

Г.А.Медведев, В.А.Морозов

Практикум на ЭВМ по анализу временных рядов

Учебное пособие

Медведев Г.А., Морозов В.А. Практикум на ЭВМ по анализу временных рядов [Электронный ресурс]: Учебное пособие. — Электрон. текст. дан. (1780 кб). — Мн.: “Электронная книга БГУ”, 2003. — Режим доступа: <http://anubis.bsu.by/publications/elresources/AppliedMathematics/morozov.pdf> . — Электрон. версия печ. публикации, 2001. — PDF формат, версия 1.4 . — Систем. требования: Adobe Acrobat 5.0 и выше.

МИНСК

«Электронная книга БГУ»

2004

© Г.А.Медведев, В.А.Морозов, 2004.

© Научно-методический центр

«Электронная книга БГУ», 2004

www.elbook.bsu.by

elbook@bsu.by

МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Г.А.Медведев, В.А.Морозов

Практикум на ЭВМ
по анализу временных рядов

Учебное пособие

Минск

2001

УДК 581.3:519.1(076.5)(075.8)
ББК 22.17я73
М 42

Рецензенты:

кафедра прикладной математики и экономической кибернетики Белорусского государственного экономического университета; Г.И.Лебедева, кандидат технических наук, доцент

Медведев Г.А.

М 42 Практикум на ЭВМ по анализу временных рядов: Учеб. пособие / Г.А.Медведев, В.А.Морозов. — Мн.: Университетское, 2001. — 192 с.
ISBN 985-09-0335-X.

Пособие подготовлено по программам курсов «Теория вероятностей и математическая статистика» и «Статистический анализ временных рядов» для студентов, обучающихся по специальности «Прикладная математика». Будет полезно студентам физико-математических, технических и экономических специальностей вузов, изучающих курс «Теория вероятностей и математическая статистика».

УДК 581.3:519.1(076.5)(075.8)
ББК 22.17я73

Учебное издание

Медведев Геннадий Алексеевич
Морозов Валерий Александрович

**ПРАКТИКУМ НА ЭВМ
ПО АНАЛИЗУ ВРЕМЕННЫХ РЯДОВ**

Учебное пособие

Редактор *А.В.Новикова*. Художественный редактор *НБ.Ярота*. Технический редактор *В.П.Безбородова*. Корректоры *Л.Н.Макейчик, Т.В.Кульнис*.

Подписано в печать 31.01.2001. Формат 70x100/16. Бумага офсетная. Гарнитура Таймс. Печать офсетная. Усл. печ. л. 15,48. Уч.-изд. л. 11,72. Тираж 230 экз. Заказ 5258.

Налоговая льгота — Общегосударственный классификатор Республики Беларусь ОКРБ 007-98, ч. 1; 22.11.20.600.

Издательское республиканское унитарное предприятие «Университетское» Государственного комитета Республики Беларусь по печати. Лицензия ЛВ № 9 от 8.09.2000. 220048, Минск, проспект Машерова, 11.

Отпечатано с оригинала-макета на республиканском унитарном предприятии «Типография «Победа». 222310, Молодечно, ул. Тавлая, 11.

ISBN 985-09-0335-X

© Медведев Г.А., Морозов В.А., 2001

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

AIC	— статистика информационного критерия Акаике
BIC	— статистика модифицированного критерия Акаике
C	— множество комплексных чисел
$\text{cor}(X, Y)$	— корреляция СВ X и Y
$\text{cov}(X, Y)$	— ковариация случайных величин X и Y
$D\{X\}$	— дисперсия СВ X
$\det(A)$	— определитель матрицы A
FRE	— статистика критерия, основанного на минимизации предельной ошибки предсказания
$F(P, N)$	— распределение Фишера с числом степеней свободы P и N (F -распределение)
I	— единичная матрица
$K_\alpha(N)$	— квантиль критерия Колмогорова — Смирнова
$M\{X\}$	— математическое ожидание СВ X
$\text{med}(x_i)$	— медиана выборки $\{x_i\}$
$N(a, \Sigma)$	— нормальное распределение со средним значением a и дисперсией (ковариационной матрицей) Σ
\mathbf{R}^n	— n -мерное евклидово пространство
$V\{\hat{X}\}$	— вариация оценки \hat{X}
Z	— множество целых чисел
$\Phi(x)$	— функция распределения $N(0, 1)$
$\Phi^{-1}(\beta), \Phi_\beta$	— квантиль распределения $\Phi(x)$ на уровне β
$\chi^2(\nu)$	— СВ с χ^2 -распределением с ν степенями свободы
$\binom{k}{l}$	— биномиальный коэффициент $\frac{k!}{l!(k-l)!}$
*	— знак транспонирования
\emptyset	— пустое множество
$\delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$	— символ Кронекера
\cong	— знак приближенного равенства

ОСНОВНЫЕ СОКРАЩЕНИЯ

АКФ	— автоковариационная функция
АР	— авторегрессия
АРСС	— авторегрессия — скользящее среднее
БПФ	— быстрое преобразование Фурье
ВКМ	— выборочная ковариационная матрица
ВКФ	— выборочная ковариационная функция
ГФП	— гауссовская функция правдоподобия
ДПФ	— дискретное преобразование Фурье
КМ	— ковариационная матрица
КФК	— квадратичная функция когерентности
ММП	— метод максимального правдоподобия
МНК	— метод наименьших квадратов
МСП	— матрица спектральных плотностей
НКФ	— нормированная ковариационная функция
НЛО	— наилучшая линейная оценка
НОР	— независимые одинаково распределенные
НСП	— нормированная спектральная плотность
ОСП	— оценка спектральной плотности
ПАРСС	— проинтегрированная авторегрессия — скользящее среднее
СВ	— случайная величина
СП	— случайный процесс
СПАРСС	— сезонная проинтегрированная авторегрессия — скользящее среднее
СС	— скользящее среднее
ССП	— стационарный случайный процесс
ТСА	— тип стохастической аппроксимации
ФР	— функция распределения

ПРЕДИСЛОВИЕ

Анализ временных рядов — одна из ветвей математической статистики, представляющая ярко выраженное практическое направление. Можно утверждать, что не существует такой области деятельности, имеющей дело с наблюдениями или измерениями, в которой не использовались бы методы анализа временных рядов.

Цель настоящего учебного пособия — ознакомление с основными методами анализа временных рядов и алгоритмами обработки данных для получения необходимых характеристик. Пособие предназначено для студентов, изучивших теорию вероятностей и математическую статистику или соответствующий курс высшей математики. Оно будет также полезно и специалистам, которые применяют методы анализа временных рядов для обработки данных в своей предметной области.

Пособие состоит из семи глав, посвященных основным направлениям в анализе временных рядов. Теоретический материал служит руководством к практическим работам. Каждая лабораторная работа составлена по следующим принципам: формулируется цель работы, затем определяется порядок ее выполнения, подразумевающий последовательность заданий. Лабораторные работы ориентированы на использование ЭВМ с графическим дисплеем, так как многие результаты полезно (а иногда и необходимо) иллюстрировать графиками. Однако при наличии определенного навыка можно организовать выдачу результатов в виде графика на любой ЭВМ. Наиболее подходящим средством выполнения лабораторных работ являются персональные ЭВМ типа IBM PC с применением соответствующей периферии и программного обеспечения. Лабораторные работы составлены таким образом, что каждая последующая производится с возможным использованием результатов предыдущих (по крайней мере, в пределах одной темы). Поэтому при прохождении практикума необходимо иметь персональную дискету, на которой бы организовывались и сохранялись все файлы данных для использования в случае необходимости.

Основная направленность лабораторных работ — не реализация методов и алгоритмов, а проведение исследования их эффективности. Поэтому следует обязательно производить сравнение применяемых методов и алгоритмов на основе полученных при работе наблюдений. В пособии излагаются, как правило, несколько известных подходов, на основе которых можно провести грамотный сравнительный анализ.

Главы 1,2,4—6 написаны Г.А.Медведевым, главы 3 и 7 — В.А.Морозовым.

Авторы выражают глубокую благодарность Е.В.Храмовой за большую помощь в подготовке к изданию настоящего учебного пособия.

СТАЦИОНАРНЫЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ И ВРЕМЕННЫЕ РЯДЫ

1.1. Основные понятия и числовые характеристики

Случайные процессы. Семейство случайных величин $\{X_t, t \in \mathfrak{S} \subset \mathbf{R}\}$ образует случайный процесс (СП), где t называется параметром СП, а \mathfrak{S} — параметрическим множеством. Если \mathfrak{S} — дискретное множество, например $\mathfrak{S} = \{0, \pm 1, \pm 2, \dots\}$, то СП называется СП с дискретным временем. Если \mathfrak{S} — интервал на числовой оси, то СП называется СП с непрерывным временем. Рассматривая X_t как случайную величину $\xi(t, \omega)$, отметим, что при фиксированном $\omega = \omega_0$ величина $x(t) = \xi(t, \omega_0)$ как функция параметра t называется *реализацией* (траекторией) СП. Значение $x(t)$ при фиксированном $t = t_k$, $x_k = x(t_k)$ называется *выборочным значением* или просто *значением* СП. Множество выборочных значений $\{x_k, t_k \in T \subseteq \mathfrak{S}\}$ называется *временным рядом*. Чаще всего, но не всегда, T — это дискретное множество. Множество T может совпадать с параметрическим множеством \mathfrak{S} . Можно сказать, что временной ряд — это множество данных о СП X_t , которыми располагает наблюдатель и на основании которых он должен делать выводы о свойствах наблюдаемого СП.

В соответствии с этим, как обычно делается в математической статистике, элементы временного ряда x_k рассматриваются, с одной стороны, как заданные, фиксированные числа при получении оценок параметров СП или принятия решений о его свойствах.

С другой стороны, при анализе свойств этих оценок или правил принятия решений элементы временного ряда изучаются как случайные величины с соответствующими вероятностными характеристиками. В каком смысле используются элементы временного ряда, становится ясно из контекста.

Для всякого целого n множество $\mathfrak{S}^n = \mathfrak{S} \times \mathfrak{S} \times \dots \times \mathfrak{S}$ является n -кратным прямым произведением параметрического множества самого на себя. Пусть $t_1 < t_2 < \dots < t_n$ и $\mathbf{t} = (t_1, t_2, \dots, t_n)$.

(Конечномерной) функцией распределения СП называется

$$F_n(\mathbf{x}, \mathbf{t}) = \mathbf{P}\{X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n\},$$

функция переменных $(x_1, x_2, \dots, x_n) = \mathbf{x}$. Функция распределения $F_n(\mathbf{x}, \mathbf{t})$ определена для всех n , $\mathbf{t} \in \mathfrak{S}^n$ и $\mathbf{x} \in \mathbf{R}^n$. Характеристическая функция СП задается равенством

$$\varphi(\mathbf{u}, \mathbf{t}) = \mathbf{M}\{e^{i\mathbf{u}^* \mathbf{x}}\} = \int_{\mathbf{R}^n} e^{i\mathbf{u}^* \mathbf{x}} dF_n(\mathbf{x}, \mathbf{t}),$$

где $\varphi(\mathbf{u}, \mathbf{t})$ определена для всех n , $\mathbf{u} \in \mathbf{R}^n$ и $\mathbf{t} \in \mathfrak{S}^n$. В показателе экспоненты под интегралом \mathbf{u} и \mathbf{x} представлены как векторы-столбцы размером n ; $*$ — знак транспонирования.

Числовые характеристики. Математическое ожидание СП и его дисперсия находятся по формулам

$$m(t) = \mathbf{M}\{X_t\} = \int_{\mathbf{R}} x dF_1(x, t),$$

$$D(t) = \mathbf{M}\{(X_t - m(t))^2\}.$$

Функции (авто)ковариации СП и (авто)корреляции СП задаются равенствами

$$\mathbf{R}^0(r, s) = \text{cov}(X_r, X_s) = \mathbf{M}\{(X_r - m(r))(X_s - m(s))\},$$

$$\mathbf{R}(r, s) = \text{cor}(X_r, X_s) = \mathbf{M}\{X_r X_s\}, \quad r, s \in \mathfrak{S}.$$

Формулы связи между этими функциями имеют вид

$$\mathbf{R}(r, s) = \mathbf{R}^0(r, s) + m(r)m(s),$$

$$\mathbf{R}^0(r, r) = D(r), \quad \mathbf{R}(r, r) = D(r) + (m(r))^2.$$

Основное свойство функций автоковариации и автокорреляции состоит в том, что они являются положительно определенными.

Функция $\mathbf{R}(s, t)$ называется *положительно определенной*, если выполняется следующее свойство. Пусть $t_1 < t_2 < \dots < t_n$ — любые вещественные числа и a_k, \bar{a}_k — любые ненулевые комплексно-сопряженные числа, $1 \leq k \leq n$. Функция $\mathbf{R}(s, t)$ положительно определенная, если для всех $n = 1, 2, \dots$ имеет место неравенство

$$\sum_{j=1}^n \sum_{k=1}^n a_j \bar{a}_k \mathbf{R}(t_j, t_k) > 0.$$

Пусть теперь $\{X_t\}, \{Y_t\}$ — два различных СП, определенных на одном и том же параметрическом множестве \mathfrak{S} ; $m_X(t), m_Y(t)$ — математические ожидания данных СП; $D_X(t), D_Y(t)$ — дисперсии СП.

Функции (взаимной) корреляции СП X_t и Y_t и их (взаимной) ковариации представляются в виде

$$\mathbf{R}_{XY}(r, s) = \text{cor}(X_r, Y_s) = \mathbf{M}\{X_r, Y_s\},$$

$$\mathbf{R}^0_{XY}(r, s) = \text{cov}(X_r, Y_s) = \mathbf{M}\{(X_r - m_X(r))(Y_s - m_Y(s))\}.$$

Соответствующая формула связи между ними задается формулой

$$\mathbf{R}_{XY}(r, s) = \mathbf{R}^0_{XY}(r, s) + m_X(r)m_Y(s).$$

Коэффициентом корреляции СП X_s, Y_t называется отношение

$$r(X_s, Y_t) = \text{cov}(X_s, Y_t) / \sqrt{D_X(s)D_Y(t)}.$$

Коэффициент корреляции обладает следующими свойствами:

- 1) $|r(X_s, Y_t)| \leq 1$;
- 2) если X_s, Y_t — независимы, то $r(X_s, Y_t) = 0$;
- 3) $|r(X_s, Y_t)| = 1$ тогда и только тогда, когда с вероятностью единица выполняется равенство $Y_t = aX_s + b$, $a \neq 0$.

В связи с этими свойствами коэффициент корреляции принято рассматривать как меру линейной (корреляционной) зависимости между СП $\{X_s\}$ и $\{Y_t\}$.

1.2. Стационарность и эргодичность

СП называется *стационарным в широком смысле* (или слабо стационарным), если выполняются следующие условия при $\mathfrak{S} = \mathbf{R}$ или $\mathfrak{S} = \mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$:

- 1) $\mathbf{M}\{|X_t|^2\} < +\infty$ для всех $t \in \mathfrak{S}$;
- 2) $\mathbf{M}\{X_t\} = m$ для всех $t \in \mathfrak{S}$ (математическое ожидание не зависит от времени);
- 3) $\text{cov}\{X_r, X_s\} = \text{cov}\{X_{r+t}, X_{s+t}\}$ для всех $r \in \mathfrak{S}$, $s \in \mathfrak{S}$, $t \in \mathfrak{S}$.

СП называется *стационарным в узком смысле* (или строго стационарным), если выполняется условие $F_n(\mathbf{x}, \mathbf{t}) = F_n(\mathbf{x}, \mathbf{t} + \mathbf{s})$ для всех $n = 1, 2, \dots$, всех $\mathbf{x} \in \mathbf{R}^n$, всех $\mathbf{t} \in \mathfrak{S}^n$, всех $\mathbf{s} = (s, s, \dots, s)$ $s \in \mathfrak{S}$.

Если $\mathbf{M}\{|X_t|^2\} < +\infty$ для всех $t \in \mathfrak{S}$, то СП, стационарные в узком смысле, являются стационарными и в широком смысле. Обратное, вообще говоря, неверно. Тем не менее имеются случаи, когда это так. В частности, из стационарности в широком смысле следует стационарность в узком смысле для гауссовских (нормальных) процессов.

Если СП стационарный (слабо или строго), то

$$\mathbf{R}(r, s) = R(r - s), \quad \mathbf{R}^0(r, s) = R^0(r - s).$$

Это значит, что функции автокорреляции и автоковариации зависят не от абсолютных значений своих переменных, а от их разности. Отсюда следует, что $R(\tau)$ и $R^0(\tau)$ — четные функции. Причем $R^0(0) = D$.

Коэффициентом корреляции стационарного СП является отношение

$$\rho(\tau) = R^0(\tau) / R^0(0) = \text{cov}(X_t, X_{t+\tau}) / D.$$

Приведем свойства коэффициента корреляции стационарного СП:

- 1) $|\rho(\tau)| \leq \rho(0) = 1$;
- 2) если $X_t, X_{t+\tau}$ — независимы, то $\rho(\tau) = 0$. Обратное справедливо не всегда;

3) если $X_{t+\tau} = aX_t + b$, $a \neq 0$, то $|\rho(\tau)| = 1$. Обратное утверждение имеет место с вероятностью единица.

Процессы $\{X_s\}$ и $\{Y_t\}$ не коррелированы, если $\text{cov}(X_s, Y_t) = 0$ при любых $s, t \in \mathfrak{Z}$. Процесс $\{X_t\}$ называется *процессом с ортогональными приращениями*, если для любых $t_1 < t_2 \leq t_3 < t_4$ разности $X_{t_2} - X_{t_1}$ и $X_{t_4} - X_{t_3}$ не коррелированы, т. е. $r(X_{t_2} - X_{t_1}, X_{t_4} - X_{t_3}) = 0$.

Среди стационарных СП выделяют класс *эргодических* процессов. Стационарный процесс $\{X_t\}$ с непрерывным временем называется *эргодическим по отношению к некоторой функции $f(x)$* , если имеет место равенство по вероятности

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(X_t) dt = \mathbf{M}\{f(X_t)\}.$$

Эргодическое свойство СП позволяет определять числовые характеристики СП не используя функций распределения, а основываясь на средних по параметру. Выяснение эргодичности СП вообще затруднительно. Однако в частных случаях могут быть сформулированы относительно простые признаки эргодичности СП.

Вещественный нормальный непрерывный стационарный процесс с нулевым средним $\{X_t\}$ является эргодическим по отношению к функции $f(x) = x$ тогда и только тогда, когда его спектральная функция $G(\lambda)$ непрерывна. По крайней мере, необходимыми условиями эргодичности СП $\{X_t\}$ по отношению к функции $f(x) = x$ являются следующие:

1) $\mathbf{M}\{|X_t|^2\} < +\infty$;

2) для каждого конечного T существуют интегралы Римана $\int_{-T}^T X_t dt$.

Эти условия должны быть дополнены достаточными, которые определяются для каждого конкретного случая. Понятие эргодичности легко модифицируется для дискретного времени. Для этого нужно операцию интегрирования по t заменить на операцию суммирования, при которой t принимает значения из множества \mathbf{Z} или его подмножества. Операция интегрирования по λ на всей числовой оси заменяется операцией интегрирования на интервале $(-\pi, \pi)$.

1.3. Спектральные свойства

Пусть $\{X_t\}$ — СП с непрерывным временем. СП X_t сходится к случайной величине X_0 при $t \rightarrow t_0$ в среднеквадратичном, если

$$\lim_{t \rightarrow t_0} \mathbf{M}\{(X_t - X_0)^2\} = 0.$$

СП $\{X_t\}$ непрерывен в среднеквадратичном в точке s , если X_t сходится к X_s в среднеквадратичном при $t \rightarrow s$.

Пусть теперь X_t — стационарный в широком смысле и непрерывный в среднеквадратичном СП с нулевым средним ($\mathbf{M}\{X_t\} = 0$). Тогда существует такой СП с ортогональными приращениями z_λ , что X_t для каждого фиксированного t допускает представление

$$X_t = \int_{-\infty}^{+\infty} e^{it\lambda} dz_\lambda, \quad t \in (-\infty, +\infty),$$

где стохастический интеграл понимается как предел в среднеквадратичном интегральной суммы (сначала $\Delta\lambda \rightarrow 0$, затем $N \rightarrow \infty$):

$$\sum_{n=-N}^N e^{itn\Delta\lambda} (z_{(n+1)\Delta\lambda} - z_{n\Delta\lambda}) \rightarrow \int_{-\infty}^{+\infty} e^{it\lambda} dz_\lambda,$$

где $z_\lambda = z'_\lambda + iz''_\lambda$ — СП с комплексными значениями; z'_λ, z''_λ — обычные СП; $\mathbf{M}\{z_\lambda\} = 0$ для всех λ . Представление СП X_t называется *спектральным представлением* СП X_t , а процесс z_λ — *спектральным процессом* СП X_t . Если X_t — СП с дискретным временем, то спектральное представление имеет вид

$$X_t = \int_{-\pi}^{+\pi} e^{it\lambda} dz_\lambda,$$

где z_λ , $-\pi \leq \lambda \leq \pi$, — непрерывный справа СП с ортогональными приращениями и $z_{-\pi} = 0$.

Справедлива теорема Бохнера, в соответствии с которой всякую положительно определенную функцию $R(\tau)$ можно представить в виде

$$R(\tau) = \int_{-\infty}^{+\infty} e^{it\lambda} dS(\lambda),$$

где $S(\lambda)$ — вещественная неубывающая и ограниченная функция. Если $R(\tau)$ — ковариационная функция СП X_t , то функция $S(\lambda)$ называется *спектральной функцией*. При этом

$$R(0) = \int_{-\infty}^{+\infty} dS(\lambda) = \lim_{\lambda \rightarrow \infty} (S(\lambda) - S(-\lambda)).$$

Спектральный процесс z_λ связан со спектральной функцией $S(\lambda)$ соотношением $S(\lambda) = \mathbf{M}\{|z_\lambda|^2\}$.

Так как спектральная функция определяется с точностью до константы, то удобно положить $S(-\infty) = 0$. Тогда $S(+\infty) = R(0)$. Если $S(\lambda)$ абсолютно непрерывна, то ее производная $g(\lambda) = dS(\lambda)/d\lambda$ называется *спектральной плотностью* СП X_t и имеют место взаимно обратные преобразования

$$R(t) = \int_{-\infty}^{+\infty} e^{it\lambda} g(\lambda) d\lambda, \quad R(0) = \int_{-\infty}^{+\infty} g(\lambda) d\lambda,$$

$$g(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-it\lambda} R(t) dt, \quad g(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R(t) dt.$$

Приведенные интегральные представления могут быть записаны через вещественные функции $u(t), v(t)$. Для процесса с непрерывным временем в среднеквадратичном справедливы следующие предельные соотношения:

$$\lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin \lambda t}{t} X_t dt = u(\lambda),$$

$$\lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{1 - \cos \lambda t}{t} X_t dt = v(\lambda).$$

Эти формулы верны для любого параметра $\lambda > 0$, который является точкой непрерывности спектральной функции $S(\lambda)$, а функции $u(\lambda)$ и $v(\lambda)$ — вещественные СП с ортогональными приращениями.

Определив вещественную функцию $G(\lambda)$ равенством $G(\lambda) = S(\lambda) - S(-\lambda)$, можно показать, что $G(\lambda)$ связана с $R(\tau)$ соотношениями

$$R(\tau) = \int_0^{\infty} \cos \lambda \tau dG\lambda, \quad G(\lambda) = \int_0^{\infty} \frac{\sin \lambda \tau}{\tau} R(\tau) d\tau.$$

Свойства СП $u(\lambda)$ и $v(\lambda)$ следующие: $\mathbf{M}\{u(\lambda)\} = \mathbf{M}\{v(\lambda)\} = 0$, $\mathbf{M}\{u(\lambda)v(\nu)\} = 0$, $\mathbf{M}\{u^2(\lambda)\} = G(\lambda)$, $\mathbf{M}\{v^2(\lambda)\} = G(\lambda) - G(+0)$ для всяких положительных λ и ν . Таким образом процессы $u(\lambda)$ и $v(\lambda)$ являются ортогональными. Они (совместно) могут рассматриваться как спектральный процесс, соответствующий СП X_t , спектральное представление которого можно переписать в виде

$$X_t = \int_0^{\infty} (\cos \lambda t du(\lambda) + \sin \lambda t dv(\lambda)).$$

1.4. Статистические критерии

Пусть в результате наблюдений СП получен временной ряд $X = \{x_t, 1 \leq t \leq N\}$.

Проверка на нормальность. Обозначим символами μ, d математическое ожидание и дисперсию элементов временного ряда. Нормированное среднее абсолютное отклонение δ , коэффициент асимметрии γ и коэффициент эксцесса β элементов временного ряда определяются равенствами

$$\delta = \frac{1}{\sqrt{d}} \mathbf{M}\{|x_t - \mu|\}, \quad \gamma = \frac{1}{\sqrt{d^3}} \mathbf{M}\{|x_t - \mu|^3\},$$

$$\beta = \frac{1}{d^2} \mathbf{M}\{(x_t - \mu)^4\}.$$

Для нормального распределения значения данных числовых характеристик следующие: $\delta = \sqrt{\frac{2}{\pi}} \cong 0,79788$; $\gamma = 0$; $\beta = 3$. Выборочные значения этих характеристик для исследуемого временного ряда вычисляются по формулам

$$\hat{\delta} = \frac{1}{\sqrt{\hat{d}N}} \sum_{t=1}^N |x_t - \hat{\mu}|, \quad \hat{\gamma} = \frac{1}{\sqrt{\hat{d}^3 N}} \sum_{t=1}^N |x_t - \hat{\mu}|^3,$$

$$\hat{\beta} = \frac{1}{\hat{d}^2 N} \sum_{t=1}^N (x_t - \hat{\mu})^4,$$

где используются выборочные математическое ожидание и дисперсия

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N x_t, \quad \hat{d} = \frac{1}{N} \sum_{t=1}^N (x_t - \hat{\mu})^2.$$

Таким образом, если выборочные значения $\hat{\delta}, \hat{\gamma}, \hat{\beta}$ существенно отличаются от значений δ, γ, β для нормального распределения, то распределение элементов ряда нельзя считать нормальным. В противном случае для проверки нормальности следует использовать более строгие критерии.

Критерий χ^2 . Разделим множество возможных значений элементов временного ряда X на несколько (например, на 11) интервалов таким образом, чтобы вероятности попадания выборочных значений в каждый интервал были примерно одинаковыми при проверяемой гипотезе. Например, проверяемой гипотезой может быть предположение о нормальном распределении элементов ряда. В этом случае при использовании $f = 11$ интервалов для стандартного нормального распределения границы интервалов u_i определяются значениями, приведенными в табл. 1.1.

Таблица 1.1

Номер интервала, i	Нижняя граница	Верхняя граница	Вероятность попадания, p_i
1	$-\infty$	-1,345	0,0893
2	-1,345	-0,915	0,0908
3	-0,915	-0,615	0,0892
4	-0,615	-0,360	0,0902
5	-0,360	-0,125	0,0908
6	-0,125	0,125	0,0994
7	0,125	0,360	0,0908
8	0,360	0,615	0,0902
9	0,615	0,915	0,0892
10	0,915	1,345	0,0908
11	1,345	$+\infty$	0,0893

Границы реальных интервалов U_i пересчитываются через нормированные границы интервалов u_i по формулам $U_i = \hat{\mu} + u_i \sqrt{\hat{d}}$, где $\hat{\mu}$ и \hat{d} — выборочные среднее значение и дисперсия временного ряда соответственно. Пусть h_i — число попаданий выборочных значений в i -й интервал, N — общее количество выборочных значений. Естественно, что $N = \sum_i h_i$. Тогда статистика критерия определяется соотношением

$$\psi_f = \sum_{i=1}^f (h_i - Np_i)^2 / Np_i.$$

При справедливости применяемой гипотезы некоррелированные выборочные значения являются независимыми и ψ_f при $N \rightarrow \infty$ сходится к случайной величине (СВ), распределенной по закону χ^2 с $(f - 1)$ степенями свободы. Проверяемая гипотеза о нормальности X на уровне значимости α отвергается, если $\psi_f > \chi_{f-1, 1-\alpha}^2$, где пороговое значение $\chi_{f-1, 1-\alpha}^2$ определяется по таблицам квантилей χ^2 -распределения. Уровень значимости λ обычно выбирается в интервале $[0,001; 0,1]$. Для $f = 11$ пороговое значение можно определить исходя из следующих данных:

α	0,001	0,005	0,01	0,025	0,05	0,1
$\chi_{f-1, 1-\alpha}^2$	29,59	25,19	23,21	20,48	18,31	15,99

Описанный критерий построен для независимых выборочных значений. Если они зависимые, то выводы могут оказаться неверными.

Критерий, основанный на порядковых статистиках. Пусть $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ — порядковая статистика выборочных данных $\{x_t, 1 \leq t \leq N\}$, подчиняющихся нормальному распределению $\mathbf{N}(\mu, d)$ со средним значением μ и дисперсией d . Если $z_{(1)} < z_{(2)} < \dots < z_{(N)}$ — порядковая статистика для СВ, имеющих стандартное нормальное распределение $\mathbf{N}(0, 1)$ выборки объемом

N , то $\mathbf{M}\{z_{(t)}\} = m_t$, а $\mathbf{M}\{x_{(t)}\} = \mu + dm_t$, $1 \leq t \leq N$. Таким образом, между x_t и m_t имеется близкая к линейной зависимость, что обеспечивает корреляцию, близкую к единице. Малое значение корреляции указывает на отсутствие линейной зависимости, в связи с чем распределение x_t нельзя считать нормальным. Величина m_t достаточно хорошо приближается значением обратной функции Лапласа в точке $(2t - 1)/2N$, т. е. $\Phi^{-1}((2t - 1)/2N)$. Поэтому в качестве статистики критерия можно взять выражение

$$R^2 = \frac{\left(\sum_{t=1}^N (x_t - \hat{\mu}) \Phi^{-1}((2t - 1)/2N) \right)^2}{\left(\sum_{t=1}^N (x_t - \hat{\mu})^2 \right) \left(\sum_{t=1}^N (\Phi^{-1}((2t - 1)/2N))^2 \right)},$$

которое по смыслу является квадратом выборочной корреляции между x_t и m_t . Гипотеза о нормальности отвергается, если $R^2 < R_\alpha^2$, где α — уровень значимости. Некоторые пороговые значения R_α^2 приведены в табл. 1.2.

Таблица 1.2

	α	
N	0,05	0,1
131	0,980	0,983
200	0,987	0,989

Критерий Аббе. Разобьем выборочные данные $X = \{x_t, 1 \leq t \leq N = nm\}$ на m подмножеств $X_{(k)}$, $1 \leq k \leq m$, где $X_k = \{x_t, (k - 1)n < t \leq kn\}$. Для этих подмножеств введем суммы

$$S_k = \frac{1}{n} \sum_{t=(k-1)n+1}^{kn} x_t, \quad 1 \leq k \leq m; \quad S = \frac{1}{m} \sum_{k=1}^m S_k.$$

Статистика Аббе представляет собой величину

$$q = \sum_{k=1}^{m-1} (S_{k+1} - S_k)^2 / 2 \sum_{k=1}^m (S_k - S)^2.$$

Проверяемой гипотезой является гипотеза равенства средних

$$\mathbf{M}\{S_1\} = \mathbf{M}\{S_2\} = \dots = \mathbf{M}\{S_m\}.$$

Альтернативой является неравенство

$$|\mathbf{M}\{S_{k+1}\} - \mathbf{M}\{S_k\}| > 0, \quad 1 \leq k \leq m - 1.$$

Согласно критерию Аббе, гипотеза о равенстве средних отвергается, если значение статистики q оказывается меньше критического значения $q_m(\alpha)$.

Критическое значение $q_m(\alpha)$ — это квантиль распределения q на уровне значимости α . При фиксированных m и α $q_m(\alpha)$ является решением уравнения $\mathbf{M}\{q < q_m(\alpha)\} = \alpha$. Для $m = 10$ критические значения равны:

α	0,001	0,01	0,05
$q_{10}(\alpha)$	0,2408	0,3759	0,5311

Критерий Аббе сформулирован для случая, когда суммы S_k , $1 \leq k \leq m$, независимы в совокупности.

Критерий Кокрена. Данный критерий используется для проверки гипотезы о равенстве m дисперсий d_k , $1 \leq k \leq m$, нормальных независимых выборок одинакового объема. Критерий основан на статистике

$$t = \max_{1 \leq k \leq m} \left(d_k / \sum_{j=1}^m d_j \right),$$

где

$$d_k = \frac{1}{n-1} \sum_{t=(k-1)n+1}^{kn} (x_t - S_k)^2;$$

$$S_k = \frac{1}{n} \sum_{t=(k-1)n+1}^{kn} x_t, \quad 1 \leq k \leq m.$$

На основе асимптотического распределения статистики t вычисляются критические значения $t_{m,n-1}(\alpha)$ для уровня значимости α .

Если $t > t_{m,n-1}(\alpha)$, то проверяемая гипотеза отвергается. При $m = 10$ критические значения могут быть найдены из табл. 1.3.

Таблица 1.3

	$n - 1$				
α	10	16	36	144	∞
0,01	0,2704	0,2297	0,1811	0,1376	0,1000
0,05	0,2353	0,2032	0,1655	0,1308	0,1000

Лабораторная работа 1. Исследование стационарного случайного процесса

Цель работы. Получить на ЭВМ реализацию случайного процесса с дискретным временем. Установить стационарность и эргодичность случайного процесса. Определить его основные характеристики.

Порядок выполнения работы

Определим последовательность случайных величин x_t рекуррентным соотношением

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \sum_{j=1}^q b_j W_{t-j}, \quad (1.1)$$

где $\{a_i\}, \{b_j\}$ — задаваемые наборы констант; $\{W_t\}$ — последовательность независимых одинаково распределенных (НОР) СВ. Заметим, что для определения стационарности последовательности (1.1) константы a_i следует задавать, пользуясь равенством

$$1 - \sum_{i=1}^p a_i z^i = \prod_{i=1}^p (1 - \alpha_i z), \quad (1.2)$$

где $\{\alpha_i\}$ — набор вспомогательных констант, таких, что $|\alpha_i| < 1$. Задав набор $\{\alpha_i\}$ из (1.2), определить набор параметров $\{a_i\}$.

Основным содержанием работы является, во-первых, получение числового материала; во-вторых, его анализ, связанный с формулировкой закономерностей, которые проявляются в этом числовом материале; в-третьих, сравнение числовых результатов, основанных на обработке полученных данных, с теоретическими результатами. Успех выполнения лабораторной работы в большой степени зависит от качества датчика случайных чисел. Поэтому, приступив к выполнению заданий, надо убедиться, что датчик обеспечивает необходимые условия. Кроме того, выбрав конкретный вид соотношения (1.1), порождающего СП, следует взять такие его параметры, которые обеспечат возможность аналитического получения характеристик СП x_t , таких, как числовые характеристики, нормированная ковариационная функция (НКФ) и нормированная спектральная плотность (НСП), эргодичность, стационарность, вид распределения.

Задание 1. Пользуясь формулой (1.1), получить реализацию СП, состоящую из N выборочных значений, $N = mn$, $X = \{x_t, 1 \leq t \leq N\}$. Расчленив выборку на m подпоследовательностей $Y_k = \{y_{tk}, 1 \leq t \leq n\}$ по правилу $y_{tk} = x_{t+n(k-1)}$, $1 \leq k \leq m$, $1 \leq t \leq n$.

Задание 2. Определить следующие выборочные числовые характеристики последовательностей X и Y_k , $1 \leq k \leq m$: $\hat{\mu}$, $\hat{\sigma}$, $\hat{\delta}$, $\hat{\gamma}$, $\hat{\beta}$ (см. § 1.4). Провести их сравнительный анализ для различных подпоследовательностей, провести проверку на нормальность.

Задание 3. Определить выборочные нормированные корреляционные функции последовательностей X и Y_k , $1 \leq k \leq m$.

Выборочная нормированная корреляционная функция находится по формуле

$$\hat{r}(\tau) = \frac{1}{\hat{d}(N - \tau - 1)} \sum_{t=1}^{N-\tau} (x_t x_{t+\tau} - \hat{\mu}^2), \quad 0 \leq \tau \leq \tau_{\max},$$

где τ_{\max} определяется из условия $\hat{r}(\tau) < \varepsilon$ для всех $\tau > \tau_{\max}$ и $\hat{r}(\tau_{\max} - 1) > \varepsilon$. Величина τ_{\max} называется временем корреляции на уровне ε . Для достаточно

малых ε можно считать, что выборочные значения СП X являются некоррелированными, если отстоят друг от друга на расстояние более чем τ_{\max} .

Вычислить числовые характеристики $\hat{\mu}, \hat{\sigma}, \hat{\gamma}, \hat{\beta}$ по некоррелированным выборочным значениям, проводя суммирование в этих формулах через τ_{\max} значений.

Провести сравнительный анализ числовых характеристик, полученных в задании 2, и по некоррелированным выборочным значениям.

Задание 4. Определить выборочные нормированные спектральные плотности (НСП) последовательностей X и Y_k , $1 \leq k \leq m$.

Для вычисления выборочных НСП использовать оценку Бартлетта, которая имеет вид

$$\hat{g}(\lambda) = \frac{1}{2\pi} \left(1 + 2 \sum_{\tau=1}^{\tau_{\max}} \left(1 - \frac{\tau}{\tau_{\max}} \right) \cos(\lambda\tau) \hat{r}(\tau) \right), \quad 0 \leq \lambda \leq \pi,$$

для СП X и Y_k , $1 \leq k \leq m$.

Эта формула определяет НСП только для неотрицательных λ . Поскольку $\hat{g}(\lambda)$ — четная функция, этого достаточно для выяснения вида $\hat{g}(\lambda)$ на всем интервале $[-\pi, \pi]$ изменения λ .

Задание 5. Для определенного класса стационарных СП $\{x_t\}$ можно показать, что значения выборочных корреляционных функций и выборочных спектральных плотностей нормально распределены. Поэтому при достаточно больших объемах выборки можно рассматривать совокупности $\{\hat{r}_k(\tau), 1 \leq k \leq m\}$ и $\{\hat{g}_k(\lambda), 1 \leq k \leq m\}$ как наблюдения НКФ $r(\tau)$ и НСП $g(\lambda)$ соответственно, распределенные по нормальному закону.

С учетом асимптотической нормальности значений НКФ и НСП построить доверительные интервалы для значений этих функций при рассмотрении каждого конкретного значения τ (или λ) в качестве наблюдений величины $r_k(\tau)$ (или $g_k(\tau)$), $1 \leq k \leq m$.

Задание 6. Проверить свойство эргодичности СП X по отношению к степенной функции. Для этого вычислить временные средние

$$\nu_l = \frac{1}{N} \sum_{t=1}^N x_t^l, \quad \nu_{lk} = \frac{1}{n} \sum_{t=1}^n y_{tk}^l, \quad 1 \leq l \leq 5.$$

Провести сравнительный анализ ν_k и ν_{kl} .

Задание 7. Основываясь на критерии χ^2 и критерии порядковых статистик, проверить нормальность СП X . Использовать эти критерии для некоррелированной части выборки и для всей выборки. Провести сравнение.

Задание 8. Основываясь на критерии Аббе, проверить гипотезу о равенстве средних значений последовательностей $\{y_{tk}\}$, $1 \leq k \leq m$, по некоррелированным выборочным значениям, имеющим нормальное распределение.

Задание 9. Основываясь на критерии Кокрена, проверить гипотезу о равенстве дисперсий последовательностей $\{y_{tk}\}$, $1 \leq k \leq m$, по некоррелированным нормально распределенным выборочным значениям.

Глава 2

ОЦЕНИВАНИЕ ФУНКЦИИ РЕГРЕССИИ. ВЫДЕЛЕНИЕ ТРЕНДОВ

2.1. Оценивание полиномиального тренда методом наименьших квадратов

Наиболее простой математической моделью временного ряда является сумма

$$x_t = m_t + s_t + y_t, \quad (2.1)$$

где m_t — некоторая медленно меняющаяся функция, называемая *трендом* (систематическая составляющая); s_t — более или менее регулярные колебания относительно тренда (сезонная составляющая); y_t — случайная (несистематическая, нерегулярная) компонента, как правило, представляющая собой стационарный случайный процесс с нулевым средним. Если $s_t = 0$, $m_t = \text{const}$ для всех t , то (2.1) является процессом, свойства которого подробно исследованы в гл.1.

Предположим теперь, что $s_t = 0$, а тренд m_t — достаточно медленно меняющаяся функция, которая может рассматриваться как "наилучшим образом сглаженный" временной ряд x_t . В такой постановке m_t удобно представлять в виде полинома по t некоторой степени q , т. е.

$$m_t = a_0 + a_1 t + a_2 t^2 + \dots + a_q t^q. \quad (2.2)$$

Если вероятностные характеристики случайной компоненты предполагаются известными, то задачей анализа такого временного ряда является определение неизвестных коэффициентов полинома (2.2).

Рассмотрим эту задачу вначале в предположении, что $\{y_t\}$ — некоррелированные СВ, т. е. $\mathbf{M}\{y_{t_1} y_{t_2}\} = 0$ для любых $t_1 \neq t_2$. Для компактности записи перейдем к матричным обозначениям. Представим, что для оценивания тренда (2.2) выполнено N наблюдений временного ряда (2.1) при значениях параметра t из множества $\{t_1, t_2, \dots, t_N\}$. Пусть $\mathbf{a} = (a_0, a_1, \dots, a_q)^*$ — вектор-столбец, составленный из коэффициентов полинома (2.2); $\Phi = (\Phi_{ij})$ — $(N \times (1 + q))$ -матрица с элементами $\Phi_{ij} = t_i^{j-1}$, $1 \leq i \leq N$, $1 \leq j \leq 1 + q$; $\mathbf{x} = (x_{t_1}, x_{t_2}, \dots, x_{t_N})^*$ — вектор-столбец наблюдений временного ряда; $\mathbf{y} = (y_{t_1}, y_{t_2}, \dots, y_{t_N})^*$ — вектор-столбец значений случайной компоненты в точках наблюдений. Тогда набор наблюдений временного ряда (2.1) представляется в матричной форме

$$\mathbf{x} = \Phi \mathbf{a} + \mathbf{y}. \quad (2.3)$$

Для того, чтобы определить тренд, нужно найти коэффициенты полинома (2.2), т. е. вектор \mathbf{a} . Наиболее распространенный метод нахождения данного вектора — метод наименьших квадратов (МНК), в соответствии с которым этот вектор находится из условия

$$\mathbf{y}^* \mathbf{y} = (\mathbf{x} - \Phi \mathbf{a})^* (\mathbf{x} - \Phi \mathbf{a}) \rightarrow \min_{\mathbf{a}}, \quad (2.4)$$

что приводит к решению системы линейных алгебраических уравнений относительно вектора \mathbf{a} , обычно называемой *нормальной* системой уравнений

$$\Phi^* \Phi \mathbf{a} = \Phi^* \mathbf{x}, \quad \hat{\mathbf{a}} = (\Phi^* \Phi)^{-1} \Phi^* \mathbf{x}. \quad (2.5)$$

Оценка $\hat{\mathbf{a}}$ вектора \mathbf{a} является несмещенной и по теореме Гаусса — Маркова обладает минимальной дисперсией в классе всех несмещенных линейных оценок (т. е. таких оценок, которые линейно связаны с наблюдениями (2.1)). Кроме того, векторы $\hat{\mathbf{a}}$ и $(\mathbf{x} - \Phi \hat{\mathbf{a}})$ некоррелированы. Причем матрица ковариации оценки $\hat{\mathbf{a}}$ имеет вид

$$\mathbf{M}\{(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^*\} = (\Phi^* \Phi)^{-1} D\{y_t\}, \quad (2.6)$$

где $D\{y_t\}$ — дисперсия случайной составляющей y_t .

Таким образом, точность оценивания вектора \mathbf{a} при помощи (2.5) определяется свойствами матрицы $(\Phi^* \Phi)^{-1}$. Для существования этой матрицы необходимо, чтобы $N \geq (1+q)$, и достаточно, чтобы Φ содержала не менее $(1+q)$ линейно независимых строк.

Обозначим диагональные элементы матрицы $\Phi^* \Phi$ через c_i^2 , а столбцы матрицы Φ — через Φ_i , т. е.

$$(\Phi^* \Phi)_{ii} = \Phi_i^* \Phi_i = \sum_{j=1}^N \Phi_{ij}^2 = c_i^2, \quad 1 \leq i \leq (1+q). \quad (2.7)$$

В этом случае имеет место неравенство

$$((\Phi^* \Phi)^{-1})_{ii} \geq c_i^{-2}, \quad 1 \leq i \leq (1+q). \quad (2.8)$$

Причем равенство достигается тогда и только тогда, когда $\Phi_i^* \Phi_j = 0$ при $i \neq j$. Из (2.6), (2.8) следует, что при фиксированных c_i^2 минимальные значения дисперсий оценки \hat{a} получаются тогда и только тогда, когда столбцы матрицы Φ ортогональны. Причем оценки \hat{a}_i коэффициентов a_i некоррелированы и вычисление как самих оценок, так и их дисперсий является существенно более простым по сравнению с (2.5) и (2.6). Действительно, если $\Phi_i^* \Phi_j = 0$ при $i \neq j$, то в (2.5) и (2.6) будем иметь

$$\hat{a}_{i-1} = \frac{\Phi_i^* \mathbf{x}}{\Phi_i^* \Phi_i}, \quad D\{\hat{a}_{i-1}\} = \frac{D\{y_t\}}{\Phi_i^* \Phi_i}, \quad 1 \leq i \leq (1+q). \quad (2.9)$$

Следовательно, матрицы Φ с ортогональными столбцами обеспечивают важное преимущество, которое заключается в том, что вызывающее затруднение

при вычислениях обращение матрицы $\Phi^*\Phi$, используемое в (2.5) и (2.6), при равенстве в (2.8) не требуется.

Итак, для получения точных и более простых с точки зрения вычислений оценок желательно, чтобы столбцы матрицы были ортогональными. Для достижения этого имеется две возможности: либо соответствующим образом задать точки наблюдения $\{t_i\}$, либо для всякого заданного набора точек наблюдения модифицировать структуру полинома (2.2) так, чтобы тренд m_t был линейной комбинацией полиномов, ортогональных на $\{t_i\}$. В первом случае, когда тренд представлен полиномом (2.2), а элементы матрицы Φ выбраны в виде $\Phi_{ij} = t_i^{j-1}$, ортогонализировать столбцы для вещественных $\{t_i\}$ невозможно, поскольку в этих условиях для всяких $l \neq k$, таких, что $l+k = 2(m+1)$, должны выполняться соотношения

$$\Phi_l^* \Phi_k = \sum_{i=1}^N t_i^{2m} = 0.$$

Остается второй случай, при реализации которого полиномиальный тренд (2.2) следует представить в виде

$$m_t = b_0 \varphi_0(t) + b_1 \varphi_1(t) + \dots + b_q \varphi_q(t), \quad (2.10)$$

где $\{\varphi_j(t), 0 \leq j \leq q\}$ — набор ортогональных полиномов на множестве $\{t_i, 1 \leq i \leq N\}$ вида

$$\varphi_j(t) = c_0(j) + c_1(j)t + \dots + c_{j-1}(j)t^{j-1} + t^j, \quad (2.11)$$

таких, что

$$\sum_{i=1}^N \varphi_l(t_i) \varphi_j(t_i) = 0 \quad \text{для всех } l \neq j. \quad (2.12)$$

В частности, (2.12) выполняется, если $\varphi_j(t)$ ортогонален по отношению к t^l , $0 \leq l \leq j-1$, на множестве $\{t_i\}$. Тогда для коэффициентов $c_k(j)$, $0 \leq k \leq j-1$, получается система уравнений

$$\sum_{k=0}^{j-1} c_k(j) \left(\sum_{i=1}^N t_i^{k+l} \right) = - \sum_{i=1}^N t_i^{l+j}, \quad 0 \leq l \leq j-1. \quad (2.13)$$

Используя обозначение

$$\tau_k = \frac{1}{N} \sum_{i=1}^N t_i^k, \quad (2.14)$$

представим (2.13) в более компактном виде:

$$\sum_{k=0}^{j-1} c_k(j) \tau_{k+l} = -\tau_{l+j}, \quad 0 \leq l \leq j-1. \quad (2.15)$$

Коэффициенты первых трех полиномов (2.11) будут иметь вид

$$\begin{aligned} c_0(0) &= 1; \\ c_1(1) &= 1, \quad c_0(1) = -\tau_1; \\ c_2(2) &= 1, \quad c_1(2) = \frac{\tau_1\tau_2 - \tau_3}{\tau_2 - \tau_1^2}, \quad c_0(2) = \frac{\tau_1\tau_3 - \tau_2^2}{\tau_2 - \tau_1^2}. \end{aligned}$$

При $t_i = i$ получаются наиболее простые полиномы. При этом предположении три первых полинома (2.11) выглядят следующим образом:

$$\begin{aligned} \varphi_0(t) &= 1, \\ \varphi_1(t) &= t - \frac{N+1}{2}, \\ \varphi_2(t) &= t^2 - (N+1)t + \frac{(N+1)(N+2)}{6}. \end{aligned}$$

Разрешив q систем уравнений (2.13) (при $1 \leq j \leq q$), можно найти коэффициенты $c_k(j)$ и набор ортогональных полиномов (2.11), которые задают матрицу Φ с элементами $\Phi_{ij} = \varphi_j(t_j)$. Свойства ортогональности полиномов $\varphi_j(t)$ позволяют найти оценки коэффициентов (2.10) и их дисперсий в виде

$$\hat{b}_j = \frac{\sum_{i=1}^N \varphi_j(t_i) x_{t_i}}{\sum_{i=1}^N (\varphi_j(t_i))^2}, \quad D\{\hat{b}_j\} = \frac{D\{y_t\}}{\sum_{i=1}^N (\varphi_j(t_i))^2}, \quad 0 \leq j \leq q. \quad (2.16)$$

Приведем явные формулы для оценки линейного тренда ($q = 1$). Обозначим

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_{t_i}, \quad \tau\bar{x} = \frac{1}{N} \sum_{i=1}^N t_i x_{t_i}, \quad \sigma^2 = D\{y_t\}. \quad (2.17)$$

Тогда оценка тренда имеет вид

$$\hat{m}_t = \frac{\tau_2 \bar{x} - \tau_1 \tau \bar{x} + (\tau \bar{x} - \tau_1 \bar{x})t}{\tau_2 - \tau_1^2}, \quad (2.18)$$

а ее дисперсия —

$$D\{\hat{m}_t - m_t\} = \frac{\sigma^2}{N} \left(1 + \frac{t^2}{\tau_2 - \tau_1^2} \right). \quad (2.19)$$

Отметим не менее важное преимущество представления тренда ортогональными полиномами (2.10), (2.11) по сравнению с представлением (2.2). Примененный здесь МНК основывается на квадратичной мере качества приближения оценки \hat{m}_t к тренду m_t . При этом предполагается, что вид m_t известен с точностью до коэффициентов, поэтому оценивание тренда связывается с оцениванием коэффициентов. Однако такая замена не всегда правомерна.

Действительно, в случае представления (2.2) и оценок (2.5) выборочное среднееквadraticное отклонение оценки \hat{m}_t от тренда m_t в точках наблюдения выражается равенством

$$\frac{1}{N} \sum_{i=1}^N (\hat{m}_{t_i} - m_{t_i})^2 = \sum_{k=0}^q \sum_{l=0}^q (\hat{a}_k - a_k)(\hat{a}_l - a_l) \tau_{k+l}. \quad (2.20)$$

Отсюда видно, что минимизация дисперсии оценок \hat{a}_j еще не обеспечивает минимизации среднего значения (2.20). В случае представления (2.10)–(2.12) и оценок (2.16) выборочное среднееквadraticное отклонение оценки \hat{m}_t от тренда m_t находится в виде

$$\frac{1}{N} \sum_{i=1}^N (\hat{m}_{t_i} - m_{t_i})^2 = \sum_{k=0}^q (\hat{b}_k - b_k)^2 \left(\frac{1}{N} \sum_{i=1}^N \varphi_k^2(t_i) \right). \quad (2.21)$$

В этом случае минимизация дисперсии оценок \hat{b}_j обеспечивает и минимизацию среднего значения (2.21).

До сих пор предполагалось, что значения случайной компоненты y_t в (2.1) являются независимыми в совокупности. К сожалению, это не всегда имеет место на практике. Пусть теперь значения y_t для различных t коррелированы, а их корреляция известна, т.е. $\mathbf{M}\{y_t y_s\} = R(t-s)$, где $R(t)$ — заданная функция. Обозначим $R_{ij} = R(t_i - t_j)$, R — $(N \times N)$ -матрица, составленная из R_{ij} . Тогда оценка (2.5) выглядит следующим образом:

$$\hat{\mathbf{a}} = (\Phi^* R^{-1} \Phi)^{-1} \Phi^* R^{-1} \mathbf{x}. \quad (2.22)$$

При этом оценка существенно усложняется, так как приходится оперировать с матрицей R^{-1} , получение которой для больших объемов выборки наблюдений N является затруднительным в вычислительном отношении. Однако когда удастся найти матрицу $R^{-1/2}$ (под матрицей $R^{1/2}$ понимается симметрическая невырожденная матрица, удовлетворяющая условию $R^{-1/2} R R^{-1/2} = I$) на основании конкретных свойств R , то, заменив $\tilde{\Phi} = R^{-1/2} \Phi$, $\tilde{\mathbf{x}} = R^{-1/2} \mathbf{x}$, можно перейти к случаю некоррелированных случайных компонент, так как исходное соотношение (2.3) превращается в равенство

$$\tilde{\mathbf{x}} = \tilde{\Phi} \mathbf{a} + \tilde{\mathbf{y}}, \quad (2.23)$$

где случайный вектор $\tilde{\mathbf{y}} = R^{-1/2} \mathbf{y}$ имеет уже некоррелированные компоненты и выполняются условия предыдущего анализа только по отношению к новым преобразованным данным.

Положительная определенность функции $R(t)$ влечет за собой положительную определенность матрицы R .

Поскольку матрица R является положительно определенной, матрица $R^{-1/2}$ всегда существует и тоже является положительно определенной.

В случае коррелированных y_t оценка (2.16) также существенно усложняется, так как ортогональные полиномы $\varphi_j(t)$ должны удовлетворять не (2.2), а более сложному требованию

$$\sum_{\alpha=1}^N \sum_{\beta=1}^N \varphi_l(t_\alpha) R_{\alpha\beta}^{-1} \varphi_j(t_\beta) = 0 \quad \text{для всех } l \neq j, \quad (2.24)$$

а это приводит к тому, что вместо уравнений (2.15) относительно коэффициентов $c_k(j)$ ортогональных полиномов $\varphi_j(t)$ получаются уравнения

$$\sum_{k=0}^{j-1} c_k(j) \tau_{k,l} = -\tau_{j,l}, \quad 0 \leq l \leq j-1, \quad (2.25)$$

где

$$\tau_{k,l} = \sum_{\alpha=1}^N \sum_{\beta=1}^N t_\alpha^k t_\beta^l R_{\alpha\beta}^{-1}. \quad (2.26)$$

В (2.24) и (2.26) $R_{\alpha\beta}^{-1}$ обозначает элемент матрицы R^{-1} , $R_{\alpha\beta}^{-1} \equiv (R^{-1})_{\alpha\beta}$. Сами уравнения (2.25) по сложности не отличаются от (2.15), но для вычисления коэффициентов $\tau_{k,l}$ приходится обращать матрицу R , что и усложняет проблему нахождения ортогональных полиномов $\varphi_j(t)$.

Если в представлении (2.10) полиномы $\varphi_j(t)$ удовлетворяют (2.11), (2.25), то оценки коэффициентов b_j и их дисперсии выражаются формулами

$$\hat{b}_j = \frac{\sum_{\alpha=1}^N \sum_{\beta=1}^N \varphi_l(t_\alpha) R_{\alpha\beta}^{-1} x_{t_\beta}}{\sum_{\alpha=1}^N \sum_{\beta=1}^N \varphi_j(t_\alpha) R_{\alpha\beta}^{-1} \varphi_j(t_\beta)}, \quad D\{\hat{b}_j\} = \frac{D\{y_t\}}{\sum_{\alpha=1}^N \sum_{\beta=1}^N \varphi_j(t_\alpha) R_{\alpha\beta}^{-1} \varphi_j(t_\beta)}. \quad (2.27)$$

До сих пор неявно предполагалось, что степень q полинома (2.2), характеризующего тренд временного ряда (2.1), известна. Однако на практике такое предположение может оказаться неоправданным. Поэтому возникает проблема оценивания не только коэффициентов полинома (2.2), но и его степени q . Это может быть сделано следующим образом. Оценка качества приближения тренда полиномом степени q по наблюдениям $\mathbf{x} = (x_{t_1}, x_{t_2}, \dots, x_{t_N})^*$ с вероятностью $(1 - \alpha)$ характеризуется величиной

$$J(q) = \frac{(\mathbf{x} - \Phi \tilde{\mathbf{b}})^* (\mathbf{x} - \Phi \tilde{\mathbf{b}})}{N - \sqrt{N \left((1+q) \left(1 + \ln \frac{N}{1+q} \right) - \ln \alpha \right)}}, \quad (2.28)$$

где Φ — $(N \times (1+q))$ -матрица с элементами $\Phi_{ij} = \varphi_j(t_i)$; $\varphi_j(t)$ — полиномы, определяемые соотношениями (2.11), (2.14) и (2.15); $\tilde{\mathbf{b}}$ — $(1+q)$ -вектор-столбец, составленный из оценок (2.16).

Та степень q , при которой $J(q)$ принимает наименьшее значение, является оптимальной степенью полиномиального приближения, а тренд m_t аппроксимируется алгебраическим полиномом (2.10) этой степени. Следует учитывать тот факт, что минимизация $J(q)$ имеет смысл только в области значений параметров, которые обеспечивают неотрицательные значения (2.28).

Если дисперсия случайной составляющей y_t неизвестна, то она может быть оценена по формуле

$$\hat{D}\{y_t\} = \frac{1}{N-1-q}(\mathbf{x} - \Phi\hat{\mathbf{b}})^*(\mathbf{x} - \Phi\hat{\mathbf{b}}), \quad (2.29)$$

где Φ — матрица с элементами $\Phi_{ij} = \varphi_j(t_i)$; $\hat{\mathbf{b}}$ — вектор-столбец оценок (2.16).

В том случае, когда для описания тренда используется степенной полином (2.2), дисперсия случайной составляющей оценивается по формуле

$$\hat{D}\{y_t\} = \frac{1}{N-1-q}(\mathbf{x} - \Phi\hat{\mathbf{a}})^*(\mathbf{x} - \Phi\hat{\mathbf{a}}), \quad (2.30)$$

где Φ — матрицы с элементами $\Phi_{ij} = t_i^{j-1}$, $1 \leq i \leq N$, $1 \leq j \leq (1+q)$; $\hat{\mathbf{a}}$ — вектор-столбец, определяемый (2.5).

Рассмотрим теперь случай, когда систематическая составляющая m_t в (2.1) отсутствует, но сезонная компонента s_t отличается от нуля. Причем

$$s_t = \sum_{j=1}^q (a_j \cos \omega_j t + b_j \sin \omega_j t), \quad t = 1, 2, \dots, N, \quad 0 \leq \omega_j \leq \pi, \quad (2.31)$$

где ω_j — набор известных параметров; a_j, b_j — неизвестны. Случайная компонента y_t представляет собой последовательность некоррелированных случайных СВ. Так что

$$x_t = s_t + y_t, \quad t = 1, 2, \dots, N, \quad (2.32)$$

а s_t определена в (2.31).

Асимптотически несмещенной оценкой коэффициентов a_j и b_j являются

$$\hat{a}_j = \frac{2}{N} \sum_{t=1}^N x_t \cos \omega_j t, \quad \hat{b}_j = \frac{2}{N} \sum_{t=1}^N x_t \sin \omega_j t. \quad (2.33)$$

Лабораторная работа 2. Выделение полиномиального тренда

Цель работы. Освоить технику выделения полиномиальных трендов и приближения функций регрессии полиномами. Научиться оценивать точность аппроксимации и выбирать оптимальные параметры аппроксимирующих полиномов.

Порядок выполнения работы

Предполагается исследовать один из двух объектов: временной ряд с полиномиальным трендом или временной ряд, тренд которого полиномом не является, но аппроксимируется полиномом. Все задания одинаковы по отношению к указанным объектам. Только в случае, когда тренд — полином, надо выяснить, насколько точно определяется степень полиномиального тренда q при помощи минимизации функционала (2.28). Когда же исследуется тренд, не являющийся полиномом, предполагается выбирать его из класса дробно-рациональных, иррациональных или трансцендентных функций.

После того, как решение о виде тренда принято, следует задать множество точек наблюдения, включая в него (не обязательно) все значения параметра t , для которых предполагается получать значения временного ряда.

Задание 1. Пользуясь датчиком случайных чисел, получить реализацию временного ряда в виде (2.3).

Задание 2. Построить оценку (2.5) коэффициентов тренда, пользуясь реализацией (2.3). Построить оценку тренда (2.2) по полученным оценкам коэффициентов в точках наблюдения $\hat{m}_t = \hat{a}_0 + \hat{a}_1 t + \dots + \hat{a}_q t^q$, $t \in \{t_1, t_2, \dots, t_N\}$. Найти $\Delta_a = \frac{1}{N} \sum_{k=1}^N |\hat{m}_{t_k} - m_{t_k}|$. Оценить дисперсию случайной составляющей по формуле (2.30).

Задание 3. Построить набор ортогональных полиномов $\{\varphi_j(t), 0 \leq j \leq q\}$ на множестве $\{t_i\}$, пользуясь уравнениями (2.15).

Задание 4. Построить оценки (2.16) коэффициентов тренда, представленного в виде (2.10). По этим оценкам найти значение тренда в точках $\{t_i\}$:

$$\hat{m}_t = \hat{b}_0 \varphi_0(t) + \hat{b}_1 \varphi_1(t) + \dots + \hat{b}_q \varphi_q(t), \quad t \in \{t_1, t_2, \dots, t_N\}.$$

Найти $\Delta_b = \frac{1}{N} \sum_{k=1}^N |\hat{m}_{t_k} - m_{t_k}|$. Сравнить Δ_a и Δ_b . Оценить дисперсию случайной составляющей по формуле (2.29).

Задание 5. Пользуясь датчиком случайных чисел, построить случайную компоненту временного ряда по типу скользящего среднего:

$$y_t = \frac{1}{\sqrt{n}} \sum_{k=1}^n \left(\xi_{t-k} - \frac{1}{2} \right),$$

где ξ_t — СВ, распределенные равномерно в $[0,1]$. В этом случае y_t являются коррелированными, так что $\mathbf{M}\{y_t y_s\} = R(t-s) = (1 - \frac{|t-s|}{n})\sigma^2, |t-s| < n$, $R(t-s) = 0, |t-s| \geq n$, где σ^2 — дисперсия СВ ξ_t .

Задание 6. Построить реализацию временного ряда с коррелированной случайной составляющей. Найти оценку (2.22) коэффициентов полиномиального тренда. Вычислить оценку тренда в точках наблюдения и определить ее отклонение Δ_a так же, как в задании 2.

Задание 7. Построить набор ортогональных полиномов $\{\varphi_j(t), 0 \leq j \leq q\}$ на множестве $\{t_i\}$, применив уравнения (2.25).

Задание 8. Построить оценки (2.27) коэффициентов представления (2.10), определить значение тренда в точках наблюдения и их отклонения от истинных значений так, как в задании 4. Дать сравнительный анализ выполнения заданий 2, 4, 6, 8.

Задание 9. Построить функционал качества (2.28) приближения тренда полиномом степени q и определить оптимальные значения q . Найти оптимальный набор полиномов (2.11)–(2.15), соответствующие ему оценки (2.16) и выборочное квадратичное отклонение Δ_b , введенное в задании 4.

Задание 10. Применяя датчик случайных чисел и задавая параметры $q, a_j, b_j, \omega_j, 1 \leq j \leq q$, получить по формулам (2.31), (2.32) реализацию временного ряда с периодическим трендом. По формулам (2.33) найти оценки коэффициентов a_j, b_j , если параметры q и ω_j известны. Исследовать изменение оценок \hat{a}_j, \hat{b}_j в зависимости от объема выборки N .

2.2. Оценивание функции регрессии. Общий случай

Функцией регрессии временного ряда $\{x_t, t \in T\}$ называется *математическое ожидание СВ x_t , рассматриваемое как функция параметра t* :

$$m_t = \mathbf{M}\{x_t\}. \quad (2.34)$$

Отличие (2.34) от (2.1) состоит в том, что, во-первых, значения временного ряда не разделяются на случайные и детерминированные компоненты. Во-вторых, m_t не представляется в виде определенной функциональной зависимости, как в (2.2), (2.10) или (2.31).

Для рассмотрения общего случая используем некоторые унифицированные обозначения. Пусть $T \subset [a, b]$, где a, b — конечные величины, т. е. t является скалярной переменной. Унифицируем изменение t , определив новую переменную

$$u = \frac{2t - a - b}{b - a}. \quad (2.35)$$

При изменении t в интервале $[a, b]$ переменная u принимает значения от -1 до $+1$ для любых a и b . Таким образом, не теряя общности, вместо (2.34) можно рассматривать функцию

$$m(u) = m\left(\frac{2t - a - b}{b - a}\right) \equiv m_t, \quad -1 \leq u \leq +1. \quad (2.36)$$

Когда верхняя граница интервала $[a, b]$ неограниченна, т. е. $b = +\infty$, полагаем $u = t - a$, так что $m(u) = m(t - a) \equiv m_t, u \in (0, +\infty)$. Наконец, когда T совпадает с числовой осью, $u = t$.

Будем предполагать, что функция $m(u)$ принадлежит некоторому линейному подпространству, заданному набором функций $\{\varphi_1(u), \varphi_2(u), \dots, \varphi_q(u)\}$, т. е.

имеет место представление

$$m(u) = \sum_{j=1}^q c_j \varphi_j(u) = \mathbf{c}^* \boldsymbol{\varphi}(u), \quad (2.37)$$

где $\mathbf{c} = (c_1, c_2, \dots, c_q)^*$ — вектор коэффициентов разложения $m(u)$ по базису $\{\varphi_j(u)\}$; $\boldsymbol{\varphi}(u) = (\varphi_1(u), \varphi_2(u), \dots, \varphi_q(u))^*$ — вектор-столбец, составленный из функций $\{\varphi_j(u)\}$. Представление (2.37) определяет функцию регрессии $m(u)$ с точностью до коэффициентов $\{c_j\}$. Эти коэффициенты обычно определяются из требования минимизации функционала

$$\int (m(u) - \mathbf{c}^* \boldsymbol{\varphi}(u))^2 du \rightarrow \min_{c_j}. \quad (2.38)$$

Здесь, как и далее, интегрирование осуществляется по всему интервалу. В общем случае оценивания функции регрессии аргумент t в (2.34), а следовательно, и u в (2.35) могут выбираться некоторым случайным образом в соответствии с некоторой функцией распределения $P(u)$. Тогда вместо функционала (2.38) вводится его обобщение

$$\int (m(u) - \mathbf{c}^* \boldsymbol{\varphi}(u))^2 dP(u) \quad (2.39)$$

и коэффициенты $\{c_j\}$ находятся из условия минимизации (2.39).

Определим функционал среднего риска соотношением

$$J(\mathbf{c}) = \mathbf{M} \left\{ \int (x(u) - \mathbf{c}^* \boldsymbol{\varphi}(u))^2 dP(u) \right\}, \quad (2.40)$$

где, как и в случае (2.35), $x(u) = x\left(\frac{2t-a-b}{b-a}\right) \equiv x_t$. Математическое ожидание $\mathbf{M}\{\cdot\}$ в (2.40) вычисляется по значениям временного ряда, т. е. по x . Можно показать, основываясь на (2.34), что вектор \mathbf{c} , доставляющий минимум функционалу (2.39), минимизирует и средний риск $J(\mathbf{c})$, т. е. представление функции регрессии (2.37) находится при помощи минимизации (2.40). Вместе с тем в реальных задачах ни $P(u)$, ни распределение значений временного ряда, как правило, не известны и средний риск $J(\mathbf{c})$ в явной форме не может быть определен. Поэтому коэффициенты разложения функции регрессии принято получать при помощи минимизации *эмпирического риска*, который вводится следующим образом. В реальных условиях временной ряд x_t наблюдается на некотором конечном множестве t_1, t_2, \dots, t_N значений параметра t . Положим для определенности $t_i < t_{i+1}$. Наблюдаемому ряду соответствует множество $\{x(u_i), 1 \leq i \leq N\}$.

Эмпирическим риском называется

$$J_{\text{э}}(\mathbf{c}) = \frac{1}{N} \sum_{i=1}^N (x(u_i) - \mathbf{c}^* \boldsymbol{\varphi}(u_i))^2. \quad (2.41)$$

Предположим, что наблюдения $\{x(u_i)\}$ независимы в совокупности. Имеет место следующее неравенство Хёфдинга. Если $0 \leq (x(u_i) - \mathbf{c}^* \boldsymbol{\varphi}(u_i))^2 \leq \tau$, $1 \leq i \leq N$, то

$$P\{|J(\mathbf{c}) - J_{\mathfrak{A}}(\mathbf{c})| > \kappa\} < \eta, \quad (2.42)$$

где κ и η связаны соотношением

$$\kappa = \tau \sqrt{\ln(1/\eta)/2N}. \quad (2.43)$$

Из (2.43) следует, что всегда найдется объем выборки N , при котором с любой заданной вероятностью $1 - \eta$ абсолютная величина разности между $J(\mathbf{c})$ и $J_{\mathfrak{A}}(\mathbf{c})$ не будет превышать κ . Аналогичный результат можно получить и из неравенства Чебышева, которое имеет тоже вид (2.42), но величины κ и η в нем связаны соотношением

$$\kappa = \sqrt{D/N\eta}, \quad (2.44)$$

где D — дисперсия $(x(u) - \mathbf{c}^* \boldsymbol{\varphi}(u))^2$. Из (2.43), (2.44) можно оценить объем выборки, обеспечивающий с заданной вероятностью $1 - \eta$ гарантированную точность κ нахождения среднего риска через эмпирический. Это является основанием для определения коэффициентов разложения (2.37) не минимизацией (2.40), а минимизацией риска (2.41), который может быть построен по наблюдениям временного ряда. Пусть $\mathbf{x} = (x(u_1), x(u_2), \dots, x(u_N))^*$ — вектор-столбец наблюдений временного ряда; Φ — $(N \times q)$ -матрица с элементами $\Phi_{ij} = \varphi_j(u_i)$. Тогда для определения коэффициентов разложения получаем уравнение

$$J_{\mathfrak{A}}(\mathbf{c}) = \frac{1}{N} (\mathbf{x} - \Phi \mathbf{c})^* (\mathbf{x} - \Phi \mathbf{c}) \rightarrow \min_{\mathbf{c}}, \quad (2.45)$$

которое в формальном смысле не отличается от уравнения (2.44). Решение (2.45) сводится к решению нормальной системы уравнений

$$\Phi^* \Phi \mathbf{c} = \Phi^* \mathbf{x}; \quad \hat{\mathbf{c}} = (\Phi^* \Phi)^{-1} \Phi^* \mathbf{x}. \quad (2.46)$$

О свойствах решения (2.46) ранее уже говорилось. Выбор функций $\{\varphi_j(u)\}$ должен быть таким, при котором матрица $\Phi^* \Phi$ не вырождена, и если Φ составлена из ортогональных столбцов, то точность оценки (2.46) наиболее высокая (см. (2.8)). Более предпочтительным является случай, когда столбцы матрицы Φ не только ортогональные, но и ортонормированные. Тогда $\Phi^* \Phi = I$ и оценка (2.46) приобретает наиболее простую форму

$$\hat{\mathbf{c}} = \Phi^* \mathbf{x}. \quad (2.47)$$

Системы ортогональных полиномов. В § 2.1 была рассмотрена процедура построения ортогональных полиномов. Приведем системы ортогональных функций, используемых чаще других.

Полиномы Лежандра $P_k(u)$ ортогональны на интервале $[-1, +1]$. Рекуррентной формулой для определения последовательности полиномов является равенство

$$P_{k+1}(u) = \frac{2k+1}{k+1}uP_k(u) - \frac{k}{k+1}P_{k-1}(u). \quad (2.48)$$

Первые шесть полиномов имеют вид

$$\begin{aligned} P_0(u) &= 1, \\ P_1(u) &= u, \\ P_2(u) &= \frac{1}{2}(3u^2 - 1), \\ P_3(u) &= \frac{1}{2}(5u^3 - 3u), \\ P_4(u) &= \frac{1}{8}(35u^4 - 30u^2 + 3), \\ P_5(u) &= \frac{1}{8}(63u^5 - 70u^3 + 15u) \end{aligned}$$

с нормировочным соотношением $\int P_k^2(u) du = (k + \frac{1}{2})^{-1}$.

Ортогональные функции, основанные на полиномах Чебышева, определяются для интервала $[-1, +1]$ формулой $\varphi_k(u) = (1 - u^2)^{-1/4}T_k(u)$, где $T_k(u)$ задаются рекуррентной зависимостью

$$T_{k+1}(u) = 2uT_k(u) - T_{k-1}(u). \quad (2.49)$$

Первые шесть полиномов Чебышева имеют вид

$$\begin{aligned} T_0(u) &= 1, \\ T_1(u) &= u, \\ T_2(u) &= 2u^2 - 1, \\ T_3(u) &= 4u^3 - 3u, \\ T_4(u) &= 8u^4 - 8u^2 + 1, \\ T_5(u) &= 16u^5 - 20u^3 + 5u \end{aligned}$$

с нормировочным соотношением $\int \varphi_0^2(u) du = \pi$, $\int \varphi_k^2(u) du = \int (1 - u^2)^{-1/2} T_k^2(u) du = \frac{\pi}{2}$, $k \geq 1$.

Тригонометрические функции, ортогональные на $[-1, +1]$, образуют ортонормированную систему

$$\varphi_0(u) = \frac{1}{\sqrt{2}}, \quad \varphi_{2k}(u) = \cos(k\pi u),$$

$$\varphi_{2k-1} = \sin(k\pi u), \quad k = 1, 2, \dots, \quad (2.50)$$

с нормировочным соотношением $\int \varphi_k^2(u) du = 1$.

Ортогональные на $(0, +\infty)$ функции, основанные на полиномах Лаггера $L_k(u)$, выражаются формулой

$$\varphi_k(u) = e^{-u/2} L_k(u),$$

где полиномы $L_k(u)$ определяются рекуррентно:

$$L_{k+1}(u) = (2k + 1 - u)L_k(u) - k^2 L_{k-1}(u). \quad (2.51)$$

Первые шесть полиномов Лаггера имеют вид

$$L_0(u) = 1,$$

$$L_1(u) = -u + 1,$$

$$L_2(u) = u^2 - 4u + 2,$$

$$L_3(u) = -u^3 + 9u^2 - 18u + 6,$$

$$L_4(u) = u^4 - 16u^3 + 72u^2 - 96u + 24,$$

$$L_5(u) = -u^5 + 25u^4 - 200u^3 + 600u^2 - 600u + 120$$

с нормировочным соотношением $\int \varphi_k^2(u) du = \int e^{-u} L_k^2(u) du = (k!)^2$.

Примером функций, ортогональных на всей числовой оси $(-\infty, +\infty)$, являются функции, построенные с использованием полиномов Эрмита $H_k(u)$, т. е.

$$\varphi_k(u) = e^{-u^2/2} H_k(u).$$

Полиномы Эрмита $H_k(u)$ находятся по рекуррентной формуле

$$H_{k+1}(u) = 2uH_k(u) - 2kH_{k-1}(u). \quad (2.52)$$

Первые шесть полиномов Эрмита имеют вид

$$H_0(u) = 1,$$

$$H_1(u) = 2u,$$

$$H_2(u) = 4u^2 - 2u,$$

$$H_3(u) = 8u^3 - 12u,$$

$$H_4(u) = 16u^4 - 48u^2 + 12,$$

$$H_5(u) = 32u^5 - 160u^3 + 120u$$

с нормировочным соотношением $\int \varphi_k^2(u) du = \int e^{-u^2} H_k^2(u) du = 2^k k! \sqrt{\pi}$.

Ортогональную систему функций можно построить для всякой конечной или бесконечной последовательности линейно независимых функций $f_1(u), f_2(u), \dots$, нормируемых на некотором интервале. Наиболее известным способом такого преобразования функций является метод ортогонализации Грама — Шмидта, который заключается в следующем. Обозначим $(f, f) = \int f^2(u) du$. Пусть

$$\begin{aligned} \varphi_1(u) &= f_1(u), \\ \varphi_{k+1}(u) &= f_{k+1}(u) - \sum_{l=1}^k \frac{(\varphi_l, f_{k+1})\varphi_l(u)}{(\varphi_l, \varphi_l)}, \quad k = 1, 2, \dots \end{aligned} \quad (2.53)$$

Функции $\varphi_k(u)$, определенные таким образом, являются ортогональными. Если, кроме того, положить $\psi_k(u) = \varphi_k(u)/\sqrt{(\varphi_k(u), \varphi_k(u))}$, то функции $\psi_k(u)$ образуют систему ортонормированных функций на рассматриваемом интервале.

До сих пор предполагалось, что функциональное подпространство, в котором может быть представлена функция регрессии, задано. Но это обычно на практике не выполняется, и речь может идти лишь о том, чтобы аппроксимировать функцию регрессии функциями из задаваемого линейного подпространства. Выбор подпространства, к сожалению, формализовать не удастся, если не считать формализацией перебор всех подпространств, доступных для использования. Поэтому считается, что аппроксимирующее подпространство функций выбирается исследователем эмпирически. Однако некоторые соображения можно сформулировать. Например, если функция регрессии, определенная на ограниченном интервале, предполагается периодической, то можно рекомендовать аппроксимацию тригонометрическими полиномами. Отметим, что при аппроксимации на ограниченном интервале чаще других используются ортогональные функции, построенные с использованием полиномов Чебышева. После того, как функциональное пространство выбрано, необходимо уточнить его размерность. Для этого может быть использован подход, описанный в § 2.1. Оценка качества приближения, справедливая для любой случайной выборки, характеризуется выражением

$$J(q) = \frac{(\mathbf{x} - \Phi \hat{\mathbf{c}})^*(\mathbf{x} - \Phi \hat{\mathbf{c}})}{N - \sqrt{N \left(q \left(1 + \ln \frac{N}{q} \right) - \ln \alpha \right)}}, \quad (2.54)$$

где $\hat{\mathbf{c}}$ — оценка (2.46), основанная на (2.45) с учетом (2.37); N — объем выборки; q — размерность аппроксимирующего подпространства (размерность q выбирается такой, чтобы (2.54) принимала минимальную величину); α — уровень значимости при определении размерности аппроксимирующего подпространства.

Лабораторная работа 3. Изучение функций регрессии

Цель работы. Изучить возможность нахождения функций регрессии в виде разложения ее по набору линейно независимых или ортогональных функций.

Провести сравнительное исследование точности оценивания функции регрессии по различным базисам.

Порядок выполнения работы

Использовать те же реализации временных рядов, которые были получены в лабораторной работе 2, поскольку оценивание в данном случае осуществляется другими средствами и появляется возможность сравнивать качество работы этих подходов. Если реализацию временного ряда лабораторной работы сохранить не удалось, то временной ряд, который должен быть исследован в настоящей работе, получается таким же образом, как и в предыдущем случае.

При выполнении лабораторной работы более важным, чем получение оценки функции регрессии, является ее исследование. Поэтому в ходе выполнения заданий следует анализировать матрицы вариаций $\Phi^*\Phi$, которые по форме одинаковы в заданиях 1–5, но могут сильно отличаться по значениям при использовании различных базисов разложения. Необходимо провести анализ точности оценивания (например, по показателям Δ) в зависимости от отношения

$$\gamma = \left(\frac{1}{t_N - t_1} \int_{t_1}^{t_N} m_t^2 dt \right) / (D\{y_t\}).$$

Задание 1. Использовать полиномы Лежандра для оценивания функции регрессии исследуемого временного ряда. Найти

$$\Delta_p = \frac{1}{2} \int_{-1}^{+1} |\hat{m}_p(u) - m(u)| du,$$

где $\hat{m}_p(u)$ — оценка функции регрессии в виде разложения по полиномам Лежандра $P_k(u)$. Число членов разложения определить с помощью минимизации (2.54).

Задание 2. Использовать функции Чебышева для оценивания функции регрессии исследуемого временного ряда. Найти $\Delta_T = \frac{1}{2} \int_{-1}^{+1} |\hat{m}_T(u) - m(u)| du$. Число членов разложения определить с помощью минимизации (2.54).

Задание 3. Использовать тригонометрические функции для оценивания функции регрессии исследуемого временного ряда. Определить

$$\Delta_\Phi = \frac{1}{2} \int_{-1}^{+1} |\hat{m}_\Phi(u) - m(u)| du.$$

Число членов разложения найти с помощью минимизации (2.54).

Задание 4. Применяя функции Лаггера, оценить функцию регрессии используемого временного ряда. Число членов разложения найти с помощью минимизации (2.54). Определить качество оценивания интегралом

$$\Delta_L = \frac{1}{b-a} \int_a^b |\hat{m}_L(u) - m(u)| du,$$

где $[a, b]$ — интервал попадания точек наблюдения u_i .

Задание 5. Использовать функцию Эрмита для оценивания функции регрессии исследуемого временного ряда. Число членов разложения определить с помощью минимизации (2.54). Вычислить качество оценивания интегралом

$$\Delta_H = \frac{1}{b-a} \int_a^b |\hat{m}_H(u) - m(u)| du.$$

2.3. Оценивание функции регрессии.

Рекуррентные методы

Использование для определения функции регрессии оценок (2.46) в явной форме имеет негативные в вычислительном отношении аспекты, связанные с обращением матрицы $\Phi^* \Phi$. Во-первых, эта матрица может иметь довольно большую размерность и, во-вторых, может быть плохо обусловленной, что приводит к значительным вычислительным погрешностям и, как следствие, некачественному восстановлению функции регрессии. Существуют методы получения оценок (2.46) в вычислительном отношении более подходящие, так как они не предполагают обращения $\Phi^* \Phi$ в явной форме. Это *рекуррентные* методы или, как их часто называют, *адаптивные*.

Покажем, как можно придать рекуррентную форму вычислениям выборочных средних. Пусть

$$F(N) = \frac{1}{N} \sum_{t=1}^N f_t$$

является выборочным средним N некоторых наблюдений f_t . Нетрудно видеть, что этой формуле можно придать рекуррентную форму

$$F(N) = \left(1 - \frac{1}{N}\right) F(N-1) + \frac{1}{N} f_N; \quad F(1) = f_1. \quad (2.55)$$

Такая формула является удобной в том случае, когда наблюдения f_t поступают вычислительно последовательно и имеется возможность до получения очередного наблюдения обработать предыдущие. Тогда очередное наблюдение только корректирует выборочное среднее, вычисленное по предыдущим наблюдениям.

Происходит адаптация выборочного среднего к поступившему новому наблюдению. Такой способ обработки наблюдений удобен в тех случаях, когда необходимо использовать выборочное среднее еще до окончания процесса наблюдения, т. е. в процессе получения наблюдений. Эта ситуация часто встречается при реализации процессов управления. Кроме того, и в вычислительном плане такой способ удобен тем, что перерасчет выборочного среднего производится с использованием минимального объема памяти: не нужно накапливать все наблюдения, а достаточно помнить только предыдущее выборочное среднее. Наконец, использование рекуррентной формы выборочных средних удобно для реализации процесса имитационного моделирования в задачах исследования методов статистического анализа. При такой обработке данных не нужно организовывать больших массивов, а следует проводить обработку этих данных одновременно с их имитацией. Воспользуемся таким подходом для вычисления оценок (2.46).

Пусть $\Phi(n)$ означает матрицу, составленную из элементов $\Phi_{ij} = \varphi_j(u_i)$, $1 \leq j \leq q$, $1 \leq i \leq n$, т. е. $\Phi(n)$ — это матрица, на основе n первых наблюдений над временным рядом. С ростом n число строк матрицы $\Phi(n)$ увеличивается. Пусть, как и прежде, $\varphi^*(u_t) = (\varphi_1(u_t), \varphi_2(u_t), \dots, \varphi_q(u_t))$ — t -я строка матрицы $\Phi(n)$. Тогда будем иметь соотношения

$$\Phi(n+1) = \begin{pmatrix} \Phi(n) \\ \varphi^*(u_{n+1}) \end{pmatrix}; \quad \Phi^*(n+1) = (\Phi(n) \dot{\vdash} \varphi(u_{n+1})),$$

в связи с чем $\Phi^*(n)\Phi(n) = \Phi^*(n-1)\Phi(n-1) + \varphi(u_n)\varphi^*(u_n)$. Полезным для дальнейшего рассмотрения является следующее матричное равенство. Пусть A и B — квадратные невырожденные матрицы, не обязательно одинаковых размеров, а X , Y , Z — матрицы соответствующих размеров, связанные с A и B соотношением $Z = A + XBY$. Тогда $Z^{-1} = A^{-1} - A^{-1}X(B^{-1} + YA^{-1}X)^{-1}YA^{-1}$. Эта формула удобна для нахождения обратной матрицы Z^{-1} , когда A^{-1} можно задать, а B^{-1} — просто вычислить. Пусть в данном случае $Z = \Phi^*(n+1)\Phi(n+1)$, $A = \Phi^*(n)\Phi(n)$, $X = Y^* = \varphi(u_{n+1})$. Тогда B оказывается скалярным параметром, равным единице, $B = 1$. Отсюда имеем

$$\begin{aligned} (\Phi^*(n+1)\Phi(n+1))^{-1} &= (\Phi^*(n)\Phi(n))^{-1} - \\ &= \frac{(\Phi^*(n)\Phi(n))^{-1}\varphi(u_{n+1})\varphi^*(u_{n+1})(\Phi^*(n)\Phi(n))^{-1}}{1 + \varphi^*(u_{n+1})(\Phi^*(n)\Phi(n))^{-1}\varphi(u_{n+1})}. \end{aligned}$$

Обозначим для краткости

$$(\Phi^*(n)\Phi(n))^{-1} = \theta(n), \quad \psi(n+1) = \theta(n)\varphi(u_{n+1}). \quad (2.56)$$

Поэтому рекуррентная формула для вычисления последовательности обратных матриц $(\Phi^*(n)\Phi(n))^{-1} = \theta(n)$ приобретает наиболее простую форму

$$\theta(n+1) = \theta(n) - \frac{\psi(n+1)\psi^*(n+1)}{1 + \varphi^*(u_{n+1})\psi(n+1)}. \quad (2.57)$$

Обозначим через $\mathbf{x}(n)$ — вектор-столбец, составленный из первых n наблюдений временного ряда x_t , $1 \leq t \leq n$. Тогда оценка (2.46), полученная с использованием n первых наблюдений временного ряда, будет иметь вид

$$\begin{aligned}\hat{\mathbf{c}}(n) &= (\Phi^*(n)\Phi(n))^{-1}\Phi^*(n)\mathbf{x}(n) = \theta(n)(\Phi^*(n-1)\mathbf{x}(n-1) + x_n\boldsymbol{\varphi}(u_n)) = \\ &= \left(\theta(n-1) - \frac{\boldsymbol{\psi}(n)\boldsymbol{\psi}^*(n)}{1 + \boldsymbol{\varphi}^*(u_n)\boldsymbol{\psi}(n)} \right) (\Phi^*(n-1)\mathbf{x}(n-1) + x_n\boldsymbol{\varphi}(u_n)) = \\ &= \hat{\mathbf{c}}(n-1) + \frac{\boldsymbol{\psi}(n)}{1 + \boldsymbol{\varphi}^*(u_n)\boldsymbol{\psi}(n)}(x_n - \hat{\mathbf{c}}^*(n-1)\boldsymbol{\varphi}(u_n)).\end{aligned}\quad (2.58)$$

Эта формула получена с учетом того, что матрица $\theta(n)$, по определению, — симметрическая матрица. Заметим, что матрица $\theta(n)$ с точностью до постоянного множителя совпадает с матрицей ковариации оценки (2.46). Указанным множителем является дисперсия временного ряда, т. е. $V\{x_t\} = \mathbf{M}\{(x_t - m_t)^2\}$, где m_t определено в (2.34). Оценке этой дисперсии соответствует величина

$$\hat{V}_n = \frac{1}{n-q}(\mathbf{x}(n) - \Phi(n)\hat{\mathbf{c}}(n))^*(\mathbf{x}(n) - \Phi(n)\hat{\mathbf{c}}(n)).\quad (2.59)$$

Таким образом, процедура рекуррентного вычисления оценки (2.46) сводится к последовательному вычислению матрицы $\theta(n)$ и вектора $\hat{\mathbf{c}}(n)$ с помощью соотношений

$$\theta(n+1) = \theta(n) - \frac{\theta(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})\theta(n)}{1 + \boldsymbol{\varphi}^*(u_{n+1})\theta(n)\boldsymbol{\varphi}(u_{n+1})};\quad (2.60)$$

$$\hat{\mathbf{c}}(n+1) = \hat{\mathbf{c}}(n) + \theta(n+1)\boldsymbol{\varphi}(u_{n+1})(x_{n+1} - \hat{\mathbf{c}}^*(n)\boldsymbol{\varphi}(u_{n+1})).\quad (2.61)$$

При реализации процедуры вычисления по формулам (2.60), (2.61) удобно использовать вектор $\boldsymbol{\psi}(n+1) = \theta(n)\boldsymbol{\varphi}(u_{n+1})$.

Для того, чтобы описанная процедура была полностью определена, необходимо задание начальных условий. В общем случае для существования матрицы $\theta(n) = (\Phi^*(n)\Phi(n))^{-1}$ необходимо, чтобы $n \geq q$, и достаточно, чтобы $\Phi(n)$ содержала не менее q линейно независимых столбцов. В предположении, что первые q наблюдений обеспечивают линейную независимость векторов $\boldsymbol{\varphi}(u_i)$, $1 \leq t \leq q$, рекуррентные формулы (2.60), (2.61) могут быть использованы только при $n \geq q$. Поэтому процедуру оценивания коэффициентов $c(n)$ удобно осуществлять в два этапа. Для малых $n < q$ определим $(q \times q)$ -матрицу $H(n)$ следующими рекуррентными соотношениями:

$$\mathbf{H}_k(n) = \mathbf{H}_k(n-1) + \frac{\mathbf{H}_n(n-1)}{\boldsymbol{\varphi}^*(u_n)\mathbf{H}_n(n-1)}(\delta_{nk} - \boldsymbol{\varphi}^*(u_n)\mathbf{H}_k(n-1)),\quad (2.62)$$

$$1 \leq k \leq q,$$

где $\mathbf{H}_k(n)$ — k -й столбец матрицы $H(n)$, причем матрица $H(0) = I$, т.е. $H_{ij}(0) = \delta_{ij}$, δ_{ij} — символ Кронекера. Оценка вектора $\mathbf{c}(n)$ при этом вычисляется по формуле

$$\hat{\mathbf{c}}(n) = \hat{\mathbf{c}}(n-1) + \frac{\mathbf{H}_n(n-1)}{\boldsymbol{\varphi}^*(u_n)\mathbf{H}_n(n-1)}(x_n - \hat{\mathbf{c}}^*(n-1)\boldsymbol{\varphi}(u_n)). \quad (2.63)$$

Если существуют какие-либо априорные данные, позволяющие составить начальный вектор $\hat{\mathbf{c}}(0)$, то это начальное приближение может быть использовано в (2.63). Если же априорных данных относительно вектора $\hat{\mathbf{c}}$ нет, то полагаем $\hat{\mathbf{c}}(0) = 0$. Применив q раз формулы (2.62), (2.63), получим $\mathbf{H}(q)$ и $\hat{\mathbf{c}}(q)$. На этом этап малых n заканчивается. В качестве $\theta(q)$ выбирается $H(q)H^*(q)$, и дальнейшие вычисления для $n \geq q$ производятся по формулам (2.60), (2.61).

Может оказаться, что первые q наблюдений не порождают линейно независимые векторы $\boldsymbol{\varphi}(u_t)$. Тогда может быть использована общая процедура, которая сводится к следующему:

$$\hat{\mathbf{c}}(n+1) = \hat{\mathbf{c}}(n) + \mathbf{K}(n+1)(x_{n+1} - \hat{\mathbf{c}}^*(n)\boldsymbol{\varphi}(u_{n+1})), \hat{\mathbf{c}}(0) = 0, \quad (2.64)$$

где $\mathbf{K}(n+1)$ — вектор, который вычисляется по формулам

$$\mathbf{K}(n+1) = \begin{cases} \frac{A(n)\boldsymbol{\varphi}(u_{n+1})}{\boldsymbol{\varphi}^*(u_{n+1})A(n)\boldsymbol{\varphi}(u_{n+1})}, & \text{если } A(n)\boldsymbol{\varphi}(u_{n+1}) \neq 0, \\ \frac{B(n)\boldsymbol{\varphi}(u_{n+1})}{1 + \boldsymbol{\varphi}^*(u_{n+1})B(n)\boldsymbol{\varphi}(u_{n+1})}, & \text{если } A(n)\boldsymbol{\varphi}(u_{n+1}) = 0. \end{cases} \quad (2.65)$$

В (2.11) $A(n)$ и $B(n)$ — $(q \times q)$ -матрицы, определяемые по рекуррентным формулам:

$$A(n+1) = \begin{cases} A(n) - \frac{A(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})A(n)}{\boldsymbol{\varphi}^*(u_{n+1})A(n)\boldsymbol{\varphi}(u_{n+1})}, & \text{если } A(n)\boldsymbol{\varphi}(u_{n+1}) \neq 0, \\ A(n), & \text{если } A(n)\boldsymbol{\varphi}(u_{n+1}) = 0; \end{cases} \quad (2.66)$$

$$B(n+1) = B(n) - \frac{B(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})A(n) + A(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})B(n)}{\boldsymbol{\varphi}^*(u_{n+1})A(n)\boldsymbol{\varphi}(u_{n+1})} + \quad (2.67)$$

$$+ \frac{\boldsymbol{\varphi}^*(u_{n+1})B(n)\boldsymbol{\varphi}(u_{n+1})}{(\boldsymbol{\varphi}^*(u_{n+1})A(n)\boldsymbol{\varphi}(u_{n+1}))^2} A(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})A(n),$$

если $A(n)\boldsymbol{\varphi}(u_{n+1}) \neq 0$;

$$B(n+1) = B(n) - \frac{B(n)\boldsymbol{\varphi}(u_{n+1})\boldsymbol{\varphi}^*(u_{n+1})B(n)}{1 + \boldsymbol{\varphi}^*(u_{n+1})B(n)\boldsymbol{\varphi}(u_{n+1})}, \quad (2.68)$$

если $A(n)\varphi(u_{n+1}) = 0$.

Начальные значения матриц следующие: $A(0) = I$, $B(0) = 0$. Из формул (2.66)–(2.68) видно, что матрицы $A(n)$, $B(n)$ являются симметрическими. При реализации вычислительного процесса по формулам (2.65)–(2.68) удобно применять векторы $\psi_A(n+1) = A(n)\varphi(u_{n+1})$ и $\psi_B(n+1) = B(n)\varphi(u_{n+1})$.

Условие $A(n)\varphi(u_{n+1}) \neq 0$, используемое в (2.65)–(2.68), эквивалентно тому, что вектор $\varphi(u_{n+1})$ не является линейной комбинацией предыдущих векторов $\varphi(u_t)$, $1 \leq t \leq n$. Значит, это условие при вычислениях может встретиться не более q раз. Если первые q наблюдений порождают линейно независимые векторы $\varphi(u_t)$, то вычисление вектора $\mathbf{K}(n+1)$ и матриц $A(n+1)$ и $B(n+1)$ для первых q наблюдений осуществляется по первым формулам в (2.65)–(2.68), а для всех последующих наблюдений ($n > q$) – по вторым формулам, так как в этом случае для ($n \geq q$) всегда будет выполняться условие $A(n)\varphi(u_{n+1}) = 0$. При этом формула (2.68) совпадает с формулой (2.60) и необходимость в вычислении матрицы $A(n)$ исчезает, поскольку она в дальнейшем не используется при вычислении оценок $\hat{\mathbf{c}}(n)$. Таким образом, как только в процессе рекуррентных вычислений условие $A(n)\varphi(u_{n+1}) \neq 0$ выполнилось q раз, необходимость в дальнейшей проверке этого условия исчезает и вычисления оценок $\hat{\mathbf{c}}(n)$ проводятся с использованием только формулы (2.64), второй формулы (2.65) и формулы (2.68), т.е. процедура вычисления оценок полностью совпадает с вычислениями по формулам (2.60), (2.61). Вместе с тем на первом этапе при обработке линейно независимых векторов, если ими являются q первых $\varphi(u_t)$, использование (2.62), (2.63) значительно упрощает вычислительный процесс по сравнению с использованием (2.65)–(2.68). Совершая N итераций вычислений по формулам (2.60)–(2.63) или (2.65)–(2.68), получаем оценку коэффициентов \mathbf{c} функции регрессии (2.37) без обращения матрицы $\Phi^*\Phi$. Если размерность вектора \mathbf{c} невелика и указанная матрица хорошо обусловлена, то использование оценки (2.46) предпочтительнее. Применение рекуррентных оценок оправдано тогда, когда обращение матрицы $\Phi^*\Phi$ составляет проблему либо когда оценки функции регрессии необходимо использовать в процессе получения наблюдений, что часто имеет место при решении задач управления в обстановке априорной неопределенности.

До сих пор предполагалось, что наблюдения временного ряда x_t являлись независимыми случайными величинами. Рассмотрим теперь случай, когда они коррелированы.

Обозначим $y_t = x_t - m_t$. Предыдущий анализ относился к случаю, когда $R = (R_{ij}) = \sigma^2 I$,

$$R_{ij} = \mathbf{M}\{y_{t_i}y_{t_j}\} = \mathbf{M}\{(x_{t_i} - m_{t_i})(x_{t_j} - m_{t_j})\} = 0, \quad i \neq j. \quad (2.69)$$

Когда это условие не выполняется, способ определения коэффициентов \mathbf{c} в (2.37) должен быть модифицирован. Если ковариации (2.15) известны, то вместо

(2.41) эмпирический риск следует записывать в виде

$$\begin{aligned} J_{\Theta}(c) &= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (x(u_i) - \mathbf{c}^* \boldsymbol{\varphi}(u_i)) R_{ij}^{-1} (x(u_j) - \mathbf{c}^* \boldsymbol{\varphi}(u_j)) = \\ &= \frac{1}{N} (\mathbf{x} - \Phi \mathbf{c})^* R^{-1} (\mathbf{x} - \Phi \mathbf{c}), \end{aligned} \quad (2.70)$$

где R^{-1} — матрица с элементами R_{ij}^{-1} , а остальные элементы (2.70) определяются так же, как и в (2.45). Минимизация (2.70) по \mathbf{c} приводит к результату, аналогичному (2.22):

$$\hat{\mathbf{c}} = (\Phi^* R^{-1} \Phi)^{-1} \Phi^* R^{-1} \mathbf{x}. \quad (2.71)$$

Для того чтобы оценку $\hat{\mathbf{c}}$ получить рекуррентным методом, можно поступить следующим образом. Пусть $R(n) = (R_{ij})$, $1 \leq i, j \leq n$. Тогда имеет место соотношение

$$R(n) = \begin{pmatrix} R(n-1) & \mathbf{R}_n \\ \mathbf{R}_n^* & R_{nn} \end{pmatrix},$$

где $\mathbf{R}_n^* = (R_{n,1} R_{n,2} \dots R_{n,n-1})$. Обозначим

$$\delta_n^{-1} = \sqrt{R_{n,n} - \mathbf{R}_n^* R^{-1}(n-1) \mathbf{R}_n}, \quad \alpha(n) = -\delta_n \mathbf{R}_n^* R^{-1}(n-1). \quad (2.72)$$

Введем в рассмотрение последовательность матриц

$$D(1) = R_{11}^{-1/2}, \quad D(n) = \begin{pmatrix} D(n-1) & 0 \\ \alpha(n) & \delta_n \end{pmatrix}, \quad n = 2, 3, \dots,$$

и вектор $\mathbf{y}(n) = (y_{t_1}, y_{t_2}, \dots, y_{t_n})^*$. Окажется справедливым равенство $D(n)R(n)D^*(n) = I$. В этом случае преобразование $D(n)\mathbf{y}(n) = \tilde{\mathbf{y}}(n)$ декоррелирует вектор $\mathbf{y}(n)$, т.е. $\mathbf{M}\{\tilde{\mathbf{y}}(n)\tilde{\mathbf{y}}^*(n)\} = I$, что позволяет, преобразуя вектор наблюдений временного ряда $D(n)\mathbf{x}(n) = \tilde{\mathbf{x}}(n)$ и матрицу известных коэффициентов $D(n)\Phi(n) = \tilde{\Phi}(n)$, получать условия, в которых решалась задача ранее. Поэтому можно показать, что явные формулы для рекуррентного вычисления оценок коэффициентов регрессии \hat{c} в отличие от (2.60), (2.61) имеют вид:

$$\hat{\mathbf{c}}(n) = \hat{\mathbf{c}}(n-1) + \gamma(n)(\bar{x}_n - \bar{\varphi}^*(n)\hat{\mathbf{c}}(n-1)); \quad (2.73)$$

$$\gamma(n) = \frac{\theta(n-1)\bar{\varphi}(n)}{R_{nn} - \mathbf{R}_n^* R^{-1}(n-1) \mathbf{R}_n + \bar{\varphi}^*(n)\theta(n-1)\bar{\varphi}(n)}; \quad (2.74)$$

$$\theta(n) = \theta(n-1) - \frac{\theta(n-1)\bar{\varphi}(n)\bar{\varphi}^*(n)\theta(n-1)}{R_{nn} - \mathbf{R}_n^* R^{-1}(n-1) \mathbf{R}_n + \bar{\varphi}^*(n)\theta(n-1)\bar{\varphi}(n)}, \quad (2.75)$$

где обозначено

$$\bar{x}_n = x_n - \mathbf{R}_n^* R^{-1}(n-1) \mathbf{x}(n-1), \quad (2.76)$$

$$\bar{\varphi}^*(n) = \varphi^*(u_n) - \mathbf{R}_n^* R^{-1}(n-1) \Phi(n-1),$$

а обратная матрица $R^{-1}(n)$ может вычисляться рекуррентно:

$$R^{-1}(1) = \frac{1}{R_{11}},$$

$$R^{-1}(n) = \begin{pmatrix} R^{-1}(n-1) + \boldsymbol{\alpha}^*(n)\boldsymbol{\alpha}(n) & \delta_n \boldsymbol{\alpha}^*(n) \\ \delta_n \boldsymbol{\alpha}(n) & \delta_n^2 \end{pmatrix}, \quad (2.77)$$

где δ_n и $\boldsymbol{\alpha}(n)$ определены в (2.72).

Формулы (2.72)–(2.77) показывают, что наличие корреляции в наблюдениях временного ряда в большой степени усложняет проблему. В частности и потому, что, как видно из (2.73), (2.76), при вычислениях на n -й итерации необходимо использовать все наблюдения, а не только те, которые получены при n -м наблюдении, как было в (2.67). Правда, в некоторых случаях усложнения не очень существенные. Для иллюстрации этого факта рассмотрим случай марковской зависимости, когда временной ряд описывается следующей математической моделью:

$$x_t = m_t + y_t, \quad y_t = \rho y_{t-1} + \varepsilon_t,$$

$$\mathbf{M}\{\varepsilon_t\} = 0, \quad D\{\varepsilon_t\} = 1, \quad |\rho| < 1, \quad (2.78)$$

где ε_t — последовательность НОР СВ. В данном случае $R_{ij} = \rho^{|j-i|}$. Матрица, составленная из таких элементов, обращается аналитически. При этом

$$(R^{-1}(n))_{11} = (R^{-1}(n))_{nn} = (1 - \rho^2)^{-1},$$

$$(R^{-1}(n))_{kk} = (1 + \rho^2)/(1 - \rho^2), \quad 1 < k < n, \quad (2.79)$$

$$(R^{-1}(n))_{ij} = \frac{-\rho}{1 - \rho^2}, \quad |i - j| = 1; \quad (R^{-1}(n))_{ij} = 0, \quad |i - j| > 1.$$

Эти свойства обратной матрицы приводят к соотношениям

$$\mathbf{R}_n^* R^{-1}(n-1) = \rho(00 \dots 01), \quad R_{nn} R^*(n-1) \mathbf{R}_n = 1 - \rho^2,$$

$$\bar{x}_n = x_n - \rho x_{n-1}, \quad \bar{\varphi}(n) = \varphi(u_n) - \rho \varphi(u_{n-1}), \quad n = 2, 3, \dots, \quad (2.80)$$

и формулы (2.73)–(2.75) существенно упрощаются.

Мы рассмотрели несколько вариантов рекуррентного построения оценок коэффициентов регрессии. Сравнивая вид рекуррентных оценок, задаваемый выражениями (2.61), (2.63), (2.64), (2.73), можно обнаружить, что все они имеют структуру типа (2.73) и отличаются вектором $\boldsymbol{\gamma}(n)$, который в том или ином случае определяется соответствующим образом. Сама структура рекуррентных оценок напоминает структуру последовательного уточнения по типу стохастической аппроксимации (ТСА), которая была разработана для определения нулей и экстремумов неизвестной функции регрессии. Покажем, как могут быть использованы идеи стохастической аппроксимации для построения оценок коэффициентов регрессии.

Оценку ТСА введем соотношением

$$\hat{\mathbf{c}}(n) = \hat{\mathbf{c}}(n-1) + \gamma(n)(x_n - \varphi^*(u_n)\hat{\mathbf{c}}(n-1)), \quad (2.81)$$

в котором вектор $\gamma(n)$ определяется из условия минимизации вариаций оценок вектора $\hat{\mathbf{c}}(n)$. Отличие такой оценки от оценок, минимизирующих (2.41) или (2.70), в том, что в случае коррелированных наблюдений оценка (2.81) является более простой по сравнению с оценкой (2.73), так как для ее построения используется только последнее наблюдение, а не все предыдущие наблюдения. Иначе говоря, вместо $\bar{\mathbf{x}}_n$ и $\bar{\varphi}(n)$, которые задаются (2.76), используются \mathbf{x}_n и $\varphi(u_n)$ соответственно. В случае некоррелированных наблюдений требование минимизации вариаций обеспечивается применением оценок МНК (2.61), (2.64), так что оценка (2.81) будет совпадать с ними. Описанный принцип приводит к тому, что вектор $\gamma(n)$ определяется по формуле

$$\gamma(n) = \frac{G(n-1)\varphi(u_n) - \lambda(n)}{R_{nn} - 2\varphi^*(u_n)\lambda(n) + \varphi^*(u_n)G(n-1)\varphi(u_n)}, \quad (2.82)$$

где $G(n)$ — матрица вариации оценки (2.81), которая вычисляется рекуррентно по формулам

$$G(0) = I, \quad G(n) = (I - \gamma(n)\varphi^*(u_n))G(n-1) + \gamma(n)\lambda^*(n); \quad (2.83)$$

$\lambda(n)$ — вектор, определяемый по формулам

$$\lambda(1) = 0, \quad \lambda(n) = B(n-1)R_n, \quad n > 1. \quad (2.84)$$

Матрица $B(n)$ находится рекуррентно по формулам

$$B(1) = \gamma(1), \quad B(n) = ((I - \gamma(n)\varphi^*(u_n))B(n-1) + \gamma(n)), \quad (2.85)$$

где $B(n)$ — матрица размером $(q \times n)$.

Лабораторная работа 4. Рекуррентные методы оценивания функции регрессии

Цель работы. Исследовать процессы последовательного оценивания коэффициентов функции регрессии. Установить скорость сходимости к истинным значениям.

Порядок выполнения работы

Использовать реализацию временного ряда (рассмотренного в лабораторных работах 2 и 3), для которой параметры функции регрессии уже определены другими методами, выбрано множество точек наблюдения $\{u_i\}$ и функциональный базис разложения функции регрессии $\{y_j(u), 1 \leq j \leq q\}$.

Задание 1. Используя рекуррентную формулу (2.66), исследовать, как часто возникает ситуация, когда q последовательных точек наблюдения $\{u_1, u_2, \dots, u_q\}$ временного ряда порождают линейно независимые векторы $\varphi(u_k)$. Для этого выбрать некоторое n_0 и, взяв u_{n_0} за начальную точку наблюдения, по формуле (2.66) последовательно вычислять матрицы $A(n_0 + k)$, $k = 1, 2, \dots$, при $A(n_0) = I$. Затем, проверяя выполнение неравенства $A(n_0 + k)\varphi(u_{n_0+k}) \neq 0$, определить такое k_0 , при котором указанное неравенство выполнится q раз для различных n_0 на непересекающихся множествах точек наблюдения u_i .

Задание 2. Построить процедуру вычисления оценок $\hat{c}(n)$ по формулам (2.62), (2.63) для первых q наблюдений и по формулам (2.60), (2.61) — для последующих наблюдений. Устанавливая скорость сходимости, параллельно вычислить $\text{tr} \theta(n)$, $n \geq q$, и $\sqrt{(\hat{c}(n) - \hat{c})^*(\hat{c}(n) - \hat{c})}$, где \hat{c} — оценка коэффициентов регрессии, полученная для исследуемой реализации при выполнении лабораторной работы 3.

Задание 3. Построить процедуру вычисления оценок $\hat{c}(n)$ по формулам (2.64)—(2.68). Провести исследование скорости сходимости по той же схеме, как и в задании 2.

Задание 4. Исследовать влияние корреляции наблюдений и по формулам (2.60), (2.61) осуществить процедуру построения оценок $\hat{c}(n)$ для последующих наблюдений. Для этого по формуле (2.62) и последовательности точек наблюдения $\{u_1, u_2, \dots, u_q\}$ определить матрицы $H(q)$ и $\theta(q) = H(q)H^*(q)$. Затем по формуле (2.75) вычислить последовательность матриц вариаций $\theta(n)$, $n > q$, предположив справедливыми соотношения (2.79), (2.80), которые характеризуют случай наблюдений, образующих марковский процесс. Для контроля скорости сходимости параллельно вычислять $\text{tr} \theta(n)$. Повторить эту процедуру для нескольких значений $\rho \in (0, 1)$. Выяснить влияние параметра корреляции ρ на поведение $\text{tr} \theta(n)$.

Задание 5. Исследовать качество алгоритма оценивания по типу стохастической аппроксимации при коррелированных наблюдениях. Для этого по формулам (2.82)—(2.85) в предположении справедливости соотношений (2.79), (2.80) вычислить последовательность матриц $H(n)$, $n = 1, 2, \dots$. Исследовать поведение $\text{tr} G(n)$ для различных ρ , принимающих те же значения, как и в задании 4.

2.4. Метод максимального правдоподобия. М-оценки

Будем предполагать, что $y_t = x_t - m_t$, $t \in T = \{t_1, t_2, \dots, t_N\}$, образуют последовательность НОР СВ с заданной плотностью вероятности $p(y)$. Функция регрессии m_t представляется в виде разложения по некоторому базису $\{\varphi_j(u)\}$ на интервале изменения унифицированной переменной u (см.(2.35)):

$$m_t = \sum_{j=1}^q c_j \varphi_j(u) = \mathbf{c}^* \boldsymbol{\varphi}(u).$$

Для простоты будем считать $x_{t_i} = x_i$.

Функция правдоподобия определяется как функция неизвестных параметров \mathbf{c} вида

$$L(\mathbf{c}) = \prod_{i=1}^N p(y_{t_i}) = \prod_{i=1}^N p(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)),$$

где $p(y)$ — плотность вероятности. Удобнее работать с логарифмом этой функции

$$\log L(\mathbf{c}) = \sum_{i=1}^N \log p(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)) = - \sum_{i=1}^N \rho(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)). \quad (2.86)$$

Оценками параметра \mathbf{c} по методу максимального правдоподобия (оценками ММП) называются такие \mathbf{c} , которые максимизируют $\log L(\mathbf{c})$, т.е.

$$\hat{\mathbf{c}}_{\text{ММП}} = \arg \min_{\mathbf{c}} \sum_{i=1}^N \rho(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)). \quad (2.87)$$

Если функция $\rho(\cdot)$ дифференцируема по \mathbf{c} , то максимизация (2.86) по \mathbf{c} соответствует решению системы уравнений

$$\sum_{i=1}^N \boldsymbol{\varphi}(u_i) \psi(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)) = 0, \quad (2.88)$$

где через $\psi(\cdot)$ обозначена производная функции $\rho(\cdot)$.

Заметим, что если y_t — нормально распределенные величины с нулевыми средними и дисперсией σ^2 , то

$$\ln p(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)) = \text{const} - \frac{1}{2\sigma^2} (x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i))^2$$

и максимизация (2.86) соответствует минимизации

$$\sum_{i=1}^N \left(\frac{x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)}{\sigma} \right)^2, \quad (2.89)$$

т.е. минимизация эмпирического риска (2.41) и получаемая оценка ММП не отличаются от оценки МНК. Таким образом, в случае нормального распределения y_t оценка ММП имеет исследованные ранее свойства. В общем случае уравнение (2.88) не может быть решено в явной форме. При довольно слабых предположениях, известных под названием *условий регулярности*, доказывається, что при $N \rightarrow \infty$ оценки ММП являются асимптотически несмещенными, асимптотически эффективными и распределенными асимптотически нормально.

Примеры для некоторых распределений. Распределение Лапласа:

$$p(y) = \frac{1}{2a} e^{-|y|/a}, \quad \rho(y) = \ln(2a) + \frac{|y|}{a}, \quad \psi(y) = \frac{1}{a} \text{sign } y.$$

При этом уравнение (2.88) имеет вид

$$\sum_{i=1}^N \varphi(u_i) \operatorname{sign}(x_i - \mathbf{c}^* \varphi(u_i)) = 0. \quad (2.90)$$

Равномерное распределение в интервале $[-a, +a]$

$$p(y) = \begin{cases} 1/2a, & |y| \leq a, \\ 0, & |y| > a, \end{cases} \quad \rho(y) = \begin{cases} \log(2a), & |y| \leq a, \\ \infty, & |y| > a. \end{cases}$$

В этом случае решение уравнения (2.88) сводится к решению системы неравенств

$$|x_i - \mathbf{c}^* \varphi(u_i)| \leq a, \quad 1 \leq i \leq N. \quad (2.91)$$

Оценкой ММП является любой вектор \mathbf{c} , который удовлетворяет всем этим неравенствам.

Распределение Коши:

$$p(y) = \frac{1}{\pi} \frac{a}{a^2 + y^2}, \quad \rho(y) = \log \frac{\pi}{a} + \log(a^2 + y^2), \quad \psi(y) = \frac{2y}{a^2 + y^2}.$$

При этом уравнение (2.88) сводится к следующему:

$$\sum_{i=1}^N \frac{(x_i - \mathbf{c}^* \varphi(u_i)) \varphi(u_i)}{a^2 + (x_i - \mathbf{c}^* \varphi(u_i))^2} = 0. \quad (2.92)$$

Из (2.90)–(2.92) видно, что уравнения максимального правдоподобия для коэффициентов регрессии относятся к различным классам уравнений и не могут быть решены какими-либо стандартными методами, кроме численных.

Численные методы решения уравнения (2.88). *Метод Гаусса – Ньютона.* Пусть $\mathbf{c}(n)$ – оценка на n -й итерации. Предположим, что $\psi(\cdot)$ в точках остаточных разностей наблюдения имеет производную. Используем представление

$$\begin{aligned} \psi(x_i - \mathbf{c}^*(n+1)\varphi(u_i)) &= \psi(x_i - \mathbf{c}^*(n)\varphi(u_i)) - \\ &- \varphi^*(u_i)(\mathbf{c}(n+1) - \mathbf{c}(n))\psi'(x_i - \mathbf{c}^*(n)\varphi(u_i)) + \dots, \end{aligned} \quad (2.93)$$

где $\psi'(\cdot)$ – производная $\psi(\cdot)$ по ее аргументу. Ограничиваясь только приведенными слагаемыми представления (2.93), уравнение (2.88) перепишем в виде

$$\begin{aligned} &\sum_{i=1}^N \varphi(u_i) \psi(x_i - \mathbf{c}^*(n)\varphi(u_i)) - \\ &- \left(\sum_{i=1}^N \varphi(u_i) \varphi^*(u_i) \psi'(x_i - \mathbf{c}^*(n)\varphi(u_i)) \right) (\mathbf{c}(n+1) - \mathbf{c}(n)) = 0. \end{aligned}$$

Из полученного уравнения находится рекуррентная формула последовательного уточнения оценки ММП

$$\begin{aligned} \mathbf{c}(n+1) = \mathbf{c}(n) + \left(\sum_{i=1}^N \varphi(u_i) \varphi^*(u_i) \psi'(x_i - \mathbf{c}^*(n) \varphi(u_i)) \right)^{-1} \times \\ \times \sum_{i=1}^N \varphi(u_i) \psi(x_i - \mathbf{c}^*(n) \varphi(u_i)). \end{aligned} \quad (2.94)$$

В качестве начального приближения можно брать оценку МНК (2.46)

$$\hat{\mathbf{c}}(0) = (\Phi^* \Phi)^{-1} \Phi^* \mathbf{x} = \left(\sum_{i=1}^N \varphi(u_i) \varphi^*(u_i) \right)^{-1} \sum_{i=1}^N \varphi(u_i) x_i. \quad (2.95)$$

Взвешенный МНК. Перепишем уравнение (2.88) следующим образом:

$$\sum_{i=1}^N \frac{\psi(x_i - \mathbf{c}^* \varphi(u_i))}{x_i - \mathbf{c}^* \varphi(u_i)} \varphi(u_i) (x_i - \varphi(u_i)^* \mathbf{c}) = 0.$$

На основе такого представления имеем

$$\sum_{i=1}^N \frac{\psi(x_i - \mathbf{c}^*(n) \varphi(u_i))}{x_i - \mathbf{c}^*(n) \varphi(u_i)} (\varphi(u_i) x_i - \varphi(u_i) \varphi(u_i)^* \mathbf{c}(n+1)) = 0,$$

откуда и получается рекуррентная формула

$$\mathbf{c}(n+1) = \left(\sum_{i=1}^N W_i(\mathbf{c}(n)) \varphi(u_i) \varphi^*(u_i) \right)^{-1} \sum_{i=1}^N W_i(\mathbf{c}(n)) \varphi(u_i) x_i, \quad (2.96)$$

в которой для краткости обозначено

$$W_i(\mathbf{c}) = \frac{\psi(x_i - \mathbf{c}^* \varphi(u_i))}{x_i - \mathbf{c}^* \varphi(u_i)}.$$

Равенства (2.95), (2.96) определяют последовательность приближений к оценке ММП.

Оценки ММП имеют и недостатки, заключающиеся в том, что эти оценки являются чувствительными к отклонениям распределения от предполагаемого. В массивах данных, отражающих наблюдения временного ряда, могут встречаться ошибки, которые порождаются либо сбоями, либо погрешностями в записи или при измерении. В данном случае оценки ММП, а также и оценки МНК могут достаточно сильно отличаться от истинных значений параметров регрессии. В связи с этим существует обобщения оценок ММП, которые образуют

класс М-оценок и формулируются таким образом, чтобы обеспечить более высокий уровень устойчивости по отношению к несоответствию выборочного массива данных предполагаемому распределению. Поскольку такое несоответствие обычно имеет место для небольшого числа наблюдений, говорят о том, что в выборке есть выбросы или что выборка "загрязнена". Имеются различные математические модели такого загрязнения выборки. Модификация оценок ММП выглядит следующим образом. В качестве М-оценки принимается (2.87), в которой функция $\rho(y) \neq -\log p(y)$, как это предполагалось в (2.86). Собственно выбор функции $\rho(\cdot)$ и определяет характер М-оценки. Обычно функция ρ выбирается дифференцируемой, так что и М-оценка находится из уравнения (2.88), в котором $\psi(\cdot)$ определяется соответствующим образом. Само уравнение (2.88) при этом в модифицированном виде выглядит так:

$$\sum_{i=1}^N \varphi(u_i) \psi \left(\frac{x_i - \mathbf{c}^* \varphi(u_i)}{s} \right) = 0, \quad (2.97)$$

где s — помехоустойчивая оценка параметра масштаба, о которой речь пойдет ниже.

Примеры выбора функции $\psi(\cdot)$. Функция Хьюбера:

$$\psi(y) = \begin{cases} y, & |y| \leq a, \\ a \operatorname{sign} y, & |y| > a. \end{cases} \quad (2.98)$$

Параметр a обычно выбирается из интервала $[1, 2]$.

Функция Хампеля:

$$\psi(y) = \begin{cases} |y|, & 0 \leq |y| < a, \\ a, & a \leq |y| < b, \\ a(c - |y|)/(c - b), & b \leq |y| < c, \\ 0, & c \leq |y|. \end{cases} \quad (2.99)$$

Рекомендуемыми значениями параметров являются: $a = 1,7$; $b = 3,4$; $c = 8,5$.

Функция Андрюса:

$$\psi(y) = \begin{cases} \sin(y/a), & |y| \leq a\pi, \\ 0, & |y| > a\pi, \end{cases} \quad (2.100)$$

где обычно $a = 2,1$.

Биквадратная функция Тьюки:

$$\psi(y) = \begin{cases} y(1 - (y/a)^2)^2, & |y| \leq a, \\ 0, & |y| > a, \end{cases} \quad (2.101)$$

где обычно $a = 6,0$.

В уравнении (2.97) чаще других используются следующие оценки параметра масштаба s . Пусть Q_α — квантиль выборочного распределения $\{x_i\}$, т.е. значение аргумента выборочной функции распределения значений временного ряда,

при котором эта функция принимает значение α . Для одних случаев в качестве помехоустойчивой оценки параметра масштаба может служить величина

$$s = (\text{med}|x_t - Q_{0,5}|)/0,6745, \quad (2.102)$$

где $\text{med}|x_t - Q_{0,5}|$ означает медиану вариационного ряда, составленного из $|x_t - Q_{0,5}|$; знаменатель 0,6745 является 75%-м квантилем стандартного нормального распределения. Для других случаев помехоустойчивого оценивания масштаба

$$s = (Q_{0,75} - Q_{0,25})/1,349. \quad (2.103)$$

Если значения $\{x_i\}$ подчиняются нормальному распределению, то s — оценка среднеквадратичного отклонения значений x_i .

Численное решение уравнения (2.97) может быть осуществлено при помощи итерационных методов (2.94), (2.95) или (2.95), (2.96) с небольшой модификацией, связанной с появлением оценки s в функции $\psi(\cdot)$. Оценка вычисляется по формуле

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \left(\sum_{i=1}^N \tilde{W}_i(\mathbf{c}(n)) \varphi(u_i) \varphi^*(u_i) \right)^{-1} \sum_{i=1}^N W_i(\mathbf{c}(n)) \varphi(u_i), \quad (2.104)$$

где в случае метода Гаусса — Ньютона

$$\tilde{W}_i(\mathbf{c}) = \frac{1}{s} \psi' \left(\frac{x_i - \mathbf{c}^* \varphi(u_i)}{s} \right), \quad W_i(\mathbf{c}) = \psi \left(\frac{x_i - \mathbf{c}^* \varphi(u_i)}{s} \right),$$

а в случае взвешенного МНК

$$\tilde{W}_i(\mathbf{c}) = \frac{1}{x_i - \mathbf{c}^* \varphi(u_i)} \psi \left(\frac{x_i - \mathbf{c}^* \varphi(u_i)}{s} \right), \quad W_i(\mathbf{c}) = \tilde{W}_i(\mathbf{c}) x_i.$$

Лабораторная работа 5. Оценивание функции регрессии. Метод максимального правдоподобия

Цель работы. Освоить технику оценивания коэффициентов регрессии при помощи ММП. Провести сравнение качества оценок ММП и оценок МНК.

Порядок выполнения работы

Выбрать некоторый базис $\{\varphi_j(u)\}$ и задать с его помощью функцию регрессии $m(u)$. Затем выбрать распределение ошибок наблюдения временного ряда необходимой длины. Установить множество точек наблюдения $\{u_i\}$ и получить реализацию временного ряда.

Поскольку при выполнении настоящей работы потребуется генерирование случайных чисел, распределенных в соответствии с различными плотностями вероятностей, приведем способы их получения. Предполагается, что в используемой ЭВМ имеется датчик случайных чисел ε_t , распределенных равномерно в интервале $[0,1]$.

СВ ξ_t , равномерно распределенная в интервале $[a, b]$, получается преобразованием

$$\xi_t = (b - a)\varepsilon_t + a.$$

СВ ξ_t , распределенная в соответствии с плотностью экспоненциального распределения

$$p(z) = \lambda e^{-\lambda z}, \quad z \geq 0,$$

получается преобразованием

$$\xi_t = -\frac{1}{\lambda} \ln \varepsilon_t.$$

СВ ξ_t , распределенная в соответствии с плотностью нормального распределения

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-a)^2}{2\sigma^2}}, \quad z \in (-\infty, +\infty),$$

получается преобразованием (практическое приближение)

$$\xi_t = \sigma \left(\sum_{j=1}^{12} \varepsilon_{jt} - 6 \right) + a.$$

СВ ξ_t , распределенная в соответствии с распределением Коши

$$p(z) = \frac{b}{\pi(b^2 + (z-a)^2)}, \quad z \in (-\infty, +\infty),$$

получается преобразованием

$$\xi_t = a + b \operatorname{tg}(2\pi\varepsilon_t).$$

СВ ξ_t , распределенная по закону Лапласа с плотностью

$$p(z) = \lambda e^{-\lambda|z|}, \quad z \in (-\infty, +\infty),$$

получается преобразованием

$$\xi_t = -\frac{1}{\lambda} \operatorname{sign} \left(\varepsilon_{1t} - \frac{1}{2} \right) \ln \varepsilon_{2t}.$$

СВ ξ_t , распределенная по закону Вейбулла с плотностью

$$p(z) = a\lambda(\lambda z)^{a-1}\exp\{-(\lambda z)^a\}, \quad z \in (0, \infty),$$

получается преобразованием

$$\xi_t = \frac{1}{\lambda}(-\ln \varepsilon_t)^{1/a}.$$

СВ ξ_t , распределенная в соответствии с логистическим распределением

$$p(z) = \frac{\exp\left\{-\frac{z-a}{\sigma}\right\}}{\sigma(1 + \exp\left\{-\frac{z-a}{\sigma}\right\})^2}, \quad z \in (-\infty, +\infty),$$

получается преобразованием

$$\xi_t = a + \sigma \ln \frac{\varepsilon_t}{1 - \varepsilon_t}.$$

СВ ξ_t , распределенная по закону Эрланга с плотностью

$$p(z) = \frac{\lambda(\lambda z)^{a-1}}{(a-1)!}e^{-\lambda z}, \quad z \in (0, +\infty),$$

получается преобразованием

$$\xi_t = -\frac{1}{\lambda} \ln \left(\prod_{i=1}^a \varepsilon_{it} \right) = -\frac{1}{\lambda} \sum_{i=1}^a \ln \varepsilon_{it}.$$

Задание. После того, как выбрано распределение ошибок и получена реализация временного ряда, определить функцию $\psi(\cdot)$ для этого распределения, составить уравнение (2.88) с использованием этой функции и решить его методом Гаусса — Ньютона и взвешенного МНК. Результаты $\hat{c}_{ГН}$ и $\hat{c}_{ВМНК}$ сравнить между собой и с оценками МНК. При помощи функций Хьюбера, Хампеля, Андриуса и Тьюки построить М-оценки коэффициентов регрессии. Сравнить все полученные оценки с истинными значениями коэффициентов регрессии.

2.5. Устойчивые процедуры оценивания параметров регрессии

Классические оценки ММП и МНК представляют собой примеры идеализированного подхода к решению практических задач оценивания коэффициентов регрессии, когда предполагается, что все условия строго выполнены. В то же время в реальных данных нередко встречаются грубые ошибки, которые существенно влияют на качество оценок, приводя к потере эффективности, увеличению смещения оценок, потере их состоятельности. При оценке регрессионных зависимостей эффект даже одной грубой ошибки может быть очень сильным. Поэтому применение классических методов должно осуществляться после

тщательной отбраковки грубых ошибок. Вместе с тем большие массивы обрабатываемых данных не позволяют делать это тщательно. Поэтому существует потребность в использовании таких методов обработки данных, которые зависят в небольшой степени от наличия указанных дефектов в обрабатываемых массивах. Такие методы известны под названием устойчивых, робастных или огрубленных. В § 2.4 уже была рассмотрена модификация ММП, обладающая чертами устойчивости. В данном параграфе эти и другие оценки используются для обработки данных, засоренных грубыми выбросами.

Применим следующую модель "засорения". Пусть плотность вероятностей остатков $y_t = x_t - m_t$ имеет вид

$$p_\varepsilon(y) = (1 - \varepsilon)p(y) + \varepsilon h(y), \quad (2.105)$$

где ε — параметр "засорения", $0 < \varepsilon < 1$; $p(y)$ — плотность вероятностей теоретического распределения, для которого строится М-оценка; $h(y)$ — плотность вероятностей "засоряющего" распределения.

Для случая, когда $p(y)$ — нормальное распределение с нулевым средним и неизвестной дисперсией, а $h(y)$ — произвольная, но симметричная относительно нулевого значения плотность, Хьюбером была получена функция (2.98), используемая в уравнении (2.97), которая определяет оценки параметров регрессии \mathbf{c} . Поскольку в рассматриваемом случае предполагаются неизвестными вектор \mathbf{c} и параметр масштаба (среднеквадратичное отклонение) σ , для их определения используется два уравнения:

$$\sum_{i=1}^N \varphi(u_i) \psi \left(\frac{x_i - \mathbf{c}^* \varphi(u_i)}{\sigma} \right) = 0; \quad (2.106)$$

$$\sigma = (\text{med } |x_i - \mathbf{c}^* \varphi(u_i)|) / 0,6745. \quad (2.107)$$

Медиана выборки находится просто. Пусть $\{z_i\}$ — некоторая выборка объемом N . Упорядочим ее (ранжируем), получив вариационный ряд $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(N)}$. Тогда

$$\text{med } z_i = \begin{cases} z_{(m)}, & \text{если } N = 2m - 1, \\ (z_{(m+1)} + z_{(m)}) / 2, & \text{если } N = 2m. \end{cases} \quad (2.108)$$

Поэтому решением (2.107) можно считать (2.108).

Определим процедуру решения уравнения (2.106) для конкретного вида функции $\psi(\cdot)$, заданного формулой (2.98).

Пусть на n -й итерации получены оценки $\mathbf{c}(n)$ и $\sigma(n)$ неизвестных параметров. Образует следующие множества индексов:

$$\begin{aligned} H &= \{i \mid x_i - \mathbf{c}^*(n) \varphi(u_i) < -k\sigma(n)\}, \\ C &= \{i \mid |x_i - \mathbf{c}^*(n) \varphi(u_i)| \leq k\sigma(n)\}, \\ B &= \{i \mid x_i - \mathbf{c}^*(n) \varphi(u_i) > k\sigma(n)\}. \end{aligned} \quad (2.109)$$

Для любого n выполняется равенство $|H| + |B| + |C| = N$, где N — объем выборки. Число k зависит от параметра засорения ε , удовлетворяя уравнению

$$\frac{\varepsilon k}{1 - \varepsilon} = \frac{2}{\sqrt{2\pi}} \left(e^{-k^2/2} - k \int_k^\infty e^{-t^2/2} dt \right).$$

Таблица 2.1

ε	k	ε	k
0,000	∞	0,20	0,826
0,001	2,630	0,25	0,766
0,002	2,435	0,30	0,685
0,005	2,160	0,40	0,550
0,01	1,945	0,50	0,436
0,02	1,717	0,65	0,291
0,05	1,309	0,80	0,162
0,10	1,140	1,00	0,000
0,15	0,980		

В табл. 2.1 приведены корни этого уравнения для некоторых значений ε . Используя формулу (2.98), определяющую функцию $\psi(\cdot)$, уравнение (2.106) можно конкретизировать:

$$\sum_{i \in C} (x_i - \mathbf{c}^*(n+1)\varphi(u_i))\varphi(u_i) + \left(\sum_{i \in B} \varphi(u_i) - \sum_{i \in H} \varphi(u_i) \right) k\sigma(n) = 0,$$

где $\sum_{i \in C}$, $\sum_{i \in B}$, $\sum_{i \in H}$ — означают, что суммирование ведется по индексам i множеств C, B, H соответственно. Из этого уравнения получаем

$$\begin{aligned} \mathbf{c}(n+1) &= \left(\sum_{i \in C} \varphi(u_i)\varphi^*(u_i) \right)^{-1} \times \\ &\times \left(\sum_{i \in C} x_i\varphi(u_i) + \left(\sum_{i \in B} \varphi(u_i) - \sum_{i \in H} \varphi(u_i) \right) k\sigma(n) \right). \end{aligned} \quad (2.110)$$

Уравнение (2.107) запишем в виде

$$\sigma(n) = (\text{med } |x_i - \mathbf{c}^*(n)\varphi(u_i)|)/0,6745. \quad (2.111)$$

В качестве нулевого приближения для \mathbf{c} можно принять оценку МНК, которая соответствует тому, что множества $B = H = \emptyset$:

$$\mathbf{c}(0) = \left(\sum_{i=1}^N \varphi(u_i)\varphi^*(u_i) \right)^{-1} \sum_{i=1}^N x_i\varphi(u_i). \quad (2.112)$$

Процедура вычислений сводится к следующему.

Шаг 1. Вычисляется $\mathbf{c}(n)$ по формуле (2.110) (для $n = 0$ по (2.112)).

Шаг 2. Вычисляется $\sigma(n)$ по формуле (2.111).

Шаг 3. Определяются множества $H(n)$, $C(n)$, $B(n)$ по формулам (2.109).

Шаг 4. Если множества $H(n)$, $C(n)$, $B(n)$ не отличаются от предыдущих $H(n-1)$, $C(n-1)$, $B(n-1)$ соответственно, то вычисления завершаются и решением признается пара $\mathbf{c}(n)$, $\sigma(n)$. Если же отличие наблюдается, то значение n увеличивается на единицу и выполняется снова шаг 1.

При выполнении этой процедуры необходимо следить за тем, чтобы не происходило заикливания при образовании множеств H , C , B .

Данная процедура существенным образом базировалась на представлении исходной плотности вероятностей ошибок (2.105), ориентированной на нормальное распределение и известный параметр засорения ε . На практике эта информация либо неизвестна исследователю, либо исходное распределение существенно отличается от (2.105). В этих случаях можно руководствоваться рекомендациями, приведенными при описании формулы (2.98).

Приведем еще два способа вычисления оценок $\mathbf{c}(n)$.

Процедура, использующая модифицированные остатки. Предположим, что на n -й итерации получены оценки $\mathbf{c}(n)$ и $\sigma(n)$ неизвестных параметров. Определим

$$r_i(n) = \sigma(n)\psi\left(\frac{x_i - \mathbf{c}^*\varphi(u_i)}{\sigma(n)}\right) = 0, \quad (2.113)$$

где $\psi(\cdot)$ — функция, заданная формулой (2.98). Решим задачу минимизации относительно вектора $\boldsymbol{\tau}$:

$$\sum_{i=1}^N (\mathbf{r}_i(n) - \boldsymbol{\tau}^* \varphi(u_i))^2 \rightarrow \min_{\boldsymbol{\tau}}.$$

Решение этой задачи следующее:

$$\boldsymbol{\tau}(n) = (\Phi^* \Phi)^{-1} \Phi^* \mathbf{r}(n) = \left(\sum_{i=1}^N \varphi(u_i) \varphi^*(u_i) \right)^{-1} \sum_{i=1}^N \varphi(u_i) \mathbf{r}_i(n). \quad (2.114)$$

В качестве оценки $\mathbf{c}(n+1)$ принимаем

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \delta \boldsymbol{\tau}(n), \quad (2.115)$$

где δ — произвольный множитель, такой, что $0 < \delta < 2$. Оценки $\sigma(n)$ находятся по формуле (2.111).

В качестве $\mathbf{c}(0)$ принимается, как и прежде, оценка МНК, вычисляемая по формуле (2.112). Процедура выполняется до достижения необходимой точности.

Процедура, использующая модифицированные веса. Предположим, что на n -й итерации получены оценки $\mathbf{c}(n)$ и $\sigma(n)$. Определим

$$W_i(n) = \psi \left(\frac{x_i - \mathbf{c}^*(n)\varphi(u_i)}{\sigma(n)} \right) / \left(\frac{x_i - \mathbf{c}^*(n)\varphi(u_i)}{\sigma(n)} \right), \quad (2.116)$$

$W(n) = \text{diag}\{W_i(n), 1 \leq i \leq N\}$ — диагональная матрица, по главной диагонали которой стоят $W_i(n)$. Пусть вектор $\boldsymbol{\tau}(n)$ задается соотношением

$$\boldsymbol{\tau}(n) = (\Phi^*W(n)\Phi)^{-1}\Phi^*W(n)(\mathbf{x} - \Phi^*\mathbf{c}(n)). \quad (2.117)$$

В качестве оценки $\mathbf{c}(n+1)$ принимаем

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \boldsymbol{\tau}(n). \quad (2.118)$$

Оценки $\sigma(n)$ вычисляются, как и прежде, по формуле (2.111).

Завершая описание процедур, основанных на М-оценках, приведем еще несколько подходов.

Процедура Андрюса. Пусть на n -й итерации получена оценка $\mathbf{c}(n)$. Оценка $\sigma(n+1)$ определяется как медиана

$$\sigma(n+1) = \text{med} |x_i - \mathbf{c}^*(n)\varphi(u_i)|. \quad (2.119)$$

Оценка $\mathbf{c}(0)$ вычисляется по формуле (2.112), а $\mathbf{c}(n+1)$ из условия

$$\mathbf{c}(n+1) = \arg \min_c \sum_{i=1}^N \varphi(u_i) \psi \left(\frac{x_i - \mathbf{c}^*\varphi(u_i)}{\sigma(n+1)} \right), \quad (2.120)$$

где $\psi(\cdot)$ задана формулой (2.90). После этого вычисления выполняются до достижения необходимой точности. Этот метод достаточно эффективен в ситуациях "тяжелого засорения" при наличии значительного числа очень грубых ошибок.

Процедура Форсайта. Оценка параметра \mathbf{c} определяется из условия

$$\hat{\mathbf{c}} = \arg \min_c \sum_{i=1}^N |x_i - \mathbf{c}^*\varphi(u_i)|^p, \quad (2.121)$$

где $p < 2$. При $p = 2$ она совпадает с оценкой МНК.

При наличии засорения в выборке $\hat{\mathbf{c}}$ предпочтительнее оценки МНК. Обычно выбирается $p = 1,5$.

Процедура Ейла — Форсайта. Данная процедура основана на так называемой винзоризации остатков. Пусть $\mathbf{c}(0)$ — оценка МНК (2.112). Определим остатки

$$r_i(n) = x_i - \mathbf{c}^*(n)\varphi(u_i), \quad n = 1, 2, \dots \quad (2.122)$$

Ранжируем их, построив вариационный ряд

$$r_{(1)}(n) \leq r_{(2)}(n) \leq \dots \leq r_{(N)}(n). \quad (2.123)$$

Винзоризируем их, т.е. приведем к виду $\tilde{r}_{(i)}(n)$ преобразованием Винзора:

$$\tilde{r}_{(i)}(n) = \begin{cases} r_{(g+1)}(n), & 1 \leq i \leq g, \\ r_{(i)}(n), & g < i \leq N - g, \\ r_{(N-g)}(n), & N - g < i \leq N. \end{cases}$$

Оценка $\mathbf{c}(n+1)$ находится из соотношения

$$\mathbf{c}(n+1) = \left(\sum_{i=1}^N \boldsymbol{\varphi}(u_{(i)}) \boldsymbol{\varphi}^*(u_{(i)}) \right)^{-1} \sum_{i=1}^N (\tilde{r}_{(i)}(n) + \mathbf{c}^*(n) \boldsymbol{\varphi}(u_{(i)})) \boldsymbol{\varphi}(u_{(i)}). \quad (2.124)$$

Процедура выполняется до достижения необходимой точности. Заметим, что суммирование в (2.124) производится не по индексам наблюдения, а по индексам ранжирования, задаваемым неравенствами (2.123).

Возможны различные формы использования винзоризации в ходе выполнения итераций.

Простой итерационный метод. Уровень винзоризации g остается постоянным на каждой итерации. При определении остатков $r_i(n)$ по формуле (2.122) вместо x_i используется $\tilde{r}_{(i)}(n+1) + \mathbf{c}^*(n) \boldsymbol{\varphi}(u_{(i)})$, $n \geq 1$.

Метод уровней. Уровень винзоризации g растет от итерации к итерации. При определении остатков $r_i(n)$ по формуле (2.122) всегда используется наблюдение x_i .

Итерационный метод с растущим уровнем. Этот метод является комбинацией двух первых. Уровень винзоризации g растет от итерации к итерации, причем в (2.122) для определения остатков $r_i(n)$ используется вместо x_i $\tilde{r}_{(i)}(n-1) + \mathbf{c}^*(n) \boldsymbol{\varphi}(u_{(i)})$.

В заключение рассмотрим применение оценок, не принадлежащих к М-оценкам. Определим для некоторого вектора \mathbf{c} остатки

$$r_i = x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i), \quad 1 \leq i \leq N, \quad (2.125)$$

и ранжируем их так: $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(N)}$. Пусть R_i — ранг остатка r_i , т.е. $r_{(R_i)} \equiv r_i$. Отталкиваясь от суммы квадратов (2.89), модифицируем эту сумму следующим образом. Один из множителей $(x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i))^2$ заменим на его ранг R_i . Тогда получим вместо (2.89)

$$\sum_{i=1}^N (x_i - \mathbf{c}^* \boldsymbol{\varphi}(u_i)) R_i. \quad (2.126)$$

Поскольку ранг R_i представляет собой функцию \mathbf{c} , эта сумма является кусочно-линейной выпуклой функцией \mathbf{c} , которую можно использовать для нахождения вектора \mathbf{c} , соответствующего минимуму этой суммы. Этот вектор принимается в качестве оценки параметров регрессии. В общем случае вместо R_i в (2.22) можно использовать некоторую монотонную неубывающую функцию $a(R_i)$, удовлетворяющую условию $\sum_{i=1}^N a(R_i) = 0$.

В этих условиях вместо уравнения (2.88) получим уравнение

$$\sum_{i=1}^N a(R_i) \varphi(u_i) = 0, \quad (2.127)$$

решение которого относительно \mathbf{c} и дает оценку параметров регрессии. Решение уравнения (2.127) затрудняется тем, что ранги R_i , рассматриваемые как функции \mathbf{c} , являются разрывными функциями.

Лабораторная работа 6. Устойчивые методы оценивания параметров регрессии

Цель работы. Ознакомиться с робастными методами оценивания коэффициентов регрессии. Освоить технику реализации робастных процедур оценивания. Провести сравнение качества различных оценок.

Порядок выполнения работы

Необходимо использовать реализацию временного ряда из лабораторной работы 5, которую надо модифицировать следующим образом. Выбрать некоторый уровень загрязнения ε . Определить случайным образом εN номеров из множества $\{1, 2, \dots, N\}$ и каждое наблюдение с номером из выбранного множества заменить "грубым выбросом". Это означает следующее. Пусть для выбранной реализации Y — такой уровень, что для некоторого t выполняется неравенство

$$\mathbf{P}\{|y_t| > Y\} = \mathbf{P}\{|x_t - m_t| > Y\} \leq \delta,$$

где δ — достаточно малая положительная величина, возможно, равная нулю. При помощи случайного механизма определим \tilde{y}_t такое, что $|\tilde{y}_t| > Y$. В качестве грубо искаженного значения временного ряда в точке t возьмем вместо x_t величину $\tilde{x}_t = m_t + \tilde{y}_t$. Полученную таким образом реализацию будем называть *загрязненной реализацией временного ряда*.

Задание. Для загрязненной реализации получить: а) оценки параметров регрессии методами, использованными в лабораторной работе 5; б) оценки параметров устойчивыми методами, т.е. методом Хьюбера по формулам (2.110)–(2.112), методом модифицированных остатков (2.113)–(2.115), методом модифицированных весов (2.116)–(2.118), методом Андриуса (2.119), (2.120), методом Форсайта (2.121), методом Ейла — Форсайта по формулам (2.122)–(2.124), а также используя ранговые критерии из уравнения (2.127).

Произвести сравнение качества оценок между группами а) и б), а также внутри группы. Исследовать зависимость качества оценок от уровня загрязнения ε . Качество оценок определять по норме отклонения оценок параметров регрессии от их истинных значений:

$$\sqrt{\sum_{j=1}^q (\hat{c}_j - c_j)^2},$$

а также по среднеквадратичному отклонению функции регрессии на интервале наблюдения:

$$\sqrt{\frac{1}{t_N - t_1} \int_{t_1}^{t_N} (\hat{m}_t - m_t)^2 dt}.$$

2.6. Аппроксимация функции регрессии сплайнами

Под приближением *сплайнами* понимают способ кусочно-полиномиального приближения непрерывных функций. В этом случае область изменения аргумента, скажем $[0, T]$, делится на $(1 + p)$ интервалов точками $0 = t^{(0)} < t^{(1)} < t^{(2)} < \dots < t^{(p)} < t^{(p+1)} = T$. На каждом интервале $(t^{(k)}, t^{(k+1)})$, $1 \leq k \leq p$, аппроксимирующая функция изображается полиномом фиксированной степени q . В точках $t^{(k)}$, $1 \leq k \leq p$, требуется выполнение условий непрерывности аппроксимируемой функции и ее $(q - 1)$ первых производных. Такие кусочно-полиномиальные функции называются *сплайнами степени q с p сопряжениями*.

Для приближения кривой регрессии будем использовать такие сплайны $S_p(t)$, которые минимизируют эмпирический риск

$$\frac{1}{N} \sum_{i=1}^N (x_{t_i} - S_p(t_i))^2.$$

Сплайн $S_p(t)$ задается $(1+q+p)$ параметрами, если разбиение области изменения аргумента делается равномерным, т.е. $t^{(k+1)} - t^{(k)} = \text{const}$, $0 \leq k \leq p$. В этом случае говорят, что $S_p(t)$ — сплайн на равномерной сетке.

Аппроксимация функции регрессии сплайнами основана на следующем важном свойстве. Пусть о функции регрессии известно только то, что она является непрерывной функцией на $[0, T]$. Какова бы ни была непрерывная исходная функция регрессии m_t , осредненные сплайны с сопряжениями в равноотстоящих узлах (на равномерной сетке) при $N \rightarrow \infty$ и $p(N) \rightarrow \infty$

$$(p(N))^4 \frac{\ln(N/p(N))}{N/p(N)} \rightarrow 0$$

равномерно приближают m_t , т.е.

$$\sup_{t \in [0, T]} |m_t - S_p(t)| \rightarrow 0.$$

Вместе с тем известно, что всегда существует такая непрерывная функция \tilde{m}_t на $[0, T]$, которую полиномиальные приближения, полученные методом наименьших квадратов, равномерно не приближают ни при каком соотношении числа наблюдений N и степени полинома $q(N)$, т.е. для всякой системы полиномов $\{\varphi_j(t)\}$, заданных на $[0, T]$, справедливо

$$\sup_{t \in [0, T]} |m_t - \sum_{j=0}^q c_j \varphi_j(t)| \rightarrow \delta \neq 0.$$

Этим и объясняется преимущество аппроксимации функции регрессии сплайнами перед приближением полиномами, если о приближаемой функции m_t отсутствует полная информация.

Обычно достаточно строить сплайны на основе полиномов третьей степени (кубические сплайны). Строить сплайны удобно, определяя систему фундаментальных сплайнов. Для кубических сплайнов с p сопряжениями на сетке $(t^{(0)}, t^{(1)}, \dots, t^{(p)}, t^{(p+1)})$ вводятся $(p+4)$ фундаментальных сплайна $\{\mu_j(t)\}$. Фундаментальные сплайны однозначно определяются следующими условиями. Пусть для краткости $\mu_{ij} = \mu_j(t^{(i)})$, μ'_{ij} — производная $\mu_j(t)$ в точке $t^{(i)}$. Тогда

$$\begin{aligned} \mu_{i1} &= 0, \quad 0 \leq i \leq 1+p, \quad \mu'_{01} = 1, \quad \mu'_{p+1,1} = 0, \\ \mu_{ij} &= \delta_{i,j-2}, \quad \mu'_{0j} = \mu'_{p+1,j} = 0, \quad 0 \leq i \leq p+1, \quad 2 \leq j \leq p+3, \\ \mu_{i,p+4} &= 0, \quad 0 \leq i \leq p+1, \quad \mu'_{0,p+4} = 0, \quad \mu'_{p+1,p+4} = 1, \end{aligned} \quad (2.128)$$

где $\delta_{i,j}$ означает символ Кронекера.

Поскольку любой кубический сплайн $S_p(t)$ полностью определяется $(p+2)$ значениями в узлах $t^{(j)}$, $0 \leq i \leq p+1$, и значениями первой производной на концах отрезка, имеет место равенство

$$S_p(t) = \sum_{j=0}^{p+1} S_p(t^{(j)}) \mu_{j+2}(t) + S'_p(t^{(0)}) \mu_1(t) + S'_p(t^{(p+1)}) \mu_{p+4}(t). \quad (2.129)$$

Именно такое представление будет использоваться в дальнейшем при восстановлении регрессии в классе кубических сплайнов с p сопряжениями. Это выражение запишем короче:

$$S_p(t) = \sum_{j=1}^{p+4} c_j \mu_j(t) = \mathbf{c}^* \boldsymbol{\mu}(t). \quad (2.130)$$

Таким образом, данная задача сведена к проблеме минимизации эмпирического риска

$$J_{\mathfrak{A}}(\mathbf{c}) = \frac{1}{N} \sum_{i=1}^N (x_i - \mathbf{c}^* \boldsymbol{\mu}(t_i))^2 \quad (2.131)$$

по вектору \mathbf{c} , содержащему $(p+4)$ компонент. Но для решения этой задачи нужно определить в явной форме систему кубических фундаментальных сплайнов. Сетку выберем равномерной, такой, что $t^{(i+1)} - t^{(i)} = \Delta$, $0 \leq i \leq p$.

Пусть $m_{i+1,j}$ и $m_{i,j}$ — значения второй производной фундаментального сплайна $\mu_j(t)$ в узлах $t^{(i+1)}$ и $t^{(i)}$ соответственно. Поскольку вторая производная полинома третьей степени — линейная функция, для $t \in [t^{(i)}, t^{(i+1)}]$ справедливо

$$\mu_j''(t) = m_{i+1,j} \frac{t - t^{(i)}}{\Delta} + m_{i,j} \frac{t^{(i+1)} - t}{\Delta}, \quad (2.132)$$

где $m_{i+1,j} = \mu_j''(t^{(i+1)})$; $m_{i,j} = \mu_j''(t^{(i)})$. Проинтегрировав дважды (2.132) с учетом непрерывности сплайна на концах отрезка $[t^{(i)}, t^{(i+1)}]$, получим

$$\begin{aligned} \mu_j(t) &= m_{i,j} \frac{(t^{(i+1)} - t)^3}{6\Delta} + m_{i+1,j} \frac{(t - t^{(i)})^3}{6\Delta} + \\ &+ \left(\mu_{i,j} - \frac{\Delta^2}{6} m_{i,j} \right) \frac{(t^{(i+1)} - t)}{\Delta} + \left(\mu_{i+1,j} - \frac{\Delta^2}{6} m_{i+1,j} \right) \frac{(t - t^{(i)})}{\Delta}, \end{aligned} \quad (2.133)$$

где для краткости обозначим $\mu_j(t^{(i)}) = \mu_{i,j}$.

Для нахождения производных $m_{i,j}$ воспользуемся условием непрерывности в узлах сопряжения первой производной кубического сплайна. Дифференцируя (2.133) для узлов $t^{(i)}$, $1 \leq i \leq p$, приходим к уравнению

$$\frac{m_{i-1,j}}{2} + 2m_{i,j} + \frac{m_{i+1,j}}{2} = \frac{6}{\Delta^2} \left(\frac{\mu_{i-1,j}}{2} - \mu_{i,j} + \frac{\mu_{i+1,j}}{2} \right). \quad (2.134)$$

Для узла $t^{(0)}$ имеем

$$2m_{0j} + m_{1j} = \frac{6}{\Delta^2} (-\Delta \mu_j'(t^{(0)}) - \mu_{0j} + \mu_{1j}). \quad (2.135)$$

Для узла $t^{(p+1)}$ получаем

$$m_{p,j} + 2m_{p+1,j} = \frac{6}{\Delta^2} (\mu_{pj} - \mu_{p+1,j} + \Delta \mu_j'(t^{(p+1)})). \quad (2.136)$$

Значения правых частей (2.134)–(2.136) определены в (2.128). Таким образом, совокупность (2.134)–(2.136) представляет собой неоднородную систему $(p+2)$ линейных алгебраических уравнений для $(p+2)$ неизвестных $m_{i,j}$, $0 \leq i \leq p+1$. Рассматривая (2.134)–(2.136) для различных j , $1 \leq j \leq p+4$, строим $(p+4)$ таких систем. Заметим, что матрицы коэффициентов при $m_{i,j}$ в этих системах являются одинаковыми и имеют вид ((2.134) удобно умножить на 2)

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 2 \end{pmatrix}. \quad (2.137)$$

Учитывая (2.137) и образуя матрицу $M = (m_{ij})$ размером $(p+2) \times (p+4)$, получаем для матрицы M уравнение

$$AM = \frac{6}{\Delta^2} B, \quad (2.138)$$

где $(p+2) \times (p+2)$ — матрица A , определенная в (2.138), а $(p+2) \times (p+4)$ — матрица B , составленная из правых частей уравнений (2.134)–(2.136), имеет структуру

$$B = \begin{pmatrix} -\Delta & -1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ \dots & & & \dots & & \dots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 & \Delta \end{pmatrix}. \quad (2.139)$$

Матрица A является матрицей Якоби, и ее можно обратить аналитически. Обозначим $A^{-1} = C = (c_{ij})$. Тогда элементы матрицы C вычисляются по формулам

$$c_{ij} = \frac{d_i d_{p+3-j}}{d_{j-1} d_{p+3-j} - d_j d_{p+4-j}}, \quad 1 \leq i \leq j, \\ c_{ij} = \frac{d_{p+3-i} d_j}{d_{j-1} d_{p+3-j} - d_j d_{p+4-j}}, \quad j \leq i \leq p+2, \quad (2.140)$$

в которых $1 \leq j \leq p+2$, а числа d_j находятся рекуррентно:

$$d_0 = -2, \quad d_1 = 1, \quad d_j = -4d_{j-1} - d_{j-2}, \quad 2 \leq j \leq p+3. \quad (2.141)$$

Таким образом, матрица M , составленная из чисел m_{ij} , определяется соотношением

$$M = \frac{6}{\Delta^2} CB, \quad (2.142)$$

где матрицы B и C задаются формулами (2.139)–(2.141). Это и завершает определение фундаментальных сплайнов $\mu_j(t)$ в виде (2.133), в котором числа μ_{ij} удовлетворяют условиям (2.128), а числа m_{ij} являются элементами матрицы (2.142).

После того, как система фундаментальных сплайнов $\mu_j(t)$ построена, вычисление сплайна $S_p(t)$, минимизирующего эмпирический риск, проводится по обычной схеме получения оценок МНК. Вектор, минимизирующий (2.131), находится в виде

$$\hat{\mathbf{c}} = \left(\sum_{i=1}^N \boldsymbol{\mu}(t_i) \boldsymbol{\mu}^*(t_i) \right)^{-1} \sum_{i=1}^N x_i \boldsymbol{\mu}(t_i). \quad (2.143)$$

Здесь, как в (2.130), (2.131), вектор-столбец $\boldsymbol{\mu}(t)$ составлен из фундаментальных сплайнов $\boldsymbol{\mu}^*(t) = (\mu_1(t), \dots, \mu_{p+1}(t))$, которые для каждого t из интервала $(t^{(i)}, t^{(i+1)})$ заданы формулой (2.133).

Остановимся теперь на определении числа сопряжений p . Как и прежде, за качество приближения функции регрессии примем функционал

$$J(p) = \frac{\sum_{i=1}^N \left(x_i - \sum_{j=1}^{p+4} c_j \mu_j(t_i) \right)^2}{N - \sqrt{N((p+4)(1 + \ln \frac{N}{p+4}) - \ln \alpha)}}, \quad (2.144)$$

где α — уровень значимости, на котором принимается решение о числе сопряжений. Наилучшим числом сопряжений считается то, которое минимизирует функционал (2.144), обеспечивая ему неотрицательное значение. При определении числа сопряжений следует сначала выяснить, имеет ли смысл приближать функцию регрессии сплайнами, так как техника приближения сплайнами все-таки более сложная, чем техника приближения полиномами. Поэтому надо проверить, насколько хорошо функция регрессии аппроксимируется полиномами нулевого, первого, второго или третьего порядка на всем интервале наблюдения $[0, T]$. Качество приближения оценивается при этом функционалом (2.28). Если оно недостаточно, то используются сплайны с оптимальным числом сопряжений p .

Лабораторная работа 7. Приближение функции регрессии сплайнами

Цель работы. Освоить технику построения кубических сплайнов, приближающих функции регрессии.

Порядок выполнения работы

Необходимо построить реализацию временного ряда, функция регрессии которого не является полиномом. В качестве реализации временного ряда в этом случае взять тот же временной ряд, который был использован в работе 2, или тот, функция регрессии которого аппроксимировалась полиномами Лежандра в работе 3. Такой выбор поможет воспользоваться уже имеющимися результатами для сравнения качества приближения. В противном случае их придется получать снова в ходе выполнения настоящей работы.

После того, как построена реализация временного ряда и выбрано число сопряжений p , необходимо построить систему фундаментальных сплайнов (2.133), параметры которых заданы (2.128) и системой уравнений (2.139)–(2.142).

Задание. Для построения системы фундаментальных сплайнов определить оценки параметров c по формуле (2.143) и по ним построить сплайн-приближение (2.130).

При помощи функционала (2.144) найти оптимальное число сопряжений. Для этого числа сопряжений оценить качество приближения по формуле

$$\frac{1}{t_N - t_1} \int_{t_1}^{t_N} |\hat{c}^* \mu(t) - m_t| dt,$$

где m_t — истинная функция регрессии. Сравнить это качество с качеством приближения полиномами, которое определялось в предыдущих работах (выделение полиномиального тренда, приближение функции регрессии полиномами Лежандра).

Глава 3

СТАЦИОНАРНЫЕ ПРОЦЕССЫ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО

3.1. Основные понятия

Пусть x_t — центрированный СП, определенный на действительной прямой \mathbf{R} или в комплексной области \mathbf{C} . Будем считать, что x_t является процессом авторегрессии — скользящего среднего (АРСС), или *ARMA* (*autoregressive — moving average*), порядка (p, q) , если существует процесс белого шума W_t и действительные или комплексные числа $a_0, \dots, a_p, b_0, \dots, b_q$ такие, что

$$\sum_{k=0}^p a_k x_{t-k} = \sum_{l=0}^q b_l W_{t-l}, \quad t \in \mathbf{Z}, \quad a_p, b_q \neq 0. \quad (3.1)$$

Определим два частных класса процесса АРСС. Будем считать, что x_t является процессом авторегрессии (АР) порядка p (АР(p)-процессом), если он удовлетворяет уравнению (3.1), где $q = 0$, а именно:

$$\sum_{k=0}^p a_k x_{t-k} = W_t, \quad t \in \mathbf{Z}.$$

Будем считать, что x_t является процессом скользящего среднего (СС) порядка q (СС(q)-процессом), если он удовлетворяет уравнению (3.1), где $p = 0$, а именно:

$$x_t = \sum_{l=0}^q b_l W_{t-l}.$$

Уравнение (3.1) можно представить в операторной форме:

$$A(\Delta)x_t = B(\Delta)W_t,$$

где Δ — оператор сдвига на шаг назад: $\Delta x_t = x_{t-1}$; $A(z)$ и $B(z)$ — полиномы вида

$$A(z) = \sum_{k=0}^p a_k z^k, \quad B(z) = \sum_{l=0}^q b_l z^l.$$

Центрированный СП x_t имеет рациональный спектр, если x_t имеет спектральную плотность $g(\lambda) = G(e^{-i\lambda})$, $\lambda \in [-\pi, \pi]$, где G — дробно-рациональная функция с коэффициентами из множества \mathbf{C} . Справедлива теорема Фейера — Рисса, в соответствии с которой центрированный СП x_t является АРСС-процессом тогда и только тогда, когда существует несократимая дробь

$B(z)/A(z)$, где полином $A(z)$ не имеет корней, по модулю равных единице, и такой, что

$$G(z) = \left| \frac{B(z)}{A(z)} \right|^2 \quad \text{для } |z| = 1, \quad z \in \mathbf{C}.$$

Если полином $G(z)$ имеет действительные коэффициенты, то можно найти полиномы $A(z), B(z)$ с действительными коэффициентами.

Процесс АРСС (p, q) — каузальный, если существует последовательность констант $\{\psi_j\}$ таких, что

$$\begin{aligned} \sum_{j=0}^{\infty} |\psi_j| < \infty, \\ x_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}, \quad t \in \mathbf{Z}, \end{aligned} \quad (3.2)$$

и обратимый, если существует последовательность констант $\{\pi_j\}$ таких, что

$$\begin{aligned} \sum_{j=0}^{\infty} |\pi_j| < \infty, \\ W_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}, \quad t \in \mathbf{Z}. \end{aligned}$$

Справедлива теорема, согласно которой процесс АРСС (p, q) является каузальным и обратимым тогда и только тогда, когда $A(z) \neq 0$, $B(z) \neq 0$ для всех $z \in \mathbf{C}$ таких, что $|z| < 1$. При этом коэффициенты определяются из соотношений

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i = B(z)/A(z), \quad |z| < 1, \quad (3.3)$$

$$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^i = A(z)/B(z), \quad |z| < 1. \quad (3.4)$$

Для определения коэффициентов ψ_j и π_j перепишем (3.3), (3.4) в виде $\psi(z)A(z) = B(z)$, $\pi(z)B(z) = A(z)$ и приравняем коэффициенты при одинаковых степенях z . Тогда получим

$$\psi_j + \sum_{k=1}^j a_k \psi_{j-k} = b_j, \quad 0 \leq j < \max(p, q + 1),$$

$$\psi_j + \sum_{k=1}^p a_k \psi_{j-k} = 0, \quad j > \max(p, q + 1).$$

Данные уравнения разрешаем относительно $\psi_0, \psi_1, \psi_2, \dots$ следующим образом:

$$\psi_0 = b_0 = 1,$$

$$\begin{aligned} \psi_1 + \psi_0 a_1 = b_1 \quad \text{и} \quad \psi_1 = b_1 - a_1, \\ \psi_2 + a_1 \psi_1 + a_2 = b_2 \quad \text{и} \quad \psi_2 = b_2 - a_1 \psi_1 - a_2. \\ \dots \quad \dots \end{aligned} \quad (3.5)$$

Аналогично находятся коэффициенты $\{\pi_j\}$ из уравнений

$$\pi_j = \sum_{k=1}^{\min(q,j)} b_k \pi_{j-k} = a_j, \quad j = 0, 1, \dots,$$

где $a_0 = 1$ и $b_k = 0$ для $k > q$; $a_j = 0$ для $j > 0$.

Рассмотрим три алгоритма вычисления ковариационной функции процесса АРСС (p, q) :

$$\gamma_s = \mathbf{M}\{x_t x_{t+s}\}, \quad s = 0, \pm 1, \dots$$

Алгоритм 1. Из соотношения (3.2) следует, что ковариационная функция удовлетворяет соотношению

$$\gamma_s = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|s|}, \quad \sigma^2 = \mathbf{M}\{W_t^2\}.$$

Коэффициенты ψ_j находятся из соотношений (3.5).

Алгоритм 2. Основан на разностных уравнениях для γ_s , $s = 0, 1, \dots$, которые получаются умножением каждого слагаемого в (3.1) на x_{t-s} и использованием операции математического ожидания:

$$\gamma_s + a_1 \gamma_{s-1} + \dots + a_p \gamma_{s-p} = \sigma^2 \sum_{k=s}^q b_k \psi_{k-s} \quad \text{для} \quad 0 \leq s < \max(p, q+1). \quad (3.6)$$

$$\gamma_s + a_1 \gamma_{s-1} + \dots + a_p \gamma_{s-p} = 0 \quad \text{для} \quad s \geq \max(p, q+1). \quad (3.7)$$

Общее решение уравнения (3.7) записывается в виде

$$\gamma_s = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \beta_{ij} s^j z_i^{-s}, \quad s \geq \max(p, q+1) - p,$$

где z_i , $1 \leq i \leq k$, — различные корни полинома $A(z)$; r_i — кратность i -го корня; $\sum_{i=1}^k r_i = p$; p констант β_{ij} и ковариации γ_s , $0 \leq s < \max(p, q+1) - p$, единственным образом определяются из граничных условий (3.6) после вычисления $\psi_0, \psi_1, \dots, \psi_q$.

Пример. Пусть $(1 - \Delta + \frac{1}{4}\Delta^2)x_t = (1 + \Delta)W_t$. Из уравнения (3.7) следует, что $\gamma_s - \gamma_{s-1} + \frac{1}{4}\gamma_{s-2} = 0$, $s \geq 2$. Общее решение данного уравнения задается соотношением

$$\gamma_n = (\beta_{10} + \beta_{11}n)2^{-n}, \quad n \geq 0. \quad (3.8)$$

Граничные условия (3.6) имеют вид

$$\begin{aligned}\gamma_s - \gamma_{s-1} + \frac{1}{4}\gamma_{s-2} &= \sigma^2(\psi_0 + \psi_1), \\ \gamma_s - \gamma_{s-1} + \frac{1}{4}\gamma_{s-2} &= \sigma^2\psi_0,\end{aligned}\tag{3.9}$$

где из (3.5) $\psi_0 = 1$, $\psi_1 = 2$. Подставляя $\gamma_0, \gamma_1, \gamma_2$ из общего решения в уравнения (3.9), получаем

$$\begin{aligned}3\beta_{10} - 2\beta_{11} &= 16\sigma^2, & -3\beta_{10} + 5\beta_{11} &= 8\sigma^2, \\ \beta_{11} &= 8\sigma^2, & \beta_{10} &= 32\sigma^2/3.\end{aligned}$$

Окончательно имеем $\gamma_s = \sigma^2 2^{-s} \left[\frac{32}{3} + 8s \right]$.

Алгоритм 3. Вычисления γ_s из уравнений (3.6), (3.7) производятся следующим образом. Сначала из (3.6) находятся $\gamma_0, \gamma_1, \dots, \gamma_p$, а затем последовательно определяются $\gamma_{p+1}, \gamma_{p+2}, \dots$ из уравнения (3.7).

Частная ковариационная функция $\alpha(k)$, $k = 1, 2, \dots$, может быть интерпретирована как ковариационная функция между x_1 и x_{k+1} при устранении влияния x_2, \dots, x_k . Частная ковариационная функция стационарного временного ряда задается так, что

$$\begin{aligned}\alpha(1) &= \rho_1 = \gamma_1\gamma_0, \\ \alpha(k) &= \text{cov}(x_{k+1} - P_{1,x_2,\dots,x_k}(x_{k+1}), x_1 - P_{1,x_2,\dots,x_k}(x_1))/\gamma,\end{aligned}$$

где $P_{1,x_2,\dots,x_k}(x)$ — проекция переменной x на линейное подпространство, образованное переменными $1, x_2, \dots, x_k$, т.е.

$$P_{1,x_1,\dots,x_n}(x) = \sum_{i=0}^n \beta_i x_i, \quad x_0 = 1.$$

Коэффициенты β_i находятся из уравнений

$$\sum_{i=0}^n \beta_i \mathbf{M}\{x_i x_j\} = \mathbf{M}\{x x_j\}, \quad j = 1, \dots, n.$$

Если процесс x_t — центрированный, то $P_{1,x_2,\dots,x_k}(\cdot) = P_{x_2,\dots,x_k}(\cdot)$. Для каузального процесса АР (p) имеет место соотношение

$$P_{x_2,\dots,x_k}(x_{k+1}) = \sum_{j=1}^p a_j x_{k+1-j}.$$

Частную ковариационную функцию можно определить иначе. Пусть $P_{x_1,\dots,x_n}(x_{n+1}) = \sum_{j=1}^n a_{nj} x_{n-j}$. Тогда из соотношения

$$\mathbf{M}\{(x_{n+1} - P_{x_1,\dots,x_n}(x_{n+1}))x_j\} = 0, \quad j = 1, \dots, n,$$

получим

$$\begin{pmatrix} \rho_0 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & \rho_0 & \dots & \rho_{n-2} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \dots & \rho_0 \end{pmatrix} \begin{pmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nn} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_n \end{pmatrix}, \quad n \geq 1, \quad (3.10)$$

где $\rho_j = \gamma_j/\gamma_0$. Частная ковариационная функция в этом случае записывается в виде $\alpha(n) = a_{nn}$, $n \geq 1$, где a_{nn} единственным образом находятся из (3.10).

Лабораторная работа 8. Вычисление ковариационных и частных ковариационных функций

Цель работы. Смоделировать временные ряды, порождаемые процессами АРСС. Оценить ковариационную функцию процесса по значениям ряда. Вычислить ковариационную и частную ковариационную функции по заданным коэффициентам процесса АРСС. Сравнить полученные функции.

Порядок выполнения работы

Используя генераторы псевдослучайных чисел, смоделировать временные ряды, порождаемые каузальными и обратимыми процессами АРСС, длиной N . Изобразить графически значения временных рядов.

Задание 1. Для временного ряда x_1, \dots, x_N , порождаемого уравнением АР (p)

$$x_t + a_1x_{t-1} + \dots + a_px_{t-p} = W_t, \quad t = p + 1, \dots, N,$$

где $\{W_t\}$ — последовательность независимых гауссовских СВ с нулевым средним и дисперсией σ^2 , начальные значения $x_1 = \dots = x_p = 0$, оценить ковариационную функцию

$$\gamma_s = \frac{1}{N-s} \sum_{t=s+1}^N x_t x_{t-s}, \quad s = 0, 1, \dots, N-1.$$

С помощью алгоритмов 1–3 вычислить ковариационную функцию γ_s . Используя формулу (3.10), определить частную ковариационную функцию $\alpha(n)$, $n = 1, 2, \dots, p+k$. Применяя формулу (3.10), в которой истинные значения ρ_n заменены на их оценки $\hat{\rho}_n = \hat{\gamma}_n/\hat{\gamma}_0$, оценить частную ковариационную функцию $\hat{\alpha}(n)$, $n = 1, 2, \dots, p+k$. Изобразить графически ковариационную и частную ковариационную функции.

Задание 2. Для временного ряда x_1, \dots, x_N , порождаемого процессом СС (q), вычислить и оценить ковариационную и частную ковариационную функции, провести сравнение вычисленных и оцененных функций. Исследовать поведение частной ковариационной функции $\alpha(n)$ при $n \rightarrow \infty$. Изобразить графически полученные функции.

Задание 3. Для временного ряда, порождаемого процессом АРСС (p, q) , вычислить и оценить ковариационную и частную ковариационную функции. Исследовать поведение ковариационной функции γ_n при $n \rightarrow \infty$ и частной ковариационной функции $\alpha(n)$ при $n \rightarrow \infty$ в зависимости от значений коэффициентов процесса АРСС (p, q) .

3.2. Оценивание параметров процесса авторегрессии и процессов скользящего среднего

Одной из первоочередных задач при статистическом исследовании процессов АР и СС является задача оценивания коэффициентов $a_1, \dots, a_p, b_1, \dots, b_q$ и дисперсии $\sigma^2 = \mathbf{M}\{W_t^2\}$ по наблюдениям x_1, \dots, x_N за процессом x_t . Рассмотрим несколько методов оценивания.

Метод наименьших квадратов. Пусть x_t каузальный процесс АР (p)

$$x_t = \sum_{i=1}^p a_i x_{t-i} + W_t, \quad (3.11)$$

где $\{W_t\}$ — последовательность независимых одинаково распределенных случайных величин с нулевым средним и дисперсией σ^2 . Обозначим

$$\tilde{\mathbf{x}}_t = (x_t, x_{t-1}, \dots, x_{t-p+1})^*, \quad t = p+1, \dots, N, \quad \mathbf{a} = (a_1, \dots, a_p)^*.$$

Тогда уравнение (3.11) примет вид

$$x_t = \mathbf{a}^* \tilde{\mathbf{x}}_{t-1} + W_t, \quad t = p+1, \dots, N.$$

Оценки МНК получаются в результате решения задачи

$$\sum_{t=p+1}^N (x_t - \mathbf{a}^* \tilde{\mathbf{x}}_{t-1})^2 \rightarrow \min_{a_1, \dots, a_p}.$$

Оценки коэффициентов a_1, \dots, a_p и дисперсии σ^2 определяются соотношениями

$$\hat{\mathbf{a}}(N) = \left(\sum_{t=p+1}^N \tilde{\mathbf{x}}_{t-1} \tilde{\mathbf{x}}_{t-1}^* \right)^{-1} \sum_{t=p+1}^N \tilde{\mathbf{x}}_{t-1} x_t; \quad (3.12)$$

$$\hat{\sigma}^2(N) = \sum_{t=p+1}^N (x_t - \hat{\mathbf{a}}^*(N) \tilde{\mathbf{x}}_{t-1})^2. \quad (3.13)$$

Найти МНК-оценки можно также с помощью рекуррентных соотношений, которые, с одной стороны, упрощают вычислительную процедуру (3.12), а с другой — позволяют пересчитывать оценки по мере поступления данных (в реальном масштабе времени). Оценка $\hat{\mathbf{a}}$ на $t+1$ шаге вычисляется следующим образом:

$$\hat{\mathbf{a}}(t+1) = \hat{\mathbf{a}}(t) - P(t)\tilde{\mathbf{x}}_t L_t(x_{t+1} - \hat{\mathbf{a}}^*(t)\tilde{\mathbf{x}}_t); \quad (3.14)$$

$$P(t+1) = P(t) - P(t)\tilde{\mathbf{x}}_{t+1} L_{t+1} \tilde{\mathbf{x}}_{t+1}^* P(t); \quad (3.15)$$

$$L_{t+1} = (1 + \tilde{\mathbf{x}}_{t+1}^* P(t)\tilde{\mathbf{x}}_{t+1})^{-1}. \quad (3.16)$$

Процедура является полностью определенной, если заданы начальные значения $\hat{\mathbf{a}}(p)$, $P(p)$.

Метод Юла — Уолкера. Умножая обе части уравнения (3.11) на x_{t-j} , $j = 0, \dots, p$, и применяя операцию математического ожидания, получаем систему уравнений Юла — Уолкера

$$\Gamma(p)\mathbf{a} = \boldsymbol{\gamma}(p); \quad (3.17)$$

$$\sigma^2 = \gamma_0 - \mathbf{a}^* \boldsymbol{\gamma}(p), \quad (3.18)$$

где $\Gamma(p)$ — ковариационная матрица $\Gamma(p) = \|\Gamma_{ij}\|$, $i, j = 1, \dots, p$; $\Gamma_{ij} = \gamma_{i-j}$; $\boldsymbol{\gamma}(p) = (\gamma_1, \dots, \gamma_p)^*$. Если заменить ковариации γ_j , $j = 0, \dots, p$, в (3.17), (3.18) соответствующими выборочными оценками, например,

$$\hat{\gamma}_j = \sum_{t=j+1}^N x_t x_{t-j},$$

то получим систему уравнений

$$\hat{\Gamma}(p)\hat{\mathbf{a}} = \hat{\boldsymbol{\gamma}}(p). \quad (3.19)$$

Если $\hat{\gamma}_0 > 0$, то матрица $\hat{\Gamma}(p)$ невырожденная. Разделив каждую часть (3.19) на $\hat{\gamma}_0$, получим

$$\hat{\mathbf{a}} = \hat{R}^{-1}(p)\hat{\boldsymbol{\rho}}(p); \quad (3.20)$$

$$\sigma^2 = \hat{\gamma}_0[1 - \hat{\boldsymbol{\rho}}^*(p)\hat{R}^{-1}(p)\hat{\boldsymbol{\rho}}(p)], \quad (3.21)$$

где $\hat{R}^{-1}(p) = \hat{\Gamma}(p)/\hat{\gamma}_0$; $\hat{\boldsymbol{\rho}}(p) = \hat{\boldsymbol{\gamma}}(p)/\hat{\gamma}_0$.

Метод Дурбина — Левинсона. Специфическая форма матрицы \hat{R} и вектора $\hat{\boldsymbol{\rho}}$ приводит к простому рекуррентному способу вычисления коэффициентов уравнения АР (m) по коэффициентам уравнения АР ($m-1$). Таким образом, можно подобрать последовательно уравнение АР такого порядка, которое наиболее точно описывает данные x_1, \dots, x_N . Этот процесс АР можно записать в виде

$$x_t = \sum_{i=1}^m \hat{a}_{mi} x_{t-i} + W_t,$$

где $\{W_t\}$ — последовательность НОР СВ,

$$\mathbf{M}\{W_t\} = 0, \quad \mathbf{M}\{W_t^2\} = \hat{v}_m.$$

Используя метод Дурбина — Левинсона, определяем рекуррентные формулы для вычисления \hat{a}_{mj}, \hat{v}_m :

$$\hat{a}_{11} = \hat{\rho}_1; \quad (3.22)$$

$$\hat{v}_1 = \gamma_0[1 - \hat{\rho}_1^2]; \quad (3.23)$$

$$\hat{a}_{mm} = \left[\hat{\gamma}_m - \sum_{j=1}^{m-1} \hat{a}_{m-1,j} \hat{\gamma}_{m-j} \right] / \hat{v}_{m-1}; \quad (3.24)$$

$$\begin{pmatrix} \hat{a}_{m1} \\ \vdots \\ \hat{a}_{m,m-1} \end{pmatrix} = \begin{pmatrix} \hat{a}_{m-1,1} \\ \vdots \\ \hat{a}_{m-1,m-1} \end{pmatrix} - \hat{a}_{mm} \begin{pmatrix} \hat{a}_{m-1,m-1} \\ \vdots \\ \hat{a}_{m-1,1} \end{pmatrix}; \quad (3.25)$$

$$\hat{v}_m = \hat{v}_{m-1}(1 - \hat{a}_{mm}^2). \quad (3.26)$$

Для процесса АР (p) коэффициент \hat{a}_{mm} является также частной ковариацией $\alpha(m) = \hat{a}_{mm}$, которая отражает корреляционную зависимость СВ x_1 и x_{m+1} при "устранении влияния" переменных x_2, \dots, x_m . Известно, что для процесса АР (p) больших значений параметра N и $m > p$ оценка a_{mm} имеет нормальное распределение с нулевым средним и дисперсией $1/N$. Это свойство позволяет эвристически оценить порядок \hat{p} уравнения АР:

$$\hat{p} = \inf\{m : |\hat{a}_{m+1,m+1}| < N^{-1/2} \Phi_{1-\alpha/2}\}, \quad (3.27)$$

где $\Phi_{1-\alpha}$ — квантиль уровня $(1 - \alpha)$ стандартного нормального распределения. Доверительные интервалы для коэффициентов a_1, \dots, a_p определяются из соотношений (для достаточно больших N)

$$\{a_j \in \mathbf{R} : |a_j - \hat{a}_{pj}| \leq N^{-1/2} \Phi_{1-\alpha/2} \hat{v}_{jj}^{-1/2}\}, \quad j = 1, \dots, p, \quad (3.28)$$

где \hat{v}_{jj} — j -й диагональный элемент матрицы $\hat{v}_p \hat{\Gamma}^{-1}(p)$.

Последовательный метод. Данный метод не гарантирует несмещенность оценок, но позволяет контролировать среднеквадратичные отклонения оценок и обоснованно выбирать момент прекращения наблюдений в зависимости от требуемой точности оценивания.

Пусть $\{c_n\}, \{\kappa_n\}$ — неограниченно возрастающие последовательности положительных чисел и таких, что

$$\sum_{n=1}^{\infty} \frac{c_n}{\kappa_n^2} < \infty.$$

Эти неравенства выполняются, например, если справедливы условия

$$c_n = \kappa_n, \quad \sum_{n=1}^{\infty} \frac{1}{c_n} < \infty.$$

Определим p последовательных моментов остановки $\{\tau_i(n), n \geq 1\}$, $i = 1, \dots, p$, по формуле

$$\tau_i(n) = \inf \left\{ N \geq 0 : \sum_{t=0}^N \psi_i(t) \geq c_n \right\},$$

где $n \geq 1$; $\psi_i(t)$ — i -й диагональный элемент матрицы $(1/\sigma^2)\tilde{x}_t\tilde{x}_t^*$.

Введем следующие статистики:

$$\mathbf{m}(n) = (m_1(n), \dots, m_p(n))^*, \quad M(n) = \|m_{ij}(n)\|,$$

$$m_i(n) = \sum_{t=0}^{\tau_i(n)} '(\tilde{\mathbf{x}}_t x_{t+1})_i, \quad m_{ij}(n) = \sum_{t=0}^{\tau_i(n)} '(\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^*)_{ij},$$

где $(a)_i$ — i -й элемент вектора \mathbf{a} ; $(B)_{ij} = b_{ij}$ — элемент матрицы B ; штрих у знака суммы означает, что последнее слагаемое берется с весом $\alpha_i(\tau_i(n))$, определяемым из уравнения

$$\sum_{t=0}^{\tau_i(n)-1} \psi_i(t) + \alpha_i(\tau_i(n))\psi_i(\tau_i(n)) = c_n.$$

Вычислим искомые оценки параметров a_i :

$$s_i \equiv s_i(H) = \inf \left\{ k \geq 1 : \sum_{l=1}^k g_i(l) \geq H \right\}, \quad H > 0; \quad (3.29)$$

$$\hat{a}_i(H) = \frac{1}{H} \left(\sum_{n=1}^{s_i-1} z_i(n)g_i(n) + \beta_i z_i(s_i)g(s_i) \right); \quad (3.30)$$

$$\beta_i = \frac{H - \sum_{n=1}^{s_i-1} g_i^2(n)}{g_i^2(s_i)}, \quad (3.31)$$

где статистики $z_i(\cdot)$, $g_i(\cdot)$ имеют вид:

$$z_i(n) = \frac{\hat{a}_i(n)\Delta_n}{R_i(n)k_n}, \quad g_i(n) = \frac{\Delta_n}{R_i(n)k_n},$$

$$\hat{\mathbf{a}}(n) = (\hat{a}_1(n), \dots, \hat{a}_p(n))^* = M^{-1}(n)\mathbf{m}(n),$$

$$R_i(n) = \max_{1 \leq k \leq p} |G_{ik}(n)|, \quad \Delta_n = \det G(n), \quad G(n) = \|G_{ik}(n)\|,$$

$G_{ik}(n)$ — алгебраическое дополнение элемента $m_{ik}(n)$ в матрице $M(n)$. Для каузального процесса АР (p) оценки вида (3.30) для любого фиксированного $H > 0$

обладают свойством равномерной ограниченности среднеквадратичного отклонения $\mathbf{M}\{\hat{a}_i(H) - a_i\}^2 \leq \rho/H$, $\rho = p^2 \sum_{n=1}^{\infty} \frac{c_n}{k_n^2}$ и все моменты $S_i(H)$ конечны с вероятностью единица.

Для одномерного процесса АР $x_t = ax_{t-1} + W_t$ последовательные оценки имеют вид

$$\tau \equiv \tau(H) = \min\{N \geq 0 : \sum_{t=0}^N x_t^2 \geq H\}; \quad (3.32)$$

$$\hat{a}(H) = \left(\sum_{t=0}^{\tau-1} [x_t x_{t+1} + \alpha_\tau(H) x(\tau) x(\tau+1)] \right) / H, \quad (3.33)$$

где коэффициент $\alpha_\tau(H)$ находится из уравнения

$$\sum_{t=0}^{\tau} x_t^2 + \alpha_\tau(H) x^2(\tau) = H. \quad (3.34)$$

При этом оценка $\hat{a}(H)$ является несмещенной, среднеквадратичное отклонение оценки равномерно ограничено:

$$\sup_{-\infty < a < +\infty} \mathbf{M}\{(\hat{a}(H) - a)^2\} \leq \frac{1}{H} \quad (3.35)$$

и момент остановки $\tau(H)$ конечен с вероятностью единица. Последовательные оценки представляют собой взвешенные суммы МНК-оценок, вычисленных в случайные моменты времени.

Метод Дурбина — Левинсона оценивания коэффициентов процесса СС. Будем подбирать уравнение СС в форме

$$x_t = W_t + \sum_{i=1}^q \hat{b}_{mi} W_{t-i}$$

для $m = 1, 2, \dots, N-1$, где $\{W_t\}$ — последовательность НОР СВ,

$$\mathbf{M}\{W_t\} = 0, \mathbf{M}\{W_t^2\} = \hat{v}_m.$$

Если известны оценки коэффициентов процесса СС $(m-1)$, то оценки коэффициентов процесса СС (m) вычисляются рекуррентно:

$$\hat{v}_0 = \hat{\gamma}_0; \quad (3.36)$$

$$\hat{b}_{m,m-k} = \hat{v}_k^{-1} \left[\gamma_{m-k} - \sum_{j=0}^{k-1} \hat{b}_{m,m-j} \hat{b}_{k,k-j} v_j \right], \quad k = 0, \dots, m-1; \quad (3.37)$$

$$\hat{v}_m = \gamma_0 - \sum_{j=0}^{m-1} \hat{b}_{m,m-j} \hat{v}_j, \quad m = 1, 2, \dots, N-1. \quad (3.38)$$

Ковариационная функция \hat{r}_m процесса СС (q) для больших значений параметра N аппроксимируется нормальным распределением с нулевым средним и дисперсией $N^{-1}(1+2\rho_1^2+\dots+2\rho_q^2)$. Этот факт позволяет предложить следующий эвристический алгоритм поиска порядка q :

$$\hat{q} = \inf\{m : |\hat{\rho}_{m+1}| < N^{-1/2}\Phi_{1-\alpha/2}\}. \quad (3.39)$$

Доверительные интервалы для коэффициентов b_1, \dots, b_q определяются из соотношений (для достаточно больших N)

$$\left\{ b_j \in \mathbf{R} : |b_j - \hat{b}_{mj}| \leq N^{-1/2}\Phi_{1-\alpha/2} \left(\sum_{k=0}^j \hat{b}_{mk}^2 \right)^{1/2} \right\}, \quad j = 1, \dots, q. \quad (3.40)$$

Наконец, рассмотрим задачу оценивания порядка p процесса АР в следующей постановке. Построим тесты для проверки нулевой гипотезы

$$H_0 : a_{m+1} = a_{m+2} = \dots = a_p = 0$$

против альтернативы

$$H_1 : a_{m+1} \neq 0 \vee \dots \vee a_p \neq 0.$$

Для этого можно использовать статистики χ_1^2 и χ_2^2 , которые для каузального процесса АР (m) имеют в пределе χ^2 -распределение с $(p-m)$ степенями свободы:

$$\chi_1^2 = \frac{\hat{\mathbf{a}}^{(2)*}(M_{22} - M_{21}(M_{11})^{-1}M_{12})\hat{\mathbf{a}}^{(2)}}{\hat{\sigma}^2}, \quad \chi_2^2 = \sum_{j=m+1}^p \hat{h}_j^2/\sigma_m^2,$$

где

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \sum_{t=p+1}^N \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^*, \quad \hat{\mathbf{a}} = \begin{pmatrix} \hat{\mathbf{a}}^{(1)} \\ \hat{\mathbf{a}}^{(2)} \end{pmatrix}.$$

Матрица M является блочной с выделением m и $p-m$ строк и столбцов, вектор $\hat{\mathbf{a}}$ разбит на два подвектора с размерностью m и $p-m$ соответственно,

$$\hat{h}_j = N^{-1/2} \sum_{k=0}^m \sum_{l=0}^m \hat{a}_k \hat{a}_l m_{l,j-k}, \quad \text{где } j = m+1, \dots, p,$$

$\hat{a}_0 = 1$; \hat{a}_i , $i = 1, \dots, m$ — МНК-оценки; σ_m^2 — оценка дисперсии σ^2 для процесса АР (m); m_{ij} — элемент матрицы M . Тогда тест для проверки гипотезы H_0 имеет следующий вид:

$$\text{принимается } \begin{cases} H_0, & \text{если } \chi_j^2 < \Delta(\alpha) \\ H_1, & \text{если } \chi_j^2 \geq \Delta(\alpha) \end{cases}, \quad j = 1, 2,$$

где $\Delta(\alpha) = \chi_{p-m}^2(1-\alpha)$ — квантиль уровня $(1-\alpha)$ χ^2 -распределения.

Лабораторная работа 9. Исследование методов оценивания процессов АР (p) и СС (q)

Цель работы. Познакомиться и освоить методы оценивания коэффициентов процесса АР (p) и дисперсии шума. Провести сравнительный анализ получаемых оценок.

Порядок выполнения работы

Используя генераторы псевдослучайных чисел, произвести моделирование временных рядов, порождаемых каузальным процессом АР (p).

Задание 1. С помощью классического (формулы (3.12), (3.13)) и рекуррентного (формулы (3.14)–(3.16), (3.13)) МНК оценить коэффициенты процесса АР (p) и дисперсию σ^2 . Исследовать поведение оценок рекуррентного МНК в зависимости от начальных оценок $\hat{a}(0)$, $p(0)$.

Задание 2. Оценить коэффициенты процесса АР (p) и дисперсию шума с помощью метода Юла — Уолкера (формулы (3.20), (3.21)) и с помощью метода Дурбина — Левинсона (формулы (3.22)–(3.26)). Провести сравнение алгоритмов по эффективности (число операций, время) и точности. Используя формулу (3.27), оценить порядок p процесса АР (p). Применяя формулу (3.28), построить доверительные интервалы для получения оценок коэффициентов a_1, \dots, a_p .

Задание 3. Оценить коэффициенты процесса АР (p) с помощью последовательного алгоритма по формулам (3.29)–(3.31) для $p > 1$ и по формулам (3.32)–(3.34) для $p = 1$. Исследовать качество полученных оценок в зависимости от выбора c_n, κ_n, H . Для $p = 1$ проверить неравенство (3.35) при помощи моделирования n временных рядов $x_{11}, \dots, x_{1,N}, x_{21}, \dots, x_{2,N}, \dots$ и $x_{n1}, \dots, x_{n,N}$. Для каждого из этих рядов оценить $\hat{a}_i(H)$, $i = 1, \dots, n$, и математическое ожидание

$$\mathbf{M}\{(\hat{a}(H) - a)^2\} \cong \frac{1}{n} \sum_{i=1}^n (\hat{a}_i(H) - a)^2.$$

Задание 4. Произвести моделирование временного ряда, порожденного процессом СС (q). С помощью метода Дурбина — Левинсона (формулы (3.36)–(3.38)) оценить неизвестные коэффициенты. Используя формулу (3.39), оценить порядок q процесса СС (q), а используя формулу (3.40), построить доверительные интервалы для коэффициентов.

Задание 5. Провести сравнительный анализ алгоритмов оценивания коэффициентов процесса АР (p) по эффективности и точности. Изобразить графически поведение остатков

$$e_t = x_t - \hat{\mathbf{a}}^* \tilde{\mathbf{x}}_{t-1}, \quad t = p + 1, \dots, N,$$

и отклонений оценок $\hat{\mathbf{a}}$ от истинных коэффициентов $\Delta_t = \|\hat{\mathbf{a}}(t) - \mathbf{a}\|^2$, где $\|\cdot\|$ — евклидова норма.

Задание 6. Провести моделирование временного ряда, порождаемого процессом АР (p) с полиномиальным трендом $u_{jt} = t^j$:

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \sum_{j=1}^q c_j u_{jt} + W_t. \quad (3.41)$$

С помощью МНК оценить коэффициенты $a_1, \dots, a_p, c_1, \dots, c_q$. Исследовать влияние тренда на точность оценок процесса АР (p). Провести сравнение с оценками a_1, \dots, a_p при отсутствии тренда.

Задание 7. Произвести моделирование временного ряда, порожденного процессом АР (p). Для различных значений m ($m < p, m = p$) проверить гипотезу H_0 о равенстве нулю коэффициентов a_{m+1}, \dots, a_p с помощью статистик χ_1^2 и χ_2^2 .

3.3. Оценивание параметров процесса АРСС

Любой стационарный ряд можно приблизить с любой степенью точности моделями АР (p) и СС (q), задав p и q достаточно большими. Смешанные модели авторегрессии и скользящего среднего позволяют строить модели, дающие хорошую аппроксимацию с помощью небольшого числа параметров.

Метод Юла — Уолкера. Для оценивания коэффициентов $a_1, \dots, a_p, b_1, \dots, b_q$ можно использовать уравнение Юла — Уолкера

$$\begin{pmatrix} \gamma_q & \gamma_{q-1} & \dots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & \dots & \gamma_{q-p+2} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{q+p-1} & \gamma_{q+p-2} & \dots & \gamma_q \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \gamma_{q+1} \\ \gamma_{q+2} \\ \vdots \\ \gamma_{q+p} \end{pmatrix},$$

матричная форма которого задается соотношением $\Gamma(p, q)\mathbf{a} = \boldsymbol{\gamma}(p, q)$. Решение имеет вид

$$\mathbf{a} = \Gamma^{-1}(p, q)\boldsymbol{\gamma}(p, q). \quad (3.42)$$

Оценка параметра \mathbf{a} получается при замене ковариационных функций γ_s в (3.42) их оценками

$$\hat{\gamma}_s = \frac{1}{N-s} \sum_{t=s+1}^N x_t x_{t-s}.$$

После того, как получена оценка \mathbf{a} , требуется оценить b_1, \dots, b_q . Положим

$$Y_n = \sum_{k=0}^p a_k x_{n-k}.$$

Пусть s_j — ковариация процесса Y , которая определяется следующим образом:

$$s_j = \mathbf{M}\{Y_n Y_{n+j}\} = \sum_{0 \leq k, l \leq p} a_k a_l \gamma_{j+k-l}.$$

Тогда из соотношения (3.1) имеем $s_j = 0$ для $j > q$ и

$$s_j = \mathbf{M} \left\{ \left(\sum_{l=0}^q b_l W_{n-l} \right) \left(\sum_{u=0}^q b_u W_{n+j-u} \right) \right\} = \sigma^2 \sum_{u=j}^q b_u b_{u-j},$$

$$0 \leq j \leq q. \quad (3.43)$$

Необходимо решить систему (3.43). Предположим, что s_j известны, а σ^2 , b_1, \dots, b_q ($b_0 = 1$) неизвестны. В качестве s_j можно использовать ее оценку

$$s_j = \sum_{0 \leq k, l \leq p} \hat{a}_k \hat{a}_l \hat{\gamma}_{j+k-l}. \quad (3.44)$$

Обозначим $y_0 = \sigma$ и $y_j = \sigma b_j$, $1 \leq j \leq q$. Тогда система (3.43) преобразуется к виду

$$\sum_{u=j}^q y_u y_{u-j} = s_j, \quad 0 \leq j \leq q. \quad (3.45)$$

Для $q \geq 2$ эта система не может быть решена аналитически, в общем случае она имеет конечное число различных решений. Заметим также, что искомое решение должно быть таким, при котором корни полинома

$$S(z) = \sum_{l=0}^q y_l z^l$$

лежат вне единичного круга $|z_i| > 1$. Покажем, что решение системы (3.45) эквивалентно нахождению полиномов $S(z)$, которые удовлетворяют соотношениям

$$S(z)S(1/z) = s_0 + \sum_{k=1}^q s_k \left(z^k + \frac{1}{z^k} \right), \quad z \in \mathbf{C},$$

или

$$S(z)S(1/z) = U \left(z + \frac{1}{z} \right) = U(v), \quad (3.46)$$

где U — полином степени q относительно переменной $v = z + 1/z$, коэффициенты которого являются линейными комбинациями s_k , $0 \leq k \leq q$. Рассмотрим три алгоритма решения системы (3.45).

Алгоритм 1. Применяется метод Ньютона нахождения корней векторного уравнения $\mathbf{g}(y) = 0$ вида $\mathbf{y}(n+1) = \mathbf{y}(n) - [g'(y(n))]^{-1} \mathbf{g}(\mathbf{y}(n))$, $\mathbf{y} = (y_0, y_1, \dots, y_q)^*$, где $\mathbf{y}(n)$ — оценка вектора \mathbf{y} на n -м шаге; $y(0)$ — некоторым образом выбранное начальное приближение,

$$\mathbf{g}(y) = \begin{pmatrix} g_0(y) \\ \vdots \\ g_q(y) \end{pmatrix}, \quad g_j(y) = -s_j + \sum_{u=j}^q y_u y_{u-j};$$

$$g'(y) = \begin{pmatrix} \frac{\partial g_0}{\partial y_0} & \frac{\partial g_0}{\partial y_1} & \cdots & \frac{\partial g_0}{\partial y_q} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_q}{\partial y_0} & \frac{\partial g_q}{\partial y_1} & \cdots & \frac{\partial g_q}{\partial y_q} \end{pmatrix}.$$

Если $\lim_{n \rightarrow \infty} \mathbf{y}(n) = \mathbf{y}(\infty)$, то $\mathbf{g}(\mathbf{y}(\infty)) = 0$. Однако это решение может не удовлетворять условию обратимости процесса АРСС. Таким же образом можно найти новые оценки параметров $b_1, \dots, b_q, \sigma^2$, если использовать начальные приближения $y(0)$, отличающиеся от предложенных.

Алгоритм 2. Для поиска решения использовать уравнение (3.46). Найти в \mathbb{C} решение уравнения

$$U(v) = 0. \quad (3.47)$$

Для $q \leq 3$ это может быть сделано явно, для $q > 3$ существует несколько численных методов. Пусть v_1, \dots, v_k — различные корни уравнения (3.47) кратности m_1, \dots, m_k соответственно. Для нахождения корней z_j решить квадратное уравнение $z + 1/z = v_j$, $1 \leq j \leq k$, корни которого обозначим z_j и $1/z_j$, полагая $|z_j| \geq 1$. Если $|z_j| > 1$, то

$$S(z) = \sigma B(z) = c \prod_{1 \leq j \leq k} (z - z_j)^{m_j},$$

где константа c находится из соотношения $S(i) = U(0)$, $i^2 = -1$. В общем случае

$$S(z) = c \prod_{|z_j| > 1} (z - z_j)^{m_j} \prod_{|z_j| = 1} [z^2 - 2z\operatorname{Re}(z_j) + 1]^{1/m_j}.$$

Алгоритм 3. Спектральная плотность g процесса Y имеет вид

$$g(\lambda) = \frac{\sigma^2}{2\pi} |B(z)|^2, \quad \lambda \in [-\pi, \pi], \quad z = e^{-i\lambda}.$$

Аппроксимировать процесс Y процессом АР большого порядка M , а именно:

$$P_M(\Delta)Y_t = W_t.$$

В этом случае спектральная плотность процесса Y_t имеет вид

$$g(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{1}{P_M(z)} \right|^2, \quad \lambda \in [-\pi, \pi], \quad z = e^{-i\lambda}.$$

Таким образом, полином $B(z)$ можно аппроксимировать полиномом

$$\begin{aligned} B_M(z) &= \frac{1}{P_M(z)} = \frac{1}{1 + P_M(z) - 1} = \\ &= 1 - (P_M - 1) + (P_M - 1)^2 + \dots + (-1)^k (P_M - 1)^k + \dots \end{aligned}$$

Коэффициенты этого полинома $b_k(M)$ сходятся к коэффициентам полинома $B(z)$ при $M \rightarrow \infty$, $0 \leq k \leq q$, и к нулю для $k \geq q + 1$. Коэффициенты полинома $P_M(z)$ находятся из уравнения Юла — Уолкера порядка M для процесса Y :

$$\begin{pmatrix} s_1 & s_2 & \dots & s_M \\ s_2 & s_1 & \dots & s_{M-1} \\ \vdots & \vdots & \dots & \vdots \\ s_M & s_{M-1} & \dots & s_1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{pmatrix} = - \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{M-1} \end{pmatrix},$$

где s_j определяется выражением (3.44) ($s_j = 0$, $j \geq q$). Представленный метод параметров по сравнению с предыдущим методом имеет преимущества в случае $q > 3$.

Метод Дурбина — Левинсона. Сначала подбирается уравнение СС порядка m , т.е. вычисляются по формулам (3.36)—(3.38) коэффициенты $\hat{B}_{m,m+j}$, $j = 0, \dots, m - 1$. Порядок m должен удовлетворять условию $m \geq p + q$, и его можно выбрать, например, следующим образом: $m = \inf\{k : |\hat{\rho}(k)| < \varepsilon\}$, где $\varepsilon > 0$ — достаточно малое число, или можно выбрать такое m , при котором достигается минимум критерия АИС (см. гл.4).

Затем оцениваются коэффициенты a_1, \dots, a_p из матричного уравнения

$$\begin{pmatrix} \hat{B}_{m,q+1} \\ \hat{B}_{m,q+2} \\ \vdots \\ \hat{B}_{m,q+p} \end{pmatrix} = \begin{pmatrix} \hat{B}_{m,q} & \hat{B}_{m,q-1} & \dots & \hat{B}_{m,q+1-p} \\ \hat{B}_{m,q+1} & \hat{B}_{m,q} & \dots & \hat{B}_{m,q+2-p} \\ \vdots & \vdots & \dots & \vdots \\ \hat{B}_{m,q+p-1} & \hat{B}_{m,q+p} & \dots & \hat{B}_{m,q} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \quad (3.48)$$

и параметры \hat{B}_j , $j = 1, \dots, q$, по формуле

$$\hat{B}_j = \hat{B}_{m,j} - \sum_{i=1}^{\min(j,p)} \hat{a}_i \hat{b}_{m,j-i}, \quad j = 1, \dots, q. \quad (3.49)$$

Дисперсия оценивается по формуле $\hat{\sigma}^2 = \hat{v}_m$, где \hat{v}_m вычисляется по (3.38).

Метод максимального правдоподобия. Предположим, что процесс АРСС (p, q) является гауссовским и $\mathbf{M}\{W_t\} = 0$, $M\{W_t^2\} = \sigma^2$, а также каузальным и обратимым. Рассмотрим два способа вычисления логарифмической функции правдоподобия.

Алгоритм Бокса — Дженкинса. Пусть векторный параметр (\mathbf{a}, \mathbf{b}) принадлежит некоторому компактному подмножеству U , $(\mathbf{a}, \mathbf{b}) \in U \subset \mathbf{R}^{p+q}$, такому, что для всех точек этого множества процесс АРСС (p, q) является каузальным и обратимым. Пусть также известны наблюдения x_1, \dots, x_N за процессом АРСС (p, q) . Тогда логарифмическую функцию правдоподобия можно записать в виде

$$L_N(\mathbf{a}, \mathbf{b}, \sigma^2) = -\frac{1}{2}[N \ln(2\pi\sigma^2) + \frac{1}{\sigma^2} F_N], \quad (3.50)$$

где

$$F_N = \sum_{-\infty < k \leq N} [\mathbf{M}\{W_k | x_1, \dots, x_N\}]^2 = \sum_{-\infty < k \leq N} |\hat{W}_k^2|.$$

ММП заключается в нахождении параметров $\mathbf{a}, \mathbf{b}, \sigma^2$, максимизирующих (3.50). Эта процедура состоит из двух шагов.

Шаг 1. Найти $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$ такие, что $F_N(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \mathbf{x}_1^N) = \inf_{(a, b) \in U} F_N(\mathbf{a}, \mathbf{b}, \mathbf{x}_1^N)$, где $\mathbf{x}_1^N = (x_1, \dots, x_N)$.

Можно предложить два способа решения задачи.

1. Оценить параметры (\mathbf{a}, \mathbf{b}) методами Юла — Уолкера и Дурбина — Левинсона для параметров процесса АРСС. Затем выбрать конечное множество точек, соседних с $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$. Методом "обратного предсказания", описанным ниже, вычислить величину $F_N(\cdot, \cdot, \mathbf{x}_1^N)$ для каждой точки этого множества. Искомая точка, таким образом, находится прямым перебором.

2. В ситуации, когда $p + q \geq 5$, предыдущий способ малоприменим. Поэтому минимизацию можно произвести одним из градиентных методов. Производные при этом можно аппроксимировать следующим образом:

$$\frac{\partial F_N}{\partial a_k} = \frac{1}{\varepsilon} [F_N(\mathbf{a} + \varepsilon \mathbf{e}_k, \mathbf{b}, \mathbf{x}_1^N) - F_N(\mathbf{a}, \mathbf{b}, \mathbf{x}_1^N)],$$

$$\frac{\partial F_N}{\partial b_k} = \frac{1}{\varepsilon} [F_N(\mathbf{a}, \mathbf{b} + \varepsilon \mathbf{e}_k, \mathbf{x}_1^N) - F_N(\mathbf{a}, \mathbf{b}, \mathbf{x}_1^N)],$$

где $\varepsilon > 0$ — некоторое малое число; \mathbf{e}_k — базисный вектор

$$\mathbf{e}_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} k.$$

В качестве начальной точки можно использовать $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$.

Шаг 2. Минимизировать выражение $N \ln(2\pi\sigma^2) + 1/\sigma^2 \cdot F_N(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \mathbf{x}_1^N)$ по σ^2 . Это даст оценку $\tilde{\sigma}^2$ в форме $\tilde{\sigma}^2 = F_N[\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \mathbf{x}_1^N]/N$.

Итак, для завершения процедуры оценивания осталось рассмотреть вычисление $F_N(\mathbf{a}, \mathbf{b}, \mathbf{x}_1^N)$ при заданных a и b . Из сделанных предположений о процессе АРСС следует, что существует белый шум V_t с дисперсией σ^2 , удовлетворяющей соотношению

$$\sum_{k=0}^p a_k x_{t+k} = \sum_{l=0}^q b_l V_{t+l}, \quad t \in \mathbf{Z}. \quad (3.51)$$

Возьмем условное математическое ожидание (при фиксированном \mathbf{x}_1^N) от соотношений (3.51), (3.1):

$$\sum_{k=0}^p a_k \hat{x}_{t+k} = \sum_{l=0}^q b_l \hat{V}_{t+l}; \quad (3.52)$$

$$\sum_{k=0}^p a_k \hat{x}_{t-k} = \sum_{l=0}^q b_l \hat{W}_{t-l}. \quad (3.53)$$

Алгоритм вычисления F_N

Шаг 1. Выбрать произвольный вектор $\alpha = (\alpha_1, \dots, \alpha_q) \in \mathbf{R}^q$ и положить $\hat{V}_{N-p+j} = \alpha_j$, $1 \leq j \leq q$. Используя формулу (3.52), вычислить V_t . При этом индекс t уменьшать с $N-p$ до единицы. Положить $V_t = 0$ для $t \leq 0$. В результате данного шага получаются значения \hat{V}_t , $t \leq N-p+q$.

Шаг 2. Используя соотношения (3.52), вычислить \hat{x}_t , $t = 0, -1, \dots, -M$. При этом $|M|$ должно иметь достаточно большое значение в случае, когда значения \hat{x}_t , $t \leq -M$, близки к нулю.

Шаг 3. Используя формулу (3.53), вычислить последовательно \hat{W}_t , $t = -M+1, \dots, N$. Положить $W_t = 0$ для $t \geq N+1$.

Шаг 4. Используя формулу (3.53), вычислить последовательно \hat{x}_t , $t = N+1, \dots, N+T$, где T — достаточно большое число и такое, что $\hat{x}_t \cong 0$ для $t > T$.

Шаг 5. Используя формулу (3.52), вычислить последовательно \hat{V}_t , $t = T, T-1, \dots, N-p+1$.

Этот шаг дает новое множество величин $\hat{\alpha}_j = \hat{V}_{N-p+j}$, $1 \leq j \leq q$, которые отличаются от начальных. Данный пошаговый алгоритм позволяет итерационно вычислить вектор $\alpha(k)$ на k -й итерации, а следовательно, и значения $\hat{W}_t(k)$, $t \leq N$, на k -й итерации. Окончательно получаем

$$F_N[\mathbf{a}, \mathbf{b}, \mathbf{x}_1^N] = \lim_{k \rightarrow \infty} \sum_{t \leq N} |\hat{W}_t(k)|^2.$$

В практической ситуации ограничиваемся некоторым достаточно большим числом итераций k . Заметим также, что в случае $q = 0$ требуется одна итерация.

Алгоритм, основанный на одношаговом прогнозе. Логарифмическую функцию правдоподобия для процесса АРСС (p, q) можно записать в виде

$$L_N(\mathbf{a}, \mathbf{b}, \sigma^2) = -\frac{1}{2} [N \ln(2\pi\sigma^2 r_0 \dots r_{N-1}) + \frac{1}{\sigma^2} S_N],$$

где $S_N = \sum_{j=1}^N (x_j - \hat{x}_j)^2 / r_{j-1}$; \hat{x}_j — одношаговый прогноз; $\mathbf{M}\{(x_{j+1} - \hat{x}_{j+1})^2\} = \sigma^2 r_j = v_j$ — дисперсия ошибки прогноза.

Одношаговый прогноз и его дисперсия вычисляются по формулам

$$\hat{x}_{i+1} = \sum_{j=1}^i \theta_{ij} (x_{i+1-j} - \hat{x}_{i+1-j}), \quad 1 \leq i < m = \max(p, q); \quad (3.54)$$

$$\hat{x}_{i+1} = a_1 x_1 + \dots + a_p x_{i+1-p} + \sum_{j=1}^q \theta_{ij} (x_{i+1-j} - \hat{x}_{i+1-j}), \quad i \geq m; \quad (3.55)$$

$$v_0 = \gamma_W(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1} \left[\gamma_W(n+1, k+1) - \sum_{j=0}^{n-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right], \quad k = 0, 1, \dots, n-1, \quad (3.56)$$

$$v_n = \gamma_W(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

Рекуррентное вычисление по формулам (3.54)–(3.56) производится в порядке $v_0; \theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \dots$.

Ковариационная функция процесса W $\mathbf{M}\{W_i W_j\} = \gamma_W(i, j)$ удовлетворяет соотношению

$$\gamma_W(i, j) = \begin{cases} \sigma^{-2} \gamma_{i-j}, & 1 \leq i, j \leq m, \\ \sigma^{-2} \left[\gamma_{i-j} - \sum_{l=1}^p a_l \gamma_{l-|i-j|} \right], & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{l=0}^q b_l b_{l+|i-j|}, & \min(i, j) > m, \\ 0, & \text{иначе,} \end{cases},$$

где $\gamma_i = \mathbf{M}\{x_t x_{t-i}\}$ — ковариационная функция процесса x_t . Процедура поиска неизвестных параметров состоит из двух шагов.

Шаг 1. $S_N(\hat{a}, \hat{b}) = \inf_{(a, b) \in U} S_N(a, b)$.

Шаг 2. $\hat{\sigma}^2 = \frac{1}{N} S_N(\hat{a}, \hat{b})$.

Поиск параметров a, b , минимизирующих $S_N(a, b)$, можно произвести с помощью какого-нибудь алгоритма нелинейной оптимизации.

Если x_t — каузальный обратимый процесс АРСС (p, q) и полиномы $A(z), B(z)$ не имеют общих корней, то оценка параметров \hat{a}, \hat{b} имеет асимптотически нормальное распределение $\sqrt{N}(\hat{\theta} - \theta) \Rightarrow N(0, V(\theta))$, где $\hat{\theta} = (\hat{a}, \hat{b})$; $V(\theta)$ — асимптотическая ковариационная матрица:

$$V(\theta) = \sigma^2 \begin{pmatrix} \mathbf{M}\{\mathbf{U}_t \mathbf{U}_t^*\} & \mathbf{M}\{\mathbf{U}_t \mathbf{V}_t^*\} \\ \mathbf{M}\{\mathbf{V}_t \mathbf{U}_t^*\} & \mathbf{M}\{\mathbf{V}_t \mathbf{V}_t^*\} \end{pmatrix}^{-1}. \quad (3.57)$$

Здесь $\mathbf{U}_t = (u_t, \dots, u_{t+1-p})^*$; $\mathbf{V}_t = (v_t, \dots, v_{t+1-q})^*$; u_t, v_t — процессы авторегрессии $A(\Delta)u_t = W_t, B(\Delta)v_t = W_t$. Для $p = 0$ $V(\theta) = \sigma^2 [\mathbf{M}\{\mathbf{V}_t \mathbf{V}_t^*\}]^{-1}$, для $q = 0$ $V(\theta) = \sigma^2 [\mathbf{M}\{\mathbf{U}_t \mathbf{U}_t^*\}]^{-1}$. Для оценивания элементов можно использовать представление процесса АР в виде бесконечного скользящего среднего:

$$u_t = A^{-1}(\Delta)W_t = \psi_0 W_t + \psi_1 W_{t-1} + \dots; \quad (3.58)$$

$$v_t = B^{-1}(\Delta)W_t = \varphi_0 W_t + \varphi_1 W_{t-1} + \dots. \quad (3.59)$$

Способы вычисления $\{\psi_i\}, \{\varphi_i\}$ даны в §3.1. Для практического вычисления матрицы $V(\theta)$ следует ограничиться конечными рядами (3.57), (3.58). Данное

асимптотическое распределение позволяет получить приближенные доверительные интервалы для оцененных коэффициентов:

$$\{\theta_j \in \mathbf{R} : |\theta_j - \hat{\theta}_j| \leq N^{-1/2} \Phi_{1-\alpha/2}^{-1} v_{jj}^{1/2}(\hat{\theta})\}, \quad (3.60)$$

где $v_{jj}(\hat{\theta})$ — j -й диагональный элемент матрицы $V(\hat{\theta})$, $\hat{\theta} = (\hat{\mathbf{a}}, \hat{\mathbf{b}})$.

Рекуррентный МНК. Если бы в каждый момент времени t величины W_1, \dots, W_t процесса АРСС (p, q) были известны, то для оценки параметров $\hat{\mathbf{a}}, \hat{\mathbf{b}}$ можно было бы применить МНК-оценки

$$\sum_{t=1}^N (x_t - \Phi_{t-1}^* \theta)^2 \rightarrow \min_{\theta},$$

$$\theta = (a_1, \dots, a_p, b_1, \dots, b_q)^*, \quad \Phi_{t-1}^* = (x_{t-1} \dots x_{t-p} W_{t-1} \dots W_{t-q}).$$

Рекуррентная форма данного алгоритма аналогична (3.14)—(3.16):

$$\hat{\theta}(t+1) = \hat{\theta}(t) - P(t) \Phi_t L_t (x_{t+1} - \Phi_t^* \hat{\theta}(t)); \quad (3.61)$$

$$P(t+1) = P(t) - P(t) \Phi_{t+1} L_{t+1} \Phi_{t+1}^* P(t); \quad (3.62)$$

$$L_{t+1} = (1 + \Phi_{t+1}^* P(t) \Phi_{t+1})^{-1}. \quad (3.63)$$

Использование в этом алгоритме вместо неизвестных СВ W_1, \dots, W_t подходящих оценок $\tilde{W}_1, \dots, \tilde{W}_t$ делает процедуру оценивания реализуемой. Оценку \tilde{W}_t можно интерпретировать как прогноз СВ W_t по данным наблюдениям $\hat{\theta}(1), \dots, \hat{\theta}(t)$, $\tilde{W}_1, \dots, \tilde{W}_{t-1}$, x_1, \dots, x_t . Будем вычислять \tilde{W}_t следующим образом:

$$\tilde{W}_t = x_t - \Phi_{t-1}^* \hat{\theta}(t). \quad (3.64)$$

Соотношения (3.61)—(3.64) после задания начальных оценок $\theta(0)$, $P(0)$, \tilde{W}_0 , \tilde{W}_{-1}, \dots полностью определяет рекуррентную процедуру оценивания. Рекуррентная процедура МНК становится малоприменимой, если приходится оценивать вектор параметров высокой размерности: основной объем вычислений связан с процедурой вычисления матриц $P(t)$. Предложим упрощенный вариант МНК, когда в алгоритме оценивания используется не сама матрица $P(t)$, а ее след. Алгоритм имеет вид

$$\hat{\theta}(t+1) = \hat{\theta}(t) - L_t P_{t-1} \Phi_t (x_{t+1} - \Phi_t^* \hat{\theta}(t)); \quad (3.65)$$

$$p_{t+1}^{-1} = p_t^{-1} + \Phi_{t+1}^* \Phi_{t+1}; \quad (3.66)$$

$$L_t = (1 + p_{t-1} \Phi_t^* \Phi_t)^{-1}. \quad (3.67)$$

Данный алгоритм является одним из вариантов *метода стохастической аппроксимации*.

Метод оценивания параметров АРСС с пропущенными наблюдениями. Пусть каузальный и обратимый процесс АРСС описывается уравнением (3.1). Предположим, что можно наблюдать не все величины x_1, x_2, \dots, x_N .

Пусть, например, известны r наблюдений x_{j_1}, \dots, x_{j_r} . Для того, чтобы вычислить логарифмическую функцию правдоподобия, надо искусственно добавить ряд наблюдений до числа N следующим образом. Перепишем уравнение АРСС в виде $Y_{t+1} = FY_t + TW_{t+1}$; $x_t = GY_t$, $t = 1, 2, \dots$, где i -я компонента вектора \mathbf{Y}_t , $i = 1, \dots, k = \max(p, q + 1)$, определяется уравнением

$$(Y_t)_i = \sum_{j=i}^k (a_j x_{t+i-1-j} + b_{j-1} W_{t+i-j}), \quad i = 1, \dots, k;$$

$a_j = 0$ для $j > k$ и $b_j = 0$ для $j > q$; $F = \|f_{ij}\|$, $f_{ij} = a_1 \delta_{ij} + \delta_{i+1,j}$, δ_{ij} — символ Кронекера, $i, j = 1, \dots, k$; $T = (1 \ b_1 \dots b_{k-1})^*$, $G = (1 \ 0 \dots 0)$.

Определим величину $z_t = G_t Y_t + \alpha_t V_t$, $t = 1, 2, \dots$, где $\{V_t\}$ — последовательность независимых гауссовских СВ с нулевым средним и единичной дисперсией, не зависит от Y_1, W_2, W_3, \dots и

$$G_t = \begin{cases} 0, & \text{если } x_t \text{ пропущено} \\ G, & \text{иначе} \end{cases} \quad \alpha_t = \begin{cases} 1, & \text{если } x_t \text{ пропущено} \\ 0, & \text{иначе.} \end{cases}$$

Таким образом, вектор $\mathbf{Z}_N = (z_1 \dots z_N)^*$ совпадает с \mathbf{x}_1^N , за исключением ненаблюдаемых компонент \mathbf{x}_1^N , которые в \mathbf{Z}_N заменяются независимыми стандартными гауссовскими величинами. Тогда одношаговый прогноз величины \hat{z}_j , $j = 1, \dots, N$, получается с помощью алгоритма калмановской фильтрации:

$$\hat{\mathbf{Y}}_{t+1} = F \tilde{\mathbf{Y}}_t; \quad (3.68)$$

$$\Sigma_{t+1} = F \tilde{\Sigma}_t F^* + TT^*; \quad (3.69)$$

$$\hat{z}_t = G_{t+1} \hat{\mathbf{Y}}_{t+1}; \quad (3.70)$$

$$s_{t+1} = G_{t+1} \Sigma_{t+1} G_{t+1}^* + \alpha_{t+1}^2; \quad (3.71)$$

$$K_{t+1} = \Sigma_{t+1} G_{t+1}^* / s_{t+1}; \quad (3.72)$$

$$\tilde{\mathbf{Y}}_{t+1} = \hat{\mathbf{Y}}_{t+1} + K_{t+1} (z_{t+1} - \hat{z}_{t+1}); \quad (3.73)$$

$$\tilde{\Sigma}_{t+1} = \Sigma_{t+1} - K_{t+1} G_{t+1} \Sigma_{t+1}, \quad (3.74)$$

с начальными условиями

$$\begin{aligned} \hat{\mathbf{Y}}_1 &= 0, \quad \Sigma_1 = \mathbf{M}\{\mathbf{Y}_1 \mathbf{Y}_1^*\}, \quad \hat{z}_1 = 0, \\ s_1 &= \alpha_1^2 + \sigma_x^2, \quad K_1 = s_1^{-1} \Sigma_1 G^*, \quad \tilde{\mathbf{Y}}_1 = K_1 z_1, \\ \tilde{\Sigma}_1 &= \Sigma_1 - K_1 G \Sigma_1, \quad \sigma_x^2 = \mathbf{M}\{x_t^2\}. \end{aligned}$$

Логарифмическая функция правдоподобия имеет вид

$$L_N(\mathbf{a}, \mathbf{b}, \sigma^2, \mathbf{x}_1^N) = -\frac{1}{2} \left[r \ln(2\pi\sigma^2) + \sum' \ln s_{i-1} + \frac{1}{\sigma^2} \sum' (x_i - \hat{z}_i)^2 / s_{i-1} \right],$$

где $'$ у суммы означает суммирование только по тем индексам i , для которых x_i наблюдается; s_i, \hat{z}_i определяются по формулам (3.68)–(3.74). Поиск неизвестных параметров осуществляется с помощью алгоритмов нелинейной оптимизации.

Лабораторная работа 10. Оценивание параметров процесса АРСС (p, q)

Цель работы. Познакомиться с методами оценивания коэффициентов процесса АРСС и дисперсии шума, освоить их и провести сравнительный анализ алгоритмов.

Порядок выполнения работы

Используя генераторы псевдослучайных чисел, произвести моделирование временных рядов, порождаемых каузальным и обратимым процессом АРСС (p, q) .

Задание 1. Оценить коэффициенты процесса АРСС (p, q) с помощью метода Юла — Уолкера (алгоритмы 1–3). Провести сравнительный анализ по эффективности и точности оценок коэффициентов, входящих в уравнение скользящего среднего.

Задание 2. Оценить коэффициенты процесса АРСС (p, q) с помощью метода Дурбина — Левинсона. Провести сравнительный анализ по эффективности и точности оценок коэффициентов с оценками из задания 1.

Задание 3. Оценить коэффициенты процесса АРСС (p, q) с помощью метода максимального правдоподобия (алгоритм Бокса — Дженкинса и алгоритм, основанный на прогнозе). Провести сравнительный анализ по эффективности и точности предложенных алгоритмов. Для алгоритма Бокса — Дженкинса подобрать оптимальные значения T и M . Для поиска экстремума функции F_N и S_N использовать градиентные прямые методы, а также методы случайного поиска.

Задание 4. Оценить ковариационную матрицу оценки ММП (формулы (3.57)–(3.59)). Построить доверительные интервалы для коэффициентов по формуле (3.60). Исследовать зависимость качества оценок (величины доверительных интервалов) от значений коэффициентов процесса АРСС. Изобразить графически поведение оценок ММП и доверительные границы.

Задание 5. Оценить коэффициенты процесса АРСС (p, q) с помощью рекуррентного алгоритма МНК (формулы (3.61)–(3.63)) и упрощенного варианта (формулы (3.65)–(3.67)). Исследовать эффективность алгоритмов для оценивания коэффициентов процесса АРСС. Используя рекуррентные алгоритмы МНК, оценить коэффициенты процесса АРСС с трендом (3.41).

Задание 6. Оценить коэффициенты процесса АРСС (p, q) для модели с пропущенными наблюдениями. В качестве модели пропуска наблюдений использовать модель пропуска через фиксированное или случайное число тактов времени. Исследовать влияния числа $m = N - r$ пропущенных наблюдений на точность оценивания.

3.4. Обнаружение разладки процессов АРСС

Пусть $\{x_k, 1 \leq k \leq t\} = x_1^t$ — реализация СП x_t , которая в момент t_0 скачкообразно меняет свои свойства. В общем случае это означает, что до момента t_0

процесс x_t имел распределение $F^0(x_1, \dots, x_{t_0})$, а после этого момента — распределение $F^1(x_{t_0+1}, \dots, x_N)$. С практической точки зрения существует два типа задач, решаемых с помощью алгоритмов обнаружения разладки. В первом случае необходимо обнаруживать разладку как можно быстрее после ее появления при заданном уровне ложных тревог и определить момент времени, когда произошла разладка. Алгоритмы подобного типа называются *последовательными алгоритмами обнаружения разладки*. Второй основной тип задач сводится к оцениванию момента появления разладки после получения всей выборки.

В некоторых случаях сам факт наличия разладки в пределах анализируемой выборки заранее неизвестен, и проверка ее наличия также является предметом решения. Алгоритмы подобного типа называются *алгоритмами апостериорного обнаружения разладки*.

Будем рассматривать параметрические свойства распределений F^0, F^1 , в частности, процессы АРСС.

Апостериорные алгоритмы. Пусть две части временного ряда $x_1^{t_0-1}$ и $x_{t_0}^N$ порождаются стационарными моделями АР (p) и $x_1^{t_0-1}$ не зависит от $x_{t_0}^N$, т.е.

$$x_t = \sum_{i=1}^p a_i^{(j)} x_{t-i} + W_t, \quad (3.75)$$

где $a_1^{(j)}, \dots, a_p^{(j)}$ — коэффициенты АР до ($j = 1$) и после ($j = 2$) разладки; σ_j^2 — дисперсия независимой последовательности W_t , $\mathbf{M}\{W_t\} = 0$, $\mathbf{M}\{W_t^2\} = \sigma_j^2$. Если векторные параметры $\theta_j^* = (a_1^{(j)} \dots a_p^{(j)} \sigma_j^2)$, $j = 1, 2$, известны, то для определения момента t_0 можно использовать ММП:

$$\begin{aligned} t_0 &= \arg \max_{p+1 \leq k \leq N-p} [\ln p(x_1^{k-1} | \theta_1) + \ln p(x_k^N | \theta_2)] = \\ &= \arg \max_{p+1 \leq k \leq N-p} \left[\ln p(x_1^p | \theta_1) + \sum_{t=p+1}^{k-1} \ln p(x_t | x_{t-p}^{t-1}, \theta_1) + \right. \\ &\quad \left. + \ln p(x_k^{k+p-1} | \theta_2) + \sum_{t=k+p}^N \ln p(x_t | x_{t-p}^{t-1}, \theta_2) \right], \end{aligned} \quad (3.76)$$

где $p(\cdot | \cdot)$ — условная плотность. Рассмотрим случай гауссовской модели, т.е. $W_t \sim N(0, \sigma_j^2)$, и предположим, что $N \gg p$. Тогда (3.76) можно аппроксимировать следующим образом:

$$\begin{aligned} t_0 &= \arg \max_{p+1 \leq k \leq N-p} \left[-\frac{k}{2} \ln(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2} \sum_{t=p+1}^k e_{t1}^2 - \right. \\ &\quad \left. - \frac{n-k}{2} \ln(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2} \sum_{t=k+1}^N e_{t2}^2 \right], \end{aligned} \quad (3.77)$$

$$e_{tj}^2 = \left(x_t - \sum_{i=1}^p a_i^{(j)} x_{t-i} \right)^2, \quad j = 1, 2.$$

Если θ_1 и θ_2 неизвестны, то в (3.77) используются оценки $\hat{\theta}_1, \hat{\theta}_2$. Оценки $\hat{\theta}_1$ рассчитываются по значениям выборки x_1^k , а $\hat{\theta}_2$ — по значениям x_k^N , в то время как k изменяется от $p+3$ до $N-p-1$.

В рамках данного подхода можно исследовать модель разладки авторегрессии для случая, когда последовательности $x_1^{t_0-1}$ и $x_{t_0}^N$ зависимы. Первую модель разладки можно рассматривать как момент переключения с наблюдения генератора авторегрессионной последовательности (3.75) с параметрами θ_1 на наблюдение генератора с параметрами θ_2 . Вторая модель разладки соответствует не переключению наблюдения, а скачкообразному изменению коэффициентов генератора (3.75) с сохранением "хвоста" памяти $x_{t_0}^{t_0-1}$. Для новой модели разладки соотношение (3.76) заменяется на следующее:

$$t_0 = \arg \max_{p+1 \leq k \leq N-p} \ln \frac{p(\mathbf{x}_1^{k-1} | \theta_1)}{p(\mathbf{x}_k^N | \theta_2)},$$

где θ_1, θ_2 известны.

В случае гауссовской АР данный алгоритм имеет вид

$$t_0 = \arg \max_{p+1 \leq k \leq N-p} \left[-k \ln \frac{\sigma_1}{\sigma_2} + \frac{1}{2\sigma_2^2} \sum_{t=p+1}^k e_{t2}^2 - \frac{1}{2\sigma_1^2} \sum_{t=p+1}^k e_{t1}^2 \right]. \quad (3.78)$$

Для отыскания неизвестных параметров можно использовать метод Юла — Уолкера или рекуррентный МНК.

Последовательные алгоритмы. Существуют два подхода для обнаружения разладки. Первый состоит в пропуске наблюдаемого сигнала x_t через выбеливающий фильтр и в проверке отклонения характеристик последовательности остатков e_t данного фильтра от параметров белого шума. Будем называть его *одномоделным*. Второй подход основан на использовании двух моделей, которые оцениваются на различных участках временного ряда и сравниваются между собой с помощью подходящей меры расстояния между распределениями. Будем называть его *двухмоделным*. Существуют также два алгоритма оценивания момента изменения свойств t_0 .

Алгоритм 1. Данный алгоритм основан на использовании порогового решающего правила для статистик, представленных в виде кумулятивных сумм:

$$S_t = \sum_{k=1}^t s_k, \quad (3.79)$$

где s_k — некоторая статистика. Статистика s_k подбирается таким образом, чтобы до разладки S_t в среднем дрейфовала вниз, а после разладки — вверх. При этом правило подачи сигнала о разладке имеет вид

$$t_0 = \inf\{t \geq 1 : g_t \geq h\}, \quad (3.80)$$

где $g_t = S_t - \min_{1 \leq k \leq t} S_k$; $h > 0$ — порог. Поведение суммы S_t до и после разладки изображено на рис. 3.1.

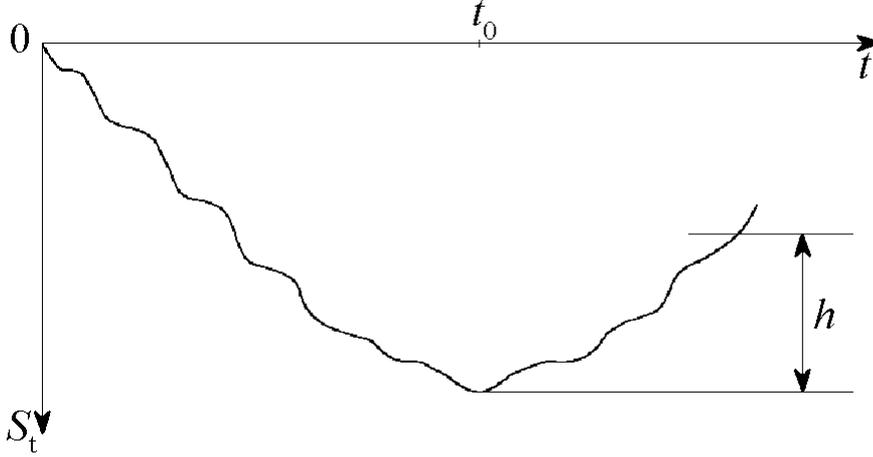


Рис. 3.1

Опишем алгоритмы обнаружения разладки временного ряда, порождаемого процессом АР (p). Все алгоритмы основаны на статистиках вида (3.79). Рассмотрим идеальный случай, когда параметры АР (p) известны до и после изменения свойств. Пусть СП x_t описывается известными условными плотностями $p^1(x_t | \mathbf{x}_1^{t-1}) = p(x_t | \mathbf{x}_1^{t-1}, \theta_1)$ до изменения и $p^2(x_t | \mathbf{x}_1^{t-1}) = p(x_t | \mathbf{x}_1^{t-1}, \theta_2)$ — после изменения.

Одномодельный подход для обнаружения разладки использует статистику

$$U_t = \sum_{k=1}^t u_k, \quad (3.81)$$

где

$$u_k = \int p^1(x | \mathbf{x}_1^{k-1}) \ln p^1(x | \mathbf{x}_1^{k-1}) dx - \ln p^1(x | \mathbf{x}_1^{k-1}).$$

Статистика U_t до изменения имеет нулевое среднее и принимает положительное значение после изменения, если выполняется условие

$$H_t^x(p^2) > H_t^x(p^1), \quad (3.82)$$

где $H_t^x(p^j)$, $j = 1, 2$ — условная энтропия: $H_t^x(p^j) = - \int p^j(x | \mathbf{x}_1^{k-1}) \times \ln p^j(x | \mathbf{x}_1^{k-1}) dx$.

В гауссовском случае для модели АР (p) статистика U_t принимает вид

$$U_t = \frac{1}{2} \sum_{k=1}^t \left(\frac{e_{k1}^2}{\sigma_1^2} - 1 \right). \quad (3.83)$$

Условие (3.82) изменяется следующим образом: $\sigma_2 > \sigma_1$. Если же $\sigma_2 < \sigma_1$, то после изменения среднее статистики U_t может быть положительным, отрицательным или даже нулевым. В этом случае не рекомендуется использовать тест U_t .

При применении двухмодельного подхода можно использовать три статистики.

1. Статистика, основанная на отношении правдоподобия:

$$V_t = \sum_{k=1}^t v_k,$$

где $v_k = \ln[p^2(x_k | \mathbf{x}_1^{k-1})/p^1(x_k | \mathbf{x}_1^{k-1})]$.

2. Статистика, основанная на расстоянии между условными распределениями, определяемом дивергенцией Кульбака:

$$D_t = \sum_{k=1}^t d_k,$$

где

$$\begin{aligned} d_k &= - \int p^1(x | \mathbf{x}_1^{k-1}) \ln \frac{p^2(x | \mathbf{x}_1^{k-1})}{p^1(x | \mathbf{x}_1^{k-1})} dx + \ln \frac{p^2(x_k | \mathbf{x}_1^{k-1})}{p^1(x_k | \mathbf{x}_1^{k-1})} = \\ &= -I_k^x(p^2 | p^1) + \ln \frac{p^2(x_k | \mathbf{x}_1^{k-1})}{p^1(x_k | \mathbf{x}_1^{k-1})}, \end{aligned}$$

где $I_k^x(p^2 | p^1)$ — информационное число Кульбака. Условные математические ожидания этих двух статистик V_t и D_t до изменения и после него имеют вид

$$\mathbf{M}_1\{v_t | \mathbf{x}_1^{t-1}\} = -I_k^x(p^1 | p^2) < 0,$$

$$\mathbf{M}_1\{d_t | \mathbf{x}_1^{t-1}\} = 0,$$

$$\mathbf{M}_2\{v_t | \mathbf{x}_1^{t-1}\} = I_k^x(p^2 | p^1) > 0,$$

$$\mathbf{M}_2\{d_t | \mathbf{x}_1^{t-1}\} = J(p^1, p^2) > 0,$$

где $\mathbf{M}_j\{\cdot | \cdot\}$ — математическое ожидание по распределению $p^j(\cdot | \cdot)$; $J(p^1, p^2)$ — дивергенция Кульбака между условными плотностями.

3. Статистика, основанная на расстоянии Чернова между условными распределениями:

$$Z_t = \sum_{k=1}^t z_k,$$

где приращения z_k таковы, что

$$\mathbf{M}_1\{z_k | \mathbf{x}_1^{k-1}\} = \text{const},$$

$$\mathbf{M}_2\{z_k | \mathbf{x}_1^{k-1}\} = f_k(\tau) = \ln \int [p^2(x | \mathbf{x}_1^{k-1})]^\tau [p^1(x_k | \mathbf{x}_1^{k-1})]^{1-\tau} dx,$$

где $f_k(\tau)$ — расстояние Чернова между условными плотностями p^1 , p^2 , $0 < \tau < 1$.

Теперь рассмотрим случай гауссовской АР. Запишем статистики v_t, d_t, z_t следующим образом:

$$v_t = \frac{1}{2} \ln \frac{\sigma_1^2}{\sigma_2^2} + \frac{e_{t1}^2}{2\sigma_1^2} - \frac{e_{t2}^2}{2\sigma_2^2}; \quad (3.84)$$

$$d_t = \frac{1}{2} \left[-2 \frac{e_{t1}^2 e_{t2}^2}{2\sigma_2^2} + \left(1 + \frac{\sigma_1^2}{\sigma_2^2} \right) \frac{e_{t1}^2}{2\sigma_1^2} - 1 + \frac{\sigma_1^2}{2\sigma_2^2} \right]; \quad (3.85)$$

$$z_t = z_t(\tau) = \frac{\tau(\tau-1)}{2} \frac{e_{t1}(e_{t1} - e_{t2})}{\tau\sigma_2^2 + (1-\tau)\sigma_1^2} - \frac{1}{2} \ln \frac{\sigma_2^{2\tau} \sigma_1^{2(1-\tau)}}{\tau\sigma_2^2 + (1-\tau)\sigma_1^2}, \quad (3.86)$$

где e_{tj} , $j = 1, 2$, определяются формулой (3.77).

В частном случае, когда коэффициенты АР не изменяются (т.е. $e_{t1} = e_{t2}$), а изменяется только дисперсия, формула (3.86) задается соотношением

$$2z_t(\tau) = \ln(\tau\sigma_2^2 + (1-\tau)\sigma_1^2) - \tau \ln \sigma_2^2 - (1-\tau) \ln \sigma_1^2. \quad (3.87)$$

В случае, когда дисперсия остается постоянной (т.е. $\sigma_1 = \sigma_2$), статистика принимает вид

$$z_t(\tau) = \frac{\tau(\tau-1)}{2} \frac{e_{t1}(e_{t1} - e_{t2})}{\sigma^2} = \frac{\tau(\tau-1)}{2} d_t.$$

На практике модели АР до и после изменения неизвестны и должны быть идентифицированы. При использовании единственной модели, т.е. когда разладку определяют с помощью U_t , АР-модель идентифицируют (оценивают параметры $a_1, \dots, a_p, \sigma^2, p$ с помощью рекуррентных алгоритмов (например, МНК).

При использовании двухмодельного подхода необходимо выбрать расположение двух участков временного ряда для идентификации моделей. Первую АР-модель \mathbf{M}_1 идентифицируют в фиксированном окне (например, T значений в начале временного ряда) и используют ее в качестве проверяемой модели. Другими словами, взяв значения x_{t1}, \dots, x_{t1+T} , оценивают параметры АР-модели $a_1^{(1)}, \dots, a_p^{(1)}, \sigma_1^2$. Затем вторую модель \mathbf{M}_2 идентифицируют в скользящем окне такого же размера и применяют в качестве проверяемой модели (т.е. оценивают коэффициенты $a_1^{(2)}, \dots, a_p^{(2)}, \sigma_2^2$). Когда модели существенно отличаются друг от друга, находят момент разладки, вторая модель становится первой и т.д. В качестве меры различия двух моделей можно использовать статистику

$$f_T = \frac{1}{T} \sum_{t=1}^T (e_{t1} - e_{t2})^2,$$

где обновляющие последовательности e_{t1} и e_{t2} определяются на окнах \mathbf{M}_1 и \mathbf{M}_2 соответственно. Этот способ имеет недостаток, заключающийся в возрастании вероятности ложной тревоги. Об этом свидетельствует простой факт: вследствие идентификации опорной модели \mathbf{M}_1 внутри временного окна заведомо

ограниченных размеров часть информации о сигнале до изменения теряется. Положение можно изменить к лучшему, если в качестве опорной использовать глобальную (т.е. оцененную в длинном временном окне) модель \mathbf{M}'_1 вместо локальной (т.е. оцененной в коротком временном окне) модели (рис. 3.2).

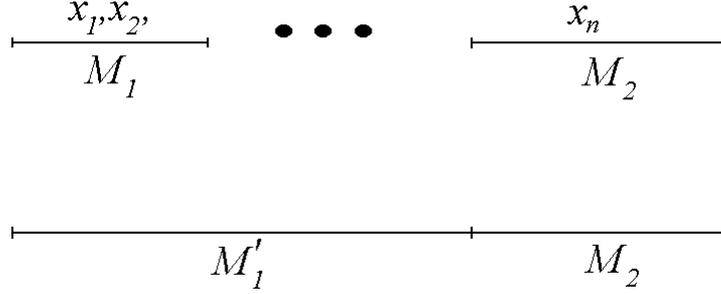


Рис. 3.2

В общем случае, значения $x_{t_1}, \dots, x_{t_1+t-k}$ определяют окно \mathbf{M}'_1 , а значения x_t, \dots, x_{t+T} — окно \mathbf{M}_2 (t — текущее время). Если $t_1 = 1$, $k = 0$, окна $\mathbf{M}'_1, \mathbf{M}_2$ соответствуют окнам, изображенном на рис. 3.2.

Алгоритм 2. Пусть T — длина временного окна \mathbf{M}_2 и пусть $\hat{\sigma}_1^2$ и $\hat{\sigma}_2^2$ — оцененные дисперсии остатков моделей \mathbf{M}'_1 и \mathbf{M}_2 соответственно:

$$\hat{\sigma}_1^2 = \frac{1}{t-k} \sum_{i=t_1}^{t_1+t-k} (\hat{e}_{i1} - \bar{e}_1)^2,$$

$$\hat{\sigma}_2^2 = \frac{1}{T} \sum_{i=t}^{t+T} (\hat{e}_{i2} - \bar{e}_2)^2,$$

$$\bar{e}_1 = \frac{1}{t-k} \sum_{i=t_1}^{t_1+t-k} \hat{e}_{i1}, \quad \bar{e}_2 = \frac{1}{T} \sum_{i=t}^{t+T} \hat{e}_{i2},$$

$$\hat{e}_{i1} = x_i - \sum_{j=1}^p \hat{a}_j^{(1)} x_{i-j}, \quad \hat{e}_{i2} = x_i - \sum_{j=1}^p \hat{a}_j^{(2)} x_{i-j},$$

где t — текущее время; $\hat{a}_j^{(1)}, \hat{a}_j^{(2)}$ — коэффициенты АР-модели, оцененные в окнах \mathbf{M}'_1 и \mathbf{M}_2 соответственно. Тогда статистика обнаружения задается формулой (3.86), в которой $\tau = T/t$, и ее можно записать в виде $z_t(T) = H(1, t) - H(1, t-T) - H(t-T+1, t)$, где $H(a, b) = (b-a+1) \ln \frac{\hat{\sigma}^2[a, b]}{b-a+1}$.

Пусть также обнаружение произошло в момент времени t_* и истинный момент изменения находится в интервале $[t_* - T + 1, t_*]$. Тогда вычисляются разности для моментов t от $t_* + 1$ до $t_* + T - 1$ по формуле $\Delta_t = Z_t(T) - Z_t(t - \tilde{t}_0 + 1)$,

где $\tilde{t}_0 = t_* - T + 1$, и оценивается момент изменения

$$\hat{t}_0 = t_M - T + 1, \quad t_M = \inf\{t_* + 1 \leq t \leq t_* + T - 1 : \Delta_t > 0\}. \quad (3.88)$$

Лабораторная работа 11. Обнаружение разладки процессов АР (р)

Цель работы. Познакомиться и освоить методы обнаружения разладки и оценивания моментов изменения свойств процессов АР (р). Провести сравнительный анализ алгоритмов.

Порядок выполнения работы

Используя генераторы псевдослучайных чисел, произвести моделирование временных рядов, порождаемых процессом АР (р). В некоторые моменты времени (детерминированные или случайные) процесс АР (р) меняет свои свойства (коэффициенты, математическое ожидание или дисперсию шума). Следует учитывать, что участки стационарности или однородности (процесс не меняет свои свойства) должны быть достаточно большими.

Задание 1. Произвести моделирование временного ряда, порождаемого процессом АР (р), с одним моментом разладки и первой моделью разладки. С помощью ММП (3.77) оценить момент разладки. Используя ММП (3.78), оценить единственный момент разладки для второй модели. Исследовать качество алгоритмов в зависимости от величины разладки $\Delta = \|a^{(1)} - a^{(2)}\|$.

Задание 2. Произвести моделирование временного ряда, порождаемого процессом АР (р), с несколькими моментами разладки. В качестве разладки использовать изменение только коэффициентов a_1, \dots, a_p , только дисперсии σ^2 и коэффициентов и дисперсии одновременно. Предположить, что параметры до и после разладки известны точно. Оценить моменты разладки, используя статистики $U_t, V_t, D_t, Z_t(\tau)$, правила остановки (3.80) и правила остановки (3.88). Исследовать качество предложенных методов в зависимости от величины "скачка параметров" $\Delta = \|\hat{\theta} - \theta\|^2$, $\theta = (a, \sigma^2)$.

Задание 3. Оценить моменты разладки временных рядов аналогично заданию 2. При этом предположить, что параметры до и после разладки неизвестны. Параметры оценить с помощью метода Дурбина — Левинсона (включая порядок р). Для оценивания использовать окна фиксированной длины в первом случае. Во втором случае опорную модель оценивать в увеличивающемся окне. Исследовать качество алгоритмов в зависимости от выбора длины фиксированного окна T (в первом случае), величины смещения k , длины T (во втором случае) и величины "скачка параметров" Δ .

Задание 4. Для каждого из последовательных алгоритмов исследовать правило остановки (3.80) в зависимости от выбора порога h . Оценить среднюю величину запаздывания при обнаружении разладки. Для этого провести моделирование n временных рядов с единственной разладкой, в один и тот же момент

времени, оценить для каждого ряда момент разладки t_i и вычислить среднюю величину

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i.$$

Глава 4

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

4.1. Прогнозирование тренда

Прогнозирование (предсказание) временного ряда является важной практической задачей. Вообще говоря, определение математического описания временного ряда, его параметров — это вспомогательная задача, которую необходимо решить, чтобы воспользоваться ею для предсказания. Наиболее типичное предсказание — прогноз погоды. Однако погода описывается сложно, зависит от многих детерминированно и случайно изменяющихся факторов и не может служить примером временного ряда. Обратимся к более простым ситуациям.

Временной ряд, как в большинстве рассмотренных случаев, представим в виде

$$x_t = m_t + y_t, \quad t \in T = \{t_1, t_2, \dots, t_N\}, \quad (4.1)$$

где m_t — детерминированная или стохастическая составляющая, которая порождается определенным закономерным образом, быть может, неизвестным исследователю; y_t — случайная составляющая, независимая от m_t , являющаяся реализацией некоторого ССП. Для определенности в дальнейшем считаем, что $t_i < t_{i+1}$ для всех i , и пишем вместо x_{t_i} более кратко x_i , полагая $t_i = i$.

В рамках такой модели могут быть сформулированы следующие задачи, связанные с предсказанием:

- 1) по наблюдениям $\{x_1, x_2, \dots, x_N\}$ предсказать значение m_{N+k} , $k > 0$;
- 2) по наблюдениям $\{x_1, x_2, \dots, x_N\}$ предсказать значение x_{N+k} , $k > 0$.

При решении первой задачи возможны два варианта: m_t является детерминированным трендом, m_t — случайный процесс.

Поскольку в (4.1) m_t и y_t независимы между собой, при решении второй задачи требуется вначале решить первую, затем, имея в виду, что y_t — ненаблюдаемый процесс, определить его как остатки $x_t - m_t$ и осуществить предсказание значения y_{N+k} по множеству остатков $\{x_i - m_i, 1 \leq i \leq N\}$. Из вышеуказанных свойств следует, что процесс $x_t - m_t$ является стационарным СП, поэтому возникает задача предсказания данного процесса. Таким образом, будут найдены предсказанное значение \hat{m}_{N+k} и предсказанное значение \hat{y}_{N+k} , сумма которых и даст предсказанное значение \hat{x}_{N+k} .

Прежде чем переходить к конкретным оценкам предсказания, рассмотрим некоторые общие вопросы. Будем считать, что временной ряд $\{x_t, t \in T\}$, $T = \{1,$

$2, \dots, N\}$, порождается СП с ограниченным вторым моментом, т.е. $\mathbf{M}\{x_t^2\} < +\infty \forall t$. Предсказание величины $x_{N+\tau}$ временного ряда, $\tau > 0$, по наблюдениям $\{x_1, x_2, \dots, x_N\}$ сводится к нахождению детерминированной функции N аргументов, значения которой можно было бы принимать в качестве $x_{N+\tau}$, т.е.

$$\hat{x}_{N+\tau} = f(x_1, x_2, \dots, x_N), \quad \tau > 0. \quad (4.2)$$

Среди всевозможных функций (4.2) представляет интерес та, которая дает наилучшее предсказание. В анализе временных рядов принято считать наилучшим такое предсказание (прогнозирование), которое обеспечивает минимальную среднеквадратичную ошибку предсказания, иными словами функция (4.2) должна быть решением следующего экстремального уравнения:

$$\mathbf{M}\{(x_{N+\tau} - \hat{x}_{N+\tau})^2\} = \mathbf{M}\{(x_{N+\tau} - f(x_1, x_2, \dots, x_N))^2\} \rightarrow \min_{f(\cdot)}. \quad (4.3)$$

Эта задача имеет единственное решение

$$\hat{x}_{N+\tau} = f(x_1, x_2, \dots, x_N) = \mathbf{M}\{(x_{N+\tau} | x_1, x_2, \dots, x_N)\}. \quad (4.4)$$

Таким образом, оптимальное предсказание — это вычисление условного математического ожидания предсказываемого значения при фиксированных наблюдаемых значениях временного ряда.

Рассмотрим случай нормального распределения. Предположим, что имеются наблюдения N значений временного ряда $\{x_1, x_2, \dots, x_N\}$. Составим из них вектор-столбец $\mathbf{x}(1, N) = (x_1, x_2, \dots, x_N)^*$. Полагаем, что структура временного ряда определена формулой (4.1), где m_t — детерминированная компонента, которая в данном случае определяет математическое ожидание x_t так, что

$$\mathbf{M}\{\mathbf{x}(1, N)\} = \mathbf{m}(1, N) = (m_1, m_2, \dots, m_N)^*. \quad (4.5)$$

Необходимо по наблюдениям $\mathbf{x}(1, N)$ построить прогноз значений временного ряда для значений параметра, определяемых неравенствами $N + k \leq t \leq N + k + n - 1$, т.е. с учетом вектора $\mathbf{x}(1, N)$ предсказать значения $\{x_{N+k}, x_{N+k+1}, \dots, x_{N+k+n-1}\}$. Для удобства обозначим $\hat{\mathbf{x}}(1, n) = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)^*$, где \hat{x}_j — оценка $x_{N+k+j-1}$. Итак, зная вектор $\mathbf{x}(1, N)$ прошлых значений временного ряда, необходимо оценить n будущих значений этого временного ряда, представленного вектором $\hat{\mathbf{x}}(1, n)$. Предполагается, что распределение значений временного ряда нормальное, так что совместная плотность $\mathbf{x}(1, N)$ и $\hat{\mathbf{x}}(1, n)$ имеет вид

$$p(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sqrt{\det K}}{\sqrt{(2\pi)^{n+N}}} e^{-\frac{1}{2} \begin{pmatrix} \mathbf{x} - \mathbf{m} \\ \hat{\mathbf{x}} - \hat{\mathbf{m}} \end{pmatrix}^* K \begin{pmatrix} \mathbf{x} - \mathbf{m} \\ \hat{\mathbf{x}} - \hat{\mathbf{m}} \end{pmatrix}}, \quad (4.6)$$

где K — матрица, обратная матрице ковариаций составного вектора $(\mathbf{x}^*, \hat{\mathbf{x}}^*)$, т.е.

$$K^{-1} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{x\hat{x}} \\ \Sigma_{\hat{x}x} & \Sigma_{\hat{x}\hat{x}} \end{pmatrix}, \quad (4.7)$$

\mathbf{m} — вектор математических ожиданий $\mathbf{x}(1, N)$; $\hat{\mathbf{m}}$ — вектор математических ожиданий вектора $\hat{\mathbf{x}}(1, n)$.

Пусть векторы \mathbf{m} , $\hat{\mathbf{m}}$ и матрица K известны. Тогда наилучшее предсказание, соответствующее (4.4), приводит к соотношению

$$\hat{\mathbf{x}}(1, n) = \hat{\mathbf{m}}(1, n) + \Sigma_{\hat{x}x} \Sigma_{xx}^{-1} (\mathbf{x}(1, N) - \mathbf{m}(1, N)). \quad (4.8)$$

При этом корреляционная матрица ошибок предсказания $\varepsilon(1, n) = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^*$, $\varepsilon_j = \hat{x}_j - x_{N+k+j-1}$, имеет структуру

$$\Sigma_{\varepsilon\varepsilon} = \Sigma_{\hat{x}\hat{x}} - \Sigma_{\hat{x}x} \Sigma_{xx}^{-1} \Sigma_{x\hat{x}}. \quad (4.9)$$

Характерной особенностью оптимальной оценки предсказания нормально распределенного временного ряда является тот факт, что предсказание значения временного ряда — линейные функции наблюдений, т.е. выражения (4.8) можно представить в виде

$$\hat{\mathbf{x}}(1, N) = \mathbf{a} + B\mathbf{x}(1, N), \quad (4.10)$$

где \mathbf{a} — среднее значение прогноза $\mathbf{a} = \hat{\mathbf{m}}(1, n) - \Sigma_{\hat{x}x} \Sigma_{xx}^{-1} \mathbf{m}(1, N)$. $B = \Sigma_{\hat{x}x} \Sigma_{xx}^{-1} - (n \times N)$ -матрица коэффициентов регрессии предсказываемых значений имеющихся наблюдений. Из полученного результата следует, что для построения прогноза на n значений временного ряда по N наблюдениям необходимо знать векторы математических ожиданий $\mathbf{m}(1, N)$, $\hat{\mathbf{m}}(1, n)$ и корреляционные матрицы Σ_{xx} , $\Sigma_{x\hat{x}}$, $\Sigma_{\hat{x}x}$, $\Sigma_{\hat{x}\hat{x}}$. Иными словами, при построении оптимального прогноза (4.4) требуется большое количество сведений о предсказываемом СП.

Если временной ряд порождается СП, отличающимся от нормального, необходимо еще знать многомерное совместное распределение данного процесса. Это требование является наиболее трудновыполнимым. Проблемы, связанные с прогнозированием случайных процессов в общем случае, будут обсуждены в последующих параграфах. Обратимся к задаче прогнозирования значений детерминированного тренда m_t в модели (4.1).

Пусть наблюдается временной ряд (4.1) с неизвестным детерминированным трендом m_t и случайной компонентой y_t . Требуется по N наблюдениям $\{x_1, x_2, \dots, x_N\}$ найти прогноз значений $\{m_{N+k}, m_{N+k+1}, \dots, m_{N+k+n-1}\}$ тренда m_t . Для этой задачи надо воспользоваться техникой оценивания функции регрессии (§ 2.2), которая сводится к следующему. Пусть наблюдения временного ряда известны для некоторого множества T значений параметра t . Необходимо определить прогноз детерминированного тренда для значений t из множества \hat{T} , $\hat{T} \cap T = \emptyset$. Объединим эти множества: $\mathfrak{S} = T \cup \hat{T}$. Выберем отрезок числовой оси $[a, b]$, в который бы погружалось множество \mathfrak{S} , т.е. $\mathfrak{S} \subseteq [a, b]$. Введем линейное преобразование параметра t так, чтобы отрезок $[a, b]$ отображался в $[+1, -1]$. Исходя из практических соображений, связанных с конкретным характером временного ряда, выберем систему ортогональных или линейно независимых функций $\{\varphi_j(u)\}$. Методом наименьших квадратов найдем коэффициенты

разложения m_t по базису $\{\varphi_j(u)\}$, а именно:

$$m(u) = \sum_{j=1}^s c_j \varphi_j(u) = \mathbf{c}^* \boldsymbol{\varphi}(u), \quad (4.11)$$

где переменная u связана с параметром t соотношением $u = (2t - b - a)/(b - a)$;

$$\mathbf{c} = \left(\sum_{i=1}^N \boldsymbol{\varphi}(u_i) \boldsymbol{\varphi}^*(u_i) \right)^{-1} \sum_{i=1}^N x_i \boldsymbol{\varphi}(u_i). \quad (4.12)$$

Таким образом, прогнозируемые значения временного ряда $m_{N+k+l-1} = m(u_l)$, $1 \leq l \leq n$, определяются по формуле

$$m(u_l) = \mathbf{c}^* \boldsymbol{\varphi}(u_l), \quad u_l \in \hat{U}, \quad (4.13)$$

где \mathbf{c} задано в (4.12); \hat{U} — образ множества \hat{T} , соответствующий описанному линейному преобразованию параметра t в u . Возвращаясь теперь к исходной переменной $t_l = ((b - a)u_l + a + b)/2$, получим оцениваемые значения $\hat{m}_{N+k+l-1}$.

Лабораторная работа 12. Прогнозирование нормально распределенных временных рядов

Цель работы. Исследовать характер изменения качества прогнозирования значений временного ряда при изменении корреляционных свойств временного ряда, а также при изменении срока прогнозирования.

Порядок выполнения работы

Задание 1. Пусть x_t — процесс СС (q), определяемый разностным уравнением

$$x_t = \sum_{j=0}^q b_j W_{t-j} + m_t, \quad (4.14)$$

где W_t — стандартный белый шум с нормальным распределением. Задать функцию m_t . Задать зависимость b_j от j (например, последовательность чисел одного знака, не увеличивающихся по абсолютной величине с ростом j). Произвести имитацию процесса (4.14).

Задание 2. По заданной выборке $\{x_t, 1 \leq t \leq N\}$ построить оценку предсказания $\{\hat{x}_{N+\tau}, \tau = 1, 2, 3\}$. Построить доверительные интервалы для каждой из оценок $\hat{x}_{N+\tau}$, задавшись уровнем значимости α . Графически представить на одном рисунке для некоторого диапазона N доверительные интервалы, реализацию x_N , оценку предсказания $\hat{x}_{N+\tau}$.

Задание 3. Вычислить матрицу корреляции $\Sigma_{\varepsilon\varepsilon}$ ошибок предсказания $\{\hat{x}_{N+\tau}, \tau = 1, 2, 3\}$. Исследовать зависимость дисперсий ошибок ($\hat{x}_{N+\tau} - x_{N+\tau}$)

и $\det \Sigma_{\varepsilon\varepsilon}$ от величины q , которую в данном случае можно назвать временем корреляции временного ряда (4.14). Учтеть, что корреляционная функция процесса (4.14) вычисляется по формуле

$$r_x(k) = \sum_{j=0}^q b_j b_{j+|k|}, \quad |k| \leq q; \quad r_x(k) = 0, \quad |k| > q.$$

Задание 4. По заданной выборке $\{x_t, 1 \leq t \leq N\}$ найти оценку предсказания $\hat{x}_{N+\tau}$ для некоторых значений q (например, $q = 1, 3, 5$). Построить доверительные интервалы для каждого из вариантов при некотором уровне значимости α . Графически представить на одном рисунке для некоторого диапазона N доверительные интервалы, реализацию x_N и оценку предсказания $\hat{x}_{N+\tau}$ во всех вариантах изменения q .

Задание 5. Вычислить матрицу корреляции $\Sigma_{\varepsilon\varepsilon}$ ошибок предсказания $(\hat{x}_{N+\tau} - x_{N+\tau})$. Исследовать зависимость дисперсий ошибок от величины τ при различных q .

Лабораторная работа 13. Прогнозирование значений тренда

Цель работы. Освоить технику оценивания функции регрессии с целью прогнозирования ее значений.

Порядок выполнения работы

Использовать временной ряд, который исследовался в лабораторной работе 12.

Задание. Определить множества значений параметра временного ряда, в которых берутся наблюдения (T) и в которых надо построить прогноз (\hat{T}). Рассматривая $\hat{T} \cup T$, определить отрезок $[a, b]$, на котором следует восстановить функцию регрессии. Осуществить прогнозирование по формулам (4.11)–(4.13). Определить качество прогнозирования, пользуясь критерием $\Delta = \sum_{t \in \hat{T}} |\hat{m}_t - m_t| / |\hat{T}|$. Исследовать зависимость качества прогнозирования Δ от объема выборки временного ряда N и от величины набора базисных функций s .

4.2. Предсказание ССП

Значения СП y_t в модели (4.1) являются ненаблюдаемыми, однако они влияют на предсказываемые значения временного ряда x_t . Если m_t — известная функция своего параметра или ее оценка уже получена способом, описанным в гл.2, можно ввести разности $x_t - m_t$, которые и характеризуют значения СП y_t . Предсказывая значения y_t для необходимого значения параметра t и используя известное или предсказанное значение m_t на основе (4.1), можно получить предсказанное значение временного ряда.

Рассмотрим проблему предсказания y_t , полагая, что $\mathbf{M}\{y_t\} = 0$, а корреляционная функция $r(\tau) = \mathbf{M}\{y_t y_{t+\tau}\}$ этого процесса является известной. Пусть значения СП $\{y_1, y_2, \dots, y_N\}$ известны и необходимо оценить y_{N+k} , $k > 0$. В формуле (4.4) определена оптимальная оценка прогнозируемого значения, в (4.8) она построена для нормального распределения. В частности, для случая оценивания СП y_t с нулевым средним и корреляционной функцией $r(\tau)$ оптимальный прогноз (4.8) трансформируется в выражение

$$\begin{aligned} \hat{y}_{N+k} &= (r(N+k-1), r(N+k-2), \dots, r(k))R^{-1}(y_1, y_2, \dots, y_N)^* = \\ &= \mathbf{r}(N+k-1, k)R^{-1}\mathbf{y}(1, N), \end{aligned} \quad (4.15)$$

где корреляционная $(N \times N)$ -матрица $R = (R_{ij})$ составлена из элементов $R_{ij} = r(i-j)$, а структура строки $\mathbf{r}(N+k-1, k)$ и столбца $\mathbf{y}(1, N)$ понятна из (4.15). В представлении (4.10) эта оценка имеет нулевое среднее a и матрицу коэффициентов регрессии, которая вырождается в строку

$$\mathbf{B} = (b_1, b_2, \dots, b_N) = \mathbf{r}(N+k-1, k)R^{-1}. \quad (4.16)$$

С учетом этого оценку предсказания для нормального ССП можно представить в виде

$$\hat{y}_{N+k} = \sum_{i=1}^N b_i y_i = \mathbf{B}\mathbf{y}(1, N), \quad (4.17)$$

где коэффициенты b_i определены равенством (4.16).

Если процесс y_t не является нормальным, вычисление условного математического ожидания (4.4) обычно затруднительно. Поэтому в среднеквадратичной теории анализа временных рядов принято строить оценки предсказания в классе линейных оценок. Предположим также, что выполняется обычно реализующееся на практике условие

$$r(0) > 0, \quad \lim_{\tau \rightarrow \infty} r(\tau) = 0. \quad (4.18)$$

Тогда оценку предсказания будем искать в форме

$$\hat{y}_{N+k} = f(y_1, y_2, \dots, y_N) = \sum_{l=1}^N b_{N,l} y_{N-l+1}. \quad (4.19)$$

Подставляя (4.19) в (4.17) и разрешая это уравнение, при выполнении условия (4.18) получаем, что коэффициенты регрессии $b_{N,k}$ оптимальной оценки предсказания определяются из равенства

$$(b_{N,1}, b_{N,2}, \dots, b_{N,N}) = (r(k), r(k+1), \dots, r(N+k-1))R^{-1}. \quad (4.20)$$

Дисперсия $D_k(N)$ ошибки предсказания по формулам (4.19), (4.20) выражается равенством

$$D_k(N) = \mathbf{M}\{(\hat{y}_{N+k} - y_{N+k})^2\} =$$

$$= r(0) - \mathbf{r}(k, N + k - 1)R^{-1}\mathbf{r}^*(k, N + k - 1). \quad (4.21)$$

Как видно из сравнения (4.15), (4.19) и (4.20), для нормального процесса y_t оптимальная линейная оценка предсказания полностью совпадает с общей оптимальной оценкой.

Часто по значениям $\{y_1, y_2, \dots, y_N\}$ необходимо предсказать следующее значение y_{N+1} . Для этого случая имеет место предельный результат. Обозначим $D_1 = D_1(+\infty)$, тогда

$$\log D_1 = \lim_{N \rightarrow \infty} \frac{1}{N} \log \det R(N), \quad (4.22)$$

где D_1 — минимальная дисперсия оценки предсказания, определяющая предельную точность предсказания на один шаг; $R(N) = (R_{ij})$, $1 \leq i, j \leq N$, $R_{ij} = r(i - j)$.

Между оптимальными коэффициентами регрессии (4.20) и точностью предсказания (4.21) для $k = 1$ может быть установлено взаимно однозначное соответствие, которое позволяет производить их вычисление по рекуррентным формулам (метод Дурбина — Левинсона). Если выполнено условие (4.18) и обозначено $D_1(0) = r(0)$, $b_{1,1} = r(1)/r(0)$, то справедливы рекуррентные формулы

$$\begin{aligned} D_1(N) &= D_1(N-1)(1 - b_{N,N}^2), \\ b_{N,N} &= \frac{1}{D_1(N-1)} \left(r(N) - \sum_{j=1}^{N-1} b_{N-1,j} r(N-j) \right), \\ b_{N,k} &= b_{N-1,k} - b_{N,N} b_{N-1,N-k}, \quad 1 \leq k \leq N-1. \end{aligned} \quad (4.23)$$

Иногда может оказаться более подходящим представление линейной оценки предсказания, которое отличается от (4.17) и для $k = 1$ записывается в виде

$$\hat{y}_1 = 0, \quad \hat{y}_{N+1} = \sum_{j=1}^N \theta_{N,j} (y_{N+1-j} - \hat{y}_{N+1-j}), \quad N > 0, \quad (4.24)$$

где коэффициенты $\theta_{N,j}$ вычисляются рекуррентно по формулам:

$$\begin{aligned} \theta_{N,N} &= v(N+1, 1)/v(1, 1); \\ \theta_{N,N-k} &= \frac{1}{D_1(k)} \left(v(N+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{N,N-j} D_1(j) \right); \\ &1 \leq k \leq N-1, \\ D_1(k) &= v(k+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}^2 D_1(j), \quad 1 \leq k \leq N, \end{aligned} \quad (4.25)$$

$$D_1(0) = v(1, 1).$$

Здесь в отличие от предыдущего обозначено $\mathbf{M}\{y_i y_j\} = v(i, j)$. Если y_t — ССП, то $v(i, j) = r(i - j)$. По рекуррентным формулам (4.25) величины $D_1(k)$ и $\theta_{k,j}$ вычисляются в следующей последовательности: $\theta_{11}, D_1(1); \theta_{22}, \theta_{21}, D_1(2); \theta_{33}, \theta_{32}, \theta_{31}, D_1(3)$ и т.д.

Рекуррентные формулы Дурбина — Левинсона определяют коэффициенты в (4.19) при прошлых значениях y_t , в то время как формулы (4.25) определяют коэффициенты при обновлениях (инновациях) $(y_t - \hat{y}_t)$ в ортогональном разложении (4.24). Это разложение является очень простым при использовании. Заметим, что если положим $\theta_{N,0} = 1$, то

$$y_{N+1} = \sum_{j=0}^N \theta_{N,j} (y_{N+1-j} - \hat{y}_{N+1-j}), \quad N = 0, 1, 2, \dots,$$

т.е. текущее значение процесса y_t представится в виде разложения по обновлениям $(y_j - \hat{y}_j)$. Оценка предсказания типа (4.24), (4.25) может быть получена и для прогноза на k шагов. Она приобретает форму

$$\hat{y}_{N+k}^{(k)} = \sum_{j=k}^{N+k-1} \theta_{N+k-1,j} (y_{N+k-j} - \hat{y}_{N+k-j}^{(1)}), \quad (4.26)$$

где коэффициенты $\theta_{N+k-1,j}$ по-прежнему вычисляются при помощи (4.25); $\hat{y}_N^{(i)}$ — оценка предсказания на i шагов по $(N-i)$ наблюдениям. Средний квадрат ошибки предсказания может быть подсчитан по формуле

$$\begin{aligned} D_k(N) &= \mathbf{M}\{(y_{N+k-j} - \hat{y}_{N+k-j})^2\} = \\ &= v(N+k, N+k) - \sum_{j=1}^N \theta_{N+k-1, N+k-j} D_1(j). \end{aligned} \quad (4.27)$$

Рассмотрим некоторые важные случаи.

Процесс АРСС (p, q) . Пусть y_t — каузальный процесс АРСС (p, q) . Пользуясь формулами (4.24), (4.25), можно было бы построить процесс предсказанных значений для \hat{y}_t . Причем в этом случае каждое \hat{y}_{t+1} определялось бы t инновациями $(y_j - \hat{y}_j)$. В данном случае удастся построить более простой процесс предсказанных значений, в котором для определения оценки \hat{y}_{t+1} требуется только $m = \max(p, q)$ инноваций $(y_j - \hat{y}_j)$ независимо от t . Этот процесс строится следующим образом. Мы предполагаем, что y_t описывается уравнением

$$y_t = \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^q b_j W_{t-j}, \quad (4.28)$$

где a_i, b_j — известные константы; W_t — процесс белого шума. Тогда

$$\begin{aligned}\hat{y}_{N+1} &= \sum_{j=1}^N \theta_{N,j} (y_{N+1-j} - \hat{y}_{N+1-j}), \quad 1 \leq N < m, \\ \hat{y}_{N+1} &= \sum_{i=1}^p a_i y_{N+1-i} + \sum_{j=1}^q \theta_{N,j} (y_{N+1-j} - \hat{y}_{N+1-j}), \quad N \geq m.\end{aligned}\quad (4.29)$$

Причем $D_1(N) = \mathbf{M}\{(y_{N+1-j} - \hat{y}_{N+1-j})^2\} = \sigma^2 D_1^0(N)$. Величины $D_1^0(N)$ и $\theta_{N,j}$ определяются формулами (4.25), в которых $D_1(k)$ необходимо заменить на $D_1^0(k)$, а $v(i, j)$ вычислить следующим образом:

$$v(i, j) = \begin{cases} r(i-j)/\sigma^2, & 1 \leq i, j \leq m; \\ \left(r(i-j) - \sum_{k=1}^p a_k r(k-|i-j|) \right) / \sigma^2, & \min(i, j) \leq m, \\ & m \leq \max(i, j) \leq 2m; \\ \sum_{k=0}^q b_k b_{k+|i-j|}, & \min(i, j) > m; \\ 0, & \text{в остальных случаях.} \end{cases}\quad (4.30)$$

Заметим, что если $v(i, j)$ определяется соотношением (4.30), а в (4.25) вместо $D_1(k)$ используется $D_1^0(k) = D_1(k)/\sigma^2$, то $D_1^0(k)$ и $\theta_{i,j}$, определяемые формулами (4.25), не зависят от σ^2 . В (4.30) $r(\tau)$ — функция ковариации процесса (4.28); σ^2 — дисперсия белого шума W_j . Напомним, что функция ковариации процесса (4.28) имеет вид

$$r(\tau) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|\tau|}, \quad \tau = 0, \pm 1, \pm 2, \dots, \quad (4.31)$$

где значения ψ_j могут быть вычислены рекуррентно (предполагается, что $b_0 = 1, a_j = 0$ для $i > p; b_j = 0$ для $j > q$) по формулам

$$\begin{aligned}\psi_0 &= 1; \\ \psi_k &= b_k + \sum_{i=0}^{k-1} a_{k-i} \psi_i, \quad 1 \leq k < \max(p, q+1), \\ \psi_k &= \sum_{i=k-p}^{k-1} a_{k-i} \psi_i = \sum_{i=1}^p a_i \psi_{k-i}, \quad k \geq \max(p, q+1).\end{aligned}\quad (4.32)$$

Таким образом, для построения процесса, предсказывающего временной ряд типа АРСС (p, q) на один шаг вперед, можно воспользоваться либо формулами (4.24), (4.25) и (4.31), (4.32) (с учетом того, что в этом случае $v(i, j) = r(i-j)/\sigma^2$), либо формулами (4.29)–(4.32).

Заметим, что в рассматриваемом случае, если процесс АРСС (p, q) является не только каузальным, но и обратимым, при $N \rightarrow \infty$ в (4.29) $D_1(N) \rightarrow \sigma^2$, $\theta_{N,j} \rightarrow b_j$, $1 \leq j \leq q$.

Процесс АР (р). Пусть y_t — процесс АР (p) , который получается из (4.28), если положить $b_j = 0$, $1 \leq j \leq q$. Тогда

$$\hat{y}_{N+1} = \sum_{i=1}^p a_i y_{N+1-i}, \quad N \geq p, \quad (4.33)$$

т.е. в оценке (4.29) $\theta_{N,j} = 0$, $1 \leq j \leq q$. Пусть теперь y_t — процесс СС (q) , который получается из (4.28), если положить $a_i = 0$, $1 \leq i \leq p$. В этом случае

$$\hat{y}_{N+1} = \sum_{j=1}^{\min(N,q)} \theta_{N,j} (y_{N+1-j} - \hat{y}_{N+1-j}), \quad N \geq 1. \quad (4.34)$$

Коэффициенты $\theta_{N,j}$ находятся из (4.25), где ковариации $v(i, j)$ вычисляются по формулам

$$v(i, j) = \sum_{k=0}^{q-|i-j|} b_k b_{k+|i-j|}, \quad v(i, j) = 0, \quad |i-j| > q. \quad (4.35)$$

Процесс АРСС (1,1). Пусть y_t — процесс АРСС $(1,1)$. В этом случае $y_t = ay_{t-1} + W_t + bW_{t-1}$, $|a| < 1$. Уравнение оценки предсказания имеет вид

$$\hat{y}_{N+1} = ay_N + \theta_{N1}(y_N - \hat{y}_N), \quad N \geq 1. \quad (4.36)$$

Используя (4.31), (4.32) и (4.30), получаем

$$v(i, j) = \begin{cases} (1 + 2ab + b^2)/(1 - a^2), & i = j = 1, \\ (1 + b^2), & i = j \geq 2, \\ b, & |i-j| = 1, \quad i \geq 1, \\ 0, & \text{в остальных случаях,} \end{cases}$$

а из (4.25) с учетом этого равенства

$$\theta_{N,1} = b/D_1^0(N-1), \quad D_1^0(N) = (1 + b^2 - (b/D_1^0(N-1))).$$

Описанные результаты можно обобщить на случай предсказания на k шагов вперед процесса АРСС (p, q) . Тогда оценка предсказания для процесса (4.28)

$$\hat{y}_{N+k}^{(k)} = \sum_{i=1}^p a_i \hat{y}_{N+k-i}^{(k-i)} + \sum_{j=k}^q \theta_{N+k-1,j} (y_{N+k-j} - \hat{y}_{N+k-j}^{(1)}), \quad (4.37)$$

где второе слагаемое обращается в нуль при $k > q$.

Среднее значение квадрата ошибки предсказания по (4.37) выражается формулой

$$\begin{aligned}
D_k(N) &= \mathbf{M}\{(y_{N+k} - \hat{y}_{N+k}^{(k)})^2\} = \\
&= \sum_{j=0}^{k-1} \left(\sum_{l=0}^j \chi_l \theta_{N+k-l-1, j-l} \right)^2 D_1(N+k-j-1), \quad (4.38)
\end{aligned}$$

где $D_1(N)$ и $\theta_{i,j}$ определяются в (4.25); χ_l может быть вычислена рекуррентно:

$$\chi_0 = 1, \quad \chi_l = \sum_{j=1}^{\min(p,l)} a_j \chi_{l-j}, \quad l = 1, 2, \dots \quad .$$

Если рассматривать процесс АРСС (p, q) не только каузальный, но и обратимый, при $N \rightarrow \infty$ получаются следующие асимптотические формулы:

$$\hat{y}_{N+k}^{(k)} = \sum_{i=1}^p a_i \hat{y}_{N+k-i}^{(k)} + \sum_{j=k}^q b_j (y_{N+k} - \hat{y}_{N+k}^{(1)}); \quad (4.39)$$

$$D_k(N) = \sigma^2 \sum_{j=0}^{k-1} \left(\sum_{l=0}^j \chi_l b_{j-l} \right)^2 = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2, \quad (4.40)$$

где ψ_j вычисляются по формулам (4.32).

Лабораторная работа 14. Предсказание стационарных временных рядов

Цель работы. Познакомиться и освоить методы предсказания значений временных рядов. Исследовать в интересах сравнительного анализа различные подходы.

Порядок выполнения работы

В качестве основы для построения стационарного временного ряда выбрать модель АРСС (p, q) (4.28), коэффициенты которой $\{a_i\}$, $\{b_j\}$ обеспечивают каузальность и обратимость моделируемому процессу y_t . По формулам (4.31), (4.32) вычислить корреляционную функцию процесса y_t .

Задание 1. Построить оценки предсказания временного ряда на один шаг вперед по формулам (4.16), (4.17); (4.17), (4.23); (4.24), (4.25). Сравнить сложность получения оценок в этих трех случаях. Оценки сравнить на некотором интервале изменения t . Вычислить дисперсию ошибок предсказания на этом интервале. По формуле (4.22) определить предельную дисперсию. Проконтролировать получение этого результата при помощи первой формулы (4.23) и второй формулы (4.25). Результаты представить в виде графиков. Повторить исследование для нескольких различных вариантов наборов параметров $p, q, \{a_i\}, \{b_j\}$. При этом установить зависимость изменения качества оценок предсказания от тенденции изменения параметров модели.

Задание 2. Воспользоваться формулой (4.26) для оценивания прогноза временного ряда на k шагов вперед. Определить точность передсказания на k шагов вперед по формуле (4.27). Исследовать точность предсказания в зависимости от срока предсказания k и объема выборки N , по которой производится предсказание. Результаты представить в виде графиков.

Задание 3. Использовать формулы (4.29), (4.30) для получения оценок предсказания на один шаг. Для интервала значений t , использованного в задании 1, сравнить эти оценки с оценками, полученными в задании 1. Сравнить также дисперсии ошибок предсказания. Установить скорости сходимости коэффициентов, вычисляемых по формулам (4.30), (4.25), к коэффициентам $\{b_j\}$ модели АРСС (p, q) . Повторить исследования тех же вариантов наборов параметров $p, q, \{a_i\}, \{b_j\}$. Провести сравнение.

Задание 4. Использовать формулы (4.37), (4.38) для построения оценок предсказания на k шагов вперед и определения точности предсказания для различных сроков прогнозирования k и объемов выборки N . Получить приближенные оценки предсказания и точности по формулам (4.39), (4.40) и установить их эффективность по сравнению с точными формулами (4.26), (4.27). Результаты представить в виде графиков.

Задание 5. Исследовать качество прогнозирования процессов АР (p) с помощью формулы (4.33). Для этого выбрать модель АР (p) , в которой зависимость $\{a_i\}$ от i обеспечивала бы каузальность модели, с одной стороны, и не очень быструю сходимость к нулю коэффициентов a_i с ростом i , с другой стороны (например, указанным свойствам удовлетворяют коэффициенты $a_i = (-a)^i$, где a — положительное число, не превышающее единицу, но близкое к ней). Исследовать качество прогнозирования выбранного процесса АР (p) в зависимости от p и $\{a_i\}$.

Задание 6. Выбрать модель СС (q) с коэффициентами $\{b_j\}$, обеспечивающими обратимость модели, аналогично способу определения $\{a_i\}$ в задании 5.

Воспользоваться формулами (4.34), (4.35) для предсказания значений процесса y_t , порождаемого выбранной моделью. Исследовать качество предсказания в зависимости от q и $\{b_j\}$.

Задание 7. Выбрать модель АРСС $(1,1)$, каузальную и обратимую. Построить оценки предсказания по формуле (4.36) для процесса, порождаемого этой моделью. Исследовать зависимость качества предсказания от значений параметров a и b .

4.3. Построение моделей на основе предсказания

Для исследования временного ряда $\{x_t, 1 \leq t \leq N\}$, когда его математическое описание неизвестно, первоочередная задача — построение математической модели процесса, порождающего этот ряд. Вначале решается задача выделения тренда (функции регрессии), которая рассматривалась в § 4.1. Важный частный

случай, когда тренд является полиномом или имеет полиномиально описываемую сезонную (циклическую) составляющую, будет описан в § 4.4. Предположим, что тренд выделен и исключен из выборки, так что наблюдаемые данные представляются в виде $\{y_t, 1 \leq t \leq N\}$, где $y_t = x_t - m_t$, и их можно рассматривать как некоторый стационарный процесс, который можно аппроксимировать подходящим процессом АРСС. Задача заключается в том, чтобы задать подходящий процесс АРСС. Процесс АРСС (p, q) , порождаемый белым шумом, определен в (4.28). Вообще говоря, для всякого фиксированного набора $\{y_t, 1 \leq t \leq N\}$ можно подобрать такие p и q , $\{a_i, 1 \leq i \leq p\}$, $\{b_j, 1 \leq j \leq q\}$, что среднеквадратичное рассогласование (выборочная дисперсия белого шума) может быть сделана сколь угодно малым.

В частности, если $y_1 \neq 0$, то можно выбрать $p = N - 1$, а в качестве $\{a_i, 1 \leq i \leq N - 1\}$ решение уравнения

$$Y\mathbf{a} = \mathbf{y}, \quad (4.41)$$

где $\mathbf{a} = (a_1, a_2, \dots, a_N)^*$; $\mathbf{y} = (y_1, y_2, \dots, y_N)^*$; $Y = (Y_{ij}) = ((N - 1) \times (N - 1))$ -матрица с элементами $Y_{ij} = y_{i-j+1}$, $i \geq j$, $Y_{ij} = 0$, $i < j$. В этом случае выборка $\{y_t, 1 \leq t \leq N\}$ будет порождаться совершенно точно при помощи модели (4.28), когда $W_j \equiv 0$. Однако построенная модель будет давать большую ошибку предсказания значения y_{N+1} при достаточно больших N .

Примем следующий принцип построения наиболее подходящих моделей для наблюдаемого множества $\{y_t\}$. Предположим, что при помощи какого-либо метода оценены $p, q, \{\hat{a}_i\}$ и $\{\hat{b}_j\}$ на основе наблюдений $\{y_t, 1 \leq t \leq N\}$. Пусть $\{Y_t, 1 \leq t \leq N\}$ — другая выборка наблюдений этого же процесса, но независимая от $\{y_t\}$, по которой вычислялись оценки параметров модели. Составим теперь по наблюдениям $\{Y_t\}$ и оценкам параметров, определенным по выборке, оценку предсказания Y_{N+1} . Дисперсия оценки такого предсказания и выбирается в качестве критерия согласованности полученных оценок параметров наблюдаемой модели. Оказывается, что эта дисперсия не уменьшается монотонно с ростом p и q , а принимает свое минимальное значение при некоторых p и q процесса АРСС (p, q) , порождающего наблюдаемую выборку $\{y_t, 1 \leq t \leq N\}$.

Опишем вначале этот подход для частного случая процесса АР (p) . Пусть $\{y_t, 1 \leq t \leq N\}$ — реализация каузального процесса АР (p) , определяемого коэффициентами $\{a_i, 1 \leq i \leq p\}$, $p < N$; $\{Y_t, 1 \leq t \leq N\}$ — независимая реализация того же самого процесса. Если $\{\hat{a}_i, 1 \leq i \leq p\}$ — оценки коэффициентов $\{a_i\}$, основанные на $\{y_t, 1 \leq t \leq N\}$, то оценка предсказания, вычисляемая по формуле (4.33), имеет дисперсию

$$d_1(p) = \mathbf{M}\left\{\left(Y_{t+1} - \sum_{i=1}^p \hat{a}_i Y_{t+1-i}\right)^2\right\} = \sigma^2 + \mathbf{M}\{((\hat{\mathbf{a}}(p) - \mathbf{a}(p))^* \mathbf{Y})^2\},$$

где $\mathbf{Y} = (Y_N, Y_{N-1}, \dots, Y_{N-p+1})^*$ — вектор последних p наблюдений из выборки $\{Y_t\}$; $\mathbf{a}(p) = (a_1, a_2, \dots, a_p)^*$ — вектор истинных значений параметров модели;

$\hat{\mathbf{a}}(p)$ — аналогичный вектор их оценок; σ^2 — дисперсия белого шума модели АР (p). Рассматривая второе слагаемое как условное математическое ожидание при фиксированных $\{y_t, 1 \leq t \leq N\}$ и используя независимость $\{Y_t\}$ от $\{y_t\}$, получаем

$$d_1(p) = \sigma^2 + \mathbf{M}\{(\hat{\mathbf{a}}(p) - \mathbf{a}(p))^* R(p)(\hat{\mathbf{a}}(p) - \mathbf{a}(p))\},$$

где $R(p) = (R_{ij}(p))$ — $(p \times p)$ -матрица корреляций $R_{ij}(p) = \mathbf{M}\{Y_i Y_j\} = r(i - j)$, $1 \leq i, j \leq N$. Если в качестве $\hat{\mathbf{a}}(p)$ выбираются оценки Юла — Уолкера, то для достаточно больших N можно считать, что вектор $\mathbf{N}^{-1/2}(\hat{\mathbf{a}}(p) - \mathbf{a}(p))$ распределен нормально с нулевым средним и матрицей ковариации $R^{-1}(p)$. Применяв это свойство, приходим к выводу, что при достаточно больших N имеет место $d_1(p) \cong \sigma^2 \left(1 + \frac{p}{N}\right)$. Осталось оценить σ^2 . Оценка МНК для σ^2 , обладает тем свойством, что при достаточно больших N отношение $N\hat{\sigma}^2/\sigma^2$ распределено приблизительно по χ^2 с $(N - p)$ степенями свободы. Оценка дисперсии белого шума имеет вид

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N ((y_t - \hat{y}_t)^2 / D_1^0(t - 1)). \quad (4.42)$$

Здесь \hat{y}_t — оценка предсказания y_t , которая находится по формулам (4.25) и (4.30), где вместо параметров $\{a_i\}$ используются их оценки, вычисленные по наблюдениям $\{y_t\}$.

Суммируя все это, получаем статистику

$$\text{FRE} = \frac{N + p}{N(N - p)} \sum_{t=1}^N \frac{(y_t - \hat{y}_t)^2}{D_1^0(t - 1)}, \quad (4.43)$$

которая и принимается в качестве критерия, определяющего наиболее подходящее значение порядка p ; наиболее подходящее p минимизирует FRE (Final Prediction Error).

Перейдем теперь к каузальным моделям АРСС (p, q), которые порождают процесс $\{y_t\}$, описываемый уравнением (4.28). Обозначим для удобства $\mathbf{c} = (a_1, \dots, a_p, b_1, \dots, b_q)^*$ — вектор, составленный из $p + q$ параметров модели АРСС (p, q). Параметры p, q в этом случае будем находить не с точки зрения минимизации дисперсии оценки предсказания, а исходя из более общего требования максимизации функции правдоподобия оценки предсказания, которую можно построить для достаточно больших N , пользуясь асимптотической нормальностью распределения оценки. Это приводит к статистике

$$\text{AIC} = N \ln \left(\frac{1}{N} \sum_{t=1}^N \frac{(y_t - \hat{y}_t)^2}{D_1^0(t - 1)} \right) + \sum_{t=1}^N \ln D_1^0(t - 1) + 2(p + q), \quad (4.44)$$

где оценка предсказания \hat{y}_t определяется по формулам (4.29), в которых коэффициенты $\theta_{N, N-k}$, дисперсии $D_1^0(t)$ вычисляются с помощью (4.30), (4.25).

Заметим, что (4.44) не зависит от σ^2 . В качестве наиболее подходящих параметров p и q выбираются такие, которые минимизируют значение статистики АИС (Akaike Information Criterion).

Практика использования АИС показала, что он имеет тенденцию переоценивать p . Для того чтобы скорректировать эту тенденцию, предлагается использовать модифицированный критерий, который для каузального обратимого процесса АРСС (p, q) с нулевым средним основан на статистике

$$\text{BIC} = (N - p - q) \ln \left(\frac{N\sigma^2}{N - p - q} \right) + (p + q) \ln \left(\frac{N\hat{r}(0) - N\hat{\sigma}^2}{p + q} \right), \quad (4.45)$$

где σ^2 вычисляется по формуле (4.42) с учетом того, что оценка предсказания определяется по общим формулам (4.29), $\hat{r}(0)$ — выборочная оценка дисперсии $\{y_t\}$. Наиболее подходящими значениями параметров p и q будут такие, которые минимизируют статистику ВИС (Binomial Information Criterion).

При обработке реальных данных необходимо иметь в виду, что подбираемая модель АРСС (p, q) служит только аппроксимацией реально существующей модели, которая вовсе не обязана относиться к классу АРСС процессов. Поэтому понятие истинных значений p и q имеет только относительный смысл. Для реального процесса может существовать несколько АРСС процессов, достаточно хорошо аппроксимирующих его. Все они могут иметь различающиеся p и q . Поэтому принято выбирать модели АРСС, для которых соответствующая статистика попадает в некоторую ε -окрестность минимального значения (для АИС типичное значение $\varepsilon = 2$), а окончательный выбор модели основывается на дополнительных соображениях, таких, как простота модели или “качество случайности” остатков. Из рассмотренных критериев чаще всего используется АИС.

После того как описанными методами определена наиболее подходящая АРСС (p, q) -модель, с учетом ее структуры (4.28) найдем предсказанное значение $\hat{y}_t(\hat{c})$ величины y_t , основанное на предыдущих значениях $\{y_i, 1 \leq i \leq t-1\}$, и вычислим остатки по формуле

$$\hat{W}_t = (y_t - \hat{y}_t(\hat{c})) / (D_1^0(t-1))^{1/2}, \quad 1 \leq t \leq N. \quad (4.46)$$

Если предполагаемая АРСС (p, q) -модель действительно порождает заданный процесс $\{y_t\}$, то $\{\hat{W}_t\}$ должны образовывать процесс белого шума. При этом дисперсия \hat{W}_t должна быть равна единице.

Первое суждение о $\{\hat{W}_t, 1 \leq t \leq N\}$ можно получить из ее графика. Несмотря на то что трудно судить о корреляционной структуре по графику, тем не менее на нем часто хорошо просматривается отклонение от среднего значения (которым в данном случае является нуль), наличие тренда или периодической составляющей, непостоянство дисперсии (обычно выражающееся в виде наличия некоторой модуляции процесса \hat{W}_t).

Далее представляет интерес исследовать выборочную корреляционную функцию (ВКФ). Нормированные значения ВКФ вычисляются по формуле

$$\hat{\rho}_W(\tau) = \sum_{t=1}^{N-\tau} (\hat{W}_t - \bar{W}_t)(\hat{W}_{t+\tau} - \bar{W}) / \sum_{t=1}^N (\hat{W}_t - \bar{W})^2, \quad \tau = 1, 2, \dots, \quad (4.47)$$

где $\bar{W} = \frac{1}{N} \sum_{t=1}^N \hat{W}_t$ — выборочное среднее значение $\{\hat{W}_t\}$. Поскольку каждое \hat{W}_t определяется параметрами \hat{c} , вычисляемыми на основе выборки $\{y_t, 1 \leq t \leq N\}$, нельзя считать, что \hat{W}_t — НОР СВ. Поэтому вид $\hat{\rho}_W(\tau)$ не такой, как в случае, когда \hat{W}_t являются НОР СВ. Можно уточнить поведение $\hat{\rho}(\tau)$. Для каузального обратимого процесса АРСС (p, q) введем $a(z) = 1 - a_1 z - \dots - a_p z^p$, $b(z) = 1 + b_1 z + \dots + b_q z^q$, $c(z) = (a(z)b(z))^{-1} = 1 + c_1 z + c_2 z^2 + \dots + c_k z^k + \dots$.

Пусть $\hat{\rho}_W = (\hat{\rho}_W(1), \hat{\rho}_W(2), \dots, \hat{\rho}_W(\tau))^*$, где τ — фиксированное положительное целое число. Положим $\tau > p + q$. Пусть далее C и S — матрицы размером $(\tau \times (p + q))$ и $(p + q) \times (p + q)$ соответственно с элементами $C = (C_{ij})$, $C_{ij} = c_{j-i}$; $S = (S_{ij})$, $S_{ij} = \sum_{k=0}^{\infty} c_k c_{k+|i-j|}$. Заметим, что S — ковариационная матрица СВ $\{Y_1, \dots, Y_{p+q}\}$, где $\{Y_t\}$ — процесс АР $(p + q)$ с полиномом АР, равным $a(z)b(z)$, и с дисперсией возбуждающего белого шума, равной единице. Пусть Q — $\tau \times \tau$ -матрица $Q = CS^{-1}C^* = Q_{ij}$. Вектор $N^{-1/2}\hat{\rho}_W$ распределен при $N \rightarrow \infty$ асимптотически нормально с нулевым средним и матрицей ковариации $(I - Q)$, так что $\hat{\rho}_W(i)$ имеет дисперсию $\frac{1}{N}(1 - Q_{ii})$.

Исследование вектора $\hat{\rho}_W$ не совсем удобно из-за его многомерности, в то время как при принятии решений желательно работать со скалярной статистикой, которую определим выражением

$$\psi_W = N\hat{\rho}_W^* \hat{\rho}_W = N \sum_{t=1}^{\tau} \hat{\rho}_W^2(t). \quad (4.48)$$

С учетом того, что $\hat{\rho}_W$ асимптотически нормален с указанной корреляционной матрицей, статистика ψ распределена приближенно по χ^2 с $(\tau - p - q)$ степенями свободы. Поэтому при проверке адекватности модели АРСС (p, q) гипотеза адекватности отвергается, если

$$\psi_W > \chi_{1-\alpha}^2(\tau - p - q), \quad (4.49)$$

где $\chi_{1-\alpha}^2(n)$ — квантиль χ^2 распределения с n степенями свободы на уровне $1 - \alpha$; α — уровень значимости теста на адекватность.

Более точные решения при отклонении неадекватных моделей получаются с помощью модифицированного теста, основанного не на корреляции остатков, а на корреляции квадратов остатков. К этому тесту приходится прибегать, потому что встречаются случаи, когда остатки некоррелированы, тогда как квадраты остатков коррелированы. Определим

$$\hat{\rho}_{WW}(\tau) = \left(\sum_{t=1}^{N-\tau} (\hat{W}_t^2 - \bar{W}^2)(\hat{W}_{t+\tau}^2 - \bar{W}^2) \right) / \sum_{t=1}^N (\hat{W}_t^2 - \bar{W}^2)^2,$$

$$\tau \geq 1, \quad (4.50)$$

где $\hat{\rho}_{WW}(\tau)$ — ВКФ квадратов остатков; $\bar{W}^2 = \frac{1}{N} \sum_{k=1}^N \hat{W}_k^2$. Можно показать, что

$$\psi_{WW} = N(N+2) \sum_{t=1}^{\tau} \hat{\rho}_{WW}^2(t) / (N-1) \quad (4.51)$$

имеет приближенно $\chi^2(\tau)$ -распределение при адекватной модели. Гипотеза адекватности модели отвергается на уровне значимости α , если

$$\psi_{WW} > \chi_{1-\alpha}^2(\tau). \quad (4.52)$$

Практика применения тестов, основанных на ВКФ остатков или их квадратов, показала, что эти тесты работают лучше, когда решается задача отклонения неадекватной модели.

Для решения проблемы “случайности” последовательности остатков (т.е. проблемы подтверждения того, что последовательность является последовательностью НОР СВ) можно использовать и классические тесты.

Анализ экстремальных точек. Рассмотрим выборку $\{y_t, 1 \leq t \leq N\}$. Точку T назовем *экстремальной*, если $y_{t-1} < y_t, y_t > y_{t+1}$ или $y_{t-1} > y_t, y_t < y_{t+1}$, $1 \leq t \leq N$. Пусть τ — число экстремальных точек в выборке. Если наблюдаемая последовательность выбрана из последовательности НОР СВ, то вероятность того, что точка t будет экстремальной, равна $2/3$, $1 \leq t \leq N$. Поэтому среднее значение и дисперсия числа τ вычисляются соответственно по формулам

$$\mu_{\tau} = \mathbf{M}\{\tau\} = 2(N-3)/3; \quad D_{\tau} = \mathbf{M}\{(\tau - \mu_{\tau})^2\} = (16N-29)/90.$$

Большое отклонение $|\tau - \mu_{\tau}|$ от нуля свидетельствует о том, что выборка имеет свойства, отличающиеся от свойств последовательности НОР СВ. Случайная величина τ асимптотически нормальна. На этом основании можно построить тест: если $(|\tau - \mu_{\tau}| / \sqrt{D_{\tau}}) > \Phi_{1-\alpha/2}$, то гипотеза о том, что последовательность $\{y_1, \dots, y_N\}$ является последовательностью НОР СВ, отвергается на уровне значимости α . Здесь $\Phi_{1-\alpha/2}$ — $(1 - \alpha/2)$ -квантиль стандартного нормального распределения.

Анализ знаков разностей. При использовании этого теста определяется количество t , таких, что $y_t > y_{t-1}$, $2 \leq t \leq N$. Обозначим число данных t через s . Если $\{y_t\}$ является выборкой из последовательности НОР СВ, то среднее значение и дисперсия s вычисляются по формулам $\mu_s = \mathbf{M}\{s\} = (N-1)/2$, $D_s = \mathbf{M}\{(s - \mu_s)^2\} = (N+1)/12$. Так же, как в предыдущем случае, s — асимптотически нормально распределено. Сильное отклонение $|s - \mu_s|$ от нуля говорит о том, что выборка $\{y_t\}$ скорее всего не соответствует НОР СВ, а имеет тренд. На этом основании гипотеза о том, что $\{y_t\}$ выбирается из последовательности НОР СВ, на уровне значимости α должна отвергаться, если $(|s - \mu_s| / \sqrt{D_s}) > \Phi_{1-\alpha/2}$.

Ранговый тест. Ранговый тест полезен на практике для обнаружения линейного тренда в выборке. Определим P как число пар (i, j) , таких, что $y_j > y_i$,

$j > i, 1 \leq i \leq N - 1$. Полное число пар (i, j) , таких, что $j > i$, равно $N(N - 1)/2$. И для каждой пары НОР СВ y_i, y_j событие $\{y_j > y_i\}$ имеет вероятность $1/2$. Среднее значение и дисперсия P определяются формулами

$$\mu_p = \mathbf{M}\{P\} = N(N - 1)/4; \quad D_p = \mathbf{M}\{(P - \mu_p)^2\} = N(N - 1)(2N + 5)/8.$$

Величина P , как и предыдущие, асимптотически нормальна, поэтому для достаточно больших N гипотеза о выборке из последовательности НОР СВ отвергается на уровне значимости α , если $(|P - \mu_p|/\sqrt{D_p}) > \Phi_{1-\alpha/2}$.

Лабораторная работа 15. Определение параметров моделей АРСС (p, q) на основе предсказания

Цель работы. Исследовать критерии FRE, AIC и BIC, определить порядки p, q моделей АРСС (p, q) с использованием оценок предсказания.

Порядок выполнения работы

Задание 1. Произвести моделирование некоторого процесса АРСС (p, q) . Используя уравнение (4.41), убедиться, что для всякого N всегда найдется набор коэффициентов $\{a_i, 1 \leq i \leq N - 1\}$, который дает аппроксимацию исследуемого процесса как процесса типа АР $(N - 1)$. Убедиться также, что предсказание значения процесса y_{N+1} производится с ошибкой. Исследовать значение этой ошибки в зависимости от N . Обратить внимание, как изменяются параметры $\{a_i\}$ с изменением N .

Задание 2. Произвести моделирование некоторого процесса АР (p_0) . По выборке наблюдений $\{y_t, 1 \leq t \leq N\}$ оценить параметры $\{a_i\}$, полагая параметр p различным $(2 \leq p \leq p_0 + 2)$. Для каждого из значений p построить значения статистик FRE, AIC и BIC по формулам (4.43)–(4.45). В двух последних случаях исследуемый процесс рассматривать как АРСС $(p, 0)$. Оценить дисперсию белого шума по формуле (4.42). Убедиться, что оценка дисперсии монотонно уменьшается с ростом p . При помощи критериев FRE, AIC, BIC принять решение о порядке p модели исследуемого процесса.

Задание 3. Произвести моделирование некоторого процесса АРСС (p_0, q_0) . По выборке наблюдений $\{y_t, 1 \leq t \leq N\}$, придав различные значения парам (p, q) , построить оценки параметров $\{a_i\}, \{b_j\}$ исследуемого процесса. Область возможных значений пар (p, q) желательно определить как множество $\{(p, q) | 0 \leq p \leq p_0 + 2, 0 \leq q \leq q_0 + 2\}$. Для каждой пары (p, q) вычислить значения статистик AIC и BIC по формулам (4.44), (4.45). Сравнить правильность решений относительно порядков p, q , основанных на рассмотренных критериях.

Лабораторная работа 16. Исследование остаточных разностей

в процессе построения моделей временного ряда

Цель работы. Построить различные критерии “белозумности” остаточных разностей, получаемых в процессе идентификации временного ряда. Произвести их сравнительный анализ.

Порядок выполнения работы

Продолжить исследования временного ряда, начатые при выполнении задания 3 лабораторной работы 15. Исходным материалом здесь служат выборка $\{y_t, 1 \leq t \leq N\}$, полученная при выполнении указанного задания, и решение о модели АРСС (\hat{p}, \hat{q}) , которое было там принято. На основании этих данных строится выборка нормированных остаточных разностей с помощью формулы (4.46).

Задание 1. Построить график $\hat{W}_t, 1 \leq t \leq N$. Проанализировать его на предмет наличия систематического тренда, периодической составляющей и непостоянства дисперсий.

Задание 2. Построить выборочную КФ $\hat{\rho}_W$ процесса остаточных разностей \hat{W}_t по формуле (4.47). Основываясь на модели АРСС (p, q) , найти параметры асимптотического распределения вероятностей значений $\hat{\rho}_W(\tau)$. Построить доверительные интервалы для этих значений на некотором уровне значимости α . Результат представить в виде графика функции $\hat{\rho}_W(\tau)$ и границ доверительных интервалов и проанализировать его. Использовать критерии (4.48), (4.49) для решения вопроса об адекватности модели АРСС (p, q) . Выполнить это задание также для истинной модели АРСС (p_0, q_0) . Сравнить результаты.

Задание 3. Построить выборочную КФ $\hat{\rho}_{WW}(\tau)$ квадратов остаточных разностей по формуле (4.50). Представить ее на графике. Сравнить с графиком для $\hat{\rho}_W(\tau)$. Использовать критерии (4.51), (4.52) для решения вопроса об адекватности модели АРСС (p, q) .

Задание 4. Исследовать выборку остатков $\{\hat{W}_t\}$ по критерию экстремальных точек.

Задание 5. Исследовать выборку остатков $\{\hat{W}_t\}$ по критерию знаков разностей.

Задание 6. Использовать ранговый тест для исследования качества случайной выборочной последовательности остаточных разностей $\{\hat{W}_t\}$.

4.4. Прогнозирование нестационарных случайных процессов, построенных на основе АРСС

В предыдущих главах и параграфах подробно описаны методы анализа класса ССП, объединенных моделями АРСС (p, q) . Эти методы оказались настолько привлекательными, что появилось стремление распространить их на какие-либо нестационарные процессы. Оказалось, что это можно сделать для процессов с

полиномиальным трендом. Процессы АРСС (p, q) определяются соотношением

$$y_t = \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^q b_j W_{t-j}, \quad (4.53)$$

где W_j — процесс белого шума с дисперсией σ^2 . Пусть $a(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p$, $b(z) = b_0 + b_1 z + \dots + b_q z^q$. Если полиномы $a(z)$ и $b(z)$ не имеют общих корней, а все корни полинома $a(z)$ лежат за пределами единичного круга, процесс называется *каузальным*.

Пусть d — положительное целое число и

$$\tilde{a}(z) = a(z)(1-z)^d = - \sum_{i=0}^{p+d} \tilde{a}_i z^i, \quad \tilde{a}_0 = -1. \quad (4.54)$$

Процесс x_t , задаваемый соотношением

$$x_t = \sum_{i=0}^{p+d} \tilde{a}_i x_{t-i} + \sum_{j=0}^q b_j W_{t-j}, \quad (4.55)$$

называется *процессом типа проинтегрированной авторегрессии — скользящего среднего* (ПАРСС) (p, d, q) , если в (4.55) коэффициенты $\{\tilde{a}_i\}$ определяются соотношением (4.54), в котором полином $a(z)$ не имеет корней внутри или на границе единичного круга. Таким образом, полином $\tilde{a}(z)$ имеет корень $z = 1$ кратности d , а остальные его корни лежат вне единичного круга. Другими словами, введем разности $\Delta x_t = x_t - x_{t-1}$, $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1}$, \dots , $\Delta^d x_t = \Delta^{d-1} x_t - \Delta^{d-1} x_{t-1}$, назвав $\Delta^d x_t$ d -кратной разностью x_t . Тогда x_t является процессом ПАРСС (p, d, q) , если его кратные разности образуют каузальный процесс АРСС (p, q) . Можно представить d -кратные разности x_t в виде

$$\Delta^d x_t = \sum_{k=0}^d (-1)^k \binom{d}{k} x_{t-k} = y_t.$$

Так что если y_t , определяемый этим разностным соотношением, является каузальным процессом АРСС (p, q) , то процесс x_t относится к классу ПАРСС (p, d, q) .

Процессы ПАРСС не обязательно стационарные. Более того, типичный представитель процесса ПАРСС (p, d, q) — случайный процесс с полиномиальным трендом. Действительно, пусть

$$x_t = x_t^{(1)} + x_t^{(2)}, \quad x_t^{(1)} = \sum_{k=1}^d c_k t^{k-1}. \quad (4.56)$$

Определим $\Delta^d x_t^{(1)}$. Имеем $\Delta x_t^{(1)} = 0$, если $d = 1$ и для $d > 1$ выполняются равенства

$$\Delta x_t^{(1)} = x_t^{(1)} - x_{t-1}^{(1)} = \sum_{k=0}^{d-1} c_{k+1} (t^k - (t-1)^k) =$$

$$= \sum_{l=0}^{d-2} t^l \left(\sum_{k=l+1}^{d-1} c_{k+1} \binom{k}{l} (-1)^{k-l} \right).$$

Здесь скобка $\binom{k}{l}$ означает биномиальный коэффициент. Обозначив

$$c_l^{(1)} = \sum_{k=l+1}^{d-1} (-1)^{k-l} c_{k+1} \binom{k}{l},$$

запишем разность

$$\Delta x_t^{(1)} = x_t^{(1)} - x_{t-1}^{(1)} = \sum_{l=0}^{d-2} c_l^{(1)} t^l, \quad d \geq 2. \quad (4.57)$$

Вообще, если $c_l^{(i+1)} = \sum_{k=l+1}^{d-i-1} (-1)^{k-l} c_{k+1}^{(i)} \binom{k}{l}$, то

$$\Delta^{(i)} x_t^{(1)} = \sum_{l=0}^{d-i-1} c_l^{(i)} t^l, \quad 1 \leq i \leq d-1. \quad (4.58)$$

Таким образом, для любого α и любого набора коэффициентов $\{c_k\}$ $\Delta^d x_t^{(1)} = 0$. Отсюда следует, что если в (4.56) компонента $x_t^{(1)}$ является полиномом степени $(d-1)$, то $\Delta^d x_t^{(1)} = 0$ и d -кратная разность процесса x_t не зависит от компоненты $x_t^{(1)}$. Если $\Delta^d x_t^{(2)} = \Delta^d x_t$ — случайный процесс типа АРСС (p, q) , то (4.58) — нестационарный случайный процесс ПАРСС (p, d, q) , обладающий полиномиальным трендом.

Обратимся теперь к проблеме предсказания значений процесса x_t . Пусть $\{y_t\}$ — процесс АРСС (p, q) , определяемый (4.53). Тогда процесс x_t , являющийся процессом ПАРСС (p, d, q) , можно представить в виде разностного уравнения

$$x_t = y_t - \sum_{k=1}^d \binom{d}{k} (-1)^k x_{t-k}, \quad t = 1, 2, \dots, \quad (4.59)$$

так как $\Delta^{(i)} x_t = \sum_{k=0}^i \binom{i}{k} (-1)^k x_{t-k}$. Из (4.59) видно, что значения процесса y_t можно однозначно определить по наблюдениям $\{x_{t-k}, 0 \leq k \leq d\}$. Прогнозирование процессов АРСС (p, q) подробно рассмотрено в §4.2. Используя этот факт, будем строить процедуру прогнозирования процесса x_t с целью определения оценки предсказания значения $x_{N+\tau}$, основывающейся на наблюдениях $\{x_{1-d}, x_{2-d}, \dots, x_N\}$. Для удобства здесь устанавливаем объем выборки $N+d$. Пусть $\hat{x}_{t+\tau}^{(\tau)}, \hat{y}_{t+\tau}^{(\tau)}$ — наилучшие в среднеквадратичном смысле линейные оценки предсказания значений $x_{t+\tau}$ и $y_{t+\tau}$ соответственно по выборке

$\{x_{1-d}, x_{2-d}, \dots, x_t\}$. Основываясь на (4.59) и свойствах оптимальных линейных оценок, получаем

$$\hat{x}_{N+\tau}^{(\tau)} = \hat{y}_{N+\tau}^{(\tau)} - \sum_{k=1}^d \binom{d}{k} (-1)^k \hat{x}_{t-k+\tau}^{(\tau-k)}; \quad (4.60)$$

$$x_{t+1} - \hat{x}_{t+1}^{(1)} = y_{t+1} - \hat{y}_{t+1}^{(1)}, \quad t \geq 0. \quad (4.61)$$

Естественно, в (4.8) предполагается, что $\hat{x}_{t+\tau}^{(\tau)} = x_{t+\tau}$, если $\tau \leq 0$. Кроме того, значения y_t , $t > 0$ не коррелированы с начальными значениями процесса x_t , т.е. с вектором $(x_{1-d}, x_{2-d}, \dots, x_0)$.

Используя выражение (4.37) для оценки предсказания $\hat{y}_{N+\tau}^{(\tau)}$ и учитывая (4.61), получаем

$$\hat{y}_{N+\tau}^{(\tau)} = \sum_{i=1}^p a_i \hat{y}_{N+\tau-i}^{(\tau-i)} + \sum_{j=\tau}^q \theta_{N+\tau-j,j} (x_{N+\tau-j} - \hat{x}_{N+\tau-j}^{(1)}), \quad (4.62)$$

где второе слагаемое обращается в нуль для $\tau > q$. Наконец, по аналогии с (4.60) имеем

$$\hat{x}_{N+\tau}^{(\tau)} = \sum_{i=1}^{p+d} \tilde{a}_i \hat{x}_{N+\tau-i}^{(\tau-i)} + \sum_{j=\tau}^q \theta_{N+\tau-1,j} (x_{N+\tau-j} - \hat{x}_{N+\tau-j}^{(1)}). \quad (4.63)$$

Коэффициенты \tilde{a}_i определены в (4.54). Коэффициенты $\theta_{i,j}$ заданы формулами (4.25), (4.30).

Дисперсия построенной оценки предсказания процесса x_t на τ шагов вперед по выборке $\{x_{1-d}, x_{2-d}, \dots, x_N\}$, которая является наилучшей в среднеквадратичном смысле линейной оценкой, определяется соотношением

$$\begin{aligned} \tilde{D}_\tau(N) &= \mathbf{M}\{(x_{N+\tau} - \hat{x}_{N+\tau}^{(\tau)})^2\} = \\ &= \sum_{j=0}^{\tau-1} \left(\sum_{r=0}^j \chi_r \theta_{N+\tau-r-1,j-r} \right)^2 D_M(N + \tau - j), \end{aligned} \quad (4.64)$$

где $D_1(k)$ — дисперсия предсказания на один шаг по k наблюдениям для процесса y_t , которая вычисляется по формуле (4.27). Коэффициент $\theta_{k0} = 1$, а χ_r находятся из равенства

$$\chi(z) = \sum_{r=0}^{\infty} \chi_r z^r = \left(1 - \sum_{k=1}^{p+d} \tilde{a}_k z^k \right)^{-1}, \quad |z| < 1.$$

Они могут быть вычислены рекуррентно по формулам

$$\chi_0 = 1; \chi_r = \sum_{j=1}^{\min(p+d,r)} \tilde{a}_j \chi_{r-j}, \quad r = 1, 2, \dots, \tau.$$

Если процесс $\{y_t\}$ не только каузальный, но и обратимый, для достаточно больших N оценку (4.63) можно выразить в приближенном виде с учетом того, что при $N \rightarrow \infty$ коэффициенты $\theta_{N,j} \rightarrow b_j$, $1 \leq j \leq q$. Кроме того, $D_1(N) \rightarrow \sigma^2$. Поэтому дисперсия $D_\tau(N)$, определяемая по (4.64), будет иметь асимптотику

$$\tilde{D}_\tau = \lim_{N \rightarrow \infty} \tilde{D}_\tau(N) = \sigma^2 \sum_{j=0}^{\tau-1} \psi_j^2, \quad (4.65)$$

где коэффициенты ψ_j удовлетворяют равенствам

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = (\tilde{a}(z))^{-1} b(z), \quad |z| < 1,$$

а полиномы $\tilde{a}(z)$ и $b(z)$ определены в (4.53) и (4.54).

Рассмотрим процесс ПАРСС (1,2,1). В этом случае (4.55) превращается в уравнение

$$x_t = (2+a)x_{t-1} - (1+2a)x_{t-2} + ax_{t-3} + W_t + bW_{t-1}, \quad t \geq 1. \quad (4.66)$$

Для удобства здесь обозначены $a_1 = a$, $b_0 = 1$, $b_1 = b$, без потери общности. Начальные значения x_1, x_0 предполагаются некоррелированными с АРСС (1,1) процессом $y_t = x_t - 2x_{t-1} + x_{t-2}$. Из § 4.2 имеем

$$\begin{aligned} \hat{y}_{N+1}^{(1)} &= ay_N + \theta_{N1}(y_N - \hat{y}_N^{(1)}), \\ \hat{y}_{N+\tau}^{(\tau)} &= a\hat{y}_{N+\tau-1}^{(\tau-1)} = a^{\tau-1}\hat{y}_{N+1}^{(1)}, \quad |\tau| > 1. \end{aligned}$$

Поэтому

$$\begin{aligned} \hat{x}_{N+1}^{(1)} &= (2+a)x_N - (1+2a)x_{N-1} + ax_{N-3} + \theta_{N1}(x_N - \hat{x}_N^{(1)}), \\ \hat{x}_{N+\tau}^{(\tau)} &= (2+a)\hat{x}_{N+\tau-1}^{(\tau-1)} - (1+2a)\hat{x}_{N+\tau-2}^{(\tau-2)} + a\hat{x}_{N+\tau-3}^{(\tau-3)}, \quad \tau > 1. \end{aligned} \quad (4.67)$$

Зафиксируем N и рассмотрим последовательность $\{g_\tau, \tau = 1, 2, \dots\}$, где $g_\tau = \hat{x}_{N+\tau}^{(\tau)}$. Очевидно, что g_τ удовлетворяет разностному уравнению

$$g_\tau - (2+a)g_{\tau-1} + (1+2a)g_{\tau-2} - ag_{\tau-3} = 0 \quad (4.68)$$

с начальными условиями $g_{-1} = x_{N-1}$, $g_0 = x_N$, $g_1 = \hat{x}_{N+1}^{(1)}$. Решение разностного уравнения (4.68) получим в виде

$$g_\tau = h_0 + h_1\tau + h_2a^\tau,$$

в котором h_0, h_1, h_2 — не зависящие от τ величины, задаваемые начальными условиями. Таким образом, для всякого фиксированного N , последовательно

вычислив g_τ , можно определить необходимое количество предсказанных значений по выборке $\{x_1, \dots, x_N\}$.

В общем случае произвольного процесса ПАРСС (p, d, q) функция $g_\tau = \hat{x}_{N+\tau}^{(\tau)}$ удовлетворяет разностному уравнению порядка $(p + d)$

$$g_\tau - \sum_{k=1}^{p+d} \tilde{a}_k g_{\tau-k} = 0, \quad \tau > q, \quad (4.69)$$

с начальными условиями $g_\tau = \hat{x}_{N+\tau}^{(\tau)}$, $\tau = q, q-1, \dots, q+1-p-d$. Решение (4.69) может быть записано для $d \geq 1$ в виде

$$g_\tau = \sum_{i=0}^{d-1} G_i \tau^i + \sum_{j=1}^p H_j A_j^\tau, \quad \tau > q, \quad (4.70)$$

где A_j — обратные значения корней полинома $a(z)$, а не зависящие от τ величины $\{G_i, 0 \leq i \leq d-1\}$ и $\{H_j, 1 \leq j \leq p\}$ определяются из начальных условий. Величины g_τ являются значениями оценок предсказания процесса на τ шагов вперед по выборке $\{x_1, x_2, \dots, x_N\}$. Таким образом, оценки предсказания значений процесса x_t могут последовательно находиться по формуле (4.63), точность этих оценок определяется по формуле (4.64), а вычисления по формуле (4.70) позволяют найти оценки для различных сроков прогнозирования.

Как правило, полиномиальные тренды используются для описания таких процессов, математические ожидания которых имеют тенденцию к монотонному изменению. Вместе с тем при описании экономических процессов или процессов протекающих в природе, часто встречаются временные ряды, математические ожидания которых периодически изменяются с ростом параметра (сезонные компоненты временного ряда). Процессы такого типа также могут быть описаны с использованием техники исследования процессов АРСС.

Предположим, что некоторый процесс x_t обладает упомянутыми свойствами, которые формально можно интерпретировать следующим образом. Будем считать, что значения процесса x_t представлены в виде таблицы, содержащей столбцы по s строк, причем последовательными значениями процесса заполняется сначала первый столбец, затем второй и т.д.:

$$\begin{array}{cccccc} x_1 & x_{1+s} & x_{1+2s} & \dots & x_{1+ns} & \dots \\ x_2 & x_{2+s} & x_{2+2s} & \dots & x_{2+ns} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_s & x_{2s} & x_{3s} & \dots & x_{(1+n)s} & \dots \end{array}$$

Пусть значения процесса x_t , попадающие в одну строку, образуют некоторый процесс ПАРСС. Причем параметры этих процессов ПАРСС для различных строк одинаковы.

Построим аналитическую модель такого процесса. Определим по аналогии с рассмотренными d -кратными разностями процесса x_t “разности процесса x_t с

циклом s — s -циклические разности процесса x_t соотношением $\Delta_s x_t \equiv \Delta_s^1 x_t = x_t - x_{t-s}$. Двукратные s -циклические разности $\Delta_s^2 x_t = \Delta_s^1 x_t - \Delta_s^1 x_{t-s}$. И вообще, D -кратные s -циклические разности $\Delta_s^D x_t = \Delta_s^{D-1} x_t - \Delta_s^{D-1} x_{t-s}$. В явной форме D -кратная s -циклическая разность вычисляется по формуле

$$\Delta_s^D x_t = \sum_{l=0}^D (-1)^l \binom{D}{l} x_{t-sl} \quad (4.71)$$

с использованием $(D+1)$ значений процесса x_t , взятых через интервал (цикл) длиной s . Понятно, что если $s=1$, то циклические разности полностью совпадают с обычными разностями, а рассмотренная таблица значений процесса x_t состоит из одной строки.

Перейдем теперь к обычным разностям s -циклических разностей процесса x_t .

Первая разность D -кратных s -циклических разностей $\Delta^1(\Delta_s^D x_t) = \Delta_s^D x_t - \Delta_s^D x_{t-1}$. Вторая разность D -кратных s -циклических разностей $\Delta^2(\Delta_s^D x_t) = \Delta^1(\Delta_s^D x_t) - \Delta^1(\Delta_s^D x_{t-1})$. Вообще, d -кратная разность D -кратных s -циклических разностей

$$\Delta^d(\Delta_s^D x_t) = \Delta^{d-1}(\Delta_s^D x_t) - \Delta^{d-1}(\Delta_s^D x_{t-1}). \quad (4.72)$$

В явной аналитической форме эта разность выглядит следующим образом:

$$\Delta^d(\Delta_s^D x_t) = \sum_{k=0}^d \sum_{l=0}^D (-1)^{l+k} \binom{d}{k} \binom{D}{l} x_{t-k-sl}. \quad (4.73)$$

Пусть d и D — целые неотрицательные числа. Будем называть СП x_t *сезонным процессом* ПАРСС $(p, d, q) \times (P, D, Q)_s$ с периодом s , если d -кратные разности D -кратных s -циклических разностей этого процесса образуют каузальный процесс АРСС $(p + sP, q + Qs)$.

Формальное аналитическое описание такого процесса сводится к следующему. Обозначим через y_t процесс, образуемый d -кратными разностями D -кратных s -циклических разностей процесса x_t , т.е. из (4.73) имеем

$$y_t = \sum_{k=0}^d \sum_{l=0}^D (-1)^{l+k} \binom{d}{k} \binom{D}{l} x_{t-k-sl}, \quad t = 1, 2, \dots \quad (4.74)$$

В свою очередь, процесс y_t порождается разностным уравнением

$$y_t = \sum_{i=1}^{p+sP} \tilde{A}_i y_{t-i} + \sum_{j=1}^{q+sQ} \tilde{B}_j W_{t-j}, \quad t = 1, 2, \dots, \quad (4.75)$$

где $\{W_t\}$ — процесс белого шума с дисперсией σ^2 , а коэффициенты \tilde{A}_i и \tilde{B}_j находятся, исходя из следующих соображений. Определим полиномы

$$\begin{aligned}
a(z) &= 1 - \sum_{i=1}^p a_i z^i, & A(z) &= 1 - \sum_{i=1}^p A_i z^i, \\
b(z) &= 1 + \sum_{j=1}^q b_j z^j, & B(z) &= 1 + \sum_{j=1}^Q B_j z^j.
\end{aligned} \tag{4.76}$$

Коэффициенты \tilde{A}_i и \tilde{B}_j вычисляются из равенств

$$\begin{aligned}
\tilde{A}(z) &= a(z)A(z^s) = 1 - \sum_{i=1}^{p+Ps} \tilde{A}_i z^i; \\
\tilde{B}(z) &= b(z)B(z^s) = 1 + \sum_{j=1}^{q+Qs} \tilde{B}_j z^j.
\end{aligned} \tag{4.77}$$

Для каузальности процесса y_t , определяемого (4.75)–(4.77), необходимо и достаточно, чтобы $a(z) \neq 0$ и $A(z) \neq 0$ для всех $|z| \leq 1$. Таким образом, сезонный ПАРСС (или СПАРСС) — это такой процесс x_t , значения которого связаны соотношением (4.74) со значениями каузального процесса y_t . Для СПАРСС, встречающихся в практических задачах, D редко бывает больше единицы, а P и Q чаще всего меньше трех. Комбинация обычных и s -циклических разностей процесса x_t значительно усложняет структуру корреляционных зависимостей процессов СПАРСС, хотя анализ этих процессов может быть представлен как совокупность некоторых фрагментов, каждый из которых используется для анализа менее сложных, чем СПАРСС, процессов. В связи с этим не будем подробно описывать весь анализ, а только перечислим основные его этапы с краткими комментариями.

Первый этап — определение таких величин d и D , которые обеспечивают стационарность последовательности y_t , задаваемой формулой (4.74), по выборочным значениям наблюдаемого временного ряда. Параметр s обычно полагается известным, так как в приложениях СПАРСС описывают данные с годовым или дневным циклом.

Второй этап — определение P и Q , когда исследуются последовательности выборочных данных, выбираемых через сезонный цикл $\{x_{j+st}, t = 0, 1, \dots\}$. Пусть $\hat{\rho}_y(\tau)$ — нормированная выборочная корреляционная функция процесса

$$\hat{\rho}_y(\tau) = \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2}, \quad \bar{y} = \frac{1}{N} \sum_{t=1}^N y_t.$$

Тогда P и Q должны быть выбраны таким образом, чтобы $\hat{\rho}_y(ks)$, $k = 1, 2, \dots$, рассматриваемая как функция k , была совместима с корреляционной функцией процесса АРСС (P, Q).

Затем фиксируются параметры p и q из расчета, чтобы $\hat{\rho}_y(1), \hat{\rho}_y(2), \dots, \hat{\rho}_y(s-1)$ соответствовали корреляционной функции процесса АРСС (p, q) .

Окончательное решение относительно параметров P, Q, p и q принимается с использованием АИС и критериев “белозвучности” остатков, описанных в § 4.3. Следует заметить, что при определенном сочетании параметров P, Q, p, q и s отдельные коэффициенты \tilde{A}_i и \tilde{B}_j должны быть нулевыми. В этом случае последнее слагаемое в (4.44) надо заменить на $2m$, где m — число нулевых параметров модели (4.75).

Для фиксированных значений параметров p, d, q, P, D, Q, s соответствующие наборы коэффициентов $\{a_i\}, \{b_j\}, \{A_i\}, \{B_j\}$ определяются по выборке наблюдений с использованием методов, изложенных в гл.3.

Метод прогнозирования, описанный для процесса ПАРСС, может быть использован также при нахождении оценок предсказания значений процесса СПАРСС. Для построенной модели процесса СПАРСС сначала предсказываются значения процесса АРСС $(p + Ps, q + Qs)$, описываемого уравнением (4.75). Затем на основе (4.74) и (4.60) записывают соотношение

$$\hat{x}_{N+\tau}^{(\tau)} = \hat{y}_{N+\tau}^{(\tau)} - \sum_{k=0}^d \sum_{l=0}^D (-1)^{l+k} \binom{d}{k} \binom{D}{l} x_{t-k-sl+\tau}^{(\tau-k)}, \quad (4.78)$$

которое рекуррентно определяет оптимальные оценки предсказания значений процесса СПАРСС на τ шагов вперед. Дисперсия ошибки предсказания значений процесса СПАРСС на τ шагов вперед вычисляется по формуле, аналогичной формуле (4.64). Если процесс, описываемый формулой (4.75), является не только каузальным, но и обратимым, при достаточно больших N дисперсия ошибки предсказания процесса СПАРСС на τ шагов вперед может приближенно вычисляться по формуле

$$\tilde{D}_\tau(N) = \sigma^2 \sum_{j=0}^{\tau-1} \psi_j^2, \quad (4.79)$$

где σ^2 — дисперсия белого шума W_t в (4.75), а коэффициенты ψ_j определяются из равенства

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{b(z)B(z^s)}{a(z)A(z^s)(1-z)^d(1-z^s)^D}, \quad |z| < 1. \quad (4.80)$$

Лабораторная работа 17. Прогнозирование процессов, порождаемых моделью ПАРСС (p, d, q)

Цель работы. Ознакомиться с моделью ПАРСС и методами исследования процессов, порождаемых этой моделью. Построение прогноза значений процессов ПАРСС (p, d, q) .

Порядок выполнения работы

Выберем параметры $p, q, \{a_i, 1 \leq i \leq p\}, \{b_j, 0 \leq j \leq q\}$ таким образом, чтобы они определяли каузальный и обратимый процесс АРСС (p, q) . На основе этого выбора и выбора положительного числа d построим процесс ПАРСС (p, d, q) в соответствии с формулой (4.55). При моделировании данного процесса необходимо задать соответствующие начальные условия, которые находятся следующим образом. Пусть в качестве полиномиального тренда выбран полином $c(t) = c_0 + c_1t + c_2t^2 + \dots + c_{d-1}t^{d-1}$. В качестве начальных условий выбирается d значений x_t , определяемых равенствами $x_t = c(t), t = 1 - d, 2 - d, \dots, -1, 0$. Использование этих значений в совокупности с (4.55) позволяет произвести моделирование процесса ПАРСС (p, d, q) с трендом в виде полинома $c(t)$.

Задание 1. Получить выборку $\{x_t, 1 \leq t \leq N\}$ процесса ПАРСС (p, d, q) с полиномиальным трендом. По формулам (4.63) произвести рекуррентное прогнозирование исследуемого процесса. Определить дисперсию ошибок такого прогноза при помощи формулы (4.64). Рассмотреть асимптотику этой дисперсии в виде (4.65) и установить быстроту сходимости (4.64) к (4.65).

Использовать для построения прогноза функцию g_τ , определяемую из (4.69), (4.70). Сравнить этот подход, используя формулу (4.63).

Задание 2. Рассмотреть случай процесса ПАРСС $(1,2,1)$, определяемого (4.66). При моделировании этого процесса использовать требования к нахождению начальных условий, сформулированные выше. При помощи (4.67) и (4.68) осуществить прогноз изучаемого процесса. На основе асимптотической дисперсии ошибок предсказания исследовать качество предсказания в зависимости от значений параметров a и b модели ПАРСС $(1,2,1)$.

Лабораторная работа 18. Исследование процессов, порождаемых моделью СПАРСС $(p, d, q) \times (P, D, Q)_s$

Цель работы. Ознакомиться с классом процессов с наличием полиномиального тренда и циклической составляющей — сезонными моделями, построенными на основе АРСС (p, q) . Произвести построение модели, осуществить моделирование, по выборке наблюдений найти параметры модели и осуществить прогноз.

Порядок выполнения работы

Задать параметры p, d, q, P, D, Q, s модели СПАРСС, а также коэффициенты полиномов (4.76). Коэффициенты следует задавать таким образом, чтобы процессы АРСС (p, q) и АРСС (P, Q) были каузальными и обратимыми. Для того чтобы модель была не очень сложной, рекомендуется выбрать $D = 1, P < 3, Q < 3$, а один из параметров p, q положить равным нулю. Для выбранных значений параметров построить полиномы (4.77) и разностное уравнение (4.75). Задать соответствующее количество начальных условий.

Задание 1. Провести моделирование процесса для выбранной модели с таким расчетом, чтобы в выборке $\{x_t, 1 \leq t \leq N\}$ было не менее 10 сезонных циклов. Построить график реализации и визуально убедиться, что процесс имеет сезонную компоненту. Моделирование процесса x_t осуществить следующим образом. Вначале при помощи разностного уравнения (4.75) моделируется процесс $\{y_t, 1 \leq t \leq N\}$. Переписать (4.74) в виде

$$y_t = \sum_{k=0}^d (-1)^k \binom{d}{k} x_{t-k} + \sum_{k=0}^d \sum_{l=0}^D (-1)^{l+k} \binom{d}{k} \binom{D}{l} x_{t-k-sl}$$

и далее

$$x_t = y_t - \sum_{k=0}^d (-1)^k \binom{d}{k} x_{t-k} + \sum_{k=0}^d \sum_{l=0}^D (-1)^{l+k} \binom{d}{k} \binom{D}{l} x_{t-k-sl}. \quad (4.81)$$

Эта формула позволит по прошлым значениям $\{x_j, t-1 \geq j \geq t-d-Ds\}$ и значению y_t определить x_t . Отсюда видно, что для получения реализации процесса x_t необходимо иметь $(d+Ds)$ начальных значений.

Задание 2. Согласно этапам, описанным в § 4.4, оценить параметры наблюдаемого процесса. Построить прогноз для процесса y_t в соответствии с подходами, описанными в § 4.2. А затем при помощи (4.78) с учетом (4.81) построить прогноз процесса x_t . Определить дисперсию ошибки данного прогноза по формуле (4.64). Вычислить асимптотическое приближение этой дисперсии, выражаемое формулой (4.79). Рекуррентно вычисленный прогноз процесса x_t вместе с истинными значениями данного процесса представить на графике.

4.5. Калмановская фильтрация и прогнозирование

До сих пор, когда говорилось о тренде временного ряда, предполагалось, что он является некоторой заданной функцией m_t (или s_t , если речь шла о сезонной компоненте). Вместе с тем более реалистичная постановка задачи следующая: тренд (включая сезонную составляющую) возникает в темпе текущего времени синхронно с наблюдениями временного ряда и может сам изменяться случайным образом. Поэтому будем предполагать, что существует некоторое разностное уравнение, определяющее тренд, — уравнение тренда

$$\mathbf{m}_{t+1} = F_t \mathbf{m}_t + \mathbf{V}_{t+1}, \quad t = 0, 1, 2, \dots \quad (4.82)$$

Здесь в качестве тренда будем рассматривать вектор $\mathbf{m}_t = (m_{t1}, m_{t2}, \dots, m_{tn})^*$, поскольку аналитическая запись результатов для векторного тренда практически несколько не усложняется по сравнению со скалярным случаем, а охват математических моделей значительно более широкий. Соответствующие результаты для скалярного тренда получаются, если в формулах положить $n = 1$. В

(4.82) F_t — известная $(n \times n)$ -матрица, а $\mathbf{V}_t = (V_{t1}, \dots, V_{tn})^*$ — векторный процесс белого шума, т.е. последовательность случайных векторов со свойствами

$$\mathbf{M}\{\mathbf{V}_t\} = 0, \quad \mathbf{M}\{\mathbf{V}_t \mathbf{V}_j^*\} = 0 \quad \text{для всех } t \neq j, \quad \mathbf{M}\{\mathbf{V}_t \mathbf{V}_t^*\} = Q, \quad (4.83)$$

Q_t — $(n \times n)$ -матрица ковариаций компонент вектора \mathbf{V}_t .

Кроме уравнения тренда, имеется уравнение наблюдений многомерного временного ряда

$$\mathbf{x}_t = H_t \mathbf{m}_t + \mathbf{W}_t, \quad t = 1, 2, \dots, \quad (4.84)$$

где $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tm})^*$ — вектор выборочных значений — наблюдений временного ряда; H_t — известная $(m \times n)$ -матрица; $\mathbf{W}_t = (W_{t1}, W_{t2}, \dots, W_{tm})^*$ — вектор белого шума, для которого аналогично (4.83) имеем

$$\begin{aligned} \mathbf{M}\{\mathbf{W}_t\} &= 0, \quad \mathbf{M}\{\mathbf{W}_t \mathbf{W}_t^*\} = R_t, \\ \mathbf{M}\{\mathbf{W}_t \mathbf{W}_j^*\} &= 0 \quad \text{для всех } t \neq j, \end{aligned} \quad (4.85)$$

R_t — $(m \times m)$ -матрица ковариаций компонент вектора W_t .

Обычно также предполагается, что

$$\mathbf{M}\{\mathbf{m}_0 \mathbf{V}_t^*\} = 0, \quad \mathbf{M}\{\mathbf{W}_t \mathbf{V}_t^*\} = \mathbf{M}\{\mathbf{W}_t \mathbf{V}_j^*\} = 0 \quad \text{для любых } t \text{ и } j. \quad (4.86)$$

При таком описании данных представляют интерес две задачи.

Задача фильтрации: основываясь на наблюдениях $\{x_1, x_2, \dots, x_N\}$, построить оценку значения тренда \mathbf{m}_N .

Задача предсказания: основываясь на наблюдениях $\{x_1, x_2, \dots, x_N\}$, построить оценку значения тренда $\mathbf{m}_{N+\tau}$, $\tau \geq 1$.

При этом указанные оценки должны быть оптимальными в том смысле, чтобы обеспечивать минимальное значение дисперсии ошибки соответствующего оценивания.

Принципиальные вопросы построения таких оценок были разработаны в классических трудах А.Н.Колмогорова и Н.Винера. Однако Р.Калманом были предложены эквивалентные оценки в рекуррентной форме, удобные в вычислительном отношении. Поэтому здесь будут описаны именно эти оценки.

Оптимальная в среднеквадратичном смысле линейная оценка предсказания $\hat{\mathbf{m}}_{N+1}$ определяется следующими рекуррентными соотношениями:

$$\hat{\mathbf{m}}_1 = F_0 \mathbf{M}\{\mathbf{m}_0\}, \quad \hat{\mathbf{m}}_{N+1} = F_N \hat{\mathbf{m}}_N + K_N (\mathbf{x}_N - H_N \hat{\mathbf{m}}_N), \quad (4.87)$$

где коэффициент усиления Калмана K_N — $(n \times m)$ -матрица, вычисляемая по формуле

$$K_N = F_N \Sigma_N H_N^* (H_N \Sigma_N H_N^* + R_N)^{-1}. \quad (4.88)$$

Здесь Σ_N — $(n \times n)$ -матрица корреляции ошибок предсказания, в свою очередь, задаваемая рекуррентно по формулам

$$\Sigma_{N+1} = (F_N - K_N H_N) \Sigma_N (F_N - K_N H_N)^* + Q_{N+1} + K_N R_N K_N^*;$$

$$\Sigma_1 = F_0 \Sigma_0 F_0^* + Q_1, \quad \Sigma_0 = \mathbf{M}\{(\mathbf{m}_0 - \mathbf{M}\{\mathbf{m}_0\})(\mathbf{m}_0 - \mathbf{M}\{\mathbf{m}_0\})^*\}. \quad (4.89)$$

Оптимальная в среднеквадратичном смысле линейная оценка фильтрации $\tilde{\mathbf{m}}_N$ определяется следующими соотношениями:

$$\tilde{\mathbf{m}}_N = \hat{\mathbf{m}}_N + \tilde{K}_N(\mathbf{x}_N - H_N \hat{\mathbf{m}}_N), \quad N = 1, 2, \dots, \quad (4.90)$$

где $(n \times m)$ -матрица \tilde{K}_N находится по формуле

$$\tilde{K}_N = \Sigma_N H_N^* (H_N \Sigma_N H_N^* + R_N)^{-1}, \quad (4.91)$$

а корреляционная матрица ошибок фильтрации

$$\tilde{\Sigma}_N = \mathbf{M}\{(\mathbf{m}_N - \tilde{\mathbf{m}}_N)(\mathbf{m}_N - \tilde{\mathbf{m}}_N)^*\}$$

определяется по формуле

$$\tilde{\Sigma}_N = (I - \tilde{K}_N H_N) \Sigma_N, \quad (4.92)$$

где I — единичная матрица.

Сравнение формул (4.87)–(4.92) позволяет заключить, что оценки предсказания и фильтрации и корреляционные матрицы ошибок предсказания и фильтрации связаны простыми соотношениями:

$$\hat{\mathbf{m}}_{N+1} = F_N \tilde{\mathbf{m}}_N, \quad N = 1, 2, \dots; \quad (4.93)$$

$$\Sigma_{N+1} = Q_{N+1} + F_N \tilde{\Sigma}_N F_N^*. \quad (4.94)$$

Это позволяет наиболее рационально построить процесс последовательного вычисления оценок предсказания и фильтрации.

Э т а п 1. Определяется $\hat{\mathbf{m}}_t$ по (4.93) (для $t = 1$ по (4.87)).

Э т а п 2. Определяется $\tilde{\Sigma}_t$ по (4.94) (для $t = 1$ по (4.89)).

Э т а п 3. Определяется \tilde{K}_t по (4.91).

Э т а п 4. Определяется $\tilde{\mathbf{m}}_t$ по (4.90).

Э т а п 5. Определяется $\tilde{\Sigma}_t$ по (4.92).

После каждой итерации значение индекса t увеличивается на единицу и процедура повторяется. Заметим, что для определения корреляционных свойств (и, в частности, дисперсий) оценок нет необходимости использовать этапы 1 и 4, на которых подсчитываются сами оценки. Добавим, что оценки фильтрации и предсказания с ростом N модернизируются в зависимости от самих выборочных значений x_N . Причем при определении очередной оценки требуется использовать только одну последнюю инновацию.

Временной ряд, заданный уравнениями (4.92), (4.93), является более общим описанием процессов типа АРСС. Покажем это. Пусть y_t — процесс АРСС (p, q) , описываемый уравнениями (4.28). Представим его в более удобной форме

$$y_t = \sum_{j=1}^n (a_j y_{t-j} + b_{j-1} W_{t+1-j}), \quad (4.95)$$

где $n = \max\{p, 1 + q\}$; $a_i = 0, i > p$; $b_i = 0, i > q$.

Определим вектор $\mathbf{m}_t = (m_{t1}, \dots, m_{tn})$ с компонентами $m_{tj} = y_{t-i+1}$; F — $(n \times n)$ -матрица с элементами $T_{ij} = b_j \delta_{1j}$; \mathbf{v}_t — вектор с компонентами $v_{tj} = W_{t+1-j}, 1 \leq j \leq n$. Составим векторное равенство

$$\mathbf{m}_{t+1} = F\mathbf{m}_t + T\mathbf{v}_{t+1}. \quad (4.96)$$

В этом равенстве первая компонента полностью совпадает с уравнением (4.95), а остальные компоненты являются тождествами $y_{t-i+1} \equiv y_{t-i+1}, 2 \leq i \leq n$. Сравнивая (4.96) с (4.82), убеждаемся в том, что уравнение (4.92) в частном случае, когда $F_t = F, \mathbf{V}_t = T\mathbf{v}_t$, приобретает вид (4.96). Причем матрица $Q = \mathbf{M}\{\mathbf{V}_t \mathbf{V}_t^*\} = \sigma^2 \mathbf{T} \mathbf{T}^*$. Она не меняется с возрастанием t и имеет единственный ненулевой элемент $Q_{11} = \sigma^2 \sum_{j=0}^q b_j^2$.

Далее, если мы определим $(1 \times n)$ -матрицу $H_t = (1 \ 0 \ 0 \dots 0)$ и $W_t = 0$, то уравнение (4.84) вырождается в уравнение наблюдения временного ряда $x_t = m_t$ с нулевыми ошибками наблюдения.

Таким образом, уравнение процесса АРСС (p, q) является частным случаем уравнений (4.82)—(4.84). Поэтому полученные в § 4.1—4.4 уравнения оценки предсказания процесса авторегрессии могут быть выражены в форме (4.90)—(4.94).

Калмановские оценки фильтрации и прогнозирования довольно просты и удобны для организации вычислений. Однако их практическое использование показало, что может возникнуть неустойчивость при определении оценок и процесс оценивания может расходиться. Это возникает тогда, когда параметры уравнений (4.82)—(4.85) известны неточно. Такая ситуация возникает, когда предсказанию процесса предшествует этап оценивания неизвестных параметров модели временного ряда и в дальнейшем используются не точные значения параметров уравнений (4.82)—(4.85), а их выборочные оценки. В результате такой неточности оценка вычисляется неверно, эти ошибки накапливаются, и оценки могут отличаться очень сильно. Проблема становится особенно острой, когда член, соответствующий шуму, в уравнениях системы мал. Тогда малы ковариации ошибок и коэффициент усиления Калмана, а последующие наблюдения незначительно влияют на оценивание. При практических применениях расходимость проявляет себя через обновление. В результате оценка больше не оптимальная, матрица Σ_t не является мерой дисперсии ошибок оценивания, и ошибка $(\hat{m}_t - m_t)$ систематически увеличивается с ростом t .

Один из способов борьбы с расходимостью оценки состоит в модификации уравнений оценки таким образом, чтобы более свежие наблюдения имели большее влияние на оценки, а воздействие старых наблюдений постепенно уменьшалось. Этого можно достичь использованием весов α_i к наблюдениям y_{t-i} , таких, что α_i монотонно уменьшается с ростом i . Наиболее удобными являются веса степенного вида $\alpha^i, 0 < \alpha < 1$. В случае применения весов уравнения (4.91) и (4.92), определяющие коэффициент усиления и матрицу ковариации оценок

фильтрации, заменяются на следующие:

$$\tilde{K}_N = \Sigma_N H_N^* (H_N \Sigma_N H_N^* + \alpha R_N)^{-1}, \quad (4.97)$$

$$\tilde{\Sigma}_N = \alpha (I - \tilde{K}_N H_N) \Sigma_N, \quad (4.98)$$

Остальные уравнения остаются без изменения. Если в (4.97), (4.98) положить $\alpha = 1$, то эти формулы превращаются в (4.89), (4.91). Каких-либо конкретных указаний выбора α сформулировать не удастся, так как расходимость оценок зависит от соотношения параметров уравнений (4.82), (4.84), а также от тех неточностей, которые допущены при оценивании этих параметров.

Лабораторная работа 19. Калмановская фильтрация и прогнозирование

Цель работы. Освоить технику вычисления оценок фильтрации и оценок предсказания при помощи рекуррентных процедур Калмана. Изучить проблему устранения расходимости рекуррентных оценок Калмана.

Порядок выполнения работы

Задание 1. Выбрать какую-либо модель ПАРСС (p, d, q) , желательно уже исследованную ранее при выполнении лабораторной работы 17. Построить для нее разностные векторно-матричные уравнения (4.82)–(4.84). Задать ненулевой шум наблюдения в уравнении (4.84). При помощи моделирования в соответствии с этими уравнениями получить реализацию процесса x_t . Сравнить эту реализацию с реализацией процесса из лабораторной работы 17. Изобразить обе реализации на графике. Установить, насколько сильно искажает шум наблюдения исследуемый процесс. По формулам (4.87)–(4.94) построить оценки фильтрации и прогнозирования исследуемого процесса, имея в виду первую компоненту вектора m_t . Вычислить точность предсказания процесса и сравнить эту точность с точностью предсказания, полученной при выполнении лабораторной работы 17. Объяснить различие в этих результатах.

Задание 2. Выбрать произвольные матрицы F_t и H_t в уравнениях (4.82), (4.84). Задать положительно определенные матрицы R_t, Q_t . (Для простоты исследования матрицы можно выбирать не зависящими от t .) Получить реализацию процессов. Построить оценки предсказания и фильтрации. Затем исказить одну из матриц F, H, R, Q и с использованием этой матрицы построить снова оценки предсказания и фильтрации процесса. (При искажении матриц R и Q необходимо это делать так, чтобы искаженные матрицы оставались положительно определенными.) Установить расходимость оценок. Исследовать, к какой расходимости приводит искажение той или иной матрицы. Использовать формулы (4.97), (4.98) с целью устранения расходимости. Исследовать влияние величины параметра α в формулах (4.97), (4.98) на качество устранения расходимости и на качество оценивания и предсказания. Исследования иллюстрировать графиками.

СПЕКТРАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

5.1. Спектральные представления СП и оценивание спектральных плотностей

Для любого стационарного СП (ССП) $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$ можно построить процесс $\{y_t, t = 0, \pm 1, \pm 2, \dots\}$, определяемый представлением

$$y_t = m + \sum_{j=1}^n (A_j \cos \lambda_j t + B_j \sin \lambda_j t), \quad t = 0, \pm 1, \dots, \quad (5.1)$$

где $m = \mathbf{M}\{x_t\}$; $\{A_j\}, \{B_j\}$ — СВ со свойствами $\mathbf{M}\{A_j\} = \mathbf{M}\{B_j\} = 0$; $\mathbf{M}\{A_j^2\} = \mathbf{M}\{B_j^2\} = \sigma_j^2$; $\mathbf{M}\{A_j B_j\} = 0$ для любых $j, 1 \leq j \leq n$; λ_j принимают значения на отрезке $[0, \pi]$. АКФ процесса $\{y_t\}$ при помощи соответствующего выбора $\{\sigma_j^2, 1 \leq j \leq n\}$ может хорошо аппроксимировать АКФ процесса $\{x_t\}$. Выбрав n достаточно большим, можно получить сколь угодно хорошую аппроксимацию исходного процесса и его АКФ. Такая декомпозиция ССП в сумму гармонических составляющих с некоррелированными случайными коэффициентами соответствует спектральному представлению ССП, рассматривавшемуся в гл.1. Декомпозиция ССП влечет за собой и декомпозицию АКФ этого ССП. Анализ ССП при помощи их спектральных представлений часто называется анализом в “частотной области”, поскольку вещественный параметр λ в (5.1) принято называть частотой (радианной, угловой и т.д.). Этот анализ является эквивалентом анализу во “временной области”, основанному на АКФ, но иногда обеспечивает способ рассмотрения процесса, который для некоторых применений более понятен. Такие примеры имеются в физике, электротехнике, акустике, геофизике и других разделах техники, а также в медицине, экономике, биологии, психологии и численном анализе.

Учитывая свойства $\{y_t\}$, определенного в (5.1) по отношению к ССП $\{x_t\}$ с нулевым средним, представим $\{x_t\}$ в виде

$$x_t = \sum_{j=1}^n a_j e^{it\lambda_j}, \quad (5.2)$$

где $-\pi < \lambda_1 < \lambda_2 < \dots < \lambda_n \leq \pi$; $\{a_j\}$ — некоррелированные, вообще говоря, комплексные СВ, такие, что $\mathbf{M}\{a_j\} = 0$, $\mathbf{M}\{a_j \bar{a}_j\} = \sigma_j^2, 1 \leq j \leq n$. Здесь \bar{a} — комплексно-сопряженная СВ по отношению к a . В частном случае, когда $\{x_t\}$ — вещественный СП, $\lambda_j = -\lambda_{n+1-j}$ и $a_j = \bar{a}_{n+1-j}, 1 \leq j \leq n$ (интересно отметить, что, несмотря на последнее равенство, a_j и a_{n+1-j} — некоррелированные СВ):

$$\mathbf{M}\{x_t x_{t+\tau}\} = r(\tau) = \sum_{j=1}^n \sigma_j^2 e^{i\tau\lambda_j}. \quad (5.3)$$

Определим $F(\lambda) = \sum_{j:\lambda_j \leq \lambda} \sigma_j^2$ — непрерывную справа неубывающую функцию, такую, что $F(-\pi) = 0$, $F(\pi) = r(0)$. Записывая (5.3) в виде интеграла Римана—Стилтьеса, получаем

$$r(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} dF(\lambda). \quad (5.4)$$

Функция $F(\lambda)$ называется *спектральной функцией* (спектральной мерой, функцией спектрального распределения).

Более общим, чем представление (5.2), является спектральное представление ССП с нулевым средним

$$x_t = \int_{-\pi}^{\pi} e^{it\lambda} dz, \quad (5.5)$$

в котором $z(\lambda)$ — случайный процесс с ортогональными приращениями, обычно называемый *спектральным процессом*. Соответственно АКФ $r(\tau)$ процесса x_t выражается в виде (5.4). Если в (5.4) $F(\lambda)$ абсолютно непрерывна, т.е.

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu, \quad -\pi \leq \lambda \leq \pi, \quad (5.6)$$

то $f(\nu)$ называется *спектральной плотностью* процесса x_t . Если $r(\tau)$ — абсолютно суммируемая функция, т.е. $\sum_{\tau=-\infty}^{+\infty} |r(\tau)| < \infty$, то имеет место представление

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} e^{-i\tau\lambda} r(\lambda), \quad -\pi \leq \lambda \leq \pi. \quad (5.7)$$

Отсюда, в частности, следует, что для вещественных x_t $f(\lambda)$ является четной.

Пусть $\{x_t\}$ — процесс АРСС (p, q) (не обязательно каузальный и обратимый), определяемый разностным уравнением

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \sum_{j=0}^q b_j W_{t-j}, \quad (5.8)$$

где W_t — белый шум с дисперсией σ^2 . Если полиномы $a(z) = 1 - a_1 z - \dots - a_p z^p$, $b(z) = 1 + b_1 z + \dots + b_q z^q$ не имеют общих корней и $a(z)$ не имеет корней на окружности $|z| = 1$, то спектральная плотность процесса АРСС (p, q) выражается равенством

$$f(\lambda) = \frac{\sigma^2 |b(e^{-i\lambda})|^2}{2\pi |a(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi, \quad (5.9)$$

т.е. является рациональной спектральной плотностью.

Если $f(\lambda)$ — непрерывная спектральная плотность некоторого вещественного ССП и $\varepsilon > 0$, то, во-первых, существует обратимый СС (q)-процесс $x_t = W_t + b_1 W_{t-1} + \dots + b_q W_{t-q}$, такой, что $|f_x(\lambda) - f(\lambda)| < \varepsilon$ для всех $-\pi \leq \lambda \leq \pi$, где W_t — белый шум с дисперсией σ^2 , причем $\sigma^2 = (1 + b_1^2 + \dots + b_q^2)^{-1} \int_{-\pi}^{\pi} f(\lambda) d\lambda$, во-вторых, существует каузальный АР (p)-процесс $x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + W_t$, такой, что $|f_x(\lambda) - f(\lambda)| < \varepsilon$ для всех $-\pi \leq \lambda \leq \pi$, где W_t — белый шум с дисперсией σ^2 .

Рассмотрим произвольное множество наблюдений $\{x_1, x_2, \dots, x_N\}$, сделанных в моменты $t = 1, 2, \dots, N$. Предположив выборку $\{x_t, 1 \leq t \leq N\}$ периодической с периодом N , представим ее элементы в виде декомпозиции по гармоникам ω_j :

$$x_t = \frac{1}{\sqrt{N}} \sum_{-\pi < \omega_j \leq \pi} a_j e^{it\omega_j}, \quad t = 1, \dots, N, \quad (5.10)$$

где частоты $\omega_j = 2\pi j/N = j(2\pi/N)$, $\omega_j \in (-\pi, \pi]$, называются частотами Фурье ряда $\{x_1, x_2, \dots, x_N\}$. Вводя вектор-столбец $\mathbf{e}_j = \frac{1}{\sqrt{N}}(e^{i\omega_j}, e^{i2\omega_j}, \dots, e^{iN\omega_j})^*$, совокупность соотношений (5.10) можно записать в векторной форме $\mathbf{x} = (x_1, \dots, x_N)^*$:

$$\mathbf{x} = \sum_{j \in J_N} a_j \mathbf{e}_j, \quad (5.11)$$

где $J_N = \{j : -\pi < \omega_j = \frac{2\pi j}{N} \leq \pi\} = \{-\lfloor \frac{N-1}{2} \rfloor, \dots, \lfloor \frac{N}{2} \rfloor\}$; $[a]$ — целая часть числа a ; J_N содержит N целых чисел. Разложение (5.11) обладает следующими свойствами: векторы $\{\mathbf{e}_j\}$ образуют ортогональный базис, коэффициенты a_j имеют вид

$$a_j = \frac{1}{\sqrt{N}} \sum_{t=1}^N x_t e^{-it\omega_j}. \quad (5.12)$$

Таким образом, $\{a_j, j \in J_N\}$ — это дискретное преобразование Фурье (ДПФ) выборки $\{x_t, 1 \leq t \leq N\}$.

Функция $I(\omega)$, определенная на множестве частот $\{\omega_j\}$, значения которой определяются по формуле

$$I(\omega_j) = |a_j|^2 = \frac{1}{N} \left| \sum_{t=1}^N x_t e^{-it\omega_j} \right|^2, \quad -\pi < \omega_j = \frac{2\pi j}{N} \leq \pi, \quad (5.13)$$

называется *периодограммой*. Периодограмма представляет разложение $\|x\|^2$ на сумму компонент, связанных с частотами Фурье:

$$\|x\|^2 = \sum_{j \in J_N} I(\omega_j). \quad (5.14)$$

Если x — вещественный ССП, то для всякого $j \in J_N$ всегда $\omega_j \in [-\pi, \pi]$ и $-\omega_j \in [-\pi, \pi]$, поэтому $a_j = \bar{a}_j$ и $I(\omega_j) = I(-\omega_j)$. Тогда (5.10) приобретает вид

$$\mathbf{x} = a_0 \mathbf{e}_0 + \sum_{j=1}^{[(N-1)/2]} (a_j \mathbf{e}_j + \bar{a}_j \mathbf{e}_{-j}) + a_{N/2} \mathbf{e}_{N/2}, \quad (5.15)$$

где последнее слагаемое равно нулю, если N — нечетное. Представим комплексные коэффициенты a_j в виде $\alpha_j e^{i\beta_j}$, тогда (5.15) перепишем следующим образом:

$$\mathbf{x} = a_0 \mathbf{e}_0 + \sum_{j=1}^{[(N-1)/2]} \sqrt{2} \alpha_j (\mathbf{c}_j \cos \beta_j - \mathbf{s}_j \sin \beta_j) + a_{N/2} \mathbf{e}_{N/2}, \quad (5.16)$$

где $\mathbf{e}_0 = \frac{1}{\sqrt{N}}(1 \ 1 \dots 1)^*$; $\mathbf{c}_j = \sqrt{2/N}(\cos \omega_j, \cos 2\omega_j, \dots, \cos N\omega_j)^*$; $\mathbf{s}_j = \sqrt{2/N}(\sin \omega_j, \sin 2\omega_j, \dots, \sin N\omega_j)^*$.

Пусть $\hat{m} = \frac{1}{N} \sum_{t=1}^N x_t$ и $\hat{r}(\tau)$ — выборочная АКФ:

$$\hat{r}(\tau) = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (x_{t+|\tau|} - \hat{m})(x_t - \hat{m}), \quad 0 \leq |\tau| < N, \quad (5.17)$$

тогда $I(\omega_j)$, определенная формулой (5.13), может быть представлена в виде

$$I(\omega_j) = \sum_{|\tau| < N} \hat{r}(\tau) e^{-i\tau\omega_j}, \quad \omega_j = \frac{2\pi j}{N}. \quad (5.18)$$

Сравнивая (5.18) с (5.7), приходим к выводу, что периодограмма $I(\omega_j)$ может быть использована для оценивания спектральной плотности $f(\lambda)$. В связи с этим отметим некоторые свойства периодограммы. Пусть $\{x_t\}$ — ССП со средним значением m и абсолютно суммируемой АКФ $r(\tau)$. Его спектральная плотность $f(\lambda)$ является непрерывной функцией и задается формулой (5.7). Назовем расширенной периодограммой $I_N(\lambda)$ функцию

$$I_N(\lambda) = \begin{cases} I(\omega_j), & \omega_j - \frac{\pi}{N} < \lambda \leq \omega_j + \frac{\pi}{N}, \quad \lambda \in [0, \pi], \\ I_N(-\lambda), & \lambda \in [-\pi, 0]. \end{cases} \quad (5.19)$$

При $N \rightarrow \infty$ будем иметь

$$\mathbf{M}\{I_N(0) - Nm^2\} \rightarrow 2\pi f(0); \quad \mathbf{M}\{I_N(\lambda)\} \rightarrow 2\pi f(\lambda), \quad \lambda \neq 0. \quad (5.20)$$

Отсюда следует, что $\frac{1}{2\pi} I_N(\lambda)$ для $\lambda \neq 0$ является асимптотически несмещенной оценкой спектральной плотности $f(\lambda)$, а $\frac{1}{2\pi}(I_N(0) - Nm^2)$ — асимптотически несмещенной оценкой $f(0)$.

Как уже отмечалось, многие случайные процессы, порождающие наблюдаемые временные ряды, могут быть представлены в виде

$$x_t = \sum_{j=-\infty}^{+\infty} \psi_j W_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots,$$

где W_t — белый шум с дисперсией σ^2 , а $\sum_{-\infty}^{+\infty} |\psi_j| < \infty$. Тогда если $f(\lambda) > 0$ для всех $\lambda \in [-\pi, \pi]$ и если $0 < \lambda_1 < \lambda_2 < \dots < \lambda_n < \pi$, то случайный вектор, составленный из значений периодограммы (5.19) $(I_N(\lambda_1), \dots, I_N(\lambda_n))^*$, сходится по распределению при $N \rightarrow \infty$ к вектору с независимыми и экспоненциально распределенными компонентами со средним значением $(2\pi f(\lambda_1), 2\pi f(\lambda_2), \dots, 2\pi f(\lambda_n))^*$.

При этом если $\sum_{j=-\infty}^{+\infty} \sqrt{|j|} |\psi_j| < \infty$, $\mathbf{M}\{W_t^4\} < \infty$, то

$$\text{cov}(I_N(\omega_j), I_N(\omega_k)) = \begin{cases} 2(2\pi)^2 f^2(\omega_j) + 0(\frac{1}{\sqrt{N}}), & \omega_j = \omega_k = 0 \text{ или } \pi, \\ (2\pi)^2 f^2(\omega_j) + 0(\frac{1}{\sqrt{N}}), & 0 < \omega_j = \omega_k < \pi, \\ 0(\frac{1}{\sqrt{N}}), & \omega_j \neq \omega_k, \end{cases} \quad (5.21)$$

где $\omega_j = 2\pi j/N$ — частоты Фурье для выборки $\{x_1, x_2, \dots, x_N\}$.

Отсюда следует, что хотя периодограмма является асимптотически несмещенной оценкой $2\pi f(\lambda)$, однако она не является состоятельной оценкой этой величины. Вместе с тем значения $I_N(\lambda)$ для различных λ , принимающих значения из множества частот Фурье, являются асимптотически некоррелированными. Поэтому, рассматривая компоненты вектора $(I_N(\omega_1), \dots, I_N(\omega_{[N/2]}))^*$ как функции ω_j , можно убедиться, что они образуют сильно флуктуирующий процесс со средним значением, аппроксимирующим $2\pi f(\lambda)$. Представляется возможность использовать сглаживающий фильтр для этого процесса, причем его характеристики выбираются таким образом, чтобы образуемая им функция была не только асимптотически несмещенной, но и состоятельной оценкой спектральной плотности. Иначе говоря, введем оценку спектральной плотности (ОСП):

$$\hat{f}_g(\omega_j) = \frac{1}{2\pi} \sum_{|k| \leq n} g_N(k) I_N(\omega_{j+k}), \quad (5.22)$$

где n — некоторое целое число; $g_N(k)$ — последовательность некоторых весовых коэффициентов. В условиях справедливости (5.21) относительно свойств (5.22) можно записать:

$$\lim_{N \rightarrow \infty} \frac{\text{cov}(\hat{f}_g(\lambda), \hat{f}_g(\nu))}{\sum_{|k| \leq n} g_N^2(k)} = \begin{cases} 2f^2(\lambda), & \lambda = \nu = 0 \text{ или } \pi, \\ f^2(\lambda), & 0 < \lambda = \nu < \pi, \\ 0, & \lambda \neq \nu. \end{cases} \quad (5.23)$$

Из (5.22) и (5.23) видно, что для асимптотической несмещенности и состоятельности необходимо и достаточно выполнение следующих условий, влияющих на характер изменения весовых коэффициентов. Для асимптотической несмещенности

$$\sum_{|k| \leq n} g_N(k) = 1. \quad (5.24)$$

Для состоятельности

$$\lim_{N \rightarrow \infty} \sum_{|k| \leq n} g_N^2(k) = 0. \quad (5.25)$$

Будем полагать также $g_N(k) = g_N(-k)$. В специальной литературе весовые коэффициенты $g_N(k)$ обычно называются *сглаживающим фильтром*.

Часто используется иная оценка спектральной плотности, которая построена на модификации представления периодограммы в виде (5.18):

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|\tau| \leq n} h\left(\frac{\tau}{n}\right) \hat{r}(\tau) e^{-i\tau\lambda}, \quad (5.26)$$

где $h(t)$ — четная кусочно-непрерывная функция, удовлетворяющая условиям $h(0) = 1$, $|h(t)| \leq 1$ для всех t , $h(t) = 0$ для $|t| > 1$; $\hat{r}(\tau)$ — выборочная АКФ.

Задав расширенную периодограмму соотношением

$$I_N(\lambda) = \sum_{|\tau| \leq N} \hat{r}(\tau) e^{-i\tau\lambda}, \quad (5.27)$$

совпадающим с (5.19) на множестве частот Фурье (см. также (5.18)), получим представление для выборочной АКФ

$$\hat{r}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\tau\lambda} I_N(\lambda) d\lambda, \quad (5.28)$$

при подстановке которого в (5.26) найдем еще одно представление оценки спектральной плотности

$$\hat{f}(\lambda) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\nu) I_N(\nu + \lambda) d\nu, \quad (5.29)$$

где функция $H(\lambda)$ определяется соотношением

$$H(\lambda) = \frac{1}{2\pi} \sum_{|\tau| \leq n} h\left(\frac{\tau}{n}\right) e^{-i\tau\lambda} \quad (5.30)$$

и называется *спектральным окном* (частотным окном). Соответственно этому $h(t)$, удовлетворяющая (5.26), называется *корреляционным окном* (временным окном).

Разбив интервал $[-\pi, \pi]$ частотами Фурье $\omega_j = \frac{2\pi j}{N}$ и заменив интеграл (5.29) интегральной суммой, получим еще одну ОСП

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{j=1-[N/2]}^{[N/2]} H(\omega_j) I_N(\lambda + \omega_j) \frac{2\pi}{N}. \quad (5.31)$$

Формулу (5.31) можно рассматривать как приближение оценки (5.26) при помощи взвешенного усреднения периодограммы $I_N(\omega)$ с весами $H_j = 2\pi H(\omega_j)/N$, $|j| \leq [N/2]$. Следует заметить, что весовые коэффициенты H_j не обязательно должны удовлетворять условиям типа (5.24), (5.25), которые требуются при построении ОСП (5.22) при помощи взвешенного усреднения периодограммы $I_N(\omega)$ с весами $g_N(k)$.

В условиях справедливости (5.21) $\lim_{N \rightarrow \infty} \mathbf{M}\{\hat{f}(\lambda)\} = f(\lambda)$, $0 \leq \lambda \leq \pi$, и выполняется предельное соотношение

$$\lim_{N \rightarrow \infty} \frac{N}{n} D\{\hat{f}(\lambda)\} = \begin{cases} 2f^2(\lambda) \int_{-1}^{+1} h^2(t) dt, & \lambda = 0 \text{ или } \pi, \\ f^2(\lambda) \int_{-1}^{+1} h^2(t) dt, & 0 < \lambda < \pi. \end{cases}$$

Отсюда следует состоятельность и асимптотическая несмещенность оценки (5.26).

В (5.26) корреляционное окно $h(t)$ задается неоднозначно. Поэтому существует достаточно много вариантов определения этих окон. Приведем наиболее известные варианты вместе с сопутствующими им спектральными окнами $H(\lambda)$.

Окно Дирихле (прямоугольное, усеченное окно):

$$h(t) = 1, \quad 0 \leq |t| \leq 1; \\ H(\lambda) = D_n(\lambda) = \frac{\sin(n+1/2)\lambda}{2\pi \sin \frac{\lambda}{2}}, \quad -\pi \leq \lambda \leq \pi, \quad (5.32)$$

$D_n(\lambda)$ — ядро Дирихле. В этом случае $D\{\hat{f}(\lambda)\} \cong \frac{2nf^2(\lambda)}{N}$.

Окно Бартлетта (треугольное окно):

$$h(t) = 1 - |t|, \quad 0 \leq |t| \leq 1, \\ H(\lambda) = \frac{1}{2\pi n} \frac{\sin \frac{n\lambda}{2}}{\sin \frac{\lambda}{2}}, \quad -\pi \leq \lambda \leq \pi. \quad (5.33)$$

В этом случае спектральное окно имеет название *ядро Фейера* и отличается от ядра Дирихле тем, что принимает только неотрицательные значения. Дисперсия оценки спектральной плотности (ОСП) в этом случае приблизительно равна $2nf^2(\lambda)/3N$.

Окно Даниэля:

$$\begin{aligned} h(t) &= (\sin \pi t)/\pi t, \quad 0 \leq |t| \leq 1, \\ H(\lambda) &= \begin{cases} n/2\pi, & 0 \leq |\lambda| \leq \pi/n, \\ 0, & \pi/n < |\lambda| \leq \pi. \end{cases} \end{aligned} \quad (5.34)$$

В этом случае имеем прямоугольное спектральное окно. Асимптотическая дисперсия оценки принимает вид $D\{\hat{f}(\lambda)\} \cong nf^2(\lambda)/N$.

Окно Блэжмена — Тьюки:

$$\begin{aligned} h(t) &= 1 - 2a + 2a \cos \pi t, \quad 0 \leq |t| \leq 1, \\ H(\lambda) &= aD_n\left(\lambda - \frac{\pi}{n}\right) + (1 - 2a)D_n(\lambda) + aD_n\left(\lambda + \frac{\pi}{n}\right), \\ & \quad 0 \leq |\lambda| \leq \pi, \end{aligned} \quad (5.35)$$

где $D_n(\lambda)$ — ядро Дирихле. Асимптотическая дисперсия равна $D\{\hat{f}(\lambda)\} \cong \frac{2n}{N}(1 - 4a + 6a^2)f^2(\lambda)$. При $a = 0,25$ это окно известно как окно Тьюки — Хэннинга, а при $a = 0,23$ (что уменьшает первые локальные максимумы спектрального окна) — как окно Хемминга.

Окно Парзена:

$$\begin{aligned} h(t) &= \begin{cases} 1 - 6t^2 + 6|t|, & 0 \leq |t| \leq 0,5, \\ 2(1 - |t|)^3, & 0,5 \leq |t| \leq 1, \end{cases} \\ H(\lambda) &= \frac{6}{\pi n^3} \frac{\sin^4 \frac{n\lambda}{4}}{\sin^4 \frac{\lambda}{2}} \left(1 - \frac{2}{3} \sin^2 \frac{\lambda}{2}\right), \quad -\pi \leq \lambda \leq \pi. \end{aligned} \quad (5.36)$$

Асимптотическая дисперсия ОСП $D\{\hat{f}(\lambda)\} \cong 0,539nf^2(\lambda)/N$.

Трапецевидальное окно:

$$\begin{aligned} h(t) &= \begin{cases} 1, & 0 \leq |t| \leq \alpha, \\ (\beta - |t|)/(\beta - \alpha), & \alpha \leq |t| \leq \beta, \\ 0, & \beta \leq |t| \leq 1, \end{cases} \\ H(\lambda) &= \left(\sin^2 \frac{\beta\lambda}{2} - \sin^2 \frac{\alpha\lambda}{2}\right) / 2\pi(\beta - \alpha) \sin^2 \frac{\lambda}{2}, \quad -\pi \leq \lambda \leq \pi. \end{aligned} \quad (5.37)$$

Это спектральное окно известно под названием “ядро Валле — Пуссена”.

Окно Алексеева:

$$\begin{aligned} H(\lambda) &= \sum_{k=0}^p a_k \left(\frac{\lambda}{\pi}\right)^{2k}, \quad \lambda \in [-\pi, \pi], \\ \sum_{k=0}^p a_k &= 0, \quad a_0 > 0. \end{aligned} \quad (5.38)$$

Окно Бозмана:

$$h(t) = (1 - |t|)\cos \pi t + \frac{1}{\pi} \sin \pi t, \quad 0 \leq |t| \leq 1,$$

$$H(\lambda) = 2\pi n(1 + \cos n\lambda)/((n\lambda)^2 - \pi^2)^2, \quad -\pi \leq \lambda \leq \pi. \quad (5.39)$$

Окно Пуассона:

$$h(t) = \rho^{|t|}, \quad 0 \leq |t| \leq 1, \quad 0 < \rho < 1,$$

$$H(\lambda) = \frac{1}{2\pi} \frac{1 - \rho^{2/n}}{1 - 2\rho^{1/n}\cos \lambda + \rho^{2/n}}, \quad -\pi \leq \lambda \leq \pi. \quad (5.40)$$

Окно Римана – Ланцоша:

$$h(t) = (\sin t)/t, \quad -1 \leq t \leq 1,$$

$$H(\lambda) = \begin{cases} n/2, & 0 \leq |\lambda| \leq 1/n, \\ 0, & 1/n < |\lambda| \leq \pi. \end{cases} \quad (5.41)$$

Окно Гаусса – Вееритрасса:

$$h(t) = \exp\{-t^2/2\}, \quad -1 \leq t \leq 1,$$

$$H(\lambda) = \frac{n}{\sqrt{2\pi}} \exp\left\{-\frac{n^2\lambda^2}{2}\right\}, \quad -\pi \leq \lambda \leq \pi. \quad (5.42)$$

Окно Рисса – Бохнера:

$$h(t) = 1 - t^2, \quad -1 \leq t \leq 1,$$

$$H(\lambda) = D_n(\lambda) + \frac{1}{n^2} \frac{d^2 D_n(\lambda)}{d\lambda^2}, \quad -\pi \leq \lambda \leq \pi. \quad (5.43)$$

Окно Тьюки:

$$h(t) = \begin{cases} 1, & 0 \leq t < \rho, \quad 0 < \rho < 1, \\ \frac{1}{2} \left(1 + \frac{\cos \pi(|t| - \rho)}{1 - \rho}\right), & \rho \leq |t| \leq 1. \end{cases} \quad (5.44)$$

Окно Бартлетта – Пристли:

$$h(t) = \frac{3}{(\pi|t|)^3} (\sin \pi t - \pi|t|\cos \pi t), \quad 0 \leq |t| \leq 1,$$

$$H(\lambda) = \begin{cases} \frac{3n}{4\pi} \left(1 - \left(\frac{n\lambda}{\pi}\right)^2\right), & 0 \leq \lambda \leq \frac{\pi}{n}, \\ 0, & \frac{\pi}{n} \leq \lambda \leq \pi. \end{cases} \quad (5.45)$$

Наряду с перечисленными оценками можно пользоваться и выборочной спектральной плотностью со взвешенными слагаемыми, которую можно записать в виде

$$\hat{h}(\lambda) = \frac{1}{2\pi} \sum_{|\tau| < N} h\left(\frac{\tau}{N}\right) \hat{r}(\tau) e^{-i\tau\lambda}. \quad (5.46)$$

В отличие от (5.26), здесь суммирование происходит без усечения по всем возможным значениям τ . Обычно в (5.46) используется окно Бартлетта (5.33). Производя усечение суммирования в (5.46), получаем

$$\hat{h}(\lambda) = \frac{1}{2\pi} \sum_{|\tau| \leq n} \left(1 - \frac{|\tau|}{N}\right) \hat{r}(\tau) e^{-i\tau\lambda}. \quad (5.47)$$

Эта оценка является промежуточной оценкой между (5.26) и оценкой (5.46) с окном Бартлетта.

Иную промежуточную оценку между (5.26) и (5.46) предложил Парзен на основе оценок Бартлетта:

$$\hat{h}(\lambda) = \frac{1}{2\pi} \sum_{|\tau| \leq n} \left(1 - \left|\frac{\tau}{N}\right|^p\right) \left(1 - \frac{|\tau|}{N}\right) \hat{r}(\tau) e^{-i\tau\lambda}. \quad (5.48)$$

Из оценок (5.48) обычно встречаются оценки с $p = 2$.

Все перечисленные оценки состоятельны. Асимптотические свойства дисперсии оценивания выражаются соотношением

$$J = \lim_{N \rightarrow \infty} \left(\frac{N}{nf^2(\lambda)} D\{\hat{f}(\lambda)\} \right) = \int_{-1}^{+1} h^2(t) dt, \quad 0 < |\lambda| < \pi. \quad (5.49)$$

В точках $\lambda = 0$, $\lambda = \pm\pi$ предел в два раза больше.

Эти оценки являются, как уже сказано, также и асимптотически несмещенными. Характер асимптотического поведения смещения, к сожалению, не удается выяснить так просто, как это сделано для асимптотической дисперсии. Будем степень гладкости корреляционного окна $h(t)$ характеризовать числом q , таким, что

$$\lim_{t \rightarrow 0} \frac{1 - h(t)}{|t|^q} = \text{const} = h_0, \quad (5.50)$$

а степень гладкости спектральной плотности числом p , таким, что все производные спектральной плотности порядка не выше p ограничены и непрерывны. Тогда если $q > p$, то смещение ОСП является величиной порядка $o(n^{-p})$, где n — число усечения в сумме (5.26). Если же $q \leq p$, то смещение имеет порядок $o(h_0 n^{-q})$. Таким образом, если рассматривать класс процессов с вполне определенным значением p , то величину смещения можно сделать малой, выбирая окно $h(t)$ с $q > p$. Если $q \leq p$, то смещение уменьшают выбором окна $h(t)$ с малым значением h_0 . Для удобства сопоставления оценок данные о характеристиках асимптотической несмещенности и состоятельности сведены в табл. 5.1 для некоторых из перечисленных оценок.

Таблица 5.1

Оценка	q	h_0 из (5.50)	J из (5.49)
Дирихле (5.32)			2,0
Бартлетта (5.33)			0,6667
Даниэля (5.34)	2	$\pi^2/6$	0,9028
Блэкмена — Тьюки (5.35)	2	$a\pi^2$	$2(1 - 4a + 6a^2)$
Тьюки — Хэннинга (5.35)	2	$\pi^2/4$	0,75
Хемминга (5.35)	2	$0,23\pi$	0,7948
Парзена (5.36)	2	6	0,5393
Усеченная оценка (5.47)	∞		2,0
Парзена — Бартлетта (5.48) $p = 1$	1	1	0,6667
Парзена — Бартлетта (5.48) $p = 2$	2	1	1,0667

Вариация ОСП выражается в виде

$$V\{\hat{f}(\lambda)\} = D\{\hat{f}(\lambda)\} + (\mathbf{M}\{\hat{f}(\lambda)\} - f(\lambda))^2. \quad (5.51)$$

Как уже отмечалось, по крайней мере, для $q \leq p$ асимптотическое смещение пропорционально h_0 , определяемому (5.50). Из табл. 5.1 видно, что для рассмотренных оценок уменьшение асимптотической дисперсии (первое слагаемое в (5.51)) сопровождается увеличением асимптотического смещения (второе слагаемое в (5.51)). По-видимому, это является общим законом в данном случае и при выборе ОСП в каждой конкретной ситуации необходимо учитывать эти две противоположные тенденции.

Надежность оценивания спектральной плотности можно определить при помощи построения доверительных интервалов. Рассмотрим их построение для одного из типов оценок, введенных выше.

Из условий, приводящих к (5.21), следует, что случайные величины $(I_N(\omega_j + \omega_k)/\pi f(\omega_j + \omega_k))$, $-j < k < N/2 - j$, асимптотически независимы и распределены в соответствии с экспоненциальным распределением (т.е. распределением χ^2 с двумя степенями свободы). Это наводит на мысль о том, что асимптотическое распределение оценки $\hat{f}_g(\omega_j)$, определенной в (5.22), оказывается распределением СВ, являющейся линейной комбинацией независимых и распределенных по $\chi^2(2)$ СВ. Однако это распределение может быть, в свою очередь, аппроксимировано распределением СВ cY , где c — константа, а Y — СВ $\chi^2(\nu)$. Параметры c и ν находятся методом моментов, т.е. путем приравнивания среднего и дисперсии СВ c к среднему и дисперсии ОСП $\hat{f}_g(\omega_j)$. Для определения c и ν эта процедура дает уравнения $c\nu = f(\omega_j)$, $2c^2\nu = \left(\sum_{|k| \leq n} g_N^2(k) \right) f^2(\omega_j)$, откуда имеем $c = \frac{1}{2} f(\omega_j) \sum_{|k| \leq n} g_N^2(k)$, $\nu = 2 \left(\sum_{|k| \leq n} g_N^2(k) \right)^{-1}$. Число ν называется *эквивалентной степенью свободы* ОСП $\hat{f}_g(\omega)$. Распределение $\nu \hat{f}_g(\omega_j)/f(\omega_j)$, таким образом,

аппроксимируется распределением $\chi^2(\nu)$ и интервал

$$\left(\frac{\nu \hat{f}(\omega_j)}{\chi_{1-\alpha/2}^2(\nu)}, \frac{\nu \hat{f}(\omega_j)}{\chi_{\alpha/2}^2(\nu)} \right), \quad 0 < \omega_j < \pi, \quad (5.52)$$

является доверительным интервалом для $f(\omega_j)$ на уровне значимости α . Удобнее пользоваться доверительным интервалом для $\ln f(\omega_j)$, который получается логарифмированием доверительных границ в (5.52). Тогда вместо (5.52) имеем

$$(\ln \hat{f}(\omega_j) + \ln \nu - \ln \chi_{1-\alpha/2}^2(\nu), \ln \hat{f}(\omega_j) + \ln \nu - \ln \chi_{\alpha/2}^2(\nu)), \quad (5.53)$$

$$0 < \omega_j < \pi.$$

Доверительный интервал (5.53) удобен тем, что его ширина одинакова для всех $\omega_j \in (0, \pi)$.

При достаточно больших ν построение доверительного интервала для $f(\omega_j)$ можно осуществить, применив нормальную аппроксимацию распределения $\hat{f}_g(\omega_j)$. Из (5.25) получается предельное соотношение $\lim_{N \rightarrow \infty} \sum_{|k| \leq n} g_N^2(k) = 0$, по-

этому следует ожидать, что ν будет достаточно большим. Из свойств распределения $\chi^2(\nu)$ видно, что при $\nu > 30$ $\sqrt{2\chi^2(\nu)}$ практически не отличается от нормальной СВ со средним $\sqrt{2\nu - 1}$ и единичной дисперсией. Однако для достаточно больших N число слагаемых в (5.22) будет тоже велико и по теореме Линдберга — Феллера оценка $\hat{f}_g(\omega_j)$ должна сходиться к нормальной СВ. Из этих соображений вытекает, что для достаточно больших N $\hat{f}_g(\omega_j)$ будет близка к нормальной СВ со средним значением $f(\omega_j)$ и дисперсией $\sum_{|k| \leq n} g_N^2(k) f^2(\omega_j)$.

Используя это, получаем доверительный интервал для $f(\omega_j)$ на уровне значимости α :

$$\hat{f}(\omega_j) \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \hat{f}(\omega_j) \sqrt{\sum_{|k| \leq n} g_N^2(k)}, \quad 0 < \omega_j < \pi. \quad (5.54)$$

Здесь $\Phi^{-1}(\beta)$ — квантиль нормального распределения на уровне β . Вновь удобно вместо (5.54) использовать доверительный интервал для $\ln f(\omega_j)$, длина которого не зависит от ω_j . Нормальная аппроксимация распределения $\hat{f}(\omega)$ подразумевает, что $\ln \hat{f}(\omega)$ при $N \rightarrow \infty$ распределен тоже по нормальному закону со средним $\ln f(\omega_j)$ и дисперсией $\sum_{|k|} g_N^2(k)$. Откуда на уровне значимости α получаем доверительный интервал для $\ln f(\omega_j)$ вида

$$\ln \hat{f}(\omega_j) \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\sum_{|k| \leq n} g_N^2(k)}, \quad 0 < \omega_j < \pi, \quad (5.55)$$

длина которого не зависит от ω_j .

Лабораторная работа 20. Оценивание спектральной плотности временного ряда

Цель работы. Освоить методы построения спектральных плотностей по выборочным значениям случайных процессов. Сравнить анализ различных ОСП.

Порядок выполнения работы

Вначале необходимо получить временной ряд $\{x_1, x_2, \dots, x_N\}$, являющийся реализацией некоторого ССП с нулевым средним и известной спектральной плотностью. Для получения такой реализации можно воспользоваться подходящей моделью АРСС (p, q) , АР (p) или СС (q) . После того, как выбраны коэффициенты модели и получена выборка, необходимо по формуле (5.9) построить спектральную плотность и, используя методы гл. 3, построить АКФ для выбранной модели СП.

Задание 1. По имеющейся выборке СП получить по формулам (5.17) \hat{m} и $\hat{r}(\tau)$. Для заданного множества частот Фурье $\{\omega_j\}$ построить периодограмму (5.18). Сравнить периодограмму с истинной спектральной плотностью СП. Исследовать изменение выборочных статистик (5.17) и (5.18) в зависимости от величины N по сравнению с истинными значениями характеристик, оценками которых являются эти статистики. Построить график периодограммы (5.18) и сравнить его с графиком спектральной плотности. Будет ли периодограмма сильно флуктуирующей?

Задание 2. Выбрать весовые коэффициенты $g_N(k)$, удовлетворяющие условиям (5.24), (5.25), и построить ОСП в виде (5.22). При построении учесть, что число усечения n должно удовлетворять условиям при $N \rightarrow \infty$, $n \rightarrow \infty$, $n/N \rightarrow 0$. Выяснить поведение оценки (5.22) как функции N . Построить график $\hat{f}_g(\omega)$ на множестве частот Фурье. Сравнить его с графиком истинной спектральной плотности.

Задание 3. Построить доверительные интервалы для спектральной плотности, оцениваемой при помощи (5.22). Рассмотреть различные аппроксимации распределения ОСП, приводящие к доверительным интервалам (5.52)—(5.55).

Задание 4. Построить ОСП с корреляционными окнами (5.26) и со спектральными окнами (5.31) для каждой пары $h(t)$ и $H(\lambda)$, определяемой формулами (5.32)—(5.48). Провести сравнение как между ОСП во временной и частотной областях, так и между ОСП с различными типами окон.

5.2. Выделение скрытых периодичностей

Как показывает практика, часто возникают задачи следующего характера. Наблюдаемый временной ряд может иметь периодическую составляющую, но может ее и не иметь. Поскольку выборочные значения представляют смесь случайной и, возможно, периодической компоненты, решение вопроса содержится

ли периодическая компонента или отсутствует, является не тривиальным. Рассмотрим несколько конкретных постановок решения вопроса о наличии периодических составляющих во временном ряде $\{x_t, 1 \leq t \leq N\}$.

Выявление наличия гармоник. Значения процесса имеют следующую структуру:

$$x_t = m + a \cos \omega t + b \sin \omega t + W_t, \quad (5.56)$$

где a, b — не случайные постоянные; ω — заданная частота; W_t — гауссовский белый шум с дисперсией σ^2 . Гипотеза $H_0 : |a| + |b| = 0$ (гармоническое колебание отсутствует). Гипотеза $H_1 : |a| + |b| > 0$ (гармоника имеется в наличии).

Рассмотрим сначала простой случай, когда ω является частотой Фурье, т.е. $\omega = \omega_k = (2\pi k/N) \in (0, \pi)$. Выборка из наблюдений (5.56) может быть записана в виде (5.16)

$$\mathbf{x} = \sqrt{N}m\mathbf{e}_0 + \sqrt{\frac{N}{2}}a\mathbf{c}_k + \sqrt{\frac{N}{2}}b\mathbf{s}_k + \mathbf{W}. \quad (5.57)$$

Гипотеза H_0 отвергается в пользу H_1 , если значение периодограммы на частоте ω_k достаточно велико. При справедливой гипотезе H_0 ($a = b = 0$) величина $2I(\omega_k)$ распределена как $\sigma^2\chi^2(2)$ и является независимой от $(\sum_{t=1}^N x_t^2 - I(0) - 2I(\omega_k))$, которая имеет распределение $\sigma^2\chi^2(N-3)$. Отношение этих величин распределено в соответствии с F -распределением. Тогда получаем критерий: H_0 отвергается в пользу H_1 на уровне α , если

$$\xi = \frac{(N-3)I(\omega_k)}{\sum_{t=1}^N x_t^2 - I(0) - 2I(\omega_k)} > F_{1-\alpha}(2, N-3), \quad (5.58)$$

где $F_{1-\alpha}(2, N-3)$ — квантиль F -распределения (распределения Фишера) с 2 и $(N-3)$ степенями свободы на уровне $1-\alpha$.

Если ω не является частотой Фурье, анализ несколько усложняется, так как векторы $\mathbf{c} = \sqrt{2/N}(\cos \omega, \cos 2\omega, \dots, \cos N\omega)^*$, $\mathbf{s} = \sqrt{2/N}(\sin \omega, \sin 2\omega, \dots, \sin N\omega)^*$ и \mathbf{e}_0 не являются ортогональными. Однако тест остается аналогичным. В отличие от (5.57) имеем

$$\mathbf{x} = \sqrt{N}m\mathbf{e}_0 + \sqrt{\frac{N}{2}}a\mathbf{c} + \sqrt{\frac{N}{2}}b\mathbf{s} + \mathbf{W}.$$

Вначале найдем оценки МНК для m, a, b из условия

$$d^*(m, a, b) \rightarrow \min_{m, a, b},$$

где

$$d(m, a, b) = \left(\mathbf{x} - \sqrt{N}m\mathbf{e}_0 - \sqrt{\frac{N}{2}}a\mathbf{c} - \sqrt{\frac{N}{2}}b\mathbf{s} \right).$$

Обозначим эти оценки \hat{m} , \hat{a} , \hat{b} . Тогда статистика критерия (5.58) (левая часть неравенства) будет вычисляться следующим образом:

$$\frac{(\bar{x}\mathbf{e}_0 - x + d(\hat{m}, \hat{a}, \hat{b}))^*(\bar{x}\mathbf{e}_0 - x + d(\hat{m}, \hat{a}, \hat{b}))}{d^*(\hat{m}, \hat{a}, \hat{b})d(\hat{m}, \hat{a}, \hat{b})}(N-3), \quad (5.59)$$

где $\bar{x} = N^{-1} \sum_{t=1}^N x_t$. И гипотеза H_0 отвергается в пользу H_1 на уровне значимости α , если $\xi > F_{1-\alpha}(2, N-3)$.

Выявление негармонической периодической составляющей. Пусть заданы целочисленный период $p < N$ и $q = [(p-1)/2]$. Всякая функция f_t , заданная на целочисленных t с периодом p , может быть представлена в виде

$$f_t = m + \sum_{k=1}^q (a_k \cos(2\pi kt/p) + b_k \sin(2\pi kt/p)) + a_{p/2}(-1)^t, \quad (5.60)$$

где $a_{p/2} = 0$, если p — нечетное. Модель данных в рассматриваемом случае примем следующей:

$$x_t = f_t + W_t, \quad 1 \leq t \leq N, \quad (5.61)$$

где f_t задана в (5.60); W_t — гауссовский белый шум с дисперсией σ^2 . Определим испытываемые гипотезы:

- 1) H_0 : $a_k = b_k = 0$, $1 \leq k \leq q$ (т.е. $\sum_{k=1}^q (|a_k| + |b_k|) = 0$);
- 2) H_1 : $\sum_{k=1}^q (|a_k| + |b_k|) > 0$.

Определим N -компонентные векторы-столбцы:

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, \dots, x_N)^*, \\ \mathbf{e}_0 &= N^{-1/2}(1, 1, \dots, 1)^*, \\ \mathbf{g}_k &= \sqrt{2/N}(\cos \psi_k, \cos 2\psi_k, \dots, \cos N\psi_k)^*, \\ \mathbf{h}_k &= \sqrt{2/N}(\sin \psi_k, \sin 2\psi_k, \dots, \sin N\psi_k)^*, \end{aligned}$$

где $\psi_k = 2\pi k/p$, $1 \leq k \leq [p/2]$. Пусть G — $(N \times p)$ -матрица $G = (\mathbf{e}_0 \mathbf{g}_1 \mathbf{h}_1 \mathbf{g}_2 \mathbf{h}_2 \dots)$. Последним столбцом в G будет $\mathbf{g}_{p/2}$, если p — четное, или $\mathbf{h}_{(p-1)/2}$, если p — нечетное. Из (5.60), (5.61) следует, что $\mathbf{x}^*(I - G(G^*G)^{-1}G^*)\mathbf{x}$ распределено как $\sigma^2\chi^2(N-p)$. Кроме того, при справедливой гипотезе H_0 величина $(\mathbf{x}^*G(G^*G)^{-1}G^* - \sqrt{N}\mathbf{x}\mathbf{e}_0^*)(G(G^*G)^{-1}G^*\mathbf{x} - \sqrt{N}\mathbf{x}\mathbf{e}_0)$ распределена как $\sigma^2\chi^2(p-1)$ и не зависит от $\mathbf{x}^*(I - G(G^*G)^{-1}G^*)\mathbf{x}$. Поэтому опять можно построить критерий типа (5.58). Гипотеза H_0 отвергается в пользу H_1 на уровне α , если

$$\xi = \frac{(N-p)(N(\bar{x})^2 + \mathbf{x}^*G(G^*G)^{-1}G^*\mathbf{x})}{(p-1)\mathbf{x}^*(I - G(G^*G)^{-1}G^*)\mathbf{x}} > F_{1-\alpha}(p-1, N-p). \quad (5.62)$$

Здесь, как и прежде, \bar{x} — выборочное среднее наблюдений; I — единичная $N \times N$ -матрица.

В случае, когда $N = kp$, где k — целое число, вычисления существенно упрощаются в связи с тем, что p векторов $\mathbf{e}_0, \mathbf{g}_1, \mathbf{h}_1, \mathbf{g}_2, \mathbf{h}_2, \dots$ становятся ортогональными. Поэтому

$$\mathbf{x}^* G (G^* G)^{-1} G^* \mathbf{x} = I(0) + 2 \sum_{1 \leq j < p/2} I(\omega_{kj}) + \delta_p I(\pi),$$

где $I(\omega)$ — периодограмма (5.13); $\delta_p = 1$, если p — четное, и $\delta_p = 0$, если p — нечетное. В связи с этим статистика критерия (5.62) сводится к виду

$$\xi = \frac{(N-p) \left(2 \sum_{1 \leq j < p/2} I(\omega_{kj}) + \delta_p I(\pi) \right)}{(p-1) \left(\sum_{t=1}^N x_t^2 - I(0) - \delta_p I(\pi) - 2 \sum_{1 \leq j < p/2} I(\omega_{kj}) \right)}, \quad (5.63)$$

где $\omega_{kj} = 2\pi k j / N$, и решение принимается так же, как и в (5.62).

Выявление скрытой периодичности с неизвестной частотой. Если $\{x_t\}$ — гауссовский белый шум с дисперсией σ^2 и $\mathbf{x} = (x_1, x_2, \dots, x_N)^*$, то случайные величины $V_k = I(\omega_k) / \sigma^2$ независимы и одинаково распределены, как $\chi^2(2)/2$, $1 \leq k \leq [(N-1)/2] = q$. Распределение V_k является экспоненциальным со средним значением, равным единице. Случайные величины

$$Y_i = \sum_{k=1}^i V_k / \sum_{k=1}^q V_k = \sum_{k=1}^i I(\omega_k) / \sum_{k=1}^q I(\omega_k), \quad 1 \leq i \leq q-1,$$

распределены также, как порядковые статистики выборки $(q-1)$ независимых случайных переменных, каждая из которых равномерно распределена в интервале $(0, 1)$. Определим $Y_0 = 0, Y_q = 1$ и

$$\eta = \max_{1 \leq i \leq q} (Y_i - Y_{i-1}) = \max_{1 \leq i \leq q} \frac{I(\omega_i)}{\sum_{i=1}^q I(\omega_i)}.$$

Тогда функция распределения (ФР) СВ η запишется в виде

$$P(\eta \leq z) = \sum_{j=0}^q (-1)^j \binom{q}{j} (1 - jz)_t^{q-1}, \quad (5.64)$$

где $(u)_t = \max\{0, u\}$.

Эти результаты могут быть использованы для построения теста Фишера при проверке справедливости одной из альтернативных гипотез:

- 1) H_0 : $\{x_t\}$ — гауссовский белый шум;

2) H_1 : $\{x_t\}$ содержит аддитивную детерминированную периодическую компоненту с нефиксированной частотой. Идея теста — отвергать H_0 , если периодограмма (5.13) содержит значение, существенно превышающее ее среднюю величину, т.е. если статистика

$$\xi = q \left(\max_{1 \leq i \leq q} I(\omega_i) \right) / \left(\sum_{i=1}^q I(\omega_i) \right) \quad (5.65)$$

принимает достаточно большое значение. Предположим, что выборка $\{x_t\}$ такова, что статистика (5.65) приняла значение z . Тогда на основе (5.64) имеем

$$P\{\xi > z\} = 1 - \sum_{j=0}^q (-1)^j \binom{q}{j} (1 - jz)^{q-1}. \quad (5.66)$$

Критерий состоит в том, что если эта вероятность меньше, чем α , то H_0 отвергается в пользу H_1 .

К построению теста в рассматриваемом случае можно подойти иначе. Построим дискретную ФР с конечными разрывами величиной $(q-1)^{-1}$ в точках, определяемых СВ Y_i , $1 \leq i \leq q-1$. Эта ФР является выборочной для выборки объемом $(q-1)$ из равномерного распределения на $[0, 1]$. Таким образом, проверка справедливости гипотезы H_0 состоит в использовании теста Колмогорова — Смирнова относительно того, совместима ли построенная ФР с равномерной ФР. Иначе говоря, не выходит ли построенная эмпирическая ФР из границ области, определяемой прямыми $y = x + K_\alpha(N)(q-1)^{-1/2}$, $0 < x < 1$, где $K_\alpha(N)$ — критические значения для наибольшего отклонения эмпирического распределения от теоретического на уровне значимости α . Если эмпирическая ФР выходит за указанные границы, то H_0 отвергается в пользу H_1 на уровне α .

Заметим, что тесты Фишера и Колмогорова — Смирнова могут быть при больших N применены и не для гауссовского белого шума W_t . Эти тесты могут быть использованы для проверки гипотезы H_0 о том, имеет ли временной ряд $\{x_t\}$ спектральную плотность $f(\lambda)$ или нет. Для этого достаточно при определении СВ Y_i и ξ (в (5.65)) заменить $I(\omega_k)$ на $I(\omega_k)/f(\omega_k)$.

Лабораторная работа 21. Выявление скрытых периодичностей

Цель работы. Ознакомиться с методами выявления скрытых периодичностей и исследовать их эффективность.

Порядок выполнения работы

Задание 1. Получить реализацию временного ряда с элементами вида (5.56), содержащими смесь гармоник с белым гауссовским шумом. Рассмотреть два варианта, когда заданная частота гармоник либо является, либо не является частотой Фурье. Исследовать несколько случаев для различных значений параметров m, a, b . При использовании критериев (5.58) и

(5.59) удобно определять критические значения $F_{1-\alpha}(P, N)$, учитывая, что $\lim_{N \rightarrow \infty} F_{1-\alpha}(P, N) = F_{1-\alpha}(P, \infty) = \chi_{1-\alpha}^2(P)/P$.

Задание 2. Получить реализацию временного ряда с элементами вида (5.60), (5.61), содержащими негармоническую периодическую составляющую с известным целочисленным периодом. Рассмотреть два варианта, когда объем выборки кратен периоду неслучайной составляющей временного ряда и когда это не выполняется. Исследовать несколько различных функций (5.60).

Задание 3. Рассмотреть применение тестов Фишера и Колмогорова — Смирнова к реализациям, которые были получены при выполнении предыдущих заданий. Используя критерий Колмогорова — Смирнова, нужно учитывать, что при $N \rightarrow \infty$ пределы имеют значения $K_{0,01}(N) \rightarrow 1,628$, $K_{0,02}(N) \rightarrow 1,517$, $K_{0,05}(N) \rightarrow 1,358$, $K_{0,1}(N) \rightarrow 1,224$, $K_{0,2}(N) \rightarrow 1,073$.

Задание 4. Получить реализацию временного ряда с известной спектральной плотностью $f(\lambda)$, например, при помощи имитации процесса АРСС (p, q) . Вычислить периодограмму временного ряда $I(\omega_k)$. Использовать тесты Фишера и Колмогорова — Смирнова для решения вопроса о соответствии временного ряда спектральной плотности $f(\lambda)$. При вычислении периодограммы применить алгоритм быстрого преобразования Фурье (БПФ).

Глава 6

МНОГОМЕРНЫЕ ВРЕМЕННЫЕ РЯДЫ

6.1. Оценивание числовых характеристик многомерных числовых рядов

Встречающиеся на практике временные ряды, как правило, могут (а в некоторых случаях и должны) изучаться как компоненты определенного многомерного СП. Так, например, количество осадков и урожайность сельскохозяйственных культур, рассматриваемые как функции времени, составляют двухмерную векторную функцию времени, компоненты которой изменяются случайным образом по соответствующей стохастической модели и стохастически зависимы между собой. В общем случае имеет смысл рассматривать временной ряд $\{\mathbf{x}_t, t \in T\}$ с элементами, являющимися n -мерными векторами $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tn})^*$. Для многомерных временных рядов второго порядка требуется, чтобы $\mathbf{M}\{x_{ti}^2\} < \infty$ для всех t и i . Математическим ожиданием такого временного ряда является вектор

$$\begin{aligned} \mathbf{m}_t = \mathbf{M}\{\mathbf{x}_t\} &= (\mathbf{M}\{x_{t1}\}, \mathbf{M}\{x_{t2}\}, \dots, \mathbf{M}\{x_{tn}\})^* = \\ &= (m_{t1}, m_{t2}, \dots, m_{tn})^*. \end{aligned} \quad (6.1)$$

Корреляционные свойства многомерного временного ряда определяются матричной функцией, каждый элемент которой является взаимной корреляционной функцией соответствующих компонент временного ряда: $R(t, t + \tau) =$

$= \mathbf{M}\{\mathbf{x}_t \mathbf{x}_{t+\tau}^*\} = (\mathbf{M}\{x_{ti} x_{t+\tau, j}\}) = (R_{ij}(t, t + \tau)), 1 \leq i \leq n, 1 \leq j \leq n$. Чаще пользуются ковариационными матрицами (КМ) временных рядов

$$R^0(t, t + \tau) = \mathbf{M}\{(\mathbf{x}_t - \mathbf{m}_t)(\mathbf{x}_{t+\tau} - \mathbf{m}_{t+\tau})^*\} = (R_{ij}^0(t, t + \tau)). \quad (6.2)$$

Если элементы временного ряда распределены по нормальному закону, то совокупность $(\mathbf{m}_t, R^0(t, t + \tau))$ полностью характеризует статистические свойства временного ряда.

Для стационарных временных рядов будем иметь вместо (6.1) и (6.2) равенства

$$\begin{aligned} \mathbf{m}_t = \mathbf{m} &= (m_1, m_2, \dots, m_n), \quad t = 1, 2, \dots; \\ R^0(t, t + \tau) &= r(\tau) = (r_{ij}(\tau)), \quad t = 1, 2, \dots, \end{aligned} \quad (6.3)$$

т.е. в этом случае вектор \mathbf{m}_t не зависит от временного параметра t и является постоянным, а матрица $R^0(t, t + \tau)$ также не зависит от параметра t и определяется только временным сдвигом τ .

Обычно удобнее пользоваться КМ, составленной из нормированных взаимных корреляционных функций компонент временного ряда

$$\rho(\tau) = (\rho_{ij}(\tau)), \quad 1 \leq i \leq n, \quad 1 \leq j \leq n, \quad (6.4)$$

где элементы матрицы $\rho(\tau)$ задаются соотношением

$$\rho_{ij}(\tau) = r_{ij}(\tau) / \sqrt{r_{ii}(0)r_{jj}(0)}. \quad (6.5)$$

Свойства функций $r_{ij}(\tau)$ и $\rho_{ij}(\tau)$ рассматривались в гл. 4.

Многомерный белый шум описывается как векторный процесс \mathbf{W}_t , для которого

$$\begin{aligned} \mathbf{M}\{\mathbf{W}_t\} &= 0, \quad \mathbf{M}\{\mathbf{W}_t \mathbf{W}_t^*\} = Q, \\ \mathbf{M}\{\mathbf{W}_t \mathbf{W}_{t+\tau}^*\} &= 0, \quad \tau > 0, \quad t = 1, 2, \dots \end{aligned} \quad (6.6)$$

Как правило, временные ряды являются реализациями СП, представляемых в виде разложения

$$x_t = m_t + \sum_{j=-\infty}^{+\infty} C_j W_{t-j}, \quad t = 1, 2, \dots, \quad (6.7)$$

где C_j — последовательность матриц с абсолютно суммируемыми элементами. Это значит, что если C_j — матрица с элементами $C_j(k, l)$, то для любой пары k, l выполняется неравенство $\sum_{j=-\infty}^{+\infty} |C_j(k, l)| < \infty$. При $C_j = 0$ для всех $j < 0$ данное разложение известно под названием “разложение Волда”. Для процесса (6.7) имеем

$$\mathbf{M}\{\mathbf{x}_t\} = \mathbf{m}_t, \quad r(\tau) = \sum_{j=-\infty}^{+\infty} C_j Q C_{j+\tau}^*. \quad (6.8)$$

Рассмотрим стационарный временной ряд $\{x_t, 1 \leq t \leq N\}$ с неизвестными \mathbf{m} и $r(\tau)$, где $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tn})^*$. Несмещенной оценкой \mathbf{m} является выборочное среднее

$$\hat{\mathbf{m}} = \frac{1}{N} \sum_{t=1}^N x_t = (\hat{m}_1, \hat{m}_2, \dots, \hat{m}_n)^*, \quad (6.9)$$

где $\hat{m}_i = \frac{1}{N} \sum_{t=1}^N x_{ti}$, $1 \leq i \leq n$.

Для стационарных временных рядов сходимость оценок (6.9) к истинным значениям характеризуется следующими результатами.

Если $r_{ii}(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$, $1 \leq i \leq n$, то

$$\lim_{N \rightarrow \infty} \mathbf{M}\{(\hat{\mathbf{m}} - \mathbf{m})^*(\hat{\mathbf{m}} - \mathbf{m})\} = 0. \quad (6.10)$$

Если $\sum_{\tau=-\infty}^{+\infty} |r_{ij}(\tau)| < \infty$, $1 \leq i, j \leq n$, то

$$\lim_{N \rightarrow \infty} N \mathbf{M}\{(\hat{\mathbf{m}} - \mathbf{m})^*(\hat{\mathbf{m}} - \mathbf{m})\} = \sum_{i=1}^n \sum_{\tau=-\infty}^{+\infty} r_{ii}(\tau) = r_*. \quad (6.11)$$

Оценка $\hat{\mathbf{m}}$ в (6.10), (6.11) находится из формулы (6.9) и является асимптотически нормальной, если выполняются некоторые дополнительные предположения.

Если \mathbf{x}_t представляется в виде (6.7), где \mathbf{W}_t — НОР СВ со свойствами (6.6), а C_j — последовательность $(n \times n)$ -матриц с абсолютно суммируемыми элементами, то $\hat{\mathbf{m}}$, вычисляемое по формуле (6.9), является асимптотически нормальным с параметрами

$$\begin{aligned} \mathbf{M}\{\hat{\mathbf{m}}\} &= \mathbf{m}, \quad \mathbf{M}\{(\hat{\mathbf{m}} - \mathbf{m})^*(\hat{\mathbf{m}} - \mathbf{m})\} = \\ &= \frac{1}{N} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} C_k Q C_l^* = G_N. \end{aligned} \quad (6.12)$$

Это свойство может быть использовано при построении доверительных областей для \mathbf{m} . Например, если КМ G_N в (6.12) является невырожденной и известной, то асимптотическая доверительная область для \mathbf{m} на уровне значимости α определяется как множество

$$\{\mathbf{m} \mid (\mathbf{m} - \hat{\mathbf{m}})^* G_N^{-1} (\mathbf{m} - \hat{\mathbf{m}}) \leq \chi_{1-\alpha}^2(n)\}, \quad (6.13)$$

где $\chi_{1-\alpha}^2(n)$ — соответствующий квантиль χ^2 -распределения с n степенями свободы. Такое построение доверительной области практически не используется, так как редко матрица G_N оказывается известной в случае, когда \mathbf{m} неизвестно. Если может быть найдена состоятельная оценка \hat{G}_N матрицы G_N , то в (6.13) подставляют ее вместо G_N и получают необходимую доверительную область. Однако в общем случае трудно находить состоятельную оценку матрицы G_N .

Более простой подход — построение доверительных интервалов для каждой компоненты m_i вектора \mathbf{m} , основанное на выборках $\{x_{1i}, x_{2i}, \dots, x_{Ni}\}$, $1 \leq i \leq n$. Комбинируя найденные таким образом доверительные интервалы, можно получить доверительную область для вектора \mathbf{m} .

Если условия справедливости (6.12) выполняются, то компонента \hat{m}_i выборочного среднего (6.9) является асимптотически нормальной с математическим ожиданием m_i и дисперсией

$$d_i = \frac{1}{N} \sum_{\tau=-\infty}^{+\infty} r_{ii}(\tau), \quad 1 \leq i \leq n. \quad (6.14)$$

Пусть $\{\tau_N\}$ — последовательность чисел, удовлетворяющих условию $(\tau_N/N) \rightarrow 0$ при $N \rightarrow \infty$. Состоятельной оценкой является

$$\hat{d}_i = \frac{1}{N} \sum_{|\tau| \leq \tau_N} \left(1 - \frac{|\tau|}{\tau_N}\right) \hat{r}_{ii}(\tau), \quad 1 \leq i \leq n, \quad (6.15)$$

где $\hat{r}_{ii}(\tau)$ — оценка автоковариационной функции i -й компоненты x_{ti} вектора \mathbf{x}_t .

Состоятельной оценкой автоковариационной функции $\hat{r}_{ii}(\tau)$ является следующая выборочная ковариационная функция:

$$\hat{r}_{ii}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (x_{ti} - \hat{m}_i)(x_{t+\tau,i} - \hat{m}_i), \quad 1 \leq i \leq n. \quad (6.16)$$

Таким образом, на уровне значимости α доверительный интервал для i -й компоненты вектора \mathbf{m} при достаточно больших N задается соотношением

$$|m_i - \hat{m}_i| \leq \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{d}_i}, \quad 1 \leq i \leq n, \quad (6.17)$$

где $\Phi^{-1}(\beta)$ — квантиль нормального распределения на уровне β ; \hat{d}_i вычисляется по формулам (6.15), (6.16).

Так как

$$P\{|m_i - \hat{m}_i| \leq \Phi^{-1}(\beta) \sqrt{\hat{d}_i}, 1 \leq i \leq n\} \geq 1 - \sum_{i=1}^n P\{|m_i - \hat{m}_i| \leq \Phi^{-1}(\beta) \sqrt{\hat{d}_i}\},$$

то доверительная область для вектора \mathbf{m} , в которой он находится с вероятностью, не меньшей чем $(1 - \alpha)$, определяется следующим образом:

$$\left\{ \mathbf{m} \mid |m_i - \hat{m}_i| \leq \Phi^{-1} \left(1 - \frac{\alpha}{2n}\right) \sqrt{\hat{d}_i}, \quad 1 \leq i \leq n \right\}. \quad (6.18)$$

Для больших значений n эта доверительная область будет значительно шире области, в которую попадает \mathbf{m} с вероятностью $(1 - \alpha)$. Тем не менее, поскольку

(6.18) строить немного легче, чем (6.13), для небольших n удобнее применять доверительную область (6.18).

В качестве оценки КМ $r(\tau)$ естественно пользоваться выборочной КМ (ВКМ), которая задается соотношением

$$\hat{r}(\tau) = (\hat{r}_{ij}(\tau)) = \frac{1}{N} \sum_{t=1}^{N-\tau} (\mathbf{x}_{t+\tau} - \hat{\mathbf{m}})(\mathbf{x}_t - \hat{\mathbf{m}})^*, \quad 0 \leq \tau \leq N-1. \quad (6.19)$$

Через элементы матрицы определим оценки элементов КМ, составленной из нормированных взаимных ковариационных функций

$$\hat{\rho}_{ij}(\tau) = \hat{r}_{ij}(\tau) / \sqrt{\hat{r}_{ii}(0)\hat{r}_{jj}(0)}, \quad 1 \leq i, j \leq n. \quad (6.20)$$

В условиях справедливости формул (6.12) оценки (6.19) и (6.20) для двумерных временных рядов являются состоятельными по вероятности, т.е. при $N \rightarrow \infty$ для всякого фиксированного $\tau \geq 0$

$$\hat{r}_{ij}(\tau) \rightarrow r_{ij}(\tau), \quad \hat{\rho}_{ij}(\tau) \rightarrow \rho_{ij}(\tau), \quad 1 \leq i, j \leq 2. \quad (6.21)$$

Скорость сходимости в этом случае характеризуется следующими соотношениями. Если матрица Q в (6.6) является диагональной (т.е. компоненты W_t независимы между собой), то для всякого $\tau \geq 0$ распределение $\hat{\rho}_{12}(\tau)$ асимптотически нормально с нулевым средним и дисперсией

$$d_{12} = \frac{1}{N} \sum_{j=-\infty}^{+\infty} \rho_{11}(j)\rho_{22}(j). \quad (6.22)$$

Если $\tau_1 \neq \tau_2$, то вектор $(\hat{\rho}_{12}(\tau_1), \hat{\rho}_{12}(\tau_2))^*$ асимптотически нормален с нулевым средним компонент, дисперсиями компонент, определяемыми (6.22), и ковариацией компонент, вычисляемой по формуле

$$\frac{1}{N} \sum_{j=-\infty}^{+\infty} \rho_{11}(j)\rho_{22}(j + \tau_1 - \tau). \quad (6.23)$$

Это дает возможность конструировать различные тесты относительно компонент при анализе их корреляционных свойств. Например, если одна из компонент является белым шумом, то $\hat{\rho}_{12}(\tau)$ имеет дисперсию $d_{12} = N^{-1}$, какими бы свойствами ни обладала другая компонента.

Рассмотрим задачу проверки на независимость двух стационарных компонент временного ряда. Как следует из (6.22), асимптотическое распределение $\hat{\rho}_{12}(\tau)$ зависит как от $\rho_{11}(\cdot)$, так и от $\rho_{22}(\cdot)$. Поэтому тест независимости двух компонент ряда не может быть основан только на оценивании значений $\rho_{12}(\tau)$ без учета других свойств компонент ряда. Эту трудность можно преодолеть

путем декорреляции рассматриваемых компонент перед тем, как оценивать выборочную взаимную ковариацию $\rho_{12}(\tau)$. В данном случае декорреляция достигается соответствующим линейным преобразованием. В частности, если $\{x_{t1}\}$ и $\{x_{t2}\}$ — обратимые процессы АРСС (p, q) , таким преобразованием является

$$z_{ti} = \sum_{j=0}^{\infty} \theta_j^{(i)} x_{t-j,i}, \quad i = 1, 2, \quad (6.24)$$

где коэффициенты $\theta_j^{(i)}$ определяются из разложения

$$\sum_{j=0}^{\infty} \theta_j^{(i)} u^j = a^{(i)}(u)/b^{(i)}(u), \quad |u| < 1, \quad i = 1, 2,$$

в котором $a^{(i)}(u)$ и $b^{(i)}(u)$ — полиномы, задающие АР и СС соответственно.

Так как на практике истинная модель обычно неизвестна и наблюдений $\{x_t, t \leq 0\}$ не имеется в наличии, удобно заменить последовательности $\{z_{ti}\}$, $i = 1, 2$, остатками $\{\hat{W}_{ti}, 1 \leq t \leq N\}$, которые рассматривались в § 4.3. Если подобранная модель АРСС (p, q) является истинной, то последовательности остатков будут последовательностями белого шума, а с учетом асимптотической нормальности — последовательностями НОР СВ. Таким образом, при справедливой гипотезе о независимости компонент временного ряда выборочные автокорреляции $\hat{\rho}_{12}(\tau_1)$ и $\hat{\rho}(\tau_2)$, $\tau_1 \neq \tau_2$, последовательностей $\{z_{t1}\}$ и $\{z_{t2}\}$ асимптотически нормальны, независимы, имеют нулевые средние и дисперсии, равные N^{-1} . Поэтому тест на независимость компонент на уровне значимости α заключается в проверке неравенства

$$|\rho_{12}(\tau)| \leq \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) N^{-1/2}, \quad \tau = 0, 1, 2, \dots, \quad (6.25)$$

где $\Phi^{-1}(\beta)$ — квантиль стандартного нормального распределения на уровне β .

Если декоррелировать только одну из двух исходных компонент (например, $\{x_{t1}\}$), то при справедливой гипотезе о независимости компонент из (6.23) следует, что выборочные автокорреляции $\hat{\rho}_{12}(\tau_1)$ и $\hat{\rho}_{12}(\tau_2)$, $\tau_1 \neq \tau_2$, последовательностей $\{z_{t1}\}$ и $\{x_{t2}\}$ асимптотически нормальны с нулевым средним, дисперсией N^{-1} и ковариацией $N^{-1}\rho_{22}(\tau_2 - \tau_1)$, где $\rho_{22}(\cdot)$ — АКФ последовательности $\{x_{t2}\}$. Следовательно, и в этом случае тест на независимость компонент на уровне значимости α заключается в проверке выполнения неравенства (6.25) для любого значения τ .

При получении (6.22) и (6.23), на которых основан описанный тест, предполагалось, что компоненты временного ряда удовлетворяют условиям разложения (6.6). Для нормальных временных рядов этого можно и не требовать, поскольку для них справедлива формула Бартлетта. Если $\{x_t\}$ — двухмерный гауссовский процесс и его АКФ удовлетворяет неравенству

$$\sum_{\tau=-\infty}^{+\infty} |r_{ij}(\tau)| < \infty, \quad 1 \leq j, i \leq 2,$$

то

$$\begin{aligned}
\lim_{N \rightarrow \infty} N \text{cov}(\hat{\rho}_{12}(\tau_1), \hat{\rho}_{12}(\tau_2)) &= \sum_{j=-\infty}^{+\infty} \left(\rho_{11}(j) \rho_{22}(j + \tau_1 - \tau_2) + \right. \\
&+ \rho_{12}(j + \tau_1) \rho_{21}(j - \tau_2) - \rho_{12}(\tau_2) \left(\rho_{11}(j) \rho_{21}(j + \tau_1) + \rho_{22}(j) \rho_{21}(j - \tau_1) \right) - \\
&- \rho_{12}(j) \left(\rho_{11}(j) \rho_{21}(j + \tau_2) + \rho_{22}(j) \rho_{21}(j - \tau_1) \right) + \\
&\left. + \rho_{12}(\tau_1) \rho_{12}(\tau_2) \left(\frac{1}{2} \rho_{11}^2(j) + \rho_{12}^2(j) + \frac{1}{2} \rho_{22}^2(j) \right) \right). \quad (6.26)
\end{aligned}$$

Отсюда, в частности, следует, что если компонентами $\{x_t\}$ являются белые шумы и $\rho_{12}(\tau) = 0$, $\tau \in [a, b]$, то

$$\lim_{N \rightarrow \infty} ND\{\hat{\rho}_{12}(\tau)\} = 1, \quad \tau \in [a, b], \quad (6.27)$$

т.е. асимптотическая дисперсия $\hat{\rho}_{12}(\tau)$ в этом случае тоже равна N^{-1} .

Среди многомерных СП своей практической распространенностью выделяются процессы АРСС, которые в многомерном варианте определяются как процессы $\{x_t\}$, удовлетворяющие многомерному разностному уравнению

$$\mathbf{x}_t = \sum_{i=1}^p A_i \mathbf{x}_{t-i} + \mathbf{W}_t + \sum_{j=1}^q B_j \mathbf{W}_{t-j}, \quad t = 1, 2, \dots \quad (6.28)$$

Здесь $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tn})^*$ — n -вектор; $A_1, A_2, \dots, A_p, B_1, \dots, B_q$ — $(n \times n)$ -матрицы; $\{\mathbf{W}_t\}$ — случайные векторы с компонентами: $\mathbf{W}_t = (W_{t1}, W_{t2}, \dots, W_{tn})^*$, для которых выполняются условия (6.6). Последовательность векторов \mathbf{x}_t , заданных уравнением (6.28), называется *многомерным* (при $n > 1$) *процессом АРСС* (p, q) .

Для рассмотрения свойства многомерного процесса АРСС (p, q) определим матричные полиномы

$$A(z) = I - A_1 z - \dots - A_p z^p, \quad B(z) = I + B_1 z + \dots + B_q z^q, \quad (6.29)$$

где I — единичная $(n \times n)$ -матрица.

Критерий каузальности процесса \mathbf{x}_t . Если $\det A(z) \neq 0$ для всех $|z| \leq 1$, то (6.28) имеет единственное стационарное решение $\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{W}_{t-j}$, где $(n \times n)$ -матрицы Ψ_j удовлетворяют равенствам

$$\sum_{j=0}^{\infty} \Psi_j z^j = \Psi(z) = A^{-1}(z) B(z), \quad |z| \leq 1. \quad (6.30)$$

Критерий обратимости процесса \mathbf{x}_t . Если $\det B(z) \neq 0$ для всех $|z| \leq 1$ и $\{\mathbf{x}_t\}$ — стационарное решение (6.28), то $\mathbf{W}_t = \sum_{j=0}^{\infty} \Phi_j x_{t-j}$, где матрицы Φ_j

определяются равенством

$$\sum_{j=0}^{\infty} \Phi_j z^j = \Phi(z) = B^{-1}(z)A(z), \quad |z| \leq 1. \quad (6.31)$$

Матрицы Ψ_j и Φ_j находятся из следующих рекуррентных соотношений:

$$\Psi_0 = I, \quad \Psi_j = \sum_{i=1}^j A_i \Psi_{j-i} + B_j, \quad j = 1, 2, \dots; \quad (6.32)$$

$$\Phi_0 = I, \quad \Phi_j = -\sum_{i=1}^j B_i \Phi_{j-i} - A_j, \quad j = 1, 2, \dots, \quad (6.33)$$

где $\Psi_j = 0$ для $j < 0$; $A_j = 0$ для $j > p$; $B_j = 0$ для $j > q$.

Ковариационная матричная функция каузального процесса АРСС (p, q) определяется равенством

$$r(\tau) = (r_{ij}(\tau)) = \mathbf{M}\{\mathbf{x}_t \mathbf{x}_{t+\tau}^*\} = \sum_{k=0}^{\infty} \Psi_k Q \Psi_{k+\tau}^*, \quad \tau \geq 0, \quad (6.34)$$

где Q — КМ белого шума \mathbf{W}_t со свойствами (6.6). Существует $\gamma \in (0, 1)$ и константа K , такие, что элементы матрицы $r(\tau)$ удовлетворяют неравенствам

$$|r_{ij}(\tau)| < K\gamma^{|\tau|}, \quad 1 \leq i, j \leq n, \quad \tau = 0, \pm 1, \pm 2, \dots \quad (6.35)$$

КМ $r(\tau)$ может быть также определена из уравнения Юла — Уолкера

$$r(\tau) - \sum_{i=1}^p A_i r(\tau - i) = \sum_{i=\tau}^q B_i Q \Psi_{i-\tau}^*, \quad \tau = 0, 1, 2, \dots \quad (6.36)$$

Первые $(p + 1)$ уравнений (6.36) могут быть разрешены относительно матриц $r(0), r(1), \dots, r(p)$, если использовать тот факт, что $r(-\tau) = r^*(\tau)$. КМ $r(\tau)$ для $\tau > p$ получаются рекуррентно из (6.36).

Производящая функция КМ, представленная

$$G(z) = \sum_{\tau=-\infty}^{+\infty} r(\tau) z^\tau,$$

выражается в виде

$$G(z) = \Psi(z)Q\Psi^*(z^{-1}) = A^{-1}(z)B(z)QB^*(z^{-1})(A^{-1}(z^{-1}))^*. \quad (6.37)$$

Таким образом, если набор матриц $\{A_j\}$ и $\{B_j\}$ известен, то корреляционные характеристики процесса (6.28) могут быть найдены из (6.32), (6.34) или (6.36). В том случае, когда $\{A_i\}$, $\{B_j\}$ неизвестны, необходимо строить процедуру их оценивания.

Рассмотрим случай гауссовского процесса АРСС (p, q) . При этом для оценивания параметров процесса может быть использован ММП. Чтобы построить гауссовскую функцию правдоподобия (ГФП) для выборки $\{x_t, 1 \leq t \leq N\}$, составим вектор с nN компонентами $\mathbf{x} = (x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{N1}, \dots, x_{Nn})^* = (x_1^*, x_2^*, \dots, x_N^*)^*$. Матрица ковариации такого вектора имеет размер $(nN \times nN)$ и может быть представлена в блочной форме:

$$R = \mathbf{M}\{\mathbf{x}\mathbf{x}^*\} = \begin{pmatrix} r(0) & r(1) & \dots & r(N-1) \\ r^*(1) & r(0) & \dots & r(N-2) \\ \dots & \dots & \dots & \dots \\ r^*(N-1) & r^*(N-2) & \dots & r(0) \end{pmatrix}. \quad (6.38)$$

В (6.38) блоками матрицы R являются матрицы $r(\tau)$, $0 \leq \tau \leq N-1$, определенные в (6.34).

ГФП с учетом введенных обозначений имеет вид

$$L(R) = (2\pi)^{-nN/2} (\det R)^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^* R^{-1} \mathbf{x} \right\}. \quad (6.39)$$

Так как матрица R — симметрическая, она задается $nN(nN+1)/2$ элементами. Для оценивания этих параметров необходимо найти такую совокупность элементов, которая бы доставляла $L(R)$ максимум. Эта задача чрезвычайно сложная в вычислительном отношении и вряд ли удобная в смысле практического воплощения. Вместе с тем элементы матрицы R , точнее ее блоки, находятся по формулам (6.32) — (6.34), выражаются через $\{A_i\}$, $\{B_j\}$, Q и содержат в общей сложности около $n^2(\frac{1}{2} + p + q)$ независимых элементов, поэтому вычислить значения ГФП (6.39) можно, если известны $n^2(\frac{1}{2} + p + q)$ чисел, при помощи (6.32) — (6.34) и (6.38). Поиск максимума ГФП (6.39) в пространстве этих переменных — задача менее трудная, чем указанная выше, так как N обычно существенно больше, чем n . Тем не менее она остается еще достаточно сложной. Более удобным является использование при построении ГФП оценок предсказания, о которых речь пойдет в § 6.2, где будет обсуждена и проблема оценивания параметров.

В случае, когда процесс $\{x_t\}$ не является гауссовским, получать оценки неизвестных корреляционных характеристик или параметров разностного уравнения (6.28) можно точно таким же путем, максимизируя ГФП (6.39). Но только в этом случае полученные оценки не будут оценками ММП, а будут М-оценками неизвестных параметров, идея построения которых рассматривалась в § 2.4.

Лабораторная работа 22. Оценивание числовых характеристик многомерных временных рядов

Цель работы. Освоить технику статистической обработки временных рядов. Провести сравнительный анализ различных подходов к построению многомерных доверительных областей.

Порядок выполнения работы

Построить модель стационарного многомерного временного ряда с $n > 2$ на основе разложения (6.6). Для этого задать конечную последовательность матриц $\{C_j, \alpha \leq j \leq \beta\}$, такую, что $\alpha < 0 < \beta$; задать положительно определенную матрицу Q , у которой отдельные, но не все, недиагональные элементы были бы нулевыми; задать вектор \mathbf{m} математических ожиданий компонент m_i ряда $\{x_t\}$. Провести имитацию этого многомерного временного ряда и получить его реализацию $\{\mathbf{x}_t, 1 \leq t \leq N\}$.

Задание 1. Основываясь на (6.8), (6.5), вычислить матрицы $r(\tau) = (r_{ij}(\tau))$, $(\rho_{ij}(\tau))$, а также величину $r_* = \sum_{i=1}^n \sum_{\tau=-\infty}^{+\infty} r_{ij}(\tau)$ и матрицу G_N по формуле (6.12). Поскольку матрицы $C_j = 0, j \notin [\alpha, \beta]$, неограниченные суммы в формулах для r_* и G_N окажутся ограниченными.

Задание 2. Построить оценки $\hat{\mathbf{m}}, \hat{r}(\tau)$ и $\hat{\rho}(\tau)$ по формулам (6.9), (6.19) и (6.20) соответственно. Выяснить, насколько сильно отличается от нуля $(\hat{\mathbf{m}} - \mathbf{m})^*(\hat{\mathbf{m}} - \mathbf{m})$ и от r_* величина $N(\hat{\mathbf{m}} - \mathbf{m})^*(\hat{\mathbf{m}} - \mathbf{m})$, что подтверждает свойства (6.10) и (6.11). Сравнить матрицы $(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})^*$ и G_N , проиллюстрировав свойство (6.12). Провести сравнительный анализ матриц $\hat{r}(\tau)$ и $r(\tau)$, а также $\hat{\rho}(\tau)$ и $\rho(\tau)$ для некоторых значений $\tau = 0, 1, 2, \dots$.

Задание 3. Для нескольких значений α построить доверительные области (6.13) и (6.18). Найти объемы этих областей и сравнить их. Объем доверительной области (6.18) вычисляется как объем n -мерного параллелепипеда

$$V_{(18)} = \left(2\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)^n \sqrt{\prod_{i=1}^n \hat{d}_i}.$$

Доверительная область (6.13) — это n -мерный эллипсоид. Его объем вычисляется по формуле

$$V_{13} = \frac{\pi^{n/2} |\det T|}{\Gamma\left(1 + \frac{n}{2}\right)} (\chi_{1-\alpha}^2(n))^{n/2} \sqrt{\prod_{i=1}^n \lambda_i},$$

где T — матрица, столбцами которой являются собственные векторы, соответствующие указанным собственным значениям; $\Gamma(\cdot)$ — гамма-функция; $\{\lambda_i\}$ — собственные значения матрицы G_N^{-1} . Иначе говоря, для вычисления объема n -мерного эллипсоида, задаваемого матрицей G_N^{-1} , необходимо диагонализировать эту матрицу, т.е. представить ее в виде $G_N^{-1} = T \text{diag}\{\lambda_i\} T^*$.

Задание 4. Выбрать две некоррелированные компоненты исследуемого временного ряда и проверить их на независимость, используя декорреляцию (6.24) и критерий (6.25). Прodelать то же самое для двух коррелированных компонент рассматриваемого временного ряда.

Задание 5. Задать две различные модели каузальных обратимых процессов АРСС (p, q) . Провести имитацию двухмерного временного ряда с компонентами,

определяемыми этими моделями, в двух вариантах: 1) белые шумы, порождающие обе компоненты, — независимые; 2) белые шумы, порождающие компоненты временного ряда, — зависимые (один способ введения зависимости: белый шум, порождающий вторую компоненту, это белый шум первой компоненты, сдвинутый на несколько тактов вперед или назад, т.е. $W_{t,2} = W_{t\pm\tau,1}$; другой способ введения зависимости: если \mathbf{W}_t — вектор белых шумов из первого варианта, то вектор белого шума во втором варианте $\tilde{\mathbf{W}}_t = A\mathbf{W}_t$, где A — матрица, отличающаяся от диагональной). В каждом варианте испытывать компоненты временного ряда на независимость с использованием декорреляции (6.24) и анализа остатков.

Задание 6. Задать модель многомерного процесса АРСС (p, q) , определив необходимые числовые величины для моделирования этого процесса в соответствии с (6.28), т.е. $p, q, \{A_i, 1 \leq i \leq p\}, \{B_j, 1 \leq j \leq q\}, Q$. Параметры должны быть выбраны так, чтобы обеспечить каузальность и обратимость процесса x_t . Использовать два метода для определения КМ процесса x_t , применяя формулы (6.32)—(6.34) в одном случае и уравнение (6.36) — в другом. Провести сравнительный анализ этих методов по сложности и точности.

Задание 7. Провести моделирование двухмерного гауссовского процесса АРСС (p, q) с небольшими значениями p и q . Получить КМ процесса выборочным методом и ММП. Провести сравнительный анализ этих методов по сложности и точности.

6.2. Прогнозирование многомерных временных рядов и оценка их параметров на основе предсказания

Пусть $\{\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tn})^*, t = 1, 2, \dots\}$ — n -мерный временной ряд со средним значением $\mathbf{M}\{\mathbf{x}_t\} = 0$ и функцией ковариации, задаваемой $(n \times n)$ -матрицей $R^0(i, j) = M\{\mathbf{x}_i \mathbf{x}_j^*\}$. Как и в скалярном случае, наилучшая линейная оценка (НЛО) предсказания $\hat{x}_{N+\tau}^{(\tau)}$ значения вектора $\mathbf{x}_{N+\tau}$ по множеству N наблюдений $\{\mathbf{x}_t, 1 \leq t \leq N\}$ — оценка, которой соответствует минимальное значение среднеквадратичного отклонения $\hat{\mathbf{x}}_{N+\tau}^{(\tau)}$ от $\mathbf{x}_{N+\tau}$ при условии, что $\hat{\mathbf{x}}_{N+\tau}^{(\tau)}$ линейно связана с наблюдениями $\{\mathbf{x}_t, 1 \leq t \leq N\}$. Обозначим НЛО предсказания на один шаг

$$\hat{\mathbf{x}}_{N+1} \equiv \hat{\mathbf{x}}_{N+1}^{(1)} = \sum_{j=1}^N C_{Nj} \mathbf{x}_{N+1-j}, \quad N = 1, 2, \dots, \quad (6.40)$$

где C_{Nj} — $(n \times n)$ -матрицы коэффициентов НЛО предсказания.

Для инноваций компонент НЛО предсказания характерно, что они ортогональны ко всем компонентам векторов наблюдаемой выборки, т.е. для всякого $k, 1 \leq k \leq n$, имеет место $\mathbf{M}\{(x_{N+1,k} - \hat{x}_{N+1,k})x_{N+1-i,j}\} = 0, 1 \leq j \leq n, 1 \leq i \leq N$, или в векторной форме $\mathbf{M}\{(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1})\mathbf{x}_{N+1-i}^*\} = 0, 1 \leq i \leq N$. Отсюда и из (6.40) для матриц коэффициентов НЛО предсказания получаются соотношения

$\sum_{j=1}^N C_{Nj} R^0(N+1-j, N+1-t) = R^0(N+1, N+1-t)$, $1 \leq t \leq N$. В случае, когда $\{x_t\}$ — стационарный СП с КМ, определяемой (6.3), это соотношение упрощается:

$$\sum_{j=1}^N C_{Nj} r(t-j) = r(t), \quad 1 \leq t \leq N. \quad (6.41)$$

Система N матричных уравнений (6.41) эквивалентна системе Nn^2 скалярных уравнений относительно элементов матричных коэффициентов НЛО предсказания.

Более популярным представлением НЛО предсказания по сравнению с (6.40) является представление \hat{x}_{N+1} через инновации, так как оно позволяет построить более удобную вычислительную процедуру определения коэффициентов. В этом случае вместо (6.40) имеем

$$\hat{x}_{N+1} = \sum_{j=1}^N \theta_{Nj} (\mathbf{x}_{N+1-j} - \hat{x}_{N+1-j}), \quad (6.42)$$

где $\{\theta_{Nj}, 1 \leq j \leq N\}$ — последовательность $(n \times n)$ -матриц, которые могут быть найдены рекуррентно при помощи следующей процедуры. Пусть МК ошибок предсказания обозначена через V_N :

$$V_N = \mathbf{M}\{(\mathbf{x}_{N+1} - \hat{x}_{N+1})(\mathbf{x}_{N+1} - \hat{x}_{N+1})^*\}. \quad (6.43)$$

Тогда для НЛО предсказания \hat{x}_{N+1} в виде (6.42) справедливы соотношения

$$\hat{x}_{N+1} = \begin{cases} 0, & N = 0, \\ \sum_{j=1}^N \theta_{Nj} (\mathbf{x}_{N+1-j} - \hat{x}_{N+1-j}), & N > 0, \end{cases} \quad (6.44)$$

где матричные коэффициенты θ_{Nj} вычисляются рекуррентно по формулам

$$\begin{aligned} V_0 &= R^0(1, 1); \\ \theta_{N, N-k} &= \left(R^0(N+1, N+1) - \sum_{j=0}^{k-1} \theta_{N, N-j} V_j \theta_{k, k-j}^* \right) V_k^{-1}, \quad 0 \leq k \leq N-1; \\ V_N &= R^0(N+1, N+1) - \sum_{j=0}^{N-1} \theta_{N, N-j} V_j \theta_{N, N-j}^*. \end{aligned} \quad (6.45)$$

Здесь $R^0(i, j) = \mathbf{M}\{\mathbf{x}_i \mathbf{x}_j^*\}$. Последовательность вычислений по рекуррентным соотношениям (6.45) следующая: V_0 ; θ_{11} , V_1 ; θ_{22} , θ_{21} , V_2 ; θ_{33} , θ_{32} , θ_{31} и т.д.

В вычислительном отношении использование НЛО предсказания в виде (6.44), (6.45) предпочтительнее, чем в виде (6.40), (6.41). При этом параллельно

находится МК ошибок предсказания V_N , характеризующая точность предсказания отдельных компонент.

Рассмотрим проблему предсказания многомерных случайных процессов АРСС (p, q) . Пусть теперь \mathbf{x}_t — n -мерный процесс, определяемый разностным уравнением (6.28).

Пусть

$$z_t = \begin{cases} \mathbf{x}_t, & 1 \leq t \leq \max(p, q), \\ \mathbf{x}_t - \sum_{j=1}^p A_j \mathbf{x}_{t-j}, & \max(p, q) < t. \end{cases} \quad (6.46)$$

Если, как и ранее, обозначить КМ стационарного процесса $\{x_t\}$ через $r(\tau)$ (см. (6.3)), то матричную АКФ процесса $\{z_t\}$ можно представить в виде

$$R_z^0(i, j) = \begin{cases} r(i-j), & 1 \leq i \leq j \leq l = \max(p, q), \\ r(i-j) - \sum_{k=1}^p A_k r(i+k-j), & 1 \leq i \leq l < j \leq 2l, \\ \sum_{k=0}^q B_k Q B_{k+j-i}^*, & l < i \leq j \leq i+q, \\ 0, & l < i, i+q < j, \\ (R_z^0(i, j))^*, & j < i. \end{cases} \quad (6.47)$$

Здесь матрицы $\{A_k\}$, $\{B_k\}$ и Q определены соотношением (6.28) и предполагается, что $B_k = 0$, если $k > q$. С процессом $\{z_t\}$ удобнее работать потому, что его АКФ $R_z^0(i, j) = 0$ для $|i-j| > q$, $i, j > l$, $l = \max(p, q)$.

Оценка предсказания на один шаг вычисляется по формулам:

$$\hat{\mathbf{x}}_{N+1} = \begin{cases} \sum_{j=1}^{N-1} \theta_{Nj} (\mathbf{x}_{N+1-j} - \hat{\mathbf{x}}_{N+1-j}), & 1 \leq N \leq l, \\ \sum_{i=1}^p A_i \mathbf{x}_{N+1-i} + \sum_{j=1}^q \theta_{Nj} (\mathbf{x}_{N+1-j} - \hat{\mathbf{x}}_{N+1-j}), & N > l; \end{cases} \quad (6.48)$$

$$\mathbf{M}\{(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1})(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1})^*\} = V_N. \quad (6.49)$$

Коэффициенты θ_{Nj} , $1 \leq j \leq N$, и V_N находятся по формулам (6.45), куда вместо $R^0(i, j)$ подставляются $R_z^0(i, j)$, вычисляемые по формулам (6.47). Как отмечалось в гл.4, в скалярном случае коэффициенты θ_{Nj} и V_N не зависели от дисперсии белого шума. В многомерном случае эти коэффициенты (уже матричные) зависят от вида МК белого шума Q . В случае, когда $\{\mathbf{x}_t\}$ является обратимым процессом, инновации $(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1})$ могут рассматриваться как аппроксимации W_{N+1} для $N \gg 1$ в том смысле, что при $N \rightarrow \infty$

$$\mathbf{M}\{(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1} - \mathbf{W}_{N+1})(\mathbf{x}_{N+1} - \hat{\mathbf{x}}_{N+1} - \mathbf{W}_{N+1})^*\} \rightarrow 0.$$

Отсюда следует, что при $N \rightarrow \infty$ $\theta_{Nj} \rightarrow B_j$, $1 \leq j \leq q$, и $V_N \rightarrow Q$.

Рассмотрим важный случай процесса АРСС $(1, 1)$. При этом разностное уравнение (6.28) упрощается:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{W}_t + B\mathbf{W}_{t-1}, \quad t = 1, 2, \dots, \quad (6.50)$$

$\det(I - Az) \neq 0$, $|z| \leq 1$. НЛО предсказания имеет вид

$$\hat{\mathbf{x}}_{N+1} = A\mathbf{x}_N + \theta_{N1}(\mathbf{x}_N - \hat{\mathbf{x}}_N), \quad N \geq 1. \quad (6.51)$$

Матричная корреляционная функция процесса (6.46) определяется соотношениями

$$R^0(i, j) = \begin{cases} r(0), & i, j = 1, \\ QB^*, & 1 \leq i, j = i + 1, \\ Q + BQB^*, & 1 < i = j, \\ 0, & 1 \leq i, j > i + 1, \\ (R^0(i, j))^*, & j < i. \end{cases} \quad (6.52)$$

Рекуррентные соотношения (6.45) упрощаются:

$$\begin{aligned} V_0 &= r(0), \\ \theta_{N1} &= BQV_{N-1}^{-1}, \quad N = 1, 2, \dots, \\ V_N &= Q + BQB^* - \theta_{N1}V_{N-1}\theta_{N1}^*. \end{aligned} \quad (6.53)$$

Для того чтобы воспользоваться рекуррентными формулами, нужно вначале получить $r(0)$. Из уравнений Юла — Уолкера (6.36) имеем

$$r(0) - Ar^*(1) = Q + BQ(A^* + B^*), \quad r(1) - Ar(0) = BQ.$$

Определив $r(1)$ из второго уравнения и подставляя его в первое, получим уравнение

$$r(0) - Ar(0)A^* = AQB^* + BQA^* + Q + BQB^*, \quad (6.54)$$

являющееся линейным уравнением относительно n^2 элементов матрицы $r(0)$. Решив (6.54) и затем последовательно разрешая (6.53), получим необходимые данные для построения НЛО предсказания в виде (6.51).

Предсказание многомерных процессов АРСС (p, q) на τ шагов вперед, $\tau > 1$, по выборке из N векторов производится также с использованием инноваций по формулам

$$\begin{aligned} \hat{\mathbf{x}}_{N+\tau}^{(\tau)} &= \sum_{j=\tau}^{N+\tau-1} \theta_{N+\tau-1,j}(\mathbf{x}_{N+\tau-j} - \hat{\mathbf{x}}_{N+\tau-j}^{(1)}), \quad 1 \leq \tau \leq l - N; \\ \hat{\mathbf{x}}_{N+\tau}^{(\tau)} &= \sum_{i=1}^p A_i \hat{\mathbf{x}}_{N+\tau-i}^{(\tau-i)} + \sum_{j=\tau}^q \theta_{N+\tau-1,j}(\mathbf{x}_{N+\tau-j} - \hat{\mathbf{x}}_{N+\tau-j}^{(1)}), \quad \tau > l - N, \end{aligned} \quad (6.55)$$

где $l = \max(p, q)$. Для $\tau > q$ второе слагаемое в последней формуле исчезает. НЛО предсказания $\hat{\mathbf{x}}_{N+\tau}^{(\tau)}$, $\tau = 1, 2, \dots$, определяются рекуррентно по формулам (6.55) для всякого фиксированного N . Понятно, что в большинстве практических случаев $N > l$, и поэтому для определения $\hat{\mathbf{x}}_{N+\tau}^{(\tau)}$ практически всегда используется только второе соотношение (6.55).

В частном случае процесса АРСС (1,1) имеем

$$\hat{\mathbf{x}}_{N+\tau}^{(\tau)} = A\hat{\mathbf{x}}_{N+\tau-1}^{(\tau-1)} = \dots = A^{\tau-1}\hat{\mathbf{x}}_{N+1}^{(1)}. \quad (6.56)$$

В более общем случае АРСС (p, q) удобно при фиксированном N воспользоваться обозначением $\hat{\mathbf{x}}_{N+\tau}^{(\tau)} = \mathbf{g}(\tau)$. Тогда из (6.55) для $\mathbf{g}(\tau)$ получаем векторное разностное уравнение

$$\mathbf{g}(\tau) - A_1\mathbf{g}(\tau-1) - \dots - A_p\mathbf{g}(\tau-p) = 0, \quad \tau > q, \quad (6.57)$$

с начальными условиями $\mathbf{g}(q-i) = \hat{\mathbf{x}}_{N+q-i}^{(q-i)}$, $0 \leq i \leq p-1$.

Таким образом, для этого случая проблема нахождения НЛО предсказания свелась к решению разностного уравнения (6.57) относительно $\mathbf{g}(\tau)$.

На основе оценок предсказания может быть построена процедура оценивания неизвестных параметров многомерного процесса АРСС (p, q) . Применяя многомерный инновационный алгоритм к процессу z_t , определяемому (6.46), можно убедиться, что $x_{t+1} - \hat{x}_{t+1} = z_{t+1} - \hat{z}_{t+1}$, $t = 1, 2, \dots$. Поэтому ГФП для реализации $\{x_t, 1 \leq t \leq N\}$ может быть записана в виде

$$L(A, B, Q) = (2\pi)^{-\frac{nN}{2}} \left(\prod_{j=1}^N \det V_{j-1} \right)^{-1/2} e^{-\frac{1}{2} \sum_{j=1}^N (\mathbf{x}_j - \hat{\mathbf{x}}_j)^* V_{j-1}^{-1} (\mathbf{x}_j - \hat{\mathbf{x}}_j)}, \quad (6.58)$$

где матрицы V_j и определяющие их матрицы θ_{Nj} , вычисляемые по формулам (6.45), связаны с матрицами $\{A_i\}$, $\{B_j\}$ и Q через соотношения (6.47). По сравнению с (6.39) такой вид ГФП является более предпочтительным, поскольку в (6.58) вычислительные операции производятся с векторами и матрицами размерности n , а не nN , как это предусматривалось при использовании (6.39). Таким образом, оценивание параметров процесса АРСС (p, q) сводится к задаче максимизации (6.58) в пространстве $(n^2(p+q) + n(n+1)/2)$ переменных, значения которых задают набор матриц $\{A_i\}$, $\{B_j\}$, Q . Такая максимизация может быть произведена только численными методами. При этом могут возникнуть существенные трудности, связанные с тем, что ГФП (6.58) имеет довольно сложную структуру с несколькими локальными минимумами. Кроме того, может оказаться, что (6.58) имеет несколько глобальных максимумов. Это соответствует тому, что наблюдаемый процесс порождается неидентифицируемой моделью АРСС. Для оценивания параметров в таких случаях большое значение имеет удачный выбор начального набора параметров, от которого отталкивается алгоритм максимизации. Опыт проведения вычислительных работ в области максимизации подобных функций говорит о том, что, прежде чем осуществлять многомерный процесс максимизации, полезно произвести анализ каждой компоненты в отдельности как скалярного временного ряда. Это соответствует тому, что в модели многомерного временного ряда (6.28) случайные векторы \mathbf{W}_t имеют некоррелированные компоненты. Полученные таким образом параметры обобщаются, по ним строятся оцениваемые матрицы $\{\hat{A}_i\}$, $\{\hat{B}_i\}$. Затем

максимизируется (6.58) по Q , что дает \hat{Q} . Таким образом, данный набор матриц $\{\hat{A}_i\}$, $\{\hat{B}_i\}$, \hat{Q} и принимается в качестве исходного набора параметров при максимизации ГФП (6.58) в пространстве полной размерности.

Проблема идентифицируемости АРСС (p, q) возникает только тогда, когда $p > 0$ и $q > 0$. Для каузальной АР (p) или обратимого СС (q) матрицы коэффициентов и МК белого шума Q определяются единственным образом. Для нахождения параметров АР (p) модели по заданной выборке наблюдений можно построить достаточно простую вычислительную процедуру, основанную на уравнениях Юла — Уолкера (6.36). Для АР (p) они превращаются в уравнения

$$r(0) - \sum_{i=1}^p A_i r(-i) = Q; \quad r(\tau) = \sum_{j=1}^p A_j r(\tau - j), \quad 1 \leq \tau \leq p. \quad (6.59)$$

Подставляя в (6.59) выборочные значения матриц $\hat{r}(\tau)$, подсчитанные по формуле (6.19), получаем систему линейных уравнений относительно элементов матриц $\{A_i\}$ и Q .

Если процесс $\{x_t\}$ является гауссовским, то, максимизируя (6.58), находим оценки ММП неизвестных параметров. Если же $\{x_t\}$ не является гауссовским, то получаемые описанным способом оценки можно рассматривать как многомерные М-оценки с функцией критерия, заданной в виде ГФП.

Оценка порядков p и q процесса АРСС (p, q) может быть произведена подобно тому, как это было описано в § 4.3. Обычно используется информационный критерий Акаике, статистика которого в данном случае вычисляется в виде

$$\text{AIC} = -2 \ln L(A_1, \dots, A_p, B_1, \dots, B_q, Q) + 2k, \quad (6.60)$$

где первое слагаемое в первой части вычисляется с помощью (6.58), а k равно полному числу оцениваемых скалярных параметров (если все элементы оцениваемых матриц неизвестны, то $k = n^2(p + q) + n(n + 1)/2$).

Лабораторная работа 23. Прогнозирование многомерных временных рядов

Цель работы. Ознакомиться с методами прогнозирования многомерных временных рядов. Исследовать проблему увеличения сложности алгоритмов предсказания с увеличением размерности. Рассмотреть возможность оценивания неизвестных параметров моделей временного ряда с использованием оценок предсказания.

Порядок выполнения работы

Задать многомерную модель АРСС (p, q) . Удобно в качестве модели взять уже рассмотренную ранее модель при выполнении задания 6 лабораторной работы 22. Естественно также воспользоваться и выборкой наблюдений из этого задания.

Задание 1. Построить одношаговые НЛО предсказания тремя способами: с помощью формул (6.40), (6.41); (6.44), (6.45); (6.47)—(6.49). Представить процессы оценок предсказания и предсказываемого процесса на одном графике и провести сравнительный анализ по точности предсказания и по сложности вычислений. Рассмотреть скорость сходимости $V_N \rightarrow Q$ и $\theta_{Nj} \rightarrow B_j$ при увеличении N .

Задание 2. Задать модель процесса АРСС (1,1). Произвести имитацию процесса в соответствии с этой моделью. Воспользоваться формулами (6.50)—(6.54) для построения НЛО предсказания. По формулам (6.55) построить НЛО предсказания на τ шагов вперед. Исследовать точность предсказания от значения величины τ . Рассмотреть возможность использования соотношения (6.57) для построения НЛО предсказания при различных τ . Произвести сравнительный анализ точности и сложности по сравнению с предыдущим подходом.

Задание 3. Использовать реализацию процесса предыдущего задания для построения оценок матриц A, B, Q с помощью максимизации ГФП (6.58). Рассмотреть различные способы определения начальных значений оцениваемых параметров.

Задание 4. Рассмотреть применение информационного критерия Акаике для оценивания порядков процесса АРСС. Изучить влияние информации о параметрах на точность определения порядка модели.

6.3. Взаимные спектры и их оценивание

Пусть $\{x_t\}$ — многомерный ССП с матричной автоковариационной функцией $r(\tau)$, элементы которой являются абсолютно суммируемыми, $r(\tau) = (r_{lj}(\tau))$. Компоненты $\{x_{tj}, 1 \leq j \leq n\}$ имеют спектральные плотности. Определим

$$f_{lj}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} r_{lj}(\tau) e^{-i\tau\lambda}, \quad -\pi \leq \lambda \leq \pi. \quad (6.61)$$

Если $l = j$, то $f_{ll}(\lambda)$ представляет собой спектральную плотность l -й компоненты $\{x_{tl}\}$ временного ряда $\{x_t\}$. Если $l \neq j$, то $f_{lj}(\lambda)$ будем называть *взаимной спектральной плотностью* (кросс-спектром) компонент $\{x_{tl}\}$ и $\{x_{tj}\}$ временного ряда $\{x_t\}$. Составим из $f_{lj}(\lambda)$ матрицу $f(\lambda) = (f_{lj}(\lambda))$, которая называется *матрицей спектральных плотностей* (МСП) или *спектром многомерного временного ряда* $\{x_t\}$. Для вещественных ССП $\{x_t\}$ спектральные плотности $f_{ll}(\lambda)$ являются вещественными функциями. Однако $f_{lj}(\lambda)$, $l \neq j$, не обязательно должны быть вещественными. Обычно они являются комплексными функциями. С одной стороны, $f_{lj}(\lambda) = c_{lj}(\lambda) + iq_{lj}(\lambda)$, где $c_{lj}(\lambda)$ — коспектр или коспектральная плотность; $q_{lj}(\lambda)$ — квадратурный спектр. В соответствии с этим $f(\lambda) = c(\lambda) + iq(\lambda)$, где $c(\lambda)$ и $q(\lambda)$ — коспектральная и квадратурная МСП соответственно. С другой стороны, $f_{lj}(\lambda) = u_{lj}(\lambda) e^{iv_{lj}(\lambda)}$, где $u_{lj} = |f_{lj}(\lambda)|$ — амплитуда спектра или амплитудный спектр; $v_{lj}(\lambda)$ — фаза спектра или фазовый спектр.

Квадратичной функцией когерентности $K_{lj}^2(\lambda)$ называется отношение $K_{lj}^2(\lambda) = |f_{lj}(\lambda)|^2 / f_{ll}(\lambda)f_{jj}(\lambda)$, а $K_{lj}(\lambda)$ — коэффициентом когерентности

$$K_{lj}(\lambda) = u_{lj}(\lambda) / \sqrt{f_{ll}(\lambda)f_{jj}(\lambda)}. \quad (6.62)$$

Справедливы соотношения $f_{lj}(\lambda) = f_{jl}(\lambda)$, $0 \leq K_{lj}(\lambda) = K_{jl}(\lambda) \leq 1$ для всех $-\pi \leq \lambda \leq \pi$. Если $\{x_{t1}\}$ и $\{x_{t2}\}$ являются входом и выходом инвариантного по времени (стационарного) линейного фильтра, то коэффициент когерентности $K_{12}(\lambda)$ временных рядов $\{x_{t1}\}$ и $\{x_{t2}\}$ равен единице, $K_{12}(\lambda) = 1$, $-\pi \leq \lambda \leq \pi$. Если $\{y_{t1}\}$ и $\{y_{t2}\}$ являются выходами стационарных линейных фильтров, входы которых — $\{x_{t1}\}$ и $\{x_{t2}\}$, т.е.

$$y_{t1} = \sum_{j=-\infty}^{+\infty} \alpha_j x_{t-j,1}, \quad y_{t2} = \sum_{j=-\infty}^{+\infty} \beta_j x_{t-j,2},$$

то коэффициент когерентности выходных рядов $\{y_{t1}\}$ и $\{y_{t2}\}$ в точности равен коэффициенту когерентности входных рядов. Коэффициент когерентности некоррелированных временных рядов равен нулю.

Рассмотрим стационарный линейный фильтр с аддитивным шумом

$$x_{t,2} = \sum_{j=-\infty}^{+\infty} \alpha_j x_{t-j,1} + y_t. \quad (6.63)$$

Тогда имеют место соотношения

$$\begin{aligned} f_{22}(\lambda) &= f_y(\lambda) + |f_{22}(\lambda)|^2 / f_{11}(\lambda); \\ f_y(\lambda) &= (1 - K_{21}^2(\lambda)) f_{22}(\lambda); \\ \sigma_y^2 &= \int_{-\pi}^{\pi} (1 - K_{21}^2(\lambda)) f_{22}(\lambda) d\lambda. \end{aligned} \quad (6.64)$$

Рассмотрим задачу оценивания МСП с помощью многомерной периодограммы многомерного временного ряда $\{x_1, x_2, \dots, x_N\}$, где \mathbf{x}_t — n -мерный выборочный вектор. Определим ДПФ $J(\omega)$ многомерного ряда соотношением

$$J(\omega_j) = \frac{1}{\sqrt{N}} \sum_{t=1}^N x_t e^{-it\omega_j}, \quad (6.65)$$

где $\left\{ \omega_j = 2\pi j/N : -\left[\frac{N-1}{2} \right] \leq j \leq \left[\frac{N}{2} \right] \right\}$ — N частот Фурье. Периодограмма временного ряда $\{x_t\}$ задается на множестве этих частот как совокупность матриц

$$I_N(\omega_j) = J(\omega_j) J^*(\omega_j), \quad (6.66)$$

где * обозначает не только транспонирование, но и комплексную сопряженность элементов транспонированной матрицы. Этот смысл будет придаваться обозначению и в дальнейшем.

Расширенную периодограмму определяем на всем множестве частот $\omega \in [-\pi, \pi]$, полагая, что для $\omega \in [0, \pi]$ выполняется равенство $I_N(\omega) = I_N(\omega_k)$, если $|\omega - \omega_k| < \frac{\pi}{N}$ или $\omega = \omega_k + \frac{\pi}{N}$, а для $\omega \in [-\pi, 0)$ считается $I_N(\omega) = I_N^*(-\omega)$. Элементы матрицы $I_N(\omega)$ будем записывать $I_{lj}(\omega)$, опуская для краткости индекс N , $1 \leq l, j \leq n$. Причем $I_{ll}(\omega)$ представляет собой периодограмму l -й компоненты многомерного ряда $\{\mathbf{x}_t\}$. Функция $I_{lj}(\omega)$ называется *взаимной* или *кросс-периодограммой*. На множестве частот Фурье она имеет значения

$$I_{lj}(\omega_k) = \frac{1}{N} \left(\sum_{t=1}^N x_{tl} e^{-it\omega_k} \right) \left(\sum_{t=1}^N x_{tj} e^{-it\omega_k} \right). \quad (6.67)$$

Если ω_k — любая частота Фурье, не равная нулю, и \bar{x} — выборочное среднее ряда $\{x_1, x_2, \dots, x_N\}$, то

$$I_N(\omega_k) = \sum_{|\tau| < N} \hat{r}(\tau) e^{-it\omega_k}, \quad \omega_k \neq 0,$$

где $\hat{r}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} (\mathbf{x}_{t+\tau} - \mathbf{x})(\mathbf{x}_t - \mathbf{x})^*$, $\tau \geq 0$; $\hat{r}(\tau) = \hat{r}^*(-\tau)$, $\tau < 0$. Периодограмма на нулевой частоте $I_N(0) = N\bar{\mathbf{x}}\bar{\mathbf{x}}^*$. Пусть $\mathbf{m} = \mathbf{M}\{\mathbf{x}_t\}$. Тогда $\lim_{N \rightarrow \infty} (I_N(0) - N\mathbf{m}\mathbf{m}^*) = 2\pi f(0)$, $\lim_{N \rightarrow \infty} \mathbf{M}\{I_N(\omega)\} = 2\pi f(\omega)$, $\omega \neq 0$, где $f(\omega)$ — МСП процесса $\{x_t\}$.

Если процесс $\{\mathbf{x}_t\}$ допускает линейное разложение

$$\mathbf{x}_t = \sum_{j=-\infty}^{+\infty} c_j \mathbf{W}_{t-j}, \quad (6.68)$$

где \mathbf{W}_t — n -мерный случайный процесс НОР векторов, то $\mathbf{M}\{\mathbf{W}_t\} = 0$, $\mathbf{M}\{\mathbf{W}_t \mathbf{W}_t^*\} = Q$. Компоненты матриц c_j удовлетворяют условию $\sum_{j=-\infty}^{+\infty} |c_j(k, l)| \sqrt{|j|} < +\infty$, $1 \leq k, l \leq n$. Периодограмма $I_N(\lambda) = (I_{ij}(\lambda))$ процесса $\{x_t\}$, $-\pi \leq \lambda \leq \pi$. Тогда, если $0 < \lambda_1 < \lambda_2 < \dots < \lambda_k < \pi$, матрицы $I_N(\lambda_1), \dots, I_N(\lambda_k)$ при $N \rightarrow \infty$ сходятся к независимым случайным матрицам, j -я из которых распределена как $\mathbf{V}_k \mathbf{V}_k^*$, где \mathbf{V}_k — n -мерный комплексный вектор с нормальным распределением, $\mathbf{M}\{\mathbf{V}_k\} = 0$, $\mathbf{M}\{\mathbf{V}_k \mathbf{V}_k^*\} = 2\pi f(\lambda_k)$.

Если ω_l и ω_j — частоты Фурье из $[0, \pi]$, то при $N \rightarrow \infty$

$$\begin{aligned} \text{cov}(I_{\alpha\beta}(\omega_l), I_{pq}(\omega_k)) = \\ = \begin{cases} (2\pi)^2 [f_{\alpha p}(\omega_l) f_{q\beta}(\omega_l) + f_{\alpha q}(\omega_l) f_{\beta p}(\omega_l)], & \omega_k = \omega_l = 0 \text{ или } \pi, \\ (2\pi)^2 f_{\alpha p}(\omega_l) f_{q\beta}(\omega_l), & 0 < \omega_l = \omega_k < \pi, \\ 0, & \omega_l \neq \omega_k. \end{cases} \end{aligned} \quad (6.69)$$

Состоятельную оценку МСП процесса (6.68) можно получить при помощи сглаживания периодограммы. Пусть $g_N(k)$ — последовательность весовых коэффициентов, определенных в § 5.1, $g_N(k) = g_N(-k)$:

$$\sum_{|k| < p} g_N(k) = 1, \quad \sum_{|k| < p} g_N^2(k) \rightarrow 0 \text{ при } N \rightarrow \infty. \quad (6.70)$$

Определим оценку МСП $f(\lambda)$ равенством

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|k| < p} g_N(k) I_N(\lambda + \omega_k), \quad 0 < \lambda \leq \pi. \quad (6.71)$$

Число p , рассматриваемое при различных N , должно вести себя следующим образом: при $N \rightarrow \infty$ $p \rightarrow \infty$, но $p/N \rightarrow 0$,

$$\hat{f}(0) = \frac{1}{2\pi} \operatorname{Re} \left\{ g_N(0) I_N(\omega_1) + 2 \sum_{k=1}^p g_N(k) I_N(\omega_{k+1}) \right\}. \quad (6.72)$$

Имеют место следующие асимптотические свойства этих оценок. Если весовые коэффициенты для всех элементов $\hat{f}(\lambda)$ одинаковы, то

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbf{M}\{\hat{f}(\lambda)\} &= f(\lambda), \\ \lim_{N \rightarrow \infty} \frac{\operatorname{cov}(\hat{f}_{\alpha\beta}(\lambda), \hat{f}_{pq}(\lambda))}{\sum_{|k| \leq p} g_N^2(k)} &= \\ &= \begin{cases} f_{\alpha p}(\lambda) f_{q\beta}(\lambda) + f_{\alpha q}(\lambda) f_{\beta p}(\lambda), & \omega = \lambda = 0 \text{ или } \pi, \\ f_{\alpha p}(\lambda) f_{q\beta}(\lambda), & 0 < \omega = \lambda < \pi, \\ 0, & \omega \neq \pi. \end{cases} \end{aligned} \quad (6.73)$$

Заметим, что для комплексных векторов ковариация вычисляется по правилу $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{M}\{\mathbf{X}\bar{\mathbf{Y}}\} - \mathbf{M}\{\mathbf{X}\}\mathbf{M}\{\bar{\mathbf{Y}}\}$. Коспектральная МСП $c(\lambda) = (f(\lambda) + f^*(\lambda))/2$ имеет состоятельную оценку вида $\hat{c}(\lambda) = (\hat{f}(\lambda) + \hat{f}^*(\lambda))/2$. Квадратурная МСП $q(\lambda) = (f(\lambda) - f^*(\lambda))/2$ имеет состоятельную оценку $\hat{q}(\lambda) = (\hat{f}(\lambda) - \hat{f}^*(\lambda))/2$. Предположение об одинаковых весовых коэффициентах $g_N(k)$ для всех элементов является несущественным, оно лишь упрощает формулировку асимптотических свойств. Оценка $\hat{f}(\lambda)$ асимптотически нормально распределена.

В качестве оценки амплитудного спектра $u_{lj}(\lambda) = |f_{lj}(\lambda)|$ выберем

$$\hat{u}_{lj}(\lambda) = \sqrt{\hat{c}_{lj}^2(\lambda) + \hat{q}_{lj}^2(\lambda)} \equiv h(\hat{c}_{lj}, \hat{q}_{lj}). \quad (6.74)$$

Если $u_{lj}(\lambda) > 0$, то при $N \rightarrow \infty$ СВ $\hat{u}_{lj}(\lambda)$ асимптотически нормальна с математическим ожиданием $\mathbf{M}\{\hat{u}_{lj}(\lambda)\} = u_{lj}(\lambda)$ и дисперсией

$$D\{\hat{u}_{lj}(\lambda)\} = \frac{1}{2} \sum_{|k| \leq p} g_N^2(k) \left(\left(\frac{\partial h}{\partial c} \right)^2 (f_u f_{jj} + c_{lj}^2 - q_{lj}^2) + \right.$$

$$+ \left(\frac{\partial h}{\partial \hat{q}} \right)^2 (f_u f_{jj} + q_{l_j}^2 - c_{l_j}^2) + 4 \left(\frac{\partial h}{\partial \hat{c}} \right) \left(\frac{\partial h}{\partial \hat{q}} \right) c_{l_j} q_{l_j}.$$

Дисперсию $\hat{u}_{l_j}(\lambda)$ можно выразить через коэффициент когерентности (асимптотически). Если $K_{l_j}(\lambda) > 0$, то

$$D\{\hat{u}_{l_j}(\lambda)\} = \frac{1}{2} u_{l_j}^2(\lambda) \left(\sum_{|k| \leq p} g_N^2(k) \right) \left(1 + \frac{1}{K_{l_j}^2(\lambda)} \right).$$

Как видно, для малых $K_{l_j}(\lambda)$ дисперсия $\hat{u}_{l_j}(\lambda)$ велика.

В качестве оценки фазового спектра $v_{l_j}(\lambda) = \arg f_{l_j}(\lambda)$ выберем

$$\hat{v}_{l_j}(\lambda) = \arg(\hat{c}_{l_j}(\lambda) + i\hat{q}_{l_j}(\lambda)) \in (-\pi, \pi]. \quad (6.75)$$

Если $K_{l_j}(\lambda) > 0$, то такая оценка при $N \rightarrow \infty$ асимптотически нормальна с параметрами $\mathbf{M}\{\hat{v}_{l_j}(\lambda)\} = v_{l_j}(\lambda)$, $D\{\hat{v}_{l_j}(\lambda)\} = \frac{1}{2} u_{l_j}^2(\lambda) \left(\sum_{|k| \leq p} g_N^2(k) \right) \left(\frac{1}{K_{l_j}^2(\lambda)} - 1 \right)$.

Если $K_{l_j}(\lambda) = 0$, то вектор $(\hat{c}_{l_j}(\lambda), \hat{q}_{l_j}(\lambda))$ распределен асимптотически нормально с нулевым средним и КМ

$$\frac{1}{2} \left(\sum_{|k| \leq p} g_N^2(k) \right) \begin{pmatrix} f_u(\lambda) f_{jj}(\lambda) & 0 \\ 0 & f_u(\lambda) f_{jj}(\lambda) \end{pmatrix}.$$

В качестве оценки квадратичной функции когерентности (КФК) $K_{l_j}^2(\lambda)$ выберем $|\hat{K}_{l_j}(\lambda)|^2$, где

$$|\hat{K}_{l_j}(\lambda)| = \sqrt{\hat{c}_{l_j}^2(\lambda) + \hat{q}_{l_j}^2(\lambda)} / \sqrt{\hat{f}_u(\lambda) \hat{f}_{jj}(\lambda)}. \quad (6.76)$$

Если $K_{l_j}(\lambda) > 0$, то $|\hat{K}_{l_j}(\lambda)|$ распределен асимптотически нормально с параметрами $\mathbf{M}\{|\hat{K}_{l_j}(\lambda)|\} = |K_{l_j}(\lambda)|$, $D\{|\hat{K}_{l_j}(\lambda)|\} = \frac{1}{2} \left(\sum_{|k| \leq p} g_N^2(k) \right) (1 - K_{l_j}^2(\lambda))^2$.

Используя асимптотическое распределение оцениваемых параметров, несложно построить доверительные интервалы для этих параметров.

Удобно пользоваться такими доверительными интервалами, ширина которых не зависит от λ . Для коэффициента когерентности такой интервал можно построить.

Возьмем вместо $|\hat{K}_{l_j}(\lambda)|$ его преобразование $\xi = \arctg |\hat{K}_{l_j}(\lambda)|$. Случайная величина ξ распределена асимптотически нормально с параметрами $\mathbf{M}\{\xi\} = \arctg |K_{l_j}(\lambda)|$, $D\{\xi\} = \frac{1}{2} \sum_{|k| \leq p} g_N^2(k)$. Поэтому асимптотический доверительный интервал для $\arctg K_{l_j}(\lambda)$ получается в виде

$$\left(\arctg |\hat{K}_{l_j}(\lambda)| \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{1}{\sqrt{2}} \left(\sum_{|k| \leq p} g_N^2(k) \right)^{\frac{1}{2}} (1 - K_{l_j}^2(\lambda))^{\frac{1}{2}} \right), \quad (6.77)$$

и ширина этого интервала не зависит от λ .

В случае, когда $g_N(k) = 1/(1 + 2p)$, $|k| \leq p$, гипотеза $K_{lj}(\lambda) = 0$ может быть использована против альтернативной гипотезы $K_{lj}(\lambda) > 0$ при помощи статистики

$$\xi = 2p|\hat{K}_{lj}(\lambda)|^2/(1 - |\hat{K}_{lj}(\lambda)|^2). \quad (6.78)$$

В условиях справедливости (6.69) $|\hat{K}_{lj}(\lambda)|^2$ будет распределен как квадрат множественного коэффициента корреляции, так, что ξ в (6.78) имеет F -распределение со степенями свободы 2 и $4p$ при справедливой гипотезе $K_{lj}(\lambda) = 0$. Поэтому эта гипотеза отвергается, если $\xi > F_{1-\alpha}(2, 4p)$, где $F_{1-\alpha}(2, 4p)$ — квантиль F -распределения со степенями свободы 2 и $4p$ на уровне $1 - \alpha$ (о свойствах этого квантиля см. задание 1 лабораторной работы 21).

Лабораторная работа 24. Оценивание взаимных спектров и коэффициентов когерентности

Цель работы. Ознакомиться со свойствами взаимных спектров многомерных временных рядов. Научиться строить оценки взаимных спектров и коэффициентов когерентности. Проанализировать их свойства.

Порядок выполнения работы

Задать модель многомерного временного ряда, например процесса АРСС (p, q) , уже исследованного в задании 6 лабораторной работы 22.

Задание 1. Вычислить МСП для выбранной модели многомерного временного ряда при помощи формулы (6.61). Вычислить все взаимные спектры: коспектр, квадратурный, амплитудный, фазовый, а также коэффициент когерентности.

Задание 2. Для выборки многомерного временного ряда построить периодограмму (6.66), (6.67). При построении периодограммы пользоваться алгоритмом БПФ.

Задание 3. Построить состоятельную оценку МСП по формулам (6.70)—(6.72). На базе этой оценки построить оценки для коспектра, квадратурного, амплитудного и фазового спектров.

Задание 4. Вычислить параметры — среднее и дисперсию — асимптотических распределений для оценок с учетом того, что асимптотически все оценки распределены по нормальному закону. Построить доверительные интервалы на выбранном уровне значимости, вычисления иллюстрировать графиками. Проанализировать полученные результаты.

Задание 5. Построить оценку коэффициента когерентности. Вычислить параметры асимптотического распределения и построить доверительный интервал на базе этих вычислений для $\arctg |K_{lj}(\lambda)|$. Проанализировать, какими доверительными интервалами удобнее пользоваться.

ОБРАБОТКА ИЗОБРАЖЕНИЙ

7.1. Основные понятия

Термин “обработка изображений” включает обычно распознавание образов, кодирование и непосредственно обработку двумерных изображений. В данной главе первые две задачи не рассматриваются.

Под обработкой изображений понимается выполнение различных операций над данными, которые носят принципиально двумерный характер. Обработку изображений можно подразделить на коррекцию геометрических искажений, улучшение визуального качества, восстановление и реконструкцию изображений.

При коррекции геометрических искажений в изображении применяются пространственные преобразования, с помощью которых устраняются геометрические искажения или обеспечивается возможно более точное совмещение изображений относительно друг друга.

Целью улучшения визуального качества является получение улучшенного (например, в результате подавления шума, оптимизации контраста, подчеркивания границ изображений) представления изображений, таких, что содержащаяся в них существенная информация искажена, но тем не менее может визуальным восприниматься и без обработки.

Восстановление изображений включает оценку параметров искажения и использование ее для коррекции исходных данных. Под реконструкцией изображений подразумевается извлечение деталей в сильно искаженных изображениях при наличии данных об искажениях.

Обозначим через $f(x, y)$ двумерное изображение, получаемое телевизионной камерой или другим устройством, дающим изображение. Здесь x и y — пространственные координаты, т.е. координаты плоскости изображения, а величина f в произвольной точке (x, y) пропорциональна яркости (интенсивности) изображения в этой точке.

Для получения удобной формы функции изображения $f(x, y)$ с точки зрения вычисления ее необходимо дискретизировать как на плоскости, так и по амплитуде (интенсивности). Дискретизацию по пространственным координатам (x, y) будем называть *дискретизацией изображения*, а амплитудную дискретизацию — *квантованием по интенсивности* или *квантованием по уровню серого*. Последний термин применяется для одноцветных изображений и подтверждает тот факт, что тон изображения меняется в серой гамме от черного к белому. Предположим, что непрерывное изображение дискретизировано равномерно на N рядов и M столбцов, причем каждая дискретная величина проквантована по интенсивности. Такая система, называемая *цифровым изображением*, может

быть представлена в виде

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, M-1) \\ f(1, 0) & f(1, 1) & \dots & f(1, M-1) \\ \dots & \dots & \dots & \dots \\ f(N-1, 0) & f(N-1, 1) & \dots & f(N-1, M-1) \end{bmatrix},$$

где x и y теперь дискретные переменные, т.е. $x = 0, 1, \dots, N-1$; $y = 0, 1, \dots, M-1$. Каждый элемент системы называется *элементом изображения*, или *элементом картинки*, или *пикселом*. Обычно на практике величины N и M , а также число уровней дискретной интенсивности каждого квантованного пиксела задают в виде целых степеней числа 2.

Рассмотрим несколько простых взаимосвязей между пикселями в цифровом изображении. При обсуждении отдельного пиксела будем использовать сокращенные обозначения p и q . Для подгруппы пикселов используется обозначение S .

Пиксел p с координатами (x, y) имеет четыре горизонтальных и вертикальных соседних пиксела с координатами $(x+1, y)$, $(x-1, y)$, $(x, y+1)$, $(x, y-1)$. Эта группа пикселов, называемая “четыре соседа p ”, обозначается $N_4(p)$. Четыре диагональных соседних пиксела p имеют координаты $(x+1, y+1)$, $(x+1, y-1)$, $(x-1, y+1)$, $(x-1, y-1)$ и обозначаются $N_D(p)$. Эти точки вместе с четырьмя указанными выше называются “восемь соседей p ” и обозначаются через $N_8(p)$.

Пусть V — ряд значений интенсивности пикселов, которые могут быть соединены. Например, если требуется связать пиксела с интенсивностью, равной 59, 60, 61, то $V = \{59, 60, 61\}$. Рассмотрим три типа связей.

1. *Четырехсвязный тип*. Два пиксела p и q со значениями интенсивностей из V являются четырехсвязными, если q относится к группе $N_4(p)$.

2. *Восьмисвязный тип*. Два пиксела p и q со значениями интенсивностей из V являются восьмисвязными, если q относится к группе $N_8(p)$.

3. *m -связный тип (смешанная связь)*. Два пиксела p и q со значениями интенсивностей из V являются m -связными, если q относится к группе $N_4(p)$ или q относится к группе $N_D(p)$ и множество $N_4(p) \cap N_4(q)$ — пустое.

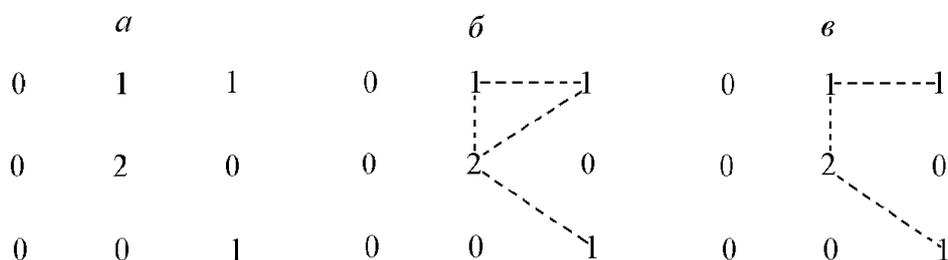


Рис. 7.1

Рассмотрим для примера систему пикселов (рис. 7.1, a). Предположим, что $V = \{1, 2\}$. Тогда связи восьми соседних пикселов с пикселом, имеющим значе-

ние 2, обозначим штриховыми линиями (рис. 7.1, б). Отметим неопределенность полученных связей ввиду их множественности. Эта неопределенность исчезает при использовании m -связной системы (рис. 7.1, в). Два изображения подгрупп S_1 и S_2 примыкают друг к другу, если несколько пикселей из S_1 примыкают к пикселям из S_2 .

Для пикселей p и q с координатами (x, y) и (s, t) соответственно определим расстояния:

- 1) Евклидово — $D(p, q) = [(x - s)^2 + (y - t)^2]^{1/2}$;
- 2) модульное — $D_4(p, q) = |x - s| + |y - t|$;
- 3) шахматное — $D_8(p, q) = \max(|x - s|, |y - t|)$.

Путь от пикселя p к пикселу q представляет собой последовательность определенных пикселей с координатами $(x_0, y_0), \dots, (x_n, y_n)$, где $(x_0, y_0) = (x, y)$; $(x_n, y_n) = (s, t)$; n — длина пути. Можно подразделить пути на 4-, 8-, m -мерные в зависимости от типа используемого примыкания. Когда имеют дело с m -связностью, величина расстояния (длина пути) между двумя пикселями зависит не только от координат, но и от значений пикселей вдоль пути, а также от их соседей. Рассмотрим для примера систему пикселей, в которой p, p_1, p_2, p_3, p_4 имеют значения 0 или 1:

$$\begin{array}{cc} p_3 & p_4 \\ p_1 & p_2 \\ p & \end{array}$$

Если предположить связность пикселей со значением 1 при p_1 и p_3 равными 0, то m -расстояние между p и p_4 равно 2. Если p_1 или p_3 равно 1, расстояние равно 3. Если же p_1 и p_3 равно 1, то расстояние будет равно 4.

Лабораторная работа 25. Моделирование цифровых изображений

Цель работы. Получить на ПЭВМ детерминированные и стохастические изображения, провести квантование отсчетов изображения, оценить расстояния между пикселями.

Порядок выполнения работы

Получение детерминированных (бинарных, двухцветных) изображений производится с помощью построения светлого (темного) объекта на темном (светлом) фоне. Получение стохастических (случайных) изображений производится с помощью искажения детерминированных изображений случайным шумом с последующим квантованием.

Пусть $g(x, y)$ ($x = 1, \dots, N$; $y = 1, \dots, N$) — заданное детерминированное изображение, для которого $g(x, y)$ принимает значение 0 для темного пикселя и 1 — для светлого. Будем полагать, что изображение $f(x, y)$ искажается гладким шумом, если $f(x, y) = g(x, y) + \xi(x, y)$, где совокупность $\{\xi(x, y)\}$ — НОР СВ с непрерывным распределением и импульсным шумом, если случайно выбранные точки объекта заменяются точками фона и наоборот.

Задание 1. Построить n -угольник, расположенный симметрично относительно центра экрана, со стороной длиной a . Параметры n, a принимают значения: $a = 4, 6, 10$ см, $n = 3, 4, 5$.

Задание 2. Построить эллипс (круг) с большой осью длиной a и малой — длиной b . Параметры a, b принимают следующие значения: $a = 4, 6, 10$ см, $b = 4, 5, 7$ см.

Задание 3. Построить произвольный сложный детерминированный объект (план дома, автомобиль и т.д.).

Задание 4. Получить стохастические изображения из заданий 1–3, искажив эти изображения шумом $\xi(x, y)$. В качестве распределения использовать равномерное распределение на интервале (a, b) : $a = 0, b = 1, 2, 3$ и нормальное распределение с нулевым средним и дисперсией $\sigma^2 = 0,1; 0,25; 0,5$.

Задание 5. Получить стохастические изображения из заданий 1–3, искажая их импульсным шумом. Зашумляемые точки расположить равномерно на прямоугольном изображении размером $N \times M$ пикселей. Количество зашумляемых точек взять равным $Q = 0,01P \cdot 0,1P \cdot 0,5P$; $P = NM$.

Задание 6. Получить квантованные изображения из заданий 4, 5. Результатом выполнения этого задания являются шесть цифровых изображений (при фиксированных значениях параметров), которые будем называть изображениями №1, 2, ..., 6. Для квантования применить равномерный код, где число уровней квантования светлоты выбирается из условия $L = 2^b$, где b — число двоичных разрядов (бит), отведенных для кодирования отсчетов. Правило квантования имеет следующий вид: отсчет $f(x, y)$ принадлежит уровню k ($\hat{f}(x, y) = k$), если $d_{k-1} < f(x, y) \leq d_k$, $d_k = f_{\min} + k(f_{\max} - f_{\min})/L$, $d_0 = f_{\min}$, где f_{\min}, f_{\max} — минимальное и максимальное значения изображения $f(x, y)$ соответственно. Параметр L принимает значения: $L = 2, 4, 6, 8, 16, 64, 256$. Найти m -расстояние между двумя любыми вершинами многоугольника для полученных квантовых изображений многоугольника. Положить $L = 2, 4$; $V = V_1, V_2$; $V_1 = 1$; $V_2 = \{3, 4\}$.

7.2. Геометрия изображений

Рассмотрим ряд геометрических преобразований изображений. Все преобразования записываются в трехмерной декартовой системе координат, в которой координаты точки обозначаются как (x, y, z) .

Смещение. Предположим, что требуется сместить точку с координатами (x, y, z) в новое место, используя перемещения (x_0, y_0, z_0) . Смещения выполняются по формулам $x^* = x + x_0$; $y^* = y + y_0$; $z^* = z + z_0$, где (x^*, y^*, z^*) — координаты новой точки. Эти выражения могут быть записаны в матричной форме

$$\begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix},$$

или

$$\begin{pmatrix} x^* \\ y^* \\ z^* \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix},$$

или $\mathbf{V}^* = T\mathbf{V}$, где \mathbf{V} — вектор, содержащий однородные начальные координаты; \mathbf{V}^* — конечные; T — матрица преобразования.

Изменение масштаба. Масштабирование на коэффициенты S_x, S_y, S_z по осям X, Y, Z производится с помощью матрицы

$$S = \begin{pmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

следующим образом: $\mathbf{V}^* = S\mathbf{V}$.

Вращение. Вращение точки относительно координатных осей X на угол α , Y на угол β , Z на угол θ выполняется соответственно с помощью матриц

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, R_\beta = \begin{pmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$R_\theta = \begin{pmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Взаимные и обратные преобразования. Реализация нескольких преобразований может быть представлена одной матрицей преобразования размерностью 4×4 . Например, смещение, масштабирование и вращение точки \mathbf{V} относительно оси Z записывается в виде $\mathbf{V}^* = R_\theta[S(T\mathbf{V})] = A\mathbf{V}$. Необходимо отметить, что эти матрицы обычно не переставляются, поэтому важен порядок их применения.

Данные приемы распространяются для одновременных преобразований группы из m точек с помощью одного преобразования. Пусть $\mathbf{v}_1, \dots, \mathbf{v}_m$ представляют координаты m точек. Если сформировать матрицу V размерностью $4 \times m$, столбцами которой являются векторы $\mathbf{v}_1, \dots, \mathbf{v}_m$, то одновременное преобразование этих точек с помощью матрицы преобразования A записывается в виде $\mathbf{V}^* = A\mathbf{V}$, где i -й столбец \mathbf{v}_i^* матрицы \mathbf{V}^* содержит координаты преобразованной точки, соответствующей \mathbf{v}_i .

Оптические преобразования. Оптические преобразования проецируют точки трехмерного пространства на плоскость, т.е. определяют способ отображения пространственных объектов. На рис. 7.2 изображена схема формирования изображения.

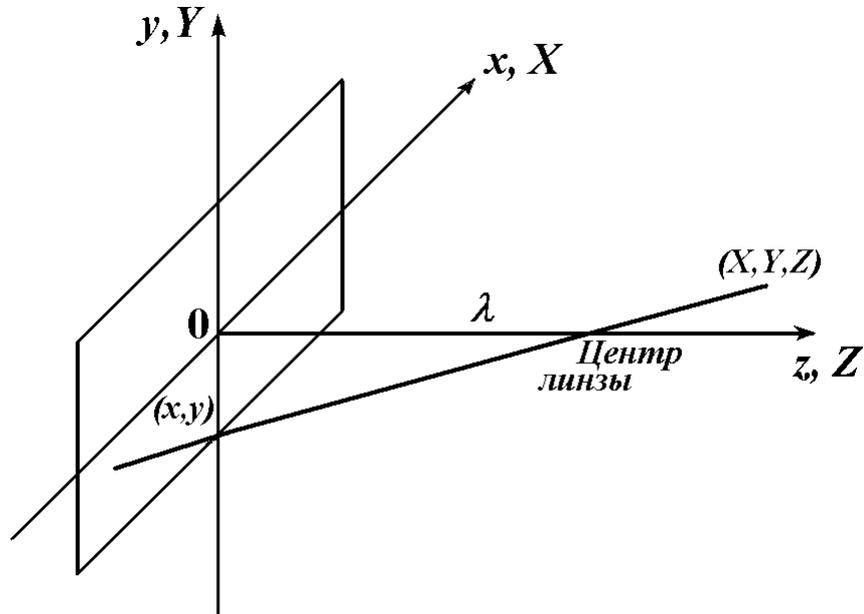


Рис. 7.2

Пусть система координат камеры (x, y, z) имеет плоскость изображения, совпадающую с плоскостью xy , а оптическая ось, установленная по центру линзы, направлена вдоль оси z . Таким образом, центр плоскости изображения является началом координат, а центр линзы имеет координаты $(0, 0, \lambda)$ (λ — фокусное расстояние линзы). Предполагается, что система координат камеры совпадает с декартовой системой координат. Спроецированные на плоскость изображения точки трехмерного пространства определяются из уравнений

$$x = \frac{\lambda X}{\lambda - Z}, \quad y = \frac{\lambda Y}{\lambda - Z}.$$

Лабораторная работа 26. Геометрические преобразования изображений

Цель работы. Получить элементарные преобразования изображения на плоскости и проекции трехмерных объектов на плоскость.

Порядок выполнения работы

Использовать шесть цифровых изображений из лабораторной работы 25.

Задание 1. Изображение №1 увеличить в 2 раза и повернуть на 45° .

Задание 2. Изображение №2 уменьшить в 1,5 раза и переместить вверх на 4 см.

Задание 3. Изображение №4 повернуть на 60° , переместить влево на 2 см и вниз на 4 см и уменьшить в 2 раза.

Задание 4. Спроецировать единичный куб на плоскость xy . Центр куба находится на расстоянии λ от плоскости xy и повернут на 45° относительно Z . Параметр λ принимает значения $\lambda = 2, 10, 20$.

Задание 5. Увеличить единичный куб в 3 раза, повернуть на 60° вокруг оси y , на 30° вокруг оси x и на 45° вокруг оси z и спроецировать его на плоскость xy . Расстояние центра куба λ от плоскости xy принимает значения $\lambda = 5, 20, 50$. Спроецированное изображение переместить на 2 см вниз и 3 см вправо.

Задание 6. Для любого сложного объекта (деталь дома и т.д.) сделать преобразования, аналогичные заданию 5.

7.3. Улучшение изображений

Процедура улучшения изображений сводится к выполнению комплекса операций с целью либо улучшения визуального восприятия изображения, либо преобразования его в форму, более удобную для визуального или машинного анализа.

Сглаживание. Данные операции используются для снижения шума и других помех, которые могут появляться на изображении в результате дискретизации, квантования, передачи или возмущений внешней среды. Для улучшения изображений применяются линейные преобразования, такие, как суперпозиция, свертка, дискретная линейная фильтрация. Дискретный оператор суперпозиции (линейной фильтрации) конечного массива отсчетов $f(x, y)$, $x, y = 0, 1, \dots, N-1$, с конечным массивом $h(i, j; p, q)$, $i, j = 0, 1, \dots, L-1$, определяется соотношением

$$Q(p, q) = \sum_{x=0}^p \sum_{y=0}^q f(x, y) h(p-x+1, q-y+1; p, q), \quad (7.1)$$

где $p, q = 0, 1, \dots, M-1$, а массивы f и h имеют нулевые значения вне областей изменения соответствующих индексов. Массив h называется *маской*, которая обычно представляет собой небольшую (например, размерность 3×3) матрицу. Элементы ее выбираются таким образом, чтобы обнаружить заданное свойство изображения. Если величины $\omega_1, \omega_2, \dots, \omega_9$ представляют собой коэффициенты маски пиксела (x, y) и его восьми соседей (табл. 7.1), операция линейной фильтрации (7.1) определяется соотношением

$$\begin{aligned} Q(x, y) = & \omega_1 f(x-1, y-1) + \omega_2 f(x-1, y) + \omega_3 f(x-1, y+1) + \\ & + \omega_4 f(x, y-1) + \omega_5 f(x, y) + \omega_6 f(x, y+1) + \\ & + \omega_7 f(x+1, y-1) + \omega_8 f(x+1, y) + \omega_9 f(x+1, y+1), \end{aligned} \quad (7.2)$$

где $x, y = 0, 1, \dots, N - 1$. Таким образом, перемещая маску от пиксела к пикселу на исходном изображении и производя операцию (7.2), получаем преобразованное изображение. Хотя иногда применяются и другие формы масок (например, круг), квадратные формы более предпочтительны из-за простоты их реализации. Кроме того, применяются маски размерностей больше, чем 3×3 . Для сглаживания изображений с целью подавления шумов используются маски трех типов:

$$1) H = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad 2) H = \frac{1}{10} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

$$3) H = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}, \quad \text{где } H = \begin{pmatrix} \omega_1 & \omega_2 & \omega_3 \\ \omega_4 & \omega_5 & \omega_6 \\ \omega_7 & \omega_8 & \omega_9 \end{pmatrix}.$$

Таблица 7.1

ω_1 ($x-1, y-1$)	ω_2 ($x-1, y$)	ω_3 ($x-1, y+1$)
ω_4 ($x, y-1$)	ω_5 (x, y)	ω_6 ($x, y+1$)
ω_7 ($x+1, y-1$)	ω_8 ($x+1, y$)	ω_9 ($x+1, y+1$)

Эти массивы нормированы для того, чтобы процедура подавления шума не вызывала смещения средней яркости обработанного изображения.

Первая маска соответствует также усреднению по окрестности $g(x, y) = \frac{1}{P} \sum_{(n,m) \in S} f(n, m)$, где P — число точек в окрестности; S — множество координат точек в окрестности точки (x, y) с учетом самой точки (x, y) .

Подавление шума на изображении можно осуществить с помощью медианной фильтрации. Медианный фильтр представляет собой скользящее окно (маску), охватывающее нечетное число элементов изображения. Центральный элемент заменяется медианой всех элементов изображения в окне. Медианой дискретной последовательности a_1, \dots, a_n для нечетного n является тот ее элемент, для которого существуют $(n - 1)/2$ элементов, меньших или равных ему по величине и больших или равных ему по величине.

Пусть окно размерностью 3×3 содержит элементы, равные 10, 20, 20, 20, 15, 20, 20, 25, 100, которые упорядочиваются по возрастанию: 10, 15, 20, 20, 20, 20, 20, 25, 100. Таким образом, медианой является значение 20, которым надо заменить значение центрального элемента окна. Медианный фильтр эффективно подавляет разрозненные импульсные помехи.

Если имеется несколько изображений одного объекта $g_i(x, y)$, то для подавления гладких шумов используется усреднение изображения. Пусть $g(x, y)$ —

исходное изображение, представляющее собой сумму неискаженного изображения $f(x, y)$ и шумового сигнала $n(x, y)$, $g(x, y) = f(x, y) + n(x, y)$. Предполагается, что шум не коррелирован и имеет нулевое среднее значение. Тогда для усредненного изображения

$$\bar{g}(x, y) = \frac{1}{k} \sum_{i=1}^k g_i(x, y)$$

получим $\mathbf{M}\{g(x, y)\} = f(x, y)$, $\sigma_{\bar{g}}^2(x, y) = \frac{1}{k} \sigma_n^2(x, y)$, где k — количество изображений; $\sigma_{\bar{g}}^2(x, y)$ и $\sigma_n^2(x, y)$ — дисперсии усредненного изображения и шума соответственно. При возрастании k функция $\bar{g}(x, y)$ приближается к неискаженному изображению.

Специальный тип сглаживания изображений применяется для бинарных изображений, т.е. принимающих значения 0 для темных точек и 1 — для светлых. Помехи в этом случае проявляются в виде таких эффектов, как наличие размытых границ, небольших окружностей, стертых углов и отдельных точек. Для сглаживания используется булева функция, вычисляемая в окрестности с центром в пикселе p . В процессе сглаживания, во-первых, заполняются небольшие (размером в один пиксел) пробелы на темных местах изображения; во-вторых, ликвидируются незначительные дефекты в виде трещин на прямоугольных сегментах; в-третьих, спрямляются небольшие выпуклости вдоль прямоугольных сегментов; в-четвертых, удаляются изолированные одиночные значения; в-пятых, восстанавливаются утраченные угловые точки. Используя табл. 7.2, два первых процесса сглаживания можно осуществлять с помощью булева выражения

$$B_1 = p + b \cdot g \cdot (d + e) + d \cdot e \cdot (b + g), \quad (7.3)$$

где точка и плюс обозначают соответственно логические операции И и ИЛИ.

Таблица 7.2

a	b	c
d	p	e
f	g	h

Тогда, если $B_1 = 1$, присваиваем пикселу p значение 1, в противном случае — 0. Уравнение (7.3) применяется одновременно ко всем пикселям.

Третий и четвертый результаты процесса сглаживания реализуются с помощью определения булева выражения

$$B_2 = p \cdot [(a + b + c) \cdot (e + g + h) + (b + c + e) \cdot (d + f + g)]. \quad (7.4)$$

Восстановление точек верхних правых углов, нижних правых, верхних левых и нижних левых производится с помощью функций:

$$B_3 = \bar{p} \cdot (d \cdot f \cdot g) \cdot (a + b + \bar{c} + e + h) + p; \quad (7.5)$$

$$B_4 = \bar{p} \cdot (a \cdot b \cdot d) \cdot (c + e + \bar{f} + g + h) + p; \quad (7.6)$$

$$B_5 = \bar{p} \cdot (e \cdot g \cdot h) \cdot (a + b + \bar{c} + d + f) + p; \quad (7.7)$$

$$B_6 = \bar{p} \cdot (b \cdot c \cdot e) \cdot (a + d + \bar{f} + g + h) + p. \quad (7.8)$$

Видоизменение гистограмм. Гистограмма распределения яркостей изображения естественного происхождения, подвергнутого линейному квантованию, обычно имеет перекосяк в сторону малых уровней и яркость большинства элементов изображения ниже средней. На темных участках подобных изображений детали часто оказываются неразличимыми. Метод видоизменения гистограммы предусматривает преобразование яркостей исходного изображения с тем, чтобы гистограмма распределения яркостей обработанного изображения приняла желаемую форму.

Процедуру видоизменения гистограммы можно рассматривать как монотонное поэлементное преобразование $s_k = T(r_k)$ входной интенсивности r_k в выходную интенсивность s_k .

Сначала рассмотрим задачу получения гистограммы обработанного изображения с равномерным распределением. Для этого используется преобразование

$$s_k = T(r_k) = \sum_{j=0}^k \frac{n_j}{n} = \sum_{j=0}^k p_r(r_j) \quad (7.9)$$

для $0 \leq r_k \leq 1$, $k = 0, 1, \dots, L-1$, где r_k — нормализованное значение интенсивности; n_k — число появлений данной интенсивности на изображении; n — общее число пикселей на изображении; $p_r(r_k)$ — оценка вероятности интенсивности r_k ; L — число дискретных уровней интенсивности.

Если требуется получить изображение с заданной гистограммой (плотностью вероятностей $p_z(z)$), то применяются следующие преобразования:

$$s_k = T(r_k) = \sum_{j=0}^k p_r(r_j);$$

$$G(z_j) = \sum_{j=0}^i p_z(z_j); \quad (7.10)$$

$$z_j = G^{-1}(s_i), \quad (7.11)$$

где $p_r(r_j)$ вычисляется из исходного изображения, а гистограмма $p_r(r_j)$ задается независимо от опыта. Два требуемых для определения гистограммы преобразования можно объединить в одно: $z = G^{-1}(s) = G^{-1}[T(r)]$, которое связывает интенсивности исходного изображения с интенсивностями выходного изображения.

Пороговое разделение. Один из путей отделения объектов от фона с сильно отличающимися интенсивностями — выбор порогового значения T , которое

разделяет эти группы интенсивностей. Тогда любая точка (x, y) , для которой $f(x, y) > T$, принадлежит объекту, а в противном случае — фону. В общем случае порог T является функцией $T = T[x, y, p(x, y), f(x, y)]$, где $f(x, y)$ — интенсивность точки (x, y) ; $p(x, y)$ — некоторое частное свойство этой точки, например, средняя интенсивность окрестности с центром в точке (x, y) . Пороговое изображение получается путем определения

$$g(x, y) = \begin{cases} 1, & f(x, y) > T, \\ 0, & f(x, y) \leq T. \end{cases}$$

Когда величина T зависит только от $f(x, y)$, порог называется *глобальным*. Если величина T зависит как от $f(x, y)$, так и от $p(x, y)$, то порог называется *локальным*. Если к тому же величина T зависит от пространственных координат x и y , порог называется *динамическим*. Глобальный порог можно оценить по гистограмме изображения (рис. 7.3).

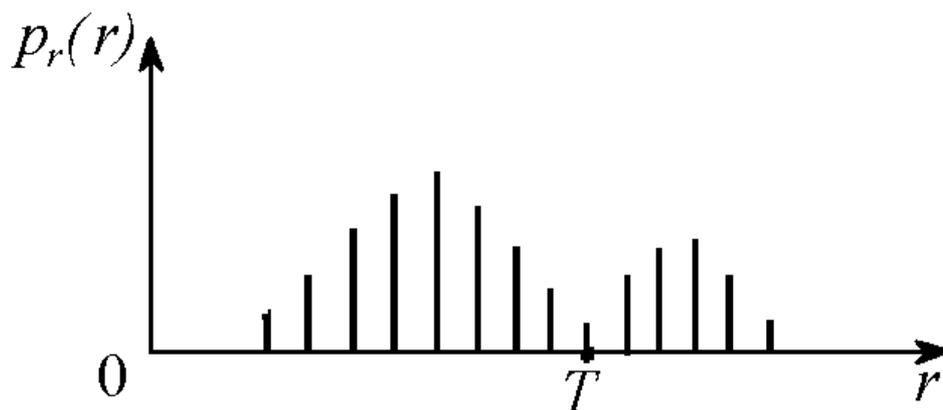


Рис. 7.3

Локальное улучшение качества. Приведенные методы преобразования гистограмм являются общими в смысле изменения интенсивности пикселей с помощью функции преобразования на всем поле изображения. Хотя методы дают общее улучшение качества, часто необходимо выделять детали на достаточно малых участках изображения. Решением данной проблемы является усовершенствование функций преобразования, основанных на распределении интенсивности или на других свойствах, в окрестности каждого пикселя заданного изображения.

В частности, рассмотренные выше гистограммные методы легко трансформируются для локального улучшения качества. Трансформация заключается в определении окрестности размером $n \times m$ и перемещении ее центра от пиксела к пикселу. В каждом положении вычисляется гистограмма точек в данной

окружности и определяется функция преобразования гистограммы. Эта функция служит для отображения пиксела, расположенного в центре окрестности. Затем центр участка размерностью $n \times m$ перемещается на соседний пиксел и процесс повторяется.

Лабораторная работа 27. Улучшение цифровых изображений

Цель работы. Преобразовать цифровые изображения на ПЭВМ для улучшения визуального восприятия.

Порядок выполнения работы

Использовать шесть цифровых изображений из лабораторной работы 25.

Задание 1. Осуществить операции подавления шума изображений №1 и 4, используя линейный и медианный фильтры. Исследовать вопрос о влиянии выбора размера и формы масок на качество обработанного изображения. Сделать вывод о применимости фильтров для подавления гладких и импульсных шумов.

Задание 2. Смоделировать K реализаций изображений №1 и 4 ($K = 10, 50, 100$). Получить усредненное изображение. Сделать вывод о применимости усреднения для подавления гладких и импульсных шумов.

Задание 3. Преобразовать яркость изображений №1, 2, 4 с помощью процедуры видоизменения гистограмм (формула (7.9)). Преобразованное изображение должно иметь равномерное на интервале (0,1) распределение. Вывести на экран дисплея гистограммы исходных и преобразованных изображений.

Задание 4. Преобразовать яркость изображений №1, 2, 4 с помощью процедуры видоизменения гистограмм (формулы (7.9)–(7.11)). Преобразованное изображение должно иметь гистограмму, изображенную на рис. 7.4.

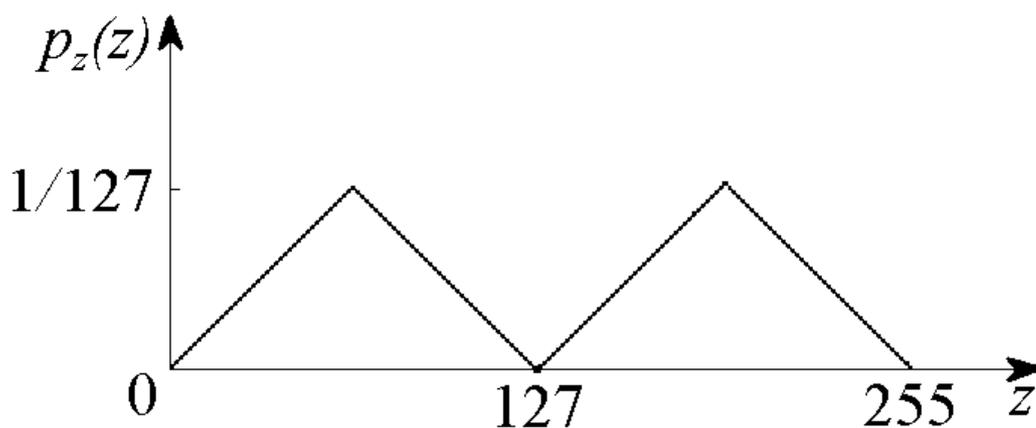


Рис. 7.4

Задание 5. Получить из изображений №1 бинарное изображение с помощью порогового разделения. В качестве порога использовать среднее значение яркости на всем изображении. Затем провести бинарное сглаживание по формулам (7.3)—(7.8).

Задание 6. Преобразовать яркость изображений №1, 2, 4 с помощью процедуры выравнивания гистограмм (см. задание 3) на локальных участках изображения. Сравнить полученное изображение с изображением из задания 3. Исследовать вопрос об оптимальном выборе размеров локальных областей. Вывести на экран дисплея гистограммы исходных и преобразованных изображений.

7.4. Выделение границ (кромки, края) изображений

Основным принципом большинства методов определения границ является вычисление частных производных. Поясним это положение с помощью рис. 7.5. Отметим, что участки кривой, соответствующей границе (переход от темной области к светлой), представляют собой наклонные, а не вертикальные линии, какие должны быть в случае изменения интенсивности. Это подтверждает тот факт, что кромки на цифровом изображении в результате дискретизации обычно слегка размыты.

В указанной модели границы, с одной стороны, первая производная всех участков кривой с постоянной интенсивностью равна нулю и является постоянной величиной на участках изменения интенсивности. С другой стороны, вторая производная равна нулю на всех участках, кроме начальных и конечных точек изменения интенсивности. Поэтому величина первой производной может быть использована для обнаружения наличия границы, а знак второй производной — для определения того, на темной (фон) или на светлой (объект) стороне границы располагается пиксел границы.

Градиент изображения $f(x, y)$ в точке (x, y) определяется как двухмерный вектор

$$\mathbf{G}[f(x, y)] = \begin{pmatrix} G_x \\ G_y \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}.$$

Модуль данного вектора задается соотношением

$$\mathbf{G}[f(x, y)] = [G_x^2 + G_y^2]^{1/2} = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2},$$

которое в дальнейшем будем называть *градиентом*. На практике, как правило, градиент аппроксимируется абсолютными значениями

$$\mathbf{G}[f(x, y)] \cong |G_x| + |G_y|. \quad (7.12)$$

Из уравнения (7.12) следует, что вычисление градиента основано на нахождении первых производных. Для цифрового изображения это достигается несколь-

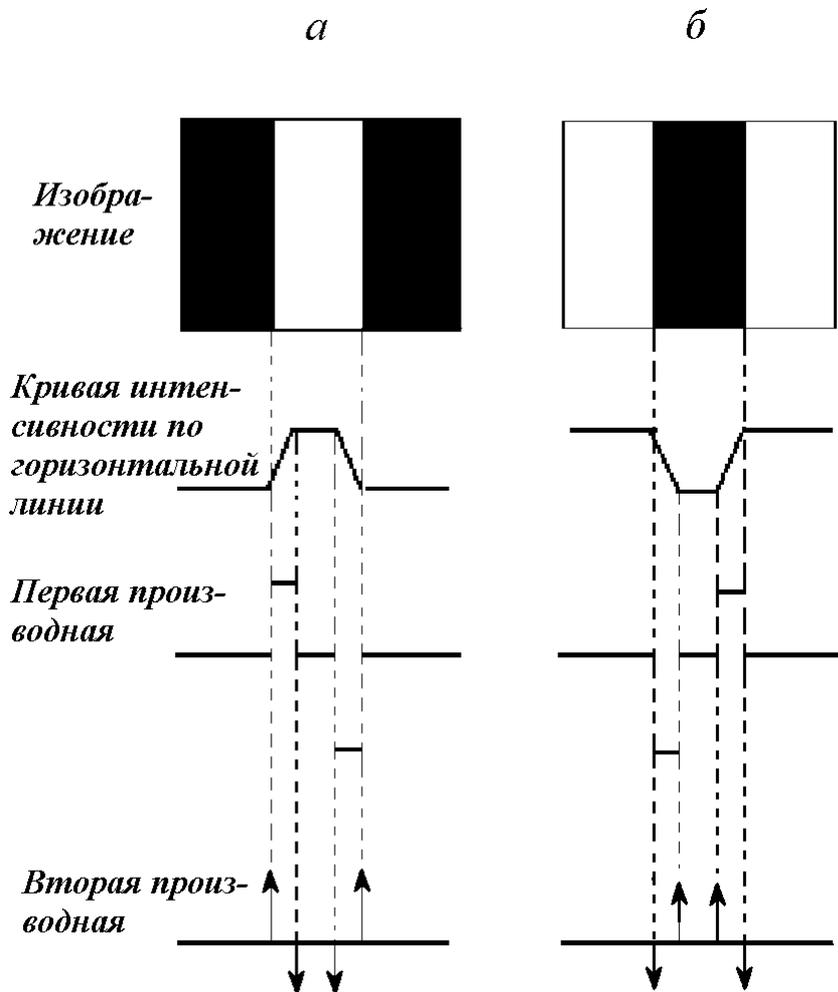


Рис. 7.5

кими путями. Один из подходов состоит в аппроксимации производных с помощью конечных разностей:

$$G_x = \frac{\partial f}{\partial x} = f(x, y) - f(x - 1, y); \quad (7.13)$$

$$G_y = \frac{\partial f}{\partial y} = f(x, y) - f(x, y - 1). \quad (7.14)$$

Несколько более сложный способ включает пиксели в окрестности размером 3×3 с центром в точке (x, y) :

$$G_x = [f(x + 1, y - 1) + 2f(x + 1, y) + f(x + 1, y + 1)] -$$

$$- [f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)]; \quad (7.15)$$

$$G_y = [f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)] - \\ - [f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)]. \quad (7.16)$$

Вычисление по формулам (7.15), (7.16) делает градиент менее чувствительным к помехам. В соответствии с изложенным в § 7.3 значения G_x и G_y можно определить с помощью масок H_x и H_y :

$$H_x = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \quad H_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}.$$

Эти маски обычно называют *операторами Собеля*. Перемещая маски по изображению $f(x, y)$, получают градиенты во всех его точках.

Существует ряд способов формирования выходного изображения $g(x, y)$, основанных на вычислении градиента. Простейшим способом является задание функции g в точке (x, y) значения, равного величине градиента входного изображения f в этой точке, т.е.

$$g(x, y) = \mathbf{G}[f(x, y)]. \quad (7.17)$$

Другой способ получения дискретного изображения — применение следующих соотношений:

$$g(x, y) = \begin{cases} 1, & \mathbf{G}[f(x, y)] > T, \\ 0, & \mathbf{G}[f(x, y)] \leq T, \end{cases} \quad (7.18)$$

где T — неотрицательная пороговая величина.

Данными способами задаются разрывы в интенсивности представления образа объекта. В идеальном случае этими методами определяются пикселы, лежащие на границе между объектом и фоном. На практике данным рядом пикселов редко полностью характеризуется граница из-за шума, разрывов на границе вследствие неравномерной освещенности и т.д. Требуются алгоритмы построения границ объектов из соответствующих последовательностей пикселов.

Локальный анализ. Все точки, являющиеся подобными в некоторой окрестности, соединяются, образуя границу из пикселов, обладающих рядом общих свойств. Для определения подобных точек необходимо найти градиент и направление градиента. Пиксел контура (x', y') подобен по величине градиента в определенной ранее окрестности (x, y) пикселу с координатами (x, y) , если справедливо неравенство

$$|\mathbf{G}[f(x, y)] - \mathbf{G}[f(x', y')]| \leq T, \quad (7.19)$$

где T — пороговое значение.

Направление градиента устанавливается по углу вектора градиента

$$\theta = \arctg \left[\frac{G_y}{G_x} \right]. \quad (7.20)$$

Угол пиксела контура с координатами (x', y') в некоторой окрестности (x, y) подобен углу пиксела с координатами (x, y) при выполнении неравенства

$$|\theta - \theta'| < A, \quad (7.21)$$

где A — пороговое значение угла.

На основе этих определений подобия соединим точку (x', y') в некоторой окрестности пиксела с точкой (x, y) , если выполняются соотношения (7.19), (7.21). Данный процесс применяется для каждой точки.

Глобальный анализ. Исследуем метод соединения граничных точек путем определения их расположения на кривой специального вида. Первоначально найдем подпоследовательность точек изображений, лежащих на прямых линиях.

Существует бесконечное число линий, проходящих через точку (x_i, y_i) , но все они удовлетворяют уравнению $y_i = ax_i + b$ при различных значениях a и b . Однако если записать это уравнение в виде $b = -x_i a + y_i$ и рассмотреть плоскость ab , то получается уравнение одной линии для фиксированной пары чисел (x_i, y_i) . Точка (x_j, y_j) также имеет в пространстве параметров связанную с ней линию, которая пересекает другую линию, связанную с точкой (x_i, y_i) в точке (a', b') , где значения a', b' — параметры линии, на которой расположены точки (x_i, y_i) , (x_j, y_j) в плоскости xy . Фактически все точки, расположенные на этой линии, в пространстве параметров будут иметь линии пересечения в точке (a', b') (рис. 7.6). Процедура определения точек, лежащих на прямых линиях, состоит в следующем. Находятся допустимые величины параметров линий (a_{\min}, a_{\max}) , (b_{\min}, b_{\max}) и k значений параметров в допустимых областях:

$$a_{\min}, \quad a_{\min} + h_1, \quad a_{\min} + 2h_1, \quad \dots, \quad a_{\max},$$

$$b_{\min}, \quad b_{\min} + h_2, \quad b_{\min} + 2h_2, \quad \dots, \quad b_{\max},$$

где $h_1 = \frac{a_{\max} - a_{\min}}{k}$, $h_2 = \frac{b_{\max} - b_{\min}}{k}$; матрица собирающих элементов

$$C = \begin{pmatrix} c_{11} & \dots & c_{1k} \\ \dots & \dots & \dots \\ c_{k1} & \dots & c_{kk} \end{pmatrix}.$$

Тогда для каждой точки (x_i, y_i) в плоскости образа полагаем параметр a равным каждому из допустимых значений на оси a и вычисляем соответствующее значение b , используя уравнение $b = -x_i a + y_i$. Полученное значение b затем округляется до ближайшего допустимого значения на оси b . Если выбор a_p приводит к вычислению b_q , то $c_{pq} = c_{pq} + 1$ (начальные значения матрицы C — нулевые). После завершения процедуры имеем c_{pq} координат точек, лежащих на линии $y = a_p x + b_q$.

Если линия принимает вертикальное положение, то параметры a и b стремятся к бесконечности. Для устранения этой трудности применяется нормальное

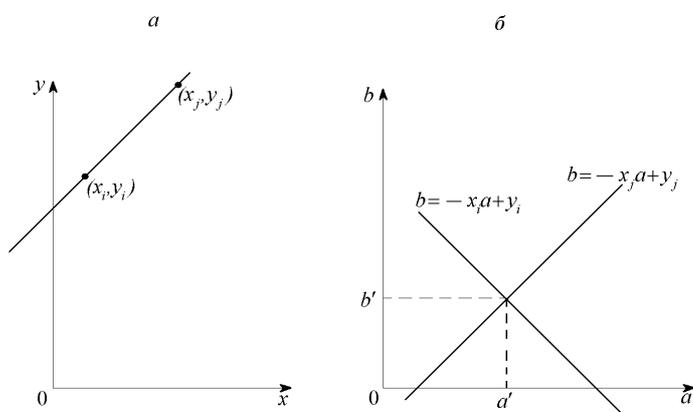


Рис. 7.6

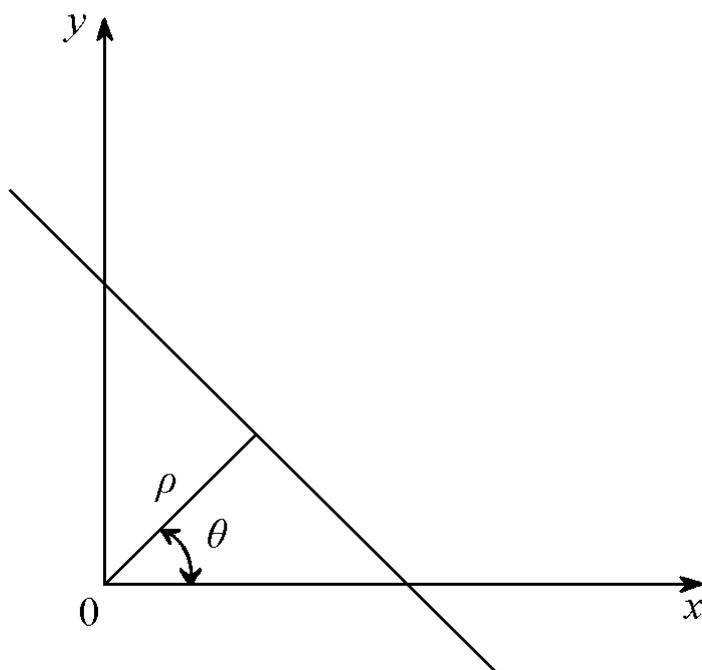


Рис. 7.7

представление линии в виде

$$x \cos \theta + y \sin \theta = \rho. \quad (7.22)$$

Смысл параметров поясняет рис. 7.7, на котором вместо параметров a и b используются параметры θ и ρ . Процедура выполняется по аналогии с выше-

изложенной.

Данная процедура применима к любой функции вида $g(\mathbf{x}, \mathbf{c}) = 0$, где \mathbf{x} — вектор координат; \mathbf{c} — вектор коэффициентов. Например, геометрическое место точек, лежащих на окружности

$$(x - c_1)^2 + (y - c_2)^2 = c_3^2, \quad (7.23)$$

может быть определено с помощью подхода, изложенного выше. Основное отличие заключается в том, что имеются три параметра c_1, c_2, c_3 .

Лабораторная работа 28. Выделение границ на изображении

Цель работы. Освоить методы выделения границ на идеальных и зашумленных изображениях, провести сравнительный анализ алгоритмов.

Порядок выполнения работы

Использовать шесть цифровых изображений из лабораторной работы 25.

Задание 1. Определить границы на идеальных (незашумленных) изображениях многоугольника, эллипса и сложного объекта и на зашумленных изображениях этих же объектов (изображения №1–3) с помощью алгоритмов (7.17), (7.18). Вычисление градиентов произвести по формуле (7.12), а производных — по формулам (7.13)—(7.16).

Задание 2. Определить границы на изображениях из задания 1 с помощью алгоритма (7.19)—(7.21). Для вычисления градиента использовать формулы (7.12), (7.15), (7.16). Исследовать качество алгоритма в зависимости от выбора параметров T, A и уровня шума.

Задание 3. Выделить границы на зашумленном изображении светлого квадрата на темном фоне, используя глобальный анализ при условии, что уравнения прямой задаются соотношением (7.22). Вывести на экран дисплея график преобразования Хоуга (график зависимости ρ от θ). Исследовать качество алгоритма в зависимости от числа уровней квантования параметров θ и ρ .

Задание 4. Определить границы на зашумленном изображении светлого круга на темном фоне, используя глобальный анализ. Для представления окружности применить уравнение (7.23). Исследовать качества алгоритма в зависимости от числа уровней квантования параметров c_1, c_2, c_3 и уровня шума.

7.5. Сегментация изображений

Сегментацией называется процесс *подразделения изображения на составляющие части или объекты*. Алгоритмы сегментации, как правило, основываются на двух принципах разрывности и подобии. В первом случае методы сегментации базируются на определении контуров, а во втором — на определении порогового уровня и устранении шумовых искажений.

Определение порогового уровня. Рассмотрим гистограмму изображения, состоящую из суммы значений функции плотности вероятности. В случае бимодальной гистограммы аппроксимирующая ее функция задается уравнением $p(z) = P_1 p_1(z) + P_2 p_2(z)$, где интенсивность z — случайная величина; P_1 и P_2 — априорные вероятности появления двух видов уравнения интенсивности на изображении (светлого и темного); $p_1(z)$ и $p_2(z)$ — плотности вероятностей. Для определения оптимального значения порога по интенсивности используется то значение $z = T$, при котором $P_1 p_1(T) = P_2 p_2(T)$. Если в качестве распределения $p_1(z), p_2(z)$ применяются гауссовские с математическими ожиданиями μ_1, μ_2 и дисперсиями σ_1^2, σ_2^2 соответственно, то оптимальный порог T находится из уравнения

$$AT^2 + BT + C = 0, \quad (7.24)$$

где

$$A = \sigma_1^2 - \sigma_2^2; \quad (7.25)$$

$$B = 2(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2); \quad (7.26)$$

$$C = \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 + 2\sigma_1^2 \sigma_2^2 \ln \frac{\sigma_2 P_1}{\sigma_1 P_2}. \quad (7.27)$$

Определение оптимального порога можно осуществить аналогичным образом в некоторых локальных подобластях всего изображения. В этом случае для каждой подобласти будет свой оптимальный порог.

Если изображение имеет многомодальную гистограмму, представляемую в виде суммы n функций плотности вероятности $p(z) = P_1 p_1(z) + \dots + P_n p_n(z)$, то задача оптимизации порогового уровня сводится к определению принадлежности данного пиксела к одному из n классов. Данный пиксел интенсивности z принадлежит k -му классу, если $P_k p_k(z) > P_j p_j(z), j = 1, \dots, n; j \neq k$. Оптимальный порог T_{kj} определяется из уравнения $P_k p_k(T_{kj}) = P_j p_j(T_{kj})$.

Возможность выбора “хорошего” порогового уровня существенно увеличивается, если пики гистограмм являются высокими, узкими, симметричными и разделены глубокими провалами. Одним из подходов к улучшению вида гистограмм может быть рассмотрение только тех пикселов, которые лежат на границе (или около нее) между объектами и фоном. Для этого можно использовать градиент (7.12) и лапласиан $L[f(x, y)] = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$. Для дискретных изображений оператор Лапласа определяется следующим образом:

$$L[f(x, y)] = [f(x + 1, y) + f(x - 1, y) + f(x, y + 1) + f(x, y - 1)] - 4f(x, y).$$

Эти два оператора можно использовать для формирования трехуровневого образа

$$s(x, y) = \begin{cases} 0, & G[f(x, y)] < T, \\ +, & G[f(x, y)] \geq T, L[f(x, y)] \geq 0, \\ -, & G[f(x, y)] \geq T, L[f(x, y)] < 0, \end{cases}$$

где символы 0, “+”, “-” представляют три различных уровня освещенности, а T — пороговый уровень. Предположим, что темный объект располагается на светлом фоне, тогда применение данного преобразования дает образ $S(x, y)$, в котором все пикселы, не лежащие на контуре, помечены 0, все пикселы на темной стороне контура помечены “+”, на светлой — “-”. Таким образом, гистограмма строится по значениям пикселов, помеченных “+” и “-”. Данная процедура может применяться для создания сегментированного, бинарного образа, в котором 1 соответствует искомым объектам, а 0 — фону.

Областно-ориентированная сегментация. Пусть R — область изображения. Рассмотрим сегментацию как процесс разбиения R на n подобластей R_1, R_2, \dots, R_n , так, что выполняются следующие условия:

- 1) $\bigcup_{i=1}^n R_i = R$;
- 2) R_i — связная область, $i = 1, 2, \dots, n$;
- 3) $R_i \cap R_j = \emptyset$ для всех i и j , $i \neq j$;
- 4) $P(R_i)$ есть истина для $i = 1, 2, \dots, n$;
- 5) $P(R_i \cup R_j)$ есть ложь для $i \neq j$,

где $P(R_i)$ — логический предикат, определенный на точках из множества R_i ; \emptyset — пустое множество. Четвертое условие определяет свойства, которым должны удовлетворять пикселы в сегментированной области. Например, $P(R_i) = \text{ИСТИНА}$, если все пикселы в R_i имеют одинаковую интенсивность.

Рассмотрим два алгоритма областно-ориентированной сегментации: расширение области за счет объединения пикселов; разбиение и объединение области.

Алгоритм 1. Расширение области сводится к процедуре группирования пикселов или подобластей в большие объединения. Простейшей из них является агрегирование пикселов. Процесс начинается с выбора множества узловых точек, с которых происходит расширение области в результате присоединения к узловым точкам соседних пикселов с похожими характеристиками (интенсивность, цвет). При использовании данного алгоритма сталкиваемся с двумя проблемами: выбором начальных узлов и определением подходящих свойств для включения точек в различные области в процессе расширения. Выбор множества, состоящего из одной или нескольких начальных точек, следует из постановки задачи. Например, выбор наиболее ярких пикселов является естественным начальным шагом в алгоритме процесса расширения области.

Алгоритм 2. Процедура расширения области имеет в качестве начальных условий множество узловых точек. Однако можно сначала разбить образ на ряд произвольных непересекающихся областей и затем объединять и (или) разбивать эти области с целью удовлетворения условий, сформулированных в начале данной главы. Итеративный алгоритм может быть изложен следующим образом.

Пусть R является полной областью образа, на которой определен предикат P . Один из способов состоит в успешном разбиении площади образа на все меньшие квадратные области, так, что для каждой области R_i , $P(R_i) = \text{ИСТИНА}$.

Процедура начинается с рассмотрения всей области R . Если $P(R) = \text{ЛОЖЬ}$, область разбивается на квадранты. Если для какого-либо квадранта P принимает значение ЛОЖЬ, этот квадрант разбивается на подквадранты и т.д. После операции разбиения производится операция объединения областей, имеющих одинаковые свойства по предикату P .

Таким образом, процедура разбиения и объединения имеет следующий вид:

- 1) разбиение области R_i , для которой $P(R_i) = \text{ЛОЖЬ}$, на четыре непересекающихся квадранта;
- 2) объединение соседних областей R_j и R_k , для которых $P(R_i \cup R_k) = \text{ИСТИНА}$;
- 3) выход на останов, когда дальнейшее объединение или разбиение невозможно.

Лабораторная работа 29. Сегментация изображений

Цель работы. Выделить объекты на изображении, используя алгоритмы сегментации, провести сравнительный анализ алгоритмов.

Порядок выполнения работы

Использовать все шесть цифровых изображений из лабораторной работы 25.

Задание 1. Отделить объект от фона на изображениях №1, 2, используя для этого пороговое решающее правило. В качестве порога T выбрать порог, вычисленный по уравнению (7.24). Предполагается, что изображения искажены гауссовским шумом с нулевым средним и дисперсией $\sigma^2 = 0,25$. Поэтому в формулах (7.25)–(7.27) следует положить $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 0,25$, $P_1 = P_2 = 0,5$.

Задание 2. Отделить объект от фона на изображениях №1, 2, используя пороговое решающее правило. Порог находится по гистограмме (значение T , разделяющее пики гистограмм), которая строится по пикселям, расположенным на границе между объектом и фоном (пиксели, принимающие значения “+” или “–” на трехуровневом образе $s(x, y)$).

Задание 3. Произвести сегментацию изображений №1, 2 с помощью алгоритма разбиения и объединения области. Присвоить значению предиката $P(R_i) = \text{ИСТИНА}$, если средний уровень яркости в этой области больше порога T . В качестве порога T использовать порог из задания 2.

7.6. Специальные модели изображений

Пространственную статистическую зависимость между пикселями в двумерном множестве соседей можно задать моделями случайного поля (МСП). Эти модели подразделяются на глобальные по признаку соответствия множества соседей, используемого в модели, целому изображению и на локальные по признаку соответствия его ограниченной (локальной) части.

Локальные МСП характеризуют статистическую зависимость интенсивности изображения в точке (m, n) от значений интенсивности в соседних точках, представляя интенсивность $x(m, n)$ как линейную комбинацию значений $\{x(m+k, n+l), (k, l) \in S\}$ и аддитивного шума, где S — множество соседей, не включающее точку (m, n) . Два неэквивалентных способа взаимодействия между отсчетами изображения приводят к двум классам локальных МСП: одновременных и условных АР моделей.

Одновременная АР модель является фундаментальной моделью класса локальных МСП. Если стационарное изображение $\{x(m, n), (m, n) \in \Omega\}$ определено на бесконечной решетке $\Omega = \{(m, n), -\infty \leq m, n \leq \infty\}$, то его одновременное АР (ОАР) представление имеет вид

$$x(m, n) = \sum_{(k, l) \in S} a_{kl} x(m+k, n+l) + \xi(m, n), \quad (7.28)$$

где $\xi(m, n)$ — двумерная последовательность независимых и одинаково распределенных значений шума с нулевым средним и дисперсией σ^2 ; $\mathbf{A} = \{a_{kl}; (k, l) \in S\}$ — вектор неизвестных коэффициентов модели. Множество соседей S должно быть симметричным.

При выборке модели из класса локальных МСП, соответствующей некоторому конечному изображению, появляются две основные группы задач. Первая из них связана с оценкой параметров модели при заданном множестве соседей, а вторая — с определением типа модели, наилучшим способом согласующейся с реальным изображением.

Реальное изображение можно рассматривать как конечный блок бесконечного изображения, заданного на бесконечной решетке Ω . Параметры ОАР модели на бесконечной решетке могут быть найдены с помощью МНК:

$$\hat{A} = \left[\sum_{(m, n)} Y(m, n) Y^*(m, n) \right]^{-1} \sum_{(m, n)} Y(m, n) x(m, n); \quad (7.29)$$

$$\sigma^2 = \frac{1}{M^2} \sum_{(m, n)} [x(m, n) - \hat{A}^* Y(m, n)]^2,$$

$$Y(m, n) = \text{col} [x(m+k, n+l); (k, l) \in S], \quad (7.30)$$

операция $\text{col}(\cdot)$ — операция развертки двумерного множества (матрицы) в вектор-столбец; M — размер квадрата изображения, заданного на решетке Ω . Оценки, полученные с помощью МНК, в общем случае являются асимптотически состоятельными и несмещенными для ОАР модели только с односторонним множеством соседей S .

Разрешение проблемы сложности при рассмотрении вопроса оценки параметров ОАР моделей осуществляется переходом к подклассу моделей на конечных решетках. Для таких моделей конечная решетка Ω_k делится на два

подмножества: внутреннее Ω_I и внешнее Ω_B , так, что $\Omega_B = \{(m, n) \mid (m, n) \in \Omega_k, (m+k, n+l) \in \Omega_k \text{ для } (k, l) \in S\}$, $\Omega_I = \Omega_k - \Omega_B$. Тогда для конечного изображения $\{x(m, n), (m, n) \in \Omega_k\}$ вводится тороидальная ОАР модель:

$$x(m, n) = \sum_{(k, l) \in S} a_{k, l} x(m+k, n+l) + \xi(m, n), \quad (m, n) \in \Omega_I; \quad (7.31)$$

$$x(m, n) = \sum_{(k, l) \in S} a_{k, l} x_1(m+k, n+l) + \xi(m, n), \quad (m, n) \in \Omega_B, \quad (7.32)$$

где

$$x_1(m+k, n+l) = \begin{cases} x(m+k, n+l), & \text{если } (m+k, n+l) \in \Omega_k, \\ x[(k \oplus m-1), (l \oplus n-1)], & \text{если } (m+k, n+l) \in \Omega_k; \end{cases}$$

знак \oplus означает суммирование по модулю $M+1$. Для гауссовского распределения отсчетов изображения оценки параметров ОАР модели на конечной решетке определяются с помощью итерационной процедуры метода максимального правдоподобия:

$$\hat{A}_{t+1} = \left(R - \frac{1}{\hat{\sigma}_t} T \right)^{-1} \left(V - \frac{1}{\hat{\sigma}_t} U \right), \quad t = 0, 1, 2, \dots; \quad (7.33)$$

$$\hat{\sigma}_t = \frac{1}{M^2} \sum_{(m, n) \in \Omega_k} [x(m, n) - \hat{A}^* Y(m, n)]^2, \quad t = 0, 1, 2, \dots, \quad (7.34)$$

где

$$T = \sum_{(m, n) \in \Omega_k} Y(m, n) Y^*(m, n);$$

$$U = \sum_{(m, n) \in \Omega_k} Y(m, n) x(m, n);$$

$$R = \sum_{(m, n) \in \Omega_k} (S_{mn} S_{mn}^* - C_{mn} C_{mn}^*);$$

$$V = \sum_{(m, n) \in \Omega_k} C_{mn};$$

$$C_{mn} = \text{col} \left\{ \cos \frac{2\pi}{M} [(m-1)k + (n-1)l], \quad (k, l) \in S \right\};$$

$$S_{mn} = \text{col} \left\{ \sin \frac{2\pi}{M} [(m-1)k + (n-1)l], \quad (k, l) \in S \right\}.$$

Начальное значение A_0 выбирается равным $T^{-1}U$, суммирование проводится по всем $(m, n) \in \Omega_k$.

Оценить коэффициенты ОАР модели можно также с использованием спектрального представления случайного поля

$$X(\lambda_1, \lambda_2) = \frac{1}{M} \sum_{(m,n) \in \Omega_k} x(m, n) e^{-i(m\lambda_1 + n\lambda_2)}. \quad (7.35)$$

Данный спектр определяется на области

$$\tilde{\Omega}_k = \left\{ (\lambda_1, \lambda_2) \mid \lambda_j = 2\pi \frac{m_j}{M}, \quad 0 \leq m_j \leq M, \quad j = 1, 2 \right\}.$$

Если Ω_k распространить на всю численную решетку, то спектральная плотность случайного поля задается соотношением

$$g(\lambda_1, \lambda_2) = M \{ \bar{X}(\lambda_1, \lambda_2) X(\lambda_1, \lambda_2) \}, \quad (7.36)$$

где черта сверху — символ комплексной сопряженности. Для спектральной плотности ОАР модели получаем формулу

$$g_1(\lambda_1, \lambda_2) = \frac{\sigma^2}{\left| 1 - \sum_{(k,l) \in S} a_{k,l} e^{i(\lambda_1 k + \lambda_2 l)} \right|^2}. \quad (7.37)$$

Спектральное представление распределено по нормальному закону в каждой точке (λ_1, λ_2) , имеет нулевое математическое ожидание, и для различных точек (λ_1, λ_2) спектральные представления статистически независимы. При этом, как следует из (7.36), дисперсией такого спектрального представления является спектральная плотность $g(\lambda_1, \lambda_2)$. Таким образом, плотность вероятностей спектрального представления случайного поля имеет вид

$$p(X(\lambda) \mid \lambda \in \tilde{\Omega}) = (2\pi)^{-M^2/2} \left(\prod_{\lambda \in \tilde{\Omega}} g(\lambda) \right)^{-1/2} \exp\left(-\frac{1}{2} \sum_{\lambda \in \tilde{\Omega}} |X(\lambda)|^2 / g(\lambda)\right),$$

где $\lambda = (\lambda_1, \lambda_2)$. Пусть $z(\lambda) = |X(\lambda)|^2$,

$$p(z \mid g_1) = (2\pi)^{-M^2/2} \prod_{\lambda \in \tilde{\Omega}} g_1(\lambda) \exp\left(-\frac{1}{2} \sum_{\lambda \in \tilde{\Omega}} (Z(\lambda) / g_1(\lambda))\right). \quad (7.38)$$

Функция $Z(\lambda) = Z(\lambda_1, \lambda_2)$ может быть вычислена по реализации случайного поля с помощью формулы (7.35) для каждой пары $(\lambda_1, \lambda_2) \in \tilde{\Omega}$. Параметры ОАР модели могут быть оценены с помощью ММП:

$$\hat{a}_{k,l} = \arg \max_{a_{k,l}} p(Z \mid g_1),$$

где $p(Z \mid g_1)$ определяются формулой (7.38). Введем для удобства функцию $F(\lambda, A)$ соотношением $g_1(\lambda) = \sigma^2 / (1 - F(\lambda, A))$. Упорядочим элементы множества S , присвоив им номера $S = \{s^{(1)}, s^{(2)}, \dots, s^{(N)}\}$, $s^{(j)} = (k, l)$, и введем в

рассмотрение векторы $\mathbf{A} = (a_{s^{(1)}}, \dots, a_{s^{(N)}})^*$, $\mathbf{B}(\lambda) = (\cos \lambda^* s^{(1)}, \dots, \cos \lambda^* s^{(N)})^*$, $\mathbf{C}(\lambda) = (\sin \lambda^* s^{(1)}, \dots, \sin \lambda^* s^{(N)})^*$, где $\lambda^* s = \lambda_1 k + \lambda_2 l$. Тогда функцию $F(\lambda, A)$ можно записать в форме $F(\lambda, A) = 2A^* B(\lambda) - A^* (B(\lambda) B^*(\lambda) + C(\lambda) C^*(\lambda)) A$. Логарифм функции правдоподобия принимает вид

$$\begin{aligned} \ln p(Z | A) &= -\frac{M^2}{2} \ln 2\pi - \frac{1}{2} \sum_{\lambda \in \tilde{\Omega}} (\ln g(\lambda) + Z(\lambda)/g(\lambda)) = \\ &= -\frac{M^2}{2} \ln 2\pi - \frac{1}{2} \sum_{\lambda \in \tilde{\Omega}} [\ln \sigma^2 - \ln(1 - F(\lambda, A)) + \frac{Z(\lambda)}{\sigma^2} (1 - F(\lambda, A))]. \end{aligned}$$

Таким образом, неизвестные параметры σ^2 и \mathbf{A} спектральной плотности $g(\lambda)$ по ММП должны находиться как решение уравнения

$$M^2 \ln \sigma^2 + \sum_{\lambda \in \tilde{\Omega}} \left[\frac{Z(\lambda)}{\sigma^2} (1 - F(\lambda, A)) - \ln(1 - F(\lambda, A)) \right] \rightarrow \min_{\sigma^2, A}. \quad (7.39)$$

По переменной σ^2 функция, стоящая в левой части (7.39), является гладкой и выпуклой книзу, поэтому оценка σ^2 вычисляется следующим образом:

$$\hat{\sigma}^2 = \frac{1}{M^2} \sum_{\lambda \in \tilde{\Omega}} Z(\lambda) (1 - F(\lambda, A)). \quad (7.40)$$

Учитывая (7.40) в (7.39), приходим к уравнению относительно вектора \mathbf{A}

$$\begin{aligned} M^2 + \sum_{\lambda \in \tilde{\Omega}} \left[\ln \left(\frac{1}{M^2} \sum_{\nu \in \tilde{\Omega}} Z(\nu) (1 - F(\nu, A)) \right) - \ln(1 - F(\lambda, A)) \right] = \\ = M^2 + \sum_{\lambda \in \tilde{\Omega}} \ln \left(\frac{\sum_{\nu \in \tilde{\Omega}} Z(\nu) (1 - F(\nu, A))}{M^2 (1 - F(\lambda, A))} \right) \rightarrow \min_A. \end{aligned} \quad (7.41)$$

Функция в левой части (7.41) является в общем случае многоэкстремальной, и поэтому при ее минимизации необходимо использовать глобальный алгоритм. Однако в качестве приближенного способа минимизации можно применять и локальные алгоритмы. Экстремальные значения параметров \mathbf{A} находятся из уравнения

$$\sum_{\lambda \in \tilde{\Omega}} \left[\frac{1}{1 - F(\lambda, A)} - \frac{M^2 Z(\lambda)}{\sum_{\nu \in \tilde{\Omega}} Z(\nu) (1 - F(\nu, A))} \right] \frac{\partial F(\lambda, A)}{\partial A} = 0. \quad (7.42)$$

Обозначим $B(\lambda) B^*(\lambda) + C(\lambda) C^*(\lambda) = H(\lambda)$, тогда (7.42) записывается в виде

$$\sum_{\lambda \in \tilde{\Omega}} \left(\frac{1}{1 - F(\lambda, A)} - \frac{M^2 Z(\lambda)}{\sum_{\nu \in \tilde{\Omega}} Z(\nu) (1 - F(\nu, A))} |B(\lambda) + H(\lambda) A| \right) = 0. \quad (7.43)$$

Приведем простую итерационную процедуру поиска параметра \mathbf{A} . Пусть $A(k)$ — оценка параметра \mathbf{A} на k -м шаге, тогда

$$A(k+1) = \sum_{\lambda \in \tilde{\Omega}} \left[\left(-B(\lambda) + \frac{M^2 Z(\lambda)}{\sum_{\nu \in \tilde{\Omega}} Z(\nu)(1 - F(\nu, A(k)))} - \frac{F(\lambda, A(k))}{1 - F(\lambda, A(k))} \right) (B(\lambda) + H(\lambda)A(k)) \right], \quad k = 0, 1, \dots,$$

где в качестве $A(0)$ берется любая точка, обеспечивающая стационарность ОАР модели, т.е. выполнение неравенства $\sum_{(k,l) \in S} |a_{k,l}(0)| < 1$.

При использовании подкласса условных АР моделей предполагается, что порождающий их шум является коррелированным. Условно-марковская модель (УМ) является одной из моделей данного подкласса. Изображение, удовлетворяющее УМ на бесконечной решетке, задается уравнением

$$x(m, n) = \sum_{(k,l) \in S} a_{k,l} x(m+k, n+l) + \nu(m, n), \quad (7.44)$$

где множество соседей симметрично, т.е. $a_{k,l} = a_{-k,-l}$, $(k, l) \in S$; $\nu(m, n)$ — стационарная гауссовская шумовая последовательность, удовлетворяющая условиям

$$\begin{aligned} \mathbf{M}\{\nu(m, n)\} &= 0, \\ \mathbf{M}\{\nu(m, n), \nu(k, l)\} &= \begin{cases} \rho, & (m, n) = (k, l), \\ -a_{m-k, n-l} \rho, & (m-k, n-l) \in S, \\ 0, & \text{в остальных случаях.} \end{cases} \end{aligned}$$

При задании УМ на конечной решетке используется, как для ОАР модели, тороидальное представление решетки. Для обеспечения стационарности модели должно выполняться условие $1 - 2A^* \Phi_{mn} > 0$, $(m, n) \in \Omega_k$, где $A = \text{col}[a_{k,l}, (k, l) \in S_1]$; $\Phi_{mn} = \text{col}\{\cos \frac{2\pi}{M}[(m-1)k + (n-1)l], (k, l) \in S_1\}$; множество S_1 является такой половиной множества соседей S , что $S = S_1 \cup \bar{S}_1$, $S_1 \cap \bar{S}_1 = \emptyset$, $\bar{S}_1 = \{(k, l) | (-k, -l) \in S_1\}$.

Оценка параметров УМ находится следующим образом:

$$\hat{A}_c = \left[\sum_{\Omega_0} Q(m, n) Q^*(m, n) \right]^{-1} \sum_{\Omega_0} Q(m, n) x(m, n), \quad (7.45)$$

где $Q(m, n) = \text{col}[x(m+k, n+l) + x(m-k, n-l), (k, l) \in S_1]$; Ω_0 — подмножество множества Ω_k . Данная оценка не является единственной, поскольку суммирование по другому подмножеству $\Omega_1 = \Omega_k - \Omega_0$ дает другую оценку, которая может отличаться от \hat{A}_c . МНК-оценка $\hat{A}_{\text{МНК}}$ модели (7.6) аналогична оценке (7.45), но суммирование производится по множеству Ω_I .

Решение задачи выбора множества соседей для локальных МСП базируется на критерии Байеса. Использование байесовского подхода приводит к следующему решению. Выбирается множество соседей S_{r^*} , если

$$r_* = \arg \min \{J_r\}, \quad (7.46)$$

где

$$J_r = \left\{ - \sum_{(m,n) \in \Omega} \ln [1 - 2\hat{A}_r^* C_{mn}(r) + \hat{A}_r^* Q_{mn}(r) \hat{A}_r + M^2 \ln \hat{\sigma}_r + m_r \ln (M^2)] \right\}; \quad (7.47)$$

$$Q(m, n) = S_{mn}(r) S_{mn}^*(r) + C_{mn}(r) C_{mn}^*(r),$$

$$C_{m,n}(r) = \text{col} \left\{ \cos \frac{2\pi}{M} [(m-1)k + (n-1)l], (k, l) \in S_r \right\},$$

$$S_{m,n}(r) = \text{col} \left\{ \sin \frac{2\pi}{M} [(m-1)k + (n-1)l], (k, l) \in S_r \right\},$$

m_r — количество точек на множестве соседей S_r .

Для УМ решающее правило выбора множества соседей также имеет вид (7.46), где

$$J_r = - \sum_{(m,n) \in \Omega} \ln [1 - 2\hat{A}_r^* \Phi_{mn}(r)] + M^2 \ln \rho_r + m_r \ln (M^2); \quad (7.48)$$

$$\Phi_{m,n}(r) = \text{col} \left\{ \cos \frac{2\pi}{M} [(m-1)k + (n-1)l], (k, l) \in S_{1r} \right\},$$

S_{1r} — такая половина множества соседей S_r , что $S_r = S_{1r} \cup \bar{S}_{1r}$, $S_{1r} \cap \bar{S}_{1r} = \emptyset$, $\bar{S}_{1r} = \{(k, l) \mid (-k, -l) \in S_{1r}\}$.

Выбор модели из класса локальных МСП, наилучшим образом соответствующих данному изображению, проводится на основе рассмотренных выше критериев при условии, что множество соседей фиксировано. Процедура выбора модели состоит в вычислении конкретного значения J_r для различных моделей, описывающих данное локальное изображение, и принятии той из них, которой соответствует наименьшее значение J_r .

Лабораторная работа 30. Оценивание параметров МСП, порождаемых уравнениями АР

Цель работы. Смоделировать изображения, удовлетворяющие уравнениям АР, оценить параметры авторегрессионного уравнения.

Порядок выполнения работы

Получить модель изображения №1, имеющего ОАР представление (7.28). Множество $S = \{(-1, 0), (0, -1), (1, 0), (0, 1)\}$. Числовые значения для коэффициентов a_{kl} могут быть взяты следующими:

- 1) $a_{10} = 0,12, a_{-10} = 0,12, a_{01} = -0,14, a_{0-1} = -0,14;$
- 2) $a_{10} = 0,24, a_{-10} = 0,24, a_{01} = 0,1, a_{0-1} = 0,1,$

где $\xi(m, n)$ — множество независимых гауссовских случайных величин с нулевым средним и дисперсией σ^2 . Дисперсия может принимать значения $\sigma^2 = 0,5; 1,0$. Значения для множества узлов решетки $(m, n) \in \Omega_I$ находятся по формуле

$$x_t(m, n) = \sum_{(k,l) \in S} a_{kl} x_{t-1}(m+k, n+l) + \xi_t(m, n), \quad t = 1, \dots, t_k,$$

$$x_0(m, n) = \xi_0(m, n), \quad (m, n) \in \Omega.$$

Параметр t_k может принимать значения $t_k = 5; 10$. Изображение $x_{t_k}(m, n)$, где $(m, n) \in \Omega_I$ — искомое изображение.

Получить модель изображения №2, имеющего ОАР представление (7.31), (7.32). Параметры a_{kl}, S, σ^2 аналогичны параметрам модели №1.

Алгоритм получения модели №2

Шаг 1. Смоделировать поле независимых гауссовских случайных величин $\xi(m, n), (m, n) \in \Omega$ с нулевым средним и дисперсией σ^2 .

Шаг 2. Произвести БПФ поля $\xi(m, n)$, которое заключается в последовательном применении одномерного БПФ для строк и столбцов (или наоборот) исходного поля.

Шаг 3. Разделить каждый элемент преобразованного изображения $\tilde{\xi}(m, n)$ на $\lambda(m, n)$:

$$\hat{\xi}(m, n) = \tilde{\xi}(m, n) / \lambda(m, n), \quad \lambda(m, n) = 1 - 2A^* \Phi_{mn}.$$

Шаг 4. Произвести обратное БПФ для изображения $\tilde{\xi}(m, n), (m, n) \in \Omega$.

Получить модель изображения №3, имеющего условно-марковское представление (7.44) на тороидальной решетке. Алгоритм аналогичен алгоритму №2, только в качестве значения $\lambda(m, n)$ принимается $\hat{\lambda}(m, n) = (\lambda(m, n))^{1/2}$.

Задание 1. Оценить множество соседей S для моделей №1–3. Для нахождения множества S моделей №1, 2 использовать (7.46), (7.47), а для модели №3 — (7.46), (7.48). Искать множество S среди множеств S_1, \dots, S_4 , где $S_1 = \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$; $S_2 = \{(2, 0), (1, 0), (-1, 0), (0, 2), (0, 1), (0, -1), (0, -2)\}$; $S_3 = \{(1, 0), (1, 1), (-1, 0), (-1, 1), (0, 1), (0, -1), (1, -1), (-1, -1)\} = N_8(p)$; $S_4 = \{(1, 1), (1, -1), (-1, 1), (-1, -1)\} = N_D(p)$.

Задание 2. Оценить параметры изображения из модели №1 по формулам (7.29), (7.30), из модели №2 — по формулам (7.33), (7.34), из модели №3 — по формулам (7.30), (7.45). Сравнить качество получаемых оценок и сделать выводы о применимости алгоритмов для оценивания параметров.

ЛИТЕРАТУРА

1. *Алберт А.* Регрессия, псевдоинверсия и рекуррентное оценивание. М.: Наука, 1977.
2. Алгоритмы и программы восстановления зависимостей / Под ред. В.Н.Вапника. М.: Наука, 1984.
3. *Андерсон Т.* Статистический анализ временных рядов. М.: Мир, 1976.
4. *Белокуров А.А., Сечко В.В.* Стохастические модели в задачах анализа и обработки изображений // Зарубежная радиоэлектроника. 1985. №5. С. 3–18.
5. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. Вып. 1; Вып. 2.
6. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1983.
7. *Бриллинджер Д.* Временные ряды. Обработка данных и теория. М.: Мир, 1980.
8. *Вапник В.Н.* Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.
9. *Кендалл М.Дж., Стьюарт А.* Многомерный статистический анализ и временные ряды. М.: Наука, 1976.
10. *Кендалл М.Дж., Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973.
11. *Конев В.В.* Последовательные оценки параметров стохастических динамических систем. Томск: Изд-во Томского ун-та, 1985.
12. *Крамер Г., Лидбеттер М.* Стационарные случайные процессы. М.: Мир, 1969.
13. *Лэйримор У.Э.* Статистические выводы в стационарных случайных полях // ТИИЭР. 1977. Т. 65. №6. С. 176–189.
14. *Медведев Г.А.* Простой алгоритм восстановления неизвестной функции // Изв. АН СССР. Техническая кибернетика. 1974. №3. С. 156–161.
15. *Мюллер П., Нойман П., Шторм Р.* Таблицы по математической статистике. М.: Финансы и статистика, 1982.
16. *Никифоров И.В.* Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1983.
17. Обнаружение изменения свойств сигналов и динамических систем / Под ред. М.Бассвиль, А.Банвениста. М.: Мир, 1989.
18. *Прэтт У.* Цифровая обработка изображений. М.: Мир, 1982.
19. *Смоляк С.А., Титаренко Б.П.* Устойчивые методы оценивания. М.: Статистика, 1980.
20. Справочник по прикладной статистике / Под ред. Э.Ллойд, У.Ледерман. М.: Финансы и статистика, 1990.
21. *Фомин В.Н.* Рекуррентное оценивание и адаптивная фильтрация. М.: Наука,

1984.

22. Фу К., Гонсалес Р., Ли К. Робототехника. М.: Мир, 1989.

23. Хеннан Э. Многомерные временные ряды. М.: Мир, 1974.

24. Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1975.

25. Хогг Р.В. Введение в помехоустойчивое оценивание // Устойчивые статистические методы оценки данных. М.: Машиностроение, 1984. С. 1–18.

26. Хорн Р., Джонсон Ч. Матