнее время. Мне особенно хотелось бы подчеркнуть интерес исследо-

[5] В кажущихся мне целесообразными обозначениях $H(\eta \mid \xi)$ есть условная энтропия η при $\xi = x$, а $\mathsf{M}H(\eta \mid \xi)$ — математическое ожидание этой условной энтропии при переменном ξ.

ваний по асимптотическому поведению ε-энтропии при $\varepsilon \to 0$. Исследованные ранее случаи

$$H_\varepsilon(\xi) \sim n \log(1/\varepsilon), \qquad \overline{H}_\varepsilon(\xi) \sim 2W \log(1/\varepsilon),$$

где n — число измерений, а W — ширина полосы спектра, являются лишь очень частными случаями могущих здесь встретиться закономерностей. Для понимания открывающихся перспектив может представлять интерес, изложенная в других терминах моя заметка [3].

§ 1
КОЛИЧЕСТВО ИНФОРМАЦИИ В ОДНОМ СЛУЧАЙНОМ ОБЪЕКТЕ ОТНОСИТЕЛЬНО ДРУГОГО

Пусть ξ и η являются случайными объектами с областями возможных значений X и Y,

$$P_\xi(A) = \mathsf{P}\{\xi \in A\}, \qquad P_\eta(B) = \mathsf{P}\{\eta \in B\}$$

— соответствующие распределения вероятностей и

$$P_{\xi\eta}(C) = \mathsf{P}\{(\xi, \eta) \in C\}$$

— совместное распределение вероятностей объектов ξ и η. По определению количество информации в случайном объекте ξ относительно случайного объекта η дается формулой

$$I(\xi, \eta) = \int_X \int_Y P_{\xi\eta}(dx\,dy) \log \frac{P_{\xi\eta}(dx\,dy)}{P_\xi(dx)\,P_\eta(dy)}. \tag{1}$$

Точный смысл этой формулы требует некоторых пояснений, а сообщаемые далее общие свойства величины $I(\xi, \eta)$ справедливы лишь при некоторых теоретико-множественного характера ограничениях на распределения P_ξ, P_η и $P_{\xi\eta}$, но здесь я не буду на этом останавливаться. Во всяком случае общая теория без больших затруднений может быть изложена таким образом, что она будет применима к случайным объектам ξ и η весьма общей природы (векторам, функциям, обобщенным функциям и т. п.). Определение (1) можно считать принадлежащим Шеннону, хотя он и ограничился случаем

$$P_\xi(A) = \int_A p_\xi(x)\,dx, \qquad P_\eta(B) = \int_B p_\eta(y)\,dy,$$

$$P_{\xi\eta}(C) = \iint_C p_{\xi\eta}(x, y)\,dx\,dy,$$

когда (1) переходит в

$$I(\xi, \eta) = \int\limits_X \int\limits_Y p_{\xi\eta}(x, y) \log \frac{p_{\xi\eta}(x, y)}{p_\xi(x)\, p_\eta(y)}\, dx\, dy.$$

Иногда бывает полезно представить распределение $P_{\xi\eta}$ в виде

$$P_{\xi\eta}(C) = \iint\limits_C a(x, y)\, P_\xi(dx)\, P_\eta(dy) + S(C), \tag{2}$$

где функция $S(C)$ сингулярна относительно произведения

$$P_\xi \times P_\eta.$$

Если сингулярная компонента S отсутствует, то формула

$$\alpha_{\xi, \eta} = a(\xi, \eta) \tag{3}$$

определяет однозначно с точностью до вероятности нуль случайную величину $\alpha_{\xi, \eta}$. Иногда бывает полезна сформулированная И. М. Гельфандом и А. М. Ягломом [13]

Теорема. *Если* $S(X \times Y) > 0$, *то* $I(\xi, \eta) = \infty$. *Если* $S(X \times Y) = 0$, *то*

$$I(\xi, \eta) = \int\limits_X \int\limits_Y a(x, y) \log a(x, y)\, P_\xi(dx)\, P_\eta(dy) =$$
$$= \int\limits_X \int\limits_Y \log a(x, y)\, P_{\xi\eta}(dx\, dy) = \mathsf{M} \log \alpha_{\xi, \eta}. \tag{4}$$

Перечислим некоторые основные свойства величины $I(\xi, \eta)$.

I. $I(\xi, \eta) = I(\eta, \xi)$.

II. $I(\xi, \eta) \geqslant 0$; $I(\xi, \eta) = 0$ лишь в случае, если ξ и η независимы.

III. Если пары (ξ_1, η_1) и (ξ_2, η_2) независимы, то

$$I((\xi_1, \xi_2), (\eta_1, \eta_2)) = I(\xi_1, \eta_1) + I(\xi_2, \eta_2).$$

IV. $I((\xi, \eta), \zeta) \geqslant I(\eta, \zeta)$.

V. $I((\xi, \eta), \zeta) = I(\eta, \zeta)$
в том и только том случае, если ξ, η, ζ есть марковская последовательность, т. е. если условное распределение ζ при фиксированных ξ и η зависит только от η.

По поводу свойства IV полезно заметить следующее. В случае энтропии

$$H(\xi) = I(\xi, \xi),$$

кроме оценок энтропии пары (ξ, η) снизу

$$H(\xi, \eta) \geqslant H(\xi), \qquad H(\xi, \eta) \geqslant H(\eta).$$

вытекающих из I и IV, имеется оценка сверху

$$H(\xi, \eta) \leqslant H(\xi) + H(\eta).$$

Для количества информации в ζ относительно пары (ξ, η) аналогичной оценки не существует: из

$$I(\xi, \zeta) = 0, \qquad I(\eta, \zeta) = 0$$

еще не вытекает, как можно показать на элементарных примерах, равенство

$$I((\xi, \eta), \zeta) = 0.$$

Для дальнейшего отметим специально случай, когда ξ и η являются случайными векторами

$$\xi = (\xi_1, \ldots, \xi_m),$$

$$\eta = (\eta_1, \ldots, \eta_n) = (\xi_{m+1}, \ldots, \xi_{m+n})$$

и величины

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

распределены нормально со вторыми центральными моментами

$$s_{ij} = \mathsf{M}\,[(\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j)].$$

Если детерминант

$$C = |s_{ij}|, \quad 1 \leqslant i, j \leqslant m + n,$$

отличен от нуля, то, как было подсчитано И. М. Гельфандом и А. М. Ягломом,

$$I(\xi, \eta) = {}^1/_2 \log (AB/C), \tag{5}$$

где

$$A = |s_{ij}|, \quad 1 \leqslant i, j \leqslant m,$$

$$B = |s_{ij}|, \quad m < i, j \leqslant m + n.$$

Часто целесообразнее, впрочем, другой подход к делу, применимый без ограничения $C > 0$. Как известно [14], после надлежащего линейного преобразования координат в пространствах X и Y все вторые моменты s_{ij}, кроме тех, для которых $i = j$ или $j = m + i$, обратятся в нуль. При таком выборе координат

$$I(\xi, \eta) = -{}^1/_2 \log \sum_k [1 - r^2(\xi_k, \eta_k)], \tag{6}$$

где суммирование берется по тем

$$k \leqslant \min(m, n),$$

для которых знаменатель в выражении коэффициента корреляции

$$r(\xi_k, \eta_k) = s_{k,\, m+k}/\sqrt{s_{k,\, k}\, s_{m+k,\, m+k}}$$

отличен от нуля.

§ 2
АБСТРАКТНОЕ ИЗЛОЖЕНИЕ ОСНОВ ТЕОРИИ ШЕННОНА

Шеннон рассматривает передачу сообщений по схеме

$$\xi \to \eta \to \eta' \to \xi',$$

где «передающее устройство»

$$\eta \to \eta'$$

характеризуется условным распределением

$$P_{\eta'|\eta}(B' \mid y) = \mathsf{P}\{\eta' \in B' \mid \eta = y\}$$

«сигнала на выходе» η' при заданном «сигнале на входе» η и некоторыми ограничениями

$$P_\eta \in V$$

распределения P_η входного сигнала. Операции «кодирования»

$$\xi \to \eta$$

и «декодирования»

$$\eta' \to \xi'$$

характеризуются условными распределениями

$$P_{\eta|\xi}(B \mid x) = \mathsf{P}\{\eta \in B \mid \xi = x\},$$

$$P_{\xi'|\eta'}(A' \mid y') = \mathsf{P}\{\xi' \in A' \mid \eta' = y'\}.$$

Основная задача Шеннона заключается в следующем. Заданы пространства X, X', Y, Y' возможных значений «сообщений на входе» ξ, «сообщений на выходе» ξ', сигналов на входе η и сигналов на выходе η', заданы характеристики передающего устройства, т. е. условные распределения $P_{\eta'|\eta}$ и класс V допустимых распределений P_η входного сигнала; заданы, наконец, распределение

$$P_\xi(A) = \mathsf{P}\{\xi \in A\}$$

сообщения на входе и «требования к точности передачи»

$$P_{\xi\xi'} \in W,$$

где W — некоторый класс совместных распределений

$$P_{\xi\xi'}(C) = \mathsf{P}\{(\xi, \xi') \in C\}$$

сообщения на входе и сообщения на выходе. Спрашивается: возможно ли, и в случае возможности — каким способом, задать правила кодирования и декодирования (т. е. условные распределения $P_{\eta|\xi}$ и $P_{\xi'|\eta'}$) таким образом, что, вычисляя распределение $P_{\xi\xi'}$ по распределениям P_{ξ}, $P_{\eta|\xi}$, $P_{\eta'|\eta}$, $P_{\xi'|\eta'}$ в предположении, что последовательность

$$\xi, \eta, \eta', \xi'$$

марковская, получим

$$P_{\xi\xi'} \in W?$$

Определим вместе с Шенноном «пропускную способность» передающего устройства формулой

$$C = \sup_{P_{\eta} \in V} I(\eta, \eta')$$

и введем величину

$$H_W(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi'),$$

которую при расчете на единицу времени Шеннон называет «скоростью создания сообщений». Тогда из свойства V § 1 сразу вытекает необходимое условие возможности передачи

$$H_W(\xi) \leqslant C. \tag{7}$$

Как уже отмечалось, несравненно глубже идея Шеннона о том, что в применении к длительной работе «каналов связи» условие (7) является в некотором смысле и при некоторых весьма широких условиях и «почти достаточным». С математической точки зрения здесь дело идет о доказательстве предельных теорем следующего типа. Предполагается, что пространства X, X', Y, Y', распределения P_{ξ} и $P_{\eta'|\eta}$, классы V и W, а следовательно, и величины C и $H_W(\xi)$ зависят от параметра T (играющего в применениях роль длительности работы передающего устройства). Требуется установить при некоторых весьма общего характера предположениях, что условие

$$\liminf_{T \to \infty} \frac{C^T}{H_W^T(\xi)} > 1 \tag{8}$$

достаточно для возможности передачи, удовлетворяющей поставленным выше условиям, при достаточно больших T. Естественно, что в такой постановке задача несколько расплывчата. Однако я умышленно избегнул здесь обращения к терминологии теории стационарных случайных процессов, так как можно получить в намеченном направлении не лишенные интереса результаты и без предположения стационарности [15].

Выводу предельных теорем указанного типа для дискретного случая посвящено много замечательных работ. Среди них особенно следует отметить уже называвшуюся работу А. Я. Хинчина [6].

Для общего непрерывного случая здесь собственно еще почти ничего не сделано.

§ 3
ВЫЧИСЛЕНИЕ И ОЦЕНКА ε-ЭНТРОПИИ В НЕКОТОРЫХ ЧАСТНЫХ СЛУЧАЯХ

Если условие

$$P_{\xi\xi'} \in W$$

выбрать в виде обязательности точного совпадения ξ и ξ'

$$\mathsf{P}\,\{\xi = \xi'\} = 1,$$

то

$$H_W(\xi) = H(\xi).$$

В соответствии с этим кажется естественным назвать в общем случае $H_W(\xi)$ «энтропией случайного объекта ξ при точности воспроизведения W».

Предположим теперь, что пространство X' совпадает с X, т. е. исследуются способы приближенной передачи сообщения о положении точки $\xi \in X$ при помощи указания точки ξ' того же пространства X, и что в пространстве введено «расстояние» $\rho(x, x')$, удовлетворяющее обычным аксиомам «метрических пространств». Кажется естественным потребовать, чтобы

$$\mathsf{P}\,\{\rho(\xi, \xi') \leqslant \varepsilon\} = 1 \qquad (W_\varepsilon^0)$$

или чтобы

$$\mathsf{M}\rho^2(\xi, \xi') \leqslant \varepsilon^2. \qquad (W_\varepsilon)$$

Эти два вида «ε-энтропии» распределения P_ξ будем обозначать

$$H_{W_\varepsilon^0}(\xi) = H_\varepsilon^0(\xi), \qquad H_{W_\varepsilon}(\xi) = H_\varepsilon(\xi).$$

Что касается ε-энтропии $H_\varepsilon^0(\xi)$, то я хочу здесь лишь отметить некоторые оценки для

$$H_\varepsilon^0(X) = \sup_{P_\xi} H_\varepsilon^0(\xi),$$

где верхняя грань берется по всем распределениям вероятностей P_ξ на пространстве X. При $\varepsilon = 0$, как известно,

$$H_0^0(X) = \sup_{P_\xi} H(\xi) = \log N_X,$$

где N_X есть число элементов множества X. При $\varepsilon > 0$

$$\log N_X^c(2\varepsilon) \leqslant H_\varepsilon^0(X) \leqslant \log N_X^a(\varepsilon),$$

где $N_X^a(\varepsilon)$ и $N_X^c(\varepsilon)$ являются характеристиками пространства X, которые введены в моей заметке [3]. Асимптотические свойства при $\varepsilon \to 0$ функций $N_X(\varepsilon)$, изученные для ряда конкретных пространств X в [3], являются интересными аналогами излагаемых далее свойств асимптотического поведения функции $H_\varepsilon(\xi)$.

Обратимся теперь к ε-энтропии $H_\varepsilon(\xi)$. Если X есть n-мерное евклидово пространство и

$$P_\xi(A) = \int_A p_\xi(x)\,dx_1\,dx_2\ldots dx_n,$$

то по крайней мере в случае достаточно гладкой функции $p_\xi(x)$ имеет место хорошо известная формула

$$H_\varepsilon(\xi) = n\log(1/\varepsilon) + [h(\xi) - n\log\sqrt{2\pi e}] + o(1), \tag{9}$$

где

$$h(\xi) = -\int_X p_\xi(x)\log p_\xi(x)\,dx_1\,dx_2\ldots dx_n$$

— «дифференциальная энтропия», введенная уже в первых работах Шеннона. Таким образом, асимптотическое поведение $H_\varepsilon(\xi)$ в случае достаточно гладких непрерывных распределений в n-мерном пространстве определяется в первую очередь размерностью пространства и лишь в виде второго члена в выражение $H_\varepsilon(\xi)$ входит дифференциальная энтропия $h(\xi)$.

Естественно ожидать, что для типичных распределений в бесконечномерных пространствах рост $H_\varepsilon(\xi)$ при $\varepsilon \to 0$ будет существенно более быстрым. В виде простейшего примера рассмотрим случайную функцию Винера $\xi(t)$, определенную для $0 \leqslant t \leqslant 1$ с нормально распределенными независимыми приращениями:

$$\Delta\xi = \xi(t + \Delta t) - \xi(t),$$

для которой

$$\xi(0) = 0, \qquad \mathsf{M}\Delta\xi = 0, \qquad \mathsf{M}(\Delta\xi)^2 = \Delta t.$$

А. М. Яглом нашел, что в этом случае в метрике L^2

$$H_\varepsilon(\xi) = \frac{4}{\pi}\frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right). \tag{10}$$

Более общим образом для процесса Маркова диффузного типа на отрезке времени $t_0 \leqslant t \leqslant t_1$ с

$$\mathsf{M}\Delta\xi = A(t, \xi(t))\,\Delta t + o(\Delta t),$$

$$\mathsf{M}(\Delta\xi)^2 = B(t, \xi(t))\,\Delta t + o(\Delta t)$$

можно при некоторых естественных допущениях получить формулу

$$H_\varepsilon(\xi) = \frac{4}{\pi}\chi\frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right), \qquad (11)$$

где

$$\chi(\xi) = \int_{t_0}^{t_1} \mathsf{M}B(t, \xi(t))\,dt.$$

Для случая нормального распределения в n-мерном евклидовом или в гильбертовом пространстве ε-энтропия H_ε может быть вычислена точно; n-мерный вектор ξ после надлежащего ортогонального преобразования координат приобретает вид

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n),$$

где координаты ξ_k взаимно независимы и распределены нормально. При заданном ε параметр θ определяется из уравнения

$$\varepsilon^2 = \Sigma \min(\theta^2, \mathsf{D}\xi_k)$$

и в случае нормально распределенного ξ

$$H_\varepsilon(\xi) = \frac{1}{2}\sum_{\mathsf{D}\xi_k > \theta^2} \log\frac{\mathsf{D}\xi_k}{\theta^2}. \qquad (12)$$

Аппроксимирующий вектор

$$\xi' = (\xi_1', \xi_2', \ldots, \xi_n')$$

следует выбирать так, что при $\mathsf{D}\xi_k \leqslant \theta^2$

$$\xi_k' = 0,$$

а при $\mathsf{D}\xi_k > \theta^2$

$$\xi_k = \xi_k' + \Delta_k, \qquad \mathsf{D}\Delta_k = \theta^2, \qquad \mathsf{D}\xi_k' = \mathsf{D}\xi_k - \theta^2$$

и векторы ξ_k и Δ_k взаимно независимы. Бесконечномерный случай ничем не отличается от конечномерного.

Наконец, весьма существенно, что максимальное значение $H_\varepsilon(\xi)$ для вектора ξ (n-мерного или бесконечномерного) при заданных вторых центральных моментах достигается в случае нормального распределения. Этот результат может быть получен непосредственно или из следующего предложения М. С. Пинскера (ср. [16]):

Теорема. *Пусть заданы положительно определенная симметричная матрица величин* s_{ij}, $0 \leqslant i, j \leqslant m + n$, *и распределение* P_ξ *вектора*

$$\xi = (\xi_1, \xi_2, \ldots, \xi_m),$$

центральные вторые моменты которого равны s_{ij} (при $0 \leqslant i, j \leqslant m$). Пусть условие W на совместное распределение $P_{\xi\xi'}$ вектора ξ и вектора

$$\xi' = (\xi_{m+1}, \xi_{m+2}, \ldots, \xi_{m+n})$$

заключается в том, что центральные вторые моменты величин

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

равны s_{ij} (при $0 \leqslant i, j \leqslant m+n$). Тогда

$$H_W(\xi) \leqslant {}^1/_2 \log(AB/C). \tag{13}$$

Обозначения в формуле (13) соответствуют изложению § 1. Из сопоставления со сказанным в § 1 видно, что неравенство (13) превращается в равенство в случае нормального распределения P_ξ.

Принципы решения вариационных задач, возникающих при вычислении «скорости создания сообщений», были указаны довольно давно Шенноном. В [10] Шеннон и Уивер пишут: «К сожалению, эти формальные решения в частных случаях трудно численно оценить, и поэтому ценность их представляется небольшой» [6]. По существу, однако, многие задачи такого рода, как видно из сказанного выше, достаточно просты. Возможно, что медленное развитие исследований в этом направлении связано с недостаточным пониманием того обстоятельства, что в типичных случаях решения вариационных задач оказываются очень часто вырожденными; например, в разобранной выше задаче вычисления $H_\varepsilon(\xi)$ для нормально распределенного вектора ξ в n-мерном случае вектор ξ' часто оказывается не n-мерным, а лишь k-мерным с $k < n$, в бесконечномерном же случае вектор ξ' оказывается всегда конечномерным.

§ 4

КОЛИЧЕСТВО ИНФОРМАЦИИ И СКОРОСТЬ СОЗДАНИЯ СООБЩЕНИЙ В СЛУЧАЕ СТАЦИОНАРНЫХ ПРОЦЕССОВ

Рассмотрим два стационарных и стационарно связанных процесса

$$\xi(t), \eta(t), \quad -\infty < t < \infty.$$

Будем обозначать через ξ_T и η_T отрезки процессов ξ и η за время $0 < t \leqslant T$ и через ξ_- и η_- течение процессов ξ и η на отрицательной полуоси $-\infty < t \leqslant 0$. Задать пару (ξ, η) стационарно связанных процессов ξ и η — это значит задать инвариантное по отношению к сдвигам вдоль оси t распределение вероятностей $P_{\xi\eta}$

[6] См. рус. пер. [10, § 28, с. 79].

в пространстве пар функций $\{x(t), y(t)\}$. Если зафиксировать ξ_-, то из распределения $P_{\xi\eta}$ возникает условное распределение

$$P_{\xi_T, \eta \mid \xi_-}(C \mid x_-) = \mathsf{P}\{(\xi_T, \eta) \in C \mid \xi_- = x_-\}.$$

При помощи этого распределения в соответствии с § 1 вычисляется условное количество информации

$$I(\xi_T, \eta \mid x_-).$$

Если математическое ожидание

$$\mathsf{M} I(\xi_T, \eta \mid \xi_-)$$

конечно при каком-либо $T > 0$, то оно конечно и при всех других $T > 0$ и

$$\mathsf{M} I(\xi_T, \eta \mid \xi_-) = T \vec{I}(\xi, \eta).$$

Величину $\vec{I}(\xi, \eta)$ естественно назвать «скоростью создания информации о процессе η при наблюдении процесса ξ». Если процесс ξ с полной точностью экстраполируется из прошлого в будущее, то

$$\vec{I}(\xi, \eta) = 0.$$

Так будет, в частности, если процесс ξ имеет ограниченный спектр. Вообще говоря, не обязательно имеет место равенство

$$\vec{I}(\xi, \eta) = \vec{I}(\eta, \xi). \qquad (14)$$

Однако при довольно широких условиях «регулярности» процесса[7] ξ имеет место равенство

$$\vec{I}(\xi, \eta) = \bar{I}(\xi, \eta),$$

где

$$\bar{I}(\xi, \eta) = \lim_{T \to \infty} \frac{1}{T} I(\xi_T, \eta_T).$$

Так как $I(\xi_T, \eta_T) = I(\eta_T, \xi_T)$, то всегда

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi)$$

и, следовательно, в случае справедливости обоих равенств $\vec{I}(\xi, \eta) = \bar{I}(\xi, \eta)$ и $\vec{I}(\eta, \xi) = \bar{I}(\eta, \xi)$ имеет место равенство (14).

Пусть теперь W есть некоторый класс совместных распределений $P_{\xi\xi'}$ двух стационарных и стационарно связанных процессов ξ и ξ'. Естественно назвать величину

$$\vec{H}_W(\xi) = \inf_{P_{\xi\xi'} \in W} \vec{I}(\xi', \xi)$$

[7] Регулярность процесса здесь и далее обозначает, грубо говоря, что отрезки процесса, соответствующие двум достаточно удаленным друг от друга отрезкам оси t, почти независимы. В случае гауссовских процессов здесь применимо хорошо известное определение регулярности, введенное в моей работе [17].

«скоростью создания сообщений в процессе ξ при точности воспроизведения W». При некоторых предположениях регулярности процесса ξ и для некоторых естественных типов условий W можно доказать, что

$$\vec{H}_W(\xi) = \overline{H}_W(\xi),$$

где

$$\overline{H}_W = \inf_{P_{\xi\xi'} \in W} \bar{I}(\xi, \xi').$$

§ 5
ВЫЧИСЛЕНИЕ И ОЦЕНКА КОЛИЧЕСТВА ИНФОРМАЦИИ И СКОРОСТИ СОЗДАНИЯ СООБЩЕНИЙ ПО СПЕКТРУ

Для случая, когда распределение $P_{\xi\eta}$ нормально и хотя бы один из процессов ξ или η регулярен, имеет место предложенная М. С. Пинскером [18] формула

$$\bar{I}(\xi, \eta) = -\frac{1}{4\pi} \int_{-\infty}^{+\infty} \log[1 - r^2(\lambda)]\, d\lambda,$$

где

$$r^2(\lambda) = \frac{|f_{\xi\eta}(\lambda)|^2}{f_{\xi\xi}(\lambda)\, f_{\eta\eta}(\lambda)}, \tag{15}$$

а $f_{\xi\xi}$, $f_{\xi\eta}$, $f_{\eta\eta}$ — спектральные плотности. В случае процессов с дискретным временем t для дифференциальной энтропии нормального процесса на единицу времени

$$\bar{h}(\xi) = \lim_{T \to \infty} \frac{1}{T} h(\xi_1, \xi_2, \ldots, \xi_T)$$

известно выражение

$$\bar{h}(\xi) = \log(2\pi \sqrt{e}) + \frac{1}{2\pi} \int_{-\pi}^{+\pi} \log f_{\xi\xi}(\lambda)\, d\lambda. \tag{16}$$

Однако в случае непрерывного времени и неограниченного спектра никакого аналога выражения $\bar{h}(\xi)$ не существует и формула Пинскера требует самостоятельного вывода.

Естественно характеризовать точность воспроизведения стационарного процесса ξ при помощи стационарного и стационарно связанного с ξ процесса ξ' величиной

$$\sigma^2 = \mathsf{M}\,[\xi(t) - \xi'(t)]^2$$

и в случае условия W вида

$$\sigma^2 \leqslant \varepsilon^2$$

называть величину

$$\overline{H}_\varepsilon(\xi) = \overline{H}_W(\xi)$$

ε-энтропией на единицу времени процесса ξ, а в предположении, что

$$\overline{H}_W(\xi) = \overline{H}_W(\xi),$$

скоростью создания сообщений в процессе ξ при средней точности передачи ε. Из соответствующего предложения для конечномерных распределений (см. § 3) можно вывести, что при заданной спектральной плотности $f_{\xi\xi}(\lambda)$ величина $\overline{H}_\varepsilon(\xi)$ достигает максимума в случае нормального процесса ξ. В нормальном случае величина $\overline{H}_\varepsilon(\xi)$ может быть легко вычислена по спектральной плотности

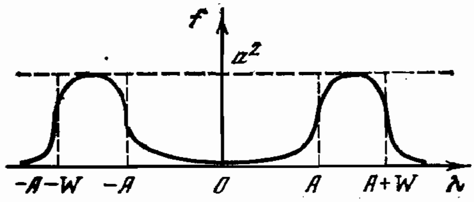


Рис. 2

$f_{\xi\xi}(\lambda)$ вполне аналогично тому, как это было объяснено в § 3 в применении к величине $H_\varepsilon(\xi)$ для n-мерных распределений. Параметр θ определяется из уравнения

$$\varepsilon^2 = \int_{-\infty}^{+\infty} \min(\theta^2, f_{\xi\xi}(\lambda))\, d\lambda. \tag{17}$$

При помощи этого параметра величина $\overline{H}_\varepsilon(\xi)$ находится по формуле

$$\overline{H}_\varepsilon(\xi) = \frac{1}{2} \int_{f_{\xi\xi}(\lambda) > \theta^2} \log \frac{f_{\xi\xi}(\lambda)}{\theta^2}\, d\lambda. \tag{18}$$

Практический интерес представляют спектральные плотности вида (рис. 2), которые хорошо аппроксимируются функцией

$$\varphi(\lambda) = \begin{cases} a^2 & \text{при } A \leqslant |\lambda| \leqslant A + W, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Легко подсчитать, что в этом случае при не слишком малых ε приближенно для нормального процесса

$$\theta^2 \sim \varepsilon^2/2W, \qquad \overline{H}_\varepsilon(\xi) \sim W \log(2Wa^2/\varepsilon^2). \tag{19}$$

Формула (19), конечно, есть не что иное, как хорошо известная формула Шеннона

$$R = W \log (Q/N). \tag{20}$$

Принципиальная новизна заключается здесь, однако, в том, что теперь мы видим, почему и в каких пределах (при не слишком малых ε) эта формула может быть применима к процессам с неограниченным спектром, а таковы все реально интересующие нас в теории передачи сообщений процессы. Записывая (19) в виде

$$\overline{H}_\varepsilon (\xi) \approx 2W (\log (1/\varepsilon) + \log (a \sqrt{W})) \tag{21}$$

и сравнивая с (9), мы видим, что двойная ширина $2W$ используемой полосы частот играет роль числа измерений. Эта идея эквивалентности двойной ширины полосы частот числу измерений, приходящихся, в некотором смысле слова, на единицу времени, была, по-видимому, впервые высказана В. А. Котельниковым (см. [12]). В обоснование этой идеи Котельников указывал на то обстоятельство, что функция, спектр которой помещается в полосе ширины $2W$, однозначно определяется значениями функции в точках

$$\ldots, -\frac{2}{2W}, -\frac{1}{2W}, 0, \frac{1}{2W}, \frac{2}{2W}, \ldots, \frac{k}{2W}, \ldots$$

Эта же аргументация сохранена и у Шеннона, использующего полученные таким образом представления и для вывода формулы (20). Так как функция с ограниченным спектром всегда сингулярна в смысле [17] и наблюдение такой функции вообще не связано со стационарным притоком новой информации, то смысл такого рода аргументации оставался не вполне ясным, так что приведенный здесь новый вывод приближенной формулы (21) представляется мне не лишенным интереса.

При малых ε для любой нормально распределенной регулярной случайной функции рост $\overline{H}_\varepsilon (\xi)$ с убыванием ε происходит существенно быстрее, чем это получалось бы по формуле (21). В частности, если $f_{\xi\xi} (\lambda)$ при $\lambda \to \infty$ имеет порядок $\lambda^{-\beta}$, то $\overline{H}_\varepsilon (\xi)$ имеет порядок $\varepsilon^{-2/(\beta-1)}$.

§ 6

ВЫЧИСЛЕНИЕ И ОЦЕНКА ПРОПУСКНОЙ СПОСОБНОСТИ В НЕКОТОРЫХ ЧАСТНЫХ СЛУЧАЯХ

Сначала рассмотрим случай, когда входной сигнал является m-мерным вектором

$$\eta = (\eta_1, \ldots, \eta_m),$$

а выходной сигнал — n-мерным вектором

$$\eta' = (\eta'_1, \ldots, \eta'_n).$$

Как уже говорилось, передающее устройство $\eta \to \eta'$ характеризуется условным распределением вероятностей $P_{\eta'|\eta}$ и некоторыми ограничениями на входной сигнал η. Предположим, что зависимость η' от η имеет характер линейной нормальной корреляции, т. е. что

$$\eta' = A\eta + \zeta, \tag{22}$$

где оператор A линеен, а вектор ζ независим от η и подчинен n-мерному гауссовскому распределению. Что касается условий на входной сигнал, то допустим, что они имеют вид

$$\mathsf{M}Q(\eta) \leqslant \varepsilon^2, \tag{23}$$

где Q есть некоторая положительно определенная квадратичная форма от координат $\eta_1, \ldots, \eta_m$. При помощи надлежащего выбора линейных преобразований координат в пространствах Y и Y' общий случай сводится к случаю, когда

$$Q(\eta) = \Sigma\eta_i^2, \tag{24}$$

а зависимость (22) записывается в виде

$$\begin{aligned} \eta'_i &= a_i\eta_i + \zeta_i,\ a_i \neq 0, \quad \text{при } 1 \leqslant i \leqslant k, \\ \eta'_i &= \zeta_i \quad \text{при } i > k, \end{aligned} \tag{25}$$

где $k = \min(m, n)$.

В сделанных предположениях пропускная способность, т. е. верхняя грань C количества информации $I(\eta, \eta')$, совместимая с наложенными условиями, легко находится. Она достигается, если η_i — взаимно независимые случайные гауссовские величины

$$\begin{aligned} \mathsf{D}\eta_i &= \theta^2 - \mathsf{D}\zeta_i/a_i^2, \qquad \text{если } i \leqslant k \text{ и } a_i^2\theta^2 > \mathsf{D}\zeta_i, \\ \mathsf{D}\eta_i &= 0 \text{ при остальных } i, \end{aligned} \tag{26}$$

где константа θ^2, как легко видеть, однозначно определяется требованиями (26) и

$$\sum_i \mathsf{D}\eta_i = \varepsilon^2. \tag{27}$$

Соответствующее значение $C = I(\eta, \eta')$ есть (см. § 3)

$$C = \frac{1}{2} \sum_{a_i^2\theta^2 > \mathsf{D}\zeta_i} \log \frac{a_i^2\theta^2}{\mathsf{D}\zeta_i}. \tag{28}$$

Аналогично дело обстоит в случае стационарных линейных каналов связи с «гауссовскими шумами», т. е. в случае

$$\eta'(t) = A\eta(t) + \zeta(t), \tag{29}$$

где A — линейный оператор, а $\zeta(t)$ — независимая от $\eta(t)$ гауссовская стационарная случайная функция. Как известно, соответствующие спектральные плотности связаны соотношением

$$f_{\eta'\eta'}(\lambda) = a^2(\lambda) f_{\eta\eta}(\lambda) + f_{\zeta\zeta}(\lambda). \tag{30}$$

Условие, ограничивающее мощность входного канала, зададим в форме

$$\int_{-\infty}^{+\infty} f_{\eta\eta}(\lambda)\, d\lambda \leqslant \varepsilon^2. \tag{31}$$

Формальное решение задачи вполне аналогично решению рассмотренной выше конечномерной задачи. Наибольшая пропускная способность канала $\overline{C}$ достигается, если

$$f_{\eta\eta}(\lambda) = \theta^2 - f_{\zeta\zeta}(\lambda)/a^2(\lambda) \ \text{при}\ a^2(\lambda)\,\theta^2 > f_{\zeta\zeta}(\lambda),$$

$$f_{\eta\eta}(\lambda) = 0 \ \text{при}\ a^2(\lambda)\,\theta^2 \leqslant f_{\zeta\zeta}(\lambda). \tag{32}$$

Соответствующее значение $\overline{I}(\eta, \eta')$ по формуле (15) равно

$$\overline{C} = \frac{1}{4\pi} \int\limits_{a^2(\lambda)\theta^2 > f_{\zeta\zeta}(\lambda)} \log \frac{a^2(\lambda)\,\theta^2}{f_{\zeta\zeta}(\lambda)}\, d\lambda. \tag{33}$$

Следует тут же заметить, что в практически интересных случаях вычисленная по формулам (32) спектральная плотность $f_{\eta\eta}(\lambda)$ обращается в нуль за пределами ограниченного участка оси λ и поэтому является спектральной плотностью сингулярного процесса. Процесс $\eta'(t)$ будет при этом (в естественном предположении регулярности процесса $\zeta(t)$) смешанным, т. е. состоять из суммы сингулярной компоненты $A\eta$ и регулярной компоненты ζ. Скорость $\vec{I}(\eta', \eta)$ создания информации о процессе η при наблюдении процесса η' в таком случае вовсе не выражается формулой $\vec{I}(\eta', \eta) = \overline{C}$, а равна нулю (см. § 5).

При помощи некоторых дополнительных рассмотрений можно, однако, показать, что, вводя в виде входного сигнала $\eta(t)$ регулярные процессы, которые в принципе практически реализуемы и действительно несут возникающую во времени информацию, можно получить скорость создания информации $\vec{I}(\eta', \eta) = \vec{I}(\eta, \eta')$, сколь угодно близкую к $\overline{C}$. Это окончательно оправдывает формулу (33).

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* Статистическая теория колебаний с непрерывным спектром.— В кн.: Общее собрание Академии наук, посвященное тридцатилетию Великой Октябрьской социалистической революции. М.; Л.: Изд-во АН СССР, 1948, с. 465—472.
2. *Колмогоров А. Н.* Статистическая теория колебаний с непрерывным спектром.— В кн.: Юбилейный сборник, посвященный тридцатилетию

Великой Октябрьской социалистической революции. М.; Л.: Изд-во АН СССР, 1947, т. 1, с. 242—252.
3. *Колмогоров А. Н.* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств.— ДАН СССР, 1956, т. 108, № 3, с. 385—388.
4. *Витушкин А. Г.* К тринадцатой проблеме Гильберта.— ДАН СССР, 1954, т. 95, № 4, с. 701—704.
5. *Хинчин А. Я.* Понятие энтропии в теории вероятностей.— УМН, 1953, т. 8, вып. 3, с. 3—20.
6. *Хинчин А. Я.* Об основных теоремах теории информации.— УМН, 1956, т. 11, вып. 1, с. 17—75.
7. Теория передачи электрических сигналов при наличии помех: Сб. пер. М.: Изд-во иностр. лит., 1953.
8. *Харкевич А. А.* Очерки общей теории связи. М.: Гостехиздат, 1955.
9. *Goldman S.* Information theory. London: Constable and Co., 1953.
10. *Shannon C. E., Weaver W.* The mathematical theory of communications. Urbana: Univ. Ill. Press, 1949, p. 3—89. Рус. пер. в кн.: Теория передачи электрических сигналов при наличии помех. М.: Изд-во иностр. лит., 1953.
11. *Крамер Г.* Математические методы статистики. М.: Изд-во иностр. лит., 1948.
12. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи.— В кн.: Матер. к I Всесоюз. съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. М., 1933.
13. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* Количество информации и энтропия для непрерывных распределений.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1958, т. 3, с. 300—320.
14. *Обухов А. М.* Нормальная корреляция векторов.— Уч. зап. МГУ, 1940, вып. 45, с. 73—92.
15. *Розенблат-Рот М.* Понятие энтропии в теории вероятностей и его применения в теории передачи по каналам связи.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1956, т. 2, с. 132—133.
16. *Пинскер М. С.* Количество информации об одном стационарном случайном процессе, содержащееся в другом стационарном случайном процессе.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1956, т. 1, с. 125.
17. *Колмогоров А. Н.* Стационарные последовательности в гильбертовом пространстве.— Бюл. МГУ. Математика, 1941, т. 2, № 6, с. 1—40.
18. *Пинскер М. С.* Количество информации о гауссовском случайном стационарном процессе, содержащейся во втором процессе, стационарно с ним связанном.— ДАН СССР, 1954, т. 99, № 2, с. 213—216.

ВЫСТУПЛЕНИЕ

Разногласия между мной и А. А. Ляпуновым по вопросу о кибернетике имеют, мне кажется, более словесный характер. Несомненно, что математическая теория конечных алгоритмов с ограниченной памятью и с полным учетом фактора времени (число шагов до получения результата, различение между быстро и медленно действующей памятью и т. п.) должна стабилизироваться в виде самостоятельной и систематически развивающейся математической дисциплины. У нас некоторые товарищи, и в том числе А. А. Ляпунов, именно такую теорию, которая объединяет математические вопросы, связанные с работой вычислительных машин и управляющих устройств дискретного действия, предлагают называть кибернетикой. Однако слово кибернетика было введено в обращение Н. Винером, который придал ему несравненно более широкий смысл. Кибернетика по Винеру должна включать в себя весь математический аппарат теории управляющих устройств дискретного или непрерывного действия.

Это означает, что в кибернетику по Винеру входит значительная часть теории устойчивости для систем дифференциальных уравнений, классическая теория регулирования, теория случайных процессов с их экстраполированием и фильтрацией, теория информации, теория игр с применениями к анализу операций, технические аспекты алгебры логики, теория программирования и еще многое другое. Легко понять, что как математическая дисциплина кибернетика в понимании Винера лишена единства, и трудно себе представить продуктивную работу по подготовке специалиста, скажем аспиранта, по кибернетике в этом смысле. Если А. А. Ляпунову удастся сформулировать отчетливо программу развития более узкой и обладающей единством методов дисциплины, то, по моему мнению, ее лучше было бы не называть кибернетикой.

Я во многом согласен с выступлением В. С. Пугачева. Хочется только отметить, что понятие ε-энтропии по существу введено еще самим Шенноном под названием скорости создания сообщений. Предложенное мною переименование и подчеркивание основной роли этого понятия, мне кажется, не лишенным значения в смысле правильного направления дальнейших работ, но особой новизны здесь нет. Больше оригинального, по моему мнению, сделали советские математики в отношении развития спектральных методов. Я думаю, что сделанное в этом направлении М. С. Пинскером и мною найдет достаточно интересное продолжение. Существенно полное математическое обоснование замечательной концепции В. А. Котельникова об эквивалентности широты полосы частот и числа степеней свободы на единицу времени.[1]

Несколько слов по вопросу о кадрах. Сейчас имеется большой недостаток в математиках, творчески владеющих вероятностными методами. В прикладных технических областях, быть может, только специалисты по теории стрельбы имеют сложившиеся кадры исследователей с подобным творческим и активным владением вероятностными методами. Ряд других традиционных применений теории вероятностей через математическую статистику в области экономики, медицины, сельского хозяйства и биологии уже давно несколько захирел, не всегда обоснован. Статистические методы в вопросах массовой продукции у нас разрабатываются, но слишком маленькой группой исследователей. Сейчас в связи с возросшим интересом к теории информации, статистическим методам изучения регулирующих устройств, методу Монте-Карло в вычислительной технике и с рядом других новых запросов мы делаем в Московском университете попытку расширения подготовки математиков, способных работать по техническим применениям вероятностных методов. Из числа окончивших в МГУ аспирантуру недавно начали работать по вопросам, связанным с теорией информации, в научно-технических институтах М. С. Пинскер и Р. Л. Добрушин. В 1957 г. мы можем дать около 20 окончивших университет с большой подготовкой по теории вероятностей, теории случайных процессов и математической статистике. Хотя это еще студенты, некоторые из них являются уже начинающими самостоятельными молодыми исследователями.

В связи с выступлением Л. В. Канторовича о математических методах планирования я хочу отметить, что центр тяжести вопроса здесь не в применении скоростной вычислительной техники. Вычислительные машины в этом деле, конечно, окажутся полезными, но пока у нас не делается в этом интересном направлении и того, что было давно возможно на базе традиционной вычислительной техники.

По вопросу о математических переводах я согласен с И. С. Бруком в том отношении, что для практического общения знание языков или употребление живых переводчиков вряд ли в обозримый срок будет заменено машинами. Но уже сейчас несомненно, что работа над машинным переводом очень многое дает и конструкторам вычислительных машин и, что особенно интересно, лингвистам, которые в процессе этой работы принуждены уточнять формулируемые ими законы языка. Они здесь приобретают критерий практики, столь важный для развития теории: если данная ими формулировка, скажем, правил расположения слов в каком-либо языке достаточна для получения при помощи машины хорошего литературного перевода, то получается действительная уверенность в том, что правила формулированы исчерпывающим образом. Здесь уже нельзя отделаться общими фразами и выдавать их за «законы», к чему в гуманитарных науках иногда имеют склонность. Но, уже в порядке шутки, позвольте здесь выразить надежду, что в следующем 1957 г. молодежь разных стран на юношеском фестивале в Москве будет разговаривать без помощи машин и найдет более простые и человеческие формы общения.

4
КОЛИЧЕСТВО ИНФОРМАЦИИ И ЭНТРОПИЯ ДЛЯ НЕПРЕРЫВНЫХ РАСПРЕДЕЛЕНИЙ*[1]

Совместно с И. М. Гельфандом и А. М. Ягломом

§ 1
ОПРЕДЕЛЕНИЕ КОЛИЧЕСТВА ИНФОРМАЦИИ, СОДЕРЖАЩЕЙСЯ В ОДНОМ СЛУЧАЙНОМ ОБЪЕКТЕ ОТНОСИТЕЛЬНО ДРУГОГО

Пусть $\mathfrak{A}$ — некоторое испытание, которое может иметь n различных исходов $A_1, A_2, \ldots, A_n$ с вероятностями $P(A_1), P(A_2), \ldots$ $\ldots, P(A_n)$; $\mathfrak{B}$ — другое испытание, имеющее исходы $B_1, B_2, \ldots$ $\ldots, B_m$ с вероятностями $P(B_1), P(B_2), \ldots, P(B_m)$. Если через $P(A_iB_k)$ мы обозначим вероятность совмещения исхода A_i испытания $\mathfrak{A}$ и исхода B_k испытания $\mathfrak{B}$, то согласно Шеннону [1] среднее количество информации, содержащейся в результатах испытания $\mathfrak{A}$ относительно результатов испытания $\mathfrak{B}$, равно

$$I(\mathfrak{A}, \mathfrak{B}) = \sum_{i,k} P(A_iB_k) \log \frac{P(A_iB_k)}{P(A_i)P(B_k)}. \tag{1.1}$$

Количество информации, содержащейся в результатах испытания $\mathfrak{A}$ относительно себя самого (т. е. полное количество информации, содержащейся в результатах испытания $\mathfrak{A}$), называется *энтро-*

---

* В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1958, т. 3, с. 300—320.

[1] Доклад посвящен в основном определению, общим свойствам и способам вычисления величин $I(\xi, \eta)$ и $H_W(\xi)$. Общее определение $I(\xi, \eta)$ содержится в добавлении 7 к работе Шеннона [1] (пользуясь русским переводом указанной работы, в котором добавление 7 пропущено, мы не обратили на это своевременно внимания, в частности при редактировании заметки [22]). Однако до нашей заметки [22] подробного изложения принципов развития общей теории на основании этого определения, по-видимому, опубликовано не было. Со сделанной выше оговоркой история возникновения результатов, изложенных в § 1 настоящего доклада, указана в работе [22]. В § 2 даны указания относительно значения величин $I(\xi, \eta)$ и $H_W(\xi)$ для теории передачи информации по каналам связи (об этом см., кроме того, в работе [21]; большая часть результатов и общих соображений, изложенных в [21], воспроизведена на английском языке в работе [24]). Результаты, данные в § 3 без ссылок на других авторов, получены И. М. Гельфандом и А. М. Ягломом (см. [23]), а в § 4,5 — А. Н. Колмогоровым (кроме констант в формулах (4.13) и (5.12), вычисленных А. М. Ягломом). Текст доклада написан А. Н. Колмогоровым и А. М. Ягломом.

пией испытания $\mathfrak{A}$ и обозначается через $H(\mathfrak{A})$:

$$H(\mathfrak{A}) = I(\mathfrak{A}, \mathfrak{A}) = -\sum_i P(A_i) \log P(A_i). \tag{1.2}$$

Если обозначить через $H(\mathfrak{A}/B_k)$ «условную энтропию» испытания $\mathfrak{A}$, получаемую при замене в формуле (1.2) вероятностей $P(A_i)$ условными вероятностями $P_{B_k}(A_i) = P(A_i/B_k)$, а через $H(\mathfrak{A}/\mathfrak{B})$ — случайную величину, принимающую с вероятностью $P(B_k)$ значение $H(\mathfrak{A}/B_k)$, то $I(\mathfrak{A}, \mathfrak{B})$ можно представить также в виде

$$I(\mathfrak{A}, \mathfrak{B}) = H(\mathfrak{A}) - \mathsf{M}H(\mathfrak{A}/\mathfrak{B}), \tag{1.3}$$

где M — знак математического ожидания.

Для того чтобы обобщить определение (1.1) на случай произвольных вероятностных объектов, удобно сперва переформулировать его на языке нормированных булевых алгебр. А именно мы будем исходить из нормированной булевой алгебры $\mathfrak{S}$ всевозможных «случайных событий» $A, B, C, \ldots$ с операциями образования дополнительного элемента A', соединения («сложения») $A \cup B$, умножения AB, единичным элементом E и нулевым элементом N; нормой здесь является распределение вероятностей $P(A)$, являющееся неотрицательной аддитивной функцией на $\mathfrak{S}$, нормированной условием $P(E) = 1$. При этом понятие «испытания» естественно идентифицировать с понятием подалгебры алгебры $\mathfrak{S}$ (подалгебры всех тех событий, исход которых становится известным после совершения данного испытания).

Известно, что каждая конечная булева алгебра $\mathfrak{A}$ совпадает с системой всевозможных сумм своих «атомов» $A_1, A_2, \ldots, A_n$. Если обе подалгебры $\mathfrak{A}$ и $\mathfrak{B}$ конечны, то количество информации, содержащейся в результатах испытания $\mathfrak{A}$ относительно результатов испытания $\mathfrak{B}$, мы определяем формулой Шеннона (1.1), где $A_1, \ldots, A_n$ — атомы $\mathfrak{A}$, а $B_1, \ldots, B_m$ — атомы $\mathfrak{B}$.

В случае произвольных подалгебр $\mathfrak{A}$ и $\mathfrak{B}$ естественно положить

$$I(\mathfrak{A}, \mathfrak{B}) = \sup_{\mathfrak{A}_1 \subseteq \mathfrak{A},\ \mathfrak{B}_1 \subseteq \mathfrak{B}} I(\mathfrak{A}_1, \mathfrak{B}_1), \tag{1.4}$$

где верхняя грань берется по всем конечным подалгебрам $\mathfrak{A}_1 \subseteq \subseteq \mathfrak{A}$ и $\mathfrak{B}_1 \subseteq \mathfrak{B}$. Символически определение (1.4) можно записать в виде

$$I(\mathfrak{A}, \mathfrak{B}) = \int_{\mathfrak{A}} \int_{\mathfrak{B}} P(d\mathfrak{A}\, d\mathfrak{B}) \log \frac{P(d\mathfrak{A}\, d\mathfrak{B})}{P(d\mathfrak{A})\, P(d\mathfrak{B})}. \tag{1.5}$$

В случае конечных алгебр $\mathfrak{A}$ и $\mathfrak{B}$ количество информации $I(\mathfrak{A}, \mathfrak{B})$ всегда действительно и неотрицательно (см. [1]); отсюда ясно, что в общем случае $I(\mathfrak{A}, \mathfrak{B})$ может быть либо действительно и неотрицательно, либо равно $+\infty$. Так же ясно, что для общего случая сохраняются следующие, хорошо известные для конечного случая свойства величины I (через $[\mathfrak{M}]$ мы будем обозначать наименьшую подалгебру алгебры $\mathfrak{S}$, содержащую множество $\mathfrak{M}$):

а) $I(\mathfrak{A}, \mathfrak{B}) = I(\mathfrak{B}, \mathfrak{A})$;

б) $I(\mathfrak{A}, \mathfrak{B}) = 0$ тогда и только тогда, когда системы событий $\mathfrak{A}$ и $\mathfrak{B}$ независимы;

в) если $[\mathfrak{A}_1 \bigcup \mathfrak{B}_1]$ и $[\mathfrak{A}_2 \bigcup \mathfrak{B}_2]$ — независимые системы событий, то

$$I([\mathfrak{A}_1 \bigcup \mathfrak{A}_2], [\mathfrak{B}_1 \bigcup \mathfrak{B}_2]) = I(\mathfrak{A}_1, \mathfrak{B}_1) + I(\mathfrak{A}_2, \mathfrak{B}_2);$$

г) если $\mathfrak{A}_1 \subseteq \mathfrak{A}$, то $I(\mathfrak{A}_1, \mathfrak{B}) \leqslant I(\mathfrak{A}, \mathfrak{B})$.

Кроме того, укажем еще следующие три свойства информации, позволяющие в ряде случаев использовать предельный переход при вычислении или оценке величины I для конкретных систем событий:

д) если алгебра $\mathfrak{A}_1 \subseteq \mathfrak{A}$ всюду плотна на алгебре $\mathfrak{A}$ в смысле метрики $\rho(A, B) = P(AB' \bigcup A'B)$, то

$$I(\mathfrak{A}_1, \mathfrak{B}) = I(\mathfrak{A}, \mathfrak{B});$$

е) если $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \ldots \subseteq \mathfrak{A}_n \subseteq \ldots$ и $\mathfrak{A} = \bigcup\limits_n \mathfrak{A}_n$, то

$$I(\mathfrak{A}, \mathfrak{B}) = \lim_{n\to\infty} I(\mathfrak{A}_n, \mathfrak{B});$$

ж) если последовательность распределений вероятностей $P^{(n)}$ сходится на $[\mathfrak{A} \bigcup \mathfrak{B}]$ к распределению P, т. е. если $\lim\limits_{n\to\infty} P^{(n)}(C) = P(C)$ для $C \in [\mathfrak{A} \bigcup \mathfrak{B}]$, то

$$I(\mathfrak{A}, \mathfrak{B}) \leqslant \varliminf_{n\to\infty} I^{(n)}(\mathfrak{A}, \mathfrak{B}),$$

где $I(\mathfrak{A}, \mathfrak{B})$ — количество информации при распределении P, а $I^{(n)}(\mathfrak{A}, \mathfrak{B})$ — при распределении $P^{(n)}$.

Все эти свойства непосредственно выводятся из определения (1.4) количества информации и являются почти очевидными.

До сих пор мы говорили лишь о понятии количества информации, содержащейся в результатах одного испытания относительно результатов другого; в приложениях, однако, чаще приходится иметь дело с понятием количества информации, содержащейся в одном случайном объекте (например, случайной величине, или случайном векторе, или случайной функции) относительно другого такого объекта. Для того чтобы перейти к этому новому понятию, будем считать, что основная булева алгебра $\mathfrak{S}$ является σ-алгеброй (см. [2]), и рассмотрим наряду с этим некоторое «измеримое пространство» X, т. е. совокупность множества X и σ-алгебры S_X (с единицей X) некоторых его подмножеств, называемых «измеримыми» (в некоторое отклонение от терминологии [2], где S_X может быть σ-кольцом). При этом ξ будет называться *случайным элемен-*

том пространства X (иначе — *случайной величиной* или *случайным объектом* из X), если каждому подмножеству $A \in S_X$ соответствует событие $B \in \mathfrak{S}$ («событие, заключающееся в том, что $\xi \in A$»); иначе говоря, случайным элементом ξ называется гомоморфное отображение

$$\xi^* (A) = B \tag{1.6}$$

σ-алгебры S_X в булеву алгебру $\mathfrak{S}$. При этом вся алгебра S_X переходит в некоторую подалгебру алгебры $\mathfrak{S}$, которую мы обозначим через

$$\mathfrak{S}_\xi = \xi^* (S_X), \tag{1.7}$$

и количество информации, заключающееся в случайном объекте ξ относительно случайного объекта η, естественно определить как

$$I (\xi, \eta) = I (\mathfrak{S}_\xi, \mathfrak{S}_\eta). \tag{1.8}$$

Формула

$$P_\xi (A) = P (\xi^* (A)) \tag{1.9}$$

определяет меру в пространстве X, которая называется *распределением* случайного объекта ξ; в соответствии с наглядным смыслом события $\xi^* (A)$ можно писать

$$P_\xi (A) = \mathsf{P} \{\xi \in A\} \tag{1.9'}$$

(где правая часть обозначает «вероятность события $\xi \in A$»). Аналогично случайному объекту η из Y соответствует мера

$$P_\eta (B) = \mathsf{P} \{\eta \in B\} \tag{1.10}$$

в пространстве Y. Если теперь обычным образом (см. [2]) определить по заданным X, Y измеримое пространство $X \times Y$, то условие

$$(\xi, \eta)^* (A \times B) = \xi^* (A)\, \eta^* (B) \tag{1.11}$$

однозначно определит гомоморфизм $(\xi, \eta)^*$ булевой σ-алгебры $S_{X \times Y}$ в $\mathfrak{S}$, т. е. новый случайный объект (ξ, η) из $X \times Y$,— пару (ξ, η) случайных объектов ξ и η. Этому новому случайному объекту отвечает распределение

$$P_{\xi, \eta} (C) = \mathsf{P} \{(\xi, \eta) \in C\}, \tag{1.12}$$

являющееся мерой на $X \times Y$; кроме того, на $X \times Y$ мы можем определить еще одну меру $P_\xi \times P_\eta$ при помощи условия

$$P_\xi \times P_\eta (A \times B) = P_\xi (A)\, P_\eta (B), \qquad A \in S_X, \quad B \in S_Y. \tag{1.13}$$

Информация $I (\xi, \eta)$ может быть выражена через меры $P_{\xi,\eta}$ и

$P_\xi \times P_\eta$ при помощи формулы

$$I(\xi, \eta) = \int_X \int_Y P_{\xi, \eta}(dx\,dy) \log \frac{P_{\xi, \eta}(dx\,dy)}{P_\xi(dx)\, P_\eta(dy)} \tag{1.14}$$

(ср. (1.5)), где интеграл понимается в общем смысле, указанном в работе А. Н. Колмогорова [3] (см. также [4]).

Вместо формулы (1.14) иногда удобно пользоваться другой формой записи $I(\xi, \eta)$ в виде интеграла, в которой интеграл уже можно понимать в обычном лебеговом смысле. А именно, применяя к мерам $P_{\xi,\eta}$ и $P_\xi \times P_\eta$ на $X \times Y$ теорему Радона—Никодима, мы можем записать

$$P_{\xi, \eta}(C) = \iint_C a(x, y)\, P_\xi(dx)\, P_\eta(dy) + S(C), \tag{1.15}$$

где мера $S(C)$ сингулярна относительно $P_\xi \times P_\eta$; в частном случае, когда $S(X \times Y) = 0$, функция $a(x, y)$ совпадает с производной Радона—Стилтьеса меры $P_{\xi,\eta}$ по мере $P_\xi \times P_\eta$

$$a(x, y) = \frac{P_{\xi, \eta}(dx\,dy)}{P_\xi(dx)\, P_\eta(dy)}. \tag{1.16}$$

При этом имеет место следующая

Т е о р е м а 1. *Количество информации $I(\xi, \eta)$ может быть конечным лишь в случае, когда $S(X \times Y) = 0$. В этом случае*

$$I(\xi, \eta) = \int_X \int_Y a(x, y) \log a(x, y)\, P_\xi(dx)\, P_\eta(dy) =$$
$$= \int_X \int_Y \log a(x, y)\, P_{\xi, \eta}(dx\,dy). \tag{1.17}$$

Заметим еще, что формулу (1.17) можно переписать в виде

$$I(\xi, \eta) = \mathsf{M} \log a(\xi, \eta), \tag{1.18}$$

где в правой части стоит математическое ожидание новой вещественной случайной величины $\log a(\xi, \eta)$ — функции от пары (ξ, η) [2].

Если распределения P_ξ, P_η и $P_{\xi,\eta}$ выражаются через плотности

$$P_\xi(A) = \int_A p_\xi(x)\, dx, \qquad P_\eta(B) = \int_B p_\eta(y)\, dy,$$
$$P_{\xi, \eta}(C) = \iint_C p_{\xi, \eta}(x, y)\, dx\, dy, \tag{1.19}$$

---

[2] Функция $\eta = f(\xi)$ от случайного элемента ξ, где $y = f(x)$ — измеримое по Борелю отображение пространства X в другое пространство Y, естественно определяется как случайный элемент Y, задаваемый отображением

$\eta^*(A) = \xi^* f^{-1}(A)$.

где dx и dy символизируют интегрирование по заданным в X и Y мерам, то формулы (1.17) и (1.14) переходят в хорошо известное в прикладной литературе выражение

$$I(\xi, \eta) = \int\limits_X \int\limits_Y p_{\xi, \eta}(x, y) \log \frac{p_{\xi, \eta}(x, y)}{p_\xi(x)\, p_\eta(y)}\, dx\, dy. \tag{1.20}$$

Любопытно отметить, что если ввести в рассмотрение «дифференциальную энтропию»

$$h(\xi) = -\int p_\xi(x) \log p_\xi(x)\, dx \tag{1.21}$$

и «условную дифференциальную энтропию» $h(\xi \mid \eta)$, определяемую условным распределением $P_{\xi|\eta}$ случайного элемента ξ при фиксированном η [3], то формулу (1.20) можно будет переписать в виде, аналогичном (1.3):

$$I(\xi, \eta) = h(\xi) - \mathsf{M}h(\xi \mid \eta). \tag{1.21'}$$

Однако в отличие от (1.3) $h(\xi)$ здесь уже не совпадает с $I(\xi, \xi)$ (последняя величина равна бесконечности).

Из свойств а) — г) величины $I(\mathfrak{A}, \mathfrak{B})$ легко вытекают следующие свойства количества информации $I(\xi, \eta)$:

а) $I(\xi, \eta) = I(\eta, \xi)$;

б) $I(\xi, \eta) \geqslant 0$; $I(\xi, \eta) = 0$ тогда и только тогда, когда ξ и η независимы;

в) если пары (ξ_1, η_1) и (ξ_2, η_2) независимы, то

$I((\xi_1, \xi_2), (\eta_1, \eta_2)) = I(\xi_1, \eta_1) + I(\xi_2, \eta_2)$;

г) если ξ_1 есть функция от ξ (в указанном выше смысле), то

$I(\xi_1, \eta) \leqslant I(\xi, \eta)$.

В частности, всегда

$I((\xi, \eta), \zeta) \geqslant I(\eta, \zeta)$.

Если по аналогии с (1.2) определить энтропию случайной величины ξ как

$I(\xi, \xi) = H(\xi)$,

то $H(\xi)$ при весьма общих условиях (например, всегда, когда распределение P_ξ есть непрерывное распределение, задаваемое плотностью p_ξ) оказывается равным бесконечности; поэтому практически полезным понятие энтропии будет лишь для сравнительно узкого класса случайных величин (в первую очередь для дискретных

---

[3] $P_{\xi|\eta}(A)$ есть действительная случайная функция $f(\eta)$ от η, удовлетворяющая требованию

$$P_{\xi\eta}(A \times B) = \int\limits_B f(y)\, P_\eta(dy), \quad B \in S_Y.$$

величин, принимающих конечное число значений). Из свойств в) — г) количества информации $I(\xi, \eta)$ немедленно следует, что

$$H(\xi, \eta) \geqslant H(\xi), \qquad H(\xi, \eta) \geqslant H(\eta) \tag{1.22}$$

и что если ξ и η независимы, то

$$H(\xi, \eta) = H(\xi) + H(\eta). \tag{1.23}$$

Хорошо известно (см., например, [1]), что для энтропии $H(\xi, \eta)$, кроме оценки (1.22) снизу, существует еще простая оценка сверху: всегда

$$H(\xi, \eta) \leqslant H(\xi) + H(\eta). \tag{1.24}$$

Следует подчеркнуть, что для количества информации в паре (ξ, η) относительно величины ζ подобной оценки не существует: на элементарных примерах легко показать, что из

$$I(\xi, \zeta) = 0, \qquad I(\eta, \zeta) = 0$$

еще не вытекает равенство

$$I((\xi, \eta), \zeta) = 0$$

(противоречащим примером будет любой пример трех случайных величин, каждые две из которых попарно независимы, но ни одна не является независимой от двух остальных).

Укажем еще два свойства количества информации $I(\xi, \eta)$, вытекающие из свойств д) — ж) величины $I(\mathfrak{A}, \mathfrak{B})$:

д) если $\xi = (\xi_1, \xi_2, \ldots, \xi_n, \ldots)$ (это есть случайный элемент пространства $X = \mathop{\mathsf{X}}\limits_{n=1}^{\infty} X_n$), то

$$I(\xi, \eta) = \lim_{n\to\infty} I((\xi_1, \ldots, \xi_n), \eta);$$

е) если X и Y — полные метрические пространства, в которых за S_X и S_Y приняты алгебры борелевских множеств, и если распределения $P^{(n)}_{\xi, \eta}$ слабо сходятся к распределению $P_{\xi, \eta}$, то для соответствующих количеств информации имеет место неравенство

$$I(\xi, \eta) \leqslant \varliminf_{n\to\infty} I^{(n)}(\xi, \eta).$$

В заключение остановимся еще специально на случае, когда X и Y — пространства всех вещественных функций на некоторых множествах A и B. Случайные элементы ξ и η здесь будут случайными функциями $\xi(a)$, $\eta(b)$ на A и B, т. е. некоторыми совокупностями вещественных случайных величин, зависящих от параметра $a \in A$ или $b \in B$. $I\{\xi, \eta\}$ будет иметь смысл количества информации, содержащейся в одной совокупности случайных величин относительно другой такой совокупности. Пространства X и Y являются здесь степенями R^A и R^B пространства R вещественных чисел (естественно, с системой его борелевских множеств в качестве S_R). Известно, что алгебра S_X порождается всевозможными цилиндрическими множествами («квазиинтервалами») пространства X, задаваемыми конечными системами неравенств вида

$$A_i \leqslant x(a_i) < B_i, \qquad a_i \in A, \quad i = 1, 2, \ldots, n. \tag{1.25}$$

Поэтому определение количества информации $I\{\xi, \eta\}$ для произвольных систем случайных величин сводится к определению количества информации $I\{\xi(a_1, \ldots, a_n), \eta(b_1, \ldots, b_m)\}$ для конечномерных случайных векторов

$$\xi(a_1, \ldots, a_n) = \{\xi(a_1), \ldots, \xi(a_n)\} \text{ и } \eta(b_1, \ldots, b_m) = \\ = \{\eta(b_1), \ldots, \eta(b_m)\},$$

а именно:

$$I\{\xi(a), \eta(b)\} = \sup I\{\xi(a_1, \ldots, a_n), \eta(b_1, \ldots, b_m)\} \tag{1.26}$$

(верхняя грань по всем целым $n, m, a_i \in A$, $i = 1, 2, \ldots, n$, и $b_j \in B$, $j = 1, 2, \ldots, m$). Таким образом, при изучении количества информации $I\{\xi(a), \eta(b)\}$ можно обойтись рассмотрением лишь конечномерных распределений вероятностей и не привлекать распределений в функциональном пространстве.

Специальным случаем «совокупностей случайных величин» являются так называемые обобщенные случайные процессы, введенные недавно К. Ито [5] и И. М. Гельфандом [6]. В этом случае множеством A является пространство Φ бесконечно дифференцируемых функций $\varphi(t)$ на множестве T вещественных чисел [4], каждая из которых обращается в нуль вне некоторого конечного интервала (пространство Φ основных функций по терминологии Л. Шварца [7]). Обобщенный случайный процесс $\xi(\varphi)$ задается совокупностью многомерных распределений вероятностей для всевозможных конечных наборов случайных величин $\xi(\varphi_i)$, $\varphi_i \in \Phi$, $i = 1, \ldots, n$, т. е. совокупностью вероятностей всевозможных «квазиинтервалов»

$$A_i \leqslant x(\varphi_i) < B_i, \qquad \varphi_i \in \Phi, \quad i = 1, \ldots, n. \tag{1.25$'$}$$

Если $\xi(\varphi)$ и $\eta(\psi)$ — два таких обобщенных процесса, отвечающих пространствам основных функций Φ и Ψ (различающимся, например, выбором множества T), то количество информации $I\{\xi(\varphi), \eta(\psi)\}$ согласно (1.8) и (1.4) определяется как

$$I\{\xi(\varphi), \eta(\psi)\} = \sup I\{\xi(\varphi_1, \ldots, \varphi_n), \eta(\psi_1, \ldots, \psi_m)\}, \tag{1.26$'$}$$

[4] Разумеется, вместо обобщенных случайных процессов (случайных функций вещественного аргумента t) можно также рассматривать «обобщенные случайные функции на произвольном дифференцируемом многообразии T»; однако такое обобщение не повлечет за собой почти никаких изменений и на нем можно не останавливаться.

где

$$\xi(\varphi_1, \ldots, \varphi_n) = \{\xi(\varphi_1), \ldots, \xi(\varphi_n)\},$$

$$\eta(\psi_1, \ldots, \psi_m) = \{\eta(\psi_1), \ldots, \eta(\psi_m)\}$$

и верхняя грань берется по всем целым n, m и $\varphi_i \in \Phi$, $i = 1, \ldots$ $\ldots, n$, $\psi_j \in \Psi$, $j = 1, \ldots, m$. Определение (1.26′) применимо также и к любому обыкновенному стохастически непрерывному и стохастически интегрируемому) случайному процессу $\xi(t)$, если положить

$$\xi(\varphi) = \int\limits_{-\infty}^{\infty} \xi(t)\varphi(t)\,dt; \tag{1.27}$$

можно доказать, что в этом случае определение (1.26′) приводит к тому же результату, как и непосредственное определение по формуле (1.26). Отметим, что часто даже для обыкновенных (не обобщенных) процессов определение (1.26′) является более удобным, чем (1.26).

§ 2
АБСТРАКТНОЕ ИЗЛОЖЕНИЕ ОСНОВ ТЕОРИИ ШЕННОНА

Шеннон рассматривает передачу сообщений по схеме

$$\xi \to \eta \to \eta' \to \xi', \tag{2.1}$$

где «передающее устройство»

$$\eta \to \eta'$$

характеризуется условным распределением

$$P_{\eta'\mid\eta}(B' \mid y) = \mathsf{P}\{\eta' \in B' \mid \eta = y\} \tag{2.2}$$

«сигнала на выходе» η' при заданном «сигнале на входе» η и некоторыми ограничениями

$$P_\eta \in V \tag{2.3}$$

распределения P_η входного сигнала (например, ограничением «мощности» сигнала η). Операции «кодирования»

$$\xi \to \eta$$

и «декодирования»

$$\eta' \to \xi'$$

характеризуются условными распределениями

$$P_{\eta\mid\xi}(B \mid x) = \mathsf{P}\{\eta \in B \mid \xi = x\}, \tag{2.4}$$

$$P_{\xi'|\eta'}(A'|y') = \mathsf{P}\{\xi' \in A' | \eta' = y'\}. \tag{2.5}$$

Основная задача Шеннона заключается в следующем. Заданы пространства X, X', Y, Y' возможных значений «сообщений на входе» ξ, «сообщений на выходе» ξ', «сигналов на входе» η и «сигналов на выходе» η'; заданы также характеристики передающего устройства, т. е. условные распределения $P_{\eta'|\eta}$ и класс V допустимых распределений P_η входного сигнала, и, наконец, распределение

$$P_\xi(A) = \mathsf{P}\{\xi \in A\} \tag{2.6}$$

сообщения на входе и «требования к точности передачи», записываемые в виде

$$P_{\xi\xi'} \in W, \tag{2.7}$$

где W — некоторый класс совместных распределений

$$P_{\xi\xi'}(C) = \mathsf{P}\{(\xi, \xi') \in C\} \tag{2.8}$$

сообщения на входе и сообщения на выходе. Спрашивается: возможно ли, и в случае возможности — каким способом, задать правила кодирования и декодирования (т. е. условные распределения $P_{\eta|\xi}$ и $P_{\xi'|\eta'}$) таким образом, что, вычисляя распределение $P_{\xi\xi'}$ по распределениям P_ξ, $P_{\eta|\xi}$, $P_{\eta'|\eta}$, $P_{\xi'|\eta'}$ в предположении, что последовательность

$$\xi, \eta, \eta', \xi'$$

марковская, получим

$$P_{\xi\xi'} \in W?$$

Определим вместе с Шенноном *пропускную способность* передающего устройства формулой

$$C = \sup_{P_\eta \in V} I(\eta, \eta') \tag{2.9}$$

и введем величину

$$H_W(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi'), \tag{2.10}$$

которую при расчете на единицу времени Шеннон называет *скоростью создания сообщений*. Тогда из важного свойства количества информации: $I((\xi, \eta), \zeta) = I(\eta, \zeta)$ в том и только том случае, когда последовательность ξ, η, ζ марковская, сразу вытекает необходимое условие возможности передачи

$$H_W(\xi) \leqslant C. \tag{2.11}$$

Несравненно глубже идея Шеннона о том, что в применении к длительной работе «каналов связи» условие (2.11) является в некотором смысле и при некоторых весьма широких условиях «почти достаточным». С математической точки зрения здесь дело идет о доказательстве предельных теорем следующего типа. Предполагается, что пространства X, X', Y, Y', распределения P_ξ и $P_{\eta'|\eta}$, классы V и W, а следовательно, и величины C и $H_W(\xi)$ зависят от параметра T (играющего в применениях роль длительности работы передающего устройства). Требуется установить при некоторых предположениях достаточно общего характера, что условие

$$\varliminf_{T\to\infty} \frac{C^T}{H_W^T(\xi)} > 1 \tag{2.12}$$

достаточно для возможности передачи, удовлетворяющей поставленным выше условиям, при достаточно больших T. Естественно, что в такой постановке задача несколько расплывчата (подобно, например, общей задаче изучения возможных предельных распределений для сумм большого числа «малых» слагаемых); более четкую формулировку можно получить, предположив, например, что ξ и ξ' являются стационарными случайными процессами. Однако мы умышленно отказались здесь от обращения к терминологии теории стационарных случайных процессов, так как в докладе на настоящем съезде молодого румынского математика Розенблат-Рот Милу [8] (см. также [9]) было показано, что можно получить в намеченном направлении не лишенные интереса результаты и без предположений стационарности.

Выводу предельных теорем указанного типа посвящено много замечательных работ; мы сошлемся на недавние работы А. Я. Хинчина [16] и М. С. Пинскера [17], содержащие также ссылки на более ранние исследования. Нам представляется, что в этом направлении остается еще многое сделать. Подчеркнем, что именно этого рода результаты предназначены для того, чтобы обосновать распространенное убеждение, что выражение $I(\xi, \eta)$ является не одним из возможных способов измерения «количества информации», а мерой количества информации, имеющей действительно принципиальное преимущество перед другими. Так как по первоначальной своей природе «информация» не есть скалярная величина, то аксиоматические исследования, позволяющие однозначно охарактеризовать $I(\xi, \eta)$ (или энтропию $H(\xi)$) при помощи простых формальных свойств, имеют в этом отношении, по нашему мнению, меньшее значение. Положение представляется здесь аналогичным тому, как из всех предложенных еще Гауссом способов обоснования нормального закона распределения ошибок мы склонны сейчас придавать наибольшее значение способу, исходящему из предельных теорем для сумм большого числа малых слагаемых. Другие способы (например, способ, основанный на принципе арифметического среднего) объясняют лишь, почему никакой другой закон распределения ошибок не мог бы быть столь же приятным и удобным, как нормальный, но не отвечают на вопрос, почему нормальный закон действительно часто встречается в реальных задачах. Точно так же красивые формальные свойства выражений $H(\xi)$ и $I(\xi, \eta)$ еще не могут объяснить, почему во многих задачах они достаточны для полного (хотя бы с асимптотической точки зрения) решения многих задач.

§ 3
ВЫЧИСЛЕНИЕ КОЛИЧЕСТВА ИНФОРМАЦИИ ДЛЯ СЛУЧАЯ ГАУССОВСКИХ РАСПРЕДЕЛЕНИЙ

В настоящем параграфе мы займемся вопросом о вычислении количества информации $I(\xi, \eta)$, причем ограничимся лишь тем специальным (но весьма важным) случаем, когда распределения P_ξ, P_η и $P_{\xi\eta}$ являются гауссовскими.

Начнем со случая, когда ξ и η — случайные векторы,

$$\begin{aligned} \xi &= (\xi_1, \ldots, \xi_m), \\ \eta &= (\eta_1, \ldots, \eta_n) \equiv (\xi_{m+1}, \ldots, \xi_{m+n}), \end{aligned} \tag{3.1}$$

распределенные нормально со вторыми центральными моментами

$$s_{ij} = \mathsf{M}\,[(\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j)]. \tag{3.2}$$

Если детерминант

$$\det C = \det \| s_{ij} \|_{1 \leqslant i, j \leqslant m+n} \tag{3.3}$$

отличен от нуля, то вычисление количества информации $I(\xi, \eta)$ в силу формулы (1.20) сводится к вычислению элементарного интеграла и приводит к результату

$$I(\xi, \eta) = \frac{1}{2} \log \frac{\det A \cdot \det B}{\det C}, \tag{3.4}$$

где

$$A = \| s_{ij} \|_{1 \leqslant i, j \leqslant m}, \qquad B = \| s_{ij} \|_{m+1 \leqslant i, j \leqslant m+n}. \tag{3.5}$$

При конкретных вычислениях иногда удобнее вместо формулы (3.4) пользоваться формулой

$$\begin{aligned} I(\xi, \eta) &= -{}^1/_2 \log \det (E - DB^{-1}D'A^{-1}) = \\ &= -{}^1/_2 \log \det (E - D'A^{-1}DB^{-1}), \end{aligned} \tag{3.6}$$

где E — единичная матрица (этим обозначением мы будем пользоваться и ниже) и

$$D = \| s_{ij} \|_{\substack{1 \leqslant i \leqslant m, \\ m < j \leqslant m+n}}, \qquad D' = \| s_{ij} \|_{\substack{m < i \leqslant m+n. \\ 1 \leqslant j \leqslant m}} \tag{3.7}$$

В простейшем случае одномерных случайных величин ξ и η формула (3.4) (или (3.7)) дает

$$I(\xi, \eta) = -{}^1/_2 \log [1 - r^2(\xi, \eta)], \tag{3.8}$$

где $r^2(\xi, \eta)$ — коэффициент корреляции между ξ и η; последняя формула, очевидно, верна и без предположения о невырожденности двумерного распределения вероятностей для ξ и η (невырожденности матрицы C). Формула (3.8) может быть обобщена и на многомерный случай (также без ограничения случаем невырожденности матрицы C). В самом деле, как известно [11, 12], при помощи надлежащего линейного преобразования координат в пространствах X и Y можно добиться того, чтобы все вторые моменты s_{ij}, кроме тех, для которых $j = i$ или $j = m + i$, обратились в нуль. При таком выборе координат в силу формулы (3.8) и свойства в) количества информации получаем [5]

$$I(\xi, \eta) = -\frac{1}{2} \sum_k \log [1 - r^2(\xi_k, \eta_k)], \tag{3.9}$$

где суммирование берется по тем $k \leqslant \min(m, n)$, для которых коэффициент корреляции между ξ_k и $\eta_k = \xi_{m+k}$ отличен от нуля.

Отметим еще, что полученное выражение для $I(\xi, \eta)$ имеет простой геометрический смысл. Будем рассматривать действительные случайные величины (точнее, классы эквивалентных между собой случайных величин, равных друг другу с вероятностью 1) с конечной дисперсией как векторы гильбертова пространства $\mathfrak{H}$ со скалярным произведением $(\xi_1, \xi_2) = \mathsf{M}[(\xi_1 - \mathsf{M}\xi_1)(\xi_2 - \mathsf{M}\xi_2)]$. Тогда случайному вектору $\xi = (\xi_1, \ldots, \xi_m)$ будет отвечать конечномерное линейное подпространство H_1 пространства $\mathfrak{H}$, натянутое на элементы $(\xi_1, \ldots, \xi_m)$ этого пространства; аналогично вектору $\eta = (\eta_1, \ldots, \eta_n)$ будет отвечать другое конечномерное линейное подпространство H_2. Рассмотрим теперь линейное преобразование B_1 в пространстве H_1, при котором каждый вектор из H_1 проектируется сперва на H_2, а затем полученный при этом вектор проектируется обратно на H_1; в подпространстве H_2 аналогичную роль будет играть линейное преобразование B_2, состоящее в проектировании каждого вектора из H_2 на H_1, а затем обратно на H_2. Преобразования B_1 и B_2 можно выразить через операторы P_1 и P_2 проектирования на подпространства H_1 и H_2 при помощи формул

$$B_1 = P_1 P_2 \quad \text{в} \quad H_1; \qquad B_2 = P_2 P_1 \quad \text{в} \quad H_2; \tag{3.10}$$

можно также доопределить B_1 и B_2 до операторов во всем пространстве $\mathfrak{H}$, предполагая их нулевыми операторами в ортогональных дополнениях подпространств H_1 и соответственно H_2 и

---

[5] Так как $\eta_k = \xi_{m+k}$.

писать

$$B_1 = P_1P_2P_1, \qquad B_2 = P_2P_1P_2. \tag{3.11}$$

Ясно, что B_1 и B_2 будут неотрицательными самосопряженными операторами, не превосходящими по норме единицы, так что все их собственные значения вещественны и заключены между 0 и 1. При этом оказывается, что квадраты коэффициентов корреляции $r^2(\xi_k, \eta_k)$ в формуле (3.9) совпадают с собственными значениями операторов B_1 и B_2 (имеющих одинаковые собственные значения), так что эту формулу можно переписать в виде

$$I(\xi, \eta) = -{}^1/_2 \log \det (E - B_1) = -{}^1/_2 \log \det (E - B_2). \tag{3.12}$$

Выражение (3.12) для информации $I(\xi, \eta)$ показывает, что эта величина зависит только от подпространств H_1 и H_2 (а не от выбора базиса в них) и что она является геометрическим инвариантом указанной пары подпространств. Но полной системой геометрических инвариантов пары конечномерных линейных подпространств является система углов между этими подпространствами (точнее говоря, «стационарных углов» между ними; см., например, [13]). Отсюда вытекает, что $I(\xi, \eta)$ должно выражаться через углы между H_1 и H_2. И действительно, можно показать, что два линейных подпространства H_1 и H_2 (одно — не более чем m-мерное, второе — не более чем n-мерное) образуют между собой $k \leqslant \min(m, n)$ различных стационарных углов $\alpha_1, \ldots, \alpha_k$, причем собственные значения операторов (3.11) равны $\cos^2 \alpha_1, \ldots, \cos^2 \alpha_k$, так что

$$I(\xi, \eta) = -\log |\sin \alpha_1 \ldots \sin \alpha_k|. \tag{3.13}$$

Формулу (3.12) нетрудно обобщить на случай, когда $\xi = \xi(a)$ и $\eta = \eta(b)$ — две произвольные гауссовские совокупности вещественных случайных величин, т. е. такие две совокупности случайных величин, что конечномерные распределения вероятностей для любого конечного числа величин, принадлежащих той или другой совокупности, все являются распределениями Гаусса. Совокупности $\xi(a)$ также можно сопоставить линейное подпространство H_1 пространства $\mathfrak{H}$ (как правило, бесконечномерное), натянутое на элементы $\xi(a)$ этого пространства; аналогично совокупности $\eta(b)$ сопоставляется линейное подпространство H_2, натянутое на элементы $\eta(b) \in \mathfrak{H}$. При этом имеет место следующая

Теорема 2. *Пусть $\xi = \xi(a)$ и $\eta = \eta(b)$ — две гауссовские совокупности вещественных случайных величин, H_1 и H_2 — отвечающие этим совокупностям линейные подпространства пространства $\mathfrak{H}$, P_1 и P_2 — операторы проектирования в $\mathfrak{H}$ на H_1 и соответственно H_2 и B_1, B_2 — операторы* (3.11). *В таком случае для того, чтобы количество информации $I\{\xi, \eta\}$ было конеч-*

ной величиной, необходимо и достаточно, чтобы по крайней мере один из операторов B_1, B_2 *был вполне непрерывным оператором с конечным следом; при этом условии и второй из операторов* B_1, B_2 *будет обладать тем же свойством и количество информации* $I\{\xi, \eta\}$ *будет даваться формулой* (3.12), *где* E — *единичный оператор.*

Особенно просто применяется теорема 2 к тому случаю, когда $\xi = \xi(\varphi)$ — случайный процесс (вообще говоря, обобщенный) на некотором интервале (может быть, бесконечном) T, а η — скалярная случайная величина. Проекция $\tilde{\eta} = P_1\eta$ величины η на подпространство H_1 здесь будет совпадать с наилучшим приближением к η, линейно зависящим от $\xi(\varphi)$, т. е. будет являться решением задачи о линейной фильтрации процесса $\xi(\varphi)$, которой посвящена обширная специальная литература (см., например, обзор [14]). Если обозначить через $\sigma = [\mathsf{M}\,|\eta - \tilde{\eta}|^2]^{1/2}$ среднеквадратичную ошибку фильтрации, а через $\sigma_1 = \sigma/[\mathsf{M}\eta^2]^{1/2}$ — относительную среднеквадратичную ошибку, то, как легко видеть, имеет место формула

$$I\{\xi(\varphi), \eta\} = -\log \sigma_1. \tag{3.14}$$

Более сложен случай, когда $\xi = \xi(\varphi)$ и $\eta = \eta(\psi)$ оба являются случайными процессами; мы здесь остановимся лишь на задаче о вычислении $I\{\xi(\varphi), \eta(\varphi)\}$ для стационарных (и стационарно связанных) процессов $\xi(\varphi)$ и $\eta(\varphi)$, заданных на одном и том же конечном интервале $-T/2 \leqslant t \leqslant T/2$ [6]. Как известно (см. [5, 6]), гауссовский стационарный обобщенный процесс $\xi(\varphi)$ однозначно характеризуется обобщенной корреляционной функцией $b_{\xi\xi}(\varphi)$ — линейным функционалом в пространстве Φ основных функций $\varphi(t)$ (в случае обыкновенного процесса $\xi = \xi(t)$

$$b_{\xi\xi}(\varphi) = \int_{-\infty}^{\infty} b_{\xi\xi}(\tau)\,\varphi(\tau)\,d\tau,$$

где $b_{\xi\xi}(\tau) = \mathsf{M}\xi(t+\tau)\,\xi(t)$ — корреляционная функция процесса $\xi(t)$). Для задания пары процессов $\xi(\varphi)$ и $\eta(\varphi)$ надо задать обобщенные корреляционные функции $b_{\xi\xi}(\varphi)$ и $b_{\eta\eta}(\varphi)$, а также обобщенную взаимную корреляционную функцию $b_{\xi\eta}(\varphi) \equiv \equiv b_{\eta\xi}(\varphi^*)$, где $\varphi^*(t) = \varphi(-t)$.

Рассмотрим теперь в пространстве Φ_T функций $\varphi(t)$ на отрезке $-T/2 \leqslant t \leqslant T/2$ следующие операторы:

$$\begin{aligned} A\varphi &= b_{\xi\xi}(\tau_t\varphi), & B\varphi &= b_{\eta\eta}(\tau_t\varphi), \\ D\varphi &= b_{\xi\eta}(\tau_t\varphi), & D'\varphi &= b_{\eta\xi}(\tau_t\varphi) = b_{\xi\eta}(\tau_{-t}\varphi), \end{aligned} \tag{3.15}$$

[6] Говоря о процессе $\xi(\varphi)$, заданном на конечном отрезке $-T/2 \leqslant \leqslant t \leqslant T/2$, мы подразумеваем, что из всех случайных величин $\xi(\varphi)$, $\varphi \in \Phi$, составляющих обобщенный процесс, нам заданы лишь те величины, которые отвечают функциям $\varphi(t)$, тождественно равным нулю вне рассматриваемого отрезка.

где τ_t — оператор сдвига: $\tau_t \varphi(s) = \varphi(s + t)$ (операторы (3.15) можно рассматривать как операторы в гильбертовом пространстве $L^2(-T/2, T/2)$, заданные на всюду плотном множестве функций $\varphi(t)$). Таким образом, в применении к рассматриваемому нами частному случаю теорема 2 принимает следующий вид:

Теорема 2′. *Для того чтобы количество информации $I\{\xi_T, \eta_T\}$, заключающееся в заданном на отрезке длины T стационарном гауссовском случайном процессе $\xi(\varphi)$ относительно другого подобного же процесса $\eta(\varphi)$ на том же отрезке, стационарно и гауссовски связанного с первым, было конечной величиной, необходимо и достаточно, чтобы хоть один из операторов*

$$B_1 = DB^{-1}D'A^{-1}, \qquad B_2 = D'A^{-1}DB^{-1}, \tag{3.16}$$

где A, B, D и D' определяются из (3.15), *был вполне непрерывным оператором с конечным следом. При этом условии и второй из этих операторов будет обладать тем же свойством и информация $I\{\xi_T, \eta_T\}$ будет даваться равенством* (3.12).

Теорема 2′ может быть использована, например, для получения явных выражений для количества информации $I\{\xi_T, \eta_T\}$ в том случае, когда преобразования Фурье обобщенных функций $b_{\xi\xi}$, $b_{\eta\eta}$ и $b_{\xi\eta}$ («спектральные плотности») существуют и являются рациональными функциями. В важном частном случае, когда $\eta = \eta(t)$ — обыкновенный стационарный случайный процесс с корреляционной функцией $B(\tau)$, а $\xi = \eta + \zeta$, где $\zeta = \zeta(\varphi)$ — некоррелированный с η «белый шум» (обобщенный стационарный процесс «с независимыми значениями», для которого

$$b_{\zeta\zeta} = 2\pi f\delta, \qquad \delta(\varphi) = \varphi(0), \tag{3.17}$$

где δ — это δ-функция Дирака, а f — постоянная спектральная плотность «белого шума» ζ), формула (3.12) дает

$$I\{\xi_T, \eta_T\} = \frac{1}{2}\log\det\left(E + \frac{1}{2\pi f}B\right), \tag{3.18}$$

где B — оператор Фредгольма с ядром $B(t - t')$:

$$B\varphi = \int_{-T/2}^{T/2} B(t - t')\varphi(t')\,dt'. \tag{3.19}$$

Таким образом, в этом случае вычисление количества информации $I\{\xi_T, \eta_T\}$ сводится к определению детерминанта интегрального оператора

$$E + B/2\pi f.$$

Если оператор $(1/2\pi f)B$ в формуле (3.18) является лишь малой добавкой к единичному оператору E (т. е. если спектральная плотность f шума много больше характерного значения f_0 спектральной плотности процесса $\eta(t)$ или если T много меньше характерного времени T_0 затухания корреляционной функции $B(\tau)$), то детерминант в правой части (3.18) можно приближенно подсчитать при помощи ряда теории возмущений:

$$\log \det \left(E + \frac{1}{2\pi f} B\right) = \operatorname{Sp} \log \left(E + \frac{1}{2\pi f} B\right) =$$
$$= \frac{1}{2\pi f} \operatorname{Sp} B - \frac{1}{2(2\pi f)^2} \operatorname{Sp} B^2 + \frac{1}{3(2\pi f)^3} \operatorname{Sp} B^3 - \ldots, \quad (3.20)$$

где Sp означает «след», так что

$$I\{\xi_T, \eta_T\} = \frac{B(0)}{4\pi f} T - \frac{T}{16\pi^2 f^2} \int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right) B^2(\tau)\, d\tau + \ldots \quad (3.21)$$

Если же корреляционная функция $B(\tau)$ является преобразованием Фурье рациональной спектральной плотности $f(\lambda)$:

$$B(\tau) = \int_{-\infty}^{\infty} e^{i\tau\lambda} f(\lambda)\, d\lambda, \quad (3.22)$$

где

$$f(\lambda) = \frac{|b_0 (i\lambda)^m + \ldots + b_m|^2}{|a_0 (i\lambda)^n + \ldots + a_n|^2} = \frac{|Q(i\lambda)|^2}{|P(i\lambda)|^2}, \quad m < n, \quad (3.23)$$

то для $I\{\xi_T, \eta_T\}$ можно получить явную формулу; в этом случае собственные значения оператора (3.19) совпадают с собственными значениями некоторой задачи Штурма—Лиувилля для уравнений с постоянными коэффициентами, т. е. являются нулями некоторой целой трансцендентной функции, и детерминант в правой части (3.18) может быть выражен через значение этой целой функции в фиксированной точке комплексной плоскости. В частности, для марковского стационарного процесса $\eta(t)$ с

$$B(\tau) = Ce^{-\alpha|\tau|}, \qquad f(\lambda) = \frac{C\alpha}{\pi(\lambda^2 + \alpha^2)} \quad (3.24)$$

получаем

$$I\{\xi_T, \eta_T\} = \frac{1}{2} \log \left\{ \frac{\sqrt{1+k} + (1 + k/2)}{2\sqrt{1+k}} e^{\tau(\sqrt{1+k}-1)} + \right.$$
$$\left. + \frac{\sqrt{1+k} - (1 + k/2)}{2\sqrt{1+k}} e^{-\tau(\sqrt{1+k}+1)} \right\}, \quad (3.25)$$

где

$$k = f(0)/f = C/\pi\alpha f, \qquad \tau = \alpha T. \quad (3.26)$$

При $k \ll 1$ («сильный шум») или $\tau \ll 1$ («малое время наблюдения») формула (3.25) переходит в результат, даваемый рядом теории возмущений (3.21), как это и должно быть. В другом крайнем случае при $\tau \gg 1$ («большое время наблюдений») будем иметь

$$I\{\xi_T, \eta_T\} \approx \frac{\sqrt{1+k}-1}{2}\tau. \tag{3.27}$$

Наконец, при $k \to 0$ («слабый шум») формула (3.25) дает

$$I\{\xi_T, \eta_T\} \approx {}^1/_2 \sqrt{k}\tau. \tag{3.28}$$

Для более сложных, чем (3.24), рациональных спектральных плотностей $f(\lambda)$ явная формула для $I\{\xi_T, \eta_T\}$, как правило, оказывается довольно громоздкой; поэтому мы здесь ограничимся лишь тем, что выпишем некоторые асимптотические результаты, относящиеся к случаю общей рациональной спектральной плотности (3.23). При $T \to 0$ или $f \to \infty$ асимптотическое поведение $I\{\xi_T, \eta_T\}$ во всех случаях дается рядом теории возмущений (3.21), так что на этом случае можно не останавливаться. При $T \to \infty$ главным членом $I\{\xi_T, \eta_T\}$ будет член

$$I\{\xi_T, \eta_T\} \approx \frac{1}{2}\sum_{k=1}^{n}(\Lambda_k - \Lambda_k^0)\,T, \tag{3.29}$$

где Λ_k, $k = 1, \ldots, n$,— расположенные в правой полуплоскости корни уравнения

$$P(\Lambda)\,P(-\Lambda) + f^{-1}Q(\Lambda)\,Q(-\Lambda) = 0, \tag{3.30}$$

а Λ_k^0 — соответствующие корни при $f = \infty$. Наконец, при $f \to 0$ асимптотически

$$I\{\xi_T, \eta_T\} \approx \frac{T}{2}\sum_{k=1}^{n}\Lambda_k, \tag{3.31}$$

где Λ_k имеют тот же самый смысл, что и в (3.29); отсюда нетрудно получить, что при $f \to 0$

$$I\{\xi_T, \eta_T\} \approx cTf^{-1/2(n-m)}, \tag{3.32}$$

где c — постоянная, зависящая от коэффициентов многочленов $P(\Lambda)$ и $Q(\Lambda)$.

Отметим, что показатель $-[2(n-m)]^{-1}$ в формуле (3.32) характеризует «степень гладкости» корреляционной функции $B(\tau)$ (а следовательно, и самого случайного процесса $\eta(t)$); в случае спектральной плотности (3.23) корреляционная функция $B(\tau)$ имеет $2(n-m)-2$ непрерывных производных, а $(2n-2m-1)$-я производная этой функции делает скачок в точке $\tau = 0$. Можно показать, что то обстоятельство, что асимптотическое поведение $I\{\xi_T, \eta_T\}$, где $\xi = \eta + \zeta$, ζ — «белый шум» со спектральной плотностью f, при $f \to 0$ определяется степенью гладкости функции $B(\tau)$, не связано со специальной формой (3.23) спектральной плотности, а имеет место при довольно общих условиях, так что для сравнительно широкого класса корреляционных функций $B(\tau)$, имеющих $(2n-2)$ непрерывных производных и $(2n-1)$-ю производную, разрывную при $\tau = 0$,

$$I\{\xi_T, \eta_T\} \approx cTf^{-1/2n} \text{ при } f \to 0. \tag{3.33}$$

§ 4
ВЫЧИСЛЕНИЕ И ОЦЕНКА ε-ЭНТРОПИИ (СКОРОСТИ СОЗДАНИЯ СООБЩЕНИЙ) В НЕКОТОРЫХ ЧАСТНЫХ СЛУЧАЯХ

Перейдем к рассмотрению «скорости создания сообщений»

$$H_W(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi').$$

Если условие $P_{\xi\xi'} \in W$ выбрать в виде обязательности точного совпадения ξ и ξ':

$$\mathsf{P}\{\xi = \xi'\} = 1, \tag{4.1}$$

то

$$H_W(\xi) = H(\xi), \tag{4.2}$$

т. е. скорость создания сообщений обращается здесь в обычную энтропию случайного элемента ξ. Как мы уже указывали, для непрерывных P_ξ последняя величина всегда равна ∞: непрерывно распределенная величина ξ содержит о себе самой бесконечную информацию. Однако поскольку непрерывные сообщения всегда наблюдаются с ограниченной точностью, то при передаче таких сообщений фактически передается лишь информация, которая может быть сделана сколь угодно близкой к, как правило, конечной величине $H_W(\xi)$. Поэтому в теории передачи непрерывных сообщений основную роль должна играть именно величина $H_W(\xi)$ (а не $H(\xi)$), которую в силу ее аналогии с энтропией можно назвать «энтропией случайного объекта ξ при точности воспроизведения W».

Условие $P_{\xi\xi'} \in W$ должно обеспечивать близость в определенном смысле элемента ξ' к ξ. Если предположить, что X — метрическое пространство, а X' совпадает с X (т. е. исследовать вопрос о способах приближенной передачи сообщения о положении точки пространства X при помощи указания точки того же пространства), то кажется естественным потребовать, чтобы было

$$\mathsf{P}\{\rho(\xi, \xi') \leqslant \varepsilon\} = 1 \tag{W_ε^0}$$

или же чтобы

$$\mathsf{M}\rho^2(\xi, \xi') \leqslant \varepsilon^2. \tag{W_ε}$$

Эти два вида ε-энтропии H_ε распределения P_ξ мы будем обозначать символами

$$H_{W_\varepsilon^0}(\xi) = H_\varepsilon^0(\xi), \tag{4.3}$$

$$H_{W_\varepsilon}(\xi) = H_\varepsilon(\xi). \tag{4.4}$$

Отметим прежде всего некоторые оценки для величины

$$H_\varepsilon^0(X) = \sup_{P_\xi} H_\varepsilon^0(\xi), \tag{4.5}$$

где верхняя грань берется по всем распределениям вероятностей P_ξ на пространстве X. При $\varepsilon = 0$, как известно,

$$H_0^0(X) = \sup_{P_\xi} H(\xi) = \log N_X, \tag{4.6}$$

где N_X есть число элементов множества X (эта верхняя грань достигается при распределении, при котором все элементы X имеют одинаковую вероятность). При $\varepsilon > 0$

$$\log N_X^c(2\varepsilon) \leqslant H_\varepsilon^0(X) \leqslant \log N_X^a(\varepsilon), \tag{4.7}$$

где $N_X^a(\varepsilon)$ и $N_X^c(\varepsilon)$ — характеристики пространства X, которые введены в недавней заметке А. Н. Колмогорова [15]. Асимптотические свойства при $\varepsilon \to 0$ функций $N_X(\varepsilon)$, изученные для ряда конкретных пространств X в [15], являются интересными аналогами излагаемых далее результатов об асимптотическом поведении функции $H_\varepsilon(\xi)$.

Если теперь X — отрезок прямой и распределение P_ξ задается непрерывной плотностью $p(x)$, всюду положительной и удовлетворяющей условию Липшица, то при $\varepsilon \to 0$

$$H_\varepsilon^0(\xi) = \log \frac{1}{2\varepsilon} + h(\xi) + O\left(\varepsilon \log \frac{1}{\varepsilon}\right), \tag{4.8}$$

где $h(\xi) = -\int p(x) \log p(x)\, dx$ — «дифференциальная энтропия» распределения P_ξ, с которой мы уже встречались в § 1 (см. (1.21)). В случае n-мерного пространства X при аналогичных условиях получаем

$$H_\varepsilon^0(\xi) = n \log \frac{1}{d_n \varepsilon} + h(\xi) + O\left(\varepsilon \log \frac{1}{\varepsilon}\right), \tag{4.9}$$

где d_n — некоторая постоянная, сравнительно слабо изменяющаяся при изменении n.

Ничего здесь не изменится также, если вместо распределений, сосредоточенных на отрезке (или па кубе), взять произвольные достаточно гладкие и достаточно быстро убывающие на бесконечности распределения.

Аналогичные результаты получаются и для энтропии $H_\varepsilon(\xi)$: оказывается, что для достаточно гладких распределений P_ξ

$$H_\varepsilon(\xi) = \inf_{P_{\xi\xi'} \in W_\varepsilon} I(\xi, \xi') = \log \frac{1}{\varepsilon} - \log \sqrt{2\pi e} + h(\xi) + O(\varepsilon). \quad (4.10)$$

Для непрерывных распределений в n-мерном пространстве X (имеющих достаточно гладкую плотность $p(x)$) аналогичным образом при широких условиях получаем

$$H_\varepsilon(\xi) = n \log \frac{1}{\varepsilon} + \left[h(\xi) - n \log \sqrt{\frac{2\pi e}{n}} \right] + O(\varepsilon). \quad (4.11)$$

Мы видим, таким образом, что асимптотическое поведение $H_\varepsilon(\xi)$ в случае достаточно гладких непрерывных распределений в n-мерном пространстве в первую очередь определяется размерностью пространства и лишь в виде второго члена входит дифференциальная энтропия $h(\xi)$. Однако поскольку первый (основной) член $H_\varepsilon(\xi)$ не зависит от распределения вероятностей, то асимптотическое поведение $H_\varepsilon(\xi)$ для данного конкретного распределения естественно характеризовать величиной $h(\xi)$; это обстоятельство (наряду с формулой (1.21′)) и определяет значение дифференциальной энтропии для теории информации.

До сих пор мы говорили лишь о распределениях вероятностей в конечномерных пространствах. Естественно ожидать, что для типичных распределений в бесконечномерных пространствах рост $H_\varepsilon(\xi)$ при $\varepsilon \to 0$ будет существенно более быстрым. В виде простейшего примера рассмотрим случайную функцию Винера $\xi(t)$ на интервале $0 \leqslant t \leqslant 1$ с нормально распределенными независимыми приращениями

$$\Delta\xi = \xi(t + \Delta t) - \xi(t),$$

для которой

$$\xi(0) = 0, \qquad \mathsf{M}\Delta\xi = 0, \qquad \mathsf{M}(\Delta\xi)^2 = \Delta t. \quad (4.12)$$

Можно показать, что в этом случае в метрике L^2, т. е. при условии

$$\mathsf{M}\left\{\int_0^1 [\xi(t) - \xi'(t)]^2\, dt\right\} \leqslant \varepsilon^2, \quad (W_\varepsilon)$$

асимптотическое поведение $H_\varepsilon(\xi)$ определяется равенством

$$H_\varepsilon(\xi) = \frac{4}{\pi} \frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right). \quad (4.13)$$

Отсюда для более общего процесса Маркова диффузионного типа на отрезке времени $t_0 \leqslant t \leqslant t_1$ с

$$\mathsf{M}\Delta\xi = A(t, \xi(t))\, \Delta t + o(\Delta t), \quad (4.14)$$

$$\mathsf{M}(\Delta\xi)^2 = B(t, \xi(t))\, \Delta t + o(\Delta t)$$

можно при некоторых естественных допущениях получить формулу

$$H_\varepsilon(\xi)=\frac{4}{\pi}\chi\frac{1}{\varepsilon^2}+o\left(\frac{1}{\varepsilon^2}\right), \tag{4.15}$$

где

$$\chi(\xi)=\int_{t_0}^{t_1} \mathsf{M}B(t, \xi(t))\,dt. \tag{4.16}$$

То обстоятельство, что в формулах (4.13), (4.15) $H_\varepsilon(\xi)$ имеет порядок ε^{-2}, может быть обобщено следующим образом: для процесса $\xi(t)$, управляемого дифференциальным уравнением

$$d\,[A_0\xi^{(n)}(t)+A_1\xi^{(n-1)}(t)+\ldots+A_n\xi(t)]=d\psi(t),$$

где $\psi(t)$ — функция Винера, при $A_0\neq 0$ энтропия $H_\varepsilon(\xi)$ имеет порядок

$$H_\varepsilon(\xi)\asymp\varepsilon^{-1/(n+1/2)}.$$

Для случая нормального распределения в n-мерном евклидовом или в гильбертовом пространстве ε-энтропия H_ε может быть вычислена точно. n-мерный вектор ξ после надлежащего ортогонального преобразования координат приобретает вид

$$\xi=(\xi_1, \xi_2, \ldots, \xi_n),$$

где координаты ξ_k взаимно независимы и распределены нормально. При заданном ε для нахождения $H_\varepsilon(\xi)$ надо только определить параметр θ из уравнения

$$\varepsilon^2=\sum_k \min(\theta^2, \mathsf{D}\xi_k), \tag{4.17}$$

после чего будем иметь

$$H_\varepsilon(\xi)=\frac{1}{2}\sum_{\mathsf{D}\xi_k>\theta^2}\log\frac{\mathsf{D}\xi_k}{\theta^2}. \tag{4.18}$$

Аппроксимирующий вектор

$$\xi'=(\xi_1', \xi_2', \ldots, \xi_n')$$

следует выбирать нормально распределенным и таким, что при $\mathsf{D}\xi_k\leqslant\theta^2$

$$\xi_k'=0,$$

а при $\mathsf{D}\xi_k>\theta^2$

$$\xi_k=\xi_k'+\Delta_k,\qquad \mathsf{D}\Delta_k=\theta^2,\qquad \mathsf{D}\xi_k'=\mathsf{D}\xi_k-\theta^2$$

и векторы ξ_k' и Δ_k взаимно независимы. Бесконечномерный случай ничем не отличается от конечномерного.

Наконец, весьма существенно, что максимальное значение $H_\varepsilon(\xi)$ для вектора ξ (n-мерного или бесконечномерного) при заданных вторых центральных моментах достигается в случае нормального распределения. Этот результат может быть получен непосредственно или из следующего предложения М. С. Пинскера (ср. [19]):

Теорема. *Пусть заданы положительно определенная симметрическая матрица величин s_{ij}, $0 \leqslant i, j \leqslant m+n$, и распределение P_ξ вектора*

$$\xi = (\xi_1, \xi_2, \ldots, \xi_m),$$

центральные вторые моменты которого равны s_{ij} (при $0 \leqslant i, j \leqslant m$). Пусть условие W на совместное распределение $P_{\xi\xi'}$ вектора ξ и вектора

$$\xi' = (\xi_{m+1}, \xi_{m+2}, \ldots, \xi_{m+n})$$

заключается в том, что центральные вторые моменты величин

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

равны s_{ij} (при $0 \leqslant i, j \leqslant m+n$). Тогда

$$H_W(\xi) \leqslant {}^1/_2 \log (AB/C). \tag{4.19}$$

Обозначения в формуле (4.19) соответствуют изложению § 3. Из сопоставления со сказанным в § 3 видно, что неравенство (4.19) превращается в равенство в случае нормального распределения P_ξ.

§ 5
КОЛИЧЕСТВО ИНФОРМАЦИИ НА ЕДИНИЦУ ВРЕМЕНИ ДЛЯ СТАЦИОНАРНЫХ ПРОЦЕССОВ

Рассмотрим два стационарных и стационарно связанных процесса $\xi(t)$, $\eta(t)$, $-\infty < t < \infty$, или более общим образом, два стационарных и стационарно связанных обобщенных процесса $\xi(\varphi)$, $\eta(\varphi)$, $\varphi \in \Phi$. Будем обозначать через ξ_T и η_T отрезки процессов ξ и η за время $0 < t \leqslant T$ (ср. сноску 6) и через ξ_- и η_- значения тех же процессов на отрицательной полуоси $-\infty < t \leqslant 0$. Задать пару (ξ, η) стационарно связанных процессов — это значит задать инвариантное по отношению к сдвигам вдоль оси t распределение вероятностей $P_{\xi\eta}$ в пространстве пар функций $\{x(t), y(t)\}$ или, более общим образом, пар обобщенных функций $\{x(\varphi), y(\varphi)\}$.

Естественно назвать выражение

$$\bar{I}(\xi, \eta) = \lim_{T \to \infty} \frac{1}{T} I(\xi_T, \eta_T) \tag{5.1}$$

количеством информации в процессе ξ относительно процесса η

на единицу времени. Так как $I(\xi_T, \eta_T) = I(\eta_T, \xi_T)$, то всегда

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi). \tag{5.2}$$

Будем считать, что распределение $P_{\xi\eta}$ нормально. В этом случае при несколько отличном от (5.1) (но, по-видимому, приводящем к той же величине) определении предела $\bar{I}(\xi, \eta)$ М. С. Пинскер [19] установил, что по крайней мере в случае обыкновенных (не обобщенных) процессов, хоть один из которых регулярен в смысле [18], $\bar{I}(\xi, \eta)$ дается формулой

$$\bar{I}(\xi, \eta) = -\frac{1}{4\pi}\int\limits_{-\infty}^{\infty} \log\left[1 - r^2(\lambda)\right] d\lambda, \tag{5.3}$$

где

$$r^2(\lambda) = \frac{|f_{\xi\eta}(\lambda)|^2}{f_{\xi\xi}(\lambda)\, f_{\eta\eta}(\lambda)}, \tag{5.4}$$

а $f_{\xi\xi}$, $f_{\eta\eta}$ и $f_{\xi\eta}$ — соответствующие спектральные плотности — преобразования Фурье корреляционных функций $b_{\xi\xi}$, $b_{\eta\eta}$ и $b_{\xi\eta}$. Можно проверить, что формула (5.3) справедлива и во многих случаях, когда процессы ξ и η (или хотя бы один из них) являются обобщенными (см., в частности, выше формулу (3.29)). Возможно даже, что во всех случаях (как для обыкновенных, так и для обобщенных процессов) величина (5.1) будет даваться формулой (5.3); пока, однако, эта гипотеза еще никем не доказана и не опровергнута.

В случае стационарных процессов $\xi(t)$ точность воспроизведения ξ при помощи стационарного же и стационарно связанного с ξ процесса ξ' естественно характеризовать величиной

$$\sigma^2 = \mathsf{M}\,[\xi(a) - \xi'(a)]^2. \tag{5.5}$$

Если в качестве условия W взять условие

$$\sigma^2 \leqslant \varepsilon^2, \tag{5.6}$$

то величину

$$\bar{H}_W(\xi) = \bar{H}_\varepsilon(\xi) \tag{5.7}$$

естественно называть *ε-энтропией* на единицу времени процесса ξ. Из соответствующего предложения для конечномерных распределений (см. § 4) можно вывести, что при заданной спектральной плотности $f_{\xi\xi}(\lambda)$ величина $\bar{H}_\varepsilon(\xi)$ достигает максимума в случае нормального процесса ξ. В нормальном же случае величина $\bar{H}_\varepsilon(\xi)$ может быть легко вычислена по спектральной плотности $f_{\xi\xi}(\lambda)$ вполне аналогично тому, как это было объяснено в § 4 в применении к величине $H_\varepsilon(\xi)$ для конечномерных распределений. Для этого надо только определить параметр θ из уравнения

$$\varepsilon^2 = \int_{-\infty}^{\infty} \min(\theta^2, f_{\xi\xi}(\lambda))\, d\lambda; \tag{5.8}$$

после этого величина $\overline{H}_\varepsilon(\xi)$ находится по формуле

$$\overline{H}_\varepsilon(\xi) = \frac{1}{2} \int_{f_{\xi\xi}(\lambda) > \theta^2} \log \frac{f_{\xi\xi}(\lambda)}{\theta^2}\, d\lambda, \tag{5.9}$$

$\overline{I}(\xi, \xi') = \overline{H}_\varepsilon(\xi)$ достигается, когда

$$\xi = \xi' + \Delta,$$

ξ' и Δ нормально распределены, независимы и (см. рис. 1)

$$f_{\Delta\Delta}(\lambda) = \min(\theta^2, f_{\xi\xi}(\lambda)), \tag{5.10}$$

$$f_{\xi'\xi'}(\lambda) = \begin{cases} f_{\xi\xi}(\lambda) - \theta^2 & \text{при } f_{\xi\xi}(\lambda) > \theta^2, \\ 0 & \text{при } f_{\xi\xi}(\lambda) \leqslant \theta^2. \end{cases} \tag{5.11}$$

Формулы (5.8) и (5.9) позволяют в ряде случаев (например, в случае спектральной плотности (3.24)) явно рассчитать $\overline{H}_\varepsilon(\xi)$

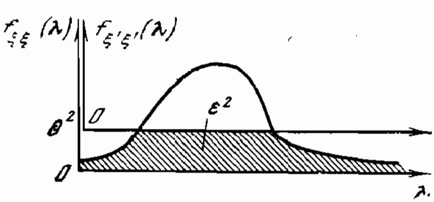


Рис. 1

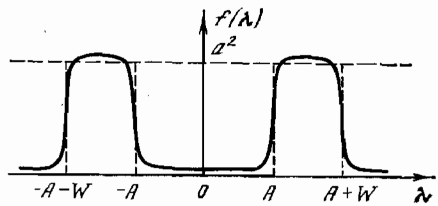


Рис. 2

как функцию от ε. Наибольший интерес, однако, представляет асимптотическое поведение $\overline{H}_\varepsilon(\xi)$ при $\varepsilon \to 0$. В частности, для плотности (3.24) это асимптотическое поведение дается формулой

$$\overline{H}_\varepsilon(\xi) = 16C\alpha/\pi\varepsilon^2 + O(1). \tag{5.12}$$

В более общем случае спектральной плотности, убывающей при $|\lambda| \to \infty$ как $c\,|\lambda|^{-k}$ (например, спектральной плотности (3.23)), мы будем иметь

$$\overline{H}_\varepsilon(\xi) \approx c_1 \varepsilon^{-2/(k-1)}. \tag{5.13}$$

Значительный практический интерес представляют спектральные плотности вида, изображенного на рис. 2, которые хорошо аппроксимируются формулой

$$f_{\xi\xi}(\lambda) = \begin{cases} a^2 & \text{при } A \leqslant |\lambda| \leqslant A + W, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{5.14}$$

Легко подсчитать, что в этом случае при не слишком малых ε (много больших той части энергии процесса ξ, которая заключена вне основной спектральной полосы $A \leqslant |\lambda| \leqslant A + W$) для нормального процесса с большой точностью будем иметь

$$\theta^2 \approx \varepsilon^2/2W, \qquad \overline{H}_\varepsilon(\xi) \approx W \log(2Wa^2/\varepsilon^2) \tag{5.15}$$

(отметим, что $\overline{H}_\varepsilon(\xi)$ здесь при $\varepsilon \to 0$ возрастает гораздо медленнее, чем в случае сравнительно медленного степенного убывания спектральной плотности на бесконечности). Формула (5.15), разумеется, есть не что иное, как хорошо известная формула Шеннона

$$R = W \log(Q/N) \tag{5.16}$$

[1, теорема 22]. Принципиальная новизна заключается здесь, однако, в том, что теперь мы видим, почему и в каких пределах (при не слишком малом ε) эта формула может быть применена к процессам с неограниченным спектром, какими являются все реально интересующие нас в теории передачи сообщений процессы. Записывая (5.15) в виде

$$\overline{H}_\varepsilon(\xi) \sim 2W[\log(a\sqrt{2W}) + \log(1/\varepsilon)] \tag{5.17}$$

и сравнивая с (4.11), мы видим, что двойная ширина $2W$ используемой полосы частот играет роль числа измерений. Эта идея эквивалентности двойной ширины полосы частот числу измерений, приходящихся в некотором смысле на единицу времени, была, по-видимому, впервые высказана В. А. Котельниковым [20]. В обоснование этой идеи В. А. Котельников указывал на то, что функция, спектр которой помещается в полосе ширины $2W$, однозначно определяется своими значениями в точках

$$\ldots, -\frac{2}{2W}, -\frac{1}{2W}, 0, \frac{1}{2W}, \frac{2}{2W}, \ldots, \frac{k}{2W}, \ldots$$

Эта же аргументация сохранена и у Шеннона, использующего полученные таким образом представления и для вывода формулы (5.16). Так как, однако, функция с ограниченным спектром всегда сингулярна в смысле [18] и наблюдение такой функции вообще не связано со стационарным притоком новой информации, то смысл такого рода аргументации оставался не вполне ясным, так что приведенный здесь новый вывод приближенной формулы (5.15) представляется не лишенным интереса.

ЛИТЕРАТУРА

1. *Shannon C. E.* A mathematical theory of communication. Pt. I, II.— Bell. Syst. Techn. J., 1948, vol. 27, N 3, p. 379—423; N 4, p. 623—656. Рус. пер.: *Шеннон К.* Математическая теория связи.— В кн.: Теория передачи электрических сигналов при наличии помех. М.: Изд-во иностр. лит., 1953.
2. *Халмош П. Р.* Теория меры. М.: Изд-во иностр. лит.. 1953.

3. *Kolmogoroff A.* Untersuchungen über den Integralbegriff.— Math. Ann., 1930, Bd. 103, S. 654—696.
4. *Смирнов В. И.* Курс высшей математики. М.; Л.: Гостехиздат, 1947, т. 5, § 90, 91.
5. *Itô K.* Stationary random distributions.— Mem. Coll. Sci. Univ. Kyoto. A, 1954, vol. 28, N 3, p. 209—223.
6. *Гельфанд И. М.* Обобщенные случайные процессы.— ДАН СССР, 1955, т. 100, № 5, с. 853—856.
7. *Schwartz L.* Théorie des distributions. Paris, 1950—1951. Vol. 1—2.
8. *Розенблат-Рот М.* Понятие энтропии в теории вероятностей и его применения в теории передачи по каналам связи.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1956, т. 2, с. 132—133.
9. *Розенблат-Рот М.* Энтропия стохастических процессов.— ДАН СССР, 1957, т. 112, № 1, с. 16—19.
10. *Витушкин А. Г.* К тринадцатой проблеме Гильберта.— ДАН СССР, 1954, т. 95, № 4, с. 701—704.
11. *Hotelling H.* Relation between two sets of variates.— Biometrica, 1936, vol. 28, p. 321—377.
12. *Обухов А. М.* Нормальная корреляция векторов.— Изв. АН СССР. Сер. мат., 1938, № 2, с. 339—370.
13. *Широков П. А.* Тензорное исчисление. М.; Л.: ГОНТИ, 1934.
14. *Яглом А. М.* Теория экстраполирования и фильтрации случайных процессов.— Укр. мат. журн., 1954, т. 6, № 1, с. 43—57.
15. *Колмогоров А. Н.* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств.— ДАН СССР, 1956, т. 108, № 3, с. 385—388.
16. *Хинчин А. Я.* Об основных теоремах теории информации.— УМН, 1956, т. 11, вып. 1, с. 17—75.
17. *Пинскер М. С.* Вычисление скорости создания сообщений стационарным случайным процессом и пропускной способности стационарного канала.— ДАН СССР, 1956, т. 111, № 4, с. 753—756.
18. *Колмогоров А. Н.* Стационарные последовательности в гильбертовом пространстве.— Бюл. МГУ. Математика, 1941, т. 2, № 6, с. 1—40.
19. *Пинскер М. С.* Количество информации о гауссовском случайном стационарном процессе, содержащейся во втором процессе, стационарно с ним связанном.— ДАН СССР, 1954, т. 99, № 2, с. 213—216.
20. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи.— В кн.: Матер. к I Всесоюз. съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. М., 1933.
21. *Колмогоров А. Н.* Теория передачи информации.— В кн.: Сессия Академии наук СССР по научным проблемам автоматизации производства, 15—20 окт. 1956 г.: Пленар. заседания. М.: Изд-во АН СССР, 1957, с. 66—99.
22. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* К общему определению количества информации.— ДАН СССР, 1956, т. 111, № 4, с. 745—748.
23. *Гельфанд И. М., Яглом А. М.* О вычислении количества информации о случайной функции, содержащейся в другой такой функции.— УМН, 1957, т. 12, вып. 1, с. 3—52.
24. *Kolmogorov A.* To the Shannon theory of information transmission in the case of continuons signales.— IRE Trans. Inform. Theory, 1956, vol. IT-2, N 4, p. 102—108.

5

НОВЫЙ МЕТРИЧЕСКИЙ ИНВАРИАНТ ТРАНЗИТИВНЫХ ДИНАМИЧЕСКИХ СИСТЕМ И АВТОМОРФИЗМОВ ПРОСТРАНСТВ ЛЕБЕГА *

Хорошо известно, что значительная часть метрической теории динамических систем может быть изложена как абстрактная теория «потоков» $\{S_t\}$ на «пространствах Лебега» M с мерой μ в терминах, инвариантных по отношению к «изоморфизмам по модулю нуль» (см. обзорную статью В. А. Рохлина [1], к которой дальнейшее изложение примыкает в отношении определений и обозначений). Меру на M мы будем предполагать нормированной условием

$$\mu(M) = 1 \tag{1}$$

и нетривиальной (т. е. предполагать существование множества $A \subset M$ с $0 < \mu(A) < 1$). Известно много примеров транзитивных автоморфизмов и транзитивных потоков с так называемым счетно-кратным лебеговым спектром (для автоморфизмов см. [1, § 4], для потоков [2—5]). Со спектральной точки зрения мы имеем здесь один тип автоморфизмов $\mathscr{L}_0^\omega$ и один тип потоков $\mathscr{L}^\omega$. Вопрос о том, не являются ли все автоморфизмы типа $\mathscr{L}_0^\omega$ (соответственно потоки типа $\mathscr{L}^\omega$) друг другу изоморфными mod 0, оставался до сих пор открытым. Мы показываем в § 3, 4, что ответ на этот вопрос отрицателен как в случае автоморфизмов, так и в случае потоков. Новый инвариант, позволяющий расщепить класс автоморфизмов $\mathscr{L}_0^\omega$ и класс потоков $\mathscr{L}^\omega$ на континуум инвариантных подклассов, есть энтропия на единицу времени. В § 1 излагаются необходимые сведения из теории информации (вводимые здесь понятия условной энтропии и условной информации и их свойства имеют, вероятно, и более широкий интерес, хотя все изложение непосредственно примыкает к определению количества информации из [7] и многочисленным работам, развивающим это определение). В § 2 дается определение характеристики h и доказывается ее инвариантность. В § 3, 4 указываются примеры автоморфизмов и потоков с произвольными значениями h в пределах $0 < h \leqslant \infty$. В случае автоморфизмов дело идет о примерах, давно построенных, в случае потоков построение примеров с конечным h — задача более деликатная и связанная с некоторыми любопытными вопросами теории марковских процессов.

---

* ДАН СССР, 1958, т. 119, № 5, с. 861—864.

§ 1
СВОЙСТВА УСЛОВНОЙ ЭНТРОПИИ И УСЛОВНОГО КОЛИЧЕСТВА ИНФОРМАЦИИ

В соответствии с [1] обозначаем через γ булеву алгебру измеримых множеств пространства M, рассматриваемых mod 0. Пусть $\mathfrak{C}$ — замкнутая в метрике $\rho(A, B) = \mu((A - B) \cup (B - A))$ подалгебра алгебры γ. Она порождает определенное mod 0 разбиение $\xi_{\mathfrak{C}}$ пространства M, определяемое тем условием, что $A \in \mathfrak{C}$ в том и только в том случае, когда mod 0 все A может быть составлено из полных элементов разбиения $\xi_{\mathfrak{C}}$. На элементах C разбиения $\xi_{\mathfrak{C}}$ определяется «каноническая система мер μ_C» [1]. Для любого $x \in C$ будем считать

$$\mu_x(A \mid \mathfrak{C}) = \mu_C(A \mid C). \tag{2}$$

С точки зрения теории вероятностей (где любая измеримая функция элемента $x \in M$ называется «случайной величиной») случайная величина $\mu_x(A \mid \mathfrak{C})$ есть «условная вероятность» события A при известном исходе «испытания» $\mathfrak{C}$ [6, гл. 1, § 7].

Для трех подалгебр $\mathfrak{A}$, $\mathfrak{B}$ и $\mathfrak{C}$ алгебры γ и $C \in \xi_{\mathfrak{C}}$ положим

$$I_C(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = \sup \sum_{i,\, j} \mu_x(A_i \cap B_j) \log \frac{\mu_x(A_i \cap B_j)}{\mu_x(A_i)\,\mu_x(B_j)}, \tag{3}$$

где верхняя грань берется по всем конечным разложениям $M = A_1 \cup A_2 \cup \ldots \cup A_n$, $M = B_1 \cup B_2 \cup \ldots \cup B_n$, для которых $A_i \cap A_j = N$, $B_i \cap B_j = N$, $i \neq j$, $A_i \in \mathfrak{A}$, $B_j \in \mathfrak{B}$ (N — пустое множество). Если $\mathfrak{C}$ есть тривиальная алгебра $\mathfrak{N} = \{N, M\}$, то (3) переходит в определение безусловной информации $I(\mathfrak{A}, \mathfrak{B})$ из приложения 7 к [7] [1]. Сама величина (3) интерпретируется как «количество информации в результатах испытания $\mathfrak{A}$ относительно испытания $\mathfrak{B}$ при известном исходе C испытания $\mathfrak{C}$». Если не фиксировать $C \in \xi_{\mathfrak{C}}$, то естественно рассматривать случайную величину $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$, которая при $x \in C$ равна $I_x(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = I_C(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$. Далее мы будем иметь дело с ее математическим ожиданием

$$\mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = \int_M I_x(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})\, \mu(dx). \tag{4}$$

Не требуют особых пояснений определения условной энтропии и средней условной энтропии $H(\mathfrak{A} \mid \mathfrak{C}) = I(\mathfrak{A}, \mathfrak{A} \mid \mathfrak{C})$, $\mathsf{M}H(\mathfrak{A}, \mathfrak{C}) = \int_M H_x(\mathfrak{A} \mid \mathfrak{C})\, \mu(dx)$.

---

[1] Авторы заметки [8] не обратили своевременно внимания на приложение 7 к [7], не включенное в русский перевод [9]. Заметка [8] должна была бы начинаться со ссылки на это приложение к [7].

Отметим те свойства условного количества информации и условной энтропии, которые нам понадобятся далее. Свойства (α) и (δ) для случая безусловных количества информации и энтропии общеизвестны, свойство (ε) для безусловного количества информации составляет содержание теоремы 2 из заметки [8]. Свойства (β) и (γ) доказываются без труда. По поводу свойства (β) следует лишь заметить, что аналогичное предложение для количества информации (из $\mathfrak{C} \supseteq \mathfrak{C}'$ вытекает: $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \leqslant I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}')$) было бы уже ошибочным. С этим связано то обстоятельство, что в свойстве (ζ) стоит нижний предел и знак $\geqslant$: соответствующий предел может не существовать, а нижний предел может в некоторых случаях оказаться больше $\mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$.

(α) $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \leqslant H(\mathfrak{A} \mid \mathfrak{C})$, равенство заведомо достигается при $\mathfrak{B} \supseteq \mathfrak{A}$.

(β) Если $\mathfrak{C} \supseteq \mathfrak{C}'$, то $H(\mathfrak{A} \mid \mathfrak{C}) \leqslant H(\mathfrak{A} \mid \mathfrak{C}')$, mod 0.

(γ) Если $\mathfrak{B} \supseteq \mathfrak{B}'$, то $\mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = \mathsf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{C}) + \mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C} \vee \mathfrak{B}')$, где $\mathfrak{C} \vee \mathfrak{B}'$ — минимальная замкнутая σ-алгебра, содержащая $\mathfrak{C}$ и $\mathfrak{B}'$.

(δ) Если $\mathfrak{B} \supseteq \mathfrak{B}'$, то $\mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \geqslant \mathsf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{C})$.

(ε) Если $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \ldots \subseteq \mathfrak{A}_n \ldots \bigcup\limits_n \mathfrak{A}_n = \mathfrak{A}$, то

$$\lim_{n\to\infty} \mathsf{M}I(\mathfrak{A}_n, \mathfrak{B} \mid \mathfrak{C}) = \mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}).$$

(ζ) Если $\mathfrak{C}_1 \supseteq \mathfrak{C}_2 \supseteq \ldots \supseteq \mathfrak{C}_n \supseteq \ldots$, $\bigcap\limits_n \mathfrak{C}_n = \mathfrak{C}$, то $\lim\limits_{n\to\infty} \inf \mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}_n) \geqslant \mathsf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$.

§ 2

ОПРЕДЕЛЕНИЕ ИНВАРИАНТА h

Будем говорить, что поток $\{S_t\}$ квазирегулярен (имеет тип $\mathscr{R}$), если [2] существует замкнутая подалгебра γ_0 алгебры γ, сдвиги которой $\gamma_t = S_t\gamma_0$ обладают следующими свойствами: (I) $\gamma_t \subseteq \gamma_{t'}$, если $t \leqslant t'$. (II) $\bigcup\limits_t \gamma_t = \gamma$. (III) $\bigcap\limits_t \gamma_t = \mathfrak{N}$.

При интерпретации потока как стационарного случайного процесса γ_t может рассматриваться как алгебра событий, «зависящих лишь от течения процесса до момента времени t». Легко доказывается, что потоки типа $\mathscr{R}$ транзитивны, а из результатов Плеснера [10, 11] можно вывести, что они имеют однородный лебегов спектр. Если кратность спектра равна ν ($\nu = 1, 2, \ldots, \omega$), то отнесем поток к типу $\mathscr{R}^\nu$. Очевидно, что $\mathscr{R}^\nu \subseteq \mathscr{L}^\nu$, где $\mathscr{L}^\nu$ — класс потоков с лебеговым спектром однородной кратности ν. Возможно, впрочем, что все $\mathscr{L}^\nu$ (и, следовательно, $\mathscr{R}^\nu$), кроме $\mathscr{L}^\omega$ ($\mathscr{R}^\omega$), пусты и что $\mathscr{L}^\omega = \mathscr{R}^\omega$.

[2] Это условие значительно слабее, чем условия «регулярности», обычно употребляемые в теории случайных процессов. См. об этом в конце § 4.

Теорема 1. *Если для потока $\{S_t\}$ существует γ_0, удовлетворяющее условиям* (I)—(III), *то при* $\Delta > 0$ $MH(\gamma_{t+\Delta} \mid \gamma_t) = h\Delta$, *где h — константа, лежащая в пределах* $0 < h \leqslant \infty$.

Теорема 2. *Константа h для данного потока $\{S_t\}$ не зависит от выбора γ_0, удовлетворяющего условиям* (I)—(III).

Наметим здесь доказательство теоремы 2. Пусть двум γ_0 и γ_0' соответствуют $h < \infty$ и h'. В силу теоремы 1 и лемм (α) и (ε) для любого $\varepsilon > 0$ можно найти такое k, что

$$h = MI(\gamma_{t+1} \mid \gamma_t) = MI(\gamma_{t+1}, \gamma \mid \gamma_t) \leqslant MI(\gamma_{t+1}, \gamma_{t+k}' \mid \gamma_t) + \varepsilon. \tag{5}$$

Из (5) в силу леммы (ζ) вытекает существование такого m, что

$$h \leqslant MI(\gamma_{t+1}, \gamma_{t+k}' \mid \gamma_t \vee \gamma_s') + 2\varepsilon \text{ при } t - s \geqslant m. \tag{6}$$

Из (6) и лемм (δ), (γ), (α), (β) (применять в указанном порядке!):

$$nh \leqslant \sum_{t=0}^{n-1} MI(\gamma_{t+1}, \gamma_{t+k}' \mid \gamma_t \vee \gamma_{-m}') + 2n\varepsilon \leqslant \tag{δ}$$

$$\leqslant \sum_{t=0}^{n-1} MI(\gamma_{t+1}, \gamma_{n+k}' \mid \gamma_t \vee \gamma_{-m}') + 2n\varepsilon = \tag{γ}$$

$$= MI(\gamma_n, \gamma_{n+k}' \mid \gamma_0 \vee \gamma_{-m}') + 2n\varepsilon \leqslant MH(\gamma_{n+k}' \mid \gamma_0 \vee \gamma_{-m}') + 2n\varepsilon \leqslant \tag{α}$$

$$\leqslant MH(\gamma_{n+k}' \mid \gamma_{-m}') + 2n\varepsilon \leqslant (n + k + m)h' + 2n\varepsilon, \tag{β}$$

$$h \leqslant \frac{n+k+m}{n} h' + 2\varepsilon. \tag{7}$$

Так как $\varepsilon > 0$ и n произвольны (причем n выбирается после фиксирования k и m), то из (7) вытекает неравенство $h \leqslant h'$. Это неравенство вполне аналогично доказывается и в случае $h = \infty$. Аналогично доказывается обратное неравенство $h' \leqslant h$, чем и заканчивается доказательство теоремы 2.

§ 3
ИНВАРИАНТЫ АВТОМОРФИЗМОВ

Если в § 2 считать, что t принимает только целые значения, то $\{S_t\}$ однозначно определяется автоморфизмом $T = S_1$. В силу теорем 1 и 2 существует инвариант $0 < h(T) \leqslant \infty$.

Легко доказывается, что любой автоморфизм типа $\mathscr{R}_0$ (индекс стоит для отличия от случая потоков с непрерывным временем) имеет счетнократный лебегов спектр, т. е. из классов $\mathscr{R}_0^\nu$ непуст только класс $\mathscr{R}_0^\omega \subseteq \mathscr{L}_0^\omega$. Он распадается по значениям $h(T)$ на классы $\mathscr{R}_0^\omega(h)$.

Теорема 3. *Для любого h, $0 < h \leqslant \infty$, существует автоморфизм, принадлежащий $\mathscr{R}_0^\omega(h)$.*

Соответствующие примеры хорошо известны и получаются, например, из схемы независимых случайных испытаний $\mathscr{L}_{-1}$, $\mathscr{L}_0, \mathscr{L}_1, \ldots, \mathscr{L}_t, \ldots$ с распределением вероятностей исхода ξ_t испытания $\mathscr{L}_t$:

$$\mathsf{P}\{\xi_t = a_i\} = p_i, \qquad -\sum_{i=1}^{\infty} p_i \log p_i = h. \tag{8}$$

Пространство M составляется из последовательностей $x = (\ldots, x_{-1}, x_0, x_1, \ldots, x_t, \ldots)$, $x_t = a_1, a_2, \ldots$, а сдвиг $Tx = x'$ определяется формулой $x'_t = x_{t-1}$. Мера μ на M определяется как прямое произведение вероятностных мер (8).

§ 4
ИНВАРИАНТЫ ПОТОКОВ

Теорема 4. *Для любого h, $0 < h < \infty$, существует поток класса $\mathscr{K}^{\omega}(h)$, т. е. поток со счетнократным лебеговым спектром и заданным значением константы h.*

По аналогии с § 3 естественно возникает идея — воспользоваться для доказательства теоремы 4 вместо схемы дискретных независимых испытаний схемой «процессов с независимыми приращениями», или обобщенных процессов «с независимыми значениями» [12, 13]. Однако этот путь приводит лишь к потокам класса $\mathscr{K}^{\omega}(\infty)$ [5]. Для получения конечных значений h приходится воспользоваться более искусственным построением. В этой заметке возможно только дать описание одного из таких построений.

Определим взаимно независимые случайные величины ξ_n, соответствующие всем целым n, распределениями их значений: $\mathsf{P}\{\xi_0 = k\} = 3\cdot 4^{-k}$, $k = 1, 2, \ldots$, а при $n \neq 0$ $\mathsf{P}\{\xi_n = k\} = 2^{-k}$, $k = 1, 2, \ldots$ Точку τ_0 на оси t расположим в случае $\xi_0 = k$ с равномерным распределением вероятностей на отрезке $-u2^{-k} \leqslant \tau_0 \leqslant 0$, а точки τ_n при $n \neq 0$ определим из соотношения $\tau_{n+1} = \tau_n + u2^{-\xi_n}$.

Положим $\varphi(t) = \xi_n$ при $\tau_n \leqslant t < \tau_{n+1}$. Легко проверить, что распределение случайной функции $\varphi(t)$ инвариантно по отношению к сдвигам $S_t\varphi(t_0) = \varphi(t_0 - t)$. Легко подсчитать, что $h\{S_t\} = 6/u$ (на единицу времени падает в среднем $3/u$ точек τ_n, а на каждое ξ_n приходится энтропия $\sum_{k=1}^{\infty} k2^{-k} = 2$).

Можно получить более наглядное представление о нашем случайном процессе, если включить в описание его состояния $\omega(t)$ в момент времени t, кроме величины $\varphi(t)$, еще значение $\delta(t) = t - \tau^*(t)$ разности между t и ближайшей слева от t точки τ_n. При таком способе описания наш процесс оказывается стационарным марковским процессом. Он заслуживает лишь название «квазирегулярного», так как, хотя соответствующая ему динамическая система транзитивна, значение разности $f(\omega(t), t) = \tau^*(t) = t - \delta(t)$ определяется с точностью до двоично-рационального слагаемого поведением реализации процесса в сколь угодно далеком прошлом.

21 января 1958 г.

ЛИТЕРАТУРА

1. *Рохлин В. А.* Избранные вопросы метрической теории динамических систем.— УМН, 1949, т. 4, вып. 2 (30), с. 57—128.
2. *Гельфанд И. М., Фомин С. В.* Геодезические потоки на многообразиях постоянной отрицательной кривизны.— УМН, 1952, т. 7, вып. 1 (47), с. 118—137.
3. *Фомин С. В.* О динамических системах в пространстве функции.— Укр. мат. журн., 1950, т. 2, № 2, с. 25—47.
4. *Itô K.* Complex multiple Wiener integral.— Jap. J. Math., 1952, vol. 22, p. 63—86.
5. *Itô K.* Spectral type of the shift transformation of differential processes with stationary increments.— Trans. Amer. Math. Soc., 1956, vol. 81, N 2, p. 253—263.
6. *Дуб Дж. Л.* Вероятностные процессы. М.: Изд-во иностр. лит., 1956. 605 с.
7. *Shannon C. E., Weaver W.* The mathematical theory of communication. Urbana: Univ. Ill. Press, 1949.
8. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* К общему определению количества информации.— ДАН СССР, 1956, т. 111, № 4, с. 745—748.
9. *Шеннон К.* Статистическая теория передачи электрических сигналов.— В кн.: Теория передачи электрических сигналов при наличии помех. М.: Изд-во иностр. лит., 1953, с. 7—87.
10. *Плеснер А. И.* Функции максимального оператора.— ДАН СССР, 1939, т. 23, № 4, с. 327—330.
11. *Плеснер А. И.* О полуунитарных операторах.— ДАН СССР, 1939, т. 25, № 9, с. 708—710.
12. *Itô K.* Stationary random distributions.— Mem. Coll. Sci. Univ. Kyoto. A, 1954, vol. 28, N 3, p. 209—223.
13. *Гельфанд И. М.* Обобщенные случайные процессы.— ДАН СССР, 1955, т. 100, № 5, с. 853—856.

6

К ОПРЕДЕЛЕНИЮ АЛГОРИТМА *[1]

Совместно с В. А. Успенским

Алгоритмом принято называть систему вычислений, которая для некоторого класса математических задач из записи

A

«условий» задачи позволяет при помощи однозначно определенной последовательности операций, совершаемых «механически»

---

* УМН, 1958, т. 13, вып. 4, с. 3—28.

[1] Статья, кроме незначительных редакционных изменений, была написана, когда авторы, стремясь для самих себя осмыслить понятия «вычисли-

без вмешательства творческих способностей человека, получить запись

B

«решения» задачи. Вопрос о путях уточнений этого традиционного приблизительного представления об алгоритмах с наибольшей широтой поставлен во вводных параграфах статьи А. А. Маркова [1]. Давая ясное и развернутое представление о существе вопроса, статья А. А. Маркова (и последовавшая за ней монография [2]) ограничивается алгоритмами, перерабатывающими «слова», а среди такого рода алгоритмов — теми, которые в статье названы «нормальными». Нашей целью является наметить возможность более широкого подхода к делу и вместе с тем более отчетливо показать, что самое общее доступное при современном состоянии науки понятие алгоритма вполне естественным образом связывается с понятием частично рекурсивной функции.

Во всех интересующих математиков случаях доступные переработке данным алгоритмом записи условий A легко включаются в занумерованную неотрицательными целыми числами последовательность

$$A_0, A_1, A_2, \ldots, A_n, \ldots,$$

а записи могущих получиться решений B — в последовательность

$$B_0, B_1, B_2, \ldots, B_n, \ldots,$$

тоже занумерованную неотрицательными целыми числами[2]. Если обозначить через G множество номеров n тех условий A_n, которые алгоритм способен переработать в решения, то резуль-

---

мая функция» и «алгоритм», пытались пересмотреть найденные в литературе варианты определения и окончательно убедиться в том, что за ними не скрывается каких-либо возможностей расширения самого объема понятия «вычислимая функция». Результат получился, естественно, отрицательным. Не без колебаний авторы публикуют отчет о своих поисках, так как предметно он не дает ничего нового по сравнению с более простыми определениями. По-видимому, однако, усвоенный нами подход к делу соответствует умонастроению многих математиков, которых публикация нашей статьи может избавить от повторного прохождения тех же путей сомнений и размышлений. Отметим еще, что обсуждался и такой вариант определений, в которых вместо пассивной, неограниченного объема памяти в машине могли возникать в неограниченном количестве параллельно работающие «активные области», каждая из которых сохраняет ограниченную сложность строения. Естественно, что и на этом пути не было получено ничего, кроме частично рекурсивных функций.

[2] Способ, позволяющий по виду записи находить ее номер, а также по номеру восстанавливать саму запись, является обычно весьма простым (так что существование алгоритма, «перерабатывающего» запись в номер, и алгоритма, «перерабатывающего» номер в запись, не вызывает сомнений).

тат работы алгоритма, осуществляющего переработку

$$A_n \to B_m,$$

однозначно определяется заданной на G числовой функцией

$$m = \varphi(n).$$

Таким образом, произвольный алгоритм сводится к алгоритму вычисления значений некоторой числовой функции (*числа* всюду далее имеются в виду целые неотрицательные).

Обратно, если для функции φ существует алгоритм, который, будучи применен к стандартной записи [3] значения аргумента n из области определения функции φ, приводит к стандартной записи значения функции $m = \varphi(n)$, то функцию φ естественно называть *алгоритмически вычислимой*, или для краткости просто *вычислимой функцией*. Поэтому вопрос об определении алгоритма по существу равносилен вопросу об определении вычислимой функции. В § 1 дается обзор существующих определений вычислимой функции и алгоритма и обсуждается степень их общности и логической завершенности. В § 2 излагается новое определение алгоритма [4]. В § 3 показывается, что любой алгоритм, подпадающий под это новое, весьма общее по форме определение, все же сводится к алгоритму вычисления значений частично рекурсивной функции. В приложении I дается сводка определений и фактов, относящихся к рекурсивным функциям, а в приложении II рассматривается пример алгоритма.

§ 1

Мы остановимся на следующих вариантах математического определения вычислимой функции или алгоритма.

А) Определение вычислимой функции как функции, значения которой выводимы в некотором логическом исчислении (Гёдель [4], Чёрч [5] [5]).

Б) Определение вычислимой функции как функции, значения которой получаются при помощи исчисления λ-конверсии Чёрча [5, 7].

В) Определение вычислимой функции как функции частично рекурсивной (см. работу Клини [8]) [6] или — для случая всюду определенной функции — как общерекурсивной (Клини [10]).

---

[3] Для чисел, больших нуля, стандартной записью можно считать запись вида $1 + 1 + \ldots + 1$ или запись по десятичной системе счисления и т. п. Без фиксирования стандартного способа записи чисел говорить об алгоритме вычисления $m = \varphi(n)$ по n не имеет смысла.

[4] Основные черты этого определения, найденного еще в 1952 г., были впервые опубликованы в [3].

[5] По поводу определения А) см. также [6, § 62, 63].

[6] По поводу определения В) см. также [6, § 62, 63].

(Термины «частично рекурсивная» и «общерекурсивная» понимаются здесь в смысле приложения I.)

Г) Вычислительная машина Тьюринга [11] [7].

Д) Финитный комбинаторный процесс Поста [13].

Е) Нормальный алгорифм А. А. Маркова [1, 2].

Рассмотрим прежде всего перечисленные определения с точки зрения естественно входящего в понятие алгоритма требования наличия однозначно определенной последовательности операций, «перерабатывающих» условия A в решение B, или значение n в значение $m = \varphi(n)$. Сразу ясно, что определения А), Б), В) являются с этой точки зрения незавершенными, так как такой однозначно определенной последовательности операций не указывают. Правда, например, определение частично рекурсивной функции по существу содержит в себе все предпосылки для построения однозначно определенного алгоритма вычисления ее значений по заданному значению ее аргумента, но в явном виде такой алгоритм не указывается.

Вычислительная машина Тьюринга производит вполне определенную последовательность операций, приводящую к последовательному развертыванию записей значений

$\varphi(0),\ \varphi(1),\ \varphi(2), \ldots$

Достаточно снабдить ее простым приспособлением, которое остановило бы ее работу при получении $\varphi(n)$ для заданного n, чтобы она могла рассматриваться как способ реализации подлинного алгоритма нахождения $\varphi(n)$ по заданному n [8].

Определения Д) и Е) полностью отвечают требованию однозначной определенности алгоритмического вычислительного процесса заданием условий A.

Перейдем теперь к обсуждению рассматриваемых определений с точки зрения расчленения алгоритмического процесса на элементарные, могущие выполняться «механически» отдельные шаги. В этом требовании расчленения вычислительного процесса на элементарные шаги *ограниченной* (в пределах данного алгоритма) *сложности* мы усматриваем вторую, не менее существенную сторону понятия алгоритма. Чтобы сделать понятными основные трудности, связанные с уточнением понятия алгоритма в смысле выполнения требования ограниченной сложности его отдельных шагов, заметим, что интересующие математиков алгоритмы применимы обычно к *бесконечному* классу задач. Таковы, например, уже алгоритмы сложения и умножения записанных по

---

[7] Здесь имеется в виду первоначальное определение вычислимой функции посредством машины Тьюринга, предложенное самим Тьюрингом [11]. Близкое к этому изложение имеется в книге Петер [12].

[8] В этом духе изложено вычисление функции на машине Тьюринга в монографии Клини [6]. Отправляясь от стандартной записи числа n, машина получает стандартную запись числа $\varphi(n)$ и останавливается.

десятичной системе натуральных чисел. Поэтому уподобление процесса предписываемых алгоритмом вычислений работе некоторой «машины» не может пониматься слишком буквально. Реальная машина допускает только конечное число четко разграниченных «состояний», и в соответствии с этим число различных «условий» A, которые реальная машина способна перерабатывать в «решения» B, неизбежным образом тоже конечно. Поэтому построение теории алгоритмов по образцу теории вычислительных машин требует во всяком случае некоторой идеализации понятия «машины».

Тем не менее наглядные образы и понятия, выработанные в связи с развитием теории вычислительных машин дискретного действия, могут быть полезны для общей ориентировки в проблеме рационального определения понятия алгоритма. Вся идеализация, необходимая для перехода от реальных вычислительных машин к математическим алгоритмам, заключается в допущении неограниченного объема «памяти» машины. Можно иллюстрировать изложение теории алгоритмов и более традиционными образами, связанными с вычислениями, записываемыми на бумаге. По сравнению с реальным процессом таких вычислений идеализация будет заключаться в допущении неограниченного объема записей, постепенно накапливаемых и вновь используемых по определенной системе в дальнейших выкладках.

Допуская неограниченный объем «памяти», понятие математического алгоритма должно сохранить из свойств реальных вычислительных машин дискретного действия и реальных письменных вычислений следующие два их свойства.

1. Сами вычислительные операции производятся дискретными шагами, причем на каждом шагу используется лишь *ограниченный* запас накопленных на предшествующих шагах данных. Если запас данных, имеющихся к началу шага, превосходит максимальную емкость «активной области», то они вовлекаются в эту активную область постепенно, по мере освобождения в ней места (и уже на следующих шагах). Можно, например, представлять себе дело так, что листки с записями, употребляемыми на каждом шаге вычислений, должны умещаться на столе и вмещают строго ограниченное число знаков, а папки, в которые складываются исписанные листки и из которых они извлекаются по мере надобности, имеют неограниченную толщину.

2. Неограниченность «объема памяти», или сохраняемых в запас записей, понимается чисто *количественным* образом: допускается лишь неограниченное накопление элементов (знаков, элементарных частей машины) ограниченного числа типов, связанных между собой связями ограниченной сложности и тоже принадлежащих к ограниченному числу различных типов. Извлечение сведений, накопленных в «памяти» машины или в отложенных в сторону записях, производится последовательно: от

попавшего в активную область элемента переходят к связанным с ним элементам и вовлекают их в активную область.

Высказанным сейчас требованиям в полной мере удовлетворяет вычислительная машина Тьюринга и финитный комбинаторный процесс Поста. В определениях А), Б) и В) остается нерасшифрованной действительная элементарность и «механическая» осуществимость таких операций, как подстановка вместо переменных тех или иных выражений, сложность которых (в смысле числа составляющих их элементарных знаков) правилами исчисления никак не ограничена. Аналогичное возражение формально может быть отнесено и к определению нормального алгорифма А. А. Маркова, хотя здесь возможность его снятия при помощи легкого дополнительного построения особенно ясна. Недостаточно расчлененной с точки зрения выдвинутого нами требования в нормальном алгорифме А. А. Маркова является только операция разыскания в перерабатываемом слове P «первого вхождения» слова X. Так как перерабатываемые слова P могут быть сколько угодно длинными, то нахождение первого вхождения в P данного слова X должно рассматриваться как процесс, требующий, вообще говоря, ряда последовательных операций. Можно сказать, что при определении нормальных алгорифмов А. А. Марков предполагает уже построенным некий специальный «алгорифм нахождений первых вхождений». Читатель может без большого труда при желании построить такой алгоритм в терминах нашего § 2.

Можно, однако, предполагать, что как А. А. Марков, так и авторы различных обсуждаемых нами определений понятия вычислимой функции оставляют некоторые операции неограниченного объема нерасчлененными лишь в силу очевидной возможности их расчленения на действительно элементарные шаги.

Последним и самым трудным вопросом является вопрос о достаточной *общности* предложенных определений алгоритма и вычислимой функции. На первый взгляд весьма общим является определение А). Однако эта общность обманчива. В самом деле, чтобы сформулировать определение вычислимой в смысле А) функции, мы неизбежно ограничиваемся некоторым фиксированным исчислением с фиксированными правилами вывода. Если же поставить вопрос о выводе в произвольном исчислении, с произвольными правилами вывода, то следует учесть, что понятие вывода в исчислении в самом общем виде может быть уточнено лишь на базе уже уточненного понятия алгоритма.

Что касается вполне четких с формальной стороны определений Б), В), Г), Д), Е), то можно сказать, что в известном смысле слова они все эквивалентны друг другу, т. е. по существу определяют классы вычислительных процессов (алгоритмов) одинаковой мощности. Эквивалентность определений Б), В), Е) установлена в работах [5, 14 — 16] [9]. Определения Д) и Е) являются специальными случаями алгоритмов в смысле нашего § 2, а эквивалентность нашего определения определению В) выясняется в § 3 нашей работы. Заметим, впрочем, что сама точная формулировка предложений об эквивалентности определений, построенных с формальной стороны столь различным образом и определяющих формально разные объекты (вычислимые функции; в случае Тьюринга — лишь вычислимые функции, принимающие только два значения 0 и 1; алгоритмы, перерабатывающие слова, и т. д.), представляет некоторые затруднения.

Как бы то ни было, именно ряд установленных или подразумеваемых предложений об эквивалентности всех предлагавшихся, начиная с 1936 г., определений алгоритма или вычислимой функции является эмпирической основой сложившегося к настоящему времени общего убеждения в окончательности полученных определений. В применении к вычислимым функциям (для определенности от натурального аргумента с натуральными значениями) сложившиеся к настоящему времени общепринятые представления могут быть изложены так.

Класс алгоритмически вычислимых числовых функций, определенных для всех натуральных n, совпадает с классом *общерекурсивных* функций (определение этих последних см. приложение I). Что касается функций, определенных лишь на некотором подмножестве G полного множества натуральных чисел, то их алгоритмическое задание можно понимать в двух различных смыслах. При первом, более узком понимании предписанная алгоритмом процедура должна для любого натурального n после конечного числа шагов привести либо к нахождению $m = \varphi(n)$, либо к установлению того, что n не входит в G. При втором, более широком понимании дела требуется только, чтобы предписанная алгоритмом процедура для любого n из G приводила к верному результату $m = \varphi(n)$, а при попытке применить ее к натуральному n, не входящему в G, не могла кончиться установлением ошибочного (так как φ вне G не определена) результата о равенстве $\varphi(n)$ какому-либо определенному натуральному числу m. В остальном при втором (широком) понимании дела от характера работы алгоритма, примененного к не входящему в G числу n, ничего не требуется: допускаются в этом случае как безрезультатная «остановка» алгоритмической процедуры после конечного числа шагов, так и случай, когда алгоритмическая процедура продолжается неограниченно, не приводя ни к какому результату.

Целесообразно (что и делается в дальнейшем изложении) считать основным второе (широкое) понимание алгоритмической вычислимости числовой функции, определенной на множестве

---

[9] См. также примечание переводчика на с. 341 книги [6].

натуральных чисел. В настоящее время это широкое понятие алгоритмически вычислимой функции идентифицируется с понятием *частично рекурсивной* функции (см. приложение I). Возможными областями определения G алгоритмически вычислимых функций $\varphi(n)$ оказываются тогда *рекурсивно перечислимые* множества натуральных чисел (определение последних см. приложение I) [10]. Соответствующим образом строится и общее понятие алгоритма, предназначенного для переработки «условий» A того или иного более общего вида в «решения» B. Для любого алгоритма Г класс $\mathfrak{A}$ (Г) формально допустимых записей «условий» A и класс $\mathfrak{B}$ (Г) возможных записей «решений» B определяются таким образом, чтобы принадлежность или непринадлежность к ним любой могущей встретиться записи могла без затруднений распознаваться. Но при этом не требуется, чтобы алгоритм Г, примененный к любой записи A класса $\mathfrak{A}$ (Г), приводил после конечного числа шагов к записи B из класса $\mathfrak{B}$ (Г). Вместо этого допускают как возможность остановки алгоритмической процедуры без получения решения, так и возможность бесконечного ее продолжения. Множество $\mathfrak{S}$ (Г) тех записей A из $\mathfrak{A}$ (Г), которые алгоритм Г перерабатывает в «решения», т. е. в записи B из $\mathfrak{B}$ (Г), будем называть *областью применимости алгоритма.*

Наиболее интересные результаты теории алгоритмов относятся как раз к алгоритмам с нетривиальной областью применимости; основные полученные до настоящего времени результаты о *невозможности* алгоритмов могут быть приведены к такому виду: для некоторого алгоритма Г доказывается невозможность «разрешающего» алгоритма Ψ, который «распознавал» бы принадлежность или непринадлежность любой записи A из $\mathfrak{A}$ (Г) к $\mathfrak{S}$ (Г) [11].

Классы $\mathfrak{A}$ (Г) и $\mathfrak{B}$ (Г) для всех имеющихся определений алгоритма легко нумеруются, как указано во введении к этой статье. Это позволяет каждому алгоритму Г поставить в соответствие числовую функцию φ_Γ, которая определена для номеров n условий A, принадлежащих $\mathfrak{S}$ (Г), и в качестве значения $m = \varphi(n)$ имеет номер решения, получаемый при переработке условия с номером n. Сложившееся к настоящему времени общее убеждение в степени общности понятия алгоритма можно резюмировать, высказав, что для любого алгоритма Г функция φ_Γ должна быть частично рекурсивной. Так и обстоит дело со всеми предложенными до настоящего времени определениями. На-

---

[10] Изложение некоторых основных фактов теории вычислимых функций, предполагающее лишь интуитивное представление об алгоритмах, дано в статье [17].

[11] Т. е. невозможность алгоритма Ψ, который перерабатывал бы любую запись A из $\mathfrak{A}$ (Г) в 1, когда A входит в $\mathfrak{S}$ (Г), и в 0, когда A не входит в $\mathfrak{S}$ (Г).

шей задачей является возможно более полное разъяснение причин такого положения вещей. С этой целью мы и предприняли попытку построения возможно более общего *по форме* определения алгоритма, удовлетворяющего выдвинутым выше требованиям *ограниченной сложности* каждого шага алгоритма. В результате, как и следовало ожидать, получилось, что, несмотря на все меры, принятые для сохранения возможной общности построения, мы не выходим за пределы алгоритмов, описываемых указанным выше способом частично рекурсивными функциями.

§ 2

Алгоритм Г задается при помощи предписания, указывающего способ перехода от «состояния» S процесса вычислений к «следующему» состоянию S^*, т. е. при помощи оператора $\Omega_\Gamma (S) = S^*$, который мы назовем оператором «непосредственной переработки». В классе $\mathfrak{S} (\Gamma) = \{S\}$ возможных состояний выделяются класс $\mathfrak{A} (\Gamma)$ «исходных» состояний, представляющих собой записи «условий» задач, и класс $\mathfrak{C} (\Gamma)$ «заключительных» состояний. При получении состояния из класса $\mathfrak{C} (\Gamma)$ алгоритмический процесс заканчивается, и из полученного «заключительного» состояния извлекается «решение». Область $\mathfrak{D} (\Gamma) \subseteq \mathfrak{S} (\Gamma)$, на которой оператор Ω определен, естественно считать не пересекающейся с $\mathfrak{C} (\Gamma)$. Класс состояний, не входящих ни в $\mathfrak{C} (\Gamma)$, ни в $\mathfrak{D} (\Gamma)$, обозначим через $\mathfrak{R} (\Gamma)$.

Алгоритмический процесс

$$S^0 = A \in \mathfrak{A} (\Gamma),$$

$$S^1 = \Omega_\Gamma (S^0),$$

$$S^2 = \Omega_\Gamma (S^1),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$S^{t+1} = \Omega_\Gamma (S^t)$$

может развиваться одним из следующих трех способов.

1. При каком-либо $t = \bar{t}$ процесс приводит к заключительному состоянию

$$S^{\bar{t}} \in \mathfrak{C} (\Gamma),$$

после чего из этого заключительного состояния извлекается решение B.

2. При каком-либо $t = \bar{t}$ процесс заканчивается безрезультатно:

$$S^{\bar{t}} \in \mathfrak{R} (\Gamma).$$

3. Процесс продолжается неограниченно, не приводя к результату.

Класс $\mathfrak{S}$ (Г) тех $A \in \mathfrak{A}$ (Г), которые приводят к первому типу течения процессов, называется «областью применимости» алгоритма.

Перейдем теперь к уточнению понятий «состояния» и «активной части» состояния. Задачей тех рассмотрений, которые сейчас будут изложены, является лишь наглядное оправдание достаточной общности того формального определения этих понятий, которое будет дано несколькими страницами дальше.

Естественно считать, что состояние S определяется наличием некоторого конечного числа элементов

$$\bigcirc_1, \bigcirc_2, \ldots, \bigcirc_v,$$

принадлежащих каждый к одному из типов

$$T_0, T_1, T_2, \ldots, T_n,$$

и наличием между некоторыми группами элементов связей, принадлежащих к одному из типов

$$R_1, R_2, \ldots, R_m.$$

Для каждого типа связи R_i будем считать фиксированным число k_i связываемых элементов. Запас типов элементов и запас типов связей (а следовательно, и числа n, m, $k_1, \ldots, k_m$) выбираются для данного алгоритма раз навсегда.

Элементы будем обозначать кружками с проставляемыми *внутри* номерами типов. Что касается индексов *около* кружков, то они будут относиться лишь к нашим рассуждениям относительно алгоритма и к составу состояния не относятся. Так как число элементов, образующих вместе с заданными между ними связями состояние, неограниченно, то и разнообразие этих внешних индексов заранее никак не ограничивается: можно, например, в качестве этих индексов употреблять сколь угодно большие натуральные числа.

Наконец, в соответствии со сказанным ранее будем считать, что в пределах одного алгоритма число связей, в которые может одновременно (т. е. при конструировании одного определенного состояния) входить один и тот же элемент, ограниченно раз навсегда выбранным числом.

Вообще говоря, порядок связываемых какой-либо связью элементов может быть существенным или несущественным. Далее нам удастся ограничиться рассмотрением только симметричеcких связей между парами элементов. Однако сначала, для того чтобы не упустить каких-либо возможностей построения мощных алгоритмов, мы предположим, что в понятие состояния существенно входит порядок, в котором связанные какой-либо связью элементы в эту связь включаются; т. е., например, наличие парной связи

$$R\,(\bigcirc_1, \bigcirc_2)$$

будем отличать от наличия парной связи

$R(\bigcirc_2, \bigcirc_1)$.

Будем также считать существенно входящим в понятие состояния и задание определенного упорядочения всех связей, в которые входит данный элемент. Таким образом, связывание элементов $\bigcirc_1, \bigcirc_2, \ldots, \bigcirc_k$ связью типа R будем мыслить себе осуществляющимся по схеме, изображенной на рис. 1.

В этой схеме индексы p_j указывают, какое место занимает рассматриваемая связь типа R в упорядоченной последовательности связей, в которые входит элемент $\bigcirc_j$. Индексы p_j при этом

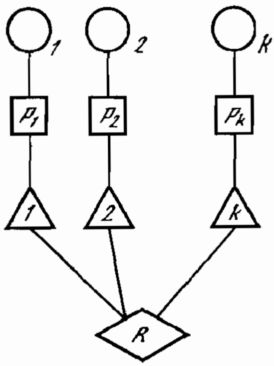


Рис. 1

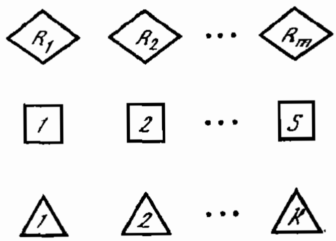
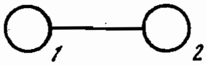


Рис. 2

можно считать принимающими только значения $1, 2, \ldots, s$. Что касается индексов $1, 2, \ldots, k$, то они принимают, вообще говоря, значения из последовательности $1, 2, \ldots, K$, где K равно максимуму чисел k_i $(i = 1, 2, \ldots, m)$.

Вместо рассмотрения состояний, составляемых из элементов, связанных по указанной сейчас общей схеме разнообразными связями $R_1, R_2, \ldots, R_m$, целесообразно ввести дополнительные типы элементов (рис. 2).

Очевидно, что это позволит заменить рассматриваемый алгоритм новым, отличающимся лишь способом записи состояний, в котором будет применяться только один тип симметричных связей между парами элементов. Такую связь между элементами $\bigcirc_1$ и $\bigcirc_2$ можно изображать просто чертой

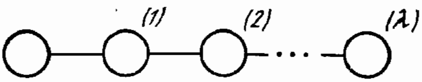


Для преобразованного алгоритма автоматически будет выполняться

Условие α). *Все элементы, связанные с каким-либо одним элементом, принадлежат различным типам.*

Естественно, что при условии α) специальная нумерация порядка вхождения элемента в различные связи излишняя: связи, в которые входит данный элемент, автоматически нумеруются номерами типов тех элементов, с которыми он связывается.

Можно без потери общности аналогичным образом стандартизировать и способ выделения «активной части» состояния, строение которой однозначно определяет те преобразования, которые переводят S в $S^* = \Omega\ (S)$. Именно будем считать, что в активной части состояния имеется отмеченный «начальный» элемент. Требование ограниченной сложности активной части состояния естественно понять так, что все входящие в нее элементы должны быть связаны с начальным элементом $\bigcirc$ цепями



ограниченной длины $\lambda \leqslant N$, где N — характеристическое число, постоянное для данного алгоритма.

Оператор

$$S^* = \Omega_\Gamma\ (S),$$

дающий переход от состояния S к «следующему» состоянию S^*, должен быть таков, что S^* строится исключительно на основании информации о виде активной части S. Точно так же на основании информации о виде активной части S должно распознаваться, достигнуто заключительное состояние или нет. Проще всего, что очевидным образом не нарушит общности, использовать для этого начальный элемент: мы примем, что к типам T_0 и T_1 может относиться только начальный элемент, причем переход начального элемента от типа T_0 к типу T_1 означает получение заключительного состояния.

Вовсе не всё заключительное состояние целесообразно считать решением. Так, при прекращении вычислений на бумаге не все написанное является решением: подавляющую часть обычно составляют промежуточные выкладки. Точно так же при остановке вычислительной машины только сравнительно небольшая часть полного «состояния» машины является решением решаемой на машине задачи. Естественно поэтому считать «решением» лишь некоторую часть «заключительного состояния», связанную с начальным элементом. Мы примем в качестве «решения» совокупность всех тех элементов, которые можно связать с начальным элементом цепями произвольной длины.

Переходим теперь к формулировке определения алгоритма.

1. Алгоритм Г предполагает наличие упорядоченного множества $\mathfrak{T} = \mathfrak{T}\ (\Gamma)$ неограниченных по объему классов

$$T_0,\ T_1, \ldots,\ T_\Gamma$$

«элементов», из которых будут строиться «состояния». Классы эти не пересекаются между собой. Соединение их обозначается T. Элементы T обозначаются кружками $\bigcirc$ с различными индексами около кружков. Для обозначения принадлежности элемента $\bigcirc$ к классу T_i внутри кружка ставится номер i. Элемент, принадлежащий классу T_i, называется также элементом типа T_i.

2. Комплексом над множеством $\mathfrak{T}$ называется обычный одномерный комплекс с вершинами из T, т. е. соединение $K = K_0 \bigcup \bigcup K_1$ конечного множества

$$K_0 = \{\bigcirc_\alpha\}$$

некоторых элементов из T, называемых «вершинами» комплекса, и множества K_1 (очевидно, тоже конечного)

$$K_1 = \{\bigcirc_{\alpha'} - \bigcirc_{\alpha''}\}$$

некоторых пар элементов из K_0. Эти пары называются «отрезками» комплекса K.

3. Выделяется множество $\mathfrak{S}_\mathfrak{T}$ таких комплексов K над $\mathfrak{T}$, которые обладают следующими свойствами:

а) вершины, соединенные отрезками с фиксированной вершиной, принадлежат разным типам T_i;

б) в K существует одна и только одна вершина типа T_0 или T_1, называемая «начальной»; остальные вершины принадлежат типам T_i с $i \geqslant 2$.

4. В множестве $\mathfrak{S}_\mathfrak{T}$ выделяются подмножество $\mathfrak{A}_\mathfrak{T}$ комплексов, начальная вершина которых принадлежит типу T_0, и подмножество $\mathfrak{E}_\mathfrak{T}$ комплексов, начальная вершина которых принадлежит типу T_1.

5. Комплексы из $\mathfrak{S}_\mathfrak{T}$ считаются «состояниями» алгоритма Г, причем комплексы из $\mathfrak{A}_\mathfrak{T}$ — «исходными» состояниями, или «условиями», а комплексы из $\mathfrak{E}_\mathfrak{T}$ «заключительными» состояниями:

$$\mathfrak{S}(\Gamma) = \mathfrak{S}_\mathfrak{T}, \qquad \mathfrak{A}(\Gamma) = \mathfrak{A}_\mathfrak{T}, \qquad \mathfrak{E}(\Gamma) = \mathfrak{E}_\mathfrak{T}.$$

6. Активной частью $U(S)$ состояния S называется подкомплекс комплекса S, состоящий из вершин и отрезков, принадлежащих цепям длины $\lambda \leqslant N$, содержащим начальную вершину. Здесь N — произвольное, но для данного алгоритма Г фиксированное число.

Термин «подкомплекс» понимается в обычном теоретико-множественном смысле: комплекс K' есть подкомплекс комплекса K, если $K_0' \subseteq K_0$, $K_1' \subseteq \subseteq K_1$. Цепь есть комплекс вида

$$\bigcirc_0 - \bigcirc_1 - \bigcirc_2 - \cdots - \bigcirc_\lambda$$

Длиной цепи называется число λ входящих в нее отрезков. Комплекс называется связным, если любые две его вершины можно соединить цепью.

7. Внешней частью $V\ (S)$ состояния S называется подкомплекс, состоящий из вершин, не соединимых с начальной вершиной цепями длины $\lambda < N$, и отрезков, не входящих в те цепи, содержащие начальную вершину, которые имеют длину $\lambda \leqslant N$. Пересечение

$$L\ (S) = U\ (S) \cap V\ (S)$$

называется границей активной части состояния S. Легко видеть, что $L\ (S)$ есть нульмерный комплекс (т. е. не содержит отрезков) и состоит из тех вершин, которые соединимы с начальной цепями длины $\lambda = N$, но не соединимы более короткими цепями (длины $\lambda < N$).

Прежде чем закончить формулировку определения алгоритма, сделаем несколько замечаний.

Сформулированное в § 1 требование 1 мы будем понимать так: оператор

$$\Omega_\Gamma\ (S) = S^*$$

должен быть таков, чтобы перестройка S в S^* требовала лишь перестройки в какой-либо новый комплекс активной части $U\ (S)$ и не меняла строения внешней части $V\ (S)$; при этом сам характер перестройки $U(S)$ должен определиться исключительно строением самого комплекса $U\ (S)$.

Два комплекса K' и K'' будем считать *изоморфными*, если их (как множества $K' = K_0' \cup K_1'$ и $K'' = K_0'' \cup K_1''$) можно поставить во взаимно однозначное соответствие так, что:

1) вершины соответствуют вершинам, а отрезки — отрезкам,

2) отрезку, соединяющему вершины $\bigcirc_\alpha'$ и $\bigcirc_\beta'$, соответствует отрезок, соединяющий соответствующие вершины $\bigcirc_\alpha''$ и $\bigcirc_\beta''$,

3) соответствующие вершины имеют одинаковый тип.

Очевидно, что число неизоморфных активных частей $U\ (S)$, возможных в данном алгоритме, ограничено. Правила их переработки поэтому легко формулируются в конечном виде. Возвращаемся к изложению определения алгоритма.

8. Правила непосредственной переработки алгоритма задаются при помощи указания пар состояний

$$U_1 \to W_1,$$

$$U_2 \to W_2,$$

$$\cdot\quad\cdot\quad\cdot\quad\cdot$$

$$U_r \to W_r$$

с установленным для каждого i изоморфным отображением φ_i подкомплекса $L(U_i)$ на некоторый подкомплекс $\tilde{L}(W_i)$ состояния W_i. Если вершины $\bigcirc_\mu \in L(U_i)$ и $\bigcirc_\nu \in \tilde{L}(W_i)$ соответствуют друг другу при изоморфизме φ_i, то произвольная не начальная вершина комплекса W_i, связанная с $\bigcirc_\nu$, имеет тип одной из вершин комплекса U_i, связанных с $\bigcirc_\mu$ [12]. Каждый комплекс U_i является состоянием класса $\mathfrak{A}(\Gamma)$ (условием), удовлетворяющим требованию $U(U_i) = U_i$. Все $U_1, U_2, \ldots, U_r$ предполагаются неизоморфными между собой. Что же касается W_i, то это — произвольные состояния, которые могут быть как исходными, так и заключительными.

9. Область $\mathfrak{D}(\Gamma)$ определения оператора $\Omega_\Gamma(S) = S^*$ состоит из тех состояний S, для которых активная часть $U(S)$ изоморфна одному из комплексов $U_1, U_2, \ldots, U_r$.

10. Пусть подкомплекс $U(S)$ изоморфен U_i. Так как $U(S)$ и U_i суть связные комплексы, принадлежащие к $\mathfrak{S}_\mathfrak{T}$, то, если они изоморфны, между ними может существовать лишь одно изоморфное соответствие. Это соответствие однозначно определяет изоморфизм между $L(S)$ и $L(U_i)$, а так как задано изоморфное отображение φ_i комплекса $L(U_i)$ на $\tilde{L}(W_i)$, то тем самым индуцируется однозначно определенный изоморфизм θ_i^S между $L(S)$ и $\tilde{L}(W_i)$.

11. Пусть $S \in \mathfrak{D}(\Gamma)$, причем подкомплекс $U(S)$ изоморфен комплексу U_i. Очевидно, можно образовать комплекс $\tilde{W}$, изоморфный комплексу W_i и удовлетворяющий следующим условиям:

1) $\tilde{W} \cap V(S) = L(S)$,

2) изоморфизм θ_i^S между подкомплексами $L(S) \subseteq \tilde{W}$ и $\tilde{L}(W_i) \subseteq W_i$ продолжается до изоморфизма между комплексами $\tilde{W}$ и W_i.

Результат применения оператора Ω_Γ к комплексу S определяется (однозначно с точностью до изоморфизма) как комплекс вида

$$S^* = \tilde{W} \cup V(S).$$

12. Если S — заключительное состояние, то связная компонента начальной вершины считается «решением» (под связной компонентой какой-либо вершины понимается максимальный связный подкомплекс, содержащий эту вершину) [13]; совокупность

---

[12] Вследствие этого требования свойство a) из п. 3 определения не нарушится в процессе работы алгоритма.

[13] Очевидно, мы получим то же самое «решение», если будем переходить к связной компоненте начальной вершины на каждом шаге работы алгоритма, т. е. если будем определять результат S^* непосредственной переработки комплекса S как *связную компоненту начальной вершины* комплекса $\tilde{W} \cup V(S)$. При таком способе задания оператора Ω_Γ все состояния алгоритмического процесса, кроме, быть может, начального состояния, окажутся

связных компонент начальных вершин заключительных состояний может быть в соответствии с ранее сказанным обозначена через $\mathfrak{B}$ (Г).

Теперь определение алгоритма заканчивается точно так, как указано в начале этого параграфа (в отношении трех типов течения алгоритмического процесса и определения их «области применимости» $\mathfrak{S}$ (Г)).

Заданный указанным способом алгоритм, характеризующийся упорядоченным множеством $\mathfrak{T}$ и числом N, мы отнесем к классу $\mathbf{A}_N$ ($\mathfrak{T}$). Всякий алгоритм класса $\mathbf{A}_N$ ($\mathfrak{T}$) мы отнесем к классу $\mathbf{A}_{\mathfrak{T}}$.

Для того чтобы сделать описание алгоритма более обозримым, мы ввели некоторые условности, которые не связаны неразрывно с общим замыслом, но нам кажется, что достаточная общность предложенного определения остается убедительной, несмотря на эти условности. Нам представляется убедительным, что произвольный алгоритмический процесс удовлетворяет нашему определению алгоритма. Хочется при этом подчеркнуть, что речь идет не о сводимости любого алгоритма к алгоритму в смысле нашего определения (так, в следующем параграфе будет устанавливаться сводимость любого алгоритма к алгоритму вычислений частично рекурсивной функции), а о том, что любой алгоритм по существу *подходит* под предложенное определение.

Трудности, могущие возникнуть у читателя, если он попытается представить, скажем, вычисление значения частично рекурсивной функции в виде определенного выше алгоритма над комплек-

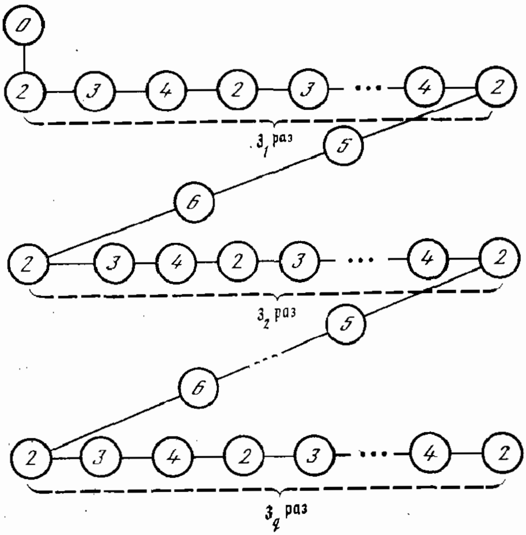


Рис. 3

связными комплексами и потому «решение» будет совпадать с заключительным состоянием.— *Примеч. авт. при подготовке статьи к публикации в данной книге.*

сами, обусловлены лишь тем, что, как мы уже отмечали в § 1, схема вычисления значений частично рекурсивной функции не задана непосредственно в виде алгоритма. Если же развернуть эти вычисления в виде алгоритмического процесса (что, конечно, нетрудно сделать), то тем самым автоматически получится некоторый алгоритм в смысле предложенного определения. Сформулируем высказанное утверждение более точно.

Рассмотрим множество типов $\mathfrak{T}_6 = \{T_0, T_1, T_2, T_3, T_4, T_5, T_6\}$. Назовем «$A$-изображением» системы q чисел $(\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q)$ комплекс $A_{\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q}$ (рис. 3), а «B-изображением» той же системы — комплекс $B_{\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q}$, получающийся из $A_{\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q}$ заменой вершины 0 на вершину 1.

Для каждой частично рекурсивной функции $f(x_1, x_2, \ldots, x_q)$ существует такое множество типов $\mathfrak{T} \supseteq \mathfrak{T}_6$ и такой алгоритм Γ класса $\mathbf{A}(\mathfrak{T})$, который применим к комплексу $A_{\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q}$ тогда и только тогда, когда f определена для системы чисел $\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q$, и в этом последнем случае перерабатывает этот комплекс в комплекс $B_{\mathfrak{t}}$, где

$$\mathfrak{t} = f(\mathfrak{i}_1, \mathfrak{i}_2, \ldots, \mathfrak{i}_q).$$

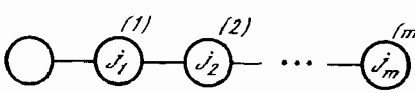
Все сказанное о рекурсивных функциях полностью относится и ко всем определениям А) — Е). Как только из любого из этих определений удается извлечь алгоритмический процесс, он, этот процесс, оказывается алгоритмом в указанном выше смысле.

§ 3

Всякий алгоритм сводится к вычислению значений некоторой частично рекурсивной функции. Это доказывается так называемым методом арифметизации.

Прежде всего заметим, что каждый комплекс S из $\mathfrak{S}_{\mathfrak{T}}$ можно задавать в виде таблицы. Для этого упорядочим множество вершин комплекса S следующим образом[14].

Каждой цепи



,

[14] А. В. Гладкий в реферате настоящей статьи, опубликованном: РЖ. Математика, 1959, № 7, с. 9 (реф. 6527), справедливо замечает: «Примененный в § 3 способ арифметизации нуждается в исправлении, так как участвующий здесь прием упорядочения множества вершин годится лишь для связных комплексов. Нужное исправление может быть сделано без труда».— *Примеч. ред.*

соединяющей начальную вершину ◯ с произвольной вершиной ◯$^{(m)}$, поставим в соответствие строчку

$$j_1, j_2, \ldots, j_m.$$

В силу условия α) установленное соответствие между исходящими от начальной вершины ◯ цепями вершины комплекса S и некоторым множеством строчек натуральных чисел будет взаимно однозначным. Если упорядочить строчки чисел в словарном порядке, то автоматически возникает порядок среди цепей, исходящих от начальной вершины. Каждой вершине из S поставим теперь в соответствие минимальную в смысле установленного порядка цепь, соединяющую начальную вершину с выбранной нами; это соответствие индуцирует порядок в множестве вершин (начальная вершина будет при этом первой). Таким образом, множество вершин каждого комплекса из S оказывается упорядоченным.

Естественно отнести каждому комплексу из $\mathfrak{S}_{\mathfrak{T}}$ симметричную таблицу, в которой i_k есть номер типа k-й (в смысле установленного порядка) вершины из S, а α_{kl} есть 1 или 0 в зависимости от того, соединены или нет k-я и l-я вершины. По таблице комплекс восстанавливается однозначно с точностью до изоморфизма.

	i_1	i_2	$\ldots$	i_v
i_1	α_{11}	α_{12}	$\ldots$	α_{1v}
i_2	α_{21}	α_{22}	$\ldots$	α_{2v}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i_v	α_{v1}	α_{v2}	$\ldots$	α_{vv}

Проведенное построение переводит всякий алгоритм класса $\mathbf{A}$ ($\mathfrak{T}$) на язык таблиц.

Каждой таблице поставим в соответствие ее «гёделевское число», для чего запишем таблицу в строчку

$$\alpha_{11}\alpha_{12} \ldots \alpha_{1v}\alpha_{21}\alpha_{22} \ldots \alpha_{2v} \ldots \alpha_{v1}, \alpha_{v2}, \ldots \alpha_{vv},$$

а строчке отнесем число

$$p_1^{i_1} p_2^{i_2} \ldots p_v^{i_v} p_{v+1}^{\alpha_{11}} p_{v+2}^{\alpha_{12}} \ldots p_{v^2+v}^{\alpha_{vv}},$$

где p_r есть r-е простое число.

В натуральном ряде возникает бесконечное подмножество Q тех чисел, которые суть «гёделевские числа» таблиц, отвечающих комплексам из $\mathfrak{S}_{\mathfrak{T}}$. Принадлежность или непринадлежность числа к Q распознается без труда. Перенумеруем элементы Q в естественном порядке натуральными числами. *Номером* таблицы назовем номер соответствующего ей «гёделевского числа». *Номером* комплекса из $\mathfrak{S}_{\mathfrak{T}}$ назовем номер представляющей этот комплекс таблицы.

По комплексу S из $\mathfrak{S}_{\mathfrak{T}}$ эффективно и однозначно находится его номер s, а по номеру s эффективно и однозначно (с точностью до изоморфизма) восстанавливается самый комплекс S.

Мы имеем здесь дело с той ситуацией, о которой говорилось в § 1. Множество $\mathfrak{S}_{\mathfrak{T}}$ нумеруется натуральными числами:

$$S_0,\ S_1,\ S_2,\ \ldots,\ S_p,\ \ldots;$$

в эту последовательность комплексов включаются как множество $\mathfrak{A}_{\mathfrak{T}}$ «условий», так и множество $\mathfrak{B}_{\mathfrak{T}}$ «решений». Каждый алгоритм Г класса $\mathbf{A}$ ($\mathfrak{T}$) задает отображение своей области применимости $\mathfrak{G}$ (Г) в множество $\mathfrak{B}_{\mathfrak{T}}$; это отображение индуцирует числовую функцию

$$m = \varphi_{\Gamma}(n),$$

заданную на множестве G тех номеров n, для которых $S_n \in \mathfrak{G}(\Gamma)$. Мы покажем, что функция φ_{Γ} частично рекурсивна.

В самом деле, нетрудно показать, что функция $s^* = \sigma(s)$, индуцированная на множестве номеров оператором непосредственной переработки $S^* = \Omega_{\Gamma}(S)$, является примитивно рекурсивной (определение примитивно рекурсивных функций см. приложение I). Отсюда сразу следует примитивная рекурсивность функции $\rho(n, t)$, дающей номер комплекса S^t, получающегося на t-м шагу переработки комплекса с номером n. Процесс продолжается до тех пор, пока комплекс $S^{\bar{t}}$, возникший на шаге $\bar{t}$, не окажется «заключительным», т. е. комплексом из $\mathfrak{C}$ (Г). Номера комплексов из $\mathfrak{C}(\Gamma) = \mathfrak{C}_{\mathfrak{T}}$, т. е. номера комплексов, начальная вершина которых принадлежит классу T_1, удовлетворяют уравнению

$$\gamma(x) = 0,$$

где функцию γ, так легко показать, всегда можно выбрать из числа примитивно рекурсивных. Процесс, таким образом, продолжается до того первого t, которое будет удовлетворять условию

$$\gamma(\rho(n, \bar{t})) = 0,$$

т. е. до числа

$$\bar{t} = \mu t\,[\gamma(\rho(n, t)) = 0]$$

(об операторе μ см. приложение I). Комплекс $S^{\bar{t}}$ принадлежит $\mathfrak{C}$ (Г), его номер $\rho(n, \bar{t})$ находится из равенства

$$\rho(n, \bar{t}) = \rho(n, \mu t\,[\gamma(\rho(n, t)) = 0]).$$

Номер $m = \varphi_{\Gamma}(n)$ решения получается из номера комплекса $S^{\bar{t}}$ при помощи функции β (тоже примитивно рекурсивной), дающей по номеру комплекса из $\mathfrak{S}_{\mathfrak{T}}$ номер связной компоненты его начальной вершины. Именно

$$\varphi_{\Gamma}(n) = \beta(\rho(n, \mu t\,[\gamma(\rho(n, t)) = 0])), \tag{1}$$

откуда следует (см. приложение I), что функция $\varphi_\Gamma(n)$ частично рекурсивная.

Отсюда вытекает, в частности, что функция $f(x_1, \ldots, x_q)$, задаваемая алгоритмом, преобразующим A-изображения наборов из q чисел в B-изображения чисел, частично рекурсивна.

Действительно, можно построить примитивно рекурсивные функции ξ и η, удовлетворяющие следующим требованиям:

1) для всякого набора чисел $\mathfrak{z}_1, \ldots, \mathfrak{z}_q$ число $\xi(\mathfrak{z}_1, \ldots, \mathfrak{z}_q)$ есть номер A-изображения набора $\mathfrak{z}_1, \ldots \mathfrak{z}_q$,

2) для всякого числа n, являющегося номером B-изображения числа $\mathfrak{t}$, значение $\eta(n)$ равно $\mathfrak{t}$.

Если теперь обозначить через Γ алгоритм, осуществляющий вычисление функции f, то, очевидно,

$$f(x_1, \ldots, x_q) = \eta(\varphi_\Gamma(\xi(x_1, \ldots, x_q))), \tag{2}$$

откуда и получается, что f — частично рекурсивная функция.

Поскольку для каждой частично рекурсивной функции f существует вычисляющий ее (в смысле, указанном в конце § 2) алгоритм Γ, то каждая частично рекурсивная функция будет допускать представление (2). Используя формулу (1), получаем

$$f(x_1, \ldots, x_q) = \\ = \eta(\beta(\rho(\xi(x_1, \ldots, x_q), \mu t\,[\gamma(\rho(\xi(x_1, \ldots, x_q), t)) = 0]))).$$

Полагая для сокращения

$$\eta(\beta(\rho(\xi(x_1, \ldots, x_q), v))) = \pi(x_1, \ldots, x_q, v),$$

$$\gamma(\rho(\xi(x_1, \ldots, x_q), t)) = \tau(x_1, \ldots, x_q, t),$$

получаем окончательно, что каждая частично рекурсивная функция представима в виде

$$f(x_1, \ldots, x_q) = \pi(x_1, \ldots, x_q, \mu t\,[\tau(x_1, \ldots, x_q, t) = 0]), \tag{3}$$

где π и τ суть примитивно рекурсивные функции.

Приложение I

СВЕДЕНИЯ ИЗ ТЕОРИИ РЕКУРСИВНЫХ ФУНКЦИЙ

Мы рассматриваем функции от одного, двух и более переменных, причем каждое переменное и сама функция принимают значения из натурального ряда. Каждое натуральное число условимся считать функцией от нулевого числа переменных. Мы не требуем, чтобы функция от n переменных была определена для всех систем из n натуральных чисел. Равенство между двумя функциями

$$f(x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_n)$$

означает, что для всякой системы из n натуральных чисел $(\mathfrak{z}_1, \mathfrak{z}_2, \ldots, \mathfrak{z}_n)$, для которой определена одна из этих функций, определена и другая и

$$f(\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_n) = g(\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_n).$$

Введем ряд операторов на множестве функций. Аргументом каждого оператора является функция или система функций.

Оператор суперпозиции. Область определения оператора суперпозиции — системы функций вида

$$\psi(x_1, x_2, \ldots, x_n);\ \varphi_1(x_1, \ldots, x_m),\ \varphi_2(x_1, \ldots, x_m), \ldots$$
$$\ldots, \varphi_n(x_1, \ldots, x_m)$$
$$(n = 1, 2, \ldots;\ m = 1, 2, \ldots).$$

Будучи применен к такой системе, оператор суперпозиции дает функцию $\theta(x_1, \ldots, x_m)$, являющуюся результатом подстановки функций $\varphi_1, \varphi_2, \ldots, \varphi_n$ в функцию ψ на места ее переменных:

$$\theta(x_1, \ldots, x_m) =$$
$$= \psi(\varphi_1(x_1, \ldots, x_m), \varphi_2(x_1, \ldots, x_m), \ldots, \varphi_n(x_1, \ldots, x_m)).$$

Функция $\theta(x_1, \ldots, x_m)$ определена для всех наборов натуральных чисел $(\mathfrak{x}_1, \ldots, \mathfrak{x}_m)$, обладающих следующими свойствами:

1) для $(\mathfrak{x}_1, \ldots, \mathfrak{x}_m)$ определены все функции φ_i $(i = 1, 2, \ldots, n)$,

2) если $\varphi_i(\mathfrak{x}_1, \ldots, \mathfrak{x}_m) = \mathfrak{t}_i$ $(i = 1, 2, \ldots, n)$, то функция ψ определена для системы $(\mathfrak{t}_1, \mathfrak{t}_2, \ldots, \mathfrak{t}_n)$.

Оператор примитивной рекурсии. Область определения — пары функций вида

$$h(x_1, \ldots, x_{n-1}), \qquad g(x_1, \ldots, x_{n-1}, x_n, x_{n+1}) \quad (n = 1, 2, \ldots).$$

Примененный к такой паре, оператор рекурсии дает функцию $f(x_1, \ldots, x_{n-1}, x_n)$, связанную с h и g следующими равенствами:

$$f(x_1, \ldots, x_{n-1}, 0) = h(x_1, \ldots, x_{n-1}),$$

$$f(x_1, \ldots, x_{n-1}, x_{n+1}) =$$
$$= g(x_1, \ldots, x_{n-1}, x_n, f(x_1, \ldots, x_{n-1}, x_n)).$$

Оператор наименьшего числа. Область определения — совокупность всех функций от одного или более переменных. Будучи применен к функции $\theta(x_1, \ldots, x_{n-1}, x_n)$, оператор наименьшего числа дает функцию $\psi(x_1, \ldots, x_{n-1})$, задаваемую следующим законом: для всякого набора натуральных чисел $\mathfrak{x}_1, \ldots, \mathfrak{x}_{n-1}$ значение функции

$$\psi(\mathfrak{x}_1, \ldots, \mathfrak{x}_{n-1})$$

есть такое число $\mathfrak{t}$, что:

1) $\theta(\mathfrak{x}_1, \ldots, \mathfrak{x}_{n-1}, \mathfrak{t}) = 0$;

2) для всякого числа $\mathfrak{x}$, меньшего чем $\mathfrak{t}$, значение $\theta(\mathfrak{x}_1, \ldots, \mathfrak{x}_{n-1}, \mathfrak{x})$ определено и не равно нулю.

Результат применения оператора наименьшего числа к функции $\theta(x_1, \ldots, x_{n-1}, x_n)$ записывают обычно в виде

$$\mu x_n [\theta(x_1, \ldots, x_{n-1}, x_n) = 0],$$

помня при этом, что запись

$$\mu x_n [\theta(x_1, \ldots, x_{n-1}, x_n) = 0]$$

обозначает функцию лишь от переменных $x_1, \ldots, x_{n-1}$.

Оператор наименьшего числа в отличие от операторов суперпозиции и рекурсии может, будучи применен ко всюду определенной функции, дать функцию, не всюду определенную (и даже нигде не определенную).

Введем следующие три группы функций:

1) Функции $O_n(x_1, \ldots, x_n)$ при $n = 0, 1, 2, \ldots$; для каждого n функция $O_n(x_1, \ldots, x_n)$ тождественно равна нулю.

2) Функции $I_{nk}(x_1, \ldots, x_n)$ при $n = 1, 2, \ldots$ и $k = 1, 2, \ldots, n$; для любого набора $(\mathfrak{z}_1, \ldots, \mathfrak{z}_k, \ldots, \mathfrak{z}_n)$ имеем $I_{nk}(\mathfrak{z}_1, \ldots, \mathfrak{z}_k, \ldots, \mathfrak{z}_n) = \mathfrak{z}_k$.

3) Функция $\lambda(x)$, равная $x + 1$.

Все эти функции назовем *базисными*.

Класс *примитивно рекурсивных функций* определяется как минимальный класс, содержащий базисные функции и замкнутый относительно применения оператора суперпозиции и оператора примитивной рекурсии [18, 6]. Каждая примитивно рекурсивная функция, как легко видеть, всюду определена. Все простейшие арифметические функции, как то:

$$x + y, \quad xy, \quad x^y, \quad |x - y|, \quad \left[\frac{x}{y}\right], \quad [\sqrt[x]{y}], \quad [\log_x y], \quad x!$$

и многие другие (например, показатель, с которым простое число входит в разложение на простые множители числа y) являются примитивно рекурсивными.

Класс *частично рекурсивных* функций определяется как минимальный класс, содержащий базисные функции и замкнутый относительно применения оператора суперпозиции, оператора примитивной рекурсии и оператора наименьшего числа [19, 6].

Частично рекурсивная функция от n переменных, определенная для всех систем из n натуральных чисел, называется *общерекурсивной* [19, 6]. В частности, всякая примитивно рекурсивная функция есть функция общерекурсивная. Существуют общерекурсивные функции, не являющиеся примитивно рекурсивными.

Каждая частично рекурсивная функция конструируется из примитивно рекурсивных функций посредством применения операторов суперпозиции, примитивной рекурсии и наименьшего числа. Представляет интерес наименьшее число операторов, требующееся для образования частично рекурсивной функции из примитивно рекурсивных. A priori это число может быть сколь угодно велико. Однако существует важная теорема Клини [19], утверждающая, что всякая частично рекурсивная функция $f(x_1, \ldots, x_n)$ может быть получена применением оператора наименьшего числа и оператора суперпозиции из двух примитивно рекурсивных функций (из которых одна является раз навсегда фиксированной, а другая зависит от функции $f(x_1, \ldots, x_n)$)[1]. Именно теорема Клини гласит, что существует такая примитивно рекурсивная функция $U(x)$, что для всякой частично рекурсивной функции $f(x_1, \ldots, x_n)$ существует примитивно рекурсивная функция $\theta(x_1, \ldots, x_n, y)$ со следующим свойством:

$$f(x_1, \ldots, x_n) = U(\mu y\,[\theta(x_1, \ldots, x_n, y) = 0]).$$

Формула Клини важна еще и тем, что она сводит к минимуму употребление оператора наименьшего числа; применение этого оператора всегда неприятно, ибо он, как отмечалось, может из всюду определенных функций конструировать не всюду определенные.

Мы скажем, что множество R порождается функцией $f(x)$, если множество значений $f(x)$ есть R. Множество называется *рекурсивно перечислимым*, если оно либо пусто, либо есть множество значений некоторой общерекурсивной функции [20]. Класс рекурсивно перечислимых множеств совпадает с классом множеств, порождаемых частично рекурсивными функциями. Класс непустых рекурсивно перечислимых множеств совпадает с классом множеств, порождаемых примитивно рекурсивными функциями [21]. Следующие множества являются рекурсивно перечислимыми: сумма и пересечение двух рекурсивно перечислимых множеств; образ и полный прообраз рекурсивно перечислимого множества при отображении, задаваемом частично рекурсивной функцией; область определения частично рекурсивной функции одного переменного.

Множество называется *рекурсивным*, если и оно и его дополнение рекурсивно перечислимы [20]. Существуют рекурсивно перечислимые, но не рекурсивные множества [20].

Характеристической функцией множества называется функция, принимающая значение 1 на множестве и 0 вне его. Характеристическая функция рекурсивного множества есть функция общерекурсивная [20]. Обратно, если характеристическая функция множества общерекурсивная, то само множество рекурсивно [20].

Если идентифицировать понятие всюду определенной алгоритмически вычислимой функции с понятием общерекурсивной функции, то рекурсивно перечислимые множества суть те множества, которые можно развернуть в алгоритмически вычислимую последовательность (быть может, с повторениями), а рекурсивные множе-

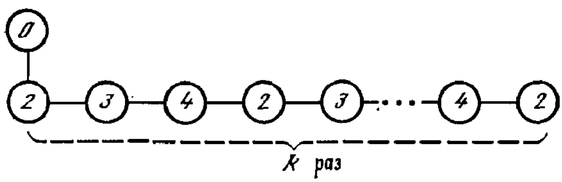
---

[1] Уже формула (3) из § 3 показывает, что любую частично рекурсивную функцию можно образовать из двух примитивно рекурсивных применением операторов наименьшего числа и суперпозиции.

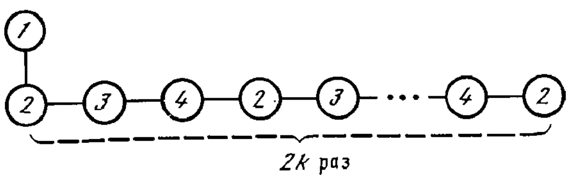
ства суть такие множества, для каждого из которых существу(алгоритм, позволяющий узнавать, принадлежит данное натурал ное число n этому множеству или нет.

Приложение II
ПРИМЕР АЛГОРИТМА

Приведем пример алгоритма «удвоения», перерабатывающего сякий комплекс A_k



в комплекс B_{2k}



Заметим, что наиболее простым алгоритмом (в интуитивном смысле), удваивающим линейно упорядоченный ряд точек, является следующая процедура: надо просматривать по очереди все точки и вместо каждой точки ставить две. Если эту процедуру формализовать, то получится как раз построенный ниже алгоритм.

Алгоритм, который мы будем строить, принадлежит типу $\mathrm{A}(\mathfrak{T}_5)$, а точнее $\mathrm{A}_4(\mathfrak{T}_5)$, где $\mathfrak{T}_5 = \{T_0, T_1, T_2, T_3, T_4, T_5\}$.

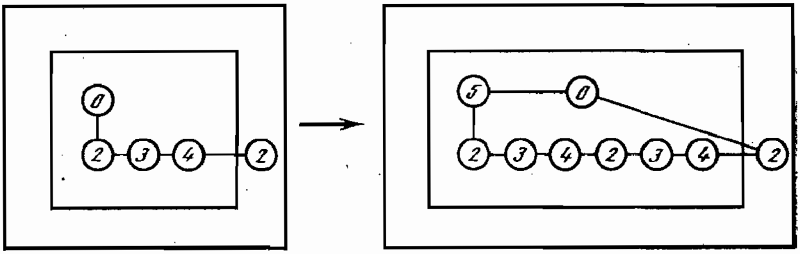
Алгоритм задается правилами непосредственной переработки, состоящими из конечного набора пар $U_i \to W_i$.

Каждую такую пары условимся графически изображать следующим образом [1]. Комплексы U_i и W_i будем чертить непересекающимися. Комплекс U_i будем помещать слева, а комплекс W_i — справа. В комплексе U_i мы выделим границу $L(U_i)$, заключив ее в рамку, состоящую из двух концентрических прямоугольников. Двумя другими (правыми) концентрическими прямоугольниками, конгруэнтными первым (левым), соответствующий под-

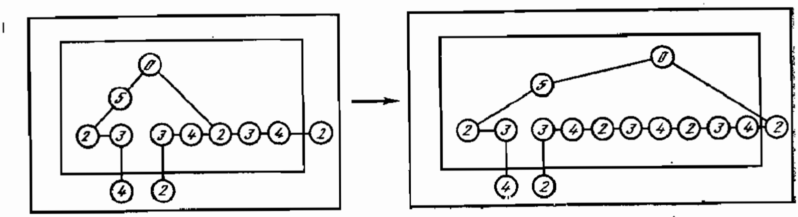
[1] Всякий комплекс изображается на чертеже с точностью до изоморфизма.

комплекс $\tilde{L}(W_i)$ выделяется в W_i. Чтобы получить изоморфное соответствие φ_i между $L(U_i)$ и $\tilde{L}(W_i)$, надо совместить правую систему прямоугольников с левой и отождествить совпавшие вершины, заключенные между прямоугольниками. После этих соглашений выпишем следующие пять правил, задающие алгоритм удвоения.

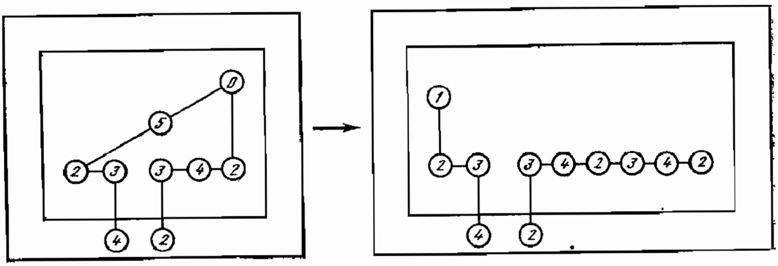
Правило 1. $U_1 \to W_1$.



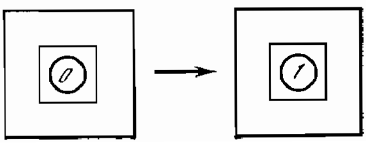
Правило 2. $U_2 \to W_2$.



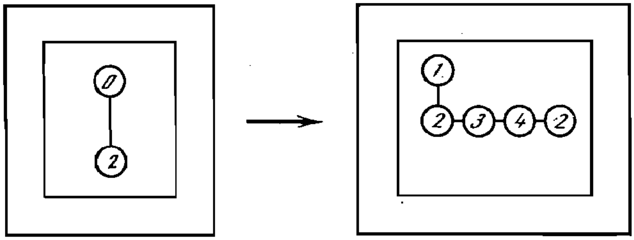
Правило 3. $U_3 \to W_3$.



Правило 4. $U_4 \to W_4$.

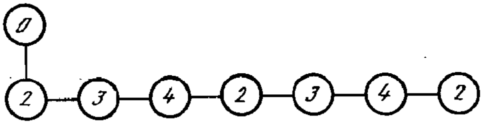


Правило 5. $U_5 \to W_5$.

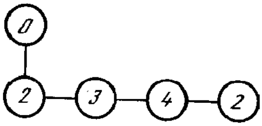


Потребность в применении правила 4 возникает тогда, когда $k = 0$, в применении правила 5 — когда $k = 1$.

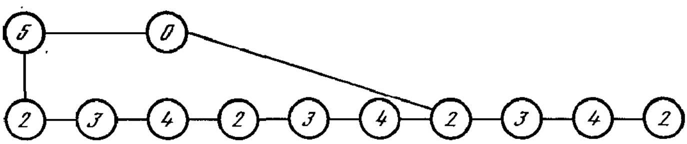
Рассмотрим в качестве примера процесс переработки построенным алгоритмом комплекса $S^0 = A_3$:



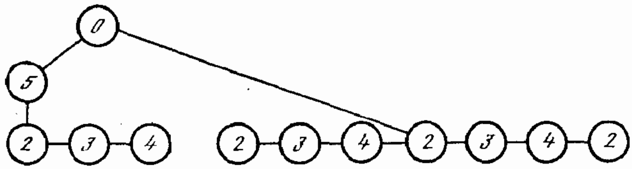
По определению к S^0 следует применить оператор непосредственной переработки Ω, заданный правилами 1—5. Для этого в S^0 выделяется комплекс $U(S^0)$. Так как $N = 4$, то комплекс $U(S^0)$ имеет вид



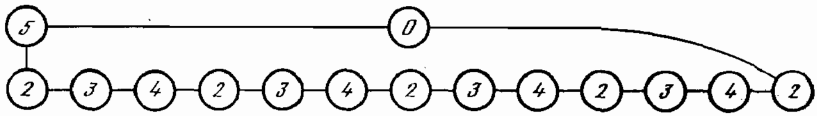
Затем среди правил 1—5 ищется та пара, у которой первый член изоморфен $U(S^0)$. Такой парой является правило 1. Производя непосредственную переработку согласно этому правилу, получаем комплекс $S^1 = \Omega(S^0)$ вида



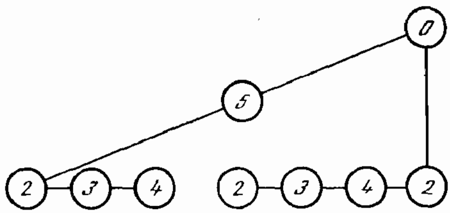
Комплекс $U(S^1)$ имеет строение



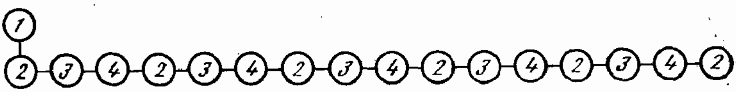
Он изоморфен первому члену U_2 правила 2, поэтому комплекс $S^2 = \Omega(S^1)$ есть следующий:



В комплексе S^2 снова выделяется комплекс $U(S^2)$; этот последний имеет вид



Он изоморфен первому члену U_3 правила 3, применяя которое получаем «решение» в виде комплекса $B_6 = S^3 = \Omega(S^2)$, имеющего строение



Заметим, что при сделанных соглашениях относительно графического изображения правил непосредственной переработки сама непосредственная переработка получает наглядное геометрическое истолкование. Проследим это на нашем примере. Мы видели, что $U(S^2)$ изоморфен U_3. Начертим комплекс S^2 так, чтобы подкомплекс $U(S^2)$ оказался конгруэнтен изображению U_3 (на чертеже, соответствующем правилу 3), и выделим, так же как это сделано для U_3, подкомплексы $U(S^2)$ и $L(S^2)$ прямоугольными рамками (рис. 1).

Тогда комплекс $\Omega(S^2)$ получается следующим простым способом: система прямоугольников, обрамляющая W_3, налагается на конгруэнтную ей систему прямоугольников, построенную для S^2;

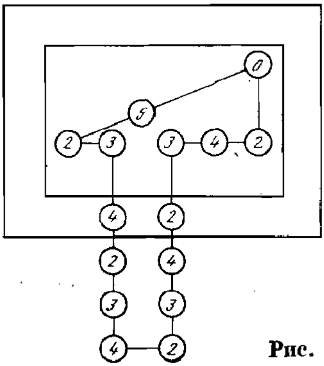


Рис. 1

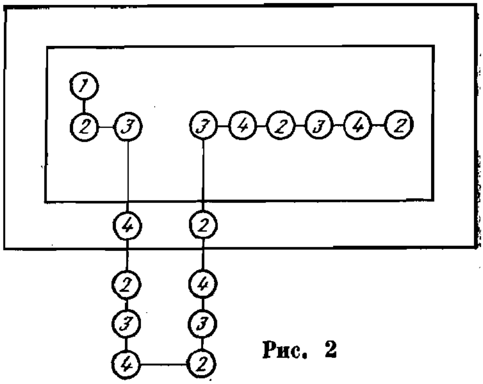


Рис. 2

совпавшие вершины, расположенные между прямоугольниками, отождествляются; все то, что лежит внутри самого внутреннего из прямоугольников, построенных для S^2, заменяется на то, что лежит внутри самого внутреннего из прямоугольников, обрамляющих W_3. Действительно, если произвести все эти операции, получится комплекс, изображенный на рис. 2.

Это и есть (с точностью изоморфизма) $S^3 = \Omega\ (S^2)$.

ЛИТЕРАТУРА

1. *Марков А. А.* Теория алгорифмов.— Тр. МИАН СССР, 1951, т. 38, с. 176—189.
2. *Марков А. А.* Теория алгорифмов.— Тр. МИАН СССР, 1954, т. 42, с. 3—375.
3. *Колмогоров А. Н.* О понятии алгоритма.— УМН, 1953, т. 8, вып. 4, с. 175—176.
4. *Gödel K.* On undecidable propositions of formal mathematical systems. Princeton, 1934.
5. *Church A.* An unsolvable problem of elementary number theory.— Amer. J. Math., 1936, vol. 58, N 2, p. 345—363.
6. *Клини С. К.* Введение в метаматематику. М.: Изд-во иностр. лит., 1957.
7. *Church A.* The calculi of lambda-conversion. Princeton, 1941.
8. *Kleene S. C.* On notation for ordinal numbers.— J. Symbol. Log., 1938, vol. 3, N 4, p. 150—155.
9. *Gödel K.* Über die Länge von Beweisen.— Ergeb. math. Kolloq. (1934—1935), 1936, H. 7, S. 23—24.
10. *Kleene S. C.* General recursive functions of natural numbers.— Math. Ann., 1936, Bd. 112, H. 5, S. 727—742.
11. *Turing A. M.* On computable numbers, with an application to the Entscheidungsproblem.— Proc. London Math. Soc. Ser. 2, 1936, vol. 42, N 3—4, p. 230—265.
12. *Петер Р.* Рекурсивные функции. М.: Изд-во иностр. лит., 1954.
13. *Post E. L.* Finite combinatory processes — formulation 1.— J. Symbol. Log., 1936, vol. 1, N 3, p. 103—105. [Рус.пер.: *Пост Э. Л.* Финитные комбинаторные процессы, формулировка 1.— В кн.: *Успенский В. А.* Машина Поста. М.: Наука, 1979, с. 89—95.]
14. *Kleene S. C.* λ-definability and recursiveness.— Duke Math. J., 1936, vol. 2, N 2, p. 340—353.

15. *Turing A. M.* Computability and λ-definability.— J. Symbol. Log., 1937, vol. 2, N 4, p. 153—163.
16. *Детловс В. К.* Нормальные алгорифмы и рекурсивные функции.— ДАН СССР, 1953, т. 90, № 5, с. 723—725.
17. *Успенский В. А.* К теореме о равномерной непрерывности.— УМН, 1957, т. 12, вып. 1(73), с. 99—142.
18. *Robinson R. M.* Primitive recursive functions.— Bull. Amer. Math. Soc., 1947, vol. 53, N 10, p. 925—942.
19. *Kleene S. C.* Recursive predicates and quantifiers.— Trans. Amer. Math. Soc., 1943, vol. 53, N 1, p. 41—73.
20. *Post E. L.* Recursively enumerable sets of positive integers and their decision problem.— Bull. Amer. Math. Soc., 1944, vol. 50, N 5, p. 284—316.
21. *Rosser J. B.* Extensions of some theorems of Gödel and Church.— J. Symbol. Log., 1936, vol. 1, N 3, p. 87—91.
22. *Post E. L.* Formal reduction of the general combinatorial decision problem.— Amer. J. Math., 1943, vol. 65, N 2, p. 197—215.

7

ε-ЭНТРОПИЯ И ε-ЕМКОСТЬ МНОЖЕСТВ В ФУНКЦИОНАЛЬНЫХ ПРОСТРАНСТВАХ*

Совместно с В. М. Тихомировым

ВВЕДЕНИЕ

Статья посвящена в основном систематическому изложению результатов, которые были в 1954—1958 гг. опубликованы К. И. Бабенко [1], А. Г. Витушкиным [2, 3], В. Д. Ерохиным [4], А. Н. Колмогоровым [5, 6], В. М. Тихомировым [7]. Естественно, что при систематизации материала были доказаны некоторые новые теоремы и просчитаны более детально некоторые примеры. Включенные в обзор не публиковавшиеся ранее результаты, выходящие за рамки подобной систематизации, принадлежат В. И. Арнольду (§ 6) и В. М. Тихомирову (§ 4, 7 и 8).

С основным направлением излагаемых исследований можно познакомиться, прочитав § 1, просмотрев примеры § 2 и прочитав § 3, который имеет характер введения к следующей за ним части статьи. Мысль о возможности характеризовать «массивность» множеств в метрических пространствах при помощи порядка роста числа элементов их наиболее экономных покрытий при $\varepsilon \to 0$ была развита в работе Л. С. Понтрягина и Л. Г. Шнирельмана (см. [8], дополнение к переводу). Из более ранних близких (но не приспособленных непосредственно для наших целей) построений

* УМН, 1959, т. 14, вып. 2, с. 3—86.

следует указать на определение мер дробных порядков в известной работе Ф. Хаусдорфа [9]. Интерес к этому кругу идей был вызван в Москве вновь, когда А. Г. Витушкин [2] получил оценку снизу (в обозначениях нашего § 1) функций $\mathscr{M}_\varepsilon(A)$ для классов функций от n переменных с ограниченными частными производными вплоть до заданного порядка p и применил эту оценку к доказательству теоремы о неизбежном понижении гладкости при представлении произвольной функции n переменных суперпозициями функций $m < n$ переменных. А. Н. Колмогоров [5] показал, что вторая часть доказательства теоремы А. Г. Витушкина, выполненная автором при помощи теории многомерных вариаций, может быть сделана очень простой при помощи оценки сверху $\mathscr{M}_\varepsilon(A)$ (см. Добавление I к нашей статье). Исходя из этой работы и сделавшихся к тому времени популярными общих идей теории информации, А. Н. Колмогоров [6] сформулировал общую программу исследования ε-энтропии и ε-емкости, интересных с точки зрения теории функций компактов в функциональных пространствах. Связь всего излагаемого направления исследований с вероятностной теорией информации имеет в настоящее время лишь характер аналогий и параллелизма (см. Добавление II). Например, результаты В. М. Тихомирова, излагаемые в § 8, вдохновлены «теоремой Котельникова» из теории информации.

Представляется несомненным, что полученные результаты могут иметь интерес в невероятностной теории информации при исследовании необходимого объема памяти и числа операций в вычислительных алгорифмах (см. работы Н. С. Бахвалова [10, 11] и А. Г. Витушкина [12]). Однако для получения практически применимых результатов известные в настоящее время оценки функций $\mathscr{E}_\varepsilon$ и $\mathscr{H}_\varepsilon$ должны быть значительно уточнены. Наш § 2 содержит некоторые еще весьма несовершенные попытки в этом направлении. В указанной ранее работе Л. С. Понтрягина и Л. Г. Шнирельмана было установлено, что асимптотика $\mathscr{M}_\varepsilon$ и $\mathscr{N}_\varepsilon$ может приводить и к топологическим инвариантам (к определению топологической размерности). В заметке А. Н. Колмогорова [13] показано, как на аналогичном пути получить некоторое определение линейной размерности топологических векторных пространств.

Нам кажется, что связи излагаемого направления исследований с различными отделами математики интересны и разнообразны. Отметим, в частности, что В. Д. Ерохин [14], решая задачу уточнения оценок функций $\mathscr{E}_\varepsilon$ и $\mathscr{H}_\varepsilon$ для некоторых классов аналитических функций, открыл новый метод приближения аналитических функций линейными формами $c_1\varphi_1(z) + \ldots + c_n\varphi_n(z)$, обладающий интересными свойствами, которые могут быть сформулированы и без понятий ε-энтропии и ε-емкости.

§ 1

ОПРЕДЕЛЕНИЕ И ОСНОВНЫЕ СВОЙСТВА ФУНКЦИЙ $\mathscr{H}_\varepsilon(A)$ И $\mathscr{C}_\varepsilon(A)$

Пусть A — непустое множество в метрическом пространстве R. Введем такие определения:

Определение 1. Система γ множеств $U \subset R$ называется *ε-покрытием* множества A, если диаметр $d(U)$ любого $U \in \gamma$ не превосходит 2ε и

$$A \subseteq \bigcup_{U \in \gamma} U.$$

Определение 2. Множество $U \subset R$ называется *ε-сетью* для множества A, если любая точка множества A находится на расстоянии, не превышающем ε, от некоторой точки из U.

Определение 3. Множество $U \subset R$ называется *ε-различимым*, если любые две его различные точки лежат на расстоянии, большем ε.

Далее выяснится, почему в определении 1 нам удобно поставить 2ε вместо обычного ε. Число ε далее без особых оговорок считается положительным.

Мы все время будем иметь дело с вполне ограниченными множествами. Целесообразно иметь в виду их три равносильных определения, заключенные в следующей теореме:

Теорема I. *Следующие три свойства множества A равносильны и зависят от метрики самого A (т. е. имеют место или не имеют места при включении A с данной метрикой в любое объемлющее пространство R):*

α) При любом ε существует конечное ε-покрытие множества A.

β) При любом ε существует конечная ε-сеть для множества A.

γ) При любом ε любое ε-различимое множество конечно.

Доказательство теоремы I может быть предоставлено читателю (см. [15, с. 312—320], где за определение взято свойство β)). Как известно, вполне ограниченными являются все *компакты*, а в предположении, что пространство полно, для полной ограниченности A необходимо и достаточно, чтобы замыкание A в R было компактом (см. [15, с. 315]).

Для вполне ограниченных A естественно ввести следующие три функции, характеризующие в некотором смысле «массивность» множества A[1]:

$\mathscr{N}_\varepsilon(A)$ — минимальное число множеств в ε-покрытии A;

$\mathscr{N}_\varepsilon^R(A)$ — минимальное число точек в ε-сети для множества A;

$\mathscr{M}_\varepsilon(A)$ — максимальное число точек в ε-различимом подмножестве множества A.

[1] При желании распространить определения на случай, когда A не вполне ограничено, их надо несколько видоизменить (см. [6]). Естественно, что для не вполне ограниченного A при достаточно малом ε все три функции $\mathscr{N}_\varepsilon$, $\mathscr{N}_\varepsilon^R$ и $\mathscr{M}_\varepsilon$ получат бесконечные значения.

Функции $\mathcal{N}_\varepsilon(A)$ и $\mathcal{M}_\varepsilon(A)$ при заданной метрике на A не зависят от выбора объемлющего пространства R (для $\mathcal{M}_\varepsilon$ это очевидно по самому определению, а для $\mathcal{N}_\varepsilon$ доказывается без труда). Наоборот, функция $\mathcal{N}_\varepsilon^R(A)$, вообще говоря, зависит от R, что и нашло отражение в ее обозначении.

Специальные названия мы присвоим двоичным логарифмам определенных выше функций [2]:

$\mathcal{H}_\varepsilon(A) = \log_2 \mathcal{N}_\varepsilon(A)$ — минимальная *ε-энтропия множества* A, или просто *ε-энтропия* множества A;

$\mathcal{H}_\varepsilon^R(A) = \log_2 \mathcal{N}_\varepsilon^R(A)$ — *ε-энтропия* A относительно R;

$\mathcal{E}_\varepsilon(A) = \log \mathcal{M}_\varepsilon(A)$ — *ε-емкость* A [3].

Названия эти связаны с представлениями теории информации. Для их пояснения достаточно следующих несколько приблизительных указаний: а) в теории информации единицей «количества информации» является количество информации в одном двоичном знаке (т. е. в указании, что он равен 0 или 1); б) «энтропией» запаса возможных «сообщений», подлежащих сохранению или передаче с определенной точностью, называется число двоичных знаков, необходимое для того, чтобы передать любое из этих сообщений с заданной точностью (т. е. такое h, что каждому сообщению x можно поставить в соответствие последовательность $s(x)$ из h двоичных знаков, по которой сообщение x может быть восстановлено с нужной точностью); в) «емкостью» передающего или запоминающего устройства называется число двоичных знаков, которое оно способно надежным образом передать или сохранить.

Если рассматривать A как множество возможных сообщений и считать, что указание x' с расстоянием $\rho(x, x') \leqslant \varepsilon$ позволяет с достаточной точностью восстановить сообщение x, то очевидно, что достаточно употреблять $\mathcal{N}_\varepsilon^R(A)$ различных «сигналов» для передачи любого из сообщений $x \in A$; этими сигналами могут служить последовательности двоичных знаков длины не более $\mathcal{H}_\varepsilon^R(A) + 1$. Обойтись же последовательностями длины меньше $\mathcal{H}_\varepsilon^R(A)$, очевидно, уже невозможно.

С другой стороны, если рассматривать A как множество возможных сигналов и считать, что два сигнала $x_1 \in A$ и $x_2 \in A$ надежным образом различимы в случае $\rho(x_1, x_2) > \varepsilon$, то очевидно, что в пределах A можно выделить $\mathcal{M}_\varepsilon(A)$ надежно различимых сигналов и при их помощи фиксировать для сохранения или передачи любую двоичную последовательность длины $\mathcal{E}_\varepsilon(A) - 1$.

---

[2] Так как A по предположению непусто, то всегда $\mathcal{N}_\varepsilon(A) \geqslant 1$ $\mathcal{N}_\varepsilon^R(A) \geqslant 1$, $\mathcal{M}_\varepsilon(A) \geqslant 1$, и поэтому $\mathcal{H}_\varepsilon(A) \geqslant 0$, $\mathcal{H}_\varepsilon^R(A) \geqslant 0$, $\mathcal{E}_\varepsilon(A) \geqslant 0$.

[3] $\log N$ всюду далее обозначает двоичный логарифм числа N.

Выделить же в A систему надежно различимых сигналов для передачи любых двоичных последовательностей длины $\mathscr{E}_\varepsilon(A)+1$, очевидно, нельзя.

Что касается функции $\mathscr{H}_\varepsilon(A)$, то роль ее такова: 1) в силу неравенства $\mathscr{H}_\varepsilon(A) \leqslant \mathscr{H}_\varepsilon^R(A)$ (см. далее теорему IV) она служит для оценки $\mathscr{H}_\varepsilon^R(A)$ снизу; 2) в случае, если пространство *центрируемо* (см. далее определение 4) $\mathscr{H}_\varepsilon^R(A) = \mathscr{H}_\varepsilon(A)$ и $\mathscr{H}_\varepsilon(A)$ получает непосредственный смысл энтропии. Любое метрическое вполне ограниченное пространство A может быть погружено в центрируемое пространство R. Поэтому, по идее А. Г. Витушкина, в известном смысле слова $\mathscr{H}_\varepsilon(A)$ можно рассматривать как энтропию A по преимуществу («абсолютную энтропию»). Практический смысл этого утверждения, впрочем, не вполне ясен, кроме тех случаев, когда реально требуемая точность воспроизведения $x \in A$ действительно заключается в указании x' с $\rho(x, x') \leqslant \varepsilon$ из некоторого центрируемого пространства R, как это бывает в случае задач на приближенное задание действительных функций $f(t)$ произвольного аргумента t с определенной равномерной точностью ε. Обо всем этом см. далее, в связи с теоремами VI, VII.

Сформулируем несколько простых теорем, выражающих основные общие свойства введенных функций.

Теорема II. *Все шесть функций $\mathscr{N}_\varepsilon(A)$, $\mathscr{N}_\varepsilon^R(A)$, $\mathscr{M}_\varepsilon(A)$, $\mathscr{H}_\varepsilon(A)$, $\mathscr{H}_\varepsilon^R(A)$, $\mathscr{E}_\varepsilon(A)$, как функции множества A, полуаддитивны, т. е. для них из*

$$A \subset \bigcup_k A_k$$

вытекает

$$F(A) \leqslant \sum_k F(A_k).$$

Доказательство может быть предоставлено читателю. Из полуаддитивности и неотрицательности наших функций вытекает, что в случае $A' \subset A$ для каждой из них справедливо неравенство $F(A') \leqslant F(A)$.

Теорема III. *Все шесть функций $\mathscr{N}_\varepsilon(A)$, $\mathscr{N}_\varepsilon^R(A)$, $\mathscr{M}_\varepsilon(A)$, $\mathscr{H}_\varepsilon(A)$, $\mathscr{H}_\varepsilon^R(A)$, $\mathscr{E}_\varepsilon(A)$, как функции от ε, являются невозрастающими (при возрастании ε, т. е. неубывающими при убывании ε). Функции $\mathscr{N}_\varepsilon(A)$, $\mathscr{M}_\varepsilon(A)$, $\mathscr{H}_\varepsilon(A)$, $\mathscr{E}_\varepsilon(A)$ непрерывны справа.*

Первая половина теоремы непосредственно вытекает из определений, так что проведение доказательства может быть предоставлено читателю.

Вторую часть теоремы достаточно доказать для функций $\mathscr{M}_\varepsilon(A)$ и $\mathscr{N}_\varepsilon(A)$. Пусть $\mathscr{M}_{\varepsilon_0}(A) = n$ и $x_1, \ldots, x_n$ — соответствующее

ε_0-различимое множество. Тогда

$$\varepsilon_1 = \min_{i \neq j} \rho(x_i, x_j) > \varepsilon_0$$

и для всех ε в пределах $\varepsilon_0 \leqslant \varepsilon < \varepsilon_1$ должно быть $\mathcal{M}_\varepsilon(A) \geqslant n$; в силу же монотонности $\mathcal{M}_\varepsilon(A) = n$.

Для $\mathcal{N}_\varepsilon(A)$ доказательство несколько сложнее и проводится от противного. Если допустить существование последовательности

$$\varepsilon_1 > \varepsilon_2 > \ldots > \varepsilon_k > \ldots \to \varepsilon_0,$$

по которой $\mathcal{N}_{\varepsilon_k}(A) = m < \mathcal{N}_{\varepsilon_0}(A) = n$, то при любом $k \geqslant 1$ должны существовать ε_k-покрытия A множествами $A_{k1}, A_{k2}, \ldots, A_{km}$. Замыкание $\bar{A}$ множества A в пополнении R^* пространства R компактно.

В метрике «отклонений» $\alpha(F, F')$ замкнутые подмножества компакта $\bar{A}$ образуют компакт (см. [16, с. 548—550]).

Рассмотрим множество

$$F_{ki} = \bar{A}_{ki} \cap \bar{A}.$$

В силу сказанного можно выбрать такую последовательность $k_s \to \infty$, что

$$F_{k_s i} \to F_i \quad (i = 1, 2, \ldots, m).$$

Легко проверить, что $F_1, \ldots, F_m$ образуют ε_0-покрытие множества A, т. е. вопреки допущению $n \geqslant m$.

Теорема IV. *Для всякого вполне ограниченного множества A в метрическом пространстве R выполнены следующие неравенства:*

$$\mathcal{M}_{2\varepsilon}(A) \leqslant \mathcal{N}_\varepsilon(A) \leqslant \mathcal{N}_\varepsilon^R(A) \leqslant \mathcal{N}_\varepsilon^A(A) \leqslant \mathcal{M}_\varepsilon(A), \tag{1}$$

а следовательно,

$$\mathcal{E}_{2\varepsilon}(A) \leqslant \mathcal{H}_\varepsilon(A) \leqslant \mathcal{H}_\varepsilon^R(A) \leqslant \mathcal{H}_\varepsilon^A(A) \leqslant \mathcal{E}_\varepsilon(A). \tag{2}$$

Будем доказывать неравенства (1) справа налево. Пусть $x_1, \ldots, x_{\mathcal{M}_\varepsilon}$ — максимальное ε-различимое в A множество. Тогда оно, очевидно, есть ε-сеть, так как в противном случае существовала бы точка $x' \in A$ такая, что $\rho(x', x) > \varepsilon$; последнее противоречило бы максимальности $x_1, \ldots, x_{\mathcal{M}_\varepsilon}$. Ввиду того что $x_i \in A$ по определению, получаем

$$\mathcal{N}_\varepsilon^A(A) \leqslant \mathcal{M}_\varepsilon(A).$$

Очевидно, что всякая ε-сеть, состоящая из точек $x_i \in A$, является ε-сетью, составленной из точек $x_i \in R \supset A$, т. е.

$$\mathcal{N}_\varepsilon^R(A) \leqslant \mathcal{N}_\varepsilon^A(A).$$

Пусть $y_1, \ldots, y_n$ есть ε-сеть по отношению к A в R. Рассмотрим множества $S_\varepsilon(y_i) \cap A = U_i$, где через $S_\varepsilon(y_i)$ мы обозначили шар радиуса ε с центром в точке y_i. $\{U_i\}$ является ε-покрытием A, т. е. всякая ε-сеть порождает ε-покрытие, откуда

$$\mathcal{N}_\varepsilon(A) \leqslant \mathcal{N}_\varepsilon^R(A).$$

И, наконец, для любого ε-покрытия и любого 2ε-различимого множества число точек в последнем не превышает числа точек в первом, так как в противном случае две точки, расстояние между которыми $> 2\varepsilon$, поместились бы в одно множество диаметра $\leqslant 2\varepsilon$; отсюда

$$\mathcal{M}_{2\varepsilon}(A) \leqslant \mathcal{N}_\varepsilon(A).$$

Как уже говорилось, взаимоотношения между введенными функциями упрощается в случае центрируемого пространства.

Определение 4. Пространство R называется *центрируемым*, если в нем для любого множества U диаметра $d = 2r$ существует точка x_0, от которой любая точка x находится на расстоянии, не большем r.

Теорема V. *В центрируемом пространстве R для любого вполне ограниченного множества A*

$$\mathcal{N}_\varepsilon^R(A) = \mathcal{N}_\varepsilon(A), \qquad \mathcal{H}_\varepsilon^R(A) = \mathcal{H}_\varepsilon(A).$$

Действительно, неравенство

$$\mathcal{N}_\varepsilon(A) \leqslant \mathcal{N}_\varepsilon^R(A)$$

входит в формулировку теоремы IV. В центрируемом же пространстве каждому ε-покрытию A множествами

$$U_1, \ldots, U_N$$

соответствует ε-сеть из того же числа точек

$$x_0^{(1)}, \ldots, x_0^{(N)}$$

и потому

$$\mathcal{N}_\varepsilon^R(A) \leqslant \mathcal{N}_\varepsilon(A).$$

Теорема VI. *Пространство D^X ограниченных функций на произвольном множестве X с метрикой*

$$\rho(f, g) = \sup_{x \in X} |f(x) - g(x)|$$

центрируемо.

Пусть $U \subset D^X$ — множество диаметра

$$d = \sup_{f, g} \sup_{x \in X} |f(x) - g(x)|.$$

Положив

$$\bar f(x) = \sup_{f \in U} f(x), \qquad \underline{f}(x) = \inf_{f \in U} f(x),$$

можно видеть, что

$$d = \sup_{x \in X} (\bar f(x) - \underline{f}(x)).$$

Теперь легко показать, что для функции

$$f_0(x) = {}^1/_2 (\bar f(x) - \underline{f}(x))$$

и любой функции $f \in U$ при любом $x \in X$

$$|f(x) - f_0(x)| \leqslant d/2,$$

т. е.

$$\rho(f, f_0) \leqslant d/2 \qquad \forall f \in U,$$

что и требовалось доказать.

Теорема VII (А. Г. Витушкин [3]). *Любое вполне ограниченное пространство A может быть вложено в центрируемое пространство R.*

В силу теоремы Мазура—Банаха [17, гл. IX] любое A (нас интересуют только вполне ограниченные, т. е. во всяком случае сепарабельные A) может быть изометрически вмещено в пространство C, а следовательно, и в содержащее его пространство D^I, где I — единичный отрезок [0, 1]; теорема доказана.

По теореме VI пространство D^I центрируемо, а по теореме V для любого расположенного в нем A

$$\mathcal{N}_\varepsilon^{D^I}(A) = \mathcal{N}_\varepsilon(A).$$

В силу неравенства

$$\mathcal{N}_\varepsilon(A) \leqslant \mathcal{N}_\varepsilon^R(A)$$

(теорема IV) и теоремы VII

$$\mathcal{N}_\varepsilon(A) = \min_{A \subseteq R} \mathcal{N}_\varepsilon^R(A), \qquad \mathcal{H}_\varepsilon(A) = \min_{A \subseteq R} \mathcal{H}_\varepsilon^R(A),$$

что оправдывает наименование $\mathcal{H}_\varepsilon(A)$ *минимальной* ε-энтропией.

§ 2

ПРИМЕРЫ ТОЧНОГО ВЫЧИСЛЕНИЯ ФУНКЦИЙ $\mathcal{H}_\varepsilon(A)$ И $\mathcal{E}_\varepsilon(A)$ И ИХ ОЦЕНКИ В НЕКОТОРЫХ ПРОСТЫХ СЛУЧАЯХ

1. A — отрезок Δ: $\{a \leqslant x \leqslant b\}$ длины $|\Delta| = b - a$ на прямой D с метрикой $\rho(x, x') = |x - x'|$.

В этом случае [4]

$$\mathcal{N}_\varepsilon(A)=\mathcal{N}_\varepsilon^D(A)=\mathcal{M}_{2\varepsilon}(A)=\begin{cases}|\Delta|/2\varepsilon & \text{при } |\Delta|/2\varepsilon \text{ целом},\\ [|\Delta|/2\varepsilon]+1 & \text{при } |\Delta|/2\varepsilon \text{ нецелом},\end{cases} \tag{3}$$

т. е.

$$\begin{aligned}\mathcal{N}_\varepsilon(A)&=\frac{|\Delta|}{2\varepsilon}+O(1), & \mathcal{M}_\varepsilon(A)&=\frac{|\Delta|}{\varepsilon}+O(1),\\ \mathcal{H}_\varepsilon(A)&=\log\frac{|\Delta|}{2\varepsilon}+O(\varepsilon), & \mathcal{E}_\varepsilon(A)&=\log\frac{|\Delta|}{\varepsilon}+O(\varepsilon).\end{aligned} \tag{4}$$

Действительно, легко видеть, что Δ является *центрированным* в смысле § 1 и поэтому в согласии с теоремой V

$$\mathcal{N}_\varepsilon^D(\Delta)=\mathcal{N}_\varepsilon(\Delta). \tag{5}$$

Число множеств ε-покрытия Δ не может быть меньшим, чем $[|\Delta|/2\varepsilon]$, или равняться $[|\Delta|/2\varepsilon]$ в случае, если $|\Delta|/2\varepsilon$ — нецелое число, так как в этом случае совокупность этих множеств имела бы меру, не большую чем $2\varepsilon[|\Delta|/2\varepsilon]<|\Delta|$.

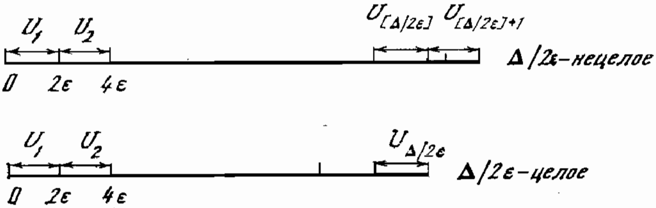


Рис. 1

Однако читатель без труда убедится в том, что из $[|\Delta|/2\varepsilon]+1$ множеств всегда можно устроить ε-покрытие, причем при $|\Delta|/2\varepsilon$ целом оказывается достаточным $|\Delta|/2\varepsilon$ множеств (рис. 1).

С другой стороны, при $|\Delta|/2\varepsilon$ нецелом мы, разделив Δ на $[|\Delta|/2\varepsilon]$ равных частей точками $a=x_1, x_2, \ldots, x_{[|\Delta|/2\varepsilon]+1}$, получим $[|\Delta|/2\varepsilon]+1$ точек, образующих 2ε-различимое множество. Если же $|\Delta|/2\varepsilon$ — целое, мы делим Δ подобным образом на $|\Delta|/2\varepsilon-1$ равных частей и получаем $|\Delta|/2\varepsilon$ точек, образующих 2ε-различимое множество. Из сказанного вытекает, что

$$\mathcal{M}_{2\varepsilon}(\Delta)\geqslant\mathcal{N}_\varepsilon(\Delta),$$

откуда, использовав (5) и неравенство (1), получим (3).

2. A — множество $F_1^\Delta(L)$ функций $f(x)$, заданных на отрезке $\Delta=[a,b]$, удовлетворяющих условию Липшица

$$|f(x)-f(x')|\leqslant L|x-x'|$$

и обращающихся в нуль в точке a.

[4] Знак $[A]$ означает целую часть числа A.

Данное множество мы рассмотрим в пространстве D^{Δ}, т. е. с метрикой

$$\rho(f(x), g(x)) = \sup_{x \in \Delta} |f(x) - g(x)|.$$

Мы покажем, что

$$\mathscr{H}_\varepsilon(A) = \mathscr{E}_{2\varepsilon}(A) = \mathscr{H}_\varepsilon^{D^\Delta}(A) = \begin{cases} \dfrac{|\Delta| L}{\varepsilon} - 1 & \text{при } \dfrac{|\Delta| L}{\varepsilon} \text{ целом}, \\ \left[\dfrac{|\Delta| L}{\varepsilon}\right] & \text{при } \dfrac{|\Delta| L}{\varepsilon} \text{ нецелом}, \end{cases} \tag{6}$$

т. е.

$$\mathscr{H}_\varepsilon(A) = \frac{|\Delta| L}{\varepsilon} + O(1), \qquad \mathscr{E}_\varepsilon(A) = \frac{2|\Delta| L}{\varepsilon} + O(1). \tag{7}$$

Отметим прежде всего, что это множество заменой независимого переменного $t = L(x - a)$ взаимно однозначно и изометрически отображается на множество $F_1^{\Delta'}(1)$, т. е. на множество функций, заданных на отрезке $[0, \Delta']$, где $\Delta' = |\Delta| L$, удовлетворяющих условию Липшица с константой, равной единице, и равных нулю в нуле. Очевидно, что при этом величины $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ не изменятся. В последнем множестве мы и будем вычислять функции $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$.

Идея наших построений будет заключаться в том, чтобы построить ε-покрытие $F_1^{\Delta'}(1)$ и 2ε-различимое в $F_1^{\Delta'}(1)$ множество, которые состояли бы из одного и того же числа элементов K_ε.

Тогда согласно определениям функций $\mathscr{N}_\varepsilon$ и $\mathscr{M}_{2\varepsilon}$ мы получим

$$\mathscr{N}_\varepsilon(F_1^{\Delta'}(1)) \leqslant K_\varepsilon, \qquad \mathscr{M}_{2\varepsilon}(F_1^{\Delta'}(1)) \geqslant K_\varepsilon, \tag{8}$$

откуда, использовав неравенство (1), мы аналогично п. 1 этого параграфа получим, что

$$\mathscr{M}_{2\varepsilon}(F_1^{\Delta'}(1)) = \mathscr{N}_\varepsilon(F_1^{\Delta'}(1)) = K_\varepsilon. \tag{9}$$

К равенствам (9) присоединится еще и равенство

$$\mathscr{N}_\varepsilon(F_1^{\Delta'}(1)) = \mathscr{N}_\varepsilon^{D^{\Delta'}}(F_1^{\Delta'}(1)),$$

ввиду того что пространство $D^{\Delta'}$ является центрированным.

Пусть $\varepsilon > 0$ и такое, что Δ'/ε не есть целое число. Число $[\Delta'/\varepsilon]$ обозначим через n, величины $k\varepsilon$ через t_k ($k = 1, 2, \ldots, n$). Разделим отрезок $[0, \Delta']$ на $n + 1$ отрезков:

$$\Delta_k = [(k-1)\varepsilon, k\varepsilon], \; k = 1, 2, \ldots, n, \quad \Delta_{n+1} = [n\varepsilon, |\Delta'|].$$

Обозначим через $\varphi(t)$ функцию, определенную на отрезке $[\varepsilon, |\Delta'|]$, равную ε в точке $t = \varepsilon$, а на отрезках Δ_k линейную с угловым коэффициентом, равным $+1$ или -1.

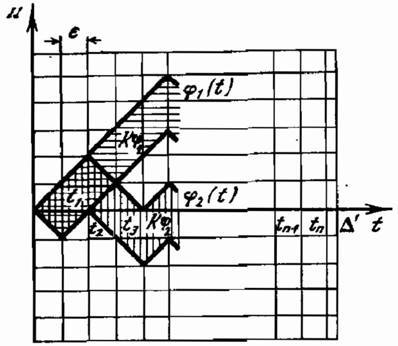


Рис. 2

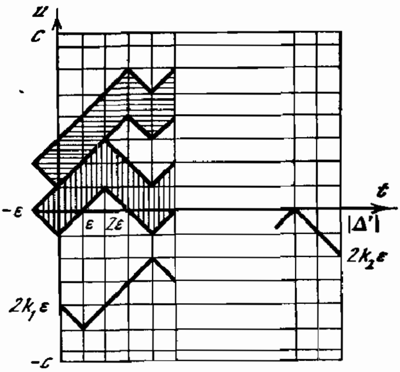


Рис. 4

Множество точек плоскости (t, u) таких, что

$$-t \leqslant u \leqslant t, \qquad t \in \Delta_1,$$

$$\varphi(t) - 2\varepsilon \leqslant u \leqslant \varphi(t), \qquad t \in [\varepsilon, |\Delta'|],$$

назовем *ε-коридором* (рис. 2) и обозначим через K_φ.

Число всех коридоров обозначим через K_ε. При нашем выборе ε K_ε, как легко видеть, равно $2^{[|\Delta'|/\varepsilon]}$.

Покажем, что совокупность всех ε-коридоров образует ε-покрытие множества $F_1^{\Delta'}(1)$. (Далее мы говорим, что функция $f(t)$ принадлежит некоторому коридору K_φ, если

$$\varphi(t) - 2\varepsilon \leqslant f(t) \leqslant \varphi(t).)$$

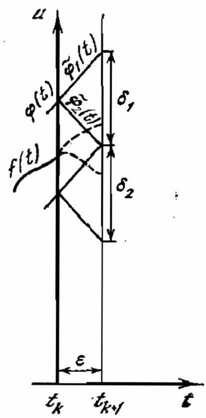


Рис. 3

Доказательство этого утверждения получим по индукции.

Действительно, на отрезке $\Delta_1 = [0, \varepsilon]$ все функции нашего множества принадлежат *всем* ε-коридорам. Пусть теперь по предположению индукции для всякой функции f из нашего множества на отрезке $[0, t_k]$ $(k \leqslant n)$ найдется ε-коридор K_φ такой, что

$$\varphi(t) - 2\varepsilon \leqslant f(t) \leqslant \varphi(t), \qquad 0 \leqslant t \leqslant t_k.$$

Тогда из условия Липшица вытекает (рис. 3), что $f(t_{k+1})$ принадлежит либо отрезку

$$\delta_1 = [\varphi(t_k) - \varepsilon, \varphi(t_k) + \varepsilon],$$

либо отрезку

$$\delta_2 = [\varphi(t_k) - 3\varepsilon, \varphi(t_k) - \varepsilon]$$

и на отрезке $[0, t_{k+1}]$ функция f принадлежит коридору $K_{\tilde\varphi}$

где $\tilde{\varphi}$ — функция, совпадающая с φ на $[0, t_k)$ и на $\Delta_k = [t_k, t_{k+1}]$ линейная с угловым коэффициентом, равным $+1$ в случае, если $f(t_{k+1}) \in \delta_1$, и -1, если $f(t_{k+1}) \in \delta_2$. Индукция закончена. Теперь нам нужно построить 2ε-различимое множество, состоящее из K_ε элементов. Делим $[0, |\Delta'|]$ на n равных отрезков $\overline{\Delta}_1, \ldots, \overline{\Delta}_n$. Длина каждого отрезка $|\overline{\Delta}_i| = |\Delta'|/n > \varepsilon$ (вспомним, что $n\ [|\Delta'|/\varepsilon]$). Рассмотрим множество M_n функций, обращающихся в нуль в нуле, а на отрезках $\overline{\Delta}_i$ линейных с угловым коэффициентом, равным ± 1. Две различные функции из M_n отличаются друг от друга по крайней мере на $2|\Delta'|/n$, т. е. они 2ε-различимы и число их равно $2^n = K_\varepsilon$.

Если $|\Delta'|/\varepsilon$ — целое, то величина K_ε, равная числу всех ε-коридоров, оказывается равной $2^{|\Delta'|/\varepsilon - 1}$. Аналогичное рассмотренному выше множество M_n, где $n = |\Delta'|/\varepsilon - 1$, как нетрудно показать, 2ε-различимо и состоит из $2^n = K_\varepsilon$ функций.

Таким образом, мы доказали, что

$$\mathcal{N}_\varepsilon(A) = \mathcal{M}_{2\varepsilon}(A) = \mathcal{N}_\varepsilon^{D\Delta}(A) =$$
$$= K_\varepsilon = \begin{cases} 2^{|\Delta'|/\varepsilon - 1} & \text{при } |\Delta'|/\varepsilon \text{ целом}, \\ 2^{[|\Delta'|/\varepsilon]} & \text{при } |\Delta'|/\varepsilon \text{ нецелом}, \end{cases}$$

откуда (6) получится логарифмированием, так как $|\Delta'| = |\Delta| L$.

3. Пусть теперь A — множество $F_1^\Delta(C, L)$ функций $f(t)$, заданных на отрезке $\Delta = [a, b]$, удовлетворяющих условию Липшица с константой L и ограниченных на Δ константой C. Это множество мы будем также рассматривать как метрическое пространство в равномерной метрике на Δ.

Пространство $F_1^\Delta(C, L)$ отличается от рассмотренного нами в предыдущем пункте пространства $F_1^\Delta(L)$ тем, что у нас появилось условие

$$|f(t)| \leqslant C, \quad t \in \Delta,$$

вместо $f(0) = 0$.

Мы получим следующие оценки:

$$\frac{|\Delta| L}{\varepsilon} + \log\frac{C}{\varepsilon} - 3 \leqslant \mathscr{E}_{2\varepsilon}(A) \leqslant \mathscr{H}_\varepsilon(A) \leqslant \frac{|\Delta| L}{\varepsilon} + \log\frac{C}{\varepsilon} + 3 \tag{10}$$

при $\varepsilon \leqslant \min(C/4, C^2/16|\Delta| L)$, т. е.

$$\mathscr{H}_\varepsilon(A) = |\Delta| L/\varepsilon + \log(C/\varepsilon) + O(1),$$
$$\mathscr{E}_{2\varepsilon}(A) = 2|\Delta| L/\varepsilon + \log(C/\varepsilon) + O(1). \tag{11}$$

Мы видим, что здесь для $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ удается получить оценки снизу и сверху, которые при не слишком больших ε расходятся между собой только на несколько единиц. Такого рода результаты можно еще рассматривать как вполне удовлетворительные с точки зрения практических применений.

Перейдем к доказательству неравенств (10).

Как и в предыдущем пункте, дело сведется к пространству $F_1^{\Delta'}(C, 1)$, где $\Delta' = [0, |\Delta| L]$.

Оценка сверху для $\mathcal{H}_\varepsilon$ получится благодаря использованию результатов предыдущего пункта. Действительно, задав $\varepsilon > 0$, рассмотрим множество точек плоскости (t, u) с координатами $(-\varepsilon, 2k\varepsilon)$, $|k| \leqslant [C/2\varepsilon]$ (рис. 4). В каждой из этих точек построим множество ε-коридоров так, как мы это делали в предыдущем пункте для нуля. Множество ε-коридоров, построенных в точке $(-\varepsilon, 2k\varepsilon)$, явится согласно предыдущему ε-покрытием для множества функций $f \in F_1^{\Delta'}(C, L)$, для которых $2k\varepsilon - \varepsilon \leqslant \leqslant f(0) \leqslant 2k\varepsilon + \varepsilon$, а следовательно, множество всех ε-коридоров является ε-покрытием для $F_1^{\Delta'}(C, L)$. Таким образом, получим

$$\mathcal{H}_\varepsilon(F_1^{\Delta'}(C, L)) \leqslant \left(2\left[\frac{C}{2\varepsilon}\right] + 1\right) \mathcal{H}_\varepsilon(F_1^{\Delta' \cup [-\varepsilon, 0]}(L)),$$

откуда, логарифмируя и используя (6), получим

$$\mathcal{H}_\varepsilon(A) \leqslant \frac{|\Delta'|}{\varepsilon} + \log\left(\frac{C}{\varepsilon} + 1\right) + 1 \leqslant \frac{|\Delta'|}{\varepsilon} + \log\frac{C}{\varepsilon} + 2, \qquad (12)$$

так как

$$\varepsilon \leqslant C \log(C/\varepsilon + 1) \leqslant \log(C/\varepsilon) + 1.$$

Оценка сверху для $\mathscr{E}_{2\varepsilon}$ получается несколько сложнее. Обозначим через $2r$ максимальное целое четное число такое, что $2r\varepsilon < < |\Delta'|$. Разделим Δ' на $2r$ равных отрезков $\Delta_1, \Delta_2, \ldots, \Delta_{2r}$. Строим множество $\overline{M}_{2r}$ функций, равных в нуле одному из значений $2k\varepsilon$, $|k| \leqslant [C/2\varepsilon]$, а на отрезках Δ_i, $i = 1, 2, \ldots, 2r$, линейных с угловым коэффициентом, равным ± 1. Функции из $\overline{M}_{2r}$ 2ε-различимы и принадлежат $F_1^{\Delta'}(1)$. Нам остается оценить число функций, принадлежащих $F_1^{\Delta'}(C, 1)$, т. е. тех, которые по модулю не превосходят C. Далее мы проводим эту оценку элементарными средствами. С точки зрения теории вероятностей дело сводится к оценке числа путей (соответствующих построенным выше функциям из $\overline{M}_{2r}$), не выходящих за заданные пределы при обычно рассматриваемом случайном блуждании с вероятностью $^1/_2$ на каждом шаге идти вверх или вниз, и нужная нам оценка могла бы быть получена гораздо быстрее благодаря использованию известных в теории вероятностей неравенств.

Проведем требуемую оценку в несколько этапов.

α) Обозначим через $N(k_1, k_2)$ число функций $\varphi(t) \in \overline{M}_{2r}$ таких, что

$$\varphi(0) = 2k_1\varepsilon, \quad \varphi(|\Delta'|) = 2k_2\varepsilon,$$

где k_1 и k_2 — любые допустимые целые числа; пусть для определенности $k_2 \geqslant k_1$; $N(k_1, k_2)$ равно числу функций φ, у которых n_1 угловых коэффициентов, равных $+1$ (число «подъемов»), минус число n_2 коэффициентов, равных -1 (число «спусков»), равно $2(k_2 - k_1)$.

Отсюда имеем

$$n_1 + n_2 = 2r, \qquad n_1 - n_2 = 2(k_2 - k_1),$$

т. е. $n_1 = r + k_2 - k_1$, а число функций с n_1 подъемами равно $C_{2r}^{n_1}$. Таким образом,

$$N(k_1, k_2) = C_{2r}^{k_2 - k_1 + r}. \tag{13}$$

β) Рассмотрим множество $U_1(k)$ функций из $\overline{M}_{2r}$ и удовлетворяющих условиям

$$\varphi(0) = 2k\varepsilon, \qquad \varphi(|\Delta'|) > 2k\varepsilon.$$

Число $N_1(k)$ таких функций удовлетворяет соотношению

$$N_1(k) = \sum_{\substack{k_1 \leqslant k \\ k_2 > k}} N(k_1, k_2) \leqslant \sum_{\substack{k_1 \leqslant k \\ k_2 > k}} C_{2r}^{k_2 - k_1 + r}. \tag{14}$$

В (14) член C_{2r}^{s+r} встречается s раз и, таким образом,

$$N_1(k) \leqslant \sum_{0 \leqslant s \leqslant r} s C_{2r}^{s+r}. \tag{15}$$

γ) Пусть $U_2(k)$ обозначает множество функций из $\overline{M}_{2r}$ таких, что

$$\varphi(0) \leqslant 2k\varepsilon, \qquad \max_{t \in \Delta'} \varphi(t) > 2k\varepsilon.$$

Пусть $\varphi(t) \in U_2(k)$, но не принадлежит $U_1(k)$. Рассмотрим последнюю из точек $s\varepsilon$, в которой $\varphi(s\varepsilon) = (k+1)\varepsilon$. Строим новую функцию $\tilde{\varphi}(t)$, которая до точки $s\varepsilon$ совпадает с $\varphi(t)$, а далее является ее зеркальным отражением в прямой $u = k + 1$. Очевидно, что $\tilde{\varphi}(t) \in U_1(k)$.

Функции $\varphi \in U_1(k) \cap U_2(k)$ ставим в соответствие ее самое. Мы получили отображение множества $U_2(k)$ в $U_1(k)$, при котором, как легко видеть, прообраз каждой функции из $U_1(k)$ состоит не более чем из *двух* функций, т. е. число функций $N_2(k)$ в $U_2(k)$ не превосходит $2N_1(k)$.

Легко понять, что число функций из $\overline{M}_{2r}$, для которых

$$|\varphi(0)| \leqslant 2k\varepsilon, \quad \max_{t \in \Delta} |\varphi(t)| > 2k\varepsilon, \tag{16}$$

не превосходит $2N_2(k) \leqslant 4N_1(k)$.

Остается только заметить, что множество функций из $\overline{M}_{2r}$, но *не принадлежащих* (16) при $k = [C/2\varepsilon]$ и есть то множество, число функций в котором нам требовалось оценить с самого начала.

Таким образом,

$$\mathscr{M}_{2\varepsilon}(F_1^{\Delta'}(C,1)) \geqslant \left(2\left[\frac{C}{2\varepsilon}\right]+1\right)2^{2r}-4N_1(k) \geqslant$$
$$\geqslant \left(2\left[\frac{C}{2\varepsilon}\right]+1\right)2^{2r}-4\sum_{s=0}^{r} sC_{2r}^{s+r}. \tag{17}$$

Проведем оценку в правой части (17). Имеем

$$\sum_{s=0}^{r} sC_{2r}^{r+s}=\left(2rC_{2r-1}^{r}-\frac{rC_{2r}^{r}}{2}\right)=\frac{r}{2}C_{2r}^{r}. \tag{18}$$

Доказательство простого равенства (18) представляется читателю.

Далее мы пользуемся элементарным неравенством

$$C_{2r}^{r}<2^{2r}/\sqrt{2r+1} \tag{19}$$

(которое содержится, например, в [18] и доказывается без труда).

Таким образом, из (17) (благодаря (18) и (19)) получим

$$\mathscr{M}_{2\varepsilon}(F_1^{\Delta'}(C,1)) \geqslant 2^{2r}\left(2\left[\frac{C}{\varepsilon}\right]+1-\frac{2r}{2r+1}\right) \geqslant$$
$$\geqslant 2^{2r}\left(\frac{C}{\varepsilon}-1-\sqrt{2r}\right). \tag{20}$$

Логарифмируя (20) и воспользовавшись тем, что $|\Delta'|-2\varepsilon \leqslant \leqslant 2r\varepsilon \leqslant |\Delta'|$, получим

$$\mathscr{E}_{2\varepsilon}(A) \geqslant \frac{|\Delta'|}{\varepsilon}-2+\log\frac{C}{\varepsilon}+\log\left(1-\frac{\varepsilon}{C}-\sqrt{\frac{\varepsilon|\Delta'|}{C^2}}\right) \geqslant$$
$$\geqslant \frac{|\Delta'|}{\varepsilon}+\log\frac{C}{\varepsilon}-3 \quad \text{при } \varepsilon \leqslant \min\left(\frac{C}{4}, \frac{C^2}{16|\Delta'|}\right),$$

что нам и требовалось доказать.

4. A — множество $A_h(C)$ функций f, заданных на D, периодических с периодом, равным 2π, аналитических в полосе ширины $h: |\operatorname{Im} z| \leqslant h$, $z=x+iy$, и ограниченных там константой C. Данное множество рассмотрим в равномерной метрике на D. Для этого множества имеют место следующие формулы:

$$\begin{aligned} \mathscr{H}_\varepsilon(A) &= 2\frac{(\log(1/\varepsilon))^2}{h\log e}+O\left(\log\frac{1}{\varepsilon}\log\log\frac{1}{\varepsilon}\right), \\ \mathscr{E}_\varepsilon(A) &= 2\frac{(\log(1/\varepsilon))^2}{h\log e}+O\left(\log\frac{1}{\varepsilon}\log\log\frac{1}{\varepsilon}\right). \end{aligned} \tag{21}$$

Соотношения (21) мы получим, как мы это делали в предыдущих пунктах, оценивая сверху число множеств, образующих

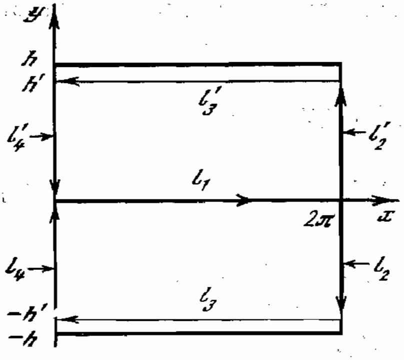


Рис. 5

ε-покрытие $A_h(C)$, и снизу число 2ε-различимых в $A_h(C)$ функций.

Для получения оценки сверху величины $\mathcal{H}_\varepsilon(A)$ разложим функцию $f \in A_h(C)$ в ряд Фурье:

$$f(x) = \sum_{k=-\infty}^{+\infty} c_k e^{ikx}. \tag{22}$$

Далее мы будем пользоваться следующими неравенствами для коэффициентов в (22):

$$|c_k| \leqslant Ce^{-|k|h}. \tag{23}$$

Неравенства (23) получаются так. Пусть $k > 0$; проинтегрируем $f(z)e^{ikz}$ по контуру, который мы изобразили на рис. 5. Ввиду периодичности f

$$\int_{l_2} f(\xi)\, e^{ik\xi}\, d\xi = -\int_{l_4} f(\xi)\, e^{ik\xi}\, d\xi,$$

откуда по теореме Коши

$$|c_k| = \left| \frac{1}{2\pi} \int_0^{2\pi} f(x)\, e^{ikx}\, dx \right| =$$

$$= \left| \frac{1}{2\pi} \int_0^{2\pi} f(x - ih')\, e^{ikx} e^{-kh'}\, dx \right| \leqslant Ce^{-|k|h'} \leqslant Ce^{-|k|h},$$

так как $h' > 0$ — произвольное число, меньшее h. Случай $k < 0$ аналогичен.

Используя (23), перепишем (22) в виде

$$\left.\begin{aligned} f(x) &= \sum_{|k| \leqslant n} c_k e^{ikx} + R_n(x), \\ |R_n(x)| &\leqslant \left| \sum_{|k|>n} c_k e^{ikx} \right| \leqslant 2Ce^{-nh} \frac{e^{-h}}{1-e^{-h}}. \end{aligned}\right\} \tag{24}$$

Пусть $\varepsilon > 0$. Подберем минимальное n так, чтобы $|R_n(x)| < \varepsilon/2$. Из (24) мы получим, что

$$n = \frac{\log(1/\varepsilon)}{h \log e} + O(1). \tag{25}$$

Приблизим, далее, каждый коэффициент $c_k = \alpha_k + i\beta_k$ по модулю с точностью до $\varepsilon/2(2n+1)$ величиной $c_k' = \alpha_k' + i\beta_k'$, полагая для этого

$$\alpha_k' = \left[\frac{2\sqrt{2}\,(2n+1)\,\alpha_k}{\varepsilon}\right]\frac{\varepsilon}{2\sqrt{2}\,(2n+1)} = m_k^1\,\frac{\varepsilon}{2\sqrt{2}\,(2n+1)}\,,$$
$$\beta_k' = \left[\frac{2\sqrt{2}\,(2n+1)\,\beta_k}{\varepsilon}\right]\frac{\varepsilon}{2\sqrt{2}\,(2n+1)} = m_k^2\,\frac{\varepsilon}{2\sqrt{2}\,(2n+1)}\,. \tag{26}$$

При этом мы, очевидно, получим, что

$$\Big| f(x) - \sum_{|k|\leqslant n} c_k' e^{ikx} \Big| \leqslant \varepsilon. \tag{27}$$

Из (27) ясно, что если две функции f и g соответствуют одному и тому же набору $\{c_k'\}$, то $\| f - g \| \leqslant 2\varepsilon$, т. е. множество элементов в ε-покрытии можно оценить числом всех наборов $\{c_k'\}$, «индуцированных» пространством $A_h(C)$.

Согласно (26) каждое множество 2ε-покрытия определяется следующей матрицей, составленной из целых чисел:

$$U = \begin{Vmatrix} m_{-n}^1 & \ldots & m_0^1 & \ldots & m_n^1 \\ m_{-n}^2 & \ldots & m_0^2 & \ldots & m_n^2 \end{Vmatrix}.$$

При этом, использовав (23), мы получим неравенства для m_k^i:

$$| m_k^i | \leqslant \frac{2\sqrt{2}C\,(2n+1)\,e^{-|k|h}}{\varepsilon} = N_k, \quad i = 1, 2, \tag{28}$$

из которых следует, что

$$\log N_k = \log (1/\varepsilon) - | k |\, h \log e + \log n + D, \tag{29}$$

где D — ограниченная константа, не зависящая от k. Из (28) вытекает, что число всех возможных матриц U не превосходит

$$N = \prod_{k=-n}^{n} N_k^2,$$

и, следовательно, вспоминая (25), мы получим

$$\mathscr{H}_\varepsilon(A) \leqslant \log N = 2 \sum_{|k|\leqslant n} \log N_k \leqslant 4n \log \frac{1}{\varepsilon} - 2n^2 h \log e +$$
$$+ O(n \log n) = 2\,\frac{(\log (1/\varepsilon))^2}{h \log e} + O\Big(\log \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\Big).$$

Теперь проведем оценку снизу для $\mathscr{E}_{2\varepsilon}(A_h(0))$. Мы используем при этом два обстоятельства.

α) Для всякого $h' > h$ выполнение неравенств

$$| c_k | \leqslant C\,\frac{1 - e^{-(h'-h)}}{1 + e^{-(h'-h)}}\, e^{-|k|h'} \tag{30}$$

влечет за собой принадлежность функции

$$f(z)=\sum_{k=-\infty}^{+\infty} c_k e^{ikz} \tag{31}$$

пространству $A_h(C)$, так как тогда ряд (31) равномерно сходится в полосе ширины h и при $|\operatorname{Im} z|<h$ имеем

$$|f(z)| \leqslant \sum_{k=-\infty}^{+\infty} |c_k| e^{|k|h} \leqslant C \frac{1-e^{-(h'-h)}}{1+e^{-(h'-h)}} \sum_{k=-\infty}^{+\infty} e^{-|k|(h'-h)}=C.$$

$$\beta)\ \|f\|=\max_{x\in D} |f(x)| \geqslant \max_k |c_k|, \tag{32}$$

где c_k — коэффициенты в (31). Последнее следует из того, что

$$|c_k|=\left|\frac{1}{2\pi}\int_0^{2\pi} f(x) e^{ikx}\, dx\right| \leqslant \max_{x\in D} |f(x)|.$$

Выбрав $\varepsilon>0$ и положив $h'=(1+(\log(1/\varepsilon))^{-1})$, подбираем максимальное m так, чтобы

$$C\frac{1-e^{-(h'-h)}}{1+e^{-(h'-h)}} e^{-mh'} \geqslant 2\sqrt{2}\varepsilon,$$

откуда мы получим, что

$$m=\frac{\log(1/\varepsilon)}{h\log e}+O\left(\log\log\frac{1}{\varepsilon}\right). \tag{33}$$

Строим теперь множество $\{\Phi\}$ многочленов степени m:

$$\varphi(z)=\sum_{|k|\leqslant m} c_k e^{ikz}, \qquad c_k=(s_k^1+is_k^2)2\varepsilon, \tag{34}$$

где s_k^j $(j=1,2)$ — любые целые числа такие, что

$$|s_k^j| \leqslant \left[\frac{C(1-e^{-(h'-h)})}{(1+e^{-(h'-h)})}\frac{e^{-|k|h'}}{2\sqrt{2}\varepsilon}\right]=M_k. \tag{35}$$

Неравенства (35) гарантируют осуществление неравенств (30), таким образом, $\{\Phi\}\subset A_h(C)$. Кроме того, два различных многочлена φ_1 и φ_2 из $\{\Phi\}$ имеют по построению хотя бы для одного k, $|k|\leqslant m$:

$$|C_k^1-C_k^2| \geqslant 2\varepsilon,$$

т. е. все многочлены из $\{\Phi\}$ 2ε-различимы.

Согласно (34) каждый многочлен из $\{\Phi\}$ определяется целочисленной матрицей

$$U=\begin{Vmatrix} s_{-m}^1 & \ldots & s_0^1 & \ldots & s_m^1 \\ s_{-m}^2 & \ldots & s_0^2 & \ldots & s_m^2 \end{Vmatrix},$$

а число всех матриц U не меньше числа

$$M = \prod_{|k| \leqslant m} M_k^2.$$

Из (35) следует также, что

$$\log M_k = \log \frac{1}{\varepsilon} - h|k| \log e + O\left(\log \log \frac{1}{\varepsilon}\right)$$

равномерно по k, откуда, использовав (33), получим, что

$$\mathscr{E}_{2\varepsilon}(A_h(C)) \geqslant \log M \geqslant 2 \sum_{|k| \leqslant m} \log M_k \geqslant$$
$$\geqslant 4m \log \frac{1}{\varepsilon} - 2m^2 h \log e + O\left(\log \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\right) =$$
$$= 2 \frac{(\log (1/\varepsilon))^2}{h \log e} + O\left(\log \frac{1}{\varepsilon} \log \log \frac{1}{\varepsilon}\right).$$

Этим доказательство формул (21) заканчивается.

§ 3
ТИПИЧНЫЕ ПОРЯДКИ РОСТА ФУНКЦИЙ $\mathscr{H}_\varepsilon(A)$ И $\mathscr{E}_\varepsilon(A)$

Рассматриваемые далее функции $f(\varepsilon)$ будут обычно положительны и определены для всех ε в пределах $0 < \varepsilon < \varepsilon_0$. Так как постоянно будет изучаться их предельное поведение при $\varepsilon \to 0$, то сам знак $\varepsilon \to 0$ будет часто опускаться. Для описания предельного поведения функций $f(\varepsilon)$, кроме обычных знаков O и o, будут употребляться еще следующие обозначения:

$f \sim g$, если $\lim (f/g) = 1$,

$f \succsim g$, $g \precsim f$, если $\overline{\lim}\,(g/f) \leqslant 1$,

$f \asymp g$, если $f = O(g)$ и $g = O(f)$,

$f \succcurlyeq g$, $g \preccurlyeq f$, если $f = O(g)$,

$f \succ\succ g$, $g \prec\prec f$, если $f = o(g)$.

Отношение $f \sim g$ называется *сильной эквивалентностью*, а отношение $f \asymp g$ — *слабой эквивалентностью* функций f и g.

Рассмотренные в § 2 примеры доставляют нам образцы трех типичных порядков роста функций $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ при $\varepsilon \to 0$.

I. Если A — ограниченное множество в n-мерном евклидовом пространстве D^n, обладающее внутренними точками, то

$$\mathscr{M}_\varepsilon(A) \asymp \mathscr{N}_\varepsilon(A) \asymp (1/\varepsilon)^n. \tag{36}$$

Формула (36) верна и в любом n-мерном банаховом пространстве (см. § 4). Из нее вытекает, что

$$\mathscr{E}_\varepsilon(A) \sim \mathscr{H}_\varepsilon(A) \sim n \log(1/\varepsilon). \tag{37}$$

Так как в силу (37) поведение $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ определяется в первую очередь размерностью n пространства D^n, то возникает естественная идея для произвольного вполне ограниченного множества A определить *нижнюю и верхнюю метрическую размерность* формулами

$$\underline{\mathrm{dm}}(A) = \underline{\lim}\, \mathscr{H}_\varepsilon(A)/\log(1/\varepsilon), \tag{38}$$

$$\overline{\mathrm{dm}}(A) = \overline{\lim}\, \mathscr{H}_\varepsilon(A)/\log(1/\varepsilon). \tag{39}$$

Из теоремы IV § 1 без труда выводится, что формулы

$$\underline{\mathrm{dm}}(A) = \underline{\lim}\, \mathscr{E}_\varepsilon(A)/\log(1/\varepsilon),$$

$$\overline{\mathrm{dm}}(A) = \overline{\lim}\, \mathscr{E}_\varepsilon(A)/\log(1/\varepsilon)$$

определяют те же самые величины $\underline{\mathrm{dm}}(A)$ и $\overline{\mathrm{dm}}(A)$. Если

$$\underline{\mathrm{dm}}(A) = \overline{\mathrm{dm}}(A) = \mathrm{dm}(A),$$

то $\mathrm{dm}(A)$ называется просто *метрической размерностью* множества A. Очевидно, что в нашем случае вполне ограниченного множества A в D^n, обладающего внутренними точками,

$$\mathrm{dm}(A) = n.$$

Легко доказывается (см. теорему XII в § 4), что для лежащего в банаховом пространстве выпуклого бесконечномерного множества A метрическая размерность всегда равна $+\infty$, т. е. порядок роста функций $\mathscr{H}_\varepsilon(A)$ и $\mathscr{E}_\varepsilon(A)$ превосходит $\log(1/\varepsilon)$:

$$\mathscr{H}_\varepsilon(A) \succ\!\succ \log(1/\varepsilon), \quad \mathscr{E}_\varepsilon(A) \succ\!\succ \log(1/\varepsilon).$$

Для различения массивности такого рода множеств метрическая размерность становится непригодной.

II. Среди важных для анализа бесконечномерных множеств наименьшей массивностью обладают типичные множества вполне ограниченных *аналитических* функций такого рода, как множество $A_h(C)$, рассмотренное в примере 3 § 2. Типичным порядком роста функций $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ здесь является их рост по закону

$$\mathscr{E}_\varepsilon \asymp \mathscr{H}_\varepsilon \asymp (\log(1/\varepsilon))^s \tag{40}$$

с некоторой конечной константой $s > 1$ или промежуточные порядки роста такого рода, как

$$\frac{(\log(1/\varepsilon))^s}{\log\log(1/\varepsilon)} \tag{41}$$

(последняя функция растет медленнее чем $(\log(1/\varepsilon))^s$, но быстрее чем $(\log(1/\varepsilon))^{s-\delta}$ при любом $\delta > 0$). Простейшей числовой характеристикой подобных порядков роста является [5]

$$df = \lim \frac{\log \mathscr{H}_\varepsilon}{\log \log (1/\varepsilon)} .$$

В § 7 мы увидим, что во многих важных и достаточно общих случаях множеств аналитических функций $df(A)$ совпадает с числом независимых переменных, или геометрически — для классов функций $f(P)$ от точки P комплексного многообразия с (комплексной) размерностью этого многообразия. Мы назовем, несколько условно, $df(A)$ *функциональной размерностью* множества A.

III. Примеры п. 2 и 3 из § 2 показывают, что рост $\mathscr{H}_\varepsilon$ и $\mathscr{E}_\varepsilon$ для функций, подчиненных лишь условию Липшица, значительно быстрее, чем указанные выше порядки роста для классов аналитических функций. Оказывается, что *порядки роста*

$$\mathscr{E}_\varepsilon \asymp \mathscr{H}_\varepsilon \asymp (1/\varepsilon)^q \tag{42}$$

с конечными положительными q типичны для классов функций, дифференцируемых некоторое конечное число p раз и имеющих производные старшего порядка p, удовлетворяющие условию Гёльдера порядка α. При этом для функций от n переменных получается формула

$$\mathscr{E}_\varepsilon \asymp \mathscr{H}_\varepsilon \asymp (1/\varepsilon)^{n/(p+\alpha)} \tag{43}$$

(см. § 5). Грубо формулу (43) можно пояснить так: массивность множества p раз дифференцируемых функций от n переменных, у которых p-е производные удовлетворяют условию Гёльдера порядка α, определяется отношением

$$q = n/(p + \alpha) \tag{44}$$

числа независимых переменных n к *«показателю гладкости» $p + \alpha$*. По поводу происхождения этих важных соотношений (43), (44) см. Добавление I. Общим определением *метрического порядка* вполне ограниченного множества A служит формула [6]

$$q(A) = \lim \frac{\log \mathscr{H}_\varepsilon(A)}{\log (1/\varepsilon)} . \tag{45}$$

---

[5] Читатель без труда напишет сам определения нижней функциональной размерности $\underline{df}(A)$ и верхней функциональной размерности $\overline{df}(A)$. Употребление вместо $\mathscr{H}_\varepsilon$ функции $\mathscr{E}_\varepsilon$ в силу теоремы IV § 1 приводит к тем же самым $df(A)$, $\underline{df}(A)$, $\overline{df}(A)$.

[6] Здесь тоже естественно ввести вполне понятным образом понятия нижнего метрического порядка $\underline{q}(A)$ и верхнего метрического порядка $\bar{q}(A)$. Употребление вместо $\mathscr{H}_\varepsilon$ функции $\mathscr{E}_\varepsilon$ не меняет значений $q(A)$, $\bar{q}(A)$, $\underline{q}(A)$.

IV. Существенно большие порядки роста $\mathcal{H}_\varepsilon$ и $\mathcal{E}_\varepsilon$, чем в (42), возникают при рассмотрении функционалов той или иной гладкости, заданных на бесконечномерных вполне ограниченных множествах (см. § 9).

§ 4
ε-ЭНТРОПИЯ И ε-ЕМКОСТЬ В КОНЕЧНОМЕРНЫХ ПРОСТРАНСТВАХ

n-мерное пространство мы будем рассматривать реализованным в виде координатного n-мерного пространства D^n с точками

$$x = (x_1, \ldots, x_n),$$

координаты которых $x_1, \ldots, x_n$ суть действительные числа, хотя по существу интересующие нас соотношения в большинстве случаев с выбором координат не связаны. С выбором системы координат связан некоторый вспомогательный аппарат, например специальная норма

$$\| x \|_0 = \max_{1 \leqslant k \leqslant n} |x_k|.$$

Единичный шар в этой норме, т. е. множество S_0 точек $x \in D^n$ с

$$\| x \|_0 \leqslant 1,$$

называется *единичным кубом.* Из единичного куба умножением на $b > 0$ и сдвигом на a получаются кубы

$$S_0(a, b) = bS_0 - a;$$

$2b$ мы назовем диаметром куба $S_0(a, b)$.

Множество точек x^i, координаты которых имеют вид $(dm_1^i, \ldots, dm_n^i)$, где m_k^i $(k = 1, 2, \ldots, n)$ — всевозможные целые числа, мы назовем кубической решеткой диаметра d.

Как известно, точка x называется *внутренней точкой* множества $A \subset D^n$, если A содержит некоторый куб $S_0(a, b)$ с центром в $a \in A$. Множества, состоящие только из внутренних точек, называются *открытыми*, что и определяет обычную топологию в D^n.

Множество $A \subset D^n$ называется *выпуклым*, если из $x \in A$, $y \in A$, $\alpha + \beta = 1$, $\alpha \geqslant 0$, $\beta \geqslant 0$ вытекает $\alpha x + \beta y \in A$. Замкнутое выпуклое множество, имеющее хотя бы одну внутреннюю точку, будем называть *выпуклым телом.*

Любое симметричное относительно начала координат ограниченное выпуклое тело может быть выбрано за единичную сферу. Норма $\| x \|_S$, соответствующая выпуклому симметричному телу S, определяется формулой

$$\| x \|_S = \inf \left\{ b \,\middle|\, \frac{x}{b} \in S,\ b > 0 \right\}.$$

В свою очередь по норме $\| x \|$ порождающий ее единичный шар определяется формулой

$$S = \{x \mid \| x \| \leqslant 1\}.$$

Таким образом определяется в D^n произвольная банахова метрика. Пространство D^n с метрикой, порожденной симметричным выпуклым ограниченным телом S, обозначается D^n_S. В случаях

$$\| x \|_p = \Big(\sum_{k=1}^{n} | x_k |^p\Big)^{1/p}, \quad p \geqslant 1,$$

единичные шары обозначаются S^n_p, а соответствующее пространство D^n_p. Пространство D^n_2 есть обычное евклидово (или «декартово») координатное пространство.

Общеизвестно, что любая банахова метрика определяет в D^n одну и ту же топологию. Не зависит от выбора метрики по существу и понятие объема и меры множеств в D^n. В стандартной теории лебеговой меры μ^n мера $^1/_2 S_0$ принимается равной 1. Если за единицу измерения принять объем $^1/_2 S$ в какой-либо иной метрике, то обычный путь построения лебеговой теории меры приведет к мере

$$\mu^n_S (A) = \mu^n (A)/\mu^n(^1/_2 S),$$

отличающейся от μ^n лишь постоянным множителем.

Множество A измеримо по Жордану, если оно измеримо по Лебегу и граница его имеет меру нуль. Меру $\mu^n (A)$ такого множества будем называть его *«объемом»* (если же мера границы не равна нулю, то объем множеству не приписывается).

Нетрудно понять, что вопрос о нахождении минимальных ε-сетей в пространстве D^n_S упирается в задачу о нахождении «наиболее экономного покрытия» подмножеств пространства D^n параллельно сдвинутыми телами εS, а задача о нахождении максимальных 2ε-различимых множеств сводится к задаче о наиболее удачном размещении в D^n параллельно сдвинутых непересекающихся тел εS с центрами на заданном множестве.

По данному вопросу существует обширная литература. Достаточно полную библиографию читатель найдет в книге Тота [19] [7].

Мы в настоящем параграфе коснемся только самых общих и простых результатов об оценках функции $\mathcal{H}_\varepsilon (A)$ и $\mathcal{E}_\varepsilon (A)$ для $A \subset D^n_S$.

Т е о р е м а VIII. *Для любого ограниченного множества* $A \subset D^n_S$ *существует константа* $\alpha > 0$ *такая, что при всех* $\varepsilon \leqslant \varepsilon_1$

$$\mathcal{N}_\varepsilon (A) \leqslant \alpha (1/\varepsilon)^n. \tag{46}$$

[7] Много существенного читатель найдет в примечаниях к этой книге, составленных И. М. Ягломом.

Если же A содержит внутреннюю точку в D^n, то существует константа $\beta > 0$ *такая, что при* $\varepsilon \leqslant \varepsilon_2$

$$\mathscr{M}_{2\varepsilon}(A) \geqslant \beta\,(1/\varepsilon)^n. \tag{47}$$

В обозначениях § 3 теорема VIII означает, что для ограниченного множества в D^n_S, имеющего внутреннюю точку,

$$\mathscr{N}_\varepsilon(A) \asymp \mathscr{M}_\varepsilon(A) \asymp (1/\varepsilon)^n. \tag{48}$$

Об этой теореме мы упоминали в § 3 (см. (36)).

Теорема IX. *Для D^n_S существуют константы* $\theta = \theta(n, S)$ *и* $\tau = \tau(n, S)$ *такие, что для всякого множества A, объема* $\mu^n(A) > 0$

$$\mathscr{N}_\varepsilon(A) \sim \theta\,\frac{\mu^n(A)}{\mu^n(S)}\left(\frac{1}{\varepsilon}\right)^n, \qquad \mathscr{M}_{2\varepsilon}(A) \sim \tau\,\frac{\mu^n(A)}{\mu^n(S)}\left(\frac{1}{\varepsilon}\right)^n. \tag{49}$$

Константу θ естественно называть *«минимальной плотностью покрытия D^n телом S»*, а константу τ — «максимальной плотностью размещения тела S в D^n».

Для θ и τ очевидны неравенства

$$\tau \leqslant 1 \leqslant \theta.$$

Доказательство теоремы VIII. Подбираем константы b и B так, чтобы одновременно

$$bS_0 \subset S \subset BS_0, \qquad A \subset BS_0.$$

Пусть $\varepsilon > 0$. Для покрытия куба BS_0 потребуется не более

$$(B/\varepsilon b + 2)^n \tag{50}$$

кубов $\varepsilon b S_0 + a^i = S_0(a^i, \varepsilon b)$, где a^i имеет координаты $2b\varepsilon\,(m_1^i, m_2^i, \ldots, m_n^i)$, m_j^i — целые числа, меньшие по модулю $[B/2\varepsilon b] + 1$. Описав вокруг кубов $\varepsilon b S_0 + a^i$ тела $\varepsilon S + a^i$, получим ε-покрытие A и, таким образом,

$$\mathscr{N}_\varepsilon(A) \leqslant (B/\varepsilon b + 2)^n,$$

откуда следует (46).

Пусть, далее, A имеет внутреннюю точку a. Подыщем такое c, чтобы куб $S_0(a, c) \subset A$. Куб $S_0(a, c)$ содержит более $(2\,[c/2\varepsilon B])^n$ кубов $S_0(a^i, \varepsilon B)$, где a^i — векторы $2\varepsilon B\,(m_1^i, \ldots, m_n^i)$, где m_j^i целые и не превосходят по модулю $[c/2\varepsilon B]$. Множество a^i 2ε-различимо, откуда следует (47).

Доказательство теоремы IX. Обозначим [8]

$$\varliminf_{\varepsilon\to 0} \mathcal{N}_\varepsilon(S_0)\mu^n(S)\left(\frac{\varepsilon}{2}\right)^n = \varliminf_{\varepsilon\to 0} \mathcal{N}_\varepsilon(S_0)\frac{\mu^n(S)}{\mu^n(S_0)}\varepsilon^n$$

через θ. Покажем, что на самом деле существует

$$\lim_{\varepsilon\to 0} \mathcal{N}_\varepsilon(S_0)\,\mu^n(S)\,(\varepsilon/2)^n = \theta.$$

Пусть $\delta > 0$. Подбираем такое ε_1, что при $\varepsilon \leqslant \varepsilon_1$

$$\theta - \delta \leqslant \mathcal{N}_\varepsilon(S_0)\,\mu^n(S)\,(\varepsilon/2)^n,$$

а

$$\mathcal{N}_{\varepsilon_1}(S_0)\,\mu^n(S)\,(\varepsilon_1/2)^n \leqslant \theta + \delta/2.$$

Куб S_0 можно покрыть не более

$$(1/\lambda + 2)^n$$

кубами $S_0(a^i, \lambda)$ при любом λ, $0 < \lambda < 1$, см. (50), а каждый из последних покрывается $\mathcal{N}_{\varepsilon_1}$ телами $\varepsilon\lambda S$.

Таким образом,

$$\mathcal{N}_{\varepsilon_1\lambda}(S_0) \leqslant (1/\lambda + 2)^n\,\mathcal{N}_{\varepsilon_1}(S_0),$$

т. е.

$$\mathcal{N}_{\varepsilon_1\lambda}(S_0)\mu^n(S)\left(\frac{\varepsilon_1}{2}\right)^n\lambda^n \leqslant \lambda^n\left(\frac{1}{\lambda}+2\right)^n\mathcal{N}_{\varepsilon_1}(S_0)\left(\frac{\varepsilon_1}{2}\right)^n\mu^n(S).$$

Выбрав λ_1 так, чтобы при $\lambda \leqslant \lambda_1$

$$\lambda^n(1/\lambda + 2)^n(\theta + \delta/2) \leqslant \theta + \delta,$$

получим для всех $\lambda \leqslant \lambda_1$

$$\theta - \delta \leqslant \mathcal{N}_{\varepsilon_1\lambda}(S_0)\,\mu^n(S)\,(\varepsilon_1\lambda/2)^n \leqslant \theta + \delta, \tag{51}$$

т. е.

$$\lim_{\varepsilon\to 0} \mathcal{N}_\varepsilon(S_0)\,\mu^n(S)\,(\varepsilon/2)^n = \theta.$$

Пусть, далее, A — произвольное тело объема $\mu^n(A) > 0$, Π_{2d} — кубическая решетка диаметра $2d$. Задавшись $\delta' > 0$, подберем d так, чтобы сумма мер кубов $S_0'(\bar{a}^i, d)$ $(i = 1, \ldots, m)$, $\bar{a}^i \in \Pi_{2d}$, имеющих общие точки *с границей* A, была бы меньше δ'. Пусть a^i $(i = 1, 2, \ldots, k)$ — центры кубов $S(a^i, d)$, каждый из которых *целиком лежит внутри* A. При этом, конечно,

$$k \geqslant (\mu^n(A) - \delta')/(2d)^n. \tag{52}$$

Выбрав произвольное $d' < d$, рассмотрим множество кубов

---

[8] Напомним, что S_0 — куб $\|x_0\| \leqslant 1$; его мера равна 2^n.

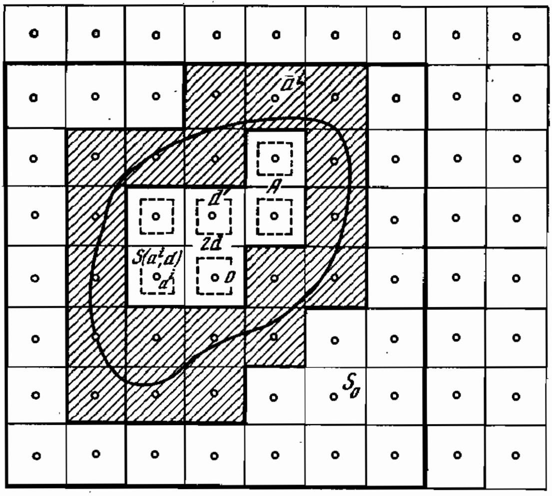


Рис. 6

S_0 (a^i, d') (рис. 6). Согласно (51) для $\varepsilon \leqslant \varepsilon_1\lambda_1$

$$\theta - \delta \leqslant \mathcal{N}_{\varepsilon d'}(d'S_0)\,\mu^n(S)\,(\varepsilon/2)^n \leqslant \theta + \delta$$

выберем ε_2 настолько малым, чтобы $2\varepsilon d' < d - d'$ для $\varepsilon \leqslant \varepsilon_2$. Тогда, очевидно,

$$\mathcal{N}_{\varepsilon d'}(A) \geqslant k\mathcal{N}_{\varepsilon d'}(d'S_0),$$

так как покрытия смежных кубов не пересекаются, и для $\varepsilon \leqslant \varepsilon_2$

$$\mathcal{N}_{\varepsilon d'}(A)\left(\frac{\varepsilon}{2}\right)^n \mu^n(S) \geqslant \frac{\mu^n(A) - \delta'}{(2d)^n}(\theta - \delta),$$

откуда

$$\varliminf_{\varepsilon \to 0} \mathcal{N}_\varepsilon(A)\,\varepsilon^n \mu^n(S) \geqslant \theta\mu^n(A)$$

ввиду произвольности δ', δ и $d' < d$. С другой стороны,

$$\mathcal{N}_{\varepsilon d}(A) \leqslant (k + m)\,\mathcal{N}_{\varepsilon d}(dS_0) \leqslant \frac{\mu^n(A) + \delta'}{(2d)^n}(\theta + \delta)\frac{1}{\mu^n(S)}\left(\frac{2}{\varepsilon}\right)^n,$$

т. е.

$$\varlimsup_{\varepsilon \to 0} \varepsilon^n \mu^n(S)\,\mathcal{N}_\varepsilon(A) \leqslant \theta\mu^n(A), \tag{53}$$

что и требовалось доказать.

Читателю предоставляется провести самостоятельно доказательство теоремы для $\mathcal{M}_{2\varepsilon}$.

Кроме теорем VIII и IX, отметим следующее весьма грубое неравенство, оценивающее сверху и снизу θ и τ.

Т е о р е м а X. *Для любого* D_S^n

$$\theta/\tau \leqslant 2^n, \tag{54}$$

откуда

$$2^{-n} \leqslant \tau \leqslant 1 \leqslant \theta \leqslant 2^n. \tag{55}$$

Выбрав $\delta > 0$, подберем ε_0 так, чтобы одновременно для некоторого A объема $\mu^n(A) > 0$ при $\varepsilon \leqslant \varepsilon_0$ выполнялись неравенства:

$$\varepsilon^n \mu^n(S) \mathscr{N}_\varepsilon(A) \geqslant (\theta - \delta) \mu^n(A),$$

$$\varepsilon^n \mu^n(S) \mathscr{M}_{2\varepsilon}(A) \leqslant (\tau + \delta) \mu^n(A).$$

Используя, далее, неравенство $\mathscr{N}_\varepsilon(A) \leqslant \mathscr{M}_\varepsilon(A)$ теоремы IV § 1, получим

$$\theta - \delta \leqslant (\tau + \delta) 2^n,$$

откуда следует теорема X.

Отметим, что если факт *существования* констант τ и θ (теорема IX) доказывается весьма просто, то вычисления их или даже лишь оценки их для разных конкретных пространств представляют нередко очень трудные задачи.

Тривиальным случаем является случай пространства D_∞^n, так как в этом пространстве единичной сферой является единичный куб S_0 и все пространство может быть разложено без перекрытий на единичные кубы так, что расстояние между центрами кубов равняется диаметру кубов (пространство «разложимо»). Во всяком разложимом пространстве, как легко видеть,

$$\theta(n, S) = \tau(n, S) = 1. \tag{56}$$

Укажем еще два примера разложимых пространств: пространство D_1^2 и пространство D_S^2, где S — правильный шестиугольник. В первом пространстве сферой является квадрат, во втором — шестиугольник, которыми можно замостить плоскость.

(Заметим, что с подобным явлением «разложимости» мы столкнулись в одном бесконечномерном пространстве — пространстве $F_1^\Delta(L)$; см. § 1, п. 2.)

Величины θ и τ вычислены в случае пространства D_2^2 [19]. Оказывается, что наилучшее покрытие плоскости кругами осуществляется следующим образом. Плоскость покрывается правильными шестиугольниками, а затем вокруг каждого из них описывается круг. Наилучшее расположение кругов на плоскости, не пересекающихся друг с другом, получится, если в каждый из шестиугольников вписать по кругу.

Нетрудно подсчитать, что при этом

$$\tau = \frac{\pi}{\sqrt{12}} = 0{,}9069\ldots, \qquad \theta = \frac{2\pi}{\sqrt{27}} = 1{,}2092\ldots$$

Уже в случае трехмерного пространства задача о нахождении плотнейшего заполнения пространства шарами и связанная с ней задача вычисления констант τ и θ для пространства D_2^3 полностью еще не решена. Тем более не вычислены значения τ_n и θ_n (так мы обозначали τ и θ для пространства D_2^n). Оценки для θ_n, наиболее точные до настоящего времени, были получены Давенпортом [20] и Ватсоном [21].

Прежде чем формулировать их результат, введем одно определение. Пусть R — решетка в n-мерном пространстве такая, что сферы $S_i = S(a^i, 1)$, $a^i \in R$, покрывают все D^n. Пусть $\sum\limits_{\lambda}$ означает суммирование по тем i, для которых $a^i \in R \cap \lambda S$. Величину

$$\lim_{\lambda\to\infty} \left(\sum_{\lambda} \mu^n (S_i)/\mu^n (\lambda S)\right)$$

в предположении существования предела естественно назвать плотностью покрытия пространства D^n телом S с центрами на R. Обозначим эту величину через $\theta_R(n, S)$. Величина $\theta_R(n, S)$, как легко видеть, оценивает значение $\theta(n, S)$ сверху:

$$\theta(n, S) \leqslant \theta_R(n, S). \tag{57}$$

Давенпорт и Ватсон построили решетку R в D^n такую, что

$$\theta_R(n, S_2^n) \leqslant (1{,}07)^n, \tag{58}$$

а следовательно,

$$\theta_n \leqslant (1{,}07)^n. \tag{59}$$

Теоремы Давенпорта и Ватсона основаны на тонких построениях в n-мерном пространстве; мы их не приводим, отсылая интересующихся к соответствующей литературе.

Оценки величины τ_n можно извлечь из более старой работы Блихфельдта [22].

Основой в ней является следующая простая

Л е м м а I (Блихфельдт). *Пусть в евклидовом пространстве задано m точек, попарно удаленных не менее чем на 2. Тогда сумма квадратов их расстояний до любой точки пространства не меньше* $2(m-1)$.

Д о к а з а т е л ь с т в о. Очевидно, что достаточно рассмотреть случай, когда данной точкой пространства является нуль. Выбрав ортонормированный базис, обозначим координаты i-й точки через $(x_1^i, \ldots, x_n^i)$. По условию

$$\sum_{k=1}^{n} (x_k^i - x_k^j)^2 \geqslant 4, \quad i \neq j. \tag{60}$$

Складывая (60) при $i \neq j$, получаем

$$m \sum_{i=1}^{m} ((x_1^i)^2 + \ldots + (x_n^i)^2) - (x_1^1 + \ldots + x_1^m)^2 -$$
$$- (x_n^1 + \ldots + x_n^m)^2 \geqslant 2m(m-1),$$

отсюда

$$\sum_{k=1}^{n} \sum_{i=1}^{m} (x_k^i)^2 \geqslant 2(m-1),$$

что и требовалось доказать.

Из доказанной леммы сразу вытекает, что если $n+1$ точек евклидова пространства имеют попарные расстояния не меньше 2, то радиус шара, содержащего все эти точки, не меньше $\sqrt{2n/(n+1)}$.

Пусть нам задана произвольная $\varepsilon\sqrt{2n/(n+1)}$-сеть. В каждой сфере помещается не более $n+1$ точек на расстоянии 2ε.

Отсюда для всякого тела A легко получить, что

$$\mathcal{M}_{2\varepsilon}(A) \leqslant (n+1)\mathcal{N}_{\varepsilon\sqrt{2n/(n+1)}}(A).$$

Из (59) и (55) немедленно следует, что

$$2^{-n} \leqslant \tau_n \leqslant (n+1)\,\theta_n \left(\frac{n+1}{2n}\right)^{n/2} \leqslant (n+1)\sqrt{e}\,2^{-n/2}(1{,}07)^n \leqslant$$
$$\leqslant (n+1)\sqrt{e}\left(\frac{11}{14}\right)^n. \tag{61}$$

Заканчивая этот параграф, покажем, что основные его результаты не допускают содержательного обобщения на бесконечномерные банаховы пространства. Ограниченные множества в бесконечномерном пространстве уже не обязаны (в отличие от случая D^n) быть вполне ограниченными. Но и для вполне ограниченных множеств в бесконечномерном пространстве нельзя получить никакой универсальной (годной для всех вполне ограниченных множеств в данном бесконечномерном E) оценки сверху для роста $\mathcal{H}_\varepsilon$ и $\mathcal{E}_\varepsilon$ (в D^n, как мы видели, они всегда $\preccurlyeq \varepsilon^{-n}$). Это видно из следующей теоремы:

Т е о р е м а XI. *Пусть* $\varphi(\varepsilon)$, *монотонно возрастая, стремится к* $+\infty$ *при* $\varepsilon \to 0$ *и банахово пространство* E *бесконечномерно. Тогда в* E *можно найти компакт* K, *для которого*

$$\mathcal{N}_\varepsilon(K) \succcurlyeq \varphi(\varepsilon), \qquad \mathcal{M}_\varepsilon(K) \succcurlyeq \varphi(\varepsilon).$$

Очевидно, в силу теоремы IV § 1 достаточно доказать теорему XI для $\mathcal{M}_\varepsilon$. Найдем в E последовательность линейных подпространств

$$E_1 \subset E_2 \subset \ldots \subset E_n \subset \ldots, \quad \dim E_n = n.$$

Покажем, что существует последовательность точек

$$x_n \in E_n$$

со свойствами

$$\| x_n \| = 1, \qquad \rho (x_n, E_{n-1}) = 1.$$

Для этого возьмем в каждом E_n по точке y_n, не принадлежаще к E_{n-1}. Найдем (это возможно, см. [23, с. 16]) в E_n одну из «бли жайших» к y_n точек z_n, т. е. точку z_n, для которой

$$\| y_n - z_n \| = \rho (y_n, E_{n-1}) = \rho_n,$$

и положим

$$x_n = \rho_n^{-1} (y_n - z_n).$$

Легко видеть, что x_n обладает требуемыми свойствами.

Введем теперь обратную к $\varphi (\varepsilon)$ функцию $\psi (n)$. Составим мн жество K из точки 0 и точек

$$\xi_n = 2\psi (n)\, x_n, \quad n = 1, 2, \ldots$$

То, что K есть компакт, очевидно, так как

$$\| \xi_n \| = 2\psi (n) \to 0 \quad \text{при } n \to \infty.$$

При любом ε для

$$n_1 = [\varphi (\varepsilon)]$$

получим

$$\psi (n_1) \geqslant \varepsilon,$$

откуда вытекает, что множество точек

$$0, \xi_1, \ldots, \xi_{n_1}$$

ε-различимо в E. Так как число его элементов

$$n_1 + 1 \geqslant \varphi (\varepsilon),$$

то

$$\mathcal{M}_\varepsilon (K) \succcurlyeq \varphi (\varepsilon).$$

Множество A, лежащее в банаховом пространстве E, называется n-мерным, если оно помещается в n-мерном линейном подпространстве, но не помещается в линейном подпространстве меньшей размерности. В силу теоремы VIII выпуклое n-мерное ограниченное множество в любом банаховом пространстве имеет метрическую размерность n и, более того, для него

$$\mathcal{N}_\varepsilon (A) \asymp \mathcal{M}_\varepsilon (A) \asymp \varepsilon^{-n}.$$

Множество A в банаховом пространстве называется бесконечномерным, если оно не умещается ни в одном линейном пространстве — подпространстве конечной размерности. Имеет место следующая

Теорема XII. *Если множество A в банаховом пространстве бесконечномерно и выпукло, то*

$$\mathcal{N}_\varepsilon(A) \succ\!\succ \left(\frac{1}{\varepsilon}\right)^n, \qquad \mathcal{M}_\varepsilon(A) \succ\!\succ \left(\frac{1}{\varepsilon}\right)^n$$

при любом n.

Доказательство. При любом $n' > n$ в A существует $n' + 1$ аффинно независимых точек. Порождаемый ими n'-мерный симплекс в силу выпуклости A целиком лежит в A, но для такого симплекса в силу теоремы VIII функции $\mathcal{N}_\varepsilon$ и $\mathcal{M}_\varepsilon$ имеют порядок роста $\varepsilon^{-n'}$.

§ 5
ε-ЭНТРОПИЯ И ε-ЕМКОСТЬ ФУНКЦИЙ КОНЕЧНОЙ ГЛАДКОСТИ

Формула (43) может быть установлена без больших лишних усилий для функций, заданных на произвольном связном компакте K, имеющем определенную метрическую размерность

$$\mathrm{dm}\,(K) = n$$

и лежащем в конечномерном банаховом пространстве. Для большей определенности будем предполагать это банахово пространство реализованным координатно как D^N с нормой $\| x \|_0$ (как описано в § 4). Очевидно, что всегда $n \leqslant N$, в силу же предположения связности компакта K всегда $n \geqslant 1$. Впрочем наибольший интерес имеет простой случай, когда K есть параллелепипед размерности n в D^n.

Как обычно, в пространстве D^K ограниченных действительных функций на K норма равна

$$\| f \| = \sup_{x \in K} | f(x) | . \tag{62}$$

Функция из этого пространства имеет гладкость $q > 0$ ($q = p + \alpha$, p — целое, $0 < \alpha \leqslant 1$), если для любых векторов $x \in K$ и $x + h \in K$

$$f(x + h) = \sum_{k=0}^{p} \frac{1}{k!} B_k(h, x) + R(h, x), \tag{63}$$

где $B_k(h, x)$ по h есть однородная форма степени k и

$$| R(h, x) | \leqslant C \| h \|_0^q, \tag{64}$$

где C — некоторая константа.

Все функции f, удовлетворяющие (62) и (63) с заданной константой C, образуют класс $F_q^K(C)$, функции же, удовлетворяющие (62) и (63) с какой-либо константой C (зависящей от f),— класс F_q^K. Если K есть n-мерный параллелепипед, то F_q^K является просто классом функций, имеющим частные производные всех порядков $k \leqslant p$, для которых частные производные порядка p удовлетворяют условию Гёльдера порядка α.

Теорема XIII. *Для любого ограниченного в смысле метрики* (62) *множества* $A \subset F_q^K(C)$ *имеют место оценки*

$$\mathscr{H}_\varepsilon(A) \preccurlyeq \left(\frac{1}{\varepsilon}\right)^{n/q}, \qquad \mathscr{E}_\varepsilon(A) \preccurlyeq \left(\frac{1}{\varepsilon}\right)^{n/q}, \tag{65}$$

которые не могут быть улучшены в том смысле, что существуют ограниченные A, *удовлетворяющие соотношению*

$$\mathscr{E}_\varepsilon(A) \asymp \mathscr{H}_\varepsilon(A) \asymp \left(\frac{1}{\varepsilon}\right)^{n/q}. \tag{66}$$

Первая часть теоремы уже включает в себя утверждение (которое легко доказать независимо), что ограниченное $A \subset F_q^K(C)$ вполне ограничено (т. е. имеет в D^K компактное замыкание). Можно, кроме того, доказать, что для функций f из ограниченного $A \subset F_q^K$ формы $B_k(h, x)$ равностепенно ограничены:

$$|B_k(h, x)| \leqslant C_k \|h\|_0^k, \quad k = 0, 1, \ldots, p, \tag{67}$$

где константы C_k зависят только от A, а не от $f \in A$. Если учесть это обстоятельство, то ясно, что для доказательства теоремы XIII достаточно рассмотреть множества

$$F^K(C_0, \ldots, C_q, C),$$

состоящие из всех функций f, удовлетворяющих условиям (63), (64), (67) с заданными константами $C_0, \ldots, C_p, C$. Следовательно, теорема XIII может быть выведена [9] из следующего предложения.

Теорема XIV. *Если все константы* $C_0, \ldots, C_p, C$ *положительны, то*

$$\mathscr{E}_\varepsilon(F_q^K(C_0, \ldots, C_p, C)) \asymp \mathscr{H}_\varepsilon(F_q^K(C_0, \ldots, C_p, C)) \asymp \left(\frac{1}{\varepsilon}\right)^{n/q}. \tag{68}$$

В одномерном случае (K есть отрезок $[a, b]$ числовой прямой) теорема XIV звучит так:

---

[9] Мы не даем доказательств (в принципе несложных, но довольно громоздких) использованных выше свойств ограниченных $A \subset F_q^K(C)$. Читатель, который не сможет их восстановить, должен будет считать доказанным результатом этого параграфа лишь теорему XIV, которая и будет использоваться далее.

Теорема XV. *Пусть*

$$A = F_q^K (C_0, \ldots, C_p, C)$$

есть множество функций $f(x)$, *определенных для* $x \in [a, b]$ *и подчиненных условиям*:

$$|f^{(k)}(x)| \leqslant C_k, \quad k = 0, \ldots, p, \tag{69}$$

$$|f^{(p)}(x+h) - f^{(p)}(x)| \leqslant C|h|^{\alpha}, \quad 0 < \alpha \leqslant 1. \tag{70}$$

Если константы $C_0, \ldots, C_p, C$ *положительны, то*

$$\mathscr{H}_\varepsilon(A) \asymp \mathscr{E}_\varepsilon(A) \asymp \varepsilon^{-1/q}, \quad q = p + \alpha. \tag{71}$$

Мы проведем подробное доказательство теоремы XV и укажем, какие изменения надо произвести в ходе доказательства в общем случае теоремы XIV. В силу теоремы IV из § 1 ясно, что заключение теоремы XV равносильно совокупности двух асимптотических неравенств

$$\mathscr{H}_\varepsilon(A) \preccurlyeq \varepsilon^{-1/q}, \tag{72}$$

$$\mathscr{E}_\varepsilon(A) \succcurlyeq \varepsilon^{-1/q}. \tag{73}$$

Доказательство неравенства (72). Положим

$$\Delta = (\varepsilon/2C)^{1/q} \tag{74}$$

и выберем точки

$$x_r = x_0 + r\Delta, \quad r = 0, \ldots, s,$$

так, чтобы отрезок $[a, b]$ заключался в отрезке $[x_0, x_s]$. Легко видеть, что число s можно взять порядка

$$s \asymp \Delta^{-1} \asymp \varepsilon^{-1/q}. \tag{75}$$

Полагая

$$\varepsilon_k = \varepsilon/e\Delta^k \asymp \Delta^{q-k} \asymp \varepsilon^{1-k/q}, \tag{76}$$

$$\beta_r^{(k)}(f) = [f^{(k)}(x_r)/\varepsilon_k], \tag{77}$$

поставим в соответствие каждой функции $f \in A$ матрицу

$$\beta = \| \beta_r^{(k)}(f) \|, \quad k = 0, \ldots, p; \quad r = 0, \ldots, s,$$

состоящую из целых чисел $\beta_r^{(k)}(f)$. Множество функций $f \in A$ с фиксированной матрицей β обозначим через U_β. Мы докажем сейчас, что диаметр каждого множества U_β не более 2ε. В самом деле, положив

$$g = f_1 - f_2, \quad f_1 \in U_\beta, \quad f_2 \in U_\beta,$$

получим в соответствии с (77)

$$|g^{(k)}(x_r)| = |f_1^{(k)}(x_r) - f_2^{(k)}(x_r)| \leqslant \varepsilon_k, \tag{78}$$

в силу (70) и (74)

$$|g^{(p)}(x+h) - g^{(p)}(x)| \leqslant 2C|h|^\alpha. \tag{79}$$

Для любой точки $x \in [a, b]$ найдем x_r, $|h| = |x - x_r| \leqslant \Delta$ и напишем разложение Тейлора

$$g(x) = \sum_{k=0}^{p} \frac{1}{k!} g^{(k)}(x_r) h^k + \frac{h^p}{p!}(g^{(p)}(\xi) - g^{(p)}(x_r)), \quad |\xi - x_r| \leqslant \Delta. \tag{80}$$

Из (78)—(80), (76), (74) получаем

$$|g(x)| \leqslant \sum_{k=0}^{p} \frac{1}{k!} \varepsilon_k \Delta^k + 2C\Delta^{p+\alpha} \leqslant$$

$$\leqslant \frac{\varepsilon}{e}\left(1 + \frac{1}{1!} + \ldots + \frac{1}{p!}\right) + \varepsilon < 2\varepsilon.$$

Таким образом, множества U_β образуют 2ε-покрытие множества A. Остается подсчитать число N *непустых* множеств U_β. Для этого заметим, что число возможных значений индекса $\beta_0^{(k)}$ в силу (69) и (77) будет порядка ε_k^{-1}, а число N' различных видов первой строки

$$\beta_0^{(0)}, \ldots, \beta_0^{(p)}$$

матрицы β:

$$N' \asymp (\varepsilon_0 \cdot \ldots \cdot \varepsilon_s)^{-1} \asymp \varepsilon^{-\omega}, \qquad \omega = \sum_{k=0}^{p}\left(1 - \frac{k}{q}\right) > 0. \tag{81}$$

Предположим теперь, что первые r строк матрицы β

$$\begin{matrix} \beta_0^{(0)} & \ldots & \beta_0^{(p)} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \beta_r^{(0)} & \ldots & \beta_r^{(p)} \end{matrix}$$

уже фиксированы, и рассмотрим число различных возможных (для непустых множеств U_β) значений индекса $\beta_{r+1}^{(k)}$. Так как

$$f^{(k)}(x_{r+1}) = f^{(k)}(x_r) + f^{(k+1)}(x_r)\Delta + \ldots + \frac{1}{(p-k)!} f^{(p)}(x_r)\Delta^{p-k} + R_k, \tag{82}$$

$$R_k = \frac{1}{(p-k)!}\Delta^{p-k}(f^{(p)}(\xi) - f^{(p)}(x_r)), \quad |\xi - x_r| \leqslant \Delta,$$

т. е. в силу (76)

$$\frac{f^{(k)}(x_{r+1})}{\varepsilon_k} = \frac{f^{(k)}(x_r)}{\varepsilon_k} + \frac{f^{(k+1)}(x_r)}{\varepsilon_{k+1}} + \ldots + \frac{1}{(p-k)!}\frac{f^{(p)}(x_r)}{\varepsilon_p} + \frac{R_k}{\varepsilon_k},$$

где

$$\frac{f^{(m)}(x_r)}{\varepsilon_m} = \beta_r^{(m)} + \tau_r^{(m)}, \quad |\tau_r^{(m)}| \leqslant 1,$$

то

$$\beta_{r+1}^{(k)} = \left[\frac{f^{(k)}(x_{r+1})}{\varepsilon_k}\right] = \left[\beta_r^{(k)} + \frac{\beta_r^{(k+1)}}{1!} + \ldots + \frac{\beta_r^{(p)}}{(p-k)!} + D\right],$$

где в силу (70), (82), (74), (76)

$$|D| \leqslant |\tau_r^{(k)}| + \frac{1}{1!}|\tau_r^{(k+1)}| + \ldots + \frac{1}{(p-k)!}|\tau_r^{(p)}| + \frac{|R_k|}{\varepsilon_k} \leqslant$$

$$\leqslant 1 + \frac{1}{1!} + \ldots + \frac{1}{(p-k)!} + \frac{\Delta^{p-k} C \Delta^{\alpha}}{\varepsilon_k} < 2e.$$

Поэтому число различных возможных значений β_{r+1}^k не более $4e + 2$, число же различных возможных $p + 1$ строк матрицы β не более

$$\gamma = (4e + 2)^{p+1}.$$

Легко теперь сообразить, что общее число N непустых множеств U_β допускает оценку

$$N \leqslant N' \gamma^s,$$

из которой получаем в силу (75) и (81)

$$\mathcal{H}_\varepsilon(A) \leqslant \log N = \log N' + s \log \gamma \preccurlyeq \log \frac{1}{\varepsilon} + \left(\frac{1}{\varepsilon}\right)^{1/q} \asymp \left(\frac{1}{\varepsilon}\right)^{1/q},$$

т. е. неравенство (72).

Доказательство неравенства (73). Положим

$$\varphi(y) = \begin{cases} a(1+y)^q(1-y)^q, & |y| \leqslant 1, \\ 0, & |y| \geqslant 1 \end{cases}$$

(рис. 7). Легко проверить, что $\varphi^{(p)}(y)$ удовлетворяют условию Гёльдера степени α. Можно подобрать $a > 0$ так, чтобы эта константа была $\leqslant C$. Выберем

$$\Delta = (\varepsilon/a)^{1/q}$$

и расположим в отрезке $[a, b]$ точки

$$x_r = x_0 + 2r\Delta, \quad r = 0, 1, \ldots, s.$$

Рис. 7

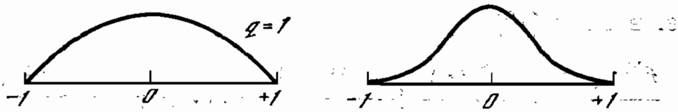
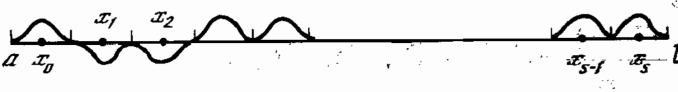


Рис. 8



Легко видеть, что их число может быть взято порядка

$$s \asymp \Delta^{-1} \asymp \varepsilon^{-1/q}.$$

Рассмотрим множество U функций вида

$$g(x) = \sum_{r=0}^{s} \gamma_s \Delta^q \varphi\left(\frac{x - x_r}{\Delta}\right), \quad \gamma_r = \pm 1$$

(пример такой функции показан на рис. 8). Легко проверить, что при достаточно малом ε функции $g(x)$ из U входят в A. Две различные такие функции имеют в какой-либо точке x_r противоположные знаки, по абсолютной же величине равны $a\Delta^q = \varepsilon$. Поэтому их расстояние в метрике (62) больше ε, т. е. множество U является ε-различимым. Число элементов множества U равно

$$N = 2^{s+1}, \quad \log N = s + 1 \asymp \varepsilon^{-1/q},$$

что и доказывает неравенство (73).

Теорема XV доказана.

Доказательство теоремы XIV может идти по тому же пути. Для множества

$$A = F_q^K(C_0, C_1, \ldots, C_p, C)$$

достаточно доказать оценки

$$\mathscr{H}_\varepsilon(A) \preccurlyeq \varepsilon^{-n/q}, \tag{83}$$

$$\mathscr{E}_\varepsilon(A) \succcurlyeq \varepsilon^{-n/q}. \tag{84}$$

Доказательство неравенства (83). Как и в одномерном случае, полагаем

$$\Delta = (\varepsilon/2C)^{1/q}$$

и выбираем в K

$$x^0, x^1, \ldots, x^s,$$

образующие Δ-связную Δ-сеть, т. е. такую Δ-сеть, что от каждой ее точки до любой другой ее точки можно добраться по цепочке ее точек, лежащих попарно на расстоянии не более Δ. Так как компакт связен, то такова будет, как легко видеть, каждая $\Delta/2$-

сеть. Ясно, что s можно выбрать порядка

$$s \asymp \Delta^{-n} \asymp \varepsilon^{-n/q}.$$

Место матрицы β займет теперь набор индексов

$$\beta_r^{(k_1, \ldots, k_n)}, \quad k_1 + k_2 + \ldots + k_n = k,$$

являющихся целыми частями от деления на ε_k коэффициентов

$$f^{(k_1, \ldots, k_n)}(x^r)$$

форм

$$B^{(k)}(x^r, h) = \frac{k_1!\, k_2! \ldots k_n!}{k!} f^{(k_1, k_2, \ldots, k_n)}(x_r)\, h_1^{k_1} \ldots h_n^{k_n}.$$

Сами ε_k приходится теперь брать равными

$$\varepsilon_k = \varepsilon / n^k e \Delta^k,$$

что не меняет их порядка: $\varepsilon_k \asymp \Delta^{q-k}$.

Доказательство того, что диаметр множеств U_β не превосходит 2ε, проводится вполне аналогично одномерному случаю.

Для оценки числа N непустых множеств U_β надо нумерацию точек

$$x^0, x^1, \ldots, x^s$$

выбрать так, чтобы для точки x^{r+1} среди точек $x^0, \ldots, x^r$ всегда существовала точка x^l с расстоянием

$$\| x^l - x^{r+1} \| \leqslant \Delta.$$

При этом условии прежним путем доказывается, что индекс $\beta_{r+1}^{(k_1, \ldots, k_n)}$ при фиксированных группах индексов $\beta_l^{(k_1, \ldots, k_n)}$ может принимать лишь ограниченное некоторой константой γ число значений. Обозначая через P число индексов β с заданным r, получаем оценки [10]

$$N \leqslant N' (\gamma^P)^s, \qquad N' \asymp \varepsilon^{-\omega}, \quad \log N \asymp s \asymp \varepsilon^{-n/q},$$

что уже непосредственно приводит к (83).

Доказательство (84). Подобно одномерному случаю положим

$$\varphi(y) = \begin{cases} a \prod_{l=1}^{n} (1 + y_l)^q (1 - y_l)^q, & \text{если } |y_l| \leqslant 1, \ l = 1, \ldots, m, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Функция φ имеет все частные производные до порядка p включительно и частные производные порядка p от нее удовлетворяют

[10] N' — здесь число различных возможных групп индексов $\beta_0^{(k_1, \ldots, k_m)}$.

условию Гёльдера степени q. Константу a можно выбрать так, что функция φ попадет в класс $F_q^K(C)$.

В компакте K выберем 2Δ-различимое множество точек

$$x^0, x^1, \ldots, x^s,$$

где, как и в одномерном случае,

$$\Delta = (\varepsilon/a)^{1/q}.$$

Очевидно, что s можно сделать порядка

$$s \asymp \Delta^{-n} \asymp \varepsilon^{-n/q}.$$

Рассмотрим теперь множество U функций вида

$$g(x) = \sum_{r=0}^{s} \gamma_r \Delta^q \varphi\left(\frac{x - x^r}{\Delta}\right), \quad \gamma_r = \pm 1.$$

Число элементов в нем равно

$$N = 2^{s+1}, \quad \log N \succcurlyeq s \asymp \varepsilon^{-n/q}.$$

Подобно одномерному случаю доказывается, что U есть ε-различимое множество в A (если ε достаточно мало).

З а м е ч а н и е 1. Для вопросов, рассматриваемых в § 8, имеет некоторый интерес сделать оценки теоремы XV равномерными по длине $T = b - a$ отрезка $[a, b]$. Анализируя ход доказательства теоремы XV, легко обнаружить, что при фиксированных $C_0, C_1, \ldots, C_p, C$ и при $T \geqslant T_0 > 0$ равномерно

$$\mathcal{H}_\varepsilon(A) \asymp \mathcal{E}_\varepsilon(A) \asymp T(1/\varepsilon)^{1/q}.$$

З а м е ч а н и е 2. При доказательстве теоремы XIV мы использовали связность континуума K лишь в доказательстве неравенства (93). Без предположения связности K приходится оценивать при каждом r число различных возможных индексов $\beta_r^{(k_1, \ldots, k_n)}$ как

$$N' \asymp (1/\varepsilon)^{\omega}.$$

Для N получается

$$s \log N' \asymp (1/\varepsilon)^{n/q} \log (1/\varepsilon)$$

и вместо (83)

$$\mathcal{H}_\varepsilon(A) \preccurlyeq (1/\varepsilon)^{n/q} \log (1/\varepsilon).$$

Вместо теоремы XIV получается (без предположения связности K)

$$\left(\frac{1}{\varepsilon}\right)^{n/q} \preccurlyeq \mathcal{E}_\varepsilon(A) \asymp \mathcal{H}_\varepsilon(A) \preccurlyeq \left(\frac{1}{\varepsilon}\right)^{n/q} \log (1/\varepsilon).$$

Замечание 3. Множество F_q^K можно рассматривать как линейное пространство. Естественная топология в нем определяется нормой

$$\| f \| = \sum_{k=0}^{p} \sup_{x \in K} \| B_f^{(k)}(x) \| + \sup_{x \in K} \| R_f(x) \|,$$

где

$$\| B_f^{(k)}(x) \| = \sup_{x+h \in K} \frac{| B_k(h, x) |}{\| h \|^k},$$

$$\| R_f(x) \| = \sup_{x+h \in K} \frac{| R(h, x) |}{\| h \|^q}.$$

Из теоремы XIV легко получается следствие: *для любого ограниченного в F_q^K множества A, имеющего хотя бы одну внутреннюю точку*, функции $\mathscr{H}_\varepsilon(A)$ и $\mathscr{E}_\varepsilon(A)$, вычисленные в метрике (62) (!), имеют порядок

$$\mathscr{H}_\varepsilon(A) \asymp \mathscr{E}_\varepsilon(A) \asymp \varepsilon^{-n/q}. \mid$$

§ 6
ε-ЭНТРОПИЯ КЛАССА ДИФФЕРЕНЦИРУЕМЫХ ФУНКЦИЙ В МЕТРИКЕ L^2

Настоящий параграф примыкает по тематике к предыдущему параграфу, хотя по методам доказательств является по существу продолжением § 4. При написании его авторы пользовались результатами исследований, принадлежащих В. И. Арнольду.

Обозначим через $P_\alpha^{[0, 2\pi]}(C)$, $\alpha > 0$, или просто $P_\alpha(C)$ *класс вещественных функций $f(x)$, периодических с периодом, равным 2π, в среднем равных нулю и имеющих в L^2 производную порядка α* (в смысле Вейля), *равномерно ограниченную в среднем константой C*. Норму в этом классе рассмотрим в L^2:

$$\| f \| = \frac{1}{\pi} \int_0^{2\pi} | f(x) |^2 dx.$$

Напомним, что функция $f^{(\alpha)}$ называется производной порядка α в смысле Вейля функции $f \in L^2$:

$$f(x) \backsim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),$$

если ей соответствует ряд Фурье

$$f^{(\alpha)}(x)=\sum_{k=1}^{\infty}k^{\alpha}\left(\left(a_k\cos\frac{\pi}{2}\alpha+b_k\sin\frac{\pi}{2}\alpha\right)\cos kx+\right.$$
$$\left.+\left(b_k\cos\frac{\pi}{2}\alpha-a_k\sin\frac{\pi}{2}\alpha\right)\sin kx\right),$$

сходящийся в среднем [11].

Равномерная ограниченность в среднем $f^{(\alpha)}$ для $f\in P_\alpha(C)$ означает, что

$$\|f^{(\alpha)}\|=\left(\sum_{k=1}^{\infty}k^{2\alpha}(a_k^2+b_k^2)\right)^{1/2}\leqslant C. \tag{85}$$

Пространства L^2 и l^2, как известно, изометричны. Рассмотрим ортонормированный базис в l^2, состоящий из векторов $\{e_1,\ldots,e_n,\ldots\}$ таких, что e_1 соответствует $\sqrt{2}/2$, $e_{2k}\leftrightarrow\sin kx$, $e_{2k+1}\leftrightarrow\cos kx$, $k\geqslant 1$, при некотором изометрическом отображении L^2 в l^2. Нетрудно убедиться, что классу $P_\alpha(C)$ будет соответствовать в этом базисе множество векторов $x=(x_1,\ldots,x_n,\ldots)$, координаты которых удовлетворяют неравенству:

$$x_1=0,\quad \sum_{k=2}^{\infty}\left[\frac{k}{2}\right]^{2\alpha}x_k^2\leqslant C^2. \tag{86}$$

Это множество, которое мы будем также обозначать через $P_\alpha(C)$, является эллипсоидом с полуосями $a_k=C/[k/2]$ [12].

Далее мы рассмотрим частный случай эллипсоидов в l^2, полуоси a_k которых равны B/k^α $(k=1,2,\ldots)$. Их мы обозначим через $Э_\alpha(B)$. Ввиду того что

$$\frac{C}{(k/2)^\alpha}\leqslant\frac{C}{[k/2]^\alpha}\leqslant\frac{C}{((k-1)/2)^\alpha},\quad k=2,3,\ldots,$$

получим

$$Э'_\alpha(C\cdot 2^\alpha)\subset P_\alpha(C)\subset Э_\alpha(C\cdot 2^\alpha), \tag{87}$$

где $Э'_\alpha(C\cdot 2^\alpha)$ — эллипсоид, у которого $a_1=0$. Включения (87) позволяют нам оценивать $\mathscr{H}_\varepsilon(P_\alpha)$ и $\mathscr{E}_\varepsilon(P_\alpha)$ через $\mathscr{H}_\varepsilon(Э_\alpha)$ и

---

[11] Для целых $\alpha=n$ $f^{(n)}(x)$ в смысле Вейля совпадает с обычной производной d^nf/dx^n.

[12] *Эллипсоидом* в l^2 называют множество x, для которых

$$\sum_{k=1}^{\infty}x_k^2/a_k^2\leqslant 1;$$

a_k называют *полуосями* этого эллипсоида.

$\mathscr{E}_\varepsilon(\mathcal{E}_\alpha)$. Для последних мы покажем, что

$$\mathscr{E}_{2\varepsilon}(\mathcal{E}_\alpha(B)) \succsim \left(\frac{B}{2\varepsilon}\right)^{1/\alpha} \alpha \log e, \tag{88}$$

$$\mathscr{H}_\varepsilon(\mathcal{E}_\alpha(B)) \precsim \left(\frac{\sqrt{2}\,B}{\varepsilon}\right)^{1/\alpha} \alpha \log e\left(1 + \frac{\gamma}{\alpha}\right), \tag{89}$$

где γ — ограниченная константа, не зависящая от α. Тогда из (87) получится

Теорема XVI.

$$\left(\frac{C}{2\varepsilon}\right)^{1/\alpha} 2\alpha \log e \precsim \mathscr{E}_{2\varepsilon}(P_\alpha(C)) \leqslant \mathscr{H}_\varepsilon(P_\alpha(C)) \precsim$$
$$\precsim \left(\frac{\sqrt{2}\,C}{\varepsilon}\right)^{1/\alpha} 2\alpha \log e\left(1 + \frac{\gamma}{\alpha}\right), \tag{90}$$

где γ — ограниченная константа, не зависящая от α.

Для $\alpha \geqslant \alpha_0$ соотношение (90) можно переписать так:

$$\left(\frac{C}{\varepsilon}\right)^{1/\alpha} 2\alpha \log e\left(1 + \frac{\gamma_1}{\alpha}\right) \precsim \mathscr{E}_{2\varepsilon}(P_\alpha(C)) \leqslant \mathscr{H}_\varepsilon(P_\alpha(C)) \precsim$$
$$\precsim \left(\frac{C}{\varepsilon}\right)^{1/\alpha} 2\alpha \log e\left(1 + \frac{\gamma_2}{\alpha}\right), \tag{91}$$

где γ_1 и γ_2 — ограниченные константы.

Результат, полученный нами в теореме XVI, интересен в том отношении, что для пространства дифференцируемых функций в метрике L^2 удается найти для $\mathscr{E}_{2\varepsilon}$ и $\mathscr{H}_\varepsilon$ достаточно узкие неравенства (90), асимптотически сходящиеся друг к другу при α (порядке гладкости), стремящемся к бесконечности (см. (91)). Это тем более интересно, если учесть, что подобного результата в метрике C не получено (см. § 5).

Итак, нам остается получить (88) и (89).

Пусть $\mathcal{E}_\alpha^n(B)$ означает у нас $\mathcal{E}_\alpha(B) \cap D_2^n$, где D_2^n есть n-мерное евклидово пространство, натянутое на первые n осей эллипсоида $\mathcal{E}_\alpha(B)$. Предположив, что $\mathscr{M}_{2\varepsilon}$ есть максимальное число непересекающихся n-мерных шаров S_2^n радиуса ε, центры которых расположены в $\mathcal{E}_\alpha^n(B)$, получим, что

$$\mathscr{M}_{2\varepsilon} \geqslant \frac{\mu^n(\mathcal{E}_\alpha^n(B))}{\mu^n(2\varepsilon S_2^n)} = \frac{\mu^n(\mathcal{E}_\alpha^n(B))}{(2\varepsilon)^n \mu^n(S_2^n)}, \tag{92}$$

так как ввиду предположенной максимальности числа $\mathscr{M}_{2\varepsilon}$, сферы радиуса 2ε должны покрыть все $\mathcal{E}_\alpha^n(B)$. Положим теперь

$$B/n^\alpha < 2\varepsilon \leqslant B/(n-1)^\alpha,$$

т. е.

$$n = (B/2\varepsilon)^{1/\alpha} + O(1), \tag{93}$$

и, воспользовавшись соотношениями

1) $\mu^n(\mathcal{E}^n_\alpha(B)) = \mu^n(S^n_2)\, B^n (n!)^{-\alpha}$,

2) $\log n! < (n + {}^1/_2) \log n - n \log e$

(формула Стирлинга), получим из (92) и (93), что

$$\mathscr{E}_{2\varepsilon}(\mathcal{E}_\alpha(B)) \geqslant \log \mathscr{M}_{2\varepsilon} \geqslant n \log \frac{B}{2\varepsilon} - \alpha n \log n -$$
$$- \frac{\alpha}{2} \log n + \alpha n \log e \geqslant \left(\frac{B}{2\varepsilon}\right)^{1/\alpha} \alpha \log e + O\left(\log \frac{1}{\varepsilon}\right) \succsim$$
$$\succsim \left(\frac{B}{2\varepsilon}\right)^{1/\alpha} \alpha \log e.$$

Для получения (89) подберем m так, чтобы

$$\frac{B}{m^\alpha} \leqslant \frac{\sqrt{2}}{2} \varepsilon < \frac{B}{(m-1)^\alpha},$$

т. е.

$$m = (\sqrt{2}B/\varepsilon)^{1/\alpha} + O(1). \tag{94}$$

Рассмотрим эллипсоид $\mathcal{E}^m_\alpha(B)$. Нетрудно видеть, что расстояние любой точки $\mathcal{E}_\alpha(B)$ до $\mathcal{E}^m_\alpha(B)$ меньше или равно $(\sqrt{2}/2)\,\varepsilon$ и, следовательно, любая $(\sqrt{2}/2)$ ε-сеть для $\mathcal{E}^m_\alpha(B)$ является ε-сетью для $\mathcal{E}_\alpha(B)$.

Строим в D^m_2 кубическую решетку диаметра $\varepsilon/\sqrt{2m}$. (Определение решетки см. в § 4.)

Пусть N — число кубов диаметра $\varepsilon/\sqrt{2m}$ [13] с центрами в точках решетки, пересекающихся с $\mathcal{E}^m_\alpha(B)$. Кубы эти лежат, очевидно, в $\varepsilon\sqrt{2}/2$-окрестности U эллипсоида $\mathcal{E}^m_\alpha(B)$. U заключается внутри гомотетичного с $\mathcal{E}^m_\alpha(B)$ эллипсоида с наименьшей осью, равной $B/m^\alpha + (\sqrt{2}/2)\,\varepsilon$, или коэффициентом подобия, меньшим 2. Таким образом, объем всех кубов решетки не превосходит объема этого эллипсоида, откуда получаем неравенство

$$N \leqslant \frac{\mu(S^m_2)\, B^m}{(m!)^\alpha} 2^m \left(\frac{\sqrt{2m}}{\varepsilon}\right)^m. \tag{95}$$

Теперь мы воспользуемся тем, что

1) $\mu^m(S_2{}^m) = \dfrac{\pi^{m/2}}{\Gamma(m/2+1)}$ (объем сферы) (см. [24, т. 3, с. 474]),

2) $\log \Gamma(a) = (a - {}^1/_2) \log a - a \log e + O(1)$ (см. [24, т. 2,

[13] При этом диаметр куба в евклидовой метрике равен $\varepsilon\sqrt{2}/2$.

с. 819]), и получим из (94) и (95)

$$\mathcal{H}_\varepsilon(\partial_\alpha(B)) \leqslant \log N \leqslant m \log \frac{\sqrt{2}\,B}{\varepsilon} - \alpha m \log m + m(\alpha \log e + O(1)) + O(\log m) \leqslant \left(\frac{\sqrt{2}\,B}{\varepsilon}\right)^{1/\alpha} \alpha \log e \left(1 + \frac{\gamma}{\alpha}\right),$$

что и требовалось доказать.

Арнольд в своих исследованиях для оценки величины $\mathcal{H}_\varepsilon$ сверху воспользовался результатами Давенпорта и Ватсона, о которых мы упоминали в § 4. Естественно, что при этом автором были получены более точные оценки, чем у нас, так как решетки, построенные Давенпортом и Ватсоном, являются гораздо более экономными, чем кубические. Приведем без доказательства неравенства, полученные Арнольдом:

$$\mathcal{H}_\varepsilon(\partial_\alpha(B)) \leqslant \left(\frac{\sqrt{2}\,B}{\varepsilon}\right)^{1/\alpha} \alpha \log e \left(1 + \frac{1{,}2}{\alpha}\right) + \alpha \log e.$$

§ 7
ε-ЭНТРОПИЯ КЛАССОВ АНАЛИТИЧЕСКИХ ФУНКЦИЙ

Кроме заметок авторов [6] и [7], содержащих оценки ε-энтропии типа слабой эквивалентности, этот параграф основан на работах А. Г. Витушкина [3] и В. Д. Ерохина [4]. Доказательства формулируемых ниже лемм II и III содержат в общей форме методы, разработанные А. Г. Витушкиным; теорема XVIII специально приспособлена к системе функций, введенных В. Д. Ерохиным. Частично результаты, полученные Ерохиным, одновременно и другими методами были доказаны К. И. Бабенко [1]. Теорема XX принадлежит А. Г. Витушкину [3], теорема XXI — В. М. Тихомирову.

1. Будем предполагать, что в линейном нормированном пространстве F дана система элементов

$$\varphi_{k_1 k_2 \ldots k_s}, \quad k_r = 0, 1, 2, \ldots, \quad r = 1, 2, \ldots, s,$$

и соответствующая система линейных функционалов

$$c_{k_1 k_2 \ldots k_s}$$

так, что каждый элемент f представляется в виде [14]

$$f = \sum_{k \in K} c_k \varphi_k,$$

[14] В этом параграфе k означает у нас целочисленный неотрицательный вектор $k = (k_1, \ldots, k_s)$. Через K мы обозначаем множество всех таких векторов. Вектор $h = (h_1, \ldots, h_s)$ мы называем положительным, если его компоненты строго больше нуля. Запись h^k означает, что для s-мерного век-

причем существует постоянная C_1 такая, что

$$\sup_{k\in K}|c_k|\leqslant\|f\|\leqslant C_1\sum_{k\in K}|c_k|. \tag{96}$$

Имеют место две основные леммы.

Лемма II. *Пусть $A\subset F$ и при этом существуют постоянная C_2 и положительный вектор $h=(h_1,\ldots,h_s)$ такие, что из $f\in A$ вытекает*

$$|c_k|\leqslant C_2e^{-(h,k)}. \tag{97}$$

Тогда

$$\mathscr{H}_\varepsilon(A)\leqslant\frac{2(\log(1/\varepsilon))^{s+1}}{(s+1)!\,(\log e)^s h^\theta}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right). \tag{98}$$

Лемма III. *Пусть при заданных C_3 и h из существования такого положительного вектора Δ, что*

$$|c_k|\leqslant C_3\,\Delta^\theta e^{-(h+\Delta,k)}, \tag{99}$$

вытекает принадлежность элемента

$$f=\sum_{k\in K}c_k\varphi_k$$

к $A\subset F$. Тогда

$$\mathscr{E}_{2\varepsilon}(A)\geqslant\frac{2(\log(1/\varepsilon))^{s+1}}{(s+1)!\,(\log e)^s h^\theta}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right). \tag{100}$$

Мы говорим, что *для множества $A\subset F$ выполнена система условий* (I), если для него выполняются три перечисленные выше условия (96), (97) и (99).

Неравенства (98) и (100) в совокупности, очевидно, дают следующую теорему.

Теорема XVII. *Если для множества A выполнена система условий* (I), *то*

$$\mathscr{H}_\varepsilon(A)=\frac{2}{(s+1)!\,(\log e)^s h^\theta}\left(\log\frac{1}{\varepsilon}\right)^{s+1}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right), \tag{101}$$

$$\mathscr{E}_\varepsilon(A)=\frac{2}{(s+1)!\,(\log e)^s h^\theta}\left(\log\frac{1}{\varepsilon}\right)^{s+1}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right).$$

тора h и $k\in K$ имеет место соотношение

$$h^k=h_1^{k_1}\ldots h_s^{k_s}=\prod_{i=1}^s h_i^{k_i}.$$

Скалярное произведение (h,k) имеет обычный смысл $\sum_{i=1}^s h_ik_i$; θ обозначает единичный вектор $(1,\ldots,1)$, так что

$$h^\theta=h_1\ldots h_s=\prod_{i=1}^s h_i.$$

Мы скажем, что *для множества $A \subset F$ выполнена система условий* (II), если для любого положительного вектора α существуют константы $C_1(\alpha)$, $C_2(\alpha)$ и $C_3(\alpha)$ такие, что:

1) $$\sup_{k\in K} |c_k| \leqslant \| f \| \leqslant C_1(\alpha) \sum_{k\in K} |c_k|, \tag{102}$$

2) из $f \in A$ вытекает

$$|c_k| \leqslant C_2(\alpha)\, e^{-(h-\alpha,\, k)},$$

3) из неравенств

$$|c_k| \leqslant C_3(\alpha)\, e^{-(h+\alpha,\, k)}$$

вытекает принадлежность

$$f = \sum_{k\in K} c_k \varphi_k$$

к множеству $A \subset F$.

Из неравенств (98) и (100) легко выводится

Теорема XVIII. *Если для множества $A \subset F$ выполнена система условий* (II), *то*

$$\mathscr{E}_\varepsilon(A) \sim \mathscr{H}_\varepsilon(A) \sim \frac{2(\log(1/\varepsilon))^{s+1}}{(s+1)!\,(\log e)^s\, h^\theta}. \tag{103}$$

Действительно, зафиксировав α, на основании неравенства (98) мы будем иметь, что

$$\overline{\lim_{\varepsilon\to 0}} \frac{\mathscr{H}_\varepsilon(A)}{(\log(1/\varepsilon))^{s+1}} \leqslant \frac{2}{(s+1)!\,(\log e)^s\,(h-\alpha)^\theta}. \tag{104}$$

На основании неравенства (100) мы получим, что

$$\underline{\lim_{\varepsilon\to 0}} \frac{\mathscr{E}_{2\varepsilon}(A)}{(\log(1/\varepsilon))^{s+1}} \geqslant \frac{2}{(s+1)!\,(\log e)^s\,(h+\alpha)^\theta}. \tag{105}$$

Неравенства (104) и (105) в совокупности дают (103), так как α произвольно и

$$\mathscr{E}_{2\varepsilon}(A) \leqslant \mathscr{H}_\varepsilon(A) \leqslant \mathscr{E}_\varepsilon(A).$$

Прежде чем переходить к доказательству лемм II и III, сделаем следующее

Примечание. В теореме XVII множество K состояло из неотрицательных векторов k. Из последующего изложения читатель легко сообразит, чтó изменится в доказательствах в случае, когда имеем дело с произвольными векторами $k = (k_1, \ldots, k_s)$, $-\infty < k_i < \infty$. Сформулируем окончательный результат. Пусть $|k|$ означает вектор с координатами $(|k_1|, \ldots, |k_s|)$, а $\widetilde{K}$ — множество всех целочисленных векторов.

Т е о р е м а XIX. *Пусть для множества A выполнена система условий* (I′), *т. е.*

1) $$\sup_{k\in\widetilde{K}}|c_k|\leqslant\|f\|\leqslant C_1\sum_{k\in\widetilde{K}}|c_k|, \tag{106}$$

2) *из* $f\in A$ *вытекает*

$$|c_k|\leqslant C_2e^{-(h,\,|k|)}, \tag{107}$$

3) *при заданной константе* C_3 *для любого положительного вектора* Δ *из осуществления неравенств*

$$|c_k|\leqslant C_3\Delta^\theta e^{-(h+\Delta,\,|k|)}$$

вытекает принадлежность элемента

$$f=\sum_{k\in\widetilde{K}}c_k\varphi_k$$

к $A\subset F$.

Тогда

$$\mathscr{H}_\varepsilon(A)=\frac{2^{s+1}}{(s+1)!\,(\log e)^s\,h^\theta}\left(\log\frac{1}{\varepsilon}\right)^{s+1}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right), \tag{108}$$

$$\mathscr{E}_\varepsilon(A)=\frac{2^{s+1}}{(s+1)!\,(\log e)^s\,h^\theta}\left(\log\frac{1}{\varepsilon}\right)^{s+1}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right).$$

Теорема XIX специально приспособлена к рядам Фурье.

2. Итак, нам остается доказать леммы II и III.

Рассмотрим множество S_n целочисленных векторов $k\in K$, для которых

$$(h,\,k)\leqslant n. \tag{109}$$

Через $\bar{S}_n$ обозначим $K\setminus S_n$ (рис. 9).

Из условия (96) мы получим, что

$$\|f\|\leqslant C_1\sum_{k\in S_n}|c_k|+C_1\sum_{k\in\bar{S}_n}|c_k|=C_1\sum_{k\in S_n}|c_k|+R_n. \tag{110}$$

Неравенства (97) дадут

$$|R_n|\leqslant C_1C_2\sum_{k\in\bar{S}_n}e^{-(h,\,k)}. \tag{111}$$

Предположив, далее, $n\geqslant(\max_i h_i/\min_i h_i)+1=\lambda$, покажем, что

$$\sum_{k\in\bar{S}_n}e^{-(h,\,k)}\leqslant\int\limits_{\substack{(h,\,x)\geqslant n-\lambda\\ x\geqslant 0}}e^{-(h,x)}\,dx. \tag{112}$$

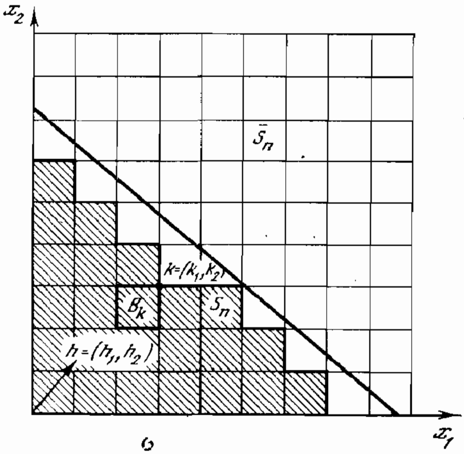


Рис. 9

Пусть B_k — куб со стороной единица:

$$B_k = \{x \mid k_r - 1 \leqslant x_r \leqslant k_r,\; r = 1, \ldots, s\}.$$

Назовем k максимальной вершиной куба B_k. Очевидно, что разность $k - x$ является неотрицательным вектором для любой точки куба x. Легко понять, что объединение всех B_k для $k \in K$ покрывает множество всех неотрицательных x. Функция $e^{-(h,x)}$ на кубе B_k имеет минимум в точке k. Действительно, проведя через любую точку $x_0 \in B_k$ гиперплоскость $(h, x) = (h, x_0)$, мы получим, что

$$e^{-(h,x)} = e^{-(h,k)} e^{(h,k-x)} \geqslant e^{-(h,k)}, \text{ ибо } (h, k - x) \geqslant 0,$$

так как h — положительный вектор, а k — максимальная вершина.

Выберем теперь минимальное λ так, чтобы гиперплоскость $(h, x) = n - \lambda$ не имела бы общих точек с B_k для $k \in S_n$. Тогда согласно доказанному

$$\sum_{k \in S_n} e^{-(h,k)} \leqslant \int\limits_{\substack{(h,x) \geqslant n-\lambda \\ x \geqslant 0}} e^{-(h,x)}\, dx.$$

Легко подсчитать, что в качестве λ достаточно взять величину $(\max_i h_i / \min_i h_i) + 1$. Для нас существенна только ограниченность λ.

Интеграл (112) подсчитывается совершенно элементарно [15]:

$$\int\limits_{\substack{(h,x) \geqslant n-\lambda \\ x \geqslant 0}} e^{-(h,x)}\, dx = \frac{1}{h^\theta} \int\limits_{n-\lambda}^{\infty} e^{-\xi} \frac{\xi^{s-1} d\xi}{(s-1)!} \leqslant B e^{-n} n^s,$$

где B — ограниченная константа. Подберем теперь минимальное n так, чтобы

$$C_1 C_2 B e^{-n} n^s \leqslant \varepsilon/2. \tag{113}$$

---

[15] Напомним, что $h^\theta = \prod_{i=1}^{s} h_i$.

Отсюда

$$n = \frac{\log(1/\varepsilon)}{\log e} + O\left(\log\log\frac{1}{\varepsilon}\right). \tag{114}$$

Оценка (113) означает, что $|R_n|$ в (110) не превосходит $\varepsilon/2$. Аппроксимируем теперь остающуюся конечномерную часть:

$$\sum_{k\in S_n} c_k\varphi_k.$$

Обозначим через P_n число векторов $k \in S_n$. Из предыдущего нетрудно понять, что для P_n выполняются неравенства

$$\frac{(n-\lambda)^s}{s!\,h^\theta} = \int\limits_{\substack{(h,x)\leqslant n-\lambda\\ x\geqslant 0}} dx \leqslant P_n \leqslant \int\limits_{\substack{(h,x)\leqslant n\\ x\geqslant 0}} dx = \frac{n^s}{s!\,h^\theta},$$

откуда

$$P_n = \frac{n^s}{s!\,h^\theta}\left(1 + \frac{O(1)}{n}\right). \tag{115}$$

Аппроксимируем, далее, координату $c_k = \alpha_k + i\beta_k$ с точностью до ε/P_nC_1 величиной $\alpha_k' + i\beta_k'$, полагая

$$\begin{aligned} \alpha_k' &= \left[\frac{\sqrt{2}\,C_1P_n\alpha_k}{\varepsilon}\right]\frac{\varepsilon}{\sqrt{2}\,P_nC_1} = m_k^1\frac{\varepsilon}{\sqrt{2}\,P_nC_1},\\ \beta_k' &= \left[\frac{\sqrt{2}\,C_1P_n\beta_k}{\varepsilon}\right]\frac{\varepsilon}{\sqrt{2}\,P_nC_1} = m_k^2\frac{\varepsilon}{\sqrt{2}\,P_nC_1}. \end{aligned} \tag{116}$$

Каждому элементу f таким образом сопоставится набор индексов

$$U = \{m_k^i\},\quad k \in S_n,\quad i = 1,\ 2.$$

Для двух элементов f_1 и f_2 из A, соответствующих одному и тому же набору U, имеем

$$\| f_1 - f_2 \| \leqslant C_1 \sum_{k\in S_n} |c_k^1 - c_k^2| + 2R_n \leqslant \sum_{k\in S_n}\frac{\varepsilon}{P_n} + \varepsilon = 2\varepsilon$$

(см. соотношения (110) и (116)), т. е. множество элементов ε-покрытия A не превосходит числа всех наборов U, соответствующих A. Нам теперь осталось только оценить число N этих наборов. Ввиду того что согласно (97) $\max\{\alpha_k, \beta_k\} \leqslant C_2\,e^{-(h,k)}$, из (116) получим, что

$$|m_i^k| \leqslant \frac{\sqrt{2}\,C_1C_2P_n}{\varepsilon}\,e^{-(h,k)} = N_k,\quad i = 1, 2. \tag{117}$$

Отсюда следует, что

$$\log N = \log(1/\varepsilon) - (h,\ k)\log e + \log P_n + D_1;$$

D_1 — здесь ограниченная константа, не зависящая от k. Число N, таким образом, оценивается величиной

$$\prod_{k\in S_n}(2N_k+1)^2,$$

откуда

$$\mathscr{H}_\varepsilon(A)\leqslant \log N\leqslant 2\sum_{k\in S_n}\log(2N_k+1)=$$

$$=2P_n\log\frac{1}{\varepsilon}-2\sum_{k\in S_n}(k,h)\log e+2P_n\log P_n+O(P_n).$$

Величина $\sum\limits_{k\in S_n}(k,h)$ по соображениям, аналогичным примененным нами для оценки R_n, не будет превосходить интеграла

$$\int\limits_{\substack{(h,x)\leqslant n-\lambda\\ x\geqslant 0}}(h,x)\,dx=\frac{s\,(n-\lambda)^{s+1}}{(s+1)!\,h^\theta}=\frac{sn^{s+1}}{(s+1)!\,h^\theta}\left(1+\frac{O(1)}{n}\right),$$

отсюда окончательно, использовав (114) и (115), получим

$$\mathscr{H}_\varepsilon(A)\leqslant\frac{2}{(s+1)!\,(\log e)^s\,h^\theta}\left(\log\frac{1}{\varepsilon}\right)^{s+1}+O\left(\left(\log\frac{1}{\varepsilon}\right)^s\log\log\frac{1}{\varepsilon}\right).$$

Доказательство леммы III. Пусть $\varepsilon>0$. Положим

$$h'=h+\Delta,$$

где вектор Δ имеет координаты

$$\Delta_i=\frac{h_i}{\log(1/\varepsilon)},\quad i=1,2,\ldots,s. \tag{118}$$

Подберем максимальное m так, чтобы

$$C_3\Delta^\theta e^{-m}=\frac{C_3h^\theta}{(\log(1/\varepsilon))^s}e^{-m}\geqslant 2\sqrt{2}\,\varepsilon, \tag{119}$$

где ε взято из условия леммы III.

Обозначим через S_m множество векторов k из K, для которых $(h',\ k)\leqslant m$.

Рассмотрим множество $\{\Psi\}$ функций $\Psi=\sum\limits_{k\in K}c_k\varphi_k$, для которых

α) $c_k=0,\quad k\notin S_m$;

β) $c_k=(s_k^1+is_k^2)\,2\varepsilon,\quad k\in S_m$; (120)

γ) s_k^i $(i=1,\,2)$ — любые целые числа такие, что

$$|s_k^i|\leqslant\frac{C_3\Delta^\theta e^{-(h',k)}}{2\sqrt{2}\,\varepsilon}=M_k. \tag{121}$$

Неравенства (121) означают согласно условию (97) леммы III, что множество $\{\Psi\} \subseteq A$.

По построению $\{\Psi\}$ из (96) вытекает 2ε-различимость множества $\{\Psi\}$. Согласно (120) каждый элемент из $\{\Psi\}$ определяется набором индексов

$$U = \{s_k^i\}, \quad k \in S_m, \quad i = 1, 2,$$

а число всех U согласно (121) превосходит

$$M = \prod_{k \in S_m} M_k^2.$$

Из (121) мы получим, что

$$\log M_k = \log (1/\varepsilon) - (h', k) \log e + \log C_3 \Delta^\theta + D_2, \tag{122}$$

где D_2 — ограниченная константа, не зависящая от k. Отсюда, вспомнив (118), преобразуем (122) к виду

$$\log M_k = \log \frac{1}{\varepsilon} - (h', k) \log e + O\left(\log \log \frac{1}{\varepsilon}\right).$$

Из (119) получим

$$m = \frac{\log (1/\varepsilon)}{\log e} + O\left(\log \log \frac{1}{\varepsilon}\right),$$

и окончательно [16]

$$\begin{aligned}
\mathscr{E}_{2\varepsilon}(A) \geqslant \log M = 2P_m \log \frac{1}{\varepsilon} - 2 \sum_{k \in S_m} (h', k) \log e + \\
+ O\left(P_m \log \log \frac{1}{\varepsilon}\right) = \frac{2}{(s+1)!\,(\log e)^s (h')^\theta} \left(\log \frac{1}{\varepsilon}\right)^{s+1} + \\
+ O\left(\left(\log \frac{1}{\varepsilon}\right)^s \log \log \frac{1}{\varepsilon}\right) = \frac{2}{(s+1)!\,(\log e)^s h^\theta} \left(\log \frac{1}{\varepsilon}\right)^{s+1} + \\
+ O\left(\left(\log \frac{1}{\varepsilon}\right)^s \log \log \frac{1}{\varepsilon}\right).
\end{aligned}$$

Доказательство леммы III закончено.

3. Переходим теперь к применениям основных теорем XVII и XVIII этого параграфа.

Всюду в дальнейшем мы рассматриваем классы $\mathscr{A}_G^K$ аналитических функций $f(z_1, \ldots, z_s)$, заданных на ограниченном континууме K, допускающих аналитическое продолжение в область $G \supset K$ s-мерного комплексного пространства. При этом все функции $f \in A_G^K(C)$ предполагаются равномерно ограниченными в G константой C.

[16] P_m имеет то же определение, что и в лемме II. Для него выполнено соотношение (115).

Классы A_G^K мы рассматриваем как *метрические пространства* в равномерной метрике

$$\|f\| = \max |f(z_1, \ldots, z_s)|, \quad (z_1, \ldots, z_s) \in K.$$

Сначала рассмотрим классы аналитических функций одного комплексного переменного.

1) Класс $A_{K_R}^{\overline{K}_r}(C)$, $\overline{K}_r$ — замкнутый круг $|z| \leqslant r$, K_R — круг $|z| < R$, $R > r$.

Функции класса $A_{K_R}^{\overline{K}_r}$ разлагаются *в ряд Тейлора*

$$f(z) = a_0 + a_1 z + \ldots + a_n z^n + \ldots = c_0 + \frac{c_1}{r} z + \ldots + \frac{c_n}{r^n} z^n + \ldots, \tag{123}$$

где $c_k = a_k r^k$. Множество K здесь состоит из целых чисел $k = 0, 1, \ldots$ Положим $e^h = R/r$, или $h = \log(R/r)/\log e$. Из неравенств Коши получим

$$|c_k| \leqslant C(r/R)^k = Ce^{-hk}. \tag{124}$$

Из (123) непосредственно следует, что если

$$|c_k| \leqslant C'\Delta e^{-(h+\Delta)k} = C'\Delta (r/R)^k e^{-\Delta k}, \tag{125}$$

то ряд (123) равномерно сходится в K_R, причем

$$|f(z)| \leqslant C'\Delta \sum_{k=0}^{\infty} e^{-\Delta k} = \frac{C'\Delta}{1 - e^{-\Delta}}.$$

Положим C' таким, чтобы

$$C'\Delta/(1 - e^{-\Delta}) \leqslant C \tag{126}$$

для всех $\Delta \leqslant \Delta_0$.

Тогда

$$f(z) \in A_{K_R}^{\overline{K}_r}(C).$$

Непосредственно из (123) получим

$$\|f\| = \max_{z \in \overline{K}_r} |f(z)| \leqslant \sum_{k=0}^{\infty} |a_k| r^k = \sum_{k=0}^{\infty} |c_k|. \tag{127}$$

Кроме того,

$$|a_k| = \left| \frac{1}{2\pi i} \int_{Z_r} \frac{f(\xi)}{\xi^{k+1}} d\xi \right| \leqslant \frac{\max\limits_{\xi \in Z_r} |f(\xi)|}{r^k} = \frac{\|f\|}{r^k}, \tag{128}$$

где Z_r — окружность радиуса r. Таким образом, из соотношений

(124), (125), (127), (128) вытекает, что $A_{K_R}^{\overline{K}_r}(C)$ удовлетворяет системе условий (I) (где $C_1 = 1$, $C_2 = C$, $C_3 = C'$ берется из (126)), и, следовательно [17],

$$\mathscr{H}_\varepsilon (A_{K_R}^{\overline{K}_r}(C)) = \frac{(\log (1/\varepsilon))^2}{\log (R/r)} + O\left(\log \frac{1}{\varepsilon} \log\log \frac{1}{\varepsilon}\right). \qquad (129)$$

2) Класс $\mathcal{A}_h(C)$ функций, периодических на D с периодом, равным 2π, аналитических и ограниченных константой C в полосе $|\operatorname{Im} z| < h$. Его мы подробно разбирали в § 2. Для этого класса выполняются условия (I′) и, таким образом, согласно формуле (108)

$$\mathscr{H}_\varepsilon (\mathcal{A}_h(C)) = \frac{2}{h \log e} \left(\log \frac{1}{\varepsilon}\right)^2 + O\left(\log \frac{1}{\varepsilon} \log\log \frac{1}{\varepsilon}\right). \qquad (130)$$

3) Класс $A_{\mathcal{Э}_\lambda}^{\Delta}(C)$. Через Δ мы здесь обозначили $[-1, 1]$, а через $\mathcal{Э}_\lambda$ — эллипс с суммой полуосей, равной λ, и фокусами в точках ± 1.

Функции этого класса разлагаются в ряд *по полиномам Чебышева*:

$$f(z) = \sum_{k=0}^{\infty} a_k T_k(z).$$

Доказательство применимости теоремы XVII можно было бы получить самостоятельно, но мы его получим как частный случай из рассмотрений следующего класса функций $A_{C_R(K)}^{K}(C)$.

Для класса $A_{\mathcal{Э}_\lambda}^{\Delta}(C)$ $h = (\log \lambda^{-1})/\log e$, т. е.

$$\mathscr{H}_\varepsilon (A_{\mathcal{Э}_\lambda}^{\Delta}(C)) = \frac{(\log (1/\varepsilon))^2}{\log \lambda^{-1}} + O\left(\log \frac{1}{\varepsilon} \log\log \frac{1}{\varepsilon}\right). \qquad (131)$$

4) Пусть K означает произвольный ограниченный континуум, содержащий более одной точки. Пусть G_∞ — та из смежных с ним областей, которая содержит точку $z = \infty$. Это — односвязная область расширенной плоскости, граница которой Γ является частью континуума K. Отобразим G_∞ на внешность единичного круга плоскости w. Прообраз окружности $|w| = R$, $R > 1$ при этом отображении назовем *R-й линией уровня континуума K* и обозначим через $C_R(K)$. Рассмотрим класс $A_{C_R(K)}^{K}(C)$.

Оказывается (см. [25, с. 414—423]), существует специальный класс рядов — *рядов по полиномам Фабера*, по которым единственным образом разлагаются функции, принадлежащие классу

---

[17] Все результаты мы пишем только для функции $\mathscr{H}_\varepsilon$. Для $\mathscr{E}_\varepsilon$ все формулы аналогичны.

$A^K_{C_R(K)}$:

$$f(z) = \sum_{k=0}^{\infty} c_k \Phi_k(z). \tag{132}$$

При этом многочлены $\Phi_k(z)$ и коэффициенты c_k в (132) удовлетворяют следующим условиям:

α) $|a_r| \leqslant M_r/r^k$,

где
$$M_r = \max_{z \in C_r(K)} |f(z)|.$$

Отсюда при $r \to R$ получим, что

$$|a_k| \leqslant C/R^k, \tag{133}$$

а при $r \to 1$ получаем, что

$$|a_k| \leqslant \max_{z \in K} |f(z)| = \|f\|. \tag{134}$$

$$\text{β)}\quad |\Phi_k(z)| \leqslant DR'^k, \quad z \in C_R(K), \quad R' > R, \tag{135}$$

где D — некоторая константа.

$$\text{γ)}\quad \max_{z \in K} |\Phi_k(z)| \leqslant B(\delta)(1+\delta)^k. \tag{136}$$

Соотношения (132)—(136), как нетрудно видеть, влекут за собой выполнимость системы условий (II) для класса $A^K_{C_R(K)}(C)$. При этом $h = \log R/\log e$. Таким образом, согласно теореме XVIII

$$\mathscr{H}_\varepsilon(A^K_{C_R(K)}(C)) \backsim \frac{(\log(1/\varepsilon))^2}{\log R}. \tag{137}$$

Полиномы Чебышева, являющиеся частным случаем полиномов Фабера, когда K есть отрезок $\Delta = [-1, 1]$, и в отличие от общих полиномов Фабера являются ограниченными на Δ:

$$|T_k(x)| \leqslant 1, \quad x \in \Delta, \tag{138}$$

а соотношения (133)—(135), (138) уже позволяют применить теорему XVIII. Можно было бы показать, что результат теоремы XVII сохранится для достаточно гладких континуумов K, но мы на этом не останавливаемся.

5) После опубликования работы Витушкина [3], в которой им были получены формулы (129)—(131) (и их обобщения на пространства многих комплексных переменных, к которым мы перейдем ниже), А. Н. Колмогоровым была высказана гипотеза, что и в случае класса $A^K_G(C)$, где G — произвольная область, а C — односвязный континуум, содержащийся в ней: $K \subset G$, существует

константа $\tau = \tau(G, K)$ такая, что

$$\mathcal{H}_\varepsilon (A_G^K (C)) \sim \tau (G, K) \left(\log \frac{1}{\varepsilon}\right)^2. \qquad (139)$$

Эта гипотеза была почти одновременно доказана В. Д. Ерохиным [4] и К. И. Бабенко [1]. Отметим при этом, что результаты Ерохина носят несколько более общий характер.

Методы Ерохина и Бабенко существенно отличаются друг от друга.

Способ получения Ерохиным формулы (139) вполне укладывается в приведенную нами схему с использованием теоремы XVIII.

Ерохиным было показано, что в области G существует последовательность функций $\{f_k(z)\}_{k \geqslant 1}$, аналитических в G, такая, что

$$f(z) = c_0 + \sum_{k=1}^{\infty} c_k f_k(z)$$

для любой функции $f(z)$, аналитической в G, и при этом выполняются следующие свойства функций $f_k(z)$ и коэффициентов c_k:

1°) $\| f_k \| \leqslant C(\delta)(1 + \delta)^k$,

2°) $\sup\limits_{z \in G} | f_k(z) | \leqslant C(\delta)(1 + \delta)^k R^k$,

3°) $| c_k | \leqslant C(\delta)(1 + \delta)^k / R^k$,

4°) $| c_k | \leqslant \| f \|$,

где $C(\delta)$ — константа, не зависящая от k, а $R = R(G, K)$ — радиус внешней окружности при отображении двусвязной области, ограниченной K и границей G, на кольцо

$1 < | w | < R$;

R иногда называют *конформным радиусом* двусвязной области $G \setminus K$. Легко видеть, что выполнение свойств 1°) — 4°) влечет за собой применимость теоремы XVIII и при этом $h = \log R / \log e$.

Таким образом, мы получаем, что

$$\mathcal{H}_\varepsilon (A_G^K (C)) \sim \frac{2}{\log R} \left(\log \frac{1}{\varepsilon}\right)^2. \qquad (140)$$

Отметим, что в заметке [4] содержится более общий результат, чем сформулированный нами.

Изложения построения функций f_k мы не проводим. Переходим к классам функции $f(z_1, \ldots, z_s)$ многих комплексных переменных.

6) Класс $A_{K_R}^{\overline{K}_r}(C)$, $R = (R_1, \ldots, R_s)$, $r = (r_1, \ldots, r_s)$, $\overline{K}_r = \overline{K}_{r_1} \times \ldots \times K_{r_s}^s$ — замкнутый полицилиндр — прямое про-

изведение кругов

$$\overline{K}^i_{r_i} = \{z_i\colon |z_i| \leqslant r_i\}, \quad i = 1, 2, \ldots, s;$$

$K_R = K^1_{R_1} \times \ldots \times K^s_{R_s}$ — полицилиндр — прямое произведение кругов

$$\overline{K}^i_{R_i} = \{z_i : |z_i| < R_i\}, \quad i = 1, 2, \ldots, s.$$

Функции этого класса разлагаются в *ряд Тейлора*:

$$f(z) = \sum_K a_K z^k, \quad z = (z_1, \ldots, z_s).$$

Применимость теоремы XVII не представляет затруднений. Вектор h в этом случае имеет вид $h = (h_1, \ldots, h_s)$, где $h_i = \log(R_i/r_i)/\log e$, $i = 1, 2, \ldots, s$, и, таким образом,

$$\mathscr{H}_\varepsilon(A^{\overline{K}_r}_{K_R}(C)) = \frac{2}{(s+1)! \prod\limits_{i=1}^{s} \log(R_i/r_i)} \left(\log\frac{1}{\varepsilon}\right)^{s+1} + O\left(\left(\log\frac{1}{\varepsilon}\right)^s \log\log\frac{1}{\varepsilon}\right). \tag{141}$$

7) Класс $A_h(C)$, $h = (h_1, \ldots, h_s)$, функций $f(z_1, \ldots, z_s)$, периодических по каждому переменному с периодом, равным 2π, ограниченных в прямом произведении

$$\Pi^1_{h_1} \times \ldots \times \Pi^s_{h_s}, \qquad \Pi^i_{h_i} = \{z_i\colon |\operatorname{Im} z_i| < h_i\}.$$

В этом случае функции разлагаются *в ряд Фурье*:

$$f(z) = \sum_{k \in \widetilde{K}} c_k e^{i(k,z)}, \quad z = (z_1, \ldots, z_s).$$

Выполнимость условий теоремы XIX (см. примечание в п. 2) проверяется без труда и, таким образом,

$$\mathscr{H}_\varepsilon(A_h(C)) = \frac{2^{s+1}}{(s+1)!(\log e)^s h^\theta} \left(\log\frac{1}{\varepsilon}\right)^{s+1} + O\left(\left(\log\frac{1}{\varepsilon}\right)^s \log\log\frac{1}{\varepsilon}\right). \tag{142}$$

Приведем далее обобщение формулы (140) (см. также [4]):

$$\mathscr{H}_\varepsilon(A^{K^s}_{G^s}(C)) \sim \frac{2}{(s+1)! \prod\limits_{i=1}^{s} \log R_i} \left(\log\frac{1}{\varepsilon}\right)^{s+1}. \tag{143}$$

В формуле (143) $G^s = G_1 \times \ldots \times G_s$ — прямое произведение односвязных областей G_i, $K^s = K_1 \times \ldots \times K_s$ — прямое произведение односвязных континуумов $K_1 \subset C_i$, R_i — объясненные при выводе формулы (140) конформные радиусы областей $G_i \setminus K_i$.

4. Закончим этот параграф рассмотрением двух классов целых аналитических функций.

В этом пункте через z мы обозначим s-мерный комплексный вектор $z = (z_1, \ldots, z_s)$; $|z|$ обозначает у нас действительный s-мерный вектор:

$$|z| = (|z_1|, \ldots, |z_s|).$$

Мы называем целую функцию f функцией *конечного порядка*, если существует положительный вектор $\bar{p} = (\bar{p}_1, \ldots, \bar{p}_s)$ такой, что неравенство

$$|f(z)| < e^{(\theta, |z|^{\bar{p}})} = e^{|z_1|^{\bar{p}_1} + \ldots + |z_s|^{\bar{p}_s}} \tag{144}$$

выполняется для всех z, у которых все координаты по модулю превосходят некоторое $r(\bar{p})$.

Точная нижняя грань чисел $\bar{p}_1, \ldots, \bar{p}_s$, для которых выполняется неравенство (144), называется *порядком целой функции* f. Порядок функции мы обозначаем далее обычно вектором $p = (p_1, \ldots, p_s)$.

Точную нижнюю границу положительных чисел $\bar{\sigma}_1, \ldots, \bar{\sigma}_s$, для которых асимптотически

$$|f(z)| < e^{(\bar{\sigma}, |z|^p)} = e^{\bar{\sigma}_1 |z_1|^{p_1} + \ldots + \bar{\sigma}_s |z_s|^{p_s}}, \tag{145}$$

мы называем *типом целой функции* f порядка p.

Тип функции мы обозначаем далее вектором $\sigma = (\sigma_1, \ldots, \sigma_s)$.

Мы изучим ε-энтропию следующих классов целых функций конечного порядка и типа:

1°. Класс $\Phi^s_{p\sigma}(C)$ (σ и p — произвольные положительные векторы) функций f, удовлетворяющих неравенству

$$|f(z)| \leqslant Ce^{(\sigma, |z|^p)}. \tag{146}$$

Класс $\Phi^s_{p\sigma}(C)$ мы рассмотрим в метрике

$$\|f\| = \max_{z \in \bar{K}^s_\theta} |f(z)|,$$

где $\bar{K}^s_\theta = \bar{K}^1_1 \times \ldots \times \bar{K}^s_1$ — полицилиндр, т. е. произведение $\bar{K}^i_1 = \{z_i \colon |z_i| \leqslant 1\}$.

2°. Класс $F^s_{p\theta, \sigma}$, где σ — произвольный положительный вектор, а $p\theta$ — вектор с компонентами $(p, p, \ldots, p)$, состоящий из целых функций, удовлетворяющих неравенству

$$|f(z)| \leqslant Ce^{(\sigma, |\operatorname{Im} z|^{p\theta})}, \quad p > 1, \tag{147}$$

и периодических с периодом 2π по каждому переменному.

Класс $F^s_{p\theta,\sigma}$ мы рассмотрим в равномерной метрике на D^s. Имеют место следующие теоремы.

Теорема XX.

$$\mathscr{E}_\varepsilon(\Phi^s_{p\sigma}(C)) \backsim \mathscr{H}_\varepsilon(\Phi^s_{p\sigma}(C)) \backsim \frac{2}{(s+1)!}\,p^\theta\,\frac{(\log(1/\varepsilon))^{s+1}}{\log\log(1/\varepsilon)}\,. \tag{148}$$

Теорема XXI.

$$\mathscr{E}_\varepsilon(F^s_{p\theta,\sigma}(C)) \backsim \mathscr{H}_\varepsilon(F^s_{p\theta,\sigma}(C)) \backsim \Delta_{s,p,\sigma}\left(\log\frac{1}{\varepsilon}\right)^{s(p-1)/p+1}, \tag{149}$$

где

$$\Delta_{s,p,\sigma} = 2^{s+1}\,\frac{\sigma^{\theta/p}p\,(p-1)^{s/p}\,(\Gamma(1-1/p))^s}{(sp-s+p)\,(\log e)^{s(1-1/p)}\,\Gamma(s-s/p+1)}\,.$$

Отметим частный случай формулы (149) для случая одного переменного, имеющий связь со следующим параграфом:

$$\mathscr{E}_\varepsilon(F^1_{p,\sigma}) \backsim \mathscr{H}_\varepsilon(F^1_{p,\sigma}) \backsim \frac{4\sigma^{1/p}p^2\,(\log(1/\varepsilon))^{2-1/p}}{(2p-1)\,((p-1)\log e)^{1-1/p}}\,. \tag{150}$$

5. Можно показать [4], что результат теоремы XX не изменится, если вместо полицилиндра $K^1_1 \times \ldots \times K^s_1$ взять прямое произведение

$$K^1 \times \ldots \times K^s,$$

где K^i — произвольные континуумы в плоскости z_i.

Функции класса $\Phi^s_{p\sigma}(C)$ разлагаются в *ряд Тейлора*, а функции класса $F^s_{p\theta,\sigma}(C)$ в *ряд Фурье*, сходящиеся при всех конечных z. Выведем неравенства для коэффициентов этих разложений. В первом случае

$$f(z) = \sum_{k\in K} a_k z^k,$$

где k и K имеют тот же смысл, что и в теореме XVII этого параграфа. Воспользуемся неравенствами Коши и соотношением (146). Получим

$$|a_k| \leqslant \frac{\max\limits_{z\in \overline{K}^1_{R_1}\times\ldots\times\overline{K}^s_{R_s}} |f(z)|}{R^k} \leqslant \frac{Ce^{(\sigma,R^p)}}{R^k}$$

при любом $R = (R_1, \ldots, R_s)$, а следовательно, неравенство сохранится и для минимума правой части по R, который нетрудно подсчитать элементарно.

В результате получим

$$|a_k| \leqslant C \prod_{i=1}^{s} \left(\frac{\sigma_i p_i e}{k_i} \right)^{k_i/p_i} =$$

$$= C \exp \left(- \frac{1}{\log e} \sum_{i=1}^{s} \left(\frac{k_i \log k_i}{p_i} - \frac{k_i B_i}{p_i} \right) \right), \tag{151}$$

где через B_i мы обозначили $\log(\sigma_i p_i e)$.

Во втором случае

$$f(z) = \sum_{k \in \widetilde{K}} c_k e^{i(z, k)},$$

где $\widetilde{K}$ имеют смысл, приданный им в примечании 14 в п. 1. В этом случае мы воспользуемся (147) и неравенствами для c_k, которые легко вывести совершенно аналогично тому, как мы это делали в § 2, п. 3:

$$|c_k| \leqslant \max_{\substack{|\operatorname{Im} z_i| \leqslant h_i \\ i=1, \ldots, s}} |f(z)| \, e^{-(h, |k|)}.$$

Отсюда получим

$$|c_k| \leqslant C e^{-((\sigma, h^p) - (h, |k|))}, \tag{152}$$

где $h = (h_1, \ldots, h_s)$ — любой положительный вектор. Неравенства (152) останутся справедливыми и для минимума правой части по h, откуда после элементарных подсчетов получим

$$|c_k| \leqslant C e^{-(A, |k|^{p/(p-1)})}, \tag{153}$$

где

$$A = (A_1, \ldots, A_s),$$

$$A_i = \left(\frac{1}{\sigma_i p} \right)^{1/(p-1)} \left(\frac{p-1}{p} \right).$$

Неравенства (151) и (153) будут у нас играть ту же роль, что и условия (97) в теореме XVIII.

Сформулируем теперь неравенства, играющие роль условий (99) в теореме XVII.

Для любого положительного вектора $\delta = (\delta_1, \ldots, \delta_s)$ существует константа $C'(\delta)$ такая, что:

α) если для всех $k \in K$ выполнены неравенства

$$|a_k| \leqslant C'(\delta) \exp \left(- \frac{1}{\log e} \sum_{i=1}^{n} \left(\frac{k_i \log k_i}{p_i} - \frac{k_i B_i}{p_i} + k_i \delta_i \right) \right), \tag{154}$$

где B_i взято из (151), то функция $f(z) = \sum_{k \in K} a_k z^k$ принадлежит классу $\Phi^s_{p\sigma}(C)$;

β) если же для всех $k \in \widetilde{K}$ выполнены неравенства

$$|c_k| \leqslant C'(\delta) \times$$
$$\times \exp\left(-\sum_{i=1}^{s} (A_i |k_i|^{p/(p-1)} + \right.$$
$$\left. + |k_i|\delta_i\right), \qquad (155)$$

где A_i взято из (153), то функция $f(z) = \sum_{k \in \widetilde{K}} c_k e^{i(z,k)}$ принадлежит классу $F^s_{p\theta,\sigma}(C)$. Неравенства (154) и (155) доказываются единообразно. Докажем для примера неравенство (155). Имеем

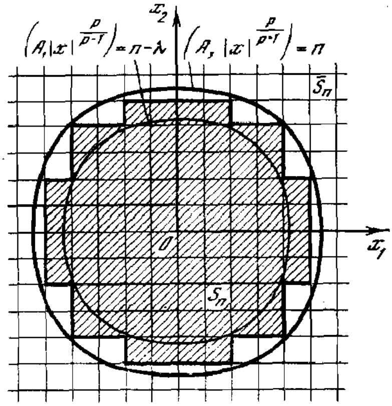


Рис. 10

$$|f(z)| \leqslant \sum_{k \in \widetilde{K}} |c_k| e^{(|\mathrm{Im}\, z|, |k|)} \leqslant C'(\delta) e^{(\sigma, |\mathrm{Im}\, z|^p)} \sum_{k \in \widetilde{K}} e^{-(\delta, k)} \leqslant C e^{(\sigma, |\mathrm{Im}\, z|^p)},$$

если положить

$$C'(\delta) = \frac{C}{\sum_{k \in \widetilde{K}} e^{-(\delta, |k|)}}.$$

Неравенства (96) оказываются справедливыми, как нетрудно заметить, и для наших классов, а именно:

$$\sup_{k \in K} |a_k| \leqslant \|f\| \leqslant \sum_{k \in K} |a_k| \qquad (156)$$

для класса $\Phi^s_{p\sigma}(C)$ и

$$\sup_{k \in \widetilde{K}} |c_k| \leqslant \|f\| \leqslant \sum_{k \in \widetilde{K}} |c_k| \qquad (157)$$

для класса $F^s_{p\theta,\sigma}$.

6. Доказательство теорем XX и XXI вполне аналогично по структуре доказательству теоремы XVII, проведенному нами в п. 2. Поэтому их доказательство мы ведем параллельно с небольшими комментариями.

Для сокращения записи введем два обозначения: k/p и x/p (где k, p и x — векторы) обозначают у нас векторы с компонентами $(k_1/p_1, \ldots, k_s/p_s)$ и соответетственно $(x_1/p_1, \ldots, x_s/p_s)$.

Запись $\log k$ для вектора k, у которого $k_i \geqslant 1$, будет обозначать у нас вектор с компонентами $(\log k_1, \ldots, \log k_s)$.

Объем встречающегося далее множества $(A, |x|^{p/(p-1)}) = 1$ мы обозначаем далее через $\Delta(A)$. Он равняется (см. [24, т. 3, с. 477]) величине

$$\Delta(A)=\frac{((1-1/p)\,\Gamma(1-1/p))^s\cdot 2^s}{\Gamma(s-s/p+1)\,A^{(1-1/p)\theta}}.$$

Приступим к доказательству.

Оценка сверху величины $\mathcal{H}_\varepsilon$

Класс $\Phi^s_{p\sigma}(C)$	Класс $F^s_{p\theta,\sigma}(C)$ (см. рис. 10)

Рассмотрим множество S_n целочисленных векторов, для которых

$\left(\frac{k}{p},\log k-B\right)\leqslant n\log n.$	$(A,\lvert k\rvert^{p/(p-1)})\leqslant n^{p/(p-1)}.$ (158)

Через $\bar S_n$ обозначим

$K\setminus S_n.$	$\tilde K\setminus S_n.$

Имеем согласно (156) и (157)

$\lVert f\rVert\leqslant\sum\limits_{k\in S_n}\lvert a_k\rvert+\sum\limits_{k\in\bar S_n}\lvert a_k\rvert=\sum\limits_{k\in S_n}\lvert a_k\rvert+R_n.$	$\lVert f\rVert\leqslant\sum\limits_{k\in S_n}\lvert c_k\rvert+\sum\limits_{k\in\bar S_n}\lvert c_k\rvert=\sum\limits_{k\in S_n}\lvert c_k\rvert+R_n.$ (159)

Неравенства (151) и (153) дадут

$\lvert R_n\rvert\leqslant C\sum\limits_{k\in\bar S_n}e^{-1/\log e\,(k/p,\,\log k-B)}$	$\lvert R_n\rvert\leqslant C\sum\limits_{k\in\bar S_n}e^{-(A,\,\lvert k\rvert^{p(p-1)})}.$ (160)

Функции $-(x/p,\ \log x-B)$ $x\geqslant 0$ и $-(A,\ x^{p/(p-1)})x\geqslant 0$ имеют минимум на кубе B_k (см. рис. 9) в точке k [18].

Доказательство этого факта мы предоставляем читателю. Таким образом, мы получаем возможность оценивать сумму (160) интегралами

$$C\int\limits_{\substack{(x/p,\,\log x-B)\geqslant(n-\lambda)\log(n-\lambda)\\ x\geqslant 0}}\exp\left(-\frac{1}{\log e}\left(\frac{x}{p},\log x-B\right)\right)dx,$$
$$C\int\limits_{(A,\,\lvert x\rvert^{p/(p-1)})\geqslant(n-\lambda)^{p/(p-1)}}\exp\left(-(A,\lvert x\rvert^{p/(p-1)})\right)dx, \tag{161}$$

где λ ограничено. Интегралы (161) оцениваются соотношениями

$\int\limits_{(x/p,\,\log x-B)\geqslant R\log R}\exp\left(-\frac{1}{\log e}\times\right.$ $\left.\times\left(\frac{x}{p},\log x-B\right)\right)dx=$ $=e^{-R\log R}R^s\left(1+\frac{O(1)}{\log R}\right).$	$\int\limits_{(A,\,\lvert x\rvert^{p/(p-1)})\geqslant R^{p/(p-1)}}\exp(-(A,$ $\lvert x\rvert^{p/(p-1)}))\,dx=O(Re^{s-R^{p/(p-1)}}),$

[18] За исключением конечного числа k в случае $\Phi^s_{p\sigma}(C)$, а именно тех k, для которых $(k/p,\ \log k-B)\leqslant 0$. Этим множеством мы пренебрегаем.

откуда для минимального n такого, чтобы $|R_n| \leqslant \varepsilon/2$, получим

$$n = \frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)} + O\left(\log\log\frac{1}{\varepsilon}\right). \quad \Bigg| \quad n = \left(\frac{\log(1/\varepsilon)}{\log e}\right)^{1-1/p} + O\left(\log\log\frac{1}{\varepsilon}\right). \tag{162}$$

Обозначим через P_n число векторов k, принадлежащих S_n. Для P_n выполнены неравенства

$$\int\limits_{\substack{(x/p,\, \log x - B) \leqslant (n-\lambda)\log(n-\lambda) \\ x \geqslant 0}} dx \leqslant P_n \leqslant \int\limits_{\substack{(x/p,\, \log x - B) \leqslant (n+\lambda)\log(n+\lambda) \\ x \geqslant 0}} dx. \quad \Bigg| \quad \int\limits_{(A,\, |x|^{p/(p-1)}) \leqslant (n-\lambda)^{p/(p-1)}} dx \leqslant P_n \leqslant \int\limits_{(A,\, |x|^{p/(p-1)}) \leqslant (n+\lambda)^{p/(p-1)}} dx. \tag{163}$$

Написанные интегралы легко поддаются оценке | Написанные интегралы равняются $\Delta(A)(n \pm \lambda)^s$, откуда

$$P_n = \frac{n^s p^\theta}{s!}\left(1 + \frac{O(1)}{\log n}\right). \quad \Bigg| \quad P_n = \Delta(A) n^s \left(1 + \frac{O(1)}{n}\right). \tag{164}$$

Аппроксимируем далее коэффициент $a_k = \alpha_k + i\beta_k$ (соответственно $c_k = \alpha_k + i\beta_k$) величиной $\alpha_k' + i\beta_k'$ с точностью до ε/P_n, полагая

$$\begin{aligned} \alpha_k' &= \left[\frac{\sqrt{2}\, P_n \alpha_k}{\varepsilon}\right] \frac{\varepsilon}{\sqrt{2}\, P_n} = m_k^1 \frac{\varepsilon}{\sqrt{2}\, P_n}, \\ \beta_k' &= \left[\frac{\sqrt{2}\, P_n \beta_k}{\varepsilon}\right] \frac{\varepsilon}{\sqrt{2}\, P_n} = m_k^2 \frac{\varepsilon}{\sqrt{2}\, P_n}. \end{aligned} \tag{165}$$

Для m_k^i, $i = 1, 2$, получаются неравенства

$$|m_k^i| \leqslant \left|\frac{C' P_n \exp(-1/\log e\,(k/p, \log k - B))}{\varepsilon}\right| = N_k, \quad \Bigg| \quad |m_k^i| \leqslant \left|\frac{C' P_n \exp(-(A,\, |k|^{p/(p-1)}))}{\varepsilon}\right| = N_k, \tag{166}$$

где C' ограничены. Отсюда

$$\log N_k = \log\frac{1}{\varepsilon} - \left(\frac{k}{p}, \log k - B\right) + \log P_n + D_1, \quad \Bigg| \quad \log N_k = \log\frac{1}{\varepsilon} - (A, |k|^{p/(p-1)}) \times \log e + \log P_n + D_2,$$

где D_1 и D_2 — ограниченные константы, не зависящие от k. Величины

$$\sum_{k\in S_n}\left(\frac{k}{p},\ \log k - B\right) \qquad \Bigg| \qquad \sum_{k\in S_n}(A, |k|^{p/(p-1)})$$

оцениваются интегралами

$$\int\limits_{(x/p,\,\log x - B)\leqslant(n-\lambda)\log(n-\lambda)} \left(\frac{x}{p},\ \log x - B\right)dx,$$

$$\int\limits_{(A,\,|x|^{p/(p-1)})\leqslant(n-\lambda)^{p/(p-1)}} (A, |x|^{p/(p-1)})\,dx,$$

главный член которого равен | который равняется (см. [24, т. 3, с. 479])

$$\frac{sp^{\theta}n^{s}n\log n}{(s+1)!} + O(n^{s}). \qquad \Bigg| \qquad \frac{s\Delta(A)\,n^{s+p/(p-1)}}{s+p/(p-1)}. \tag{167}$$

Суммируя все полученное, будем иметь, использовав (162),

$$\mathscr{H}_{\varepsilon}(\Phi^{s}_{p\sigma}(C)) \lesssim \frac{2n^{s}p^{\theta}}{s!}\log\frac{1}{\varepsilon} - \frac{2sp^{\theta}n^{s}}{(s+1)!}n\log n \sim \frac{2p^{\theta}}{(s+1)!}\frac{(\log(1/\varepsilon))^{s+1}}{(\log\log(1/\varepsilon))}.$$

$$\mathscr{H}_{\varepsilon}(F^{s}_{p\sigma}(C)) \lesssim 2\Delta(A)\,n^{s}\log\frac{1}{\varepsilon} - 2\Delta(A)\frac{s}{s+p/(p-1)}n^{s+p/(p-1)} \times \log e \sim \Delta_{sp\sigma}\left(\log\frac{1}{\varepsilon}\right)^{s(p-1)/p+1}. \tag{168}$$

Оценка снизу для $\mathscr{E}_{\varepsilon}$

Задавшись произвольным $\delta > 0$, подберем максимальное m так, чтобы

$$C'(\delta)\,m^{-m} \geqslant 2\sqrt{2}\,\varepsilon, \qquad \Big| \qquad C'(\delta)\,e^{-m^{p/(p-1)}} \geqslant 2\sqrt{2}\,\varepsilon, \tag{169}$$

где $C'(\delta)$ — константа из (154) и (155).

Рассмотрим множество S_m целочисленных векторов, введенное выше, для m, удовлетворяющего (169). Пусть $\{\Psi\}$ означает множество функций

$$f(z) = \sum_{k\in K} a_k z^k \qquad \Big| \qquad f(z) = \sum_{k\in \widetilde{K}} c_k e^{i(k,z)}$$

таких, что

α) $a_k = 0,\quad k \notin S_m$; | α) $c_k = 0,\quad k \notin S_m$;

β) $a_k = (s_k^1 + is_k^2)\,2\varepsilon,\quad k \in S_m$; | β) $c_k = (s_k^1 + is_k^2)\,2\varepsilon,\quad k \in S_m$;

γ) $s_k^i,\ i = 1,\ 2,$ — целые и | γ) $s_k^i,\ i = 1,\ 2,$ — целые и

$$|s_k^i| \leqslant \frac{C'(\delta)}{2\sqrt{2}\varepsilon} \exp\left(-\frac{1}{\log e}\left(\frac{p}{k}, \log k - B + p\delta\right)\right) = M_k.$$

$$|s_k^i| \leqslant \frac{C'(\delta)}{2\sqrt{2}\,\varepsilon} \times \exp(-(A, |k|^{p/(p-1)}) - (\delta, |k|)) = M_k. \tag{170}$$

Таким образом, согласно (154) и (155)

$$\{\Psi\} \subset \Phi_{p\sigma}^{s}(C). \qquad \{\Psi\} \subset F_{p\theta,\sigma}^{s}(C).$$

Из (168) и (170) получим, что

$$m = \frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)} + O(1),$$

$$\log M_k = \log\frac{1}{\varepsilon} - \left(\frac{k}{p}, \log k - B + p\delta\right) + \log C'(\delta) + D_1',$$

$$m = \left(\frac{\log(1/\varepsilon)}{\log e}\right)^{1-1/p} + O(1),$$

$$\log M_k = \log\frac{1}{\varepsilon} - (A, |k|^{p/(p-1)}) \times \log e - (\delta, |k|)\log e + \log C'(\delta) + D_2', \tag{171}$$

где D_1' и D_2' — ограниченные константы. Отсюда окончательно выкладки, аналогичные выкладкам при оценке сверху, дадут

$$\mathscr{E}_{2\varepsilon}(\Phi_{p\sigma}^{s}(C)) \succsim \frac{2p^{\theta}}{(s+1)!} \times \frac{(\log(1/\varepsilon))^{s+1}}{(\log\log(1/\varepsilon))^{s}}.$$

$$\mathscr{E}_{2\varepsilon}(F_{p\theta,\sigma}^{s}(C)) \succsim \Delta_{sp\sigma} \times \left(\log\frac{1}{\varepsilon}\right)^{s((p-1)/p)+1}. \tag{172}$$

Из (168) и (172) вытекают (148) и (149), что и требовалось доказать.

§ 8
ε-ЭНТРОПИЯ КЛАССОВ АНАЛИТИЧЕСКИХ ФУНКЦИЙ, ОГРАНИЧЕННЫХ НА ДЕЙСТВИТЕЛЬНОЙ ОСИ[19]

1. Результаты данного параграфа принадлежат В. М. Тихомирову. Частично они были опубликованы в заметке [7].

Проблематика настоящего параграфа возникла в связи с некоторыми идеями теории информации. В теории связи большую роль играет то обстоятельство, что сигналы с ограниченным спектром, помещающимся в полосе частот ширины 2σ, определяются по дискретной совокупности своих значений, взятых в равноотстоящих друг от друга точках [20].

---

[19] Печатается в сокращении, причем нумерация формул, теорем и лемм сохранена.

[20] В применении к функциям, преобразование Фурье (спектр) которых обращается в нуль вне некоторого отрезка длины 2σ, эта идея восстанавливаемости функции по дискретной совокупности ее значений, взятых в арифметической прогрессии с разностью π/σ, была впервые обоснована Валле-Пуссеном.

В. А. Котельниковым в работе [26] было указано важное применение этого свойства для теории связи. Им было отмечено, что количество информации, содержащееся в задании на отрезке длины T функции со спектром, ограниченным полосой частот ширины 2σ, при больших T эквивалентно количеству информации в задании $2\sigma T/\pi$ действительных чисел. В литературе это утверждение часто именуется *теоремой Котельникова.* Эту же идею в несколько иной форме высказал также и К. Шеннон [27, Добавл. 7].

Класс $\mathscr{B}_\sigma(C)$, с которым мы встретимся в этом параграфе, является замыканием в равномерной метрике на отрезке длины T (при любом $T > 0$) класса $B_\sigma(C)$ конечных сумм:

$$f(t) = \sum_{k=-n}^{n} c_k e^{i\lambda_k t},$$

подчиненных на действительной оси условию $|f| \leqslant C$ с частотами λ_k из отрезка $-\sigma \leqslant \lambda \leqslant \sigma$ (см. [23, с. 160—164]). Поэтому формулируемая нами теорема XXII может служить одним из вариантов обоснования теоремы Котельникова.

В настоящем параграфе мы рассматриваем следующие классы одного комплексного переменного:

$\mathscr{B}_\sigma(C)$ — класс целых аналитических функций, удовлетворяющих соотношению

$$|f(z)| \leqslant Ce^{\sigma|\operatorname{Im} z|}.$$

Обозначение $\mathscr{B}_\sigma$ мы заимствуем из [23].

$\mathscr{F}_{p\sigma}(C)$ — класс целых функций порядка p и типа σ, удовлетворяющих соотношению

$$|f(z)| \leqslant Ce^{\sigma|\operatorname{Im} z|^p}.$$

$\mathscr{A}_h(C)$ — класс аналитических функций $f(z)$, ограниченных в полосе $|\operatorname{Im} z| < h$ константой C.

Все перечисленные классы, как легко видеть, ограничены на вещественной оси D.

Отметим, что класс $\mathscr{B}_\sigma$ совпадает с классом $\mathscr{F}_{1\sigma}$, но мы его выделяем отдельно ввиду его важной роли в дальнейшем [21].

Перечисленные классы функций мы рассматриваем как нормированные пространства, вводя нормы и метрики формулами

$$\|f\|_T = \max_{t\in\Delta_T} |f(t)|, \qquad \rho_T(f_1, f_2) = \|f_1 - f_2\|_T, \tag{173}$$

где Δ_T — отрезок действительной прямой $-T \leqslant t \leqslant T$.

Всюду в этом параграфе ε-энтропию и ε-емкость класса $\mathscr{F}$ функций, определенных на D с нормой (173), мы обозначаем соответственно через $\mathscr{H}_\varepsilon^T(\mathscr{F})$ и $\mathscr{E}_\varepsilon^T(\mathscr{F})$.

[21] Как обычно, мы считаем, что $f \in \mathscr{B}_\sigma$, $\mathscr{F}_{p\sigma}$ или $\mathscr{A}_h$, если f принадлежит соответственно $\mathscr{B}_\sigma(C)$, $\mathscr{F}_{p\sigma}(C)$ или $\mathscr{A}_h(C)$ с некоторой константой C.

Для формулировки результатов, которые мы будем доказывать в этом параграфе, нам потребуется определение верхней и нижней ε-энтропии и ε-емкости на единицу длины.

Верхней ε-энтропией на единицу длины для класса $\mathcal{F}$, определенного на D, назовем функцию

$$\overline{\mathcal{H}}_\varepsilon^B(\mathcal{F}) = \limsup_{T\to\infty} (2T)^{-1} \mathcal{H}_\varepsilon^T(\mathcal{F}).$$

Нижней ε-энтропией на единицу длины назовем функцию

$$\overline{\mathcal{H}}_\varepsilon^H(\mathcal{F}) = \liminf_{T\to\infty} (2T)^{-1} \mathcal{H}_\varepsilon^T(\mathcal{F}).$$

Аналогично определяются *верхняя и нижняя ε-емкость на единицу длины* $\overline{\mathcal{E}}_\varepsilon^B(\mathcal{F})$ и $\overline{\mathcal{E}}_\varepsilon^H(\mathcal{F})$.

Имеют место следующие теоремы.

Теорема XXII.

$$\overline{\mathcal{H}}_\varepsilon^B(\mathcal{B}_\sigma(C)) \backsim \overline{\mathcal{H}}_\varepsilon^H(\mathcal{B}_\sigma(C)) \backsim \frac{2\sigma}{\pi} \log\frac{1}{\varepsilon}. \tag{174}$$

Теорема XXIII.

$$\overline{\mathcal{H}}_\varepsilon^B(\mathcal{A}_h(C)) \backsim \overline{\mathcal{H}}_\varepsilon^H(\mathcal{A}_h(C)) \backsim \frac{1}{\pi h \log e}\left(\log\frac{1}{\varepsilon}\right)^2. \tag{175}$$

Теорема XXIV.

$$\overline{\mathcal{H}}_\varepsilon^B(\mathcal{F}_{p\sigma}(C)) \backsim \overline{\mathcal{H}}_\varepsilon^H(\mathcal{F}_{p\sigma}(C)) \backsim \frac{2\sigma^{1/p} p^2 (\log(1/\varepsilon))^{2-1/p}}{\pi(2p-1)((p-1)\log e)^{1-1/p}}. \tag{176}$$

Соотношения для $\overline{\mathcal{E}}_\varepsilon^B$ *и* $\overline{\mathcal{E}}_\varepsilon^H$ *имеют тот же вид.*

Рекомендуем читателю сравнить написанные выше формулы (175) и (176) с формулами (130) и (149) обычной ε-энтропии классов $F_{p\sigma}^1$ и A_h. (Напомним, что эти классы состоят из периодических с периодом, равным 2π, функций, принадлежащих классам $F_{p\sigma}(C)$ и $A_h(C)$.) Сходство этих формул дополнится сходством методов их доказательств, о чем речь будет идти в дальнейшем.

Отметим также соотношение

$$\mathcal{H}_\varepsilon(B_\sigma(C)) \backsim (4[\sigma]+2)\log(1/\varepsilon) \tag{177}$$

для класса функций из $\mathcal{B}_\sigma$, периодических с периодом, равным 2π, которое естественно сопоставить с соотношением (174). Эквивалентность (177) вытекает из того факта, что класс $B_\sigma(C)$ состоит из тригонометрических многочленов степени, не большей $[\sigma]$ (см. [23, с. 190]).

2. Переходим к доказательству теоремы XXII. Следуя стилю предыдущего параграфа, мы выделяем отдельно в виде двух лемм те свойства класса $\mathcal{B}_\sigma$, которые используем в дальнейшем при доказательстве.

В леммах IV и V α — произвольное положительное число (в лемме V — меньшее σ), $C_1(\alpha)$ и $C_2(\alpha)$ — положительные константы, зависящие только от α.

Лемма IV. *Функция $f(z) \in \mathcal{B}_\sigma(C)$ может быть представлена в виде ряда*

$$f(z) = \sum_{k=-\infty}^{+\infty} c_k f_k(z). \tag{178}$$

При этом:

а) $|c_k| \leqslant C$.

б) *Существует сходящийся ряд из положительных членов*

$$\sum_{k=-\infty}^{+\infty} a_k = A \tag{179}$$

такой, что для f, представимой рядом (178),

$$|f(t)| \leqslant C_1(\alpha) \sum_{k=-\infty}^{+\infty} a_k |c_{[t(\sigma+\alpha)/\pi]+k}|. \tag{180}$$

Из леммы IV мы получим, что

$$\limsup_{\varepsilon\to 0} \frac{\overline{\mathcal{H}}_\varepsilon^B(\mathcal{B}_\sigma(C))}{\log(1/\varepsilon)} \leqslant \frac{2\sigma}{\pi}. \tag{181}$$

Лемма V. *Существуют функции $\varphi_k(z)$, $k = 0, \pm 1, \ldots$, такие, что если*

$$|c_k| \leqslant C_2(\alpha)\, C,$$

то функция

$$f(z) = \sum_{|k| \leqslant [T(\sigma-\alpha)/\pi]} c_k \varphi_k(z) \tag{182}$$

принадлежит $\mathcal{B}_\sigma(C)$ и

$$\| f \|_T \geqslant \max_{|k| \leqslant [T(\sigma-\alpha)/\pi]} |c_k|. \tag{183}$$

Из леммы V мы получим, что

$$\liminf_{\varepsilon\to 0} \frac{\overline{\mathcal{E}}_{2\varepsilon}^H(\mathcal{B}_\sigma(C))}{\log(1/\varepsilon)} \geqslant \frac{2\sigma}{\pi}. \tag{184}$$

Ввиду того что

$$\overline{\mathcal{E}}_{2\varepsilon}^H \leqslant \overline{\mathcal{E}}_{2\varepsilon}^B \leqslant \overline{\mathcal{H}}_\varepsilon^H \leqslant \overline{\mathcal{H}}_\varepsilon^B,$$

соотношения (181) и (184) дают в совокупности теорему XXII.

Выведем неравенства (181) и (184) из соответствующих лемм.

Фиксируем α. Подберем n таким образом, чтобы

$$R_n < \varepsilon/2C_1(\alpha)\,C, \tag{185}$$

где $R_n = \sum\limits_{|k|>n} a_k$ — остаток ряда (179).

Положим $n_T = [(\sigma+\alpha)\,T/\pi] + n + 1$. Коэффициенты $c_k = \alpha_k + i\beta_k$, $|k| \leqslant n_T$, ряда (178) аппроксимируем с точностью до $\varepsilon/2AC_1(\alpha)$ (A — сумма ряда (179)), положив

$$c_k' = \left[\frac{\alpha_k 2\sqrt{2}\,AC_1(\alpha)}{\varepsilon}\right]\frac{\varepsilon}{2\sqrt{2}\,AC_1(\alpha)} + i\left[\frac{\beta_k 2\sqrt{2}\,AC_1(\alpha)}{\varepsilon}\right]\frac{\varepsilon}{2\sqrt{2}\,AC_1(\alpha)}.$$

Пусть

$$\tilde f(z) = \sum_{|k|\leqslant n_T} c_k' f_k(z).$$

Имеем при $|t| \leqslant T$

$$|f(t) - \tilde f(t)| \leqslant C_1(\alpha)\sum_{k=-n}^{n} a_k\,|C_{[t(\sigma+\alpha)/\pi]+k} - C'_{[t(\sigma+\alpha)/\pi]+k}| +$$
$$+ C_1(\alpha)\sum_{|k|>n} a_k\,|C_{[t(\sigma+\alpha)/\pi]+k}| =$$
$$= \frac{\varepsilon}{2A}\sum_{|k|\leqslant n} a_k + C_1(\alpha)\,C\sum_{|k|>n} a_k \leqslant \varepsilon. \tag{186}$$

В выкладке (186) мы пользовались тем, что для всех $s = [t(\sigma+\alpha)/\pi] + k$, $|t| \leqslant T$, $|k| \leqslant n$, c_s аппроксимировано с точностью до $\varepsilon/2AC_1(\alpha)$, и соотношением (185).

Соотношение (186) означает, что $\|f - \tilde f\|_T \leqslant \varepsilon$, т. е. множество всех функций $\{\tilde f\}$ образует ε-сеть для класса $\mathcal{B}_\sigma(C)$ в метрике ρ_T. Легко подсчитать, что число $N(T)$ элементов в этом множестве $\{\tilde f\}$ удовлетворяет неравенству[22]

$$N(T) \leqslant \left(2\left[\frac{2\sqrt{2}\,AC_1(\alpha)\,C}{\varepsilon}\right] + 1\right)^{4n_T+2} =$$
$$= \left(2\left[\frac{C'C}{\varepsilon}\right] + 1\right)^{4[T(\sigma+\alpha)/\pi]+4n+6}. \tag{187}$$

Логарифмируя (187) и используя то, что $\mathcal{H}_\varepsilon^T(\mathcal{B}_\sigma(C)) \leqslant \log N(T)$, а также произвольность α, получим (181).

Неравенство (184) получается еще проще. Фиксируя α, положим $m_T = [T(\sigma-\alpha)/\pi]$. Рассмотрим матрицу целочисленных

---

[22] Через C' мы обозначили константу $2AC_1(\alpha)$.

векторов

$$U=\left\|\begin{matrix} s_m^1 & \ldots & s_0^1 & \ldots & s_m^1 \\ s_{-m}^2 & \ldots & s_0^2 & \ldots & s_m^2 \end{matrix}\right\|.$$

При этом потребуем, чтобы

$$|s_k^i| \leqslant CC_2(\alpha)/\sqrt{2}\varepsilon. \tag{188}$$

Каждой матрице U из данного множества поставим в соответствие функцию $f_U(t)$:

$$f_U(t)=2\varepsilon\sum_{|k|\leqslant m_T}(s_k^1+is_k^2)\varphi_k(z).$$

Из (183) следует 2ε-различимость множества $\{f_U(t)\}$ в метрике ρ_T.

Число $M(T)$ функций из этого множества согласно (188) при достаточно малых ε удовлетворяет неравенству

$$M(T)\geqslant\left(2\left[\frac{CC_2(\alpha)}{2\sqrt{2}\varepsilon}\right]\right)^{4m_T}\geqslant\left(2\left[\frac{CC_2(\alpha)}{2\sqrt{2}\varepsilon}\right]\right)^{2[T(\sigma-\alpha)/\pi]}. \tag{189}$$

Прологарифмировав (189) и используя то, что $C_{2\varepsilon}^T(\mathscr{B}_\sigma(C))\geqslant \geqslant \log M(T)$, а также произвольность α, получим (184).

3. Для завершения доказательства теоремы XXII нам остается получить леммы IV и V п. 2.

Воспользуемся интерполяционной формулой Картрайт. Пусть $g(t)\in\mathscr{B}_{\sigma'}(C)$, $\sigma'<\pi$, $0<\omega<\pi-\sigma'$. Тогда имеет место следующая формула:

$$g(t)=\frac{\sin\pi t}{\pi\omega}\sum_{k=-\infty}^{+\infty}(-1)^k\frac{g(k)\sin\omega(t-k)}{(t-k)^2}. \tag{190}$$

Доказательство формулы (190) содержится в [28, с. 269]. Легко понять, что если $f\in\mathscr{B}_\sigma(C)$, то $g(t)=f(\sigma' t/\sigma)\in\mathscr{B}_{\sigma'}(C)$. Используя это, перепишем формулу (190) для функции $f\in\mathscr{B}_\sigma$ при произвольном σ:

$$f(t)=\frac{\sin(\pi\sigma t/\sigma')}{\pi\omega}\sum_{k=-\infty}^{+\infty}(-1)^k\frac{f(\sigma' k/\sigma)\sin\omega(\sigma t/\sigma'-k)}{(\sigma t/\sigma'-k)^2}. \tag{191}$$

Отметим, что формула (191) выражает ту самую «восстановимость» функции по ее дискретной последовательности в применении к классу $\mathscr{B}_\sigma$, о которой мы упоминали в начале п. 1.

Положим теперь для произвольного α

$$\sigma'=\frac{\pi\sigma}{\sigma+\alpha},\quad \text{или}\quad \frac{\sigma}{\sigma'}=\frac{\sigma+\alpha}{\pi},\qquad \omega=\frac{\pi-\sigma'}{2},$$

$$f_k(z)=\frac{(-1)^{k+1}\sin(\sigma+\alpha)t\sin\omega(k-(\sigma+\alpha)t/\pi)}{\pi\omega(k-t(\sigma+\alpha)/\pi)^2},$$

$$c_k=f(\pi k/(\sigma+\alpha)).$$

Получим

$$f(z)=\sum_{k=-\infty}^{+\infty} c_k f_k(z). \tag{192}$$

Покажем, что ряд (192) удовлетворяет поставленным в лемме IV условиям.

Условие а) удовлетворяется вследствие того, что $f \in \mathcal{B}_\sigma(C)$. Далее при фиксированном t в сумме (192) выделим отдельно слагаемые с номерами

$$k_0=[t(\sigma+\alpha)/\pi], \quad k_0+1 \quad \text{и} \quad k_0-1.$$

Для остальных k, т. е. k, задаваемых формулой

$$k=\pm s+k_0, \quad s=2, 3, \ldots,$$

воспользуемся тем, что

$$\frac{1}{(k-t(\sigma+\alpha)/\pi)^2} \leqslant \frac{1}{(|s|-1)^2},$$

и получим

$$|f(t)| \leqslant \sum_{s=-1}^{+1} |c_{k_0+s}|\,|f_{k_0+s}(t)| + \sum_{|s|\geqslant 2} |c_{k_0+s}| \left(\frac{1}{|s|-1}\right)^2.$$

Обозначив, далее, через $a_{-1}=a_0=a_1$ максимум функции

$$f_0(t)=\frac{\sin(\sigma+\alpha)t\sin(\omega(\sigma+\alpha)t/\pi)}{\pi\omega t^2},$$

а через $a_{\pm s}$ величину $(|s|-1)^{-2}$, $s=2, 3, \ldots$, получим то, что требовалось в условии б) леммы IV.

В качестве функции φ_k леммы V возьмем функции

$$\frac{\sin(\sigma-\alpha)(z-k\pi/(\sigma-\alpha))\sin\alpha(z-k\pi/(\sigma-\alpha))}{\alpha(\sigma-\alpha)(z-k\pi/(\sigma-\alpha))^2}.$$

Лемма V есть следствие следующих очевидных фактов:

1) $\varphi_k \in \mathcal{B}_\sigma$ и, следовательно, $\sum\limits_{k=-m}^{m} c_k\varphi_k \in \mathcal{B}_\sigma$;

2) при равномерно ограниченных $|c_k|$ суммы $\sum\limits_{k=-m}^{m} c_k\varphi_k$ равномерно ограничены на D.

3) $\sum\limits_{k=-m}^{m} c_k\varphi_k\left(\frac{l\pi}{\sigma-\alpha}\right)=c_l, \quad |l|\leqslant m.$

Этим заканчивается доказательство лемм IV и V и, следовательно, теоремы XXII.

§ 9
ε-ЭНТРОПИЯ ПРОСТРАНСТВ ДЕЙСТВИТЕЛЬНЫХ ФУНКЦИОНАЛОВ

1. В этом пункте мы рассматриваем пространства функционалов, удовлетворяющих условию Гёльдера.

Пусть X — метрическое пространство, $F(x)$ — действительный функционал на X. Функционал $F(x)$ называется *удовлетворяющим условию Гёльдера с функцией* $\omega = \omega(\lambda)$, если

$$|F(x_1) - F(x_2)| \leqslant \omega(\rho(x_1, x_2)), \tag{232}$$

где $\omega(\lambda)$ — определенная для положительных λ монотонная положительная функция.

Мы будем говорить, что $F(x)$ *удовлетворяет условию Гёльдера с показателем* β $(0 < \beta \leqslant 1)$, если $\omega(\lambda) = L\lambda^{\beta}$.

Через $\mathscr{D}_{\omega}^{A}(C)$, $A \subset X$, мы обозначаем множество всех функционалов над X, удовлетворяющих условию Гёльдера с функцией ω и ограниченных на A константой C.

Все функционалы мы рассматриваем, как всегда, в равномерной метрике:

$$\rho(F_1, F_2) = \sup_{x \in A} |F_1(x) - F_2(x)|.$$

Имеет место

Теорема XXV. *Если A вполне ограничено и*

$$\log\log(1/\varepsilon) \prec\!\!\prec \mathscr{H}_{\varepsilon}(A) \asymp \mathscr{H}_{\varepsilon k}(A)$$

для любого $k > 0$, то

$$\log \mathscr{H}_{\varepsilon}(\mathscr{D}_{\omega}^{A}(C)) \asymp \mathscr{H}_{\omega^{-1}(\varepsilon)}(A).$$

Докажем, что в предположениях теоремы выполняются неравенства:

$$\mathscr{H}_{\delta_2}(A) \leqslant \log \mathscr{H}_{\varepsilon}(\mathscr{D}_{\omega}^{A}(C)) \leqslant \mathscr{H}_{\delta_1}(A) + \log\log s, \tag{233}$$

где $\delta_1 = {}^1/_2\, \omega^{-1}(\varepsilon/2)$, $\delta_2 = 2\omega^{-1}(2\varepsilon)$, $s = 2\,[2C/\varepsilon] + 1$.

Действительно, для всякого $\delta_1 > 0$ существует δ_1-покрытие A, состоящее из $N_{\delta_1}(A)$ элементов: $U_1, \ldots, U_{N_{\delta_1}}$. Выберем в каждом из элементов покрытия U_i по точке x_i. Функционал $F(x)$ на множестве A мы аппроксимируем функционалом:

$$\tilde{F}(x) = \left[\frac{2F(x_i)}{\varepsilon}\right]\frac{\varepsilon}{2} \quad \text{для } x \in U_i.$$

Нетрудно убедиться, что если $\delta_1 = {}^1/_2\, \omega^{-1}(\varepsilon/2)$, то

$$\rho(F, \tilde{F}) \leqslant \varepsilon.$$

Функционал $\tilde{F}$ принимает на каждом множестве U_i не более s значений, а значит, число всех функционалов не превосходит числа $s^{N_{\delta_1}}$, т. е.

$$N_\varepsilon(\mathscr{D}_\omega^A(C)) \leqslant \left(2\left[\frac{2C}{\varepsilon}\right]+1\right)^{N_{\delta_1}(A)}. \tag{234}$$

Пусть, с другой стороны, $x_1, \ldots, x_{M_{\delta_2}(A)}$ образуют максимальное $\delta_2/2$-различимое множество в A, состоящее из $M_{\delta_2}(A)$ элементов. Положим $\delta_2 = 2\omega^{-1}(2\varepsilon)$. Вокруг каждого x_i опишем сферу S_i радиуса $\delta_2/4$. Построенные сферы, очевидно, не пересекаются.

Строим множество функционалов, зависящих от наборов $\alpha_1, \ldots, \alpha_{M_{\delta_2}(A)}$, где α_i принимают значения 0 или 1:

$$F_{\alpha_1, \ldots, \alpha_{M_{\delta_2}(A)}}(x) = \begin{cases} 2\varepsilon\left(1 - \dfrac{\omega(\rho(x, x_i))}{2\varepsilon}\right) & \text{при } \alpha_i = 1,\ x \in S_i, \\ 0 \quad \text{при всех остальных } x \text{ или } \alpha_i = 0. \end{cases}$$

Число всех функционалов равно, очевидно, $2^{M_{\delta_2}(A)}$, в силу выбора все они 2ε-различимы. Нетрудно проверить, что они удовлетворяют условию Гёльдера с функцией ω, т. е. при $2\varepsilon < C$

$$M_{2\varepsilon}(\mathscr{D}_\omega^A(C)) \geqslant 2^{M_{\delta_2}(A)}. \tag{235}$$

Неравенства (233) получаются из (234) и (235) двойным логарифмированием. Применим полученную теорему и неравенства (233) к пространству функционалов, удовлетворяющих условию Гёльдера с показателем β над изученными нами ранее пространствами.

1) A есть единичный n-мерный куб D_n. Пространство $\mathscr{D}_\beta^A(C)$ совпадает с пространством $F_\beta^{D_n}(C)$ или $F_\beta^n(C)$ в обозначении § 5.

Из неравенств (233) следует, что

$$\mathscr{H}_\varepsilon(F_\beta^n(C)) \asymp (1/\varepsilon)^{n/\beta}. \tag{236}$$

Неравенство (232) — частный случай основной теоремы, выведенной нами в § 5.

2) A есть введенное в § 5 пространство $F_q^n(C_1, \ldots, C_p, L) = F_q^n$. Из результата § 5

$$\mathscr{H}_\varepsilon(F_q^n) \asymp \left(\frac{1}{\varepsilon}\right)^{n/q}$$

следует применимость нашей теоремы, откуда

$$\log \mathscr{H}_\varepsilon(\mathscr{D}_\beta^{F_q^n}) \asymp \left(\frac{1}{\varepsilon}\right)^{n/q\beta}. \tag{237}$$

3) A есть какое-либо из пространств аналитических функций $A_G^K(C)$, для которого

$$\mathscr{H}_\varepsilon(A_G^K(C)) \backsim \tau(K, G)\left(\log\frac{1}{\varepsilon}\right)^{\lambda+1}$$

(см. § 7). Из неравенств (233) получим, что

$$\log \mathscr{H}_\varepsilon(\mathscr{D}_\beta^{A_G^K(C)}) \backsim \frac{\tau(K, G)}{\beta^{\lambda+1}}\left(\log\frac{1}{\varepsilon}\right)^{\lambda+1}.$$

2. В этом пункте мы даем некоторые незначительные уточнения неравенств (233) для функционалов, удовлетворяющих условию Липшица.

Функционал $F(x)$ мы назовем *удовлетворяющим условию Липшица*, если он удовлетворяет условию Гёльдера с показателем $\beta = 1$.

Далее для простоты мы считаем константу Липшица $L = 1$. Пространство функционалов, удовлетворяющих условию Липшица над пространством A, мы обозначаем $\mathscr{D}_1^A$.

Покажем, что для *связного вполне ограниченного множества A, лежащего в центрируемом* [23] *пространстве, и для любого целого s выполнены следующие неравенства*:

$$2^{M_{2\varepsilon}(A)} \leqslant N_\varepsilon(\mathscr{D}_1^A(C)) \leqslant \left(2\left[\frac{C(s+1)}{\varepsilon}\right]+1\right)(s+1)^{N_{s\varepsilon/(s+1)}(A)}. \quad (238)$$

Правое неравенство получается так. Рассмотрим $s\varepsilon/(s+1)$-покрытие A, состоящее из $N_{s\varepsilon/(s+1)}(A) = N$ элементов $U_1, \ldots, U_N$. Через x_i обозначим центры множеств U_i. Связность множества A дает нам возможность соединить любые два множества U_i и U_j цепочкой пересекающихся множеств U_k. Строим теперь аппроксимирующий функционал $\tilde{F}$ для функции F. В качестве его значения на множестве U_1 берем число

$$\left[\frac{F(x_1)(s+1)}{\varepsilon}\right]\frac{\varepsilon}{s+1},$$

т. е. приближаем значение $F(x_1)$ с точностью до $\varepsilon/(s+1)$. Легко видеть, что значение функционала F в центрах x_i, примыкающих к U_1 множеств U_i, удовлетворяют неравенству

$$\begin{aligned}|F(x_i) - \tilde{F}(x_1)| &\leqslant |F(x_i) - F(x_1)| + |F(x_1) - \tilde{F}(x_1)| \leqslant \\ &\leqslant \frac{s\varepsilon}{s+1} + \frac{\varepsilon}{s+1} = \varepsilon. \end{aligned} \quad (239)$$

Неравенство (239) означает, что для приближения $F(x_i)$ в примыкающих к U_1 центрах x_i множеств U_i с точностью до $\varepsilon/(s+1)$ нам потребуется при знании $F(x_1)$ одно из $s+1$ чисел:

$$\tilde{F}(x_1) \pm (2k-1)\frac{\varepsilon}{s+1} \quad \left(k = 1, \ldots, \frac{s+1}{2}\right) \text{ при } s \text{ нечетном}$$

или

$$\tilde{F}(x_1) \pm 2k\frac{\varepsilon}{s+1}\left(k = 1, \ldots, \frac{s}{2}\right) \text{ при } s \text{ четном.}$$

[23] Определение см. § 1.

Ввиду оговоренной связности множества A мы сумеем однозначно построить весь функционал $\tilde{F}$, последовательно переходя от центра к центру, выбирая при каждом шаге одно из $s+1$ значений. Таким образом, получим

$$\rho(F, \tilde{F}) \leqslant \varepsilon,$$

т. е. ε-сеть в пространстве $\mathscr{D}_1^A(C)$. Правое неравенство в (238) и означает подсчет всевозможных функционалов $\tilde{F}$.

Для получения левого неравенства в (238) построим максимальное 2ε-различимое множество $x_1, \ldots, x_M$, состоящее из $M_{2\varepsilon}(A) = M$ элементов. Как и при доказательстве теоремы XXV, строим множество функционалов, зависящих от $\alpha_1, \ldots, \alpha_M$, $\alpha_i = \pm 1$:

$$F_{\alpha_1, \ldots, \alpha_M}(x) = \begin{cases} \varepsilon(1 - \rho(x, x_i)/\varepsilon) & \text{при } \alpha_i = 1,\ \rho(x, x_i) \leqslant \varepsilon, \\ -\varepsilon(1 - \rho(x, x_i)/\varepsilon) & \text{при } \alpha_i = -1,\ \rho(x_i, x) \leqslant \varepsilon, \\ 0 & \text{при всех остальных } x. \end{cases}$$

Аналогично доказывается, что все функционалы принадлежат $\mathscr{D}_1^A(C)$ и 2ε-различимы. Число всех функционалов равно $2^{M_{2\varepsilon}(A)}$, что доказывает неравенства (238).

Применим полученные неравенства к подсчетам в конкретных пространствах.

1) Пространство $F_1^n(C)$ функций, заданных на единичном кубе E^n и удовлетворяющих условию Липшица:

$$|f(x) - f(x')| \leqslant \max_{1 \leqslant i \leqslant n} |x_i - x_i'|.$$

В этом случае неравенства (238) дают

$$2^{(1/2\varepsilon)^n} \leqslant N_\varepsilon(F_1^n(C)) \leqslant \left(2\left[\frac{C(s+1)}{\varepsilon}\right] + 1\right)(s+1)^{((s+1)/2\varepsilon s)^n} \qquad (240)$$

при любом s.

Отсюда для величин, подобных плотностям из § 4,

$$\underline{\delta} = \liminf_{\varepsilon \to 0} \mathscr{H}_\varepsilon(F_1^n(C))\,\varepsilon^n, \qquad \overline{\delta} = \limsup_{\varepsilon \to 0} \mathscr{H}_\varepsilon(F_1^n(C))\,\varepsilon^n$$

получим

$$2^{-n} \leqslant \underline{\delta} \leqslant \overline{\delta} \leqslant \left(1 + \frac{1}{s}\right)^n \log s\, 2^{-n} \leqslant \frac{e \log n}{2^n}.$$

Более точные оценки $\overline{\delta}$, $\underline{\delta}$ их асимптотик, а также сам факт их совпадения или несовпадения неизвестны и представляют собой имеющие известный интерес задачи.

2) Пространство $\mathscr{D}_1^{F_1^\Delta(C)}$ *функционалов*, удовлетворяющих условию Липшица над пространством $F_1^\Delta(C)$ *функций* $f(x)$ на отрезке $\Delta = [0, 1]$, удовлетворяющих условию Липшица (см. § 2).

В этом случае неравенства (238) дают

$$2^{2^{[1/\varepsilon]-1}} \leqslant N_\varepsilon(\mathscr{D}_1^{F_1^\Delta(C)}(C_1)) \leqslant \left(2\left[\frac{C_1(s+1)}{\varepsilon}\right]+1\right)(s+1)^{2^{[(1+s)/s\varepsilon]+1}}. \tag{241}$$

Положив $s = [1/\varepsilon]$, мы получим

$$2^{1/\varepsilon} \preccurlyeq \mathscr{H}_\varepsilon(\mathscr{D}_1^{F_1^\Delta(C)}) \preccurlyeq \log(1/\varepsilon)\,2^{1/\varepsilon}. \tag{242}$$

Из (242), в частности, следует, что

$$\log \mathscr{H}_\varepsilon(\mathscr{D}_1^{F_1^\Delta(C)}) \sim 1/\varepsilon; \tag{243}$$

(243) представляет собой усиление формулы (237) для рассматриваемого пространства.

Добавление I

ТЕОРЕМА А. Г. ВИТУШКИНА О НЕВОЗМОЖНОСТИ ПРЕДСТАВЛЕНИЯ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ СУПЕРПОЗИЦИЯМИ ФУНКЦИЙ МЕНЬШЕГО ЧИСЛА ПЕРЕМЕННЫХ

Первый набросок излагаемой в этой статье теории был дан одним из авторов в качестве комментария к работе А. Г. Витушкина [2], посвященной представлению функций нескольких переменных функциями меньшего числа переменных.

В этом добавлении мы даем изложение задачи, которой занимался А. Г. Витушкин, и ее решение, использующее результаты § 5 этой статьи.

Пусть функция

$$y = f(x_1, \ldots, x_n)$$

задается последовательностью формул S:

$$\begin{aligned}
&y = \varphi(y_1, \ldots, y_m),\\
&y_k = \varphi_k(y_1^k, \ldots, y_{m_k}^k),\\
&y_{k_2}^{k_1} = \varphi_{k_1k_2}(y_1^{k_1k_2}, \ldots, y_{m_{k_1k_2}}^{k_1k_2}),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&y_{k_s}^{k_1,\ldots,k_{s-1}}\varphi_{k_1,\ldots,k_s}(y_1^{k_1,\ldots,k_s}, \ldots, y_{m_{k_1,\ldots,k_s}}^{k_1,\ldots,k_s}),\\
&y_k^{k_1,\ldots,k_s} = x_{r_k^{k_1,\ldots,k_s}}.
\end{aligned} \tag{244}$$

Областью определения функции f является множество тех точек $(x_1, \ldots, x_n)$, для которых при последовательном вычислении по формулам (244) мы всегда, вычислив все y с $r+1$ индексами, попадаем в области определения функций φ с r индексами. Функция f называется *суперпозицией функций φ порядка s, составленной по схеме S*. При этом считается, что схема суперпозиции определяется заданием таблиц натуральных чисел T:

n

m

$m_k,\ k = 1, \ldots, m,$

$m_{k_1k_2},\ k_1 = 1, \ldots, m;\ k_2 = 1, 2, \ldots, m_{k_1},$

$\ldots\ldots\ldots\ldots\ldots\ldots$

$m_{k_1, \ldots, k_s},\ k_1 = 1, 2, \ldots, m; \ldots; k_s = 1, 2, \ldots, m_{k_1 \ldots, k_{s-1}},$

$r_k^{k_1, \ldots, k_s} = 1, \ldots, n;\ k_1 = 1, 2, \ldots, m;\ k_s = 1, \ldots, m_{k_1, \ldots, k_{s-1}}.$

Мы будем предполагать, что все функции $\varphi_{k_1, \ldots, k_r}$ имеют в качестве областей определения единичные кубы

$$0 \leqslant y_k^{k_1, \ldots, k_r} \leqslant 1$$

соответствующего числа измерений. Тогда автоматически область определения f *содержится* в n-мерном единичном кубе. Мы будем, однако, рассматривать лишь суперпозиции, определенные на *всем* этом кубе.

Поставим теперь каждой последовательности индексов $k_1, \ldots$ $\ldots, k_r$ с $0 \leqslant r \leqslant s$ (при $r = 0$ «пустая» последовательность индексов), удовлетворяющей условиям таблицы T, в соответствие класс $\Phi_{k_1, \ldots, k_r}$ функций $\varphi_{k_1, \ldots, k_r}$ от $m_{k_1, \ldots, k_r}$ переменных. Класс получаемых при этом функций f будем обозначать $S(\Phi)$.

Теорема XXVI. *Пусть все функции $\varphi = \varphi_{k_1, \ldots, k_r}$ от $m = m_{k_1, \ldots, k_r}$ классов $\Phi_{k_1, \ldots, k_r}$ удовлетворяют условию Липшица*

$$|\varphi(y_1, \ldots, y_m) - \varphi(y_1', \ldots, y_m')| \leqslant C \sum_{k=1}^{m} |y_k - y_k'|$$

с одной и той же константой C; тогда в равномерной метрике

$$\|f\| = \sup |f|$$

(как для функций f, так и для функций φ)

$$\mathcal{H}_\varepsilon(S(\Phi)) \leqslant \mathcal{H}_{\varepsilon'}(\Phi_{k_1, \ldots, k_r}),$$

где

$$\varepsilon' = \varepsilon / C^s$$

и суммы берутся по всем наборам индексов $k_1, \ldots, k_r$, удовлетворяющим условиям таблицы T.

Для доказательства покроем каждое $\Phi_{k_1, \ldots, k_r}$

$$N_{k_1, \ldots, k_r} = N_{\varepsilon'}(\Phi_{k_1, \ldots, k_r})$$

множествами $U_{k_1,\ldots,k_r}$ диаметра $2\varepsilon'$. Выбрав для каждого набора индексов $k_1, \ldots, k_r$ одно определенное множество $U_{k_1,\ldots,k_r}$, образуем множество U тех функций f, которые можно получить суперпозициями по схеме S из функций $\varphi_{k_1,\ldots,k_r}$, принадлежащих выбранным $U_{k_1,\ldots,k_r}$. Легко видеть, что U имеет диаметр $\leqslant 2\varepsilon$. Число же различных множеств $U_{k_1,\ldots,k_r}$, очевидно, равно

$$N = \prod_{k_1,\ldots,k_r} N_{k_1,\ldots,k_r}.$$

Переходя к логарифмам, получаем требуемую оценку.

Допустим теперь (эти допущения далее будем именовать условиями (А)), что часть из классов $\Phi_{k_1,\ldots,k_r}$ состоит из одной-единственной функции (в каждом классе), удовлетворяющей условию Липшица с соответствующей константой $Z_{k_1,\ldots,k_r}$, часть же является классами

$$F_{q_{k_1,\ldots,k_r}}^{m_{k_1,\ldots,k_r}} (C_0^{k_1,\ldots,k_r}, \ldots, C_p^{k_1,\ldots,k_r}, C^{k_1,\ldots,k_r})$$

всех функций гладкости $q = q_{k_1,\ldots,k_r} = p + \alpha \geqslant 1$ с заданными константами $C_0, \ldots, C_p, C$[1]. Легко видеть, что в силу теоремы XXVI

$$\mathscr{H}_\varepsilon (S (\Phi)) \preccurlyeq (1/\varepsilon)^\rho, \tag{245}$$

где

$$\rho = \sup_{k_1,\ldots,k_r} (m_{k_1,\ldots,k_r}/q_{k_1,\ldots,k_r}).$$

Рассмотрим пересечение $S (\Phi)$ с классом F_q^n определенных на n-мерном единичном кубе функций f гладкости q. В F_q^n введем топологию при помощи нормы $\| f \|$ (см. замечание 3 из конца § 5).

Имеет место

Лемма VIII. *При условиях* (А) $\rho < n/q$ *пересечение* $S (\Phi) \cap F_q^n$ *нигде не плотно в* F_q^n.

Лемма является непосредственным следствием оценки (245) и замечания 3 из конца § 5. Так как с указанной там метрикой F_q^n является полным пространством, то дополнение соединения счетного числа нигде не плотных в F_q^n множеств всюду плотно. Это и позволяет вывести из нашей леммы следующую далее теорему XXVII, являющуюся некоторым обобщением теоремы, доказанной в 1954 г. А. Г. Витушкиным с привлечением существенно других средств.

---

[1] См. § 5.

Обозначим через

$$\Sigma_\rho^n(\psi_1, \ldots, \psi_t)$$

класс функций f, определенных в n-мерном единичном кубе и представимых в виде суперпозиций (по любой схеме любого конечного порядка s) конечного числа фиксированных функций

$$\psi_1, \ldots, \psi_t$$

и конечного числа функций φ, для каждой из которых отношение m_φ/q_φ числа переменных к гладкости $\leqslant \rho$.

Теорема XXVII. *Если $\rho < n/q$ и функции $\psi_1, \ldots, \psi_t$ удовлетворяют условию Липшица (с любыми константами $Z_1, \ldots, Z_t$), то множество функций f из F_q^n, не входящих в $\Sigma_\rho^n(\psi_1, \ldots, \psi_t)$, всюду плотно в F_q^n.*

Для доказательства надо лишь представить $\Sigma_\rho^n(\psi_1, \ldots, \psi_t)$ в виде счетного соединения множеств, подходящих под условия леммы. Такими множествами могут служить множества функций, удовлетворяющих условиям (A) при той или иной (фиксированной для каждого множества) схеме суперпозиции, фиксированных наборов индексов, для которых $\Phi_{k_1, \ldots, k_r}$ состоит из той или иной из функций $\psi_1, \ldots, \psi_t$ и фиксированных натуральных значений констант $C_0^{k_1, \ldots, k_r}, \ldots, C_p^{k_1, \ldots, k_r}, C^{k_1, \ldots, k_r}$. Очевидно, что число различных таких множеств лишь счетно.

Добавление II

СВЯЗЬ С ВЕРОЯТНОСТНОЙ ТЕОРИЕЙ ПРИБЛИЖЕННОЙ ПЕРЕДАЧИ СИГНАЛОВ

1. Каждый элемент x множества A, лежащего в метрическом пространстве R, будем рассматривать как возможный «сигнал на входе» передающего или запоминающего устройства. Будем считать, что соответствующий «сигнал на выходе» есть случайный элемент η пространства R, имеющий распределение вероятностей

$$\mathcal{P}(\eta \in Q) = \Pi(x, Q),$$

зависящее от x. При помощи ядра $\Pi(x, Q)$ из любого распределения

$$\mathcal{P}(\xi \in B) = \mathcal{P}_\xi(B)$$

на множестве A образуем распределение

$$\mathcal{P}(\eta \in Q) = \int_A \mathcal{P}_\xi(dx)\, \Pi(x, Q) = \mathcal{P}_\eta(Q)$$

и совместное распределение

$$\mathscr{P}((\xi, \eta) \in U) = \iint_U \mathscr{P}_\xi(dx)\, \Pi(x, dy) = \mathscr{P}_{\xi\eta}(U).$$

Пусть, наконец,

$$\mathscr{I}_{\mathscr{P}_\xi}(\xi, \eta) = \int_A \int_R \mathscr{P}_{\xi\eta}(dx, dy) \log \frac{\mathscr{P}_{\xi\eta}(dx, dy)}{\mathscr{P}_\xi(dx)\, \mathscr{P}_\eta(dy)}$$

есть шенноновское количество информации, содержащейся в η с условным распределением Π относительно ξ, распределенного по закону $\mathscr{P}_\xi$ (см. [27]). В соответствии с Шенноном величина

$$\mathscr{E}(\Pi) = \sup_{\mathscr{P}_\xi} \mathscr{I}^{\Pi}_{\mathscr{P}_\xi}$$

является *емкостью* устройства, определяемого распределением Π.

Нас будет занимать случай, когда с вероятностью единица выходной сигнал η находится на расстоянии не более ε от входного сигнала x:

$$\Pi(x, \{y: \rho(x, y) \leqslant \varepsilon\}) = 1. \tag{246}$$

При условии (246) будем говорить, что ядро Π принадлежит классу K_ε.

Теорема XXVIII. *Если* $\Pi \in K_\varepsilon$, *то*

$$\mathscr{E}(\Pi) \geqslant \mathscr{E}_{2\varepsilon}(A).$$

В самом деле, если расположить входной сигнал в $M_{2\varepsilon}(A)$ точках максимального (по числу элементов) 2ε-различимого множества с вероятностью $M_{2\varepsilon}^{-1}(A)$ для каждой точки, то значение η будет определяться по значению ξ с достоверностью (вероятность ошибки равна нулю), т. е. количество информации $\mathscr{I}(\xi, \eta)$ будет равно энтропии распределения ξ

$$\log M_{2\varepsilon}(A) = \mathscr{E}_{2\varepsilon}(A).$$

Теорема XXIX. *Существует такое ядро* $\Pi \in K_\varepsilon$, *что*

$$\mathscr{E}(\Pi) = \mathscr{H}^R_\varepsilon(A).$$

Для построения ядра, удовлетворяющего теореме XXIX, достаточно взять в R множество U из $N^R_\varepsilon(A)$ точек, являющееся ε-сетью для A, и поставить каждому $x \in A$ в соответствие с вероятностью единица одну из точек $y \in U$ с $\rho(x, y) \leqslant \varepsilon$ в качестве выходного сигнала η. Тогда энтропия η, а следовательно и $\mathscr{I}(\xi, \eta)$, будет не больше $\mathscr{H}^R_\varepsilon(A)$ (независимо от выбора $\mathscr{P}_\xi$).

Рассмотрим теперь величину

$$\mathscr{E}^R_\varepsilon(A) = \inf_{\Pi \in K_\varepsilon} \mathscr{E}(\Pi).$$

Это гарантированная емкость передающего устройства при точности передачи ε (когда, кроме (246), об условных распределениях П ничего неизвестно). Из теорем XXVIII и XXIX непосредственно вытекают оценки

$$\mathscr{E}_{2\varepsilon}(A) \leqslant \mathscr{E}_{\varepsilon}^{R}(A) \leqslant \mathscr{H}_{\varepsilon}^{R}(A). \tag{247}$$

2. Наряду с величиной $\mathscr{E}$ (П) в вопросах передачи сигналов с точностью до ε имеет значение величина

$$\mathscr{H}_{\varepsilon}^{R}(\mathscr{P}_{\xi}) = \inf_{\Pi \in K_{\varepsilon}} \mathscr{I}_{\mathscr{P}_{\xi}}^{\Pi},$$

называемая «энтропией распределения $\mathscr{P}_{\xi}$ при условии (246) на точность передачи», или, короче, ε-энтропией распределения $\mathscr{P}_{\xi}$.

Теорема XXX.

$$\widetilde{\mathscr{H}}_{\varepsilon}^{R}(A) = \sup_{\mathscr{P}_{\xi}} \mathscr{H}_{\varepsilon}^{R}(\mathscr{P}_{\xi}) \leqslant \mathscr{E}_{\varepsilon}^{R}(A).$$

Теорема XXX непосредственно вытекает из известной общей формулы

$$\sup_{x} \inf_{y} F(x, y) \leqslant \inf_{y} \sup_{x} F(x, y).$$

Теорема XXXI. *Существует* $\mathscr{P}_{\xi}$, *для которого*

$$\mathscr{H}_{\varepsilon}^{R}(\mathscr{P}_{\xi}) \geqslant \mathscr{E}_{2\varepsilon}(A).$$

Для доказательства достаточно рассмотреть распределение $\mathscr{P}_{\xi}$, приписывающее каждой из $M_{2\varepsilon}(A)$ точек максимального 2ε-различимого множества U вероятность $M_{2\varepsilon}^{-1}(A)$.

Из теорем XXX и XXXI и формулы (247) получаем

$$\mathscr{E}_{2\varepsilon}(A) \leqslant \widetilde{\mathscr{H}}_{\varepsilon}^{R}(A) \leqslant \mathscr{H}_{\varepsilon}^{R}(A). \tag{248}$$

Из теоремы XXX и формул (247), (248) и теоремы IV получим, наконец,

$$\mathscr{E}_{2\varepsilon}^{R}(A) \leqslant \widetilde{\mathscr{H}}_{\varepsilon}^{R}(A) \leqslant \mathscr{E}_{\varepsilon}^{R}(A). \tag{249}$$

Правое неравенство (249) совпадает с теоремой XXX. Для доказательства левого надо заметить, что в силу (247), теоремы IV и (248)

$$\mathscr{E}_{2\varepsilon}^{R}(A) \leqslant \mathscr{H}_{2\varepsilon}^{R}(A) \leqslant \mathscr{E}_{2\varepsilon}(A) \leqslant \widetilde{\mathscr{H}}_{\varepsilon}^{R}(A).$$

15 декабря 1958 г.

ЛИТЕРАТУРА

1. *Бабенко К. И.* О энтропии одного класса аналитических функций.— Науч. докл. высш. шк., 1958, т. 1, № 2.
2. *Витушкин А. Г.* К тринадцатой проблеме Гильберта.— ДАН СССР, 1954, т. 95, № 4, с. 701—704.
3. *Витушкин А. Г.* Абсолютная энтропия метрических пространств.— ДАН СССР, 1957, т. 117, № 5, с. 745—748.

4. *Ерохин В. Д.* Об асимптотике ε-энтропии аналитических функций.— ДАН СССР, 1958, т. 120, № 5, с. 949—952.
5. *Колмогоров А. Н.* Оценки минимального числа элементов ε-сетей в различных функциональных классах и их применение к вопросу о представимости функций нескольких переменных суперпозициями функций меньшего числа переменных.— УМН, 1955, т. 10, вып. 1, с. 192—194.
6. *Колмогоров А. Н.* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств.— ДАН СССР, 1956, т. 108, № 3, с. 385—389.
7. *Тихомиров В. М.* Об ε-энтропии некоторых классов аналитических функций.— ДАН СССР, 1957, т. 117, № 2, с. 191—194.
8. *Гуревич В., Волмен Г.* Теория размерности. М.: Изд-во иностр. лит., 1948.
9. *Hausdorff F.* Dimention und äußeres Maß.— Math. Ann., 1919, Bd. 79, S. 157—179.
10. *Бахвалов Н. С.* К вопросу о числе арифметических действий при решении уравнения Пуассона для квадрата методом конечных разностей.— ДАН СССР, 1957, т. 113, № 2, с. 252—255.
11. *Бахвалов Н. С.* О составлении уравнений в конечных разностях при приближенном решении уравнения Лапласа.— ДАН СССР, 1957, т. 114, № 6, с. 1146—1149.
12. *Витушкин А. Г.* Некоторые оценки теории табулирования.— ДАН СССР, 1957, т. 114, № 5, с. 923—926.
13. *Колмогоров А. Н.* О линейной размерности топологических векторных пространств.— ДАН СССР, 1958, т. 120, № 2, с. 239—241.
14. *Ерохин В. Д.* О конформных преобразованиях колец и об основном базисе пространства функций, аналитических в элементарной окрестности произвольного континуума.— ДАН СССР, 1958, т. 120, № 4, с. 689—692.
15. *Александров П. С.* Введение в общую теорию множеств и функций. М.; Л.: Гостехиздат, 1948.
16. *Урысон П. С.* Труды по топологии и другим областям математики. М.: Гостехиздат, 1951. Т. 2.
17. *Banach S.* Théorie des operation linéaires. Warszawa, 1932.
18. *Бернштейн С. Н.* Теория вероятностей. М.; Л.: Гостехиздат, 1946.
19. *Тот Л. Ф.* Расположения на плоскости, сфере и в пространстве. М.: Физматгиз, 1958.
20. *Davenport H.* The covering of space by spheres Reneli.— Rend. Circ. Palermo, 1952, vol. 1, N 2, p. 92—107.
21. *Watson L.* The covering of space by spheres.— Rend. Circ. Palermo, 1956, vol. 5, N 1, p. 93—100.
22. *Blichfeldt H. F.* The minimum value of quadratic forms, and the dosest packing of spheres.— Math. Ann., 1929, Bd. 101, S. 605—608.
23. *Ахиезер Н. И.* Лекции по теории аппроксимации. М.; Л.: Гостехиздат, 1947.
24. *Фихтенгольц Г. М.* Курс дифференциального и интегрального исчисления. М.: Гостехиздат, 1951. Т. 2; 1949. Т. 3.
25. *Маркушевич А. И.* Теория аналитических функций. М.; Л.: Гостехиздат, 1950.
26. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи.— В кн.: Матер. к I Всесоюз. съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. М., 1933.
27. *Shannon C.* A mathematical theory of communication.— Bell Syst. Techn. J., 1948, vol. 27, N 3, p. 379—423; N 4, p. 623—656.
28. *Левин Б. Я.* Распределение корней целых функций. М.: Гостехиздат, 1956.
29. *Лаврентьев М. А., Шабат Б. В.* Методы теории функций комплексного переменного. М.: Физматгиз, 1958.

8
РАЗЛИЧНЫЕ ПОДХОДЫ К ОЦЕНКЕ ТРУДНОСТИ ПРИБЛИЖЕННОГО ЗАДАНИЯ И ВЫЧИСЛЕНИЯ ФУНКЦИЙ *

Чтобы избежать технических осложнений, мы будем рассматривать действительные функции

$$y = f(x),$$

заданные на отрезке

$$\Delta = \{x: \ -1 \leqslant x \leqslant 1\}.$$

Все наиболее существенные черты проблематики, которая нас будет интересовать, можно понять на примере функциональных классов W_p функций, удовлетворяющих неравенству

$$\operatorname{vrai} \max |f^{(p)}(x)| \leqslant 1,$$

A_r-функций, которые могут быть аналитически продолжены во внутренность эллипса с фокусами в точках -1 и $+1$ и суммой полуосей $r > 1$.

Мы будем изучать способы задания и вычисления функций указанных классов с точностью до некоторого фиксированного $\varepsilon > 0$

1
ПРИБЛИЖЕНИЕ АЛГЕБРАИЧЕСКИМИ МНОГОЧЛЕНАМИ

Будем обозначать через $n_a(f, \varepsilon)$ наименьшее n, при котором существует многочлен

$$P(x) = c_1 + c_2 x + \ldots + c_n x^{n-1},$$

удовлетворяющий на Δ неравенству

$$|f - P| < \varepsilon,$$

а для функционального класса F положим

$$n_a(F, \varepsilon) = \sup_{f \in F} n_a(f, \varepsilon).$$

Вот наиболее окончательные результаты об асимптотическом поведении функций $n_a(F, \varepsilon)$ для классов W_p и A_r, которые дает классическая теория наилучших приближений:

$$n_a(W_p, \varepsilon) \sim C_p (1/\varepsilon)^{1/p} \tag{1}$$

---

* Proc. Intern. Congr. Math. Stockholm, 1963, p. 369—376.

где C_p — некоторые константы, аналитическое выражение которых мы приводить не будем (Никольский, Бернштейн, 1946—1947);

$$n_a(A_r, \varepsilon) \backsim \frac{1}{\log r} \log \frac{1}{\varepsilon} \tag{2}$$

(Бернштейн, 1913).

2
ПРИБЛИЖЕНИЕ ЛИНЕЙНЫМИ ФОРМАМИ[1]

Естественно поставить вопрос о том, насколько удачен выбор алгебраических многочленов в качестве аппарата для приближенного представления функций классов W_p и A_r (кажется, проблематика этого рода была впервые выдвинута в [3]). Обозначим через $n(F, \varepsilon)$ наименьшее n, при котором можно найти такие функции $\varphi_1, \varphi_2, \ldots, \varphi_n$, что для любой функции $f \in F$ найдется линейная форма

$$L(x) = c_1\varphi_1(x) + \ldots + c_n\varphi_n(x),$$

удовлетворяющая на Δ неравенству

$$|f - L| < \varepsilon.$$

Легко понять, что всегда

$$n(F, \varepsilon) \leqslant n_a(F, \varepsilon).$$

В случае класса A_r

$$n(F, \varepsilon) \backsim n_a(F, \varepsilon) \backsim \frac{1}{\log r} \log \frac{1}{\varepsilon} \tag{3}$$

(Витушкин, Ерохин, 1957—1958).

В случае класса W_p алгебраические многочлены оказываются нерациональным способом приближения, так как

$$n(F, \varepsilon) \backsim C_p (1/\varepsilon)^{1/p}, \tag{4}$$

где

$$C'_p / C_p = 2/\pi \tag{5}$$

(Тихомиров, 1960).

Иначе говоря, выбирая рационально функции φ_k, можно при той же точности ε ограничиться формами, содержащими в

$$\pi/2 = 1{,}57 \ldots$$

раз меньше членов, чем при приближении многочленами.

[1] Современное изложение проблематики наших § 1,2 см. [1]. Вместо (2) в [1] сообщаются найденные позднее обобщения. В более старомодной формулировке (2) можно найти уже в § 46 классического мемуара Бернштейна [2]. В (1), (2) и далее $n(\varepsilon) \backsim m(\varepsilon)$, если $n(\varepsilon)/m(\varepsilon) \to 1$ при $\varepsilon \to 0$; $n(\varepsilon) \succcurlyeq m(\varepsilon)$, если $\liminf_{\varepsilon\to 0} (n(\varepsilon)/m(\varepsilon)) > 0$; $n(\varepsilon) \asymp m(\varepsilon)$, если одновременно $n(\varepsilon) \succcurlyeq m(\varepsilon)$ и $m(\varepsilon) \succcurlyeq n(\varepsilon)$.

3
НЕОБХОДИМОЕ КОЛИЧЕСТВО ИНФОРМАЦИИ

Чтобы фиксировать с точностью до ε функцию f, принадлежащую к классу F, в котором все функции допускают представление линейными формами

$$L(x) = c_1\varphi_1(x) + \ldots + c_n\varphi_n(x),$$

с ошибкой, не превышающей

$$|f - L| \leqslant \varepsilon/2,$$

достаточно указать коэффициенты c_k с ошибкой не больше

$$\varepsilon' = \varepsilon/2nM,$$

где M — верхняя грань абсолютных величин функций φ_k.

Приближенные значения коэффициентов можно записывать по двоичной системе счисления. Таким образом, задание функции сводится к заданию определенного числа знаков, каждый из которых может принимать лишь значения 0 и 1. Если ограничиться функциями, подчиненными дополнительному условию

$$|f(x)| \leqslant 1,$$

то при приближении функций классов W_p и A_r алгебраическими многочленами при $\varepsilon \to 0$ число знаков, с которым необходимо задавать коэффициенты, оказывается порядка $\log(1/\varepsilon)$.

Это позволяет получить следующие оценки «количества информации», необходимого для фиксации с точностью до ε произвольной функции f класса A_r или класса W_p:

$$I(A_r, \varepsilon) \preccurlyeq (\log(1/\varepsilon))^2, \tag{6}$$

$$I(W_p, \varepsilon) \preccurlyeq (1/\varepsilon)^{1/p}\log(1/\varepsilon). \tag{7}$$

Первая из этих оценок не может быть усилена в смысле порядка величины $I(A_r, \varepsilon)$, но может быть уточнена:

$$I(A_r, \varepsilon) \sim \frac{1}{2\log r}\left(\log\frac{1}{\varepsilon}\right)^2 \tag{8}$$

(Витушкин, 1957).

Оценка же (7) дает немного завышенный порядок величины $I(W_p, \varepsilon)$. Истинный порядок величины таков:

$$I(W_p, \varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{1/p} \tag{9}$$

(Колмогоров, 1955).

Более подробно о величине $I(F, \varepsilon)$ см. во введении к статье [4]. Там показывается, что $I(F, \varepsilon)$ есть целое число, определяемое неравенствами

$$\log_2 N(F, \varepsilon) \leqslant I(F, \varepsilon) < \log_2 N(F, \varepsilon) + 1, \tag{10}$$

где $N(F, \varepsilon)$ — минимальное число элементов покрытия множества F множествами диаметра 2ε в метрике

$$\rho(f_1, f_2) = \sup_{x \in \Delta} | f_1(x) - f_2(x) |.$$

4
СЛОЖНОСТЬ ВЫЧИСЛЕНИЯ

Для простоты ограничимся в этом разделе функциями, которые, кроме условия

$$| f(x) | \leqslant 1,$$

введенного в предшествующем разделе, подчинены условию Липшица

$$| f(x'') - f(x') | \leqslant 1.$$

Пусть

$$s \geqslant \log(1/\varepsilon) + 3.$$

Легко видеть, что любая таблица, которая каждому двоичному числу

$$x = -1 + x_1, x_2, \ldots, x_s \quad (x_i = 0, 1)$$

ставит в соответствие двоичное число

$$y = -1 + y_1, y_2, \ldots, y_s = f_\varepsilon(x) \quad (y_i = 0, 1), \tag{11}$$

удовлетворяющее условию

$$| f(x) - f_\varepsilon(x) | \leqslant \varepsilon/2,$$

может рассматриваться как таблица, дающая значения $y = f(x)$ с точностью до ε для любого $x \in \Delta$.

Зависимость (11) может интерпретироваться как дискретная векторная функция

$$Y = \Phi(X),$$

отображающая s-мерные векторы

$$X = (x_1, \ldots, x_s)$$

в s-мерные векторы

$$Y = (y_1, \ldots, y_s),$$

где переменные x_j и y_i принимают лишь значения 0 и 1.

Такого рода векторные «функции алгебры логики» естественно представлять в виде суперпозиций элементарных функций

$$\zeta = f(\xi, \eta),$$

где вспомогательные переменные ξ, η, ζ принимают лишь значения 0, 1. Известно, что таких элементарных функций имеется всего 16. Будем называть *сложностью* векторной функции алгебры логики $Y = \Phi(X)$ минимальное число членов в представлении ее в виде суперпозиции элементарных функций (см. [5]).

Сложностью ε-*задания* функции f мы будем называть минимальную сложность функций Φ, соответствующих таблицам, задающим $y = f(x)$ с точностью до ε.

Соединяя результаты заметок [5] и [6] с некоторыми приемами, разработанными Шенноном, и оценками, изложенными в разделе 3, Офман установил, что максимальные ε-сложности функций из W_p и A_p допускают оценки

$$K(W_p, \varepsilon) \asymp (1/\varepsilon)^{1/p}/\log(1/\varepsilon), \tag{12}$$

$$\frac{(\log(1/\varepsilon))^2}{\log\log(1/\varepsilon)} \preccurlyeq K(A_r, \varepsilon) \preccurlyeq \left(\log\frac{1}{\varepsilon}\right)^{\alpha}, \quad \alpha = \log_2 3 + 1 = 2{,}58\ldots \tag{13}$$

Незавершенность оценок (13) связана с тем обстоятельством, что до настоящего времени, по-видимому, остается неизвестной асимптотика числа $m(s)$ элементарных операций над парами двоичных знаков, необходимого для выполнения умножения двух s-значных чисел. Обычный школьный рецепт умножения дает лишь

$$m(s) \preccurlyeq s^2,$$

а оценки Офмана и Карацубы таковы:

$$s \preccurlyeq m(s) \preccurlyeq s^{\log_2 3} \tag{14}$$

5
ОЦЕНКА СЛОЖНОСТИ ИНДИВИДУАЛЬНЫХ ФУНКЦИЙ

В классической теории приближений алгебраическими многочленами известны «обратные теоремы», показывающие, что ни одна индивидуальная функция, не обладающая достаточной «гладкостью», не может хорошо приближаться многочленами не слишком высоких степеней на всем отрезке Δ.

Все сказанное в разделах 2 и 3 по существу относится к классам функций, а не к индивидуальным функциям. Но при подходе к делу, изложенному в разделе 4, задача оценки ε-сложностей $K(f, \varepsilon)$ индивидуальных функций содержательна и очень интересна. К сожалению, получение оценок снизу здесь, видимо, очень трудно. Нарушение гладкости не делает еще неизбежным образом функцию сложной. Например, для функции Ван дер Вардена, не имеющей ни в одной точке производной, легко получается оценка

$$K(f, \varepsilon) \preccurlyeq (\log(1/\varepsilon))^2. \tag{15}$$

ЛИТЕРАТУРА

1. *Тихомиров В. М.* Поперечники множеств в функциональных пространствах и теория наилучших приближений.— УМН, 1960, т. 15, вып. 3, с. 81—120.
2. *Bernstein S.* Sur la valeur asymptotique de la meilleure approximation des fonctions analytiques admettant des singularités données.— Bull. Acad. Roy. Belg. Ser. 2, 1913, vol. 4, p. 76—90.
3. *Kolmogoroff A.* Ueber die beste Annäherung von Funktionen einer gegebenen Funktionenklasse.— Ann. Math., 1936, vol. 37, p. 107—110.
4. *Колмогоров А. Н., Тихомиров В. М.* ε-энтропия и ε-емкость множеств в функциональных пространствах.— УМН, 1959, т. 14, вып. 2, с. 3—86.
5. *Офман Ю.* Об алгоритмической сложности дискретных функций.— ДАН СССР, 1962, т. 145, № 1, с. 48—51.
6. *Карацуба А., Офман Ю.* Умножение многозначных чисел на автоматах.— ДАН СССР, 1962, т. 145, № 2, с. 293—294.

9

О ТАБЛИЦАХ СЛУЧАЙНЫХ ЧИСЕЛ *[1]

От автора

Редакция «Семиотики и информатики» сочла целесообразным опубликовать русский перевод моей статьи, отражающей определенный этап моих попыток осмыслить частотную интерпретацию вероятностей Р. фон Мизеса. Статья была написана во время моего пребывания в Индии в 1962 г. и опубликована в индийском журнале «Санкхиа», весьма авторитетном у специалистов по математической статистике англоязычных стран, но мало распространенном у нас в СССР. Знакомый с работами Мизеса читатель, может быть, обратит внимание на более широкое определение алгоритма формирования выборок A. Но главное отличие от Мизеса состоит в строго финитном характере всей концепции и во введении количественной оценки устойчивости частот. Задача сближения оценок (4.3) и (4.4) при всей ее элементарности, кажется, еще ждет своего решения. Дальнейшие этапы моих поисков отражены в двух заметках, опубликованных в 1965 и 1969 гг. в журнале «Проблемы передачи информации». В «Семиотике и информатике» (1981, вып. 16) см. обзор В. В. Вьюгина.

1
ВВЕДЕНИЕ

Теоретико-множественные аксиомы теории вероятностей, в формулировке которых мне довелось принять участие [1], позволили устранить большинство трудностей в построении математи-

---

* On tables of random numbers.— Sankhya. Indian J. Statist. Ser. A, 1963, vol. 25, N 4, p. 369—376. *Перевод А. Х. Шеня.*

[1] Эта статья была опубликована в периодическом сборнике «Семиотика и информатика» (М.: ВИНИТИ, 1982, вып. 18, с. 3—13), причем публикации перевода был предпослан специально написанный А. Н. Колмогоровым и воспроизводимый здесь текст «От автора».— *Примеч. ред.*

ческого аппарата, пригодного в большом числе приложений вероятностных методов, причем настолько успешно, что проблема отыскания причин применимости математической теории вероятностей стала казаться многим исследователям второстепенной.

Я уже высказывал точку зрения [1, гл. 1] [2], что основой применимости математической теории вероятностей к случайным явлениям реального мира является *частотный подход к вероятности* в той или иной форме, неизбежность обращения к которому горячо отстаивал фон Мизес. Тем не менее в течение длительного времени я считал, что:

(1) частотный подход, основанный на понятии *предельной частоты* при стремящемся к бесконечности числе испытаний, не позволяет обосновать применимость результатов теории вероятностей к практическим задачам, в которых мы имеем дело с конечным числом испытаний;

(2) частотный подход в случае большого, но конечного числа испытаний не может быть развит строго формально, чисто математически.

В соответствии с этим я иногда выдвигал частотный подход к вероятности, включавший сознательное использование некоторых не вполне формальных соображений о «практической достоверности», «приблизительного постоянства частот при больших сериях испытаний» без точного описания того, какие серии «достаточно велики» и т. д. (см. [1—3]: «Основные понятия теории вероятностей», глава 1, и более подробно в БСЭ статью «Вероятность» и главу о теории вероятностей в книге «Математика, ее содержание, методы и значение»).

Я по-прежнему придерживаюсь первого из указанных положений. Что касается второго, то я пришел к выводу, что понятие случайного распределения может быть введено строго формально, а именно: можно показать, что в достаточно больших совокупностях распределение некоторого свойства может быть таким, что частота его появления будет примерно одинаковой для всех достаточно больших выборок, если только *закон выбора достаточно прост*. Полное развитие такого подхода предполагает введение меры сложности алгоритмов. Я рассчитываю обсудить этот вопрос в другой статье. В настоящей статье я буду использовать лишь тот факт, что *число простых алгоритмов не может быть очень велико*.

Для определенности мы будем рассматривать таблицу

$$T = (t_1, \ldots, t_N)$$

из N нулей и единиц: $t_k = 0$ или 1. Такая таблица будет назы-

[2] Впервые эта книга появилась на немецком языке: *Kolmogoroff A.* Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin, 1933. Рус. пер.: *Колмогоров А. Н.* Основные понятия теории вероятностей. М.; Л.: ОНТИ, 1936; 2-е изд. (под тем же загл.). М.: Наука, 1974.— *Примеч. пер.*

ваться случайной, если при различных способах выбора подмножества A достаточно большого объема n из $\{1, \ldots, N\}$ частота

$$\pi(A) = \frac{1}{n} \sum_{k \in A} t_k$$

появления единиц в A приблизительно одинакова. Можно, например, выбрать в качестве A:

(а) множество первых n четных чисел;

(б) множество первых n простых чисел $p_1, \ldots, p_n$ — и т. д.

Обычное понятие «случайности» не сводится к постоянству частот при не зависящих от строения таблицы способах выбора. Можно, например, выбрать множество A так:

(в) множество первых n значений $k \geqslant 2$, для которых $t_{k-1} = 0$;

(г) множество первых n значений $k > s$, для которых $t_{k-1} = a_1, t_{k-2} = a_2, \ldots, t_{k-s} = a_s$;

(д) множество первых n четных чисел $k = 2i$, для которых $t_i = 1$;

(е) множество чисел $k_1, k_2, \ldots, k_n$, выбранных по правилу

$$k_1 = 1;$$

$$k_{i+1} = k_i + 1 + t_{k_i} p_i$$

и так далее.

Точное определение понятия «допустимого алгоритма» выбора будет дано в разделе 2.

Если для таблицы достаточно большого размера N по крайней мере один простой тест на случайность такого типа с достаточно большим размером выборки n показывает «значительное» отклонение от принципа постоянства частот, мы немедленно отвергаем гипотезу о «чисто случайном» происхождении рассматриваемой таблицы.

2
ДОПУСТИМЫЕ АЛГОРИТМЫ ВЫБОРА И (n, ε)-СЛУЧАЙНЫЕ ТАБЛИЦЫ

Допустимый алгоритм выбора множества

$$A = R(T) \subset \{1, \ldots, N\}$$

в соответствии с таблицей T размера N определяется функциями [3]

$$F_0, G_0, H_0;$$

$$F_1(\xi_1, \tau_1), \quad G_1(\xi_1, \tau_1), \quad H_1(\xi_1, \tau_1);$$

[3] Функции в первой строке являются константами (функциями от пустого множества аргументов).

$$F_2(\xi_1, \tau_1; \xi_2, \tau_2), \quad G_2(\xi_1, \tau_1; \xi_2, \tau_2), \quad H_2(\xi_1, \tau_1; \xi_2, \tau_2);$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$F_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1}),$$

$$G_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1}),$$

$$H_{N-1}(\xi_1, \tau_1; \xi_2, \tau_2; \ldots; \xi_{N-1}, \tau_{N-1}),$$

где переменные τ_k и функции G_k и H_k принимают значения 0 и 1, а переменные ξ_k и функции F_k принимают значения из множества $\{1, \ldots, N\}$. Функции F_k должны удовлетворять дополнительному условию

$$F_k(\xi_1, \tau_1; \ldots; \xi_k, \tau_k) \neq \xi_i. \tag{2.1}$$

Применение алгоритма состоит в образовании последовательности

$$x_1 = F_0,$$

$$x_2 = F_1(x_1, t_{x_1}),$$

$$x_3 = F_2(x_1, t_{x_1}; x_2, t_{x_2}),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$x_s = F_{s-1}(x_1, t_{x_1}; \ldots; x_{s-1}, t_{x_{s-1}}) \tag{2.2}$$

и определении того, какие элементы этой последовательности включаются в A. Последовательность оканчивается, как только значение [4]

$$H_s(x_1, t_{x_1}; \ldots x_s, t_{x_s}) = 1 \tag{2.3}$$

появляется. В этом случае последним членом последовательности будет x_s. Если

$$H_k(x_1, t_{x_1}; \ldots; x_k, t_{x_k}) = 0$$

при всех $k < N$, то последовательность оканчивается элементом x_s при $s = N$, т. е. исчерпывает все элементы множества $\{1, \ldots, N\}$: ведь согласно условию (2.1) все элементы последовательности (2.2) различны.

Те элементы x_k, для которых

$$G_{k-1}(x_1, t_{x_1}; \ldots; x_{k-1}, t_{x_{k-1}}) = 1, \tag{2.4}$$

образуют множество A. Мне кажется, что эта конструкция правильно отражает замысел фон Мизеса во всей его общности; при этом сохраняется основное ограничение — при определении того, входит ли $x \in \{1, \ldots, N\}$ в A, не используется значение t_x.

---

[4] В частности, если $H_0 = 1$, выбор не может начаться и множество A пусто.

Пусть дана система

$$\mathscr{R}_N = \{\mathrm{R}\}$$

допустимых алгоритмов выбора (размер N таблицы фиксирован).

Определение. Таблица T размера N называется (n, ε)-*случайной* по отношению к системе $\mathscr{R}_N$, если существует такая константа p, $0 \leqslant p \leqslant 1$, что для всякого $A = \mathrm{R}(T)$, $\mathrm{R} \in \mathscr{R}_N$ с числом элементов $v \geqslant n$ частота $\pi(A) = \frac{1}{v} \sum\limits_{k \in A} t_k$ удовлетворяет неравенству $|\pi(A) - p| \leqslant \varepsilon$.

Иногда удобно говорить о (n, ε, p)-случайности при фиксированном p. Имеет место следующая

Теорема 1. *Если число элементов системы $\mathscr{R}_N$ не превосходит*

$$\tau(n, \varepsilon) = {}^1/_2 e^{2n\varepsilon^2(1-\varepsilon)}, \tag{2.5}$$

то для любого p, $0 \leqslant p \leqslant 1$, существует таблица T размера N, являющаяся (n, ε, p)-случайной по отношению к $\mathscr{R}_N$.

Смысл входящей в теорему оценки станет более ясным, если мы введем двоичный логарифм $\lambda(\mathscr{R}_N) = \log_2 \rho(\mathscr{R}_N)$ числа элементов ρ системы $\mathscr{R}_N$. Число $\lambda(\mathscr{R}_N)$ равно количеству информации, необходимому для указания конкретного элемента R из $\mathscr{R}_N$. Ясно, что в случае большого $\lambda(\mathscr{R}_N)$ система $\mathscr{R}_N$ должна содержать алгоритмы, само описание (а не только действительное выполнение) которых сложно (требует не менее $\lambda(\mathscr{R}_N)$ двоичных знаков).

В нашей теореме условие существования (n, ε)-случайных по отношению к $\mathscr{R}_N$ таблиц при произвольном p записано в виде неравенства

$$\lambda(\mathscr{R}_N) \leqslant 2(\log_2 e) n\varepsilon^2 (1 - \varepsilon) - 1.$$

Такая количественная формулировка содержащегося в теореме результата поучительна сама по себе. Если отношение λ/n достаточно мало, то для любого заранее выбранного ε и для любых N и p и любой системы допустимых алгоритмов с

$$\lambda(\mathscr{R}_N) \leqslant \lambda$$

существуют (n, ε)-случайные по отношению к ней таблицы.

Доказательство этой теоремы будет дано в разделе 3. В разделе 4 мы исследуем возможность улучшения данных в ней оценок. Сделаем еще два замечания.

Замечание 1. Поскольку алгоритм выбора множества определяется функциями F_k, G_k, H_k, естественно считать два алгоритма совпадающими, если и только если соответствующие функции F_k, G_k, H_k совпадают. Уже с этой точки зрения число различных возможных алгоритмов выбора для данного N конечно.

Можно встать на другую точку зрения, считая два алгоритма различными только в том случае, если они дают различные множества $A = \mathrm{R}\,(T)$ хотя бы для одной таблицы T. С этой точки зрения число различных алгоритмов еще сокращается и не превосходит

$$(2^N)^{2^N} = 2^{N \cdot 2^N}.$$

Вопрос о точной оценке числа допустимых алгоритмов (со второй точки зрения) не так прост. Легко оценить лишь число алгоритмов, образующих множество A независимо от свойств таблицы T. Число различных таких алгоритмов равно 2^N в соответствии с числом различных подмножеств $A \subset \{1, \ldots, N\}$.

Замечание 2. Допустимые алгоритмы выбора из множества всевозможных натуральных чисел рассматривались А. Чёрчем [4]. Заменим в нашем определении конечную таблицу T на бесконечную последовательность нулей и единиц

$$t_1, t_2, \ldots$$

Мы предполагаем, что значения переменных ξ_k и функций F_k являются натуральными числами. Однако мы отказываемся от требования окончания выбора при $s = N$, предполагая вместо этого, что (бесконечная) последовательность функций F_k, G_k, H_k вычислима в традиционном для формальных определений подобного рода смысле. Исходным пунктом исследований Чёрча является существование (для любого p) последовательностей $t_1, t_2, \ldots, t_N, \ldots$, плотность которых равна p по любому бесконечному [5] множеству A, получаемому с помощью допустимого алгоритма.

3
ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1

Этот результат относится к конечным объектам; его формулировка не включает никаких понятий теории вероятностей. Использование некоторых результатов теории вероятностей в доказательстве не нарушит его формального характера, если мы ограничимся рассмотрением распределения «весов» в множестве таблиц T размера N, приписывая вес

$$P\,(T) = p^M\,(1 - p)^{N-M}$$

таблице, содержащей M единиц. Такой метод доказательства не затрагивает логической природы самой теоремы и не мешает ссылаться на нее при обсуждении области применимости теории вероятностей.

---

[5] В этом понятии Чёрча существенный интерес представляют лишь бесконечно продолжающиеся алгоритмы выбора, поэтому функции G_k и все связанное с ними должны быть опущены.

Может быть доказано следующее неравенство, относящееся к «схеме Бернулли»:

$$\mathsf{P}\left\{\sup_{k\geqslant n}\left|\frac{\mu_k}{k}-p\right|>\varepsilon\right\}<2e^{-2n\varepsilon^2(1-\varepsilon)}. \qquad (3.1)$$

Здесь p — вероятность успеха в каждом из последовательности независимых испытаний; μ_k — число успехов в первых k испытаниях. Мы легко можем вывести следствие из (3.1).

Следствие. *Пусть*[6]

$$\mathsf{P}\{\xi_k = 1 \mid k\leqslant\nu, \xi_1, \ldots, \xi_{k-1}\} = p,$$

где $\xi_1, \xi_2, \ldots, \xi_\nu$ — *последовательность случайного числа случайных величин и* p — *постоянная. Тогда*

$$\mathsf{P}\{\nu\geqslant n, \mid \mu_\nu/\nu - p\mid > \varepsilon\} < 2e^{-2n\varepsilon^2(1-\varepsilon)}. \qquad (3.2)$$

Рассмотрим теперь систему $\mathscr{R}_N$ из ρ допустимых алгоритмов.

Рассмотрим таблицу, образованную случайно, причем $t_x = 1$ с вероятностью p независимо от значений других t_m. Если мы зафиксируем $\mathrm{R}\in\mathscr{R}_N$ и обозначим через

$$\xi_1, \xi_2, \ldots, \xi_\nu$$

те элементы последовательности

$$t_1, t_2, \ldots, t_N,$$

номера которых попадают в $A = \mathrm{R}(T)$ (перечисленные в порядке их появления в ходе применения алгоритма), то, как легко видеть, условия справедливости (3.2) будут выполнены. Поэтому для любого данного $\mathrm{R}\in\mathscr{R}_N$ вероятность того, что число элементов ν множества A будет не меньше n и что неравенство $|\pi(A) - p|\geqslant\varepsilon$ также будет выполнено, меньше чем $2e^{-2n\varepsilon^2(1-\varepsilon)}$. Если

$$\rho\leqslant {}^1/_2 e^{2n\varepsilon^2(1-\varepsilon)},$$

то сумма вероятностей, с которыми нарушается неравенство

$$|\pi(A) - p|\leqslant\varepsilon$$

по всем алгоритмам, приводящим к множествам из n или более элементов, будет меньше 1. Поэтому с положительной вероятностью таблица T окажется (n, ε, p)-случайной в смысле определения из раздела 2. Отсюда следует утверждение о существовании (n, ε, p)-случайных по отношению к $\mathscr{R}_N$ таблиц (не апеллирующее к вероятностным допущениям о распределении $P(T)$ в пространстве таблиц).

---

[6] Мы рассматриваем условную вероятность того, что $\xi_k = 1$ при $k\leqslant\nu$ и данных $\xi_1, \ldots, \xi_{k-1}$.

4
О ВОЗМОЖНЫХ УЛУЧШЕНИЯХ ОЦЕНКИ, ДАВАЕМОЙ ТЕОРЕМОЙ 1

Пусть n, ε, N, p фиксированы. Тогда для каждого неотрицательного ρ имеет место одно из двух:

(а) для всякой системы $\mathcal{R}_N$ из ρ допустимых алгоритмов выбора найдется таблица T размера N, которая (n, ε, p)-случайна относительно $\mathcal{R}_N$;

(б) существует система $\mathcal{R}_N$ из ρ допустимых алгоритмов выбора, относительно которой никакая таблица T размера N не (n, ε, p)-случайна.

Легко видеть, что если ситуация (а) имеет место для некоторого ρ, то та же ситуация имеет место и для всех $\rho' < \rho$. Ясно также, что для $\rho = 0$ имеет место ситуация (а). Следовательно, существует верхняя грань

$$\tau(n, \varepsilon, N, p) = \sup_{\rho \in a} \rho$$

тех ρ, для которых выполняется (а). Для всех ρ, больших $\tau(n, \varepsilon, N, p)$, имеет место (б).

Если мы положим

$$\tau(n, \varepsilon) = \inf_{p,\, N} \tau(n, \varepsilon, N, p),$$

то утверждение теоремы 1 раздела 2 запишется в виде неравенства

$$\tau(n, \varepsilon) \geqslant {}^1/_2 e^{2n\varepsilon^2(1-\varepsilon)}. \tag{4.1}$$

Переходя к логарифмам

$$l(n, \varepsilon, N, p) = \log_2 \tau(n, \varepsilon, N, p), \qquad l(n, \varepsilon) = \log_2 \tau(n, \varepsilon),$$

мы можем записать (2.6) в форме

$$l(n, \varepsilon) \geqslant 2n\varepsilon^2(1 - \varepsilon) - 1. \tag{4.2}$$

В действительности основной интерес представляет асимптотически точная оценка $l(n, \varepsilon)$ при малых ε и больших n и $l(n, \varepsilon)$. При $\varepsilon \to 0$, $n\varepsilon^2 \to \infty$ из (4.2) вытекает

$$l(n, \varepsilon) \geqslant 2n\varepsilon^2 + o(n\varepsilon^2). \tag{4.3}$$

С другой стороны, как мы увидим, при $\varepsilon \to 0$, $n\varepsilon \to \infty$ имеет место соотношение

$$l(n, \varepsilon) \leqslant 4n\varepsilon + o(n\varepsilon). \tag{4.4}$$

К сожалению, мне не удалось устранить разрыв между степенями ε в (4.3) и (4.4).

Оценка (4.4) является простым следствием следующей теоремы, формулировка которой, к сожалению, довольно сложна и станет ясной из даваемого нами доказательства.

Теорема 2. *Если* $k < (1-2\varepsilon)/4\varepsilon$, $n \leqslant (k-1)m$, $N \geqslant km$, *то*

$$\tau(n, \varepsilon, N, {}^1/_2) \leqslant k \cdot 2^m.$$

Для доказательства теоремы достаточно при выполнении условий

$$k < (1-2\varepsilon)/4\varepsilon, \qquad n = (k-1)m, \qquad N = km$$

построить систему $\mathscr{R}_N$ из $\rho = k \cdot 2^m + 1$ допустимых алгоритмов, для которой не существует $(n, \varepsilon, {}^1/_2)$-случайной таблицы T.

Мы разбиваем множество $\{1, \ldots, N\}$ на k множеств Δ_i, $i = 1, \ldots, k$, по m элементов в каждом. Каждое Δ_i имеет 2^m подмножеств. Образуем множество A_{is}, $i = 1, 2, \ldots, k$; $s = 1, 2, \ldots, 2^m$, взяв объединение всех Δ_j при $j \neq i$ и s-го подмножества Δ_i. В систему $\mathscr{R}_N$ войдут:

$k \cdot 2^m$ алгоритмов, выбирающих множества A_{is},

один алгоритм, выбирающий $A = \{1, \ldots, N\}$.

Мы докажем, что никакая таблица не является $(n, \varepsilon, {}^1/_2)$-случайной по отношению к $\mathscr{R}_N$.

Предположим, что таблица T является $(n, \varepsilon, {}^1/_2)$-случайной по отношению к $\mathscr{R}_N$. Она должна содержать по крайней мере $({}^1/_2 - \varepsilon)N$ нулей и $({}^1/_2 - \varepsilon)N$ единиц. Поэтому можно так выбрат i и j, чтобы Δ_i содержало $\alpha \geqslant ({}^1/_2 - \varepsilon)m$ нулей, а Δ_j содержало $\beta \geqslant ({}^1/_2 - \varepsilon)m$ единиц.

Обозначим через γ наименьшее из чисел α и β; тогда $\gamma \geqslant ({}^1/_2 - \varepsilon)m$. Существует алгоритм $R' \in \mathscr{R}_N$ ($R'' \in \mathscr{R}_N$), выбирающий множество A' (A''), содержащее все элементы $\{1, \ldots, N\}$, за исключением γ элементов в Δ_i (Δ_j), соответствующих нулям (единицам) таблицы T. Легко видеть, что соответствующие частоты равны

$$\pi(A') = \frac{M}{N-\gamma}, \qquad \pi(A'') = \frac{M-\gamma}{N-\gamma},$$

где M — общее число единиц в таблице T. Оценим разницу между этими частотами

$$\pi(A') - \pi(A'') = \frac{\gamma}{N-\gamma} \geqslant \frac{({}^1/_2 - \varepsilon)m}{km} > 2\varepsilon.$$

Эта оценка противоречит системе неравенств

$$|\pi(A') - {}^1/_2| \leqslant \varepsilon, \qquad |\pi(A'') - {}^1/_2| \leqslant \varepsilon,$$

вытекающей из предположения $(n, \varepsilon, {}^1/_2)$-случайности таблицы T. Это противоречие доказывает теорему.

ЛИТЕРАТУРА

1. *Kolmogorov A. N.* Foundations of the theory of probability. Chelsea, 1950.
2. *Колмогоров А. Н.* Вероятность.— В кн.: БСЭ. 2-е изд. 1951, т. 7, с. 508—510.
3. *Колмогоров А. Н.* Теория вероятностей.— В кн.: Математика, ее содержание, методы и значение. М.: Изд-во АН СССР, 1956, т. 2, с. 252—284.
4. *Church A.* On the concept of a random sequence.— Bull. Amer. Math. Soc., 1940, vol. 46, N 2, p. 130—135.

10

ТРИ ПОДХОДА К ОПРЕДЕЛЕНИЮ ПОНЯТИЯ «КОЛИЧЕСТВО ИНФОРМАЦИИ» *

§ 1

КОМБИНАТОРНЫЙ ПОДХОД

Пусть переменное x способно принимать значения, принадлежащие конечному множеству X, которое состоит из N элементов. Говорят, что «энтропия» переменного x равна

$$H(x) = \log_2 N.$$

Указывая определенное значение

$$x = a$$

переменного x, мы «снимаем» эту энтропию, сообщая «информацию»

$$I = \log_2 N.$$

Если переменные $x_1, x_2, \ldots, x_k$ способны независимо пробегать множества, которые состоят соответственно из $N_1, N_2, \ldots, N_k$ элементов, то

$$H(x_1, x_2, \ldots, x_k) = H(x_1) + H(x_2) + \ldots + H(x_k). \tag{1}$$

Для передачи количества информации I приходится употреблять

$$I' = \begin{cases} I & \text{при } I \text{ целом,} \\ [I] + 1 & \text{при } I \text{ дробном} \end{cases}$$

двоичных знаков. Например, число различных «слов», состоящих из k нулей и единиц и одной двойки, равно

$$2^k (k + 1).$$

---

* Проблемы передачи информации, 1965, т. 1, № 1, с. 3—11.

Поэтому количество информации в такого рода сообщении равно

$$I = k + \log_2 (k + 1),$$

т. е. для «кодирования» такого рода слов в чистой двоичной системе требуется [1]

$$I' \approx k + \log_2 k$$

нулей и единиц.

При изложении теории информации обычно не задерживаются надолго на таком комбинаторном подходе к делу. Но мне кажется существенным подчеркнуть его логическую независимость от каких бы то ни было вероятностных допущений. Пусть, например, нас занимает задача кодирования сообщений, записанных в алфавите, состоящем из s букв, причем известно, что частоты

$$p_r = s_r/s \tag{2}$$

появления отдельных букв в сообщении длины n удовлетворяют неравенству

$$\chi = -\sum_{r=1}^{s} p_r \log_2 p_r \leqslant h. \tag{3}$$

Легко подсчитать, что при больших n двоичный логарифм числа сообщений, подчиненных требованию (3), имеет асимптотическую оценку

$$H = \log_2 N \sim nh.$$

Поэтому при передаче такого рода сообщений достаточно употребить примерно nh двоичных знаков.

Универсальный метод кодирования, который позволит передавать любое достаточно длинное сообщение в алфавите из s букв, употребляя не многим более чем nh двоичных знаков, не обязан быть чрезмерно сложным, в частности не обязан начинаться с определения частот p_r для всего сообщения. Чтобы понять это, достаточно заметить: разбивая сообщение S на m отрезков $S_1, S_2, \ldots, S_m$, получим неравенство

$$\chi \geqslant n^{-1} [n_1\chi_1 + n_2\chi_2 + \ldots + n_m\chi_m]. \tag{4}$$

Впрочем, я не хочу входить здесь в детали этой специальной задачи. Мне важно лишь показать, что математическая проблематика, возникающая на почве чисто комбинаторного подхода к измерению количества информации, не ограничивается тривиальностями.

Вполне естественным является чисто комбинаторный подход к понятию «энтропии речи», если иметь в виду оценку «гибкости»

---

[1] Всюду далее $f \approx g$ обозначает, что разность $f - g$ ограничена, а $f \sim g$, что отношение $f : g$ стремится к единице.

речи — показателя разветвленности возможностей продолжения речи при данном словаре и данных правилах построения фраз. Для двоичного логарифма числа N русских печатных текстов, составленных из слов, включенных в «Словарь русского языка» С. И. Ожегова и подчиненных лишь требованию «грамматической правильности» длины n, выраженной в «числе знаков» (включая «пробелы»), М. Ратнер и Н. Светлова получили оценку

$$h = (\log_2 N)/n = 1{,}9 \pm 0{,}1.$$

Это значительно больше, чем оценки сверху для «энтропии литературных текстов», получаемые при помощи различных методов «угадывания продолжений». Такое расхождение вполне естественно, так как литературные тексты подчинены не только требованию «грамматической правильности».

Труднее оценить комбинаторную энтропию текстов, подчиненных определенным содержательным ограничениям. Представляло бы, например, интерес оценить энтропию русских текстов, могущих рассматриваться как достаточно точные по содержанию переводы заданного иноязычного текста. Только наличие такой «остаточной энтропии» делает возможным стихотворные переводы, где «затраты энтропии» на следование избранному метру и характеру рифмовки могут быть довольно точно подсчитаны. Можно показать, что классический четырехстопный рифмованный ямб с некоторыми естественными ограничениями на частоту «переносов» и т. п. требует допущения свободы обращения со словесным материалом, характеризуемой «остаточной энтропией» порядка 0,4 (при указанном выше условном способе измерения длины текста по «числу знаков, включая пробелы»). Если учесть, с другой стороны, что стилистические ограничения жанра, вероятно, снижают приведенную выше оценку «полной» энтропии с 1,9 до не более чем 1,1—1,2, то ситуация становится примечательной как в случае перевода, так и в случае оригинального поэтического творчества.

Да простят мне утилитарно настроенные читатели этот пример. В оправдание замечу, что более широкая проблема оценки количеств информации, с которыми имеет дело творческая человеческая деятельность, имеет очень большое значение.

Посмотрим теперь, в какой мере чисто комбинаторный подход позволяет оценить «количество информации», содержащееся в переменном x относительно связанного с ним переменного y. Связь между переменными x и y, пробегающими соответственно множества X и Y, заключается в том, что не все пары x, y, принадлежащие прямому произведению $X \times Y$, являются «возможными». По множеству возможных пар U определяются при любом $a \in X$ множества Y_a тех y, для которых

$$(a, y) \in U.$$

x	y			
	1	2	3	4
1	+	+	+	+
2	+	−	+	−
3	−	+	−	−

Естественно определить условную энтропию равенством

$$H\ (y \mid a) = \log_2 N\ (Y_a) \qquad (5)$$

(где $N\ (Y_x)$ — число элементов в множестве Y_x), а информацию в x относительно y формулой

$$I\ (x : y) = H\ (y) - H\ (y \mid x). \qquad (6)$$

Например, в случае, изображенном в таблице, имеем

$$I\ (x = 1 : y) = 0, \qquad I\ (x = 2 : y) = 1, \qquad I\ (x = 3 : y) = 2.$$

Понятно, что $H\ (y \mid x)$ и $I\ (x : y)$ являются функциями от x (в то время как y входит в их обозначение в виде «связанного переменного»).

Без труда вводится в чисто комбинаторной концепции представление о «количестве информации, необходимом для указания объекта x при заданных требованиях к точности указания». (См. по этому поводу обширную литературу об «ε-энтропии» множеств в метрических пространствах.)

Очевидно,

$$H\ (x \mid x) = 0, \qquad I\ (x : x) = H\ (x). \qquad (7)$$

§ 2
ВЕРОЯТНОСТНЫЙ ПОДХОД

Возможности дальнейшего развития теории информации на основе определений (5) и (6) остались в тени ввиду того, что придание переменным x и y характера «случайных переменных», обладающих определенным совместным распределением вероятностей, позволяет получить значительно более богатую систему понятий и соотношений. В параллель к введенным в § 1 величинам имеем здесь

$$H_W\,(x) = -\sum_x p\,(x) \log_2 p\,(x), \qquad (8)$$

$$H_W\,(y \mid x) = -\sum_y p\,(y \mid x) \log_2 p\,(y \mid x), \qquad (9)$$

$$I_W\ (x : y) = H_W\ (y) - H_W\ (y \mid x). \qquad (10)$$

По-прежнему $H_W\ (y \mid x)$ и $I_W\ (x : y)$ являются функциями от x. Имеют место неравенства

$$H_W\ (x) \leqslant H\ (x), \qquad H_W\ (y \mid x) \leqslant H\ (y \mid x), \qquad (11)$$

переходящие в равенства при равномерности соответствующих распределений (на X и на Y_x). Величины $I_W\ (x : y)$ и $I\ (x : y)$

не связаны неравенством определенного знака. Как и в § 1,

$$H_W(x \mid x) = 0, \qquad I_W(x : x) = H_W(x). \tag{12}$$

Но отличие заключается в том, что можно образовать математические ожидания

$$\mathsf{M} H_W(y \mid x), \qquad \mathsf{M} I_W(x : y),$$

а величина

$$I_W(x, y) = \mathsf{M} I_W(x : y) = \mathsf{M} I_W(y : x) \tag{13}$$

характеризует «тесноту связи» между x и y симметричным образом.

Стоит, однако, отметить и возникновение в вероятностной концепции одного парадокса: величина $I(x : y)$ при комбинаторном подходе всегда неотрицательна, как это и естественно при наивном представлении о «количестве информации», величина же $I_W(x : y)$ может быть и отрицательной. Подлинной мерой «количества информации» теперь становится лишь осредненная величина $I_W(x, y)$.

Вероятностный подход естествен в теории передачи по каналам связи «массовой» информации, состоящей из большого числа не связанных или слабо связанных между собой сообщений, подчиненных определенным вероятностным закономерностям. В такого рода вопросах практически безвредно и укоренившееся в прикладных работах смешение вероятностей и частот в пределах одного достаточно длинного временнóго ряда (получающее строгое оправдание при гипотезе достаточно быстрого «перемешивания»). Практически можно считать, например, вопрос об «энтропии» потока поздравительных телеграмм и «пропускной способности» канала связи, требующегося для их своевременной и неискаженной передачи, корректно поставленным в его вероятностной трактовке и при обычной замене вероятностей эмпирическими частотами. Если здесь и остается некоторая неудовлетворенность, то она связана с известной расплывчатостью наших концепций, относящихся к связям между математической теорией вероятностей и реальными «случайными явлениями» вообще.

Но какой реальный смысл имеет, например, говорить о «количестве информации», содержащемся в тексте «Войны и мира»? Можно ли включить разумным образом этот роман в совокупность «возможных романов», да еще постулировать наличие в этой совокупности некоторого распределения вероятностей? Или следует считать отдельные сцены «Войны и мира» образующими случайную последовательность с достаточно быстро затухающими на расстоянии нескольких страниц «стохастическими связями»?

По существу не менее темным является и модное выражение «количество наследственной информации», необходимой, скажем, для воспроизведения особи вида кукушка. Опять в пределах принятой вероятностной концепции возможны два варианта. В первом варианте рассматривается совокупность «возможных видов» с неизвестно откуда берущимся распределением вероятностей на этой совокупности [2]. Во втором варианте характеристические свойства вида считаются набором слабо связанных между собой случайных переменных. В пользу второго варианта можно привести соображения, основанные на реальном механизме мутационной изменчивости. Но соображения эти иллюзорны, если считать, что в результате естественного отбора возникает система согласованных между собой характеристических признаков вида.

§ 3
АЛГОРИТМИЧЕСКИЙ ПОДХОД

По существу наиболее содержательным является представление о количестве информации «в чем-либо» (x) и «о чем-либо» (y). Не случайно именно оно в вероятностной концепции получило обобщение на случай непрерывных переменных, для которых энтропия бесконечна, но в широком круге случаев

$$I_W(x, y) = \iint P_{xy}(dx\,dy) \log_2 \frac{P_{xy}(dx\,dy)}{P_x(dx)\,P_y(dy)}$$

конечно. Реальные объекты, подлежащие нашему изучению, очень (неограниченно?) сложны, но связи между двумя реально существующими объектами исчерпываются при более простом схематизированном их описании. Если географическая карта дает нам значительную информацию об участке земной поверхности, то все же микроструктура бумаги и краски, нанесенной на бумагу, никакого отношения не имеет к микроструктуре изображенного участка земной поверхности.

Практически нас интересует чаще всего количество информации в индивидуальном объекте x относительно индивидуального объекта y. Правда, уже заранее ясно, что такая индивидуальная оценка количества информации может иметь разумное содержание лишь в случаях достаточно больших количеств информации. Не имеет, например, смысла спрашивать о количестве информации в последовательности цифр

0 1 1 0

относительно последовательности

1 1 0 0.

---

[2] Обращение к множеству видов, существующих или существовавших на Земле, даже при чисто комбинаторном подсчете дало бы совершенно неприемлемо малые оценки сверху (что-либо вроде <100 бит!).

Но если мы возьмем вполне конкретную таблицу случайных чисел обычного в статистической практике объема и выпишем для каждой ее цифры цифру единиц ее квадрата по схеме

$$0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9$$

$$0\ 1\ 4\ 9\ 6\ 5\ 6\ 9\ 4\ 1$$

то новая таблица будет содержать примерно

$$\left(\log_2 10 - \frac{8}{10}\right) n$$

информации о первоначальной (n — число цифр в таблицах).

В соответствии с только что сказанным предлагаемое далее определение величины

$$I_A(x:y)$$

будет сохранять некоторую неопределенность. Разные равноценные варианты этого определения будут приводить к значениям, эквивалентным лишь в смысле $I_{A_1} \approx I_{A_2}$, т. е.

$$|I_{A_1} - I_{A_2}| \leqslant C_{A_1 A_2},$$

где константа $C_{A_1 A_2}$ зависит от положенных в основу двух вариантов определения универсальных методов программирования A_1 и A_2.

Будем рассматривать «нумерованную область объектов», т. е. счетное множество

$$X = \{x\},$$

каждому элементу которого поставлена в соответствие в качестве «номера» $n(x)$ конечная последовательность нулей и единиц, начинающаяся с единицы. Обозначим через $l(x)$ длину последовательности $n(x)$. Будем предполагать, что:

1) соответствие между X и множеством D двоичных последовательностей описанного вида взаимно однозначно;

2) $D \subset X$, функция $n(x)$ на D общекурсивна [1], причем для $x \in D$

$$l(n(x)) \leqslant l(x) + C,$$

где C — некоторая константа;

3) вместе с x и y в X входит упорядоченная пара (x, y), номер этой пары есть общерекурсивная функция номеров x и y и

$$l(x, y) \leqslant C_x + l(y),$$

где C_x зависит только от x.

Не все эти требования существенны, но они облегчают изложение. Конечный результат построения инвариантен по отношению к переходу к новой нумерации $n'(x)$, обладающей теми же свойствами и выражающейся общерекурсивно через старую, и по отношению к включению системы X в более обширную систему X' (в предположении, что номера n' в расширенной системе для элементов первоначальной системы общерекурсивно выражаются через первоначальные номера n). При всех этих преобразованиях новые «сложности» и количества информации остаются эквивалентными первоначальным в смысле $\approx$.

«Относительной сложностью» объекта y при заданном x будем считать минимальную длину $l(p)$ «программы» p получения y из x. Сформулированное так определение зависит от «метода программирования». Метод программирования есть не что иное, как функция

$$\varphi(p, x) = y,$$

ставящая в соответствие программе p и объекту x объект y.

В соответствии с универсально признанными в современной математической логике взглядами следует считать функцию φ частично рекурсивной. Для любой такой функции полагаем

$$K_\varphi(y \mid x) = \begin{cases} \min\limits_{\varphi(p, x)=y} l(p), \\ \infty, \text{ если нет такого } p, \text{ что } \varphi(p, x) = y. \end{cases}$$

При этом функция

$$v = \varphi(u)$$

от $u \in X$ со значениями $v \in X$ называется частично рекурсивной, если она порождается частично рекурсивной функцией преобразования номеров

$$n(v) = \Psi[n(u)].$$

Для понимания определения важно заметить, что частично рекурсивные функции, вообще говоря, не являются всюду определенными. Не существует регулярного процесса для выяснения того, приведет применение программы p к объекту x к какому-либо результату или нет. Поэтому функция $K_\varphi(y \mid x)$ не обязана быть эффективно вычислимой (общерекурсивной) даже в случае, когда она заведомо конечна при любых x и y.

Основная теорема. *Существует такая частично рекурсивная функция $A(p, x)$, что для любой другой частично рекурсивной функции $\varphi(p, x)$ выполнено неравенство*

$$K_A(y \mid x) \leqslant K_\varphi(y \mid x) + C_\varphi,$$

где константа C_φ не зависит от x и y.

Доказательство опирается на существование универсальной частично рекурсивной функции

$$\Phi(n, u),$$

обладающей тем свойством, что, фиксируя надлежащий номер n, можно получить по формуле

$$\varphi(u) = \Phi(n, u)$$

любую другую частично рекурсивную функцию. Нужная нам функция $A(p, x)$ определяется формулой [3]

$$A((n, q), x) = \Phi(n, (q, x)).$$

В самом деле, если

$$y = \varphi(p, x) = \Phi(n, (p, x)),$$

то

$$A((n, p), x) = y, \quad l(n, p) \leqslant l(p) + C_n.$$

Функции $A(p, x)$, удовлетворяющие требованиям основной теоремы, назовем (как и определяемые ими методы программирования) *асимптотически оптимальными*. Очевидно, что для них при любых x и y «сложность» $K_A(y \mid x)$ конечна. Для двух таких функций A_1 и A_2

$$|K_{A_1}(y \mid x) - K_{A_2}(y \mid x)| \leqslant C_{A_1A_2},$$

где $C_{A_1A_2}$ не зависит от x и y, т. е. $K_{A_1}(y \mid x) \approx K_{A_2}(y \mid x)$.

Наконец,

$$K_A(y) = K_A(y \mid 1)$$

можно считать просто «сложностью объекта y» и определить «количество информации в x относительно y» формулой

$$I_A(x : y) = K_A(y) - K_A(y \mid x).$$

Легко доказать [4], что величина эта всегда в существенном положительна:

$$I_A(x : y) \gtrsim 0,$$

что понимается в том смысле, что $I_A(x : y)$ не меньше некоторой отрицательной константы C, зависящей лишь от условностей избранного метода программирования. Как уже говорилось, вся теория рассчитана на применение к большим количествам информации, по сравнению с которыми $|C|$ будет пренебрежимо мал.

Наконец, $K_A(x \mid x) \approx 0$, $I_A(x : x) \approx K_A(x)$.

Конечно, можно избегнуть неопределенностей, связанных с константами C_φ и т. д., остановившись на определенных областях объектов X, их нумерации и функции A, но сомнительно, чтобы

---

[3] $\Phi(n, u)$ определена только в случае $n \in D$, $A(p, x)$ только в случае, когда p имеет вид (n, q), $n \in D$.

[4] Выбирая в виде «функции сравнения» $\varphi(p, x) = A(p, 1)$, получим $K_A(y \mid x) \leqslant K_\varphi(y \mid x) + C_\varphi = K_A(y) + C_\varphi$.

это можно было сделать без явного произвола. Следует, однако, думать, что различные представляющиеся здесь «разумные» варианты будут приводить к оценкам «сложностей», расходящимся на сотни, а не на десятки тысяч бит. Поэтому такие величины, как «сложность» текста романа «Война и мир», можно считать определенными с практической однозначностью. Эксперименты по угадыванию продолжений литературных текстов позволяют оценить сверху условную сложность при заданном запасе «априорной информации» (о языке, стиле, содержании текста), которой располагает угадывающий. В опытах, проводившихся на кафедре теории вероятностей Московского государственного университета, такие оценки сверху колебались между 0,9 и 1,4. Оценки порядка 0,9—1,1, получившиеся у Н. Г. Рычковой, вызывали у менее удачливых угадчиков разговоры о ее телепатической связи с авторами текстов.

Я думаю, что и для «количества наследственной информации» предлагаемый подход дает в принципе правильное определение самого понятия, как бы ни была трудна фактическая оценка этого количества.

§ 4
ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Изложенная в § 3 концепция обладает одним существенным недостатком: она не учитывает «трудности» переработки программы p и объекта x в объект y. Введя надлежащие определения, можно доказать точно формулируемые математические предложения, которые законно интерпретировать как указание на существование таких случаев, когда объект, допускающий очень простую программу, т. е. обладающий очень малой сложностью $K(x)$, может быть восстановлен по коротким программам лишь в результате вычислений совершенно нереальной длительности. В другом месте я предполагаю изучить зависимость необходимой сложности программы

$$K^t(x)$$

от допустимой трудности t ее переработки в объект x. Сложность $K(x)$, которая была определена в § 3, появится при этом в качестве минимума $K^t(x)$ при снятии ограничений на величину t.

За пределами этой заметки остается и применение построений § 3 к новому обоснованию теории вероятностей. Грубо говоря, здесь дело идет о следующем. Если конечное множество M из очень большого числа элементов N допускает определение при помощи программы длины пренебрежимо малой по сравнению с $\log_2 N$, то почти все элементы M имеют сложность $K(x)$, близкую к $\log_2 N$. Элементы $x \in M$ этой сложности и рассматриваются как «случайные» элементы множества M. Не вполне завершенное изложение этой идеи можно найти в статье [2].

ЛИТЕРАТУРА

1. *Успенский В. А.* Лекции о вычислимых функциях. М.: Физматгиз, 1960.
2. *Kolmogorov A. N.* On tables of random numbers.— Sankhya. A, 1963, vol. 25, N 4, p. 369—376.

11

О РЕАЛИЗАЦИИ СЕТЕЙ В ТРЕХМЕРНОМ ПРОСТРАНСТВЕ *[1]

Совместно с Я. М. Барздинем

Под (d, n)-сетью будем понимать направленный граф с n нумерованными вершинами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и dn отмеченными дугами, такой, что в каждую вершину входит ровно d дуг, одна из которых помечена буквой x_1, другая — буквой x_2 и т. д. и, наконец, последняя — буквой x_d.

Примерами таких сетей являются логические сети и нейронные сети. Именно по этой причине представляет интерес вопрос о реализации таких сетей в обычном трехмерном пространстве при условии, что вершины являются шарами, а дуги — трубками с некоторым положительным диаметром.

В дальнейшем без существенного ограничения общности (ввиду того что наши оценки будут делаться только с точностью до порядка) будем ограничиваться рассмотрением только $(2, n)$-сетей, которые будем называть просто сетями.

Будем говорить, что сеть $\mathfrak{A}$ реализована в трехмерном пространстве [2], если:

1) каждой вершине $\alpha \in \mathfrak{A}$ сопоставлена некоторая точка $\varphi\alpha$ данного пространства (для разных α и β $\varphi\alpha$ и $\varphi\beta$ разные); эту точку будем называть φ-точкой;

2) каждой дуге $p = (\alpha, \beta)$, исходящей из α и входящей в β, сопоставлена некоторая непрерывная кривая K_p, соединяющая точки $\varphi\alpha$ и $\varphi\beta$; эту кривую будем называть проводником;

3) расстояние между любыми двумя различными φ-точками не меньше 1;

---

* Проблемы кибернетики, 1967, вып. 19, с. 261—268.

[1] В начале 60-х годов А. Н. Колмогоров доказал теорему 1* для сетей с ограниченным ветвлением и такое приближение к теореме 2: существует сеть с $n > 1$ элементами, любая реализация которой имеет диаметр больше $C\sqrt{n}/\log n$, где C — некоторая константа, не зависящая от n. Окончательные варианты приведенных теорем принадлежат Я. М. Барздиню.— *Примеч. авт.*

[2] Имеется в виду обычное евклидово трехмерное пространство.

4) расстояние между любыми проводниками, соответствующими неинцидентным дугам [3], не меньше 1;

5) расстояние между любыми двумя проводниками, соответствующими дугам вида (α, β) и (γ, α) или (β, α) и (γ, α), $\beta \neq \gamma$, соответственно, вне окрестности радиуса 1 точки $\varphi\alpha$ не меньше 1.

Пусть R — некоторая реализация сети $\mathfrak{A}$ в трехмерном пространстве. Будем говорить, что тело W содержит реализацию R, если все φ-точки и проводники, возникающие при этой реализации, лежат в теле W и находятся от его поверхности на расстоянии, не меньшем 1. Под объемом реализации R будем понимать минимальный объем тел, содержащих данную реализацию R. Под объемом $V(\mathfrak{A})$ сети $\mathfrak{A}$ будем понимать минимальный объем ее реализаций в трехмерном пространстве.

Будем говорить, что почти все сети обладают некоторым свойством E, если

$$E(n)/S(n) \to 1 \quad \text{при } n \to \infty,$$

где $S(n)$ — число всех попарно различных [4] сетей с n вершинами и $E(n)$ — число тех попарно различных сетей с n вершинами, которые обладают свойством E.

Теорема 1 (верхняя оценка). *Для всех сетей $\mathfrak{A}$ с n вершинами*

$$V(\mathfrak{A}) \leqslant C_1 n \sqrt{n},$$

где C_1 — некоторая константа, не зависящая от n.

Теорема 2 (нижняя оценка). *Для почти всех сетей $\mathfrak{A}$ с n вершинами*

$$V(\mathfrak{A}) \geqslant C_2 n \sqrt{n},$$

где C_2 — некоторая константа > 0, не зависящая от n.

Теорема 1 вытекает из более сильного утверждения:

Теорема 1*. *Любая сеть с n вершинами может быть реализована в сфере радиуса $C\sqrt{n}$, где C — константа, не зависящая от n.*

Доказательство теоремы 1*. Сначала докажем теорему 1* для сетей, обладающих тем свойством, что из каждой вершины исходит не более двух дуг. Такие сети будем называть сетями с ограниченным ветвлением.

Итак, пусть $\mathfrak{A}$ — некоторая сеть с n вершинами и ограниченным ветвлением. Покажем, что ее можно реализовать в па-

[3] Две дуги (α, β) и (γ, δ) считаются инцидентными тогда и только тогда, когда $(\alpha = \gamma) \vee (\alpha = \delta) \vee (\beta = \gamma) \vee (\beta = \delta)$.

[4] Две сети $\mathfrak{A}$ и $\mathfrak{B}$ считаются одинаковыми тогда и только тогда, когда любые две вершины с одинаковыми номерами как в $\mathfrak{A}$, так и в $\mathfrak{B}$ связаны между собой дугами, одинаково ориентированными и с одинаковыми метками.

раллелепипеде $ABCDA'B'C'D'$, изображенном на рис. 1 (без ограничения общности будем считать, что $\sqrt{n}$ — целое число). Это будет означать, что $\mathfrak{A}$ можно реализовать в сфере радиуса $C\sqrt{n}$, где $C = \sqrt{21}$.

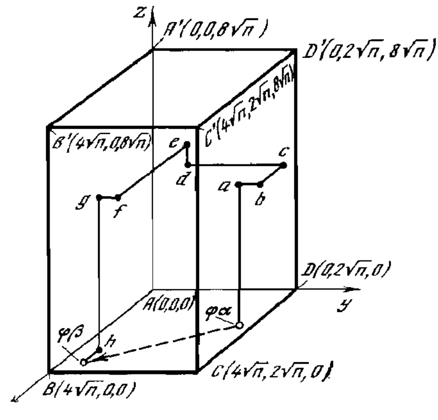


Рис. 1

Пусть U — множество точек основания $ABCD$, имеющих вид $(4i + 2,\ 2j,\ 0)$, $i = 0,\ 1, \ldots, \sqrt{n} - 1$; $j = 0,\ 1, \ldots, \sqrt{n} - 1$. Сопоставим каждой вершине α сети $\mathfrak{A}$ свою точку $\varphi\alpha$ из U, а каждой дуге $p = (\alpha, \beta)$ ломаную $K_p(\zeta) = (\varphi\alpha)abcdefgh(\varphi\beta)$ (рис. 1), зависящую от натурального параметра ζ следующим образом: пусть координаты точек $\varphi\alpha$ и $\varphi\beta$ соответственно $(4i + 2, 2j, 0)$ и $(4i' + 2, 2j', 0)$, тогда:

координаты точки h: $(4i' + 2 + \tau, 2j', 0)$, где $\tau = 1$, если дуга p помечена буквой x_1, и $\tau = -1$, если дуга p помечена буквой x_2,

координаты точек a и g соответственно $(4i + 2, 2j, 2\zeta - 1)$ и $(4i' + 2 + \tau, 2j', 2\zeta)$,

координаты точек b и f соответственно $(4i + 2, 2j + 1, 2\zeta - 1)$ и $(4i' + 2 + \tau, 2j' + 1, 2\zeta)$,

координаты точки c: $(4i, 2j + 1, 2\zeta - 1)$,

координаты точек d и e соответственно $(4i, 2j' + 1, 2\zeta - 1)$ и $(4i, 2j' + 1, 2\zeta)$.

Дуги $p = (\alpha, \beta)$ и $p' = (\alpha', \beta')$ будем называть родственными, если или точки $\varphi\alpha$ и $\varphi\alpha'$ имеют одинаковые абсциссы, или точки $\varphi\beta$ и $\varphi\beta'$ имеют одинаковые ординаты, или то и другое. Из ограниченности ветвления вытекает, что число дуг сети $\mathfrak{A}$, родственных дуге p, не превышает $4\sqrt{n}$.

Из определения ломаной $K_p(\zeta)$ вытекают следующие важные следствия.

1) Если дуги $p = (\alpha, \beta)$ и $p' = (\alpha', \beta')$ не являются родственными, то соответствующие им ломаные $K_p(\zeta)$ и $K_{p'}(\zeta')$ при любых значениях параметров ζ и ζ' находятся друг от друга на расстоянии $\geqslant 1$ (за исключением окрестности радиуса 1 точки $\varphi\alpha$ $(\varphi\beta)$ в случае, когда $\alpha = \beta'$ $(\beta = \alpha')$).

2) Если дуги $p = (\alpha, \beta)$ и $p' = (\alpha', \beta')$, $\alpha \neq \alpha'$, являются родственными, то соответствующие им ломаные $K_p(\zeta)$ и $K_{p'}(\zeta')$ обладают тем свойством, что если $\zeta \neq \zeta'$, то эти ломаные находятся друг от друга на расстоянии $\geqslant 1$ (за исключением окрест-

ности радиуса 1 точки $\varphi\alpha$ ($\varphi\beta$) в случае, когда $\alpha = \beta'$ ($\beta = \alpha'$ или $\beta = \beta'$)).

Поэтому для доказательства реализуемости сети $\mathfrak{A}$ в параллелепипеде $ABCDA'B'C'D'$ достаточно показать, что для дуг сети $\mathfrak{A}$ можно так подобрать натуральные значения параметра ζ, чтобы:

а) эти значения для разных родственных дуг были разными,

б) соответствующие ломаные находились в параллелепипеде $ABCDA'B'C'D'$, т. е. $1 \leqslant \zeta \leqslant 4\sqrt{n}$.

Такой выбор значений параметра ζ всегда возможен ввиду того, что для любой дуги p число дуг, родственных ей, не превышает $4\sqrt{n}$, т. е. не превышает числа допустимых различных значений параметра ζ.

Таким образом, мы показали, что для сетей с ограниченным ветвлением теорема 1* имеет место. Справедливость теоремы 1* для произвольных сетей вытекает из того, что любую сеть $\mathfrak{A}$ с n вершинами, добавляя не более $2n$ новых вершин, можно перестроить в сеть $\mathfrak{A}'$ с ограниченным ветвлением. Например, дуги вида рис. 2 можно заменить деревом вида рис. 3 (γ_i —новые вершины). В результате мы получили сеть $\mathfrak{A}'$ с ограниченным ветвлением и $n' \leqslant 3n$ вершинами. Эту сеть можно реализовать в сфере радиуса $C\sqrt{n'}$. Для того чтобы из реализации сети $\mathfrak{A}'$ получить реализацию первоначальной сети $\mathfrak{A}$, точки $\varphi\gamma_i$ надо интерпретировать как точки ветвления ломаных, соединяющих $\varphi\alpha$ с $\varphi\beta_1, \ldots, \varphi\beta_k$.

Пусть ω — некоторое разбиение вершин сети $\mathfrak{A}$ на три упорядоченные части, которые обозначим соответственно через ω_1, ω_2, ω_3. Разбиение ω сети $\mathfrak{A}$ с n вершинами будем называть (ε, δ)-разбиением, если ω_1 содержит $[\varepsilon n]$ вершин, ω_2 содержит $[\delta n]$ вершин и ω_3 содержит $n - [\varepsilon n] - [\delta n]$ вершин. Под степенью связности $S(\mathfrak{A}, \omega)$ сети $\mathfrak{A}$ относительно разбиения ω будем понимать максимальное число неинцидентных дуг, идущих из вершин ω_3 к вершинам ω_1.

Лемма 1. *Существуют такие* $0 < \varepsilon_0 < 1$, $a_0 > 0$ *и* $b_0 > 0$, *что для любых* $\varepsilon \leqslant \varepsilon_0$ *и* $a \leqslant a_0$ *степень связности почти всех сетей* $\mathfrak{A}$ *с* n *вершинами относительно любого* $(\varepsilon, a\varepsilon)$*-разбиения не меньше* $b_0\varepsilon n$.

Докажем эту лемму. Под степенью инцидентности $Z(\mathfrak{A}, \omega)$ сети $\mathfrak{A}$ относительно разбиения ω будем понимать число вершин, принадлежащих ω_3, из которых идут дуги к вершинам, принадлежащим ω_1. Так как в одну вершину может входить не более двух дуг, то $S(\mathfrak{A}, \omega) \geqslant {}^1/_2 Z(\mathfrak{A}, \omega)$. Таким образом, наша лемма будет доказана, если мы докажем аналогичное утверждение для степени инцидентности (это утверждение обозначим через А).

С этой целью процесс построения сетей с n вершинами $\alpha_1, \alpha_2, \ldots, \alpha_n$ представим следующим образом. Вначале дано n вершин $\alpha_1, \alpha_2, \ldots, \alpha_n$ таких, что в каждую из них входит ровно две дуги со свободными вторыми концами (рис. 4). Затем свободный конец каждой дуги случайно подсоединяем к какой-нибудь вершине[5] из $(\alpha_1, \ldots, \alpha_n)$. В результате мы получаем некоторую сеть с вершинами $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Пусть вначале было фиксировано некоторое (ε, δ)-разбиение ω множества $(\alpha_1, \ldots, \alpha_n)$. Обозначим через $P_\omega(\varepsilon, \delta, c, n)$ вероятность того, что в результате описанной выше процедуры мы получим сеть, которая относительно ω имеет степень инцидентности $Z(\mathfrak{A}, \omega) < cn$. Заметим, что при фиксированных ε и δ вероятность

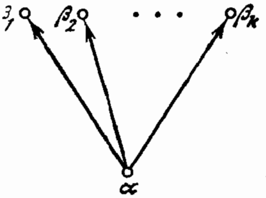


Рис. 2

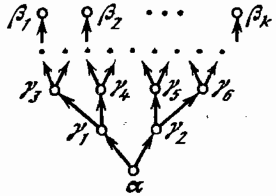


Рис. 3

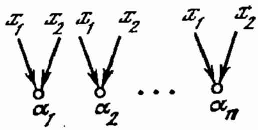


Рис. 4

$P_\omega(\varepsilon, \delta, c, n)$ для всех (ε, δ)-разбиений ω множества $(\alpha_1, \ldots, \alpha_n)$ одна и та же и что число всевозможных (ε, δ)-разбиений множества $(\alpha_1, \ldots, \alpha_n)$ равно[6] $C_n^{\varepsilon n} C_{n-\varepsilon n}^{\delta n}$. Обозначим через $P(\varepsilon, \delta, c, n)$ вероятность того, что в результате описанной выше процедуры мы получим сеть, которая хотя бы при одном (ε, δ)-разбиении имеет степень инцидентности $< cn$. Из сказанного выше следует, что

$$P(\varepsilon, \delta, c, n) < C_n^{\varepsilon n} C_{n-\varepsilon n}^{\delta n} P_\omega(\varepsilon, \delta, c, n). \tag{1}$$

Утверждение А, очевидно, будет доказано, если мы докажем следующее утверждение В:

Существуют такие $0 < \varepsilon_0' < 1$, $a_0' > 0$ и $b_0' > 0$, что для любых $\varepsilon \leqslant \varepsilon_0'$ и $a \leqslant a_0'$

$$P(\varepsilon, a\varepsilon, b_0'\varepsilon, n) \to 0 \quad \text{при} \quad n \to \infty.$$

Оценим $P_\omega(\varepsilon, \delta, c, n)$. С этой целью рассмотрим следующую вероятностную модель. Имеется n ящиков, из которых εn белых, δn черных и $(1 - \varepsilon - \delta)\, n$ красных. Рассмотрим серию испытаний $\Omega = \xi_1, \xi_2, \ldots, \xi_{2\varepsilon n}$, состоящую в последовательном случайном бросании $2\varepsilon n$ шариков в ящики (одно испытание — это бросание одного шарика). Под благоприятным исходом испытания будем понимать попадание шарика в пустой красный ящик. Обозначим

[5] Предполагается, что вероятности выбора у всех вершин одинаковы, т. е. равны $1/n$.

[6] Здесь и в дальнейшем для простоты будем считать, что εn, δn и cn — целые числа.

через $P'(\varepsilon, \delta, c, n)$ вероятность того, что число благоприятных исходов в серии испытаний Ω будет $< cn$. Очевидно,

$$P_\omega(\varepsilon, \delta, c, n) = P'(\varepsilon, \delta, c, n) \tag{2}$$

(в роли шариков выступают свободные концы дуг, входящих в вершины ω_1, в роли белых ящиков — вершины ω_1, в роли черных ящиков — вершины ω_2, в роли красных ящиков — вершины ω_3).

Оценим $P'(\varepsilon, \delta, c, n)$. Вероятность благоприятного исхода испытания ξ_i в серии Ω хотя и зависит от результатов предыдущих испытаний, но всегда $>1 - 3\varepsilon - \delta$ (самую малую вероятность благоприятного исхода может иметь испытание $\xi_{2\varepsilon n}$ при условии, что исходы всех предыдущих испытаний были благоприятными; в этом случае, как нетрудно видеть, эта вероятность будет $n^{-1}[(1 - \varepsilon - \delta)n - (2\varepsilon n - 1)] > 1 - 3\varepsilon - \delta$). Поэтому

$$P'(\varepsilon, \delta, c, n) < P''(\varepsilon, \delta, c, n), \tag{3}$$

где $P''(\varepsilon, \delta, c, n)$ — вероятность того, что в $2\varepsilon n$ испытаниях Бернулли, каждое из которых имеет вероятность благоприятного исхода $1 - 3\varepsilon - \delta$, число благоприятных исходов будет $\leqslant cn$.

Оценим $P''(\varepsilon, \delta, c, n)$. Как известно,

$$P''(\varepsilon, \delta, c, n) = \sum_{i=0}^{cn} C_{2\varepsilon n}^{i}(1 - 3\varepsilon - \delta)^i (3\varepsilon + \delta)^{2\varepsilon n - i}.$$

Дальнейшие оценки будем проводить при следующих предположениях:

(а) $0 < \varepsilon \leqslant 1/2$ и $0 < \delta < 1 - \varepsilon$;

(б) $0 < cn \leqslant$ математического ожидания $2\varepsilon n(1 - 3\varepsilon - \delta)$, т. е. $0 < c \leqslant 2\varepsilon(1 - 3\varepsilon - \delta)$.

В таком случае

$$P''(\varepsilon, \delta, c, n) < cnC_{2\varepsilon n}^{cn}(1 - 3\varepsilon - \delta)^{cn}(3\varepsilon + \delta)^{2\varepsilon n - cn} <$$
$$< nC_{2\varepsilon n}^{cn}(3\varepsilon + \delta)^{2\varepsilon n - cn}.$$

Отсюда и из (1)—(3) получаем, что при предположениях (а) и (б)

$$P(\varepsilon, \delta, c, n) < C_n^{\varepsilon n} \cdot C_{n-\varepsilon n}^{\delta n} \cdot nC_{2\varepsilon n}^{cn}(3\varepsilon + \delta)^{2\varepsilon n - cn}.$$

Введем теперь параметры a и b такие, что $\delta = a\varepsilon$ и $c = b\varepsilon$. Тогда вместо предположений (а) и (б) получим следующие неравенства:

(а′) $0 < \varepsilon \leqslant 1/2$ и $a > 0$, $1 - \varepsilon - a\varepsilon > 0$;

(б′) $0 < b \leqslant 2 - 6\varepsilon - 2a\varepsilon$.

Кроме того, из формулы Стирлинга вытекает, что при достаточно больших k и r

$$C_k^r = \frac{k!}{r!(k-r)!} <$$
$$< \exp\{k \ln k + 1/2 \ln k - r \ln r - 1/2 \ln r -$$
$$- (k - r)\ln(k - r) - 1/2 \ln(k - r)\}.$$

Поэтому при достаточно больших n:

$$P(\varepsilon, a\varepsilon, b\varepsilon, n) < C_n^{\varepsilon n}\cdot C_{n-\varepsilon n}^{a\varepsilon n}\cdot n C_{2\varepsilon n}^{b\varepsilon n}(3\varepsilon + a\varepsilon)^{2\varepsilon n - b\varepsilon n} <$$
$$< \exp\{[n\ln n + {}^1/_2\ln n - \varepsilon n\ln \varepsilon n - {}^1/_2\ln \varepsilon n - (n - \varepsilon n)\ln(n-\varepsilon n) - {}^1/_2\ln(n-\varepsilon n)] + [(n-\varepsilon n)\ln(n - \varepsilon n) + {}^1/_2\ln(n-\varepsilon n) - a\varepsilon n\ln a\varepsilon n - {}^1/_2\ln a\varepsilon n - (n - \varepsilon n - a\varepsilon n)\ln(n - \varepsilon n - a\varepsilon n) - {}^1/_2\ln(n-\varepsilon n - a\varepsilon n)] + \ln n + [2\varepsilon n\ln 2\varepsilon n + {}^1/_2\ln 2\varepsilon n - b\varepsilon n\ln b\varepsilon n - {}^1/_2\ln b\varepsilon n - (2\varepsilon n - b\varepsilon n)\ln(2\varepsilon n - b\varepsilon n) - {}^1/_2\ln(2\varepsilon n - b\varepsilon n)] + (2\varepsilon n - b\varepsilon n)\ln(3\varepsilon + a\varepsilon)\} =$$
$$= \exp\{n[\varepsilon(1-a-b)\ln\varepsilon - \varepsilon a\ln a + \varepsilon\ln 4 - \varepsilon b\ln b - \varepsilon(2-b)\ln(2-b) + \varepsilon(2-b)\ln(3+a) - (1-\varepsilon-\varepsilon a)\ln(1-\varepsilon-\varepsilon a)] + [-{}^1/_2\ln n - {}^1/_2\ln\varepsilon - {}^1/_2\ln a - {}^1/_2\ln\varepsilon - {}^1/_2\ln(1-\varepsilon-\varepsilon a) + {}^1/_2\ln 2 + {}^1/_2\ln\varepsilon - {}^1/_2\ln b - {}^1/_2\ln\varepsilon - {}^1/_2\ln\varepsilon - {}^1/_2\ln(2-b)]\} <$$

(ввиду того что ε, a, b не зависят от n)

$$< \exp\{n[\varepsilon(1-a-b)\ln\varepsilon - \varepsilon a\ln a + \varepsilon\ln 4 - \varepsilon b\ln b - \varepsilon(2-b)\ln(2-b) + \varepsilon(2-b)\ln(3+a) - (1-\varepsilon-\varepsilon a)\ln(1-\varepsilon-\varepsilon a)]\} <$$

(ввиду предположений (а′) и (б′))

$$< \exp\{n[\varepsilon[-(1-a-b)\,|\ln\varepsilon| - a\ln a + \ln 4 - b\ln b - (2-b)\ln(2-b) + (2-b)\ln(3+a)] + |\ln(1-\varepsilon(1-a))|\,]\}.$$

Пусть $a = a_0' > 0$ и $b = b_0' > 0$ зафиксированы произвольным образом, но так, чтобы $1 - a_0' - b_0' > 0$. Теперь учтем, что при $\varepsilon \to 0$ $|\ln(1-\varepsilon(1+a))| \to \varepsilon(1+a)$ и $|\ln\varepsilon| \to \infty$. Теперь нетрудно видеть, что найдется такое ε_0', удовлетворяющее неравенствам (а′) и (б′), при котором выражение

$$[\varepsilon_0'[-(1-a_0'-b_0')\,|\ln\varepsilon_0'| - a_0'\ln a_0' + \ln 4 - b_0'\ln b_0' - (2-b_0')\ln(2-b_0') + (2-b_0')\ln(3+a_0')] + |\ln(1-\varepsilon_0'(1+a_0'))|\,]$$

будет меньше некоторой отрицательной константы. Это будет означать, что при данных a_0', b_0' и ε_0' величина $P(\varepsilon_0', a_0', \varepsilon_0', b_0'\varepsilon_0', n) \to 0$, когда $n \to \infty$.

Для полного доказательства утверждения В остается только убедиться, что $P(\varepsilon, a\varepsilon, b_0'\varepsilon, n) \to 0$ при $n \to \infty$ для любых $\varepsilon \leqslant \varepsilon_0'$

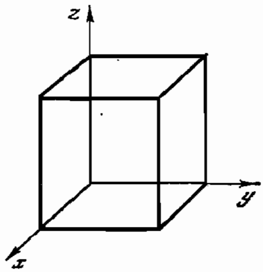


Рис. 5

и $a \leqslant a_0'$. Справедливость этого вытекает из того, что если выражение

$$[\varepsilon[-(1-a-b)\,|\ln\varepsilon|-a\ln a+\ln 4-b\ln b-(2-b)\ln(2-b)+(2-b)\ln(3+a)]+|\ln(1-\varepsilon(1+a))|]$$

было меньше некоторой отрицательной константы при $\varepsilon=\varepsilon_0'$, $a=a_0'$ и $b=b_0'$, то оно будет обладать этим же свойством и при $\varepsilon\leqslant\varepsilon_0'$, $a\leqslant a_0'$ и $b=b_0'$.

Лемма доказана.

Теперь учтем следующих два обстоятельства:

1) согласно определению реализации расстояние между любыми двумя проводниками, соответствующими неинцидентным дугам, $\geqslant 1$;

2) согласно лемме 1 у почти всех сетей с n вершинами любые $\varepsilon_0 n$ вершин связаны с остальными вершинами по меньшей мере с помощью $b_0\varepsilon_0 n$ неинцидентных дуг (константы ε_0 и b_0 из леммы 1).

Отсюда вытекает справедливость следующей леммы.

Лемма 2. *Почти все сети $\mathfrak{A}$ с n вершинами обладают следующим свойством: при любой реализации сети $\mathfrak{A}$ в трехмерном пространстве в любом параллелепипеде с площадью поверхности $L\leqslant {}^1/_2 b_0\varepsilon_0 n$ находится меньше чем $\varepsilon_0 n$ φ-точек (константы ε_0 и b_0 из леммы 1).*

Доказательство теоремы 2. Пусть μ и s — некоторые числа (их значения мы подберем позже). Пусть $\mathfrak{A}$ — некоторая сеть с n вершинами, для которой имеют место утверждения лемм 1 и 2 (такие сети — почти все). Пусть, далее, сеть $\mathfrak{A}$ реализована в трехмерном пространстве и пусть T — некоторый параллелепипед с длиной ребер $\geqslant s\mu\sqrt{n}$, в котором лежат все φ-точки и проводники, возникающие при данной реализации сети $\mathfrak{A}$ (рис. 5).

Тогда:

а) С помощью плоскостей $z=i\mu\sqrt{n}$ $(i=0, 1, 2, \ldots)$ разбиваем параллелепипед T на параллельные слои толщины $\mu\sqrt{n}$ (без ограничения общности будем считать что последний слой тоже имеет толщину $\mu\sqrt{n}$). Перенумеруем эти слои в направлении оси Oz числами 1, 2, 3, ... (таким образом, самый низкий слой будет иметь номер 1, следующий по высоте — номер 2 и т. д.). Затем на основании этого разбиения строим разбиение параллелепипеда T на части $A_1, A_2, \ldots, A_s$; к A_i относим те и только те слои, которые имеют номер вида $ks+i$, где k — целое число. Затем из совокупности этих частей выбираем ту, которая содержит наименьшее число φ-точек. Пусть это будет часть A_z. Нетрудно видеть, что

A_z содержит не больше чем n/s φ-точек. Слои, принадлежащие A_z, будем называть отмеченными.

б) С помощью плоскостей $y = i\mu\sqrt{n}$ ($i = 0, 1, 2, \ldots$) разбиваем параллелепипед T на параллельные слои, затем строим разбиение $B_1, B_2, \ldots, B_s$ аналогично пункту а) и слои, принадлежащие части B_y, содержащей наименьшее число φ-точек, объявляем отмеченными.

в) С помощью плоскостей $x = i\mu\sqrt{n}$ ($i = 0, 1, 2, \ldots$) разбиваем параллелепипед T на параллельные слои, затем строим разбиение $C_1, C_2, \ldots, C_s$ аналогично пункту а) и слои, принадлежащие части C_x, содержащей наименьшее число φ-точек, объявляем отмеченными.

Из предыдущего следует, что отмеченные слои параллелепипеда T содержат не больше чем $3n/s$ φ-точек.

Пусть теперь числа μ и s такие, что кубики, ограниченные отмеченными слоями, имеют площадь поверхности не больше $^1/_2 b_0 \varepsilon_0 n$ (имеется в виду каждый кубик в отдельности), т. е. пусть μ и s удовлетворяют неравенству

$$6\,(s-1)^2\,\mu^2 n \leqslant {}^1/_2 b_0 \varepsilon_0 n \tag{4}$$

(константы ε_0 и b_0 из леммы 1).

Тогда согласно лемме 2 каждый из упомянутых кубиков содержит меньше чем $\varepsilon_0 n$ φ-точек. Отсюда в свою очередь следует, что если число φ-точек, содержащихся в неотмеченных слоях, не меньше $\varepsilon_0 n$, т. е.

$$n - 3n/s \geqslant \varepsilon_0 n, \tag{5}$$

то можно построить такое множество этих кубиков G, которое содержит не меньше чем $^1/_2\varepsilon_0 n$ и не больше чем $\varepsilon_0 n$ точек [7].

Пусть, далее, число s такое, что отмеченные слои содержат не больше чем $a_0\,(\varepsilon_0/2)n$ φ-точек, т. е. пусть s удовлетворяет неравенству

$$3n/s \leqslant a_0\,(\varepsilon_0/2)\,n. \tag{6}$$

Рассмотрим следующее (ε, δ)-разбиение сети $\mathfrak{A}$:

а) ω_1 — это множество тех вершин сети $\mathfrak{A}$, которым при рассматриваемой реализации сети $\mathfrak{A}$ сопоставлены точки из G; число элементов εn множества ω_1 равно числу φ-точек, содержащихся в G, т. е. $\varepsilon_0/2 \leqslant \varepsilon \leqslant \varepsilon_0$;

б) ω_2 — это множество тех вершин сети $\mathfrak{A}$, которым при рассматриваемой реализации сети $\mathfrak{A}$ сопоставлены точки из отмеченных слоев; таким образом, число элементов δ_n множества ω_2 равно числу φ-точек, содержащихся в отмеченных слоях, т. е. $\delta \leqslant a_0\varepsilon_0/2$.

---

[7] Здесь и в дальнейшем под выражением «множество G содержит φ-точки» будем понимать «элементы множества G (т. е. кубики) содержат φ-точки».

По лемме 1 получаем, что сеть $\mathfrak{A}$ при таком (ε, δ)-разбиении имеет степень связности $\geqslant b_0 \varepsilon n \geqslant b_0 (\varepsilon/2) n$. Это означает, что φ-точки, содержащиеся в множестве кубиков G и φ-точки, содержащиеся в остальных кубиках, ограниченных отмеченными слоями, соединены между собой по меньшей мере $b_0(\varepsilon_0/2)n$ неинцидентными проводниками. Множество этих проводников обозначим через E. Согласно определению реализации расстояние между проводниками из E должно быть $\geqslant 1$. Так как толщина отмеченных слоев равна $\mu \sqrt{n}$, то каждый из проводников множества E соединяет точки, находящиеся на расстоянии $\geqslant \mu \sqrt{n}$. Отсюда получаем, что любое тело W', которое содержит все проводники из E и при том так, что они находятся от его поверхности на расстоянии $\geqslant 1$, имеет объем больше чем

$$b_0 (\varepsilon_0/2) n\pi (^1/_2)^2 \mu \sqrt{n} = (\pi/8)\, b_0 \varepsilon_0 \mu n \sqrt{n}.$$

Теперь вспомним, что эти оценки были получены при условии, что s и μ удовлетворяют неравенствам (4), (5) и (6). Нетрудно убедиться, что существуют s и $\mu > 0$, не зависящие от n, которые удовлетворяют этим неравенствам. Это означает, что объем сети $\mathfrak{A}$

$$V(\mathfrak{A}) > (\pi/8)\, b_0 \varepsilon_0 \mu n \sqrt{n} = C_2 n \sqrt{n},$$

где C_2 — некоторая константа > 0, не зависящая от n.

Теорема 2 доказана.

12

К ЛОГИЧЕСКИМ ОСНОВАМ ТЕОРИИ ИНФОРМАЦИИ И ТЕОРИИ ВЕРОЯТНОСТЕЙ *

1. Мы будем иметь дело с основными начальными понятиями теории информации. Исходным будем считать понятие условной энтропии объекта x при заданном объекте y, $H(x \mid y)$, которую можно интерпретировать как количество информации, необходимое для задания объекта x в обстановке, когда объект y уже задан. Обозначая через ϕ «заведомо заданный объект», получим безусловную энтропию

$$H(x \mid \phi) = H(x).$$

Информация, содержащаяся в объекте y относительно объекта x, определяется формально при помощи вычитания

$$I(x \mid y) = H(x) - H(x \mid y).$$

Естественно, что при этом

$$I(x \mid x) = H(x).$$

* Проблемы передачи информации, 1969, т. 5, № 3, с. 3—7.

Обычно пользуются вероятностным определением энтропии, которое относится, однако, не к индивидуальным объектам, а к «случайным», т. е. по существу к распределениям вероятностей в каких-либо классах объектов. Чтобы подчеркнуть это различие, будем обозначать случайные объекты греческими буквами. Ограничиваясь случаем дискретных распределений, напомним стандартное определение

$$H(\xi \mid \eta) = -\sum_{x,\,y} p(\xi = x, \eta = y) \log_2 p(\xi = x \mid \eta = y) \ldots \qquad (1)$$

Как известно, «вероятностные» высказывания реально интерпретируются через высказывания статистические. Иначе говоря, наше определение (1) может быть практически использовано лишь в применении к обширным статистическим совокупностям пар объектов

$$(x_1, y_1),\ (x_2, y_2), \ldots, (x_n, y_n), \ldots$$

Далеко не все применения теории информации укладываются разумным образом в такую интерпретацию ее основных понятий. Я думаю, что потребность в придании определенного смысла выражениям $H(x \mid y)$ и $I(x \mid y)$ в случае индивидуальных объектов x и y, не рассматриваемых как реализации случайных испытаний с определенным законом распределения, была понятна давно многим, занимавшимся теорией информации.

Насколько я знаю, первой печатной публикацией, содержащей замысел перестройки теории информации, удовлетворяющей высказанному пожеланию, является статья Соломонова [1], опубликованная в 1964 г. Я пришел к аналогичной концепции, еще не зная работ Соломонова в 1963—1964 гг., и опубликовал первую заметку на эту тему [2] в начале 1965 г. Разработкой новой концепции энергично занялся молодой шведский математик Мартин-Лёф, работавший в 1964—1965 гг. в Москве. Прочитанные им в Эрлангене (1966) лекции [3] представляют, кажется, лучшее введение в весь круг идей моей статьи.

Замысел определения очень прост: энтропия $H(x \mid y)$ есть минимальная длина записанной в виде последовательности нулей и единиц «программы» P, которая позволяет построить объект x, имея в своем распоряжении объект y

$$H(x \mid y) = \min_{A(P,\,y)=x} l(P). \qquad (2)$$

Точное проведение этого замысла опирается на общую теорию «вычислимых» (частично рекурсивных) функций, т. е. на общую теорию алгоритмов во всей их общности. Мы вернемся далее к расшифровке записи $A(P, y) = x$.

Хотя мне и Мартин-Лёфу важность новой концепции представлялась несомненной, ее развитие тормозилось тем, что самые простые формулы, которые в вероятностной концепции получаются в результате очень простых алгебраических преобразований выр жения (1), в новой концепции не удавалось восстановить. К числ таких не восстанавливаемых в новой концепции формул относитс формула

$$I\ (x \mid y) = I\ (y \mid x). \qquad (3$$

Впрочем, если поразмыслить, ее реальное содержание и в само деле не тривиально, а скорее должно бы вызвать сомнения в бе условной применимости этой формулы, как и формулы «разложе ния» информации

$$H\ (x, y) = H\ (x) + H\ (y \mid x). \qquad (4$$

Формулы (3) и (4) так привычны, что не сразу мы обратил внимание на то, что в новой концепции они просто неверны и м гут быть восстановлены в правах лишь в форме приближенных ра венств

$$|\ I\ (x \mid y) - I\ (y \mid x)\ | = O\ (\log_2 H\ (x, y)), \qquad (3'$$

$$H\ (x, y) = H\ (x) + H\ (y \mid x) + O\ (\log_2 H\ (x, y)). \qquad (4'$$

2. Рассмотрим длинную последовательность

$$x = (x_1, x_2, \ldots, x_l), \qquad l = l\ (x),$$

состоящую из нулей и единиц. Легко понять, что существуют та кие последовательности, для которых энтропия $H\ (x)$ не меньш их длины,

$$H\ (x) \geqslant l\ (x).$$

Такие последовательности нельзя определить программой, кото рая короче, чем их длина. Чтобы их задать, надо их просто выпи сать. Никакой более просто выражаемой закономерностью их нель зя определить. Естественно вспомнить, что отсутствие закономе ности по здравому смыслу является признаком случайности. М практически исходим из допущения, что так устроены помещае мые в руководствах по математической статистике и теории вероят ностей «таблицы случайных чисел».

Чтобы пойти немного дальше, рассмотрим, как мы представля ем себе последовательность нулей и единиц, возникающую в ре зультате независимых испытаний с вероятностью p получения еди ницы при каждом испытании. Если l велико, то число единиц при близительно равно lp, т. е. частота $k/l \approx p$ дает представление о некоторой закономерности, имеющейся в последовательности x. В силу этой закономерности условная энтропия x может быть оце нена неравенством

$$H\ (x \mid l, k) \leqslant \log_2 C_l^k + O\ (1)$$

(о смысле добавки члена O (1) в нашей формуле будет сказано далее). Если энтропия $H(x|\, l,\, k)$ близка к этой верхней границе, то не существует существенно более экономного способа задания x, чем указать l, k и номер последовательности x среди всех C_l^k последовательностей с данными l и k. Примерно так мы и представляем себе «бернуллиевские последовательности», в которых отдельные знаки «независимы», возникая с некоторой «вероятностью» p.

Мы видим, таким образом, что некоторый аналог понятия «бернуллиевской» последовательности может быть сформулирован на языке намеченной выше алгоритмической теории информации.

Специально бернуллиевским последовательностям в этом новом смысле посвящена работа Мартин-Лёфа [4]. Естественно, что понятие бернуллиевской конечной последовательности должно быть релятивизированным. Для конечных последовательностей нет резкой грани между «закономерным» и «случайным». По Мартин-Лёфу последовательность является «m-бернуллиевской», если

$$H\,(x \mid l\,(x),\, k\,(x)) \geqslant \log_2 C_{l(x)}^{k(x)} - m.$$

Строгое различение между «бернуллиевскими» и «небернуллиевскими» последовательностями возможно лишь после предельного перехода для бесконечных последовательностей нулей и единиц

$$x = (x_1,\, x_2,\, \ldots,\, x_n,\, \ldots).$$

Будем для такой последовательности обозначать через x^l их начальный отрезок длины l

$$x^l = (x_1,\, \ldots,\, x_l).$$

Было бы соблазнительно определить x как «бернуллиевскую» требованием существования такого m, что все x^l являются m-бернуллиевскими, т. е. что всегда

$$H(x^l \mid l,\, k_l) \geqslant \log_2 C_l^{k_l} - m.$$

Но таких бесконечных последовательностей нулей и единиц, как показал Мартин-Лёф, не существует. Причины легко объяснимы. Ведь и из традиционной теории вероятностей мы знаем, что в настоящих случайных последовательностях встречаются сколь угодно длинные последовательности сплошных единиц и сплошных нулей. Понятно, что описание заканчивающихся таким образом отрезков бесконечной последовательности может быть существенно упрощено по сравнению со стандартным.

Наиболее естественное определение бесконечной бернуллиевской последовательности таково: x считается m-бернуллиевской, если существует такое m, что все x^l являются начальными участками конечных m-бернуллиевских последовательностей. Мартин-Лёф дает другое, быть может немного более узкое, определение.

Более частная концепция (с которой начал и Мартин-Лёф) $^1/_2$-бернуллиевской последовательности была независимо развита Шайтиным [5].

Возможно, что некоторые читатели уже заметили, что, обратившись к бесконечным последовательностям, мы занимаемся задачей, которая была поставлена уже Мизесом в его концепции «коллективов». Как известно, концепция Мизеса была формализована на языке теории вычислимых функций Чёрчем [6]. Существенное дополнение концепции Мизеса (расширение понятия «допустимой системы выбора») было дано в моей работе [7]. Строго формальное изложение дополненной таким образом теории можно найти в статьях Ловеланда [8, 9].

Однако класс «бернуллиевских» последовательностей в смысле Чёрча или Ловеланда слишком широк. Их отрезки могут быть еще довольно «закономерными». Существуют последовательности бернуллиевские в смысле обоих этих определений с отрезками, имеющими лишь логарифмически возрастающую энтропию:

$$H(x^l) = O(\log_2 l).$$

Наоборот, бернуллиевские последовательности по Мартин-Лёфу (и по указанному выше определению) имеют отрезки, сложность которых почти максимальна в смысле неравенства

$$H(x^l) \geqslant l - O((\log_2 l)^{1+\varepsilon}), \quad \varepsilon > 0 \text{ произвольно.}$$

Последовательности, бернуллиевские по Мартин-Лёфу, обладают всеми конструктивными свойствами, которые в современной теории вероятностей доказываются (при любой вероятности p) «с вероятностью единица». Ничего аналогичного нельзя утверждать о последовательностях, бернуллиевских по Чёрчу или Ловеланду.

3. Предшествующее краткое изложение должно оправдать два общих тезиса:

1) основные понятия теории информации должны и могут быть обоснованы без помощи обращения к теории вероятностей и так, что понятия «энтропия» и «количество информации» оказываются применимы к индивидуальным объектам;

2) введенные таким образом понятия теории информации могут лечь в основу новой концепции случайного, соответствующей естественной мысли о том, что случайность есть отсутствие закономерности.

Сделаем важное замечание.

Изложение в первой части статьи (информация) было несколько упрощено. Лишь во второй части статьи мы подчеркнули неизбежную относительность различения между «случайным» и «неслучайным» в применении к конечным объектам. Аналогично дело обстоит и в основах теории информации. По существу она применима к большим массивам информации, при рассмотрении которых начальная информация, заключенная в способе построения самой теории, исчезающе мала. Наша основная формула (2) подразумевает «универсальный метод программирования» A. Такие существуют в том смысле, что имеются методы программирования A, обладающие свойством $H_A(x \mid y) \leqslant H_{A'}(x \mid y) + C_{A'}$ и позволяющие программировать «что угодно» с длиной программ, которая превосходит длину при любом другом методе программирования не более чем на константу, зависящую лишь от этого второго метода программирования, а не от объектов x и y. В обращении внимания на это довольно простое обстоятельство и состоит, по-видимому, заслуга моя и Соломонова.

Поэтому все предложения алгоритмической теории информации в их общей формулировке верны лишь с точностью до членов вида $O(1)$. В применении к формулам (3) и (4) неожиданностью явилось лишь появление членов логарифмического порядка.

Важно лишь понимать, что, обращаясь к теории вероятностей, мы прибегаем к значительно более грубой релятивизации. Реальное истолкование вероятностных результатов всегда статистическое, и оценки ошибок, получающихся при применении вероятностных результатов к конечным объектам, значительно грубее, чем в развиваемом нами изложении теории информации.

ЛИТЕРАТУРА

1. *Solomonoff R.* A formal theory of inductive inference. I.— Inform. and Contr., 1964, vol. 7, N 1, p. 1—22.
2. *Колмогоров А. Н.* Три подхода к определению понятия «количество информации».— Проблемы передачи информации, 1965, т. 1, № 1, с. 3—11.
3. *Martin-Löf P.* Algorithms and random sequences. Erlangen: Univ. Press, 1966.
4. *Martin-Löf P.* The definition of random sequences.— Inform. and Contr., 1966, vol. 9, N 6, p. 602—619.
5. *Chaitin G.* On the length of programs for computing finite binary sequences.— J. Assoc. Comput. Mach., 1966, vol. 13, N 4, p. 547—569.
6. *Church A.* On the concept of a random sequence.— Bull. Amer. Math. Soc., 1940, vol. 46, p. 130—135.
7. *Kolmogorov A. N.* On tables of random numbers.— Sankhya. A, 1963, vol. 25, N 4, p. 369—376.
8. *Loveland D. A.* A new interpretation of the von Mises concept of random sequence.— Ztschr. math. Log. und Grundl. Math., 1966, Bd. 12, S. 279—294.
9. *Loveland D.* The Kleene hierarchy classification of recursively random sequences.— Trans. Amer. Math. Soc., 1966, vol. 125, N 3, p. 497—510.

13

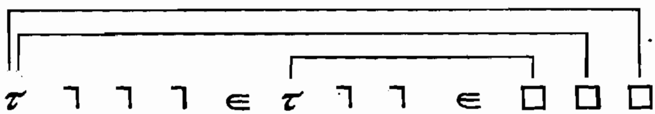
КОМБИНАТОРНЫЕ ОСНОВАНИЯ ТЕОРИИ ИНФОРМАЦИИ И ИСЧИСЛЕНИЯ ВЕРОЯТНОСТЕЙ *[1]

§ 1
О ВОЗРАСТАЮЩЕЙ РОЛИ ФИНИТНОЙ МАТЕМАТИКИ

Мне хочется начать с некоторых соображений, выходящих за рамки основной темы моего доклада. Формализованная по Гильберту математика является не чем иным, как теорией операций над специального вида схемами, состоящими из конечного числа знаков, расположенных в том или ином порядке и связанных теми или иными связями. Например, по концепции Н. Бурбаки вся теория множеств изучает исключительно выражения, составленные из знаков

$$\square,\ \tau,\ \vee,\ \urcorner,\ =,\ \in,\ \supset$$

и «букв», связанных еще «связями» ⌐, как это имеет место, например, в выражении



называемом «пустым множеством». Оставаясь на финитной точке зрения, было бы логично для неограниченной последовательности «букв» усвоить те или иные стандартные обозначения, например

$$\Pi_0,\ \Pi_1,\ \Pi_{10},\ \Pi_{11},\ \Pi_{100},\ \ldots;$$

Любопытно, что из-за наличия «связей» ⌐ выражения формализованной математики Бурбаки являются не вытянутыми в одну строчку «словами», как, например, в теории нормальных алгорифмов А. А. Маркова, а по существу одномерными комплексами с обозначенными определенными знаками вершинами.

Но к реальному, воспринимаемому интуитивно содержанию математики такая концепция математики, как занятия перестройкой по определенным правилам однозначных одномерных комплексов, имеет лишь косвенное отношение. Н. Бурбаки замечает, что в его концепции выражение, имеющее смысл «число единица», содержит несколько десятков тысяч знаков, но от этого понятие

---

* УМН, 1983, т. 38, вып. 4, с. 27—36.

[1] Публикуемый текст был подготовлен в 1970 г. в связи с моим выступлением на Международном математическом конгрессе в Ницце.

«число единица» не делается недоступным нашему наглядному пониманию.

Чистая математика благополучно развивается как по преимуществу наука о бесконечном. И сам основатель концепции формализованной полностью финитной математики Гильберт предпринял свой титанический труд лишь для того, чтобы обеспечить за математиками право оставаться в «канторовом парадизе» теории множеств. По-видимому, это положение вещей глубоко обосновано устройством нашего сознания, с большой легкостью оперирующего с наглядными представлениями о неограниченных последовательностях, предельных переходах, непрерывных и даже «гладких» многообразиях и т. п.

До недавнего времени и в математическом естествознании господствовало моделирование реальных явлений при помощи математических моделей, построенных на математике бесконечного и непрерывного. Например, изучая процесс молекулярной теплопроводности, мы представляем себе непрерывную среду, в которой температура подчинена уравнению

$$\frac{\partial u}{\partial t} = K\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right). \tag{1}$$

Математики привыкли рассматривать соответствующую разностную схему

$$\Delta_t u = K\,(\Delta_{xx} u + \Delta_{yy} u + \Delta_{zz}\, u) \tag{2}$$

лишь как возникающую при приближенном решении «точного» уравнения (1). Но реальный процесс теплопроводности не более похож на свою непрерывную модель, выраженную уравнением (1), чем на дискретную модель, выраженную непосредственно уравнением (2).

Весьма вероятно, что с развитием современной вычислительной техники будет понято, что в очень многих случаях разумно изучение реальных явлений вести, избегая промежуточный этап их стилизации в духе представлений математики бесконечного и непрерывного, переходя прямо к дискретным моделям. Особенно это относится к изучению сложно организованных систем, способных перерабатывать информацию. В наиболее развитых таких системах тяготение к дискретности работы вызвано достаточно разъясненными в настоящее время причинами. Является требующим объяснения парадоксом то обстоятельство, что человеческий мозг математика работает в существенном по дискретному принципу и тем не менее математику значительно доступнее интуитивное постижение, скажем, свойств геодезических на гладких поверхностях, чем могущих их приблизительно заменить свойств комбинаторных схем.

Пользуясь своим мозгом, как данным от господа бога, математик мог не интересоваться комбинаторными основами его работы. Но искусственный интеллект машин должен быть создан человеком, и человеку приходится погрузиться в неизбежную при этом комбинаторную математику. Пока еще рано делать окончательные выводы о том, что это будет значить для общей архитектуры математики будущего.

§ 2
ТЕОРИЯ ИНФОРМАЦИИ

Дискретные формы хранения и переработки информации являются основными. На них основана и сама мера «количества информации», выражаемого в «битах» — числе двоичных знаков. В силу сказанного ранее дискретная часть теории информации призвана в некотором роде играть ведущую, организующую роль в развитии комбинаторной финитной математики. Из развитых вкратце общих соображений не видно, почему теория информации должна столь существенно основываться на теории вероятностей, как это представляется по большинству руководств. Моей задачей является показать, что эта зависимость от заранее созданной теории вероятностей в действительности не является неизбежной. Ограничусь, впрочем, двумя простыми примерами.

Реальное содержание формулы для энтропии [2]

$$H = -\Sigma p_i \log p_i \tag{1}$$

таково. Если мы произведем большое число n независимых испытаний с распределением вероятностей

$$(p_1, p_2, \ldots, p_s)$$

s исходов каждого испытания, то для фиксирования исхода всей серии из n испытаний нам понадобится приблизительно

$$nH$$

двоичных знаков. Но этот результат справедлив при несравненно более слабых и чисто комбинаторных допущениях. Для записи результата наших испытаний достаточно указать, что каждый из исходов появился соответственно

$$m_1, m_2, \ldots, m_s$$

раз, и уже после этого указать номер того из

$$C(m_1, m_2, \ldots, m_s) = \frac{n!}{m_1!\, m_2! \ldots m_s!}$$

расположений исходов, который имел место. Для этого потребует-

---

[2] Логарифмы всюду двоичные.

ся не более

$$s \log n + \log C(m_1, m_2, \ldots, m_s)$$

двоичных знаков, а при большом n это и будет приблизительно

$$n\left(-\sum \frac{m_i}{n} \log \frac{m_i}{n}\right) \sim nH. \qquad (2)$$

В силу закона больших чисел в случае независимых испытаний с указанным выше распределением вероятностей $m_i/n \sim p_i$. Но наши допущения при выводе формулы (2) несравненно слабее.

Второй пример с энтропией цепи Маркова вполне аналогичен. Здесь тоже допущение о том, что требуется записать информацию о реализации марковского процесса, сильно избыточно.

§ 3
ОПРЕДЕЛЕНИЕ «СЛОЖНОСТИ»

Если какой-либо объект «просто» устроен, то для его описания достаточно небольшого количества информации; если же он «сложен», то его описание должно содержать много информации. По некоторым соображениям (см. далее § 7) нам удобно назвать вводимую сейчас величину «сложностью».

Стандартным способом задания информации считаем двоичные последовательности, начинающиеся с единицы,

$$1, 10, 11, 100, 101, 110, 111, 1000, 1001, \ldots,$$

являющиеся двоичными записями натуральных чисел. Будем обозначать через $l(n)$ длину последовательности n.

Пусть мы имеем дело с какой-либо областью объектов D, в которой уже имеется некоторая стандартная нумерация объектов номерами $n(x)$. Однако указание номера $n(x)$ далеко не всегда будет наиболее экономным способом выделения объекта x. Например, двоичная запись числа

$$9^{9^9}$$

необозримо длинна, но мы его определили достаточно просто. Надо подвергнуть сравнительному изучению различные способы задания объектов из D. Достаточно ограничиться способами задания, которые каждому двоично записанному числу p ставят в соответствие некоторый номер

$$n = S(p).$$

Таким образом, способ задания объекта из D становится не чем иным, как функцией S от натурального аргумента с натуральными значениями. Немного далее мы обратимся к случаю, когда эта функция вычислима. Такие методы задания можно назвать «эффективными». Но пока сохраним полную общность. Для каждого объекта из D естественно рассмотреть приводящие к нему p наименьшей длины $l(p)$. Эта наименьшая длина и будет «сложностью» объекта x при «способе задания S»:

$$K_S(x) = \min l(p), \qquad S(p) = n(x).$$

На языке вычислительной математики можно назвать p «программой», а S — «методом программирования». Тогда можно будет сказать, что p есть минимальная длина программы, по которой можно получить объект x при методе программирования S.

Если имеется несколько различных способов задания элементов из D

$$S_1, S_2, \ldots, S_r,$$

то легко построить новый метод S, который будет доставлять нам любой объект $x \in D$ со сложностью $K_S(x)$, лишь примерно на $\log r$ превосходящей минимум из сложностей

$$K_{S_1}(x), K_{S_2}(x), \ldots, K_{S_r}(x).$$

Построение такого метода очень просто. Надо резервировать достаточное число начальных знаков последовательности p для фиксации метода S_i, которому надо следовать, пользуясь как программой оставшимися знаками p.

Скажем, что метод S «с точностью до l поглощает метод S'», если всегда

$$K_S(x) \leqslant K_{S'}(x) + l.$$

Выше показано, как построить метод S, который с точностью до l сильнее любого из методов $S_1, S_2, \ldots, S_r$, где приблизительно $l \backsim \log r$.

Методы S_1 и S_2 назовем «l-эквивалентными», если каждый из них l-поглощает другой. Все это построение было бы мало продуктивным, если бы иерархия методов в отношении поглощения одних другими была слишком причудливой. Лишь сравнительно недавно было замечено, что при некоторых достаточно естественных предположениях это не так. Я далее следую своей работе [1], но примерно те же идеи можно найти в работах [3—5], в работе [3], впрочем, в несколько завуалированной форме.

Теорема. *Среди вычислимых функций $S(p)$ существуют оптимальные, т. е. такие, что для любой другой вычислимой функции $S'(p)$*

$$K_S(x) \leqslant K_{S'}(x) + l(S, S').$$

Ясно, что все оптимальные способы задания объектов из D эквивалентны:

$$|K_{S_1}(x) - K_{S_2}(x)| \leqslant l(S_1, S_2).$$

Таким образом, с асимптотической точки зрения сложность $K(x)$ элемента x при ограничении эффективными способами задания не зависит от случайных особенностей выбранного оптимального метода. Конечно, чисто практический интерес этого результата зависит от того, насколько велики расхождения в сложностях при различных достаточно гибких, но вместе с тем и достаточно удобных и естественных методах программирования.

§ 4
ЗАКОНОМЕРНОСТЬ И СЛУЧАЙНОСТЬ

Вполне традиционно представление, что «случайность» состоит в отсутствии «закономерности». Но, по-видимому, только сейчас возникла возможность основать точные формулировки условий применимости к реальным явлениям результатов математической теории вероятностей непосредственно на этой простой идее.

Любые результаты наблюдений могут быть запротоколированы в виде конечной, хотя иногда и весьма длинной записи. Поэтому, когда говорят об отсутствии в результатах наблюдений закономерности, имеют в виду лишь отсутствие *достаточно простой* закономерности. Например, последовательность из 1000 цифр

1274031274031274031...,

сменяющихся с периодом в шесть цифр, мы с несомненностью отнесем к «закономерным», а не «случайным» последовательностям. Последовательность первой тысячи десятичных знаков дробной части числа π

1415...,

как известно, обладает многими свойствами «случайных последовательностей». Узнав ее закон образования, мы и ее откажемся признать «случайной». Но если нам выпишут многочлен 999-й степени, значения которого при $x = 1, 2, 3, \ldots, 1000$ доставляют последовательность целых чисел $p(x)$, заключенных в пределах от 0 до 9, полученных в результате частных случайных испытаний типа игры в рулетку, то наличие такого многочлена не помешает нам продолжать считать последовательность «случайной».

Если тем или иным способом мы пришли к выводу, что последовательность результатов каких-либо испытаний не допускает полного описания в приемлемой для нас в отношении сложности форме, то мы скажем, что эта последовательность разве что лишь частично закономерна, отчасти же «случайна». Но это еще не та «случайность», которая нужна для применимости выводов теории вероятностей. Применяя теорию вероятностей, мы не ограничиваемся отрицанием закономерности, а делаем из гипотез о случайности наблюдаемых явлений определенные положительные выводы.

Мы сейчас увидим, что практические выводы теории вероятностей могут быть обоснованы в качестве следствий гипотез о *предельной* при данных ограничениях сложности изучаемых явлений.

§ 5
УСТОЙЧИВОСТЬ ЧАСТОТ

Следуя Мизесу, часто выдают за основу приложений теории вероятностей признание гипотезы об устойчивости частот. В приближенной к практике форме эта концепция принята и в моей известной книжке по основаниям теории вероятностей (1933). Пусть результат последовательности большого числа N испытаний записан в виде последовательности нулей и единиц

$$1101001011100101 1\ldots$$

Говорят, что выпадение единицы случайно с вероятностью p, если доля единиц

$$M/N \sim p \tag{1}$$

и эта частота не может быть существенно изменена при выборе из нашей последовательности не слишком короткой подпоследовательности, произведенной по достаточно простому закону и притом так, что зачисление какого-либо элемента основной последовательности в подпоследовательность производится без использования значений этого элемента [3].

Но оказывается, что такое требование может быть заменено другим, формулируемым существенно проще. Сложность последовательности нулей и единиц с условием (1) не может быть существенно больше чем

$$nH(p) = n(-p \log p - (1-p) \log (1-p)).$$

Можно доказать, что мизесовская устойчивость частот автоматически выполняется, если сложность нашей последовательности достаточно близка к указанной верхней оценке.

У меня нет здесь возможности дать количественные уточнения этого результата [4] и рассмотреть с той же точки зрения более сложные задачи теории вероятностей. Но принцип является общим. Например, предполагая, что последовательность нулей и единиц образует цепь Маркова с матрицей переходных вероятностей

$$\begin{Vmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{Vmatrix},$$

---

[3] Наиболее тщательную финитную формулировку этого принципа Мизеса см. в работе [17].

[4] См. в [17], хотя там еще нет появившегося в [1] определения сложности.

мы по существу задаем приближенные значения частот единиц после единиц, единиц после нулей, нулей после единиц и нулей после нулей. Можно вычислить максимальную сложность такой последовательности длины n. Если сложность конкретной последовательности с заданными частотами переходов близка к этой максимальной, то для нее автоматически выполняются все предсказания вероятностной теории цепей Маркова.

§ 6
БЕСКОНЕЧНЫЕ СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

Только что намеченная программа пока еще не выполнена, хотя у меня нет сомнений в ее выполнимости. Именно ее выполнение должно более совершенным образом, чем построения мизесовского типа, связать математическую теорию вероятностей с ее применениями. При этом я имею в виду, что нет никакой необходимости изменять сложившееся построение математической теории вероятностей на основе общей теории меры. Исследованиям, к обзору которых я перехожу, я не склонен придавать значения необходимых основ теории вероятностей. Но сами по себе они весьма интересны.

Для математика является привлекательной задачей определить, какие бесконечные последовательности нулей и единиц следует называть «случайными». Ограничусь простейшим случаем последовательностей с частотами нулей и единиц, равными $1/2$. Хотелось бы потребовать от последовательности

$$x = (x_1, x_2, \ldots, x_n, \ldots),$$

чтобы ее конечные отрезки

$$x^n = (x_1, x_2, \ldots, x_n)$$

имели сложность

$$K(x^n) \geqslant n - C,$$

где C — некоторая (различная для разных x) константа. Но Мартин-Лёф доказал теорему.

Первая теорема Мартин-Лёфа. *Если $f(n)$ — такая вычислимая функция, что*

$$\Sigma 2^{-f(n)} = \infty, \tag{1}$$

то для любой двоичной последовательности

$$x = (x_1, x_2, \ldots, x_n, \ldots)$$

существует бесконечно много значений n, для которых

$$K(x^n) < n - f(n).$$

Условию теоремы удовлетворяет, например, функция $f(n) = l(n)$. Но если ряд

$$\Sigma 2^{-f(n)} \tag{2}$$

«конструктивно сходится» [5], то почти все (в смысле диадической меры) последовательности x обладают свойством

$$K(x^n) \geqslant n - f(n) \tag{3}$$

начиная с некоторого n. Было бы нелогично принять свойство (3) за определение случайной последовательности. Определение Мартин-Лёфа глубже. Я лишен здесь возможности его привести полностью. Случайные по Мартин-Лёфу двоичные последовательности обладают всеми «эффективно проверяемыми» [6] свойствами, которые с точки зрения обычной современной теории вероятностей имеют место «с вероятностью единица» в случае независимых испытаний, в которых $x_n = 1$ с вероятностью $^1/_2$. Для таких случайных последовательностей Мартин-Лёф доказал вторую теорему.

Вторая теорема Мартин-Лёфа. *Для случайных последовательностей нулей и единиц начиная с некоторого n выполняется неравенство* (3), *если только функция f такова, что ряд* (2) *конструктивно сходится.*

Я привел эти тонкие, но довольно специальные результаты Мартин-Лёфа, чтобы показать, что здесь образовалось поле для весьма интересных математических изысканий (см. в связи с этим другие работы Мартин-Лёфа и Шнорра, например [9]).

§ 7
УСЛОВНАЯ СЛОЖНОСТЬ И КОЛИЧЕСТВО ИНФОРМАЦИИ

Сложность задания какого-либо объекта может быть облегчена тем, что уже задан какой-либо другой объект. Этот факт отражает такое определение условной сложности объекта x при заданном объекте y:

$$K_S(x \mid y) = \min_{S(n(y),\, p) = n(x)} l(p).$$

Здесь метод условных определений S есть функция двух аргументов — номера объекта y и номера p программы вычисления номера $n(x)$ при заданном y. По поводу условных сложностей можно повторить все сказанное в § 3.

Если условная сложность $K(x \mid y)$ много меньше, чем безусловная сложность $K(x)$, то это естественно понять как указание на то, что в объекте y содержится некоторая «информация» об

---

[5] Подробности см. в [8].

[6] То же.

объекте x. Разность

$$\mathcal{I}_S(x \mid y) = K_S(x) - K_S(x \mid y)$$

и естественно признать за количественную меру информации об x, содержащейся в y.

Будем допускать в качестве второго аргумента функции $S(n, p)$ число нуль и положим

$$S(n, 0) = n$$

(нулевая программа из n производит n). Тогда

$$K_S(x \mid x) = 0, \qquad \mathcal{I}_S(x \mid x) = K_S(x).$$

Таким образом, сложность $K_S(x)$ можно называть и информацией, содержащейся в объекте о себе самом.

Наше определение количества информации в прикладном отношении имеет то преимущество, что оно относится к индивидуальным объектам, а не к объектам, рассматриваемым в качестве включенных в множество объектов с заданным на нем распределением вероятностей. Вероятностное определение убедительным образом можно применять к информации, содержащейся, например, в потоке поздравительных телеграмм. Но было не слишком понятно, как его применить, например, к оценке количества информации, содержащейся в романе или в переводе романа на другой язык относительно подлинника. Я думаю, что новое определение способно внести в подобные применения теории хотя бы принципиальную ясность.

Возникает вопрос о том, позволяет ли новое определение доказать ряд основных положений теории информации, достаточно себя зарекомендовавших. Заранее ясно, что они должны иметь место лишь с точностью до аддитивных констант, соответствующих неопределенности, возникавшей уже в § 3. Нельзя было ожидать, например, точного соблюдения равенства

$$\mathcal{I}(x \mid y) = \mathcal{I}(y \mid x), \tag{1}$$

но сначала казалось, что разность между левой и правой частями здесь должна быть ограничена. В действительности Колмогоров и Левин установили только более слабое неравенство типа

$$|\mathcal{I}(x \mid y) - \mathcal{I}(y \mid x)| = O(\log K(x, y)) \tag{2}$$

(см. [16]). При этом установлено, что разность в самом деле может иметь этот порядок.

Но в применениях, доступных вероятностному подходу, (2) вполне заменяет (1). В самом деле, строгое равенство (1) вероятностной теории информации позволяет делать реальные выводы лишь в применении к большому числу пар (x_i, y_i), т. е. по существу об информации в

$$(x_1, x_2, \ldots, x_r)$$

относительно

$$(y_1, y_2, \ldots, y_r)$$

и наоборот. А такого рода выводы получаются и из (2), где в этом случае выражение в правой части оказывается пренебрежимо малым.

§ 8
ТЕОРЕМА БАРЗДИНЯ

Новый ряд понятий оказывается интересным и за пределами теории вероятностей и прикладной теории информации. Чтобы дать тому пример, изложу одну теорему Барздиня. Она относится к бесконечным двоичным последовательностям

$$x = (x_1, x_2, \ldots, x_n, \ldots),$$

в которых множество номеров n с $x_n = 1$ перечислимо. Если бы было перечислимо и дополнительное множество номеров нулей, то функция $f(n) = x_n$ была бы вычислима и условная сложность $K(x^n \mid n)$ была бы ограничена. Но в общем случае (при перечислимости множества номеров единиц) $K(x^n \mid n)$ может неограниченно возрастать.

Теорема Барздиня [15]. *Для двоичной последовательности с перечислимым множеством единиц M*

$$K(x^n \mid n) \leqslant \log n + C_M;$$

существуют такие последовательности, для которых при любом n

$$K(x^n \mid n) \geqslant \log n.$$

Этот результат представляется мне имеющим принципиальный интерес с точки зрения исследований по основаниям математики. Для определенности рассмотрим такую проблему. Занумеруем все диофантовы уравнения натуральными числами. Недавно было доказано Матиясевичем, что не существует общего алгоритма для решения вопроса о том, имеет ли уравнение D_n решения в целых числах. Но можно поставить вопрос о существовании алгоритма, который позволял бы решить вопрос о существовании или несуществовании решений для первых n диофантовых уравнений при помощи некоторой дополнительной информации $\mathcal{I}_n$ при том или ином порядке роста количества этой информации с ростом n. Теорема Барздиня показывает, что этот рост может быть очень медленным:

$$\log n + C.$$

§ 9
ЗАКЛЮЧЕНИЕ

Доклад неизбежно был крайне неполным. Подробную библиографию примыкающих работ можно найти в [16]. Повторю некоторые выводы.

1. Теория информации должна предшествовать теории вероятностей, а не опираться на нее. Основы теории информации имеют по самому существу этой дисциплины финитный комбинаторный характер.

2. Применения теории вероятностей могут получить единообразное обоснование. Дело идет всегда о следствиях гипотез о невозможности теми или иными указанными средствами сократить сложность описания изучаемых объектов. Естественно, что такой подход к делу не мешает тому, чтобы теория вероятностей как часть математики развивалась как специализация общей теории меры.

3. Понятия теории информации в применении к бесконечным последовательностям дают повод к весьма интересным исследованиям, которые, не будучи необходимы в качестве основ теории вероятностей, могут получить некоторое значение в исследовании алгоритмической стороны математики в целом.

Московский государственный университет
им. М. В. Ломоносова
20 августа 1982 г.

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* Три подхода к определению понятия «количество информации».— Проблемы передачи информации, 1965, т. 1, № 1, с. 3—11.
2. *Kolmogoroff A. N.* Logical basis for information theory and probability theory.— IEEE Trans. Inform. Theory, 1968, vol. IT-14, p. 662—664. Рус. пер.: *Колмогоров А. Н.* К логическим основам теории информации и теории вероятностей.— Проблемы передачи информации, 1969, т. 5, № 3, с. 3—7.
3. *Solomonoff R. L.* A formal theory of inductive inference.— Inform. and Contr., 1964, vol. 7, N 1, p. 1—22.
4. *Chaitin G.* On the length of programs for computing finite binary sequences.— J. Assoc. Comput. Mach., 1966, vol. 13, N 4, p. 547—569.
5. *Loveland D.* A new interpretation of the von Mises concept of random sequence.— Ztschr. math. Log. und Grundl. Math., 1966, Bd. 12, S. 279—294.
6. *Church A.* On the concept of a random sequence.— Bull. Amer. Math. Soc., 1940, vol. 46, p. 130—135.
7. *Martin-Löf P.* The definition of random sequences.— Inform. and Contr., 1966, vol. 9, N 6, p. 602—619.
8. *Martin-Löf P.* Algorithms and random sequences. Erlangen: Univ. Press, 1966.
9. *Schnorr C. P.* Ein Bemerkung zum Bergiff der zufälligen Folge.— Ztschr. Wahrscheinlichkeitstheor. verw. Geb., 1969, Bd. 14, S. 27—35.
10. *Трахтенброт Б. А.* Сложность алгоритмов и вычислений. Новосибирск: Наука, 1967.

11. *Колмогоров А. Н., Успенский В. А.* К определению алгоритма.— УМН, 1958, т. 13, вып. 4, с. 3—28.
12. *Барздинь Я. М.* Проблемы универсальности в теории растущих автоматов.— ДАН СССР, 1964, т. 157, № 3, с. 542—545.
13. *Офман Ю. П.* Универсальный автомат.— Тр. Моск. мат. о-ва, 1965, т. 14, с. 186—199.
14. *Офман Ю. П.* Моделирование самоконструирующейся системы на универсальном автомате.— Проблемы передачи информации, 1966, т. 2, № 1, с. 68—73.
15. *Барздинь Я. М.* Сложность программ, распознающих принадлежность натуральных чисел, не превышающих n, рекурсивно перечислимому множеству.— ДАН СССР, 1968, т. 182, № 6, с. 1249—1252.
16. *Звонкин А. К., Левин Л. А.* Сложность объектов и обоснование понятий информации и случайности с помощью теории алгоритмов.— УМН, 1970, т. 25, вып. 6, с. 85—127.
17. *Kolmogorov A. N.* On tables of random numbers.— Sankhya. A, 1963, vol. 25, N 4, p. 369—376. Рус. пер.: *Колмогоров А. Н.* О таблицах случайных чисел.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1982, вып. 18, с. 3—13.

КОММЕНТАРИИ И ПРИЛОЖЕНИЯ

К РАБОТАМ ПО ТЕОРИИ ИНФОРМАЦИИ И НЕКОТОРЫМ ЕЕ ПРИМЕНЕНИЯМ

А. Н. Колмогоров

Возникновение теории информации как самостоятельной науки связано с идеями кибернетики. Кибернетика изучает в самом общем виде процессы управления, а рациональное управление требует определенной информации, ее получения, хранения, передачи по каналам связи и переработки. Поэтому естественно, что цикл моих работ по теории информации был создан под большим влиянием публикаций Норберта Винера и Клода Шеннона (1948) в конце 50-х и 60-х годов [1].

Значительную часть этих работ удобно обозреть в порядке существующих к настоящему времени трех подходов к введению основных понятий теории информации:

I — чисто комбинаторный подход,
II — вероятностный подход,
III — алгоритмический подход.

Математический аппарат теории информации имеет и применения, выходящие за пределы этих трех интерпретаций. Таковы работы, в которых этот аппарат применяется в теории динамических систем. Из моих работ сюда относится работа № 5 (см. ниже раздел IV).

I. ЧИСТО КОМБИНАТОРНЫЙ ПОДХОД. РАБОТЫ № 7—10, 12, 13

При комбинаторном подходе количество информации, сообщаемой при указании определенного элемента в множестве из N объектов, принимается равным двоичному логарифму N (Р. Хартли, 1928). Например, имеется

$$C(m_1, \ldots, m_s) = \frac{n!}{m_1! \ldots m_s!}$$

различных слов в алфавите из s элементов, содержащих по m_i вхождений i-й буквы нашего алфавита $(m_1 + \ldots + m_s = n)$.

---

[1] Норберт Винер с большой широтой признал мой приоритет в спектральной теории стационарных процессов и их экстраполяции. Что же касается общих идей кибернетики, то в многочисленных публикациях я занимался лишь пропагандой этих идей. В задачу же настоящего издания входит публикация лишь работ, составляющих мой оригинальный вклад в математическую теорию информации.

Поэтому интересующее нас количество информации равно

$$H = \log_2 C(m_1, \ldots, m_s).$$

При $n, m_1, \ldots, m_s$, стремящихся к бесконечности, действует асимптотическая формула

$$H \sim n\left(\sum_{i=1}^{s} \frac{m_i}{n} \log_2 \frac{m_i}{n}\right). \tag{1}$$

Читатель, вероятно, уже заметил сходство этой формулы с формулой вероятностной теории информации

$$H = n\left(\sum_{i=1}^{s} p_i \log_2 p_i\right). \tag{2}$$

Если наше слово образовано по хорошо известной схеме при помощи независимых испытаний, то асимптотическая формула (1) является очевидным следствием формулы (2) и закона больших чисел, но круг практических применений формулы (1) значительно шире (см., например, работы по передаче информации по нестационарным каналам). Вообще мне представляется важной задача освобождения всюду, где это возможно, от излишних вероятностных допущений. На независимой ценности чисто комбинаторного подхода к задачам теории информации я неоднократно настаивал в своих лекциях.

На чисто комбинаторном подходе к понятию энтропии основаны мои работы и работы моих сотрудников по ε-энтропии и ε-емкости компактных классов функций. Здесь ε-энтропия $H_\varepsilon(K)$ есть количество информации, необходимое для выделения из класса функций какой-нибудь индивидуальной функции, а ε-емкость $C_\varepsilon(K)$ есть количество информации, которое может быть закодировано элементами из K при условии, что надежно различимы элементы K, расположенные друг от друга на расстоянии, не меньшем ε (по поводу относящихся сюда работ № 7 и № 8 см. комментарий В. М. Тихомирова). В конце работы № 8 указаны без доказательства некоторые мои результаты о сложности приближенного вычисления функций из разных функциональных классов. В работе Е. А. Асарина даны полные доказательства теорем моего стокгольмского доклада (№ 8 наст. изд.) и некоторые их усиления и обобщения. Интересно отметить неожиданный результат, что оценку сложности вычисления аналитических функций удается существенно понизить, пользуясь алгоритмами быстрого умножения.

В работе № 9 на уровне элементарных подсчетов в рамках чисто комбинаторного подхода к понятию информации анализируется возможность формального определения понятия случайности. Естественно, что при этом не удается определение какого-либо абсолютного понятия случайности, а удается только определение случайности по отношению к ограниченному классу законов выбора подпоследовательности и мере ε устойчивости частот (замечание 2 § 2 этой работы выходит за пределы строго финитного подхода и освещено в комментарии А. Х. Шеня к работам № 10, 12, 13).

II. ВЕРОЯТНОСТНЫЙ ПОДХОД. РАБОТЫ № 2—4

Исчерпывающий комментарий к этим работам дан Р. Л. Добрушиным.

III. АЛГОРИТМИЧЕСКИЙ ПОДХОД. РАБОТЫ № 10, 12, 13

Алгоритмический подход основан на применении теории алгоритмов для определения понятия энтропии, или сложности, конечного объекта и понятия информации в одном конечном объекте о другом. Интуитивное различие между «простыми» и «сложными» объектами ощущалось, по-видимому, давно. На пути его формализации встает очевидная трудность: то, что просто описывается на одном языке, может не иметь простого описания на другом, и непонятно, какой способ описания выбрать. Основное открытие, сделанное мной и одновременно Р. Соломоновым, состоит в том, что с помощью теории алгоритмов можно ограничить этот произвол, определив сложность почти инвариантно (замена одного способа описания на другой приводит лишь к добавлению ограниченного слагаемого) [2]. Среди моих работ наиболее полное изложение этих идей содержится в работе № 13 (см. также комментарий А. Х. Шеня к работам № 10, 12, 13).

IV

Работа № 5 содержит некоторые применения математического аппарата теории информации к теории динамических систем. Вводимое в этой работе понятие «количества информации» не получает какого-либо реального истолкования. Оно служит лишь математическим аппаратом теории динамических систем (см. комментарий к этой работе Я. Г. Синая).

---

[2] Насколько я знаю, первой печатной публикацией, содержащей перестройки теории информации на алгоритмической основе, является статья Р. Соломонова, опубликованная в 1964 г. Я пришел к аналогичной концепции в 1963—1964 гг., еще не зная работ Соломонова, и опубликовал первую заметку № 10 на эту тему в начале 1965 г.

ТЕОРИЯ ИНФОРМАЦИИ
(к № 2—4)

(Р. Л. Добрушин)

Есть одна черта, которую цикл работ А. Н. Колмогорова по теории информации разделяет со всеми по-настоящему важными научными исследованиями. По прошествии достаточного времени их основные мысли воспринимаются специалистами как прописные истины, но совсем не так обстояло дело в период их создания. Надо вспомнить обстановку 50-х годов, когда А. Н. Колмогоров заинтересовался вопросами теории информации. Это было время, когда под давлением надвигавшейся эры вычислительных машин некритически подозрительное отношение к идеям кибернетики внезапно сменилось столь же некритическим и восторженным пиететом, и в результате зачастую на первый план выступали наиболее поверхностные и спекулятивные из этих идей. Такая шумиха порождала и естественную негативную реакцию.

Теория информации воспринималась тогда как одна из глав кибернетики. На самом деле она была создана исключительно глубокими работами (см. [1]) американского ученого Шеннона, инспирированными конкретными нуждами техники связи, еще до того как Винер придумал термин «кибернетика». Теория информации вызвала бурный интерес как среди инженеров-связистов, так и среди представителей очень многих других наук. Однако этот бум не был столь уж оправдан. В первую очередь увлекала сама мысль о возможности математического описания количества информации, и при этом не было должного понимания того, что столь общий многообразный объект, как информация, не может допускать единого метода численного измерения, а идеи Шеннона обоснованы лишь в применении к той важной, но все же ограниченной ситуации, когда рассматриваются оптимальные методы кодирования и декодирования информации в целях ее передачи по каналам связи или ее хранения. Как писал в то время сам Шеннон (см. [2]): «В результате всего этого значение теории информации было, быть может, преувеличено и раздуто до пределов, превышающих ее реальные достижения... Сейчас теория информации, как модный опьяняющий напиток, кружит голову всем вокруг». Однако это увлечение не охватило математиков. Первые работы Шеннона были написаны на уровне физической строгости, противопоказанной познавательным возможностям среднего математика. Как писал А. Н. Колмогоров в предисловии к изданному в 1963 г. русскому переводу статей Шеннона: «Значение работ Шеннона для чистой математики не сразу было достаточно оценено. Мне вспоминается, что еще на международном съезде математиков в Амстердаме (1954) мои американские коллеги, специалисты по теории вероятностей, считали мой интерес к работам Шеннона несколько преувеличенным, так как это более техника, чем математика. Сейчас такие мнения вряд ли нуждаются в опровержении. Правда, строгое математическое «обоснование» своих идей Шеннон в сколько-нибудь трудных случаях предоставил своим продолжателям. Однако его математическая интуиция изумительно точна...».

Ситуация усугублялась еще тем, что в единственно доступном в 50-х годах первом русском переводе статьи Шеннона [3] были сделаны произвольные купюры, причем не только в связи с господствовавшим тогда представлением о принципиальной недопустимости приложений математики к гуманитарным наукам, но и просто потому, что некоторые математически наиболее интересные разделы статьи показались издателям ее перевода слишком абстрактными (см. примечание в начале статьи [4]).

Были нужны работы математического уровня, закладывающие математический фундамент теории информации. Эта задача была выполнена статьями Хинчина [5, 6], Гельфанда, Колмогорова и Яглома [7]. Была нужна популяризация идей теории информации, совмещенная с авторитетным разъяснением границ ее приложимости. Эта задача была решена известным докладом А. Н. Колмогорова [8] на сессии АН СССР, посвященной автоматизации производства. Первые пионерские работы Хинчина были посвящены доказательству основных теорем теории информации для дискретного случая. В последовавших за ними работах Гельфанда, Колмогорова и Яглома [4, 7] был рассмотрен более общий случай непрерывных систем, лишь слегка затронутый в первоначальных работах Шеннона. Впервые установленные в них общие свойства количества информации, явные формулы для информационных характеристик в гауссовском случае, формулировки теорем кодирования для воспроизведения сообщений с заданной точностью остались незыблемым элементом всех современных работ по теории информации. И, что, может быть, еще важнее, эти работы московских ученых заложили традицию изложения результатов теории информации на уровне математической строгости, которой с тех пор неизменно придерживаются как те из специалистов по теории информации, которые считают себя математиками, так и те, кто считает себя инженерами. (На уровне математической строгости написаны и более поздние, тоже очень содержательные, работы Шеннона (см. [9]).)

Часто бывает, что после вылущения математического ядра идей, возникших в недрах другой науки, открываются неожиданные связи этих идей с иными разделами математики. Так случилось и с теорией информации. А. Н. Колмогоровым были открыты очень важные и плодотворные возможности применения понятия энтропии к проблеме изоморфизма динамических систем (см. [10]) и энтропийной характеризации функциональных пространств (см. [11]). Выдвинутый А. Н. Колмогоровым тезис о том, что наряду с уже признанным вероятностным подходом к определению количества информации правомерны, а во многих ситуациях и более естественны, иные подходы: комбинаторный и алгебраический, привел к созданию нового раздела науки — алгоритмической теории информации. Эти исследования подробно характеризуются в других комментариях.

Перейдем теперь к более детальному обсуждению дальнейших исследований, связанных с развитием идей А. Н. Колмогорова по вероятностной теории информации. Список дальнейших работ, в которых используются понятия теории информации, впервые исследованные в работах Колмогорова и его соавторов, был бы слишком длинен. Поэтому мы ограничимся лишь некоторыми работами московских ученых, непосредственно инспирированными влиянием А. Н. Колмогорова. Подробное изложение результатов доклада [4] содержится в статье Гельфанда и Яглома [12]. В монографии Пинскера [13] было продолжено начатое А. Н. Колмогоровым асимптотическое исследование информационных характеристик случайных процессов. Добрушин [14] развил идею А. Н. Колмогорова об общей формулировке основной теоремы теории информации для произвольной последовательности передающих устройств и сообщений с растущей информацией. Эти исследования были продолжены Пинскером [15]. Отметим также работы китайского ученого Ху Го Дина, стажировавшегося в то время в Московском университете [16, 17] и работу безвременно погибшего талантливого математика Ерохина [18]. В работах Фитингофа [19, 20], создававшихся под непосредственным влиянием А. Н. Колмогорова, был развит новый подход к проблеме оптимального кодирования сообщений в ситуации, когда статистика сообщений неизвестна. В работе А. Н. Колмогорова [4] отмечен как открытый вопрос о том, верна ли для произвольной пары стационарно связанных гауссовских процессов естественная явная формула для среднего количества информации. В работе [21] было обнаружено, что это может быть не так, если оба процесса сингулярны. Исследования этой трудной проблемы продолжаются и в последние годы. Так Солевым (см. [22]) найдены весьма общие условия верности этой формулы для процессов с непрерывным временем.

ЛИТЕРАТУРА

1. *Shannon C. E.* A mathematical theory of communication. Pt I, II.— Bell Syst. Techn. J., 1948, vol. 27, N 3, p. 379—423; N 4, p. 623—656. Рус. пер.: *Шеннон К.* Математическая теория связи.— В кн.: Работы по теории информации и кибернетике. М.: Изд-во иностр. лит., 1963, с. 243—332.
2. *Shannon C. E.* The bandwagon.— IRE Trans. Inform. Theory, 1956, vol. IT-2, N 1, p. 3. Рус. пер.: *Шеннон К.* Бандвагон.— В кн.: Работы по теории информации и кибернетике. М.: Изд-во иностр. лит., 1963, с. 667—668.
3. *Шеннон К.* Математическая теория связи.— В кн.: Теория передачи электрических сигналов при наличии помех. М.: Изд-во иностр. лит., 1953.
4. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* Количество информации и энтропия для непрерывных распределений.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1958, т. 3, с. 300—320.
5. *Хинчин А. Я.* 1. Понятие энтропии в теории вероятностей.— УМН, 1953, т. 8, вып. 3, с. 3—20.
6. *Хинчин А. Я.* 2. Об основных теоремах теории информации.— УМН, 1956, т. 11, вып. 1, с. 17—75.
7. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* К общему определению количества информации.— ДАН СССР, 1956, т. 111, № 4, с. 745—748.
8. *Колмогоров А. Н.* Теория передачи информации.— В кн.: Сессия АН СССР по научным проблемам автоматизации производства; 15—20 окт. 1956 г.: Пленар. заседания. М.: Изд-во АН СССР, 1957, с. 66—69.
9. *Шеннон К.* Работы по теории информации и кибернетике. М.: Изд-во иностр. лит., 1963.
10. *Колмогоров А. Н.* Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега.— ДАН СССР, 1958, т. 119, № 5, с. 861—864.

11. *Колмогоров А. Н.* Об энтропии на единицу времени как метрическом инварианте автоморфизмов.— ДАН СССР, 1959, т. 124, № 4, с. 754—755.
12. *Гельфанд С. И., Яглом А. М.* О вычислении количества информации случайной функции, содержащейся в другой такой функции.— УМН, 1957, т. 12, вып. 1, с. 3—52.
13. *Пинскер М. С.* Информация и информационная устойчивость случайных величин и процессов. М.: Изд-во АН СССР, 1960.
14. *Добрушин Р. Л.* Общая формулировка основной теоремы Шеннона в теории информации.— УМН, 1959, т. 14, вып. 6, с. 3—104.
15. *Пинскер М. С.* Источники сообщений.— Проблемы передачи информации, 1963, вып. 4, с. 5—20.
16. *Ху Го Дин.* 1. Три обратные теоремы к теореме Шеннона в теории информации.— Acta math. Sinica, 1961, vol. 11, N 3, p. 260—294 (на кит. яз.).
17. *Ху Го Дин.* 2. Об информационной устойчивости последовательности каналов.— Теория вероятностей и ее применения, 1962, т. 7, № 3, с. 271—282.
18. *Ерохин В. Д.* ε-энтропия дискретного случайного объекта.— Теория вероятностей и ее применения, 1958, т. 3, № 1, с. 103—107.
19. *Фитингоф Б. М.* Оптимальное кодирование при неизвестной и меняющейся статистике сообщений.— Проблемы передачи информации, 1966, т. 2, № 2, с. 3—11.
20. *Фитингоф Б. М.* Сжатие дискретной информации.— Проблемы передачи информации, 1967, т. 3, № 3, с. 26—36.
21. *Grood I. J., Doog K. C.* A paradox concerning rate of information.— Inform. and Contr., 1958, vol. 1, N 2, p. 113—126 (поправка: Inform. and Contr., 1959, vol. 2, p. 195—197).
22. *Солев В. Н.* О среднем на единицу времени количестве информации, содержащейся в одном гауссовском стационарном процессе относительно другого.— Зап. науч. семинаров ЛОМИ, 1972, т. 29, с. 18—26.

АЛГОРИТМИЧЕСКАЯ ТЕОРИЯ ИНФОРМАЦИИ (к № 10, 12, 13)

(А. Х. Шень)

В этом комментарии мы перечислим основные результаты, полученные А. Н. Колмогоровым и его учениками и последователями в области алгоритмической теории информации.

Напомним, что через $K(x)$ обозначается простая колмогоровская энтропия (=сложность) объекта x. Мы предпочитаем говорить «энтропия», как в [1], а не «сложность», как в [2], чтобы избежать смешения со «сложностями вычисления» — временнóй, емкостной и т. п. Через $K(y \mid x)$ обозначается условная колмогоровская энтропия объекта y при известном объекте x. (Говоря об «объектах», мы — если это не оговорено особо — имеем в виду натуральные числа или какие-то конструктивные объекты, вычислимым образом закодированные натуральными числами, например двоичные слова.) Понятие условной энтропии может, как показано в [2], служить основой для определения понятия «количество информации».

Рассмотрим, как связан этот «алгоритмический подход» к понятиям энтропии и информации с двумя другими подходами, изложенными в [2]: комбинаторным и вероятностным. (Отметим сразу же, что комбинаторный подход был впоследствии развит в работах Гоппы [3, 4].)

Комбинаторный подход. Нетрудно показать, что количество объектов, (простая колмогоровская) энтропия которых не превосходит n, примерно равно 2^n (с точностью до ограниченного и отделенного от нуля множителя). С другой стороны, если просто устроенное множество состоит из M объектов, то энтропия его элементов не намного больше $\log_2 M$. Один из вариантов уточнения этого высказывания таков: пусть A_0, A_1, ...— вычислимая последовательность перечислимых множеств (вычислимость понимается в том смысле, что по i можно указать программу алгоритма, область определения которого есть A_i), причем число элементов в A_i не превосходит 2^i. Тогда существует такое C, что для любого i и для любого $x \in A_i$ выполнено неравенство $K(x) \leqslant i + C$. (Это утверждение легко вытекает из результатов [5].) Грубо говоря, для установления неравенства $K(x) \leqslant i$ достаточно включить x в (просто устроенное) множество из не более чем 2^i элементов.

Вероятностный подход. Смысл приводимых ниже результатов о связи алгоритмического подхода с вероятностным можно коротко сформулировать так: при алгоритмическом подходе учитываются не только все вероятностные закономерности, но и другие (если они есть); если же других закономерностей нет, то алгоритмический подход приводит к тем же результатам, что и вероятностный.

Пусть x — последовательность нулей и единиц длины n, частота единиц в которой равна p, а нулей — $q = 1 - p$. Пусть H — соответствующая шенноновская энтропия, т. е.

$$H = -p \log_2 p - q \log_2 q.$$

Тогда

$$K(x) \leqslant nH + C \log_2 n.$$

(Здесь K — простая колмогоровская энтропия, C — константа, зависящая только от выбора p, q и способа описания и не зависящая от n, см. [5, теорема 5.1].) Таким образом, «удельная колмогоровская энтропия» $K(x)/n$ не может сильно превосходить H (но может быть значительно меньше H, если в последовательности x имеются закономерности, отличные от вероятностных). Следующий результат показывает, что если последовательность имеет только вероятностные закономерности, то ее удельная колмогоровская энтропия близка к шенноновской. Пусть p, $q \in [0, 1]$, $p + q = 1$. Рассмотрим на множестве бесконечных последовательностей нулей и единиц бернуллиевскую меру с вероятностями появления нуля и единицы, равными p и q соответственно. Тогда для почти всех (по этой мере) последовательностей $x = x_0 x_1 \ldots$ справедливо соотношение

$$\lim (K(x_0 x_1 \ldots x_{n-1})/n) = H,$$

где $H = -p \log_2 p - q \log_2 q$. (Это утверждение легко вытекает из [5, формула (5.18)].)

Наряду с простой и условной колмогоровскими энтропиями различными авторами были введены другие алгоритмические варианты энтропии конечных объектов — энтропия разрешения (см. [5]), монотонная и префиксная энтропии (см. [6]).

Различные виды энтропии оказываются связанными с так называемой априорной вероятностью (см. [5, 6]) и с введенным Шнорром понятием сложности вычислимой функции как логарифма ее номера в «оптимальной нумерации» (см. [7—9]).

В 1966 г. Мартин-Лёф [10] предложил определение случайности бесконечной последовательности нулей и единиц относительно данной вычислимой меры P на пространстве Ω всех бесконечных последовательностей нулей и единиц. Приведем это определение.

Пусть P — мера на пространстве Ω бесконечных последовательностей нулей и единиц, обладающая свойством вычислимости, т. е. такая, что мера множества

$$\Gamma_x = \{\omega \in \Omega \mid x \text{ есть начало } \omega\}$$

для любого слова x в алфавите $\{0, 1\}$ является вычислимым действительным числом, которое можно эффективно найти по x. Определим понятие конструктивно нулевого относительно меры P множества — понятие, являющееся эффективизацией обычного понятия нулевой меры. Назовем множество $A \subset \subset \Omega$ конструктивно нулевым относительно меры P, если по всякому рациональному $\varepsilon > 0$ можно эффективно указать перечислимое множество конечных слов $\{x_0, x_1, \ldots\}$, для которого

$$\text{(а) } A \subset \bigcup \Gamma_{x_i}, \qquad \text{(б) } \Sigma P(\Gamma_{x_i}) \leqslant \varepsilon.$$

(Имеется в виду существование алгоритма, указывающего по ε номер соответствующего перечислимого множества в стандартной нумерации.) Можно доказать (теорема Мартин-Лёфа), что для вычислимых мер существует максимальное по включению конструктивно нулевое (относительно данной меры) множество. Другими словами, объединение всех конструктивно нулевых (относительно вычислимой меры P) множеств снова оказывается конструктивно нулевым относительно меры P. Доказательство этого факта см. в [5] и [6].

Случайными относительно вычислимой меры P называются согласно Мартин-Лёфу те последовательности, которые не входят в максимальное конструктивно нулевое множество (или, что то же, не входят ни в одно конструктивно нулевое множество). Таким образом, для любого свойства A последовательностей нулей и единиц условия:

(а) все случайные относительно меры P последовательности обладают свойством A;

(б) множество последовательностей, не обладающих свойством A, является конструктивно нулевым относительно меры P

— равносильны.

Обычные законы теории вероятностей (типа закона больших чисел или закона повторного логарифма) приводят к множествам меры нуль, которые являются также и конструктивно нулевыми. Это дает основание говорить, что случайные по Мартин-Лёфу последовательности «обладают всеми конструктивными свойствами, которые в современной теории вероятностей доказываются... „с вероятностью единица“» [1].

Отметим, что впервые предложение классифицировать бесконечные последовательности нулей и единиц на «случайные» и «неслучайные» было выдвинуто фон Мизесом (см. [11, 12]). Схема Мизеса впоследствии уточнялась и модифицировалась разными авторами (см. [13, 14]); одна из таких модификаций предложена в [15, § 2, замеч. 2] (позже она была предложена также в [16, 17]). Об истории «частотного подхода к случайности», возникшего из идей Мизеса, см. подробно в [9, 18]; впрочем этот подход имеет скорее исторический интерес.

Как указано в [1, с. 5], естественно стремиться получить критерий случайности последовательности в терминах энтропий ее начальных отрезков. (О трудностях, встретившихся при этом, Колмогоров пишет в [1] в связи с определением понятия бернуллиевской последовательности.) Эти трудности были преодолены в 1973 г. независимо Левиным и Шнорром (см. [19, 8]), указавшими такой критерий в терминах монотонной энтропии. Этот критерий состоит в следующем: последовательность ω случайна относительно вычислимой меры P тогда и только тогда, когда существует такая константа C, что монотонная энтропия $KM(x)$ любого ее начального отрезка x не меньше $-\log_2 P(\Gamma_x) - C$. Можно доказать, что $KM(x)$ не может сильно превосходить $-\log_2 P(\Gamma_x)$: существует такое C, что для любого слова x справедливо неравенство

$$KM(x) \leqslant -\log_2 P(\Gamma_x) + C;$$

таким образом, случайными оказываются те и только те последовательности ω, у которых

$$KM(x) = -\log_2 P(\Gamma_x) + O(1)$$

для всех x, являющихся началами ω. Доказательства указанных фактов см. в [6]. В случае равномерной бернуллиевской меры, когда $P(\Gamma_x) = 2^{-l(x)}$, где $l(x)$ — длина x, случайными оказываются те последовательности, у которых монотонная энтропия начального отрезка длины n равна $n + O(1)$.

В заключение укажем, где можно найти доказательства некоторых результатов, сформулированных в комментируемых статьях. Доказательства формул (3) и (4) из [1] опубликованы в [5]. На с. 6 в [1] сформулированы утверждения о том, что существуют бернуллиевские по Чёрчу последовательности, сложность начальных отрезков которых растет логарифмически, а также более сильное утверждение о том, что существуют бернуллиевские по Ловеланду последовательности, сложность начальных отрезков которых также растет логарифмически. Доказательство первого утверждения можно найти в [4], доказательство второго не опубликовано. Отметим, что в [9] и [18] бернуллиевские по Чёрчу последовательности называются случайными по Мизесу—Чёрчу, а бернуллиевские по Ловеланду последовательности называются случайными по Мизесу—Колмогорову—Лавлэнду.

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* К логическим основам теории информации и теории вероятностей.— Проблемы передачи информации, 1969, т. 5, № 3, с. 3—7.
2. *Колмогоров А. Н.* Три подхода к определению понятия «количество информации».— Проблемы передачи информации, 1965, т. 1, № 1, с. 3—11.
3. *Гоппа В. Д.* Невероятностная взаимная информация без памяти.— Проблемы управления и теории информации, 1975, т. 4, № 2, с. 97—102.
4. *Гоппа В. Д.* Информация слов (начальное приближение — информация без памяти).— Проблемы передачи информации, 1978, т. 14, № 3, с. 3—17.
5. *Звонкин А. К., Левин Л. А.* Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов.— УМН, 1970, т. 25, вып. 6, с. 85—127.
6. *Вьюгин В. В.* Алгоритмическая энтропия (сложность) конечных объектов и ее применение к определению случайности и количества информации.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1981, вып. 16, с. 14—43.
7. *Schnorr C. P.* Optimal Gödel numberings.— In: Information processing 71: Proc. IFIP Congr. 71. Lubljana, 1971, vol. 1, p. 56—58.
8. *Schnorr C. P.* Optimal enumerations and optimal Gödel numberings. — Math. Syst. Theory, 1975, vol. 8, N 2, p. 182—191.
9. *Успенский В. А., Семенов А. Л.* Теория алгоритмов: ее основные открытия и приложения.— В кн.: Алгоритмы в современной математике и ее приложениях: Матер. Междунар. симпоз. Ургенч (УзССР), 1979 г. Новосибирск: ВЦ СО АН СССР, 1982, ч. 1, с. 99—342.
10. *Martin-Löf P.* The definition of random sequence.— Inform. and Contr., 1966, vol. 9, N 6, p. 602—619.
11. *Mises R. von.* Grundlagen der Wahrscheinlichkeitsrechnung.— Math. Ztschr., 1919, Bd. 5, S. 52—99.
12. *Мизес Р.* Вероятность и статистика. М.; Л.: Госиздат, 1930. 250 с. Пер. с нем.: *Mises R. von.* Wahrscheinlichkeit, Statistik und Wahrheit. Wien: J. Springer, 1928.
13. *Church A.* On the concept of a random sequence.— Bull. Amer. Math. Soc., 1940, vol. 46, N 2, p. 130—135.
14. *Daley R. P.* Minimal-program complexity of pseudorecursive and pseudorandom sequences.— Math. Syst. Theory, 1975, vol. 9, N 1, p. 83—94.
15. *Kolmogorov A. N.* On tables of random numbers.— Sankhya. A, 1963, vol. 25, N 4, p. 369—376. Рус. пер.: *Колмогоров А. Н.* О таблицах случайных чисел.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1982, вып. 18, с. 3—13.
16. *Loveland D.* A new interpretation of the von Mises concept of random sequence.— Ztschr. math. Log. und Grundl. Math., 1966, Bd. 12, H. 4, S. 279—284.
17. *Loveland D.* The Kleene hierarchy classification of recursively random sequences.— Trans. Amer. Math. Soc., 1966, vol. 125, N 3, p. 497—510.
18. *Шень А.* Частотный подход к определению понятия случайной последовательности.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1982, вып. 18, с. 14—42.
19. *Левин Л. А.* О понятии случайной последовательности.— ДАН СССР, 1973, т. 212, № 3, с. 548—550.

ε-ЭНТРОПИЯ И ε-ЕМКОСТЬ (к № 7, 8)

(В. М. Тихомиров)

Непосредственно к теме данного комментария относятся работы [1—5]. Однако их естественно рассматривать в чуть более расширенном цикле, к которому следует присоединить еще работы [6—10]. Стержнем, связующим весь цикл статей [1—10], является понятие *энтропии*.

В п. 1 мы приводим несколько различных определений энтропии (возникших в 50-е и 60-е годы под влиянием работ К. Шеннона и А. Н. Колмогорова) для того, чтобы читателю легче было ориентироваться в проблематике всего цикла в целом.

Статьи [1—3] имели предварительный характер, и их содержание фактически поглотилось обзорной статьей [4], помещенной в настоящей книге (см. № 7). В докладе [5] намечена лишь программа дальнейших исследований, в которой по замыслу А. Н. Колмогорова должны воссоединиться концепции теории приближений и вычислительной математики.

Циклу работ [1—10] посвящены два обзора [11, 12], написанные к шестидесятилетнему и восьмидесятилетнему юбилеям А. Н. Колмогорова. Работы [6—8, 10] прокомментированы Р. Л. Добрушиным и Я. Г. Синаем. Работа [5] практически не получила пока развития. Поэтому в п. 2—10 мы ограничиваемся лишь комментированием работы [4], указывая на основные направления развития ее тематики.

1. О понятии энтропии. Термин «энтропия» был введен (в термодинамике) немецким ученым Р. Клаузиусом в 1865 г. Впоследствии понятие энтропии сыграло фундаментальную роль в статистической физике. Л. Больцман вскрыл суть понятия энтропии как «меры неопределенности» состояния газа.

Создавая основы теории информации, Шеннон [13] воспользовался термином «энтропия» для характеризации неопределенности источников сообщений. По сути дела простейшее шенноновское определение энтропии относится к случайному объекту $\xi = (X, \Sigma, P)$, где X включает в себя конечное число элементарных событий $X = \bigcup_{i=1}^{N} A_i$, $A_i \cap A_j = \varnothing$, $i \neq j$, $P(A_i) = p_i$. *Энтропией этого случайного объекта* (*по Шеннону*) называется величина

$$H(\xi) = \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i}.$$

Разумеется, это понятие тривиально распространяется на случай счетного числа элементарных событий. В случае, когда число элементарных событий равно N и все они равновероятны, получаем, что $H(\xi) = \log_2 N$. Поэтому число $\log_2 N$ принято называть энтропией «неслучайного объекта», состоящего из N элементов.

Следует заметить, что и до Шеннона делались попытки ввести величины подобного содержания (Р. Фишер — в статистике, Р. Хартли — в теории информации), но именно Шеннону довелось связать с введенным им понятием замечательные математические результаты, которые заложили основания нового направления — теории информации.

При переходе к непрерывным сообщениям возникает трудность: все естественные аналоги шенноновской энтропии оказываются равными бесконечности. А. Н. Колмогоров неоднократно указывал (см., например, [7]), что основным понятием, допускающим обобщение на совершенно произвольные непрерывные сообщения и сигналы, является не понятие энтропии, а понятие *количества информации* $I(\xi, \eta)$ в случайном объекте ξ относительно объекта η. В простейших случаях величина $I(\xi, \eta)$ была определена Шенноном в [13, прил. 7], а в общем случае — А. Н. Колмогоровым [6, 7; 4, доп. II]. В случае дискретного случайного объекта, описанного выше, $H(\xi) = I(\xi, \xi)$.

Отправляясь от определения количества информации, Колмогоров [7] дает определение *ε-энтропии случайного объекта*. Суть дела состоит в следующем. Пусть источник информации доставляет величину ξ и от нас требуется *закодировать полученную информацию с точностью до* ε. Это по определению означает, что по сигналу ξ нужно выбрать другой сигнал ξ' так, чтобы совместное распределение $P_{\xi\xi'}$ принадлежало бы некоторому классу распределений W_ε, определенному параметром ε. Тогда ε-энтропией источника естественно считать величину

$$H_\varepsilon(\xi) = \inf\{I(\xi, \xi') \mid P(\xi, \xi') \in W_\varepsilon\}.$$

Подход А. Н. Колмогорова к количественным характеристикам неопределенности непрерывных объектов является весьма общим. Различные частные варианты возникали во многих работах. Приведем один из них. Пусть $\xi = ((X, \rho), \Sigma, \mathsf{P})$ — случайный объект, где X есть, с одной стороны, вероятностное, а с другой — метрическое пространство. Разобьем X на N дизъюнктных измеримых частей $X = \bigcup_{i=1}^{N} A_i$, $A_i \cap A_j = \varnothing$. Величина

$$H^1_\varepsilon(\xi) = \inf \sum_{i=1}^{N} P(A_i) \log_2 1/P(A_i),$$

где нижняя грань берется по всем дизъюнктным измеримым разбиениям на множества диаметра $\leqslant \varepsilon$, является ε-энтропией в выше определенном смысле. Величина $H^1_\varepsilon(\xi)$ изучалась, например, в работе [14].

В работе [3] Колмогоров дает аналогичное только что приведенному определение ε-энтропии $\mathcal{H}_\varepsilon(C)$ *неслучайного* объекта C, являющегося подмножеством метрического пространства X. Так был назван двоичный логарифм величины $N_\varepsilon(C)$ — минимального числа элементов покрытия множества C множествами диаметра $\leqslant 2\varepsilon$.

Наряду с этой ε-энтропией (названной впоследствии абсолютной) А. Н. Колмогоров дает и другое определение — относительной ε-энтропии $\mathcal{H}_\varepsilon(C, X)$ как двоичного логарифма $N_\varepsilon(C, X)$ минимального числа элементов ε-сети для X. Эти величины подробно исследовались в [4] и затем во многих других работах.

В упомянутом уже приложении 7 работы Шеннона [13] содержалась конструкция «(ε, δ)-энтропии» $H_{\varepsilon\delta}(X)$ метрического пространства с мерой $((X, \rho), \Sigma, \mathsf{P})$, равной двоичному логарифму минимального числа элементов ε-сети во множестве, мера дополнения которого $\leqslant \delta$. Это понятие развивалось в статье [15] и других.

Для классов функций на всей вещественной прямой Колмогоров ввел понятие *ε-энтропии на единицу длины* (см. § 8 в [4]). Аналогичное понятие *(ε, δ)-энтропии на единицу длины* по сути дела было введено ранее Шенноном (в том же приложении 7 к [13]). Об этих понятиях далее в п. 8 у нас еще пойдет речь.

В работе [8] Колмогоров делает еще один шаг в расширении понятия энтропии. Он вводит понятие *энтропии динамической системы.* Динамическая система (X, S) — это пространство с мерой (X, Σ, μ), в котором задано преобразование $S: X \to X$, сохраняющее меру. Со всяким разбиением $A = \{A_1, \ldots, A_N\}$, $X = \bigcup_{i=1}^{N} A_i$, $A_i \cap A_j = \phi$, $i \neq j$, связана энтропия этого разбиения

$$H_1(A) = \sum_{i=1}^{N} \mu(A_i) \log_2 \mu^{-1}(A_i).$$

Обозначив через $\mu_{i_1, \ldots, i_r}$ величину

$$\mu(A_{i_1} \cap SA_{i_2} \cap \ldots \cap S^{r-1}A_{i_r}),$$

положим

$$H_r(A) = \sum_{\{i_1, \ldots, i_r\}} \mu_{i_1 \ldots i_r} \log \mu^{-1}_{i_1, \ldots, i_r}.$$

По известной в теории информации теореме Мак-Миллана существует предел

$$\overline{H}(A) = \lim_{r\to\infty} H_r(A)/r.$$

Энтропией динамической системы (X, S) *на единицу времени* называется число

$$\overline{H} = \sup_A \overline{H}(A).$$

О выдающейся роли введенного понятия для развития теории динамических систем см. в комментарии Я. Г. Синая.

Впоследствии был сделан еще один шаг. Пусть X — топологическое пространство и S — гомеоморфизм его. С каждым разбиением $A = \{A_1, \ldots, A_N\}$ пространства X связана его «неслучайная энтропия» $H_1(A) = \log_2 N$. Обозначим через $A_{i_1 \ldots i_r}$ множество $A_{i_1} \cap SA_{i_2} \cap \ldots \cap S^{r-1}A_{i_r}$ и через $H_r(A)$ обозначим двоичный логарифм числа таких (непустых) множеств. Снова доказывается, что существует предел

$$\overline{H}(A) = \lim_{r\to\infty} H_r(A)/r.$$

Топологической энтропией пары (X, S) *на единицу времени* называется число

$$\overline{H} = \sup_A \overline{H}(A).$$

Между всеми введенными понятиями (включая, быть может, и энтропию в статистической физике), по-видимому, существует генетическая связь, и можно ожидать появления работы, где все они будут исследоваться одновременно.

Одно из основных назначений введенных понятий состояло в различении отдельных представителей в классе тех объектов, на которых энтропии определялись. Энтропия источника определяет необходимую скорость его передачи (см. [13]), т. е. различает источники по этой скорости, энтропия динамической системы дает способ построить инвариант динамических систем (см. [8] и комментарий Я. Г. Синая), ε-энтропия функциональных классов позволяет различать линейные топологические пространства (см. [9] и далее п. 7) и т. п.

В дальнейшей части комментария обсуждаются энтропийные характеристики величины типа $\mathscr{H}_\varepsilon(C)$ и $\mathscr{H}_\varepsilon(C, X)$.

2. ε-энтропия и поперечники. Во введении к работе [4] сказано о тех причинах, которые побудили А. Н. Колмогорова сформулировать «общую программу исследования ε-энтропии и ε-емкости, интересных с точки зрения теории функций компактов в функциональных пространствах». Там в качестве побудительных стимулов были названы идеи, заложенные в работах Хаусдорфа, Понтрягина—Шнирельмана, Витушкина и Шеннона.

Однако следует сказать, что понятие энтропии формируется по той же традиционной схеме, по которой в топологии и теории приближений образуют асимптотические характеристики, называемые *поперечниками*. Поясним это.

Пусть X — нормированное пространство, C — подмножество в X, $\mathfrak{A} = \{A\}$ — совокупность аппроксимирующих множеств A. Положим

$$\mathscr{E}(C, \mathfrak{A}, X) = \inf_{A\in\mathfrak{A}} \sup_{x\in C} \inf_{y\in A} \| x - y \|.$$

Если в качестве $\mathfrak{A}$ взять $\mathscr{L}_N$ совокупность всех линейных подпространств размерности $m \leqslant N$, то величина $\mathscr{E}(C, \mathscr{L}_N, X)$ называется *колмогоровским N-поперечником* C. Эта величина (обозначаемая обычно $d_N(C, X)$) была введена Колмогоровым еще в 1936 г. Но если в качестве $\mathfrak{A}$ взять Σ_N — совокупность всех N-точечных множеств, то получится величина, которую обозначим $\varepsilon_N(C, X)$. Двоичный логарифм обратной функции к $\varepsilon_N(C, X)$ есть не что иное, как (относительная) ε-энтропия $\mathscr{H}_\varepsilon(C, X)$.

Пусть, далее, Z — некоторое множество, $\varphi: C \to Z$ — некоторое отображение («кодирование» элементов из C элементами из Z) и $\Phi = \{\varphi\}$ — некоторая совокупность «кодирований». Положим

$$\mathscr{K}(C, \Phi) = \inf_{\varphi\in\Phi} \sup_{x\in C} \{\operatorname{diam} \varphi^{-1}(\varphi(x))\}.$$

Если в качестве Φ взять совокупность $\mathscr{U}_N$ всех непрерывных отображений из C в (любые) комплексы размерности $\leqslant N$, то величина $\mathscr{K}(C, \mathscr{U}_N)$ окажется совпадающей с так называемым урысоновским N-поперечником, введенным в 1922 г. (это — исторически первая величина, названная поперечником).

Если же в качестве Φ взять совокупность Φ_N всевозможных отображений в множество $\{1, 2, \ldots, N\}$, то получится величина, которую обозначим $\varepsilon^N(C)$. Двоичный логарифм обратной функции к $\varepsilon^N(C)$ есть не что иное, как абсолютная ε-энтропия $\mathscr{H}_\varepsilon(C)$. Подробнее об этом см. в [16]. Связь поперечников и энтропии изучалась Брудным, Тиманом, Тихомировым, Митягиным и др.

3. О точных решениях задачи об ε-энтропии и ε-емкости. Вопрос о нахождении точного значения ε-энтропии или ε-емкости принадлежит к числу трудных. Многие старые и взвешенно трудные задачи геометрии допускают истолкование в терминах ε-энтропии. Приведем в качестве примера проблему Борсука (см. [17]), которую мы выразим в терминах величины N_ε.

Пусть X есть n-мерное евклидово пространство и $C \subset X$ — некоторое множество диаметра единица. Вопрос Борсука состоит в том, на какое минимальное число частей меньшего диаметра может быть разбито множество C. Иначе говоря, требуется найти величину

$$\beta(X) = \max\{N_{1-0}(C) \mid \operatorname{diam} C \leqslant 1\}.$$

Гипотеза Борсука состоит в том, что $\beta(X) \leqslant n+1$. Она в полном объеме не доказана до сих пор. Кроме примера, приведенного в § 2 статьи [4], автору этого комментария неизвестны содержательные примеры бесконечномерных компактов, для которых точно решена задача об ε-энтропии или ε-емкости. Отметим попутно, что теоремы существования наилучших ε-сетей были получены Гаркави.

4. О функциональных размерностях (определенных в § 3 статьи [4]) **решений дифференциальных уравнений** см. диссертацию Казаряна [18], где приведены ссылки на предшествующие работы (Зилезни, Казарян, Комура, Танака и др.).

5. О конечномерных задачах. Задачи, равносильные асимптотике ε-энтропии и ε-емкости в конечномерных пространствах, с очень давних времен занимают геометров. На геометрическом языке — это проблемы наилучших укладок и покрытий. Большой прогресс в оценке величины θ_n (§ 2 из [4]) был достигнут в сравнительно недавнее время (об этом см., например, монографию Роджерса [19]).

Произошел сдвиг и в оценке снизу величины τ_n (§ 2 из [4]). См. об этом [20, 21].

6. Энтропия функций конечной гладкости изучалась в работах Бабенко, Бахвалова, Бирмана—Соломяка, Борзова, Брудного, Котляра, Головкина, Двейрина, Колля, Олейника, Смоляка, Тихомирова и др. Принципиальный результат об ε-энтропии функций конечной гладкости был получен Бирманом и Соломяком для соболевских классов $\mathring{W}_p^r$, заданных на единичном кубе I^n в $\mathbf{R}^n$ и рассматриваемых в пространстве $L_q(I^n)$, ими в [22] было показано, что если $\mathring{W}_p^r$ компактно вкладывается в $L_q(I^n)$, то

$$\mathscr{H}_\varepsilon(\mathring{W}_p^r, L_q(I^n)) \asymp \varepsilon^{-n/r}$$

(метрики p и q в ответе не участвуют). Нетрудно показать, что на самом деле существует

$$\lim_{\varepsilon\to 0}\varepsilon^{n/r}\mathscr{H}_\varepsilon(\overset{\circ}{W}{}^r_p,\ L_q(I^n))=\varkappa(r,p,q,n).$$

В § 2 статьи [4] вычислена величина $\varkappa(1,\infty,\infty,1)$ (для липшицевых функций). Больше ни в одном случае задача о сильной асимптотике $\mathscr{H}_\varepsilon$ (насколько нам известно) не решена. (Правда, Тихомировым [15] было показано, что

$$\varkappa(r,p,p,1)\sim 2r\log e;$$

случай $p=2$ был разобран ранее Арнольдом, см. § 6 в [4].)

7. ε-энтропия классов аналитических функций. Пусть A_G^K есть класс функций $x(\cdot)$ на компакте K, допускающих аналитическое продолжение $\tilde{x}(\cdot)$ в область $G\supset K$ такое, что $|\tilde{x}(z)|\leqslant 1$ $\forall z\in G$. Задача об ε-энтропии и колмогоровских поперечниках класса A_G^K привлекла многих исследователей (Альпер, Бабенко, Ерохин, Левин—Тихомиров, Захарюта—Скиба, Нгуен Тхан Ван, Фишер—Миччелли, Ганелиус и др.). Наиболее общий результат принадлежит Видому [23], который доказал, что

$$\lim_{n\to\infty}(d_n(A_G^K,\ C(K)))^{1/n}=\exp(-1/c(K,G)),$$

где $c(K,G)$ — емкость Грина компакта K относительно G при самых малых допущениях относительно G и K. Из этой формулы вытекает соотношение для ε-энтропии

$$\lim_{\varepsilon\to 0}\frac{\mathscr{H}_\varepsilon(A_G^K,C(K))}{\log^2(1/\varepsilon)}=\exp\left(-\frac{1}{c(K,G)}\right).$$

Исследовались и другие классы аналитических функций, а также классы функций, в определении которых участвуют и гладкость и аналитичность (Бабенко, Тихомиров, Тайков, Двейрин, Парфенов, Олейник, Фарков, Альпер, Боянов, Хюбнер, Стегбухнер и др.).

8. Теорема Котельникова. В § 8 статьи [4] найдена ε-энтропия на единицу времени класса функций с ограниченным спектром. Понятие спектра возникло в физике. Оно играет важную роль в технике и, в частности, в теории связи. В математике функции с ограниченным спектром изучаются в гармоническом и комплексном анализе. Если функция $x(\cdot)$ имеет спектр (т. е. носитель преобразования Фурье) на $[-\sigma,\sigma]$, то имеет место формула

$$x(t)=\sum_{k\in\mathbf{Z}}x\left(\frac{k\pi}{\sigma}\right)\frac{\sin\sigma(t-k\pi/\sigma)}{\sigma(t-k\pi/\sigma)},$$

которую можно интерпретировать так: *функция с ограниченным спектром определяется своими значениями, отсчитываемыми через равные промежутки времени длины* π/σ. Теоретико-информационный смысл этого утверждения (*«число измерений, приходящихся на единицу времени, равно двойной ширине частот»*) был вскрыт Котельниковым в 1933 г. и в нашей советской литературе получил название «теоремы Котельникова». Эти же вещи пробовал осмыслить и Шеннон. Точный смысл такого рода результатам может быть придан по-разному. Одно из таких истолкований и содержится в [4, § 8].

С другой стороны, сам Колмогоров исследовал вопрос о том, какой смысл можно придать утверждению Котельникова для *случайных процессов*. Важно подчеркнуть, что случайная функция с ограниченным спектром всегда *сингулярна* и ее наблюдение не связано со стационарным притоком новой информации. Колмогоров в [7] дает формулу для ε-энтропии гауссовского процесса со спектральной плотностью, хорошо аппроксимируемой плотностью, задаваемой характеристической функцией. Отсюда приближенно вытекают формулы, выражающие идею Котельникова—Шеннона.

Следует отметить, что в приложении 7 из [13] Шеннон наметил путь получения формулы, подобной той, которая была доказана в [4, § 8], но также в применении к случайным процессам. Этот путь был связан с понятием (ε, δ)-энтропии на единицу времени, о которой говорилось выше в п. 1.

Формула Шеннона оказалась верной [24]. Этот результат был усилен Пинскером и Софманом, которыми было показано, что (ε, δ)-энтропия (по Шеннону) на единицу времени $\overline{H}_{\varepsilon\delta}(\xi)$ не зависит от δ и равна ε-энтропии процесса ξ по Колмогорову $H_\varepsilon(\xi)$ (см. п. 1). Но тогда оказывается, что

$$\lim_{\varepsilon\to 0}\frac{\overline{H}_{\varepsilon\delta}(\xi)}{\log(1/\varepsilon)} = \operatorname{mes}\{\lambda \mid f_\xi(\lambda) > 0\} = \operatorname{supp} f_\xi,$$

где f_ξ — спектральная функция процесса ξ. Но в точности это и формулировал Шеннон в качестве гипотезы в [13].

9. **ε-энтропии пространств функционалов** посвящено несколько работ. Укажем, в частности, работу Тимана [25]. Но следует сказать, что многие принципиальные вопросы об ε-энтропии функционалов конечной гладкости и аналитических функционалов еще ждут своего решения.

10. **О некоторых приложениях понятия ε-энтропии.** В терминах ε-энтропии можно дать критерий ядерности линейных топологических пространств X (Митягин): для ядерности F-пространства X необходимо и достаточно, чтобы для любого компакта K и любой уравновешенной окрестности нуля U

$$\varlimsup_{\varepsilon\to 0} \log H_\varepsilon(K, X_U)/\log(1/\varepsilon) = 0.$$

(Здесь X_U — пространство X, снабженное полунормой, порожденной U.)

С помощью ε-энтропии Колмогоров ввел топологический инвариант — аппроксимативную размерность (несколько раньше подобный инвариант ввел Пельчинский), с помощью которой он доказал неизоморфизм пространств аналитических функций от разного числа переменных.

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* Оценки минимального числа элементов ε-сетей в различных функциональных классах и их применение к вопросу о представимости функций нескольких переменных суперпозициями функций меньшего числа переменных.— УМН, 1955, т. 10, вып. 1, с. 192—194.
2. *Колмогоров А. Н.* О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных.— ДАН СССР, 1956, т. 108, № 2, с. 179—182.
3. *Колмогоров А. Н.* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств.— ДАН СССР, 1956, т. 108, № 3, с. 385—388.

4. *Колмогоров А. Н., Тихомиров В. М.* ε-энтропия и ε-емкость множеств в функциональных пространствах.— УМН, 1959, т. 14, вып. 2, с. 3—86.
5. *Колмогоров А. Н.* Различные подходы к оценке трудности приближенного задания и вычисления функций.— In: Proc. Intern. Congr. Math. Stokholm, 1963, p. 369—376.
6. *Гельфанд И. М., Колмогоров А. Н., Яглом А. М.* Количество информации и энтропия для непрерывных распределений.— В кн.: Тр. III Всесоюз. мат. съезда. М.: Изд-во АН СССР, 1958, т. 3, с. 300—320.
7. *Колмогоров А. Н.* Теория передачи информации.— В кн.: Сессия Академии наук СССР по научным проблемам автоматизации производства, 15—20 окт. 1956 г.: Пленар. заседания. М.: Изд-во АН СССР, 1957, с. 66—99.
8. *Колмогоров А. Н.* Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега.— ДАН СССР, 1958, т. 119, с. 861—864.
9. *Колмогоров А. Н.* О линейной размерности топологических векторных пространств.— ДАН СССР, 1958, т. 120, с. 239—241.
10. *Колмогоров А. Н.* Об энтропии на единицу времени как метрическом инварианте автоморфизмов.— ДАН СССР, 1959, т. 124, с. 754—755.
11. *Тихомиров В. М.* Работы А. Н. Колмогорова по ε-энтропии функциональных классов и суперпозициям функций.— УМН, 1963, т. 18, вып. 5, с. 55—92.
12. *Тихомиров В. М.* Поперечники и энтропия.— УМН, 1983, т. 38, вып. 4, с. 91—99.
13. *Shannon C. E.* A mathematical theory of communication. Pt I, II.— Bell Syst. Techn. J., 1948, vol. 27, N 3, p. 379—423; N 4, p. 623—656. Рус. пер.: *Шеннон К.* Математическая теория связи.— В кн.: Работы по теории информации и кибернетике. М.: Изд-во иностр. лит., 1963, с. 243—332.
14. *Renyi A.* On the dimention and entropy of probability distributions.— Acta math. Acad. sci. hung., 1959, vol. 10, N 1—2, p. 193—215.
15. *Posner E. C., Rodemish E. R., Rumsey H. G.* ε-entropy of stochastic processes.— Ann. Math. Statist., 1967, vol. 38, N 4, p. 1000—1020.
16. *Тихомиров В. М.* Некоторые вопросы теории приближений. М.: Изд-во МГУ, 1976.
17. *Болтянский В. Г., Гохберг И. Ц.* Разбиения фигур на меньшие части. М.: Наука, 1971.
18. *Казарян Г. Г.* Неравенства между производными и дифференциальными операторами и некоторые их приложения к уравнениям в частных производных: Докт. дис. Ереван, 1980.
19. *Роджерс К. А.* Укладки и покрытия. М.: Мир, 1968.
20. *Сидельников В. М.* О плотнейшей упаковке шаров на поверхности n-мерной евклидовой сферы и числе векторов двоичного кода с данным кодовым расстоянием.— ДАН СССР, 1973, т. 213, № 5, с. 1029—1032.
21. *Левенштейн В. М.* О максимальной плотности заполнения n-мерного евклидова пространства равными шарами.— Мат. заметки, 1975, т. 18, № 2, с. 301—311.
22. *Бирман М. Ш., Соломяк М. З.* Кусочно-полиномиальные приближения функций классов W_p^α.— Мат. сб., 1967, т. 73, № 3, с. 331—335.
23. *Widom H.* Rational approximation and n-dimentional diameter.— J. Approxim. Theory, 1972, vol. 5, N 1, p. 343—361.
24. *Софман Л. Б.* Задача Шеннона.— ДАН СССР, 1974, т. 215, № 6, с. 1313—1316.
25. *Тиман А. Ф.* О порядке роста ε-энтропии пространств действительных непрерывных функционалов, определенных на связном множестве.— УМН, 1964, т. 19, вып. 1, с. 173—177; Функцион. анализ и его приложения, 1971, т. 5, № 3, с. 104—105; ДАН СССР, 1974, т. 218, № 3, с. 505—507.

ТАБЛИЦЫ СЛУЧАЙНЫХ ЧИСЕЛ
(к № 9)

(А. Х. Шень)

1. С точки зрения классической теории вероятностей появление любой последовательности нулей и единиц при n-кратном бросании симметричной монеты имеет вероятность 2^{-n}. Тем не менее появление некоторых из них (например, последовательности из одних нулей или последовательности 010101...) представляется нам удивительным и заставляет заподозрить что-то неладное, в то время как появление других не вызывает такой реакции. Другими словами, одни последовательности кажутся нам менее «случайными», чем другие, так что можно пытаться классифицировать конечные последовательности нулей и единиц по «степени их случайности». Этой классификации и посвящена в основном комментируемая статья.

Если при рассмотрении конечных последовательностей возможно говорить лишь о большей или меньшей степени случайности, то для бесконечных последовательностей естественно пытаться провести четкую границу между случайным и неслучайным, выделив в множестве всех последовательностей некоторое подмножество и назвав его элементы случайными последовательностями. Впервые идея необходимости такого деления была высказана фон Мизесом [8, 9]. Им же была предложена и конкретная схема определения случайной последовательности.

2. Изложим эту схему, ограничиваясь для простоты случаем последовательностей из нулей и единиц.

Пусть дана некоторая последовательность $x_0x_1\ldots$ Согласно Мизесу для случайности этой последовательности необходимо прежде всего, чтобы существовала средняя частота единиц, т. е. предел

$$\lim n^{-1}(x_0 + \ldots + x_{n-1}).$$

(Это — необходимое, но не достаточное условие; ему, например, удовлетворяет последовательность 01010101..., которую вряд ли стоит считать случайной!) Кроме того, средняя частота единиц должна сохраняться, если перейти от всей последовательности к ее бесконечной подпоследовательности, полученной с помощью любого допустимого правила выбора. Мизес не дает точного определения допустимости правила, ограничиваясь несколькими примерами допустимых правил и общим указанием на то, что «нужно... иметь возможность решать вопрос о принадлежности каждого отдельного бросания к выбранной частичной последовательности независимо от результата этого бросания, т. е. при помощи некоторого правила, относящегося только к номеру данного бросания и установленного прежде, чем стал известен его исход» [9, с. 32]. (В приведенной цитате Мизес называет «бросаниями» то, что мы назвали бы членами последовательности.) Высказанное ограничение запрещает считать допустимым правило «выбрать все те члены, которые равны единице». Однако оно не запрещает выбрать те члены, которым предшествует единица (это правило рассматривается на с. 36 цитированной книги Мизеса [9]) или все те члены, номера которых являются простыми числами (там же, с. 32).

3. В 1940 г. Чёрч [11] предложил формальное определение случайности, уточняющее изложенный замысел Мизеса. Приведем это определение, дав точное определение «допустимого правила выбора».

Пусть перед нами выложен бесконечный (в одну сторону) ряд карточек, верхние стороны которых одинаковы, а на нижних написаны нули и единицы — члены последовательности. Мы переворачиваем все карточки по очереди начиная с первой и отбираем некоторые из них. Числа, написанные на отобранных карточках, образуют некоторую подпоследовательность исходной последовательности. Перед переворачиванием каждой из карточек мы объявляем, желаем ли мы включить ее в число отбираемых. При этом мы имеем право использовать свои знания о том, какие числа оказались на уже перевернутых карточках. Таким образом, при этой схеме выбора допустимое правило выбора задается некоторым множеством R конечных последовательностей нулей и единиц; применить это правило к последовательности $x_0x_1 \ldots$ означает выбрать из нее последовательность, состоящую из тех членов x_{n+1}, для которых $x_0x_1 \ldots x_n \in R$. Естественно требовать, чтобы допустимые правила выбора задавались разрешимыми множествами R. Таким образом, случайность по Мизесу—Чёрчу последовательности $x_0x_1 \ldots$ означает, что:

1) существует предел $p = \lim n^{-1} (x_0 + \ldots + x_{n-1})$,

2) для любого разрешимого множества R конечных последовательностей нулей и единиц либо подпоследовательность $y_0y_1 \ldots$, полученная из исходной выбором (с сохранением порядка) тех членов x_{n+1}, для которых $x_0x_1 \ldots \ldots x_n \in R$, конечна, либо предел $\lim k^{-1} (y_0 + \ldots + y_{k-1})$ существует и равен p.

4. В комментируемой статье А. Н. Колмогорова предложена некоторая более общая схема выбора. В терминах переворачивания карточек это обобщение можно объяснить так: разрешается переворачивать карточки в любом порядке (не обязательно по очереди). Более точно. Правило выбора задается двумя вычислимыми функциями F и G. Первая функция определяет, в каком порядке мы будем переворачивать карточки. Ее аргументами являются конечные последовательности нулей и единиц, а значениями — натуральные числа. Вначале переворачивается карточка с номером $F(\Lambda)$. (Через Λ обозначается пустая последовательность.) Обозначим написанное на ней число (0 или 1) через t_0. Затем переворачивается карточка с номером $F(t_0)$; написанное на ней число обозначим t_1. Следующей переворачивается карточка с номером $F(t_0t_1)$, написанное на ней число обозначается t_2; переворачивается карточка с номером $F(t_0t_1t_2)$ и т. д. Вторая функция G имеет аргументами конечные последовательности нулей и единиц, а значениями 0 и 1. Она определяет, будет ли включен очередной член в подпоследовательность. Именно член с номером $F(\Lambda)$ будет включен в подпоследовательность, если и только если $G(\Lambda) = 1$, член с номером $F(t_0)$ будет включен, если и только если $G(t_0) = 1$, член с номером $F(t_0t_1)$ — если и только если $G(t_0t_1) = 1$ и т. д. Порядок членов в подпоследовательности определяется порядком их выбора с помощью функции F. Если в некоторый момент описанного процесса значения функции F или функции G оказываются неопределенными, то процесс выбора обрывается и выбираемая подпоследовательность оказывается конечной. То

же самое происходит, если очередное значение функции F совпадает с одним из предыдущих.

Определение допустимого (по Колмогорову) правила выбора закончено. Тем самым согласно общей схеме Мизеса возникает определение случайности: последовательность α называется случайной, если существует предел p частоты единиц в последовательности α и любая бесконечная последовательность, полученная из α с помощью допустимого (в описанном смысле) правила выбора, также имеет предел частоты единиц, равный p.

Сказанное выше в этом п. 4 является подробным изложением замечания 2 в конце п. 2 комментируемой статьи. Единственное отличие состоит в том, что мы исключили из числа аргументов функций F и G переменные ξ_i (см. начало п. 2 комментируемой статьи); это не меняет класса допустимых правил, так как значения ξ_i суть предыдущие значения функции F и могут быть восстановлены по остальным аргументам. То же самое определение допустимого правила выбора позже (но независимо) предложил Ловеланд [4, 5]. В [5, с. 499, примеч. при кор.] Ловеланд пишет: «Подобная модификация была предложена и использована А. Н. Колмогоровым» (имеется в виду модификация определения случайности по Мизесу—Чёрчу).

Поэтому допустимые в указанном только что смысле правила выбора мы будем называть допустимыми по Колмогорову—Ловеланду, а возникающий класс случайных последовательностей — классом случайных по Мизесу—Колмогорову—Ловеланду последовательностей. Очевидно, всякая случайная по Мизесу—Колмогорову—Ловеланду последовательность случайна и по Мизесу—Чёрчу; обратное, как вытекает из результатов Ловеланда, неверно (см. об этом подробнее в [12]).

5. К сожалению, определение случайности по Мизесу—Колмогорову—Ловеланду, как и определение случайности по Мизесу—Чёрчу, не вполне отвечает интуиции случайности (см. об этом подробнее в [12, § 4, 7]). Позднее А. Н. Колмогоровым и его последователями были развиты другие, отличные от мизесовского подходы к определению понятия случайной последовательности (см. [6, 7]). Подход из [7] можно назвать теоретикомерным, а подход из [6] — сложностным. Оба эти подхода изложены в настоящем издании в комментариях к работам № 10, 12. Эти подходы приводят к одному и тому же классу случайных последовательностей; этот класс можно назвать классом случайных по Мартин-Лёфу последовательностей (в честь Мартин-Лёфа, предложившего в [7] одно из определений этого класса). Нетрудно показать что всякая случайная по Мартин-Лёфу последовательность случайна и по Мизесу—Колмогорову—Ловеланду (и тем более по Мизесу—Чёрчу). В [3, с. 6] сформулировано утверждение о том, что существуют случайные по Мизесу—Колмогорову—Ловеланду последовательности, сложность начальных отрезков которых растет логарифмически. Из этого результата вытекает, что существуют случайные по Мизесу—Колмогорову—Ловеланду последовательности, неслучайные по Мартин-Лёфу. Доказательство упомянутого результата А. Н. Колмогорова из [3] не опубликовано и неизвестно автору настоящего комментария. На этом мы заканчиваем обсуждение понятия случайной (бесконечной) последовательности, отсылая читателя за более подробным изложением к [10, 12].

6. Перейдем теперь к обсуждению понятия случайности (точнее, степени случайности) конечного объекта. (О существенности рассмотрения именно конечных объектов говорит А. Н. Колмогоров в заметке «От автора», предпосланной публикации русского перевода комментируемой статьи: «Главное отличие от Мизеса состоит в строго финитном характере всей концепции и во введении количественной оценки устойчивости частот».) Предлагаемый в комментируемой статье подход к определению случайности естественно назвать частотным: последовательность называется (n, ε, p)-случайной по отношению к данной конечной системе правил выбора, если для каждой ее подпоследовательности, выбранной согласно одному из правил рассматриваемой системы и содержащей не менее чем n элементов, средняя частота единиц уклоняется от p не более чем на ε.

Позднее [2, 3] А. Н. Колмогоровым был предложен сложностной подход к определению случайности. В [2] Колмогоров пишет: «Грубо говоря, здесь дело идет о следующем. Если конечное множество M из очень большого числа элементов N допускает определение при помощи программы длины, пренебрежимо малой по сравнению с $\log_2 N$, то почти все элементы M имеют сложность $K(x)$, близкую к $\log_2 N$. Элементы $x \in M$ этой сложности и рассматриваются как „случайные" элементы множества M». (О понятии сложности, или энтропии, конечного объекта см. в [2, 3] и комментарии к этим статьям.) Таким образом, здесь определяется случайность данной последовательности x не самой по себе, а относительно данного множества M, содержащего x. Желая сравнить этот подход с подходом из комментируемой статьи, рассмотрим в качестве M множество всех последовательностей фиксированной длины, содержащих фиксированное число единиц. Более подробно. Пусть $x_0x_1 \ldots$ — последовательность нулей и единиц длины n, содержащая k нулей. Рассмотрим множество $C(n, k)$, состоящее из всех последовательностей длины n, содержащих k нулей. Сложность (= энтропия) большинства элементов $C(n, k)$ близка к $\log_2 C_n^k$ и тем самым к $H(k/n) \cdot n$, где $H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$ — шенноновская энтропия случайной величины с двумя значениями, вероятности которых равны p и $1 - p$. Так вот последовательность $x_0x_1 \ldots$ следует считать случайной, если она попадает в это большинство, т. е. если ее энтропия близка к $H(k/n) \cdot n$. Говоря более точно, ее следует считать тем менее случайной, чем дальше ее энтропия от $H(k/n) \cdot n$. А. Н. Колмогоровым установлена связь между случайностью последовательностей в этом сложностном смысле и их случайностью в смысле комментируемой работы (1982, не опубл.).

ЛИТЕРАТУРА

1. *Kolmogorov A. N.* On tables of random numbers.— Sankhya. A, 1963, vol. 25, N 4, p. 369—376. Рус. пер.: *Колмогоров А. Н.* О таблицах случайных чисел.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1982, вып. 18, с. 3—13.
2. *Колмогоров А. Н.* Три подхода к определению понятия «количество информации».— Проблемы передачи информации, 1965, т. 1, № 1, с. 3—11.
3. *Колмогоров А. Н.* К логическим основам теории информации и теории вероятностей.— Проблемы передачи информации, 1969, т. 5, № 3, с. 3—7.

4. *Loveland D.* A new interpretation of the von Mises concept of random sequence.— Ztschr. math. Log. und Grundl. Math., 1966, Bd. 12, H. 4, S. 279—294.
5. *Loveland D.* The Kleene hierarchy classification of recursively random sequences.— Trans. Amer. Math. Soc., 1966, vol. 125, N 3, p. 497—510.
6. *Левин Л. А.* О понятии случайной последовательности.— ДАН СССР, 1973, т. 212, № 3, с. 548—550.
7. *Martin-Löf P.* The definition of random sequence.— Inform. and Contr., 1966, vol. 9, N 6, p. 602—619.
8. *Mises R. von.* Grundlagen der Wahrscheinlichkeitsrechnung.— Math. Ztschr., 1919, Bd. 5, S. 52—99.
9. *Mises R. von.* Wahrscheinlichkeit, Statistik und Wahrheit. Wien: J. Springer, 1928. Рус. пер.: *Мизес Р.* Вероятность и статистика. М.; Л.: Госиздат, 1930. 250 с.
10. *Uspensky V. A., Semenov A. L.* What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its application in mathematics.— In: Algorithms in modern mathematics and computer science: Proc. Conf. Urgench (UzSSR), 1979. Berlin etc.: Springer-Verl., 1981, p. 100—234 (Lect. Notes in Comput. Sci.; Vol. 122).
11. *Church A.* On the concept of a random sequence.— Bull. Amer. Math. Soc., 1940, vol. 46, N 2, p. 130—135.
12. *Шень А.* Частотный подход к определению понятия случайной последовательности.— В кн.: Семиотика и информатика. М.: ВИНИТИ, 1982, вып. 18, с. 14—42.

РЕАЛИЗАЦИЯ СЕТЕЙ В ТРЕХМЕРНОМ ПРОСТРАНСТВЕ (к № 11)

(Я. М. Барздинь)

В оригинале статьи стоит следующее подстрочное примечание, написанное самим Андреем Николаевичем:

«Несколько лет тому назад А. Н. Колмогоров доказал теорему 1* для сетей с ограниченным ветвлением и такое первое приближение к теореме 2: существует сеть с $n > 1$ элементами, любая реализация которой имеет диаметр больше $C\sqrt{n}/\log n$, где C — некоторая константа, не зависящая от n. Окончательные варианты всех теорем (и сама идея постановки вопроса о свойствах «почти всех» сетей) принадлежат Я. М. Барздиню».

В данном варианте (см. примеч. 1 к статье № 11) я несколько переделал его и также оставил его в виде подстрочного примечания. Но на самом деле я лишь передоказал (и несколько обобщил) теоремы, полученные Андреем Николаевичем уже ранее, так что мои заслуги здесь не очень большие. К сожалению, я не помню, что это было за мероприятие, на котором Андрей Николаевич впервые упомянул эти результаты (я сам не присутствовал). Знаю только, что там речь шла об объяснении того факта, что мозг (например, человека) устроен так, что основную его массу занимают нервные волокна (аксоны), а нейроны расположены лишь на его поверхности. Конструкция теоремы 1 как раз и подтверждает оптимальность (в смысле объема) такого расположения нервной сети.

К РАБОТЕ О ДИНАМИЧЕСКИХ СИСТЕМАХ
(к № 5)

А. Н. Колмогоров

В. А. Рохлин заметил, что теорема 2 моей работы № 5 неверна. В моей работе [1] предложено другое определение энтропии автоморфизма. Наиболее короткий путь введения энтропии произвольного автоморфизма предложен Я. Г. Синаем (см. [2]). Хочу обратить также внимание на работу А. Г. Кушниренко [3], показавшего, что энтропия любого гладкого диффеоморфизма или векторного поля гладкого компактного многообразия по отношению к абсолютно непрерывной инвариантной мере конечна.

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* Об энтропии на единицу времени как метрическом инварианте автоморфизмов.— ДАН СССР, 1959, т. 124, № 4, с. 754—755.
2. *Синай Я. Г.* О понятии энтропии динамической системы.— ДАН СССР, 1959, т. 124, № 4, с. 768—771.
3. *Кушниренко А. Г.* Оценка сверху энтропии классической динамической системы.— ДАН СССР, 1965, т. 161, № 1, с. 37—38.

ЭРГОДИЧЕСКАЯ ТЕОРИЯ
(к № 5)

(Я. Г. Синай)

Эта работа А. Н. Колмогорова сыграла выдающуюся роль в развитии эргодической теории. Во-первых, она содержала решение хорошо известной проблемы, стоявшей фактически более 25 лет, причем успех был достигнут в результате привлечения в эргодическую теорию абсолютно новых для нее идей и методов из теории информации. Во-вторых, после этой работы неожиданно открылось совершенно ранее неизвестное, чрезвычайно плодотворное направление в эргодической теории, богатое по своим глубоким интересным внутренним проблемам и по разнообразию приложений к анализу конкретных динамических систем. Нет сомнения в том, что проникновение идей эргодической теории в физику, все более широкое использование их при анализе систем со стохастическим поведением, странных аттракторов и т. п. имеют одним из своих источников именно данную работу А. Н. Колмогорова.

Основная проблема общей эргодической теории есть проблема метрической классификации, или проблема метрического изоморфизма, динамических систем. Под динамической системой ниже понимается однопараметрическая группа преобразований пространства с вероятностной мерой, сохраняющая эту меру. Две динамические системы называются изоморфными mod 0, если существует определенное почти всюду отображение фазового пространства одной системы на подмножество полной меры фазового пространства другой системы, перестановочное с действием группы (см. [6]). Проблема метрического изоморфизма динамических систем состоит в нахождении полной системы метрических инвариантов, т. е. такого набора инвариантов, что их совпадение влечет метрический изоморфизм. Первым примером метрического инварианта служит спектр группы унитарных операторов, сопряженной с динамической системой.

Интерес к проблеме метрического изоморфизма возник после работ Неймана [23] и Неймана и Халмоша [21], где было показано, что в классе эргодических динамических систем с чисто точечным спектром полная система метрических инвариантов исчерпывается спектром. Гельфанд заметил, что результат фон Неймана может быть получен как простое следствие тривиальности второй группы когомологий спектра, который всегда является счетной абелевой группой, с коэффициентами в S^1. Казалось, что для систем с непрерывным спектром надо понять, в каком смысле спектр образует группу, ввести для нее группы когомологий с коэффициентами в группе унитарных операторов, после чего проблема изоморфизма сведется к вычислению соответствующей второй группы когомологий.

В определенной степени эта идеология оказалась полезной лишь для систем с непрерывным простым или конечнократным спектром, где она связана с некоторыми проблемами теории банаховых алгебр (см. [6]). Однако уже довольно давно было понято, что основной класс динамических систем с непрерывным спектром — это класс систем с лебеговым спектром бесконечной кратности. Систем с лебеговым спектром конечной кратности не удалось, насколько нам известно, построить до сих пор. Системы же со счетно-кратным лебеговым спектром к моменту появления работы А. Н. Колмогорова были уже хорошо известны: это автоморфизмы Бернулли и родственные им примеры из теории вероятностей (см. [18]), эргодические автоморфизмы коммутативных компактных групп (см. [10]), геодезические потоки на поверхностях постоянной отрицательной кривизны (см. [5]) и ряд других. Для таких систем не удавалось найти никакого обобщения рассуждений фон Неймана.

В работе № 5 был введен новый метрический инвариант динамической системы — метрическая энтропия. Энтропия — это числовой инвариант, принимающий любые неотрицательные значения, включая ∞. Ее независимость от спектра состоит в том, что в классе динамических систем со счетнократным лебеговым спектром энтропия может принимать любые допустимые значения. Иными словами, возможны динамические системы со счетно-кратным лебеговым спектром и разными значениями энтропии, а поэтому неизоморфные.

После работы А. Н. Колмогорова проблема метрического изоморфизма для систем со счетно-кратным лебеговым спектром стала выглядеть совершенно иначе, чем для систем с дискретным спектром. В результате открытия энтропии стало ясно, что ее следует рассматривать как проблему кодирования в стиле теории информации и первые примеры такого изоморфизма были построены именно как «коды», отображающие одно пространство последовательностей — на другое (см. [7]).

Не менее важным, чем энтропия, оказалось введенное в этой же работе № 5 понятие квазирегулярной динамической системы, или, как теперь принято говорить, K-системы. В этом названии отражен и первоначальный термин и имя автора. Понятие K-системы навеяно общими размышлениями А. Н. Колмогорова о регулярности случайных процессов, т. е. о том, как происходит и чем характеризуется ослабление статистических связей между значениями процесса на удаленных друг от друга промежутках времени. С этой точки зрения K-системы соответствуют случайным процессам с самыми слабыми свойствами регулярности.

Все K-системы имеют счетно-кратный лебегов спектр (для автоморфизмов это было отмечено в самой работе № 5, для потоков было показано в [11]) и положительную энтропию. Никаких метрических инвариантов, кроме энтропии, различающих K-системы, не было известно, и проблема метрического изоморфизма в классе K-систем приобрела следующий вид: верно ли, что K-системы с одинаковым значением энтропии метрически изоморфны.

Большого успеха в решении этой проблемы добился американский математик Орнстейн. Вначале он показал, что автоморфизмы Бернулли с равной энтропией метрически изоморфны, а затем перенес этот результат на гораздо более широкий класс динамических систем (так называемые очень слабо бернуллиевские системы), включающий в себя автоморфизмы Маркова и многие классические динамические системы, такие, как эргодические автоморфизмы коммутативных компактных групп, геодезические потоки на компактных многообразиях отрицательной кривизны, некоторые бильярдные системы. С другой стороны, Орнстейн построил примеры K-систем, неизоморфных никакому автоморфизму Бернулли. Недавно был предложен красивый и довольно простой пример подобной системы (см. [22]). Тем самым оказалось, что и в классе K-систем энтропия не составляет полной системы метрических инвариантов. Орнстейн и Шилдс показали, что число неизоморфных типов K-систем с одинаковой энтропией несчетно. Теория Орнстейна изложена в его монографии [9].

В § 4 работы № 5 строятся примеры потоков, т. е. динамических систем с непрерывным временем, с конечной энтропией. Заметим, что эти примеры с точки зрения традиционной теории вероятностей довольно неестественны. В этих примерах типичные реализации таких процессов принимают конечное или счетное число значений, и поэтому кусочно-постоянны на отрезках случайной длины, а меняют свое состояние в соответствии с некоторым распределением вероятностей. Оказалось, что именно процессы типа примеров А. Н. Колмогорова возникают при анализе гладких динамических систем с сильными статистическими свойствами, например геодезических потоков на компактных многообразиях отрицательной кривизны (см. [2, 12]) и некоторых бильярдных систем (см. [13, 19]). Для этого имеются достаточно глубокие причины. В конечномерных динамических системах, динамика которых определяется строго детерминированными уравнениями движения, вся случайность заключена только в выборе начальных данных. Возникающие при этом стационарные случайные процессы неизбежно оказываются процессами «с дискретным вмешательством случая». Тем не менее для целого ряда важных классов динамических систем, которым отвечают подобные случайные процессы, удается установить ряд сильных статистических свойств, таких, как центральную предельную теорему для временны́х флуктуаций, быстрое убывание временны́х корреляционных функций и т. п. (см., например, [3, 26, 14, 20]). Основным динамическим механизмом для этого служит внутренняя неустойчивость движения (см. [15, 16]). Свойством неустойчивости обладают системы Аносова (см. [1, 8]), A-системы Смейла (см. [17]), модель Лоренца со странным аттрактором (см. [15, 4]), рассеивающие бильярды (см. [13]).

Через несколько лет после появления работы № 5 была обнаружена тесная связь теории неустойчивых динамических систем с равновесной статистической механикой классических одномерных решетчатых моделей (см. [14, 3, 26]). Чисто формально эта связь возникает, если большой параметр T — время — эргодической теории рассматривать как большой параметр числа степеней свободы в статистической механике. При этом появляется возможность прямого перенесения понятий и методов статистической механики в новую обстановку. Возникающая при этом теория называется «топологической термодинамикой». С ее помощью уже удалось получить ряд новых глубоких результатов.

ЛИТЕРАТУРА

1. *Аносов Д. В.* Геодезические потоки на замкнутых римановых многообразиях отрицательной кривизны.— Тр. МИАН СССР, 1967, т. 90, с. 3—209.
2. *Аносов Д. В., Синай Я. Г.* Некоторые гладкие динамические системы.— УМН, 1967, т. 22, вып. 5, с. 107—172.
3. *Боуэн Р.* Методы символической динамики. М.: Мир, 1979. (Математика; Вып. 13).
4. *Бунимович Л. А., Синай Я. Г.* Стохастичность аттрактора Лоренца.— В кн.: Нелинейные волны. М.: Наука, 1979.
5. *Гельфанд И. М., Фомин С. В.* Геодезические потоки на многообразиях постоянной отрицательной кривизны.— УМН, 1952, т. 47, вып. 1, с. 118—137.
6. *Корнфельд И. П., Синай Я. Г., Фомин С. В.* Эргодическая теория. М.: Наука, 1980.
7. *Мешалкин Л. Д.* Один случай изоморфизма схем Бернулли.— ДАН СССР, 1959, т. 128, № 1, с. 41—44.
8. *Нитецки З.* Введение в дифференциальную динамику. М.: Мир, 1975.
9. *Орнстейн О.* Эргодическая теория, случайность и динамические системы. М.: Мир, 1978. (Математика; Вып. 8).
10. *Рохлин В. А.* Об эндоморфизмах коммутативных компактных групп.— Изв. АН СССР. Сер. мат., 1949, т. 13, с. 323—340.
11. *Синай Я. Г.* Динамические системы со счетно-кратным лебеговским спектром.— Изв. АН СССР. Сер. мат., 1961, т. 25, № 6, с. 899—924.
12. *Синай Я. Г.* Классические динамические системы со счетно-кратным лебеговским спектром.— Изв. АН СССР. Сер. мат., 1967, т. 30, № 1, с. 15—68.
13. *Синай Я. Г.* Динамические системы с упругими отражениями. Эргодические свойства рассеивающих бильярдов.— УМН, 1970, т. 25, вып. 2, с. 141—192.
14. *Синай Я. Г.* Гиббсовские меры в эргодической теории.— УМН, 1972, т. 27, вып. 4, с. 21—64.
15. *Синай Я. Г.* Стохастичность динамических систем.— В кн.: Нелинейные волны. М.: Наука, 1979, с. 192—211.

16. *Синай Я. Г.* Случайность неслучайного.— Природа, 1981, № 3, с. 72—80.
17. *Смейл С.* Дифференцируемые динамические системы.— УМН, 1970, т. 25, вып. 1, с. 113—185.
18. *Халмош П.* Лекции по эргодической теории. М.: Изд-во иностр. лит., 1959.
19. *Bunimovich L. A.* On the ergodic properties of nowhere dispersing billiards.— Communs Math. Phys., 1979, vol. 65, N 3, p. 295—312.
20. *Bunimovich L. A., Sinai Ya. G.* Statistical properties of Lorentz gas with periodic configuration of scatterers.— Communs Math. Phys., 1981, vol. 78, N 4, p. 479—497.
21. *Halmos P., Neumann J. von.* Operator methods in classical mechanics. II.— Ann. Math., 1942, vol. 43, p. 332—338.
22. *Kalikow S. A.* T, T^{-1} transformation is not loosely Bernoulli.— Ann. Math. Ser. 2, 1982, vol. 115, N 2, p. 393—409.
23. *Neumann J. von.* Zur Operatoren methode in der klassichen Mechanik.— Ann. Math., 1932, vol. 33, p. 587—642.

АЛГОРИТМЫ, ИЛИ МАШИНЫ, КОЛМОГОРОВА (к № 1, 6)

(В. А. Успенский, А. Л. Семенов)

I. ОБЩИЙ ПОДХОД К ПОНЯТИЮ АЛГОРИТМА: ЦЕЛИ И ИДЕИ

В конце 1951 г. студент V курса механико-математического факультета МГУ В. А. Успенский получил от своего научного руководителя А. Н. Колмогорова страницу машинописного текста с описанием некоторой абстрактной машины. Состояниями этой машины служили одномерные топологические комплексы, а точнее комплексы с некоторой дополнительной информацией; эта дополнительная информация заключалась в том, что на вершинах комплекса задавалась функция со значениями из заранее выбранного конечного множества, так что комплекс оказывался размеченным (значениями этой функции). Предложенная А. Н. Колмогоровым формулировка имела две цели:

1) найти возможно более общее математическое определение понятия алгоритма (столь общее, чтобы оно непосредственно охватывало все алгоритмы);

2) убедиться, что и это более общее определение не приводит к расширению того класса вычислимых функций, который уже сформировался к этому времени на основе предшествующих, более узких определений (Тьюринга, Поста, Маркова и других) и который, если ограничиться числовыми функциями, оказался совпадающим с классом частично рекурсивных функций.

Разработке формулировки А. Н. Колмогорова была посвящена дипломная работа В. А. Успенского «Общее определение алгоритмической вычислимости и алгоритмической сводимости», выполненная в первой половине 1952 г.

Впоследствии А. Н. Колмогоров изложил свои идеи в докладе «О понятии алгоритма», сделанном 17 марта 1953 г. на заседании Московского математического общества; публикация [3] (№ 1 наст. изд.) представляет собой резюме этого доклада.

Среди выдвинутых в докладе идей следующие три являются наиболее фундаментальными:

1) указание общего строения произвольного алгоритмического процесса; см. п. 1, 2 и 3 в [3];

2) уточнение понятия конструктивного объекта — *состояния алгоритмического процесса*;

3) указание на «локальность» алгоритмического процесса; см. п. 4 в [3].

Предложенная А. Н. Колмогоровым общая схема строения произвольного алгоритма легла в основу представления о семи параметрах алгоритма, изложенного в статьях «Алгоритм» в «Большой Советской Энциклопедии» (3-е изд.) и «Математической энциклопедии» [8, 9]. Что касается так называемых конструктивных объектов, то это — те объекты, с которыми работают алгоритмы. Таковыми являются, например, конечные цепочки хорошо различимых символов, матрицы с целочисленными элементами и т. п.— в отличие от, скажем, действительных чисел или бесконечных множеств, не являющихся конструктивными объектами. Любой общий подход к понятию алгоритма неизбежно должен включать в себя и общий подход к понятию конструктивного объекта. Точка зрения А. Н. Колмогорова состояла в том, что конструктивный объект (А. Н. Колмогоров говорил о «состояниях» алгоритмического процесса) — это некоторое конечное множество элементов, связанных друг с другом некоторыми связями; и элементы, и связи могут быть нескольких различных типов, причем как число типов элементов, так и число типов связей заранее ограничено (каждое конкретное такое ограничение приводит к своему типу конструктивных объектов).

Наконец, «локальность» означает локальный характер *непосредственной переработки*, т. е. преобразования, осуществляемого на каждом отдельном шаге; это предполагает выделение в каждом конструктивном объекте заранее ограниченной *активной части* (слово «заранее» означает, что объем этой части зависит только от рассматриваемого алгоритма, но не от состояний, возникающих в процессе его работы).

Итоги более детального анализа ситуации были изложены в статье А. Н. Колмогорова и В. А. Успенского [5] [1].

II. СВЯЗЬ С РЕКУРСИВНЫМИ ФУНКЦИЯМИ

Второй из целей, упомянутых в начале настоящего комментария, в работе [5] посвящен § 3.

Поясним, какой подход при этом используется. Прежде всего нужно уточнить саму постановку задачи. Действительно, поскольку аргументами и

[1] Реферат на эту статью опубликован в реферативном журнале: Математика, 1959, № 7, с. 9 (реф. 6527). В настоящем издании учтены примечания референта (А. В. Гладкого), в частности исправлены указанные в реферате опечатки. Реферат на эту статью опубликован также в журнале: Math. Rev., 1959, vol. 20, N 9, p. 948 (реф. 5735). Англ. пер. этой статьи [5]: *Kolmogorov A. N.*, *Uspenskii V. A.* On the definition of an algorithm.— Amer. Math. Soc. Transl. Ser. 2, 1963, vol. 29, p. 217—245.

результатами алгоритмов Колмогорова являются не числа, а совсем другие конструктивные объекты, на первый взгляд неясно, как можно сравнивать, скажем, класс функций, вычислимых алгоритмами Колмогорова, и класс частично рекурсивных функций, у которых аргументами и значениями являются по определению только натуральные числа. Конечно, эта трудность возникает не только в связи с алгоритмами Колмогорова, но и при изучении других вычислительных моделей. Рассмотрим, например, семейство всех машин Тьюринга, у которых входной и выходной алфавит есть $\{a, b, c, d\}$, и попытаемся объяснить, что значит, что класс вычислимых такими машинами функций не шире класса частично рекурсивных функций. Для этого фиксируем какое-либо естественное — и потому заведомо вычислимое в обе стороны — взаимно однозначное соответствие между множеством всех слов в алфавите $\{a, b, c, d\}$ и некоторым множеством натуральных чисел. Это соответствие позволяет сопоставить со всякой функцией, перерабатывающей слова в алфавите $\{a, b, c, d\}$ в слова в том же алфавите, некоторую одноместную числовую функцию. Пусть оказалось (а на самом деле так и оказывается), что при этом сопоставлении всякой функции, вычислимой машиной Тьюринга, соответствует частично рекурсивная функция. Тогда мы можем свести задачу вычисления произвольной функции на машине Тьюринга к задаче вычисления подходящей частично рекурсивной функции. Это и означает, что класс вычислимых на машинах Тьюринга функций не шире класса частично рекурсивных функций. Такая же операция с заменой машин Тьюринга на алгоритмы Колмогорова и проделывается в § 3 из [5] (конечно, с соответствующими изменениями).

Заметим, что к достижению обсуждаемой сейчас второй цели можно подойти и иначе. Проиллюстрируем сперва этот иной подход на уже рассмотренном примере машин Тьюринга с входным и выходным алфавитом $\{a, b, c, d\}$. Следующим образом определим понятие «числовая функция от q аргументов вычисляется данной машиной Тьюринга». Закрепим какой-либо естественный способ, посредством которого произвольный «входной» числовой кортеж $\langle m_1, \ldots, m_q\rangle$ записывается на входной ленте какой-либо машины Тьюринга, а произвольное «выходное» число n на выходной ленте той же машины. (При этом мы не предполагаем, что входная и выходная ленты непременно отличаются друг от друга и от рабочей ленты — это просто те ленты, на которых записываются исходные данные и впоследствии формируются результаты.) Каждая машина Тьюринга вычисляет следующую числовую функцию f от q аргументов: кортеж $\langle m_1, \ldots, m_q\rangle$ записывается на входной ленте и машина пускается в ход; если работа машины заканчивается и на выходной ленте при этом оказывается записанным число n, то это n и объявляется значением $f(m_1, \ldots, m_q)$; во всех остальных случаях функция f считается неопределенной для аргументов $m_1, \ldots, m_q$. (Без ограничения общности можно считать, что если машина останавливается, то на выходной ленте непременно записано некоторое число.) Как хорошо известно, все числовые функции, вычислимые машинами Тьюринга, частично рекурсивны. Аналогичная схема применяется и в [5]. Именно: в § 2 указывается способ, посредством которого «входные» числовые кортежи изображаются в виде «исходных состояний» колмогоровских алгоритмов, а «выходные» числа — в виде «заключительных состояний»; тем самым возникает понятие числовой функции, вычислимой алгоритмами Колмогорова; указывается, что каждая частично рекурсивная функция вычислима некоторым алгоритмом Колмогорова; в § 3 показывается, что всякая числовая функция, вычислимая каким-либо алгоритмом Колмогорова, является частично рекурсивной.

III. ОБЩИЙ ПОДХОД К ПОНЯТИЮ КОНСТРУКТИВНОГО ОБЪЕКТА

В дальнейшем речь пойдет о первой из целей, упомянутых в начале комментария,— найти для понятия алгоритма наиболее общее математическое определение.

Для понимания замысла А. Н. Колмогорова существенно осознавать, что каждый алгоритм имеет дело лишь с конструктивными объектами того или иного фиксированного типа, например со словами, составленными из букв заданного, жестко закрепленного алфавита (а не со словами во всех алфавитах). Всевозможные конструктивные объекты заданного типа образуют так называемый *ансамбль конструктивных объектов*. Важное, несмотря на его простоту, наблюдение, сделанное А. Н. Колмогоровым, состояло в том, что сами связи (в которых участвуют образующие конструктивный объект элементы) можно рассматривать как элементы новых типов и, таким образом, свести все дело к одному-единственному типу бинарных связей и по-прежнему конечному (хотя и увеличенному) числу типов элементов. А тогда конструктивные объекты естественно изображаются графами, или одномерными топологическими комплексами, вершины которых помечены символами из некоторого конечного алфавита Б, а ребра никак не помечены. Все такие комплексы с заданным Б образуют один ансамбль.

Разметка ребер, т. е. надписывание на каждом ребре некоторого символа, не приводит в силу только что изложенного наблюдения А. Н. Колмогорова к принципиальному расширению класса рассматриваемых объектов. Действительно, символ на ребре, соединяющем вершины A и B, можно считать вершиной нового типа (т. е. вершиной с особой пометой), соединенной с вершинами A и B простыми (не размеченными) ребрами.

При общем рассмотрении связей между элементами конструктивного объекта аргументы каждой такой связи в [5] упорядочивались; однако при переходе к комплексам (когда прежние элементы и связи делались вершинами и допускалась лишь одна новая связь — смежность) концы ребер не упорядочивались (не указывалось, который из концов ребра являлся началом, а который концом), т. е. комплексы рассматривались неориентированные: как обнаруживалось в [5], нужды в ориентации не возникало. Можно, однако, рассматривать и ориентированные комплексы, в которых для каждого ребра указаны его начало и конец. Тогда у каждой вершины возникают две степени — исходящая и входящая (представляющие собой соответственно число исходящих и число входящих в эту вершину ребер). Если все эти степени ограничены некоторым числом k, возникающая совокупность ориентированных колмогоровских комплексов естественно вкладывается в подходящий ансамбль неориентированных комплексов. Другое дело, если считать ограниченными лишь исходящие степени, а степени входящие допускать сколь угодно большие. Возникает новый вид конструктивных объектов; в объектах этого вида один и тот же элемент может участвовать — в роли конца ребра — в произвольно большом числе связей. Можно ли считать, что при фиксированном алфавите разметки и фиксированном ограничении исходящих степеней все объекты нового вида принадлежат одному и тому же ансамблю? Любой ответ на этот вопрос является спорным; по-видимому, наша интуиция конструктивного объекта и ансамбля конструктивных объектов еще недостаточно прояснена. Если принять положительный ответ на заданный вопрос и считать, что ориентированные колмогоровские комплексы с фиксированным алфавитом и фиксированным ограничением исходящих степеней (но без ограничения входящих степеней!) образуют в своей совокупности ансамбль, то можно рассматривать и алгоритмы Колмогорова, работающие с членами этого ансамбля. Такие алгоритмы называются «ориентированными алгоритмами (или машинами) Колмогорова» в отличие от «неориентированных алгоритмов (или машин) Колмогорова», работающих с неориентированными комплексами.

IV. КОЛМОГОРОВСКИЕ КОМПЛЕКСЫ

Мы приведем сейчас параллельно определения двух типов конструктивных объектов — *ориентированных* и *неориентированных колмогоровских комплексов*, а затем в разделе V определения соответствующих алгоритмов. Эти определения взяты из обзоров [13] и [14].

1. Колмогоровский комплекс является *ориентированным или неориентированным графом*, в соответствии с чем колмогоровский комплекс называется *ориентированным* или *неориентированным*.

2. Колмогоровский комплекс является *инициальным графом*, т. е. в нем как-то выделена ровно одна вершина — начальная.

3. Колмогоровский комплекс является *связным графом*, т. е. в каждую его вершину можно попасть по (ориентированному — для ориентированных комплексов) пути из начальной вершины.

4. Колмогоровский комплекс является *размеченным графом*, т. е. каждая его вершина помечена буквой из какого-либо алфавита.

5. Для каждой вершины a все вершины b, такие, что $\langle a, b\rangle$ является (ориентированным — для ориентированных комплексов) ребром, помечены различными буквами.

Требование связности комплекса является основным отличием сформулированного только что определения от определения из работы [5], допускавшего и несвязные комплексы — как в качестве входов, или исходных данных, так и в качестве «промежуточных результатов», или состояний, алгоритмического процесса. Однако и в [5] окончательный результат в силу самой конструкции алгоритма мог оказаться лишь связным комплексом.

Поясним происхождение пятого условия (соответствующего свойству «а» п. 3 определения алгоритма из § 2 работы [5]). Естественно считать, что при обработке конструктивного объекта мы обозреваем лишь ограниченную его часть, находящуюся вблизи начальной вершины. В то же время мы хотели бы иметь способ попасть в любую наперед заданную вершину комплекса. Это возможно, только если любая вершина *однозначно* определяется последовательностью пометок на каком-нибудь ведущем в нее из начальной вершины пути. Двигаясь именно по этому пути, мы можем ее достичь. Отметим, что в отличие от [5] мы не требуем, чтобы метка начальной вершины отличалась от меток других вершин комплекса, а считаем, что начальная вершина выделена каким-то другим способом.

Таким образом, объекты переработки, или «состояния» (в частности, исходные данные и результаты), алгоритмов Колмогорова из работы [5] и обзоров [13, 14] различаются. Рассмотрим, однако, только такие функции, для которых аргументы и значения — связные неориентированные комплексы, размеченные в соответствии с условиями на разметку начальных вершин у исходных данных и результатов, приведенными в работе [5]. Можно показать, что для таких функций понятия вычислимости алгоритмами Колмогорова из [5] и неориентированными машинами Колмогорова из [13] совпадают.

Если вершины колмогоровского комплекса помечены буквами конечного алфавита Б, мы называем этот комплекс *колмогоровским комплексом над* Б, или Б-*комплексом*. Ориентированный (соответственно неориентированный) Б-комплекс называется (Б, k)-*комплексом*, если все исходящие степени (соответственно все степени) его вершин ограничены числом k.

Ансамбль ориентированных (соответственно *неориентированных*) (Б, k)-*комплексов* — это множество всех ориентированных (соответственно неориентированных) (Б, k)-комплексов для заданных конечного алфавита Б и натурального k.

В связи с приведенным общим определением конструктивного объекта возникает естественная с точки зрения приложений задача реализации колмогоровских комплексов в реальном физическом пространстве. Эта задача рассматривалась в [4] в следующей постановке. Фиксируем некоторый ансамбль неориентированных колмогоровских комплексов. Предположим теперь, что вершины комплекса реализуются шарами положительного (и единого для всех шаров) диаметра, а ребра — гибкими трубами положительного (и единого для всех труб) диаметра. Тогда можно так выбрать диаметры труб и шаров, что для некоторых положительных действительных чисел c и d будет выполнено следующее: 1) всякий комплекс с n вершинами может быть реализован внутри сферы радиуса $c\sqrt{n}$; 2) объем любой реализации почти всякого комплекса с n вершинами не меньше $dn\sqrt{n}$ («почти всякого» означает, что отношение количества комплексов с таким свойством к количеству всех комплексов с n вершинами стремится к единице при стремлении n к бесконечности).

V. ОРИЕНТИРОВАННЫЕ И НЕОРИЕНТИРОВАННЫЕ МАШИНЫ КОЛМОГОРОВА

Определения ориентированных и неориентированных машин Колмогорова мы дадим параллельно. В случае неориентированных машин наше определение по существу совпадает с определением алгоритма в работе [5]. Главное различие состоит в том, что в [5] множество комплексов, получаемых в результате работы алгоритмов, не пересекалось с множеством комплексов, подаваемых на вход алгоритмов. Действительно, в [5] ««входные» комплексы имели одну, «входную» метку начальной вершины, а «результирующие» комплексы — другую, «результирующую» метку. Чтобы результат работы одного алгоритма подать на вход другого, необходимо было изменить метку начальной вершины. Предлагаемое ниже определение делает ненужным такое изменение (напомним, что у нас начальная вершина не несет никакой особой, «начальной» метки). Из соображения простоты определения мы также слегка меняем предложенную в [3] (и использованную в [5]) общую схему описания всякого алгоритма. Именно потребуем, чтобы после того, как появился сигнал о получении решения — сигнал окончания, еще ровно один раз применился оператор непосредственной переработки. Наконец, как уже отмечалось, алгоритмы из [5] в отличие от определяемых ниже могли применяться и к несвязным комплексам; связность результата обеспечивалась в [5] тем, что из заключительного состояния, возникающего на последнем шаге работы алгоритма, извлекалась связная компонента начальной вершины и она-то и объявлялась «решением» (результатом, выходом) алгоритма. Приводимое ниже определение требует выделения связной компоненты начальной вершины на *каждом* шаге работы алгоритма (так что все промежуточные состояния являются колмогоровскими комплексами). Заметим в этой связи, что и в [5] непосредственная переработка затрагивала на каждом шаге лишь связную компоненту начальной вершины (поскольку именно в ней помещалась активная часть состояния).

Машина Колмогорова (как *ориентированная*, так и *неориентированная*) может быть описана теперь следующим образом. (Смысл терминов «состояние» и «активная часть» такой же, как в [3].) Работа машины Г состоит из последовательности шагов, состоящих в применении к обрабатываемому на очередном шаге объекту (т. е. состоянию) *оператора непосредственной переработки* Ω_Γ. В начальный момент работы машины обрабатываемый объект — это исходное данное. Перед выполнением каждого шага проверяется, не поступил ли *сигнал об окончании* работы («сигнал о решении»). Как только такой сигнал появляется, оператор Ω_Γ применяется еще один раз, после чего работа машины считается законченной, а образовавшееся состояние объявляется результатом работы машины («решением»).

Перейдем теперь к определению того, что является состоянием алгоритмического процесса, какой вид имеет оператор непосредственной переработки Ω_Γ и как формируется сигнал окончания.

Состояниями алгоритмического процесса являются колмогоровские комплексы из некоторого ансамбля (Б, k)-комплексов.

Активной частью состояния является подкомплекс, вершины которого достижимы из начальной посредством (ориентированных) путей длины, не большей, чем определенное число, фиксированное для данной машины.

Обозначим множество всех вершин комплекса G через $v(G)$. Оператор Ω_Γ определяется конечным множеством *команд* вида $U \to \langle W, \gamma, \delta \rangle$, где U, W — комплексы; γ — отображение из $v(U)$ в $v(W)$, сохраняющее разметку; δ — взаимно однозначное отображение из $v(U)$ в $v(W)$, причем для каждого $x \in v(U)$ множество меток тех вершин, куда идет ребро из $\delta(x)$, включено в множество меток тех вершин, куда идет ребро из x. В случае неориентированных комплексов γ должно совпадать с δ. Все команды имеют различные левые части. Для того чтобы получить новое состояние $S^* = \Omega_\Gamma(S)$, мы должны:

1) найти команду, скажем, $U \to \langle W, \gamma, \delta \rangle$, левая часть которой совпадает с активной частью состояния S; после этого

2) удалить эту активную часть U и

3) заменить ее комплексом W; отображения γ, δ используются для того, чтобы связать вершины $v(S) \setminus v(U)$ с вершинами $v(W)$ следующим образом: для всех $a \in v(U)$, $b \in v(S) \setminus v(U)$, если $\langle b, a\rangle$ является ребром S и $\gamma(a)$ определено, то $\langle b, \gamma(a)\rangle$ будет ребром S^*; если $\langle a, b\rangle$ является ребром S и $\delta(a)$ определено, то $\langle \delta(a), b\rangle$ будет ребром S^*; после того как эти связи будут установлены, нужно

4) удалить все вершины из $v(S)$, не достижимые из начальной вершины, а также все ребра, инцидентные этим вершинам. Таким образом, новое состояние S^* также будет связным графом.

Сигнал окончания появляется, если активная часть оказывается принадлежащей данному конечному множеству комплексов; в этом случае, как указано выше, оператор Ω_Γ применяется еще один только раз. Более подробно: если активная часть состояния S принадлежит выделенному конечному множеству, состояние $S^* = \Omega_\Gamma(S)$ объявляется результатом работы алгоритма Г, и с появлением S^* алгоритмический процесс прекращается.

Изложению неориентированных алгоритмов Колмогорова был посвящен § 3 главы I монографии В. М. Глушкова [1]. На с. 34 автор писал: «Внимательный анализ описания алгоритмической схемы Колмогорова — Успенского показывает, что эта схема по форме в весьма значительной степени напоминает работу, фактически выполняемую человеком, когда он преобразовывает задаваемую ему извне информацию в соответствии с теми или иными правилами запоминаемого им алгоритма. Авторы этой схемы предпринимали специальные меры, чтобы не потерять общности в характере выполняемых преобразований».

VI. МАШИНЫ ШЁНХАГЕ (ОНИ ЖЕ АЛГОРИТМЫ КОЛМОГОРОВА —ШЁНХАГЕ)

В 1970 г. западногерманский математик А. Шёнхаге предложил вычислительную модель в виде «машин с модифицируемой памятью» (это есть предложенный в [7] перевод для термина Шёнхаге «storage modification machines», сокращенно SMM) (см. [11]). Во введении к статье [12] Шёнхаге пишет: «Как нам было указано многими коллегами, Колмогоров и Успенский ввели в рассмотрение машинную модель, весьма близкую к машинам с модифицируемой памятью, значительно ранее (в 1958 г., см. [5])». Машины Шёнхаге можно рассматривать как специальный вид ориентированных машин Колмогорова, и потому для обозначения этих машин в обзоре [7] (см. гл. 3, § 1) использовался термин «алгоритмы Колмогорова—Шёнхаге» (тогда как для обозначения неориентированных машин Колмогорова в [7] использовался термин «алгоритмы Колмогорова—Успенского»). Возможно, было бы правильнее называть ориентированные машины Колмогорова (а тем самым, и алгоритмы Колмогорова—Шёнхаге) машинами и алгоритмами с полулокальным преобразованием информации: действительно, при осуществлении одного шага работы алгоритма вновь появившаяся вершина может оказаться концом сколь угодно большого числа ребер.

Предложенное в [12] определение машин с модифицируемой памятью (машин Шёнхаге, или в терминологии обзора [7] алгоритмов Колмогорова—Шёнхаге) воспроизведено на с. 130—132 очерка [14] в качестве добавления к § 2 части I. Вкратце, основные отличия этой вычислительной модели от ориентированных машин Колмогорова состоят в следующем: 1) помечены ребра, а не вершины (несущественность этого момента отмечалась нами в начале раздела III); 2) команды являются специальным видом команд ориентированных машин Колмогорова; 3) на командах задана управляющая структура с помощью операторов перехода; 4) имеются две ленты — входная и выходная — на которых записываются слова в алфавите $\{0,1\}$, причем сразу вслед за входным словом пишется специальный символ, означающий его конец.

Основные результаты, связанные с алгоритмами Колмогорова—Шёнхаге, подробно изложены в [12].

VII. МОДЕЛИРОВАНИЕ В РЕАЛЬНОЕ ВРЕМЯ

Выше уже отмечалось, что определение алгоритма, предложенное А. Н. Колмогоровым, было призвано охватывать любые мыслимые виды алгоритмов. И действительно, машины Колмогорова непосредственно моделируют работу всех других известных видов алгоритмов с локальным преобразованием информации. Для алгоритмов с нелокальным преобразованием информации, таких, как нормальные алгорифмы А. А. Маркова или машины с произвольным доступом к памяти [10], требуется предварительное разбиение их шагов на локальные шаги. Тезис о возможности прямого моделирования алгоритмов с локальным преобразованием информации машинами Колмогорова может, на первый взгляд, вызвать возражения. Действительно, существование машины Колмогорова, непосредственно моделирующей машину Тьюринга с плоскостной памятью («плоской лентой»), не является самоочевидным, а требует весьма тонких построений. Так обстоит дело, например, при конструировании машины Колмогорова, решающей задачу «о распознавании самопересечения плоской траектории» (см. [6]); существование машины Тьюринга с плоской лентой, решающей эту задачу, очевидно. Дело здесь, однако, в том, что в действительности на вход рассматриваемой машины Тьюринга с плоской лентой подается не только аргумент, но и необходимая часть рабочей памяти, т. е. плоскости, стандартным образом разбитой на клетки. Это разбиение и возникающее в силу него отношение соседства клеток машина может использовать в своей работе. Если на вход машины Колмогорова подавать вместе с аргументом такую же часть плоскости, разбитой на клетки, то моделирование машины Тьюринга машиной Колмогорова становится очевидным. Тем не менее даже и без подачи на вход алгоритма Колмогорова участка плоскости моделирование оказывается в некотором, уточняемом ниже смысле возможным — машина Колмогорова оказывается в состоянии «достраивать» нужные участки памяти по ходу вычисления; как это делается, показано в [12].

Прежде чем сформулировать соответствующую теорему работы [12], приведем некоторые дополнительные определения. Наряду с машинами Шёнхаге будем рассматривать машины Тьюринга с многомерной памятью, описание которых сейчас будет дано. Предполагается, что каждая такая машина имеет входную и выходную ленты, движение головок по этим лентам одностороннее, со входной ленты возможно только чтение, на выходную — только запись. Рабочая память может содержать произвольное фиксированное число массивов фиксированной размерности (лент, плоскостей и т. д.), головки по этим массивам способны передвигаться за один шаг на одну клетку в любом направлении. Во всяком вычислении такой машины Тьюринга, так же как и во всяком вычислении машины Шёнхаге, выделяются моменты сдвига головки по входной ленте и моменты печати (очередного символа на выходной ленте). Пусть теперь $\mathfrak{A}$ и $\mathfrak{B}$ — две машины со входным и выходным алфавитом $\{0,1\}$, каждая из которых может быть машиной Тьюринга с многомерной памятью или машиной Шёнхаге. Будем в соответствии с [12] говорить, что машина $\mathfrak{A}$ *в реальное время моделирует* машину $\mathfrak{B}$, если выполнено следующее. Для всякого исходного данного x, на котором $\mathfrak{B}$ кончает работу, пусть $t_0 < t_1 < \ldots < t_p$ — такая последовательность моментов времени, что $t_0 = 0$, t_p — момент окончания работы машины, а $t_1, \ldots, t_{p-1}$ — все моменты времени, в которые машина $\mathfrak{B}$ производила сдвиг входной ленты или запись на выходную. Тогда должны существовать такие моменты времени $\tau_0 < \tau_1 < \ldots < \tau_p$, что $\tau_0 = 0$, τ_p — момент окончания работы машины $\mathfrak{A}$ для аргумента x и при всяком $i = 1, \ldots, p$;

1) при вычислении с исходным данным x часть этого x, прочитанная машиной $\mathfrak{A}$ к моменту τ_i, совпадает с частью x, прочитанной машиной $\mathfrak{B}$ к моменту t_i;

2) при вычислении с исходным данным x часть результата, напечатанная машиной $\mathfrak{A}$ к моменту τ_i, совпадает с частью результата, напечатанной машиной $\mathfrak{B}$ к моменту t_i;

3) для некоторого c, зависящего от $\mathfrak{A}$ и $\mathfrak{B}$, но не от x, имеет место

$$\tau_i - \tau_{i-1} \leqslant c\,(t_i - t_{i-1}).$$

Из этих условий, очевидно, следует, что машина $\mathfrak{A}$ вычисляет ту же функцию, что и машина $\mathfrak{B}$, или некоторое продолжение этой функции; при этом время вычисления для машины $\mathfrak{A}$ не превосходит времени вычисления $\mathfrak{B}$, умноженного на константу c.

В теореме из [12], о которой шла речь выше, доказана возможность моделирования в реальное время всякой машины Тьюринга с многомерной памятью подходящей машиной Шёнхаге. Как отмечено в [7], аналогичный результат может быть получен относительно моделирования машин Тьюринга с многомерной памятью алгоритмами Колмогорова—Успенского. Из результатов работы [2] вытекает невозможность обратных моделирований.

ЛИТЕРАТУРА

1. *Глушков В. М.* Введение в кибернетику. Киев: Изд-во АН УССР, 1964. 324 с.
2. *Григорьев Д. Ю.* Алгоритмы Колмогорова сильнее машин Тьюринга.— Зап. науч. семинаров ЛОМИ, 1976, т. 60, с. 29—37.
3. *Колмогоров А. Н.* О понятии алгоритма.— УМН, 1953, т. 8, вып. 4, с. 175—176.
4. *Колмогоров А. Н., Барздинь Я. М.* О реализации сетей в трехмерном пространстве.— Проблемы кибернетики, 1967, вып. 19, с. 261—268.
5. *Колмогоров А. Н., Успенский В. А.* К определению алгоритма.— УМН, 1958, т. 13, вып. 4, с. 3—28.
6. *Кубинец М. В.* Распознавание самопересечения плоской траектории алгорифмом Колмогорова.— Зап. науч. семинаров ЛОМИ, 1972, т. 32, с. 35—44.
7. *Слисенко А. О.* Сложностные задачи теории вычислений.— УМН, 1981, т. 36, вып. 6, с. 21—103.
8. *Успенский В. А.* Алгоритм.— В кн.: БСЭ. 3-е изд., 1970, т. 1, с. 400—401.
9. *Успенский В. А.* Алгоритм.— В кн.: Мат. энцикл., 1977, т. 1, с. 202—206.
10. *Aho A. V., Hopcroft J. E., Ullman J. D.* The design and analysis of computer algorithms. Reading (Mass.) etc.: Addison—Wesley Publ. Co., 1974. X + 470 p. Рус. пер.: *Ахо А., Хопкрофт Дж., Ульман Дж.* Построение и анализ вычислительных алгоритмов. М.: Мир, 1979. 536 с.
11. *Schönhage A.* Universelle Turing Speicherung.— In: Automatentheorie und formale Sprachen/Eds J. Dörr, G. Hotz. Mannheim, 1970, S. 369—383.
12. *Schönhage A.* Storage modification machines.— Soc. Industr. and Appl. Math. J. Comput., 1980, vol. 9, N 3, p. 490—508.
13. *Uspensky V. A., Semenov A. L.* What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its application in mathematics.— In: Algorithms in modern mathematics and computer science: Proc. Conf. Urgench (UzSSR). Berlin etc.: Springer-Verl., 1981, p. 100—234 (Lect. Notes in Comput. Sci.; Vol. 122).
14. *Успенский В. А., Семенов А. Л.* Теория алгоритмов: ее основные открытия и приложения.— В кн.: Алгоритмы в современной математике и ее приложениях. Новосибирск: ВЦ СО АН СССР, 1982, ч. 1. с. 99—342.

ИЗ ВОСПОМИНАНИЙ А. Н. КОЛМОГОРОВА

Теорию аналитических функций в этот год читали два лектора: Болеслав Корнелиевич Млодзеевский и Николай Николаевич Лузин.

Лекции Николая Николаевича посещались в принципе всеми членами Лузитании, включая старшекурсников и даже доцентов. С курсом Лузина связано первое мое достижение, после которого на меня было обращено некоторое внимание.

Николай Николаевич любил импровизировать на лекциях. На лекции, посвященной доказательству теоремы Коши, ему пришло в голову использовать такую лемму. Пусть квадрат разделен на конечное число квадратов. Тогда для любой константы C найдется такое число C', что для всякой кривой длины не больше C сумма периметров задевающих кривую квадратов не превосходит C'. Через две недели я обратился к председателю студенческого математического кружка Семену Самсоновичу Ковнеру с небольшой рукописью [1], где это утверждение было опровергнуто.

Приложение 1

ДОКЛАД МАТЕМАТИЧЕСКОМУ КРУЖКУ О КВАДРИЛЬЯЖЕ

Вопрос, о котором я буду говорить, возник на почве теории функций комплексного переменного, но по существу он чисто геометрический, и как такой я буду его рассматривать.

1.

На одной из лекций Н. Н. Лузина была доказана следующая теорема.

Дана на плоскости непрерывная спрямляемая кривая. Как бы ни была разделена плоскость на равные квадраты, сумма периметров всех квадратов, задеваемых, т. е. пересекаемых или касающихся, кривой, будет меньше постоянной величины K, лишь бы стороны квадратов оставались меньше некоторого $\mathscr{E}$.

Доказательство. Пусть ε — стороны квадратов деления, s — длина кривой. Разделим s на равные части меньше ε, для этого достаточно взять число частей $P > s/\varepsilon$ и всегда возможно $P \leqslant s/\varepsilon + 1$. Каждый отрезок кривой займет не более четырех квадратов, так как если бы он занял пять, то среди них были бы два не смежных, и, следовательно, число квадратов $N \leqslant 4P \leqslant 4\,(s/\varepsilon + 1)$. Периметр каждого квадрата 4ε, и, следовательно, сумма периметров $u = 4\varepsilon N \leqslant 4\varepsilon 4\,(s/\varepsilon + 1) = 16\,(s + \varepsilon)$, а так как $\varepsilon < \mathscr{E}$, то $u < 16\,(s + \varepsilon) = K$.

2.

Для доказательства обратного положения мне необходима следующая теорема. Дана непрерывная спрямляемая или неспрямляемая кривая $x = \varphi\,(t)$, $y = \psi\,(t)$ при изменении t от a до l.

Как бы ни была разделена кривая на конечное число частей, всегда расстояние между каждыми двумя точками несмежных частей будет больше некоторого постоянного α.

---

[1] Эта рукопись сохранилась (см. Приложение 1). Она датирована 4 января 1921 г. Мы предлагаем ее вниманию читателя.— *Примеч. ред.*

Доказательство. Допустим обратное. Тогда каждому положительному числу z соответствует множество T_z значений τ, точки соответствующие которым лежат на расстоянии, меньшем z, от точек несмежных с ними частей. Очевидно, что при $z_1 > z_2$ множество T_{z_2} заключается в множестве T_{z_1}. Разделим интервал (a, b) пополам и возьмем первую половину, в которой есть элементы всех множеств T; такая половина существует, так как если бы в одной половине не было элементов T_α, а в другой T_β, то в обеих бы не было элементов множеств с указателями, меньшими и α и β. Эту половину разделим опять пополам и также выберем первую долю, заключающую в себя элементы всех множеств. Продолжая этот процесс, получим последовательность включенных друг в друга безгранично убывающих отрезков. Все они включают в себя или имеют границей единственную точку τ_ω. От точки на кривой, соответствующей τ_ω, лежат сколь угодно близко точки, сами находящиеся на сколь угодно близком расстоянии от точек несмежных с ними частей, и, следовательно, на сколь угодно близком расстоянии от нее находятся точки несмежных частей или по крайней мере, если она является границей двух частей, противоположные границы их. Составим теперь множества T'_z точек, находящихся на расстоянии, меньшем z, от точки, соответствующей τ_ω, и лежащих в несмежных с ней частях или на противоположных с ней границах частей. Так же при $z_1 > z_2$ T'_{z_2} заключается в T'_{z_1}, так же найдем τ'_ω, от точки соответствующей которому лежат на сколь угодно близком расстоянии элементы всех T'. Эта точка не может лежать на конечном расстоянии от точки, соответствующей τ_ω, и, следовательно, с ней совпадает, что противоречит условию отсутствия двойных точек.

3.

Для неспрямляемой кривой без кратных точек сумма периметров квадратов, задаваемых кривой, безгранично возрастает с уменьшением сторон квадратов.

Покажем, что она может быть сделана больше любой данной величины K. Найдем, что всегда возможно, вписанный в кривую полигон длиною, большей K. Выделим из него все прямолинейные отрезки, только четные или нечетные, сумма которых больше $K/2$. Разобьем плоскость на столь мелкие квадраты, чтобы диагонали их были меньше минимального расстояния между точками частей кривой, соответствующих несмежным отрезкам полигона, какое по предыдущей теореме всегда существует. Сумма периметров квадратов, задаваемых такой частью кривой, будет больше удвоенной длины соответствующего отрезка. Выделенные несмежные части не могут проходить через один и тот же квадрат, и, следовательно, сумма периметров квадратов, задаваемых всеми ими, больше их общей удвоенной длины, или больше K.

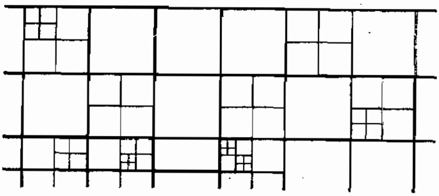


Рис. 1

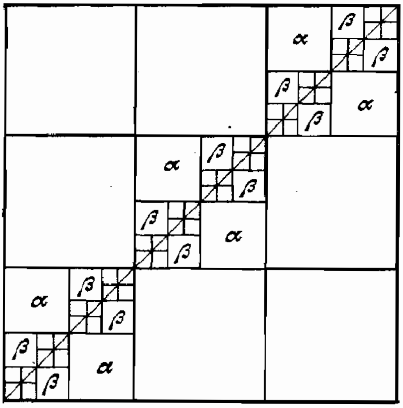


Рис. 2

4.

Для теории функций комплексного переменного важно было бы следующее расширение первой теоремы.

Дана непрерывная спрямляемая кривая. Плоскость разбивается сначала на равные квадраты, и затем некоторые из этих квадратов разбиваются далее путем последовательного разделения на четыре равные части. При всяком разбиении сумма периметров квадратов, задаваемых кривой, будет меньше постоянного K, лишь бы первоначальное разбиение на равные кьадраты было проведено достаточно далеко. Без вреда для применения можно видоизменить это положение, принимая во внимание только квадраты, пересекаемые кривой, а не касающиеся ее.

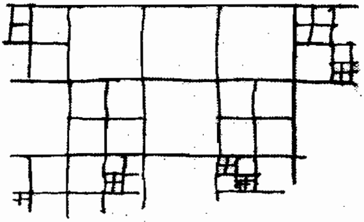
На двух примерах я попытаюсь показать ложность этого положения как в первоначальном, так и в видоизмененном виде.

А. Вместо кривой возьмем диагональ квадрата со стороной M. Разделим его на более мелкие равные квадраты. Как бы мелки эти квадраты ни были, можно подразделить их таким образом, чтобы сумма периметров квадратов, задаваемых диагональю, была сколь угодно велика. Для этого разделим диагональные квадраты. Квадраты α оставим без изменения, сумма их периметров равна $4M$. С новыми диагональными квадратами повторяем то же самое; квадраты β оставляем без изменения, сумма их периметров также $4M$. Продолжая достаточно далеко такое деление, получаем общую сумму периметров сколь угодно большой.

В. Для опровержения видоизмененного положения возьмем кривую, получаемую на прямой предыдущего примера, путем замены отрезка вокруг диагонали $^1/_{10}$ ее длины полуокружностью; отрезков вокруг точек, отмечающих $^1/_4$ и $^3/_4$ ее длины,— длиною в $^1/_{100}$; вокруг $^1/_8$, $^3/_8$, $^5/_8$ и $^7/_8$ — в $^1/_{1000}$ и т. д., поскольку они не налегают друг на друга. Общая длина замененных отрезков:

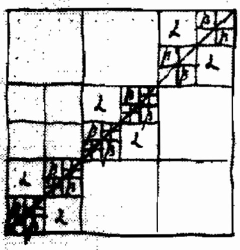
$$\frac{1}{10}+\frac{2}{100}+\frac{4}{1000}+\ldots=\frac{^1/_{10}}{1-{^2/_{10}}}=\frac{1}{8}.$$

периметров квадратов, касающихся кривой, будет меньше постоянного K, если бы первоначальное разбиение на равные квадраты было приведено достаточно далеко. Без вреда для применения можно видоизменять это положение, принимая во внимание равно квадраты пересекающие кривою а не касающиеся ее.



На двух примерах я попытаюсь показать ложность этого положения как в первоначальном так и в видоизмененном виде.

A.



Возьмем кривой в виде диагонали квадрата со стороною M. Разделим его на более мелкие равные квадраты. Как бы мелки эти квадраты ни были, мы можем подразделить далее, избежав чтобы сумма периметров квадратов, касающихся диагонали была сколь угодно велика. Для этого разделим диагональные квадраты. Квадраты L оставим без изменения сумма их периметров равна 4M. С новыми диагональными квадратами повторим то же самое; квадраты B оставим без изменения, сумма их периметров тоже 4M. Продолжая достаточно долго такие деления, можем общую сумму периметров сделать сколь угодно большой.

Доклад математическому кружку

о квадрильяже

А. Колмогоров

4 января 1921 г.

Страница рукописи «Доклада математическому кружку о квадрильяже», 1921 г.

Деление на квадраты оставим то же самое с тем ограничением, что при первоначальном разбиении на равные квадраты, каждую сторону большого квадрата разбиваем на число частей, являющееся степенью двух. Половина квадратов α, β и т. д. пересекается кривой, поскольку они окажутся против дуг с радиусом, большим их диагонали, но таких в каждом ряду будет не более $^1/_8$.

4 января 1921 г.

Приложение 2
ОБ ОПЕРАЦИЯХ НАД МНОЖЕСТВАМИ. II

ОТ АВТОРА

Первая часть моей работы об операциях над множествами, выполненная в 1921—1922 гг., была опубликована в 1928 г. в томе 35 «Математического сборника» (см. работу № 13 в первой книге моих избранных трудов «Математика и механика». М.: Наука, 1985). Вторая часть, оставаясь в рукописи, была доступна ряду исследователей по дескриптивной теории множеств, развивавших изложенную в этой второй части теорию R-операций. Ниже публикуется с небольшими редакционными изменениями вторая часть — по рукописи, обнаруженной уже после того, как первая книга избранных трудов была сдана в печать.

Как отмечено в тексте упомянутой статьи № 13 на с. 86, начала теории были независимо развиты Ф. Хаусдорфом и изложены им во 2-м немецком издании его «Теории множеств» (1927). В русском издании (1937) книги Хаусдорфа соответствующие параграфы написаны мной. Из-за отсутствия общего определения дополнительной операции теория Хаусдорфа менее полна. Образцом для моего общего определения дополнительной операции послужило данное П. С. Александровым определение Г-операции.

Далее идет текст второй части моей работы об операциях над множествами [1]. Так как первая часть закончилась разделом V, вторая начинается с раздела VI.

VI

Напомним, что все рассматриваемые в дальнейшем множества расположены на замкнутом сегменте[2] [0, 1] действительной прямой. За основной класс элементарных множеств принимаются замкнутые относительно основного сегмента множества, включая в их число «пустое» множество.

---

[1] Редакция благодарит Л. Б. Шапиро за подготовку к печати этой части работы А. Н. Колмогорова.— *Примеч. ред.*

[2] В этой части открытый интервал (0,1) заменен сегментом [0,1]. — *Примеч. ред.*

Для операции X над последовательностью множеств, заданной своей системой числовых цепей, определим *операцию RX над системой множеств со всеми возможными кортежами* вида $(n_1, n_2, \ldots, n_k)$, где $n_1, n_2, \ldots, n_k$ — натуральные числа:

$$E = RX \, \{E_{n_1 n_2 \ldots n_k}\}_{n_1 n_2 \ldots n_k}.$$

Цепью кортежей U^{RX} назовем всякую совокупность кортежей, обладающую такими свойствами:

1. Кортежи первого ранга[3] (λ), входящие в U^{RX}, образуют цепь U^X.

2. Если $(n_1 n_2 \ldots n_k)$ входит в U^{RX}, то индексы λ кортежей вида $(n_1 n_2 \ldots n_k \lambda)$, входящих в U^{RX}, образуют цепь U^X.

3. Если $(n_1 n_2 \ldots n_{k-1} n_k)$ входит в U^{RX}, то и $(n_1 n_2 \ldots n_{k-1})$ входит в U^{RX}.

Очевидно, что в случае, когда X есть операция сложения Σ, RX есть A-операция: $R\Sigma = A$.

Легко видеть, что в случае $E_{n_1 n_2 \ldots n_k} = E_{n_1}$

$$RX \, \{E_{n_1 n_2 \ldots n_k}\}_{n_1 n_2 \ldots n_k} = X \, \{E_n\}_n,$$

т. е. X-операция есть частный случай RX-операции.

Т е о р е м а I. *X-операция над RX-множествами дает всегда RX-множества.*

Это утверждение есть частный случай сформулированной далее теоремы II, но проще оно следует из таких формул: если $E = X \, \{E_n\}_n$, где

$$E_n = RX \, \{E^n_{m_1 m_2 \ldots m_k}\}_{m_1 m_2 \ldots m_k},$$

то, положив

$$\mathfrak{E}^*_{p_1 p_2 \ldots p_k} = E^{p_1}_{p_2 p_3 \ldots p_k} \text{ и } \mathfrak{E}^*_{p_1} = [0, 1],$$

получим

$$E = RX \, \{\mathfrak{E}^*_{p_1 p_2 \ldots p_k}\}_{p_1 p_2 \ldots p_k},$$

доказательство которых не представляет затруднений.

Если имеется конечное (например, X и $\bar{X}$) или счетное число операций, то, строя по методу, использованному в IV при доказательстве теоремы о непустоте классов (п. 2), операцию Z, их заменяющую, и затем RZ, получаем операцию, дающую из сегментов все множества области, инвариантной по отношению к данным операциям и содержащей сегменты (в силу теоремы I).

Т е о р е м а II. *Какова бы ни была операция X, операция RX нормальна.*

Пусть функции $i = \varphi(k)$, $j = \psi(k)$ и $k = f(i, j)$ осуществляют взаимно однозначное соответствие между натуральными числа-

---

[3] Число чисел, входящих в кортеж, называется его *рангом*. Пустой кортеж имеет нулевой ранг и входит в любую цепь кортежей.— *Примеч. ред.*

ми $k = f(i, j) = f(\varphi(k), \psi(k))$ и парами натуральных чисел (i, j) и удовлетворяют условиям:

$$f(i, j) < f(i, (j+1)), \tag{1}$$

$$f(i, j) < f((i+1), j), \tag{2}$$

$$f(i, j) > f(1, (i-1)). \tag{3}$$

Такое соответствие легко построить, положив, например, $f(i, j) = 2^{i-1}(2j-1)$. Из (1) и (3) вытекает, что

$$f(1, 1) = 1, \tag{4}$$

$$f(i, j) \geqslant i, \tag{5}$$

$$f(i, j) \geqslant j. \tag{6}$$

Предположим, что даны

$$E = RX\{E_{n_1n_2\ldots n_k}\}, \qquad E_{n_1n_2\ldots n_k} = RX\{E_{n_1n_2\ldots n_k}^{m_1m_2\ldots m_l}\}_{m_1m_2\ldots m_l}.$$

Пусть $\varphi(k) = i$. Положим при $i > 1$

$$\mathfrak{E}^*_{p_1p_2\ldots p_k} = E^{p_{f(i,1)}p_{f(i,2)}\cdots p_{f(i,j)}}_{p_{f(1,1)}p_{f(1,2)}\cdots p_{f(1,i-1)}}.$$

При $i = 1$ положим $\mathfrak{E}^*_{p_1p_2\ldots p_k} = [0,1]$. Обозначим через

$$\mathfrak{E}^* = RX\{\mathfrak{E}^*_{p_1p_2\ldots p_k}\}_{p_1p_2\ldots p_k}.$$

Достаточно доказать, что $\mathfrak{E}^* = E$.

1. Доказательство включения $E \subset \mathfrak{E}^*$. Пусть x входит в E. Значит, существует цепь $\mathfrak{U}^{RX}$ множеств $E_{n_1n_2\ldots n_k}$, заключающих x, и для каждого из них цепь $\mathfrak{U}^{RX}_{n_1n_2\ldots n_k}$ множеств $E_{n_1n_2\ldots n_k}^{m_1m_2\ldots m_l}$, содержащих x.

Выделим все $\mathfrak{E}^*_{p_1p_2\ldots p_q}$, удовлетворяющие таким условиям:

а) если $\varphi(q) = 1$, то множество

$$E_{p_{f(1,1)}p_{f(1,2)}\cdots p_{f(1,\psi(q))}} = E_{p_{f(1,1)}p_{f(1,2)}\cdots p_q}$$

входит в цепь $\mathfrak{U}^{RX}$;

б) если $\varphi(q) > 1$, то множество

$$E^{p_{f(\varphi(q),1)}p_{f(\varphi(q),2)}\cdots p_{f(\varphi(q),\psi(q))}}_{p_{f(1,1)}p_{f(1,2)}\cdots p_{f(1,\varphi(q)-1)}} = E^{p_{f(\varphi(q),1)}p_{f(\varphi(q),2)}\cdots p_q}_{p_{f(1,1)}p_{f(1,2)}\cdots p_{f(1,\varphi(q)-1))}}$$

входит в цепь $\mathfrak{U}^{RX}_{p_{f(1,1)}p_{f(1,2)}\cdots p_{f(1,\varphi(q)-1)}}$;

в) условия а) и б) выполнены также для всех $\mathfrak{E}^*_{p_1p_2\ldots p_r}$, $r < q$.

Я утверждаю, что совокупность $\mathfrak{S}$ всех $\mathfrak{E}^*_{p_1p_2\ldots p_k}$, удовлетворяющих выставленным условиям, образует цепь $\mathfrak{U}_1^{RX}$. В самом деле.

1. В силу в) если $\mathfrak{E}^*_{p_1p_2\ldots p_q}$ входит в $\mathfrak{U}_1^{RX}$, то в $\mathfrak{U}_1^{RX}$ входит и $\mathfrak{E}^*_{p_1p_2\ldots p_r}$, когда $r < q$.

2. Из $\mathfrak{E}^*_{p_1}$ в $\mathfrak{S}$ входят те, для которых E_{p_1} входит в $\mathfrak{U}^{RX}$ (так как $\varphi(1) = 1$), т. е. образующие своими индексами цепь U^X.

3. Если $\mathfrak{E}^*_{p_1p_2\ldots p_q}$ входит в $\mathfrak{S}$, то из множеств $\mathfrak{E}^*_{p_1p_2\ldots p_{q+1}}$ в $\mathfrak{S}$ входят те и только те, которые удовлетворяют условиям а) и б). Здесь представляются такие случаи.

(i) $\varphi(q) = 1$. В этом случае $\mathfrak{E}^*_{p_1p_2\ldots p_{f(1,\psi(q+1)-1)}}$ входит в $\mathfrak{S}$ по условию в), поэтому $E_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\psi(q+1)-1)}}$ в силу условия а) входит в $\mathfrak{U}^{RX}$, так как $\varphi(f(1, \psi(q+1)-1)) = 1$. Из множеств вида

$$E_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\psi(q+1)-1)}p_{f(1,\psi(q+1))}} = E_{p_{f(1,1)}p_{f(1,2)}\ldots p_{q+1}}$$

входят в $\mathfrak{U}^{RX}$ только те, последние индексы которых образуют цепь U^X. Поэтому в силу условия а) в $\mathfrak{U}_1^{RX}$ входят такие $\mathfrak{E}^*_{p_1p_2\ldots p_qp_{q+1}}$, что их индексы p_{q+1} образуют цепь U^X.

(ii) $\varphi(q+1) > 1$, $\psi(q+1) = 1$. В этом случае $\mathfrak{E}^*_{p_1p_2\ldots p_{f(1,\varphi(q+1)-1)}}$ входит в $\mathfrak{U}_1^{RX}$, так как $f(1, \varphi(q+1)-1) \leqslant q$. Следовательно, $E_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$ входит в $\mathfrak{U}^{RX}$. Но множества $\mathfrak{E}^*_{p_1p_2\ldots p_{q+1}}$, входящие в $\mathfrak{U}_1^{RX}$, суть по условию те, для которых множества

$$E^{p_{f(\varphi(q+1),1)}}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}} = E^{p_{q+1}}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$$

входят в цепь $\mathfrak{U}^{RX}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$. Их индексы p_{q+1} образуют, следовательно, цепь U^X.

(iii) $\varphi(q+1) > 1$, $\psi(q+1) > 1$. В этом случае множество

$$\mathfrak{E}^*_{p_1p_2\ldots p_{f(\varphi(q+1),\psi(q+1)-1)}} = E^{p_{f(\varphi(q+1),1)}p_{f(\varphi(q+1),2)}\ldots p_{f(\varphi(q+1),\psi(q+1)-1)}}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$$

входит в $\mathfrak{U}^{RX}$, так как $f(\varphi(q+1), \psi(q+1)-1) \leqslant q$. Следовательно, второе множество входит в $\mathfrak{U}^{RX}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$. Множества $\mathfrak{E}^*_{p_1p_2\ldots p_qp_{q+1}}$ входят в $\mathfrak{U}_1^{RX}$, если множества

$$E^{p_{f(\varphi(q+1),1)}p_{f(\varphi(q+1),2)}\ldots p_{f(\varphi(q+1),\psi(q+1)-1)}p_{f(\varphi(q+1),\psi(q+1))}}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}} =$$

$$= E^{p_{f(\varphi(q+1),1)}p_{f(\varphi(q+1),2)}\ldots p_{f(\varphi(q+1),\psi(q+1)-1)}p_{q+1}}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$$

входят в $\mathfrak{U}^{RX}_{p_{f(1,1)}p_{f(1,2)}\ldots p_{f(1,\varphi(q+1)-1)}}$, а это выполняется для тех p_{q+1}, которые образуют цепь U^X.

Итак, $\mathfrak{U}_1^{RX}$ есть действительно цепь. Все ее элементы содержат x, так как они или равны сегменту $[0, 1]$, или множеству

$E^{m_1 m_2 \ldots m_l}_{n_1 n_2 \ldots n_k}$, содержащему x. Следовательно, x входит в $\mathfrak{E}^*$.

2. Доказательство включения $\mathfrak{E}^* \subset E$. Пусть теперь x входит в $\mathfrak{E}^*$. Значит, есть цепь $\mathfrak{U}_1^{RX}$ множеств $\mathfrak{E}^*_{p_1 p_2 \ldots p_k}$, содержащая x.

A. Для каждого $\mathfrak{E}^*_{p_1 p_2 \ldots p_k}$ отметим $\mathfrak{E}^*_{p_1 p_2 \ldots p_k \bar{p}_{k+1}}$ — множество с наименьшим индексом p_{k+1}, равным $\bar{p}_{k+1}$, входящим в $\mathfrak{U}_1^{RX}$; очевидно, для каждого $\mathfrak{E}^*_{p_1 p_2 \ldots p_k}$, входящего само в $\mathfrak{U}_1^{RX}$, такое найдется.

B. Установим между множествами с одним и тем же $\varphi(k) = i$ такое подчинение (B):

множеству $\mathfrak{E}^*_{p_1 p_2 \ldots p_{f(i,j)}}$ непосредственно подчинены (B) множества вида:

$$\mathfrak{E}^*_{p_1 p_2 \ldots p_{f(i,\, j)} \bar{p}_{f(i,\, j)+1} \bar{p}_{f(i,\, j)+2} \ldots \bar{p}_{f(i,\, j+l)-1} p_{f(i,\, j+l)}},$$

где $p_{f(i,\, j+l)}$ принимает цепь U^X значений, при которых полученное множество входит в цепь $\mathfrak{U}_1^{RX}$, а $\bar{p}_{f(i,\, j)+n}$ вполне определены условием A.

C. Для $i = 1$ выделим все $\mathfrak{E}^*_{p_1}$ первого ранга, входящие в $\mathfrak{U}_1^{RX}$, и все им подчиненные (B). Множества $E_{p_{f(1,1)} p_{f(1,2)} \ldots p_{f(1,\, j)}}$, соответствующие выделенным, как нетрудно убедиться, образуют цепь $\mathfrak{U}^{RX}$.

D. Для каждого выделенного (C)

$$E_{p_{f(1,1)} p_{f(1,2)} \ldots p_{f(1,\, j)}}$$

выделим множества вида

$$\mathfrak{E}^*_{p_1 p_2 \ldots p_{f(1,\, j)} \bar{p}_{f(1,\, j)+1} \bar{p}_{f(1,\, j)+2} \ldots \bar{p}_{f(j+1,1)-1} p_{f(j+1,1)}},$$

где $p_{f(j+1,1)}$ принимает цепь U^X значений, при которых полученное множество входит в $\mathfrak{U}_1^{RX}$. Выделим, далее, все подчиненные (B) последним множествам. Нетрудно видеть, что они тождественны множествам цепи $\mathfrak{U}^{RX}_{p_{f(1,1)} p_{f(1,2)} \ldots p_{f(1,j)}}$ системы, определяющей $E_{p_{f(1,\, 1)} p_{f(1,\, 2)} \ldots p_{f(1,\, j)}}$, которое, следовательно, содержит x.

Отсюда: E содержит x.

VII

Если X-операция есть Г, то назовем RГ-операцию H-операцией (рассматривая при этом Г-операцию, определенную отделяющими цепями, см. с. 88 из № 13).

В соответствии с теоремой I из VI Г-операция может быть заменена H-операцией.

Теорема III. *A-операция заменяется H-операцией.*

В самом деле, ряд чисел $1, 2, 3, \dots$ распадается на счетное множество непересекающихся Г-цепей, в k-ю цепь

$$U_k^{\Gamma} = \{(a_1^k a_2^k a_3^k \dots)\}$$

будем считать входящими номера a_i^k кортежей k-го ранга. Если $E = A\,\{E_{p_1 p_2 \dots p_k}\}$, то

$$\mathfrak{E}^*_{n_1 n_2 \dots n_k} = E_{p_1 p_2 \dots p_k}$$

при $n_1 = a_{q_1}^{p_1}$, $n_2 = a_{q_2}^{p_2}$, $\dots$, $n_k = a_{q_k}^{p_k}$ с любыми $q_1, q_2, \dots, q_k$. Получим $E = H\,\{\mathfrak{E}^*_{n_1 n_2 \dots n_k}\}$.

В самом деле, легко видеть, что $\mathfrak{E}^*_{n_1 n_2 \dots n_k}$, равные $E_{p_1 p_2 \dots p_k}$, из какой-либо A-цепи образуют цепь $\mathfrak{U}^H$. Обратно, из любой цепи $\mathfrak{U}^H$ можно выделить множество, элементы которого равны членам некоторой A-цепи множеств $E_{p_1 p_2 \dots p_k}$.

Так как H-операция нормальна, то отсюда следует, что область H-множеств инвариантна по отношению к A- и Г-операциям и, следовательно, содержит C-область.

Если рассматривать Г-операцию не как операцию над последовательностью множеств, а как операцию над системой, занумерованной кортежами, то H-операция определится как операция над системой множеств со всеми возможными двойными кортежами:

$$E = H \left\{ E \left| \begin{matrix} n_1^1 & n_2^1 & \dots & n_{k_1}^1 \\ n_1^2 & n_2^2 & \dots & n_{k_2}^2 \\ \dots & \dots & \dots & \dots \\ n_1^{\nu} & n_2^{\nu} & \dots & n_{k_\nu}^{\nu} \end{matrix} \right| \right\},$$

где ν, k_i и n_j^i принимают независимо всевозможные значения.

Определение цепи двойных кортежей подобно данному в VI, если ранг считать равным числу строк, а кортежами, непосредственно подчиненными данному, получаемые из него прибавлением к нему одной строки снизу.

Пользуясь этим обозначением, можно получить удобную формулировку H-операции, но пока я сохраняю сокращенное обозначение, рассматривающее Г-операцию над последовательностью.

VIII

Определяющая система H-множества:

$$E = H\,\{\delta^1_{n_1 n_2 \dots n_k}\}_{n_1 n_2 \dots n_k}, \qquad \varphi = \{\delta^1_{n_1 n_2 \dots n_k}\}_{n_1 n_2 \dots n_k},$$

где δ^1 — сегменты, как и для A-множеств, может быть заменена равносильной ей «правильной» определяющей системой, в которой $\delta_{n_1 n_2 \dots n_k} \supset \delta_{n_1 n_2 \dots n_k n_{k+1}}$.

Для этого достаточно положить

$\delta_{n_1n_2\ldots n_k} = \delta^1_{n_1}\cdot\delta^1_{n_1n_2}\cdot\ldots\cdot\delta^1_{n_1n_2\ldots n_k}$.

Очевидно, что $E = H\,\{\delta_{n_1n_2\ldots n_k}\}$. В дальнейшем я буду всегда предполагать систему правильной.

1. Положим

$$E_{n_1n_2\ldots n_k} = H\,\{\delta_{n_1n_2\ldots n_kp_1p_2\ldots p_l}\}_{p_1p_2\ldots p_l}.$$

Очевидно, что

$$E = \Gamma\,\{E_{n_1}\}_{n_1},\qquad E_{n_1n_2\ldots n_k} = \Gamma\,\{E_{n_1n_2\ldots n_kn_{k+1}}\}_{n_{k+1}}.$$

Определим еще

$$\varphi_{n_1n_2\ldots n_k} \equiv \{\delta_{n_1n_2\ldots n_kp_1p_2\ldots p_l}\}_{p_1p_2\ldots p_l}.$$

2. Назовем *индексом* точки x относительно системы сегментов, определяющей H-множество, $\mathrm{Ind}_x\,\{\delta_{n_1n_2\ldots n_k}\}_{n_1n_2\ldots n_k} = \mathrm{Ind}_x\,\varphi$ число, определяемое таким индуктивным процессом:

(i) $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k} = 0$, если $\overline{\delta}_{n_1n_2\ldots n_k}$ содержит x.

(ii) Если определены все системы $\varphi_{n_1n_2\ldots n_k}$ с индексами, меньшими натурального числа m, то считаем $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k} = m$, если он еще не определен и если во всякой цепи U^Γ есть число λ такое, что $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k\lambda} < m$.

Иначе: если есть цепь U^A, состоящая только из таких λ.

(iii) Вообще если определены все системы с индексами, меньшими некоторого числа второго класса [4] α, то считаем $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k} = \alpha$, если он еще не определен и во всякой цепи U^Γ есть число λ такое, что $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k\lambda} = \beta < \alpha$.

(iv) Всем точкам, не получившим в силу предшествующих условий индекса второго класса, припишем условно индекс ω_1. Поскольку система φ занумерована пустым кортежем, то $\mathrm{Ind}_x\varphi$ корректно определен [5].

3. Т е о р е м а IV. $\mathrm{Ind}_x\varphi = \omega_1$ *тогда и только тогда, когда x входит в* $H_\varphi = H\,\{\delta_{n_1n_2\ldots n_k}\}$.

Д о к а з а т е л ь с т в о. В самом деле, если

$$\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k} = \omega_1,$$

то есть цепь U^Γ такая, что если $\lambda \in U^\Gamma$, то $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_k\lambda} = \omega_1$. Если это не так, то в каждой цепи U^Γ найдется элемент n_{k+1} такой, что $\mathrm{Ind}_x\varphi_{n_1n_2\ldots n_kn_{k+1}} < \omega_1$.

---

[4] α — трансфинитное число второго класса, если $\omega_0 \leqslant \alpha < \omega_1$.— *Примеч. ред.*

[5] По индуктивному определению $\mathrm{Ind}_x\varphi = \alpha < \omega_1$ тогда и только тогда, когда во всякой цепи U^Γ есть число λ такое, что $\mathrm{Ind}_x\,\varphi_\lambda < \alpha$.— *Примеч. ред.*

Но тогда в силу пункта (iii) определения индекса было бы

$\mathrm{Ind}_x \varphi_{n_1 n_2 \dots n_k} < \omega_1$.

Значит, совокупность $\delta_{n_1 n_2 \dots n_k}$, для которых $\mathrm{Ind}_x \varphi_{n_1 n_2 \dots n_k} = \omega_1$, удовлетворяет условиям 1 и 2 (см. VI) определения цепи. Следовательно, она содержит цепь $\mathfrak{U}^H$, в ядро которой входит точка x, т. е. $x \in H_\varphi$.

Обратно, покажем, что если $\mathrm{Ind}_x \varphi < \omega_1$, то x не входит в H_φ.

а) Утверждение очевидно в случае $\mathrm{Ind}_x \varphi = 1$ (см. определение индекса).

б) Пусть утверждение доказано для всех $\beta < \alpha$ и $\mathrm{Ind}_x \varphi = \alpha$. В каждой цепи $\mathfrak{U}^\Gamma$ множеств E_n есть такое E_λ, что $\mathrm{Ind}_x \varphi_\lambda = \beta < \alpha$, следовательно, x не входит в E_λ. Отсюда в силу формул раздела VIII, п. 1, x не входит в E.

Итак, дополнение к H-множеству состоит из $\aleph_1$ множеств

$$P_\alpha = \{x \mid \mathrm{Ind}_x \varphi = \alpha\}.$$

IX

Теорема V. *Дополнения к H-множествам, в частности все C-множества, суть $\aleph_1$-суммы все возрастающих A-множеств.*

Следуя обозначениям раздела VIII (п. 3), имеем

$$\bar{E} = P_1 + P_2 + \dots + P_{\omega_0} + \dots + P_\alpha + \dots,$$
$$\bar{E}_n = P_1^n + P_2^n + \dots + P_{\omega_0}^n + \dots + P_\alpha^n + \dots$$

Обозначим, далее

$$Q_\alpha = P_1 + P_2 + \dots + P_{\omega_0} + \dots + P_\alpha,$$
$$Q_\alpha^n = P_1^n + P_2^n + \dots + P_{\omega_0}^n + \dots + P_\alpha^n.$$

Если x входит в Q_α, то во всякой цепи $\mathfrak{U}^\Gamma$ множеств E_n найдется такое E_λ, что $x \in Q_\beta^\lambda$ для некоторого $\beta < \alpha$. Обратно, если это условие выполнено, то x входит в Q_α. Из этого следует формула

$$Q_\alpha = A\,\{\Sigma\,\{Q_\beta^n \mid \beta < \alpha\}\}_n$$

(см. определение дополнительной операции в разделе II первой части, кн. 1, № 13, с. 88).

Множества Q_α суть A-множества. В самом деле:

1. Это верно для $\alpha = 1$, так как $Q_1 = P_1 = \overline{\Gamma\,\{\delta_n\}_n} = A\,\{\bar{\delta}_n\}_n$.

2. Если это верно для $\beta < \alpha$, то Q_β^n суть A-множества, также их счетные суммы, а в силу нормальности A-операции Q_α также есть A-множество.

$$E = \Sigma \{Q_\alpha \mid \alpha < \omega_1\}, \qquad Q_\alpha \supset Q_\beta, \quad \text{если} \quad \alpha > \beta.$$

Сами H-множества, также и C-множества суть $\aleph_1$-произведения Г-множеств:

$$E = \Pi \{\bar{Q}_\alpha \mid \alpha < \omega_1\}.$$

Замечание 1. Вопрос о существовании H-множеств, дополнения которых были бы тоже H-множествами, не являющимися C-множествами, остается открытым. Методы, примененные для A-множеств, в этом случае неприменимы.

Замечание 2. Выводы теорем IV и V легко могут быть обобщены на RX- и $\overline{RX}$-множества. Определение индексов остается при этом без изменений, но вместо теоремы V получается

Теорема V*. *Если X-операция нормальна (следовательно, и $\bar{X}$) и содержит как частный случай операцию умножения (следовательно, $\bar{X}$ — сложения), то $\overline{RX}$-множества суть $\aleph_1$-суммы $\bar{X}$-множеств, а RX-множества — $\aleph_1$-произведения X-множеств.*

Теорема VI. *Если X-операция над измеримыми множествами всегда дает измеримое множество, то то же верно для RX-операции.*

Доказательство достаточно длинно, но по существу является лишь обобщением доказательства измеримости множеств, получаемых A-операцией над измеримыми множествами [6].

Теорема VII. *Если каждой точке сегмента* [0, 1] *приведена в соответствие определенная операция X_τ, то можно назвать операцию T, заменяющую* (см. раздел IV, п. 2 первой части, кн. 1, № 13, с. 91) *все операции X_τ.*

A. Занумеруем нечетными числами все рациональные сегменты (с рациональными концами). Назовем *цепью чисел* U^T всякую бесконечную последовательность чисел $n_1, n_2, \ldots, n_k, \ldots$, удовлетворяющую таким условиям:

(i) Последовательность сегментов, соответствующих нечетным элементам U^T, сходится к некоторой точке τ.

(ii) Четные элементы U^T, деленные на 2, образуют цепь из системы цепей $\{U^{X_\tau}\}$, определяющей операцию X_τ.

B. Пусть $E = X_\tau \{E_n\}_n$. Полагаем:

а) $\mathfrak{E}_{2n} = E_n$,

б) для номеров $2n + 1$ некоторой фиксированной последовательности рациональных сегментов, сходящейся к τ, $\mathfrak{E}_{2n+1} = [0, 1]$,

в) остальные $\mathfrak{E}_{2n+1}$ равными нулевому сегменту.

Нетрудно видеть, что в этих условиях $E = T\{\mathfrak{E}_n\}_n$.

---

[6] См.: *Lusin N., Sierpinski W.* Sur quelques propriétés des ensembles (A).— Bull. Acad. sci. Cracovie, 1918, vol. 4, p. 35—48.

Теорема VIII. *Если каждому трансфиниту второго класса соответствует операция, то можно дать операцию, заменяющую их все.*

Для этого достаточно разбить сегмент [0, 1] на $\aleph_1$ множеств P_α и привести в соответствие каждой точке P_α операцию, соответствующую α; затем построить операцию P.

Заключительное замечание. Если дано множество, то легко дать операцию, дающую его из сегментов; для этого достаточно занумеровать рациональные сегменты и назвать цепью операции совокупность номеров, соответствующую последовательности сегментов, сходящейся к точке множества. В силу теоремы VII можно эффективно континуальному классу множеств также найти операцию, дающую из сегментов все множества этого класса. Но так как мощность класса всех операций равна мощности класса всех точечных множеств, то можно думать, что с точки зрения изучения произвольных точечных множеств мы мало выигрываем, сводя их исследование к изучению классов множеств, порождаемых различными операциями рассмотренного в этой работе типа. Существенный интерес предлагаемый метод исследования имеет в применении к специальным классам множеств, порождаемых достаточно простыми операциями.

Февраль 1922 г.

ПОСЛЕСЛОВИЕ

Предлагаемые читателю три книги «Избранных трудов» включают в себя в существенном все мои работы по математике, классической механике, теории турбулентного движения, теории вероятностей, математической логике и теории информации. Сюда не вошли работы по методике преподавания и истории математики, стиховедению и статьи общего характера.

В некоторых направлениях сделанное мною представляется мне достаточно цельным и законченным, так что в моем 82-летнем возрасте я с удовольствием оставляю сделанное продолжателям.

В других направлениях дело обстоит иначе и опубликованное мною представляется мне лишь фрагментами будущих работ, относительно которых я надеюсь только, что они будут выполнены другими. То, что в этих направлениях уже сделано, во многих случаях описано в комментариях, составленных группой моих учеников, которым я выражаю сердечную благодарность.

25 апреля 1985 г. *А. Н. Колмогоров*

СОДЕРЖАНИЕ

КОММЕНТАРИИ И ПРИЛОЖЕНИЯ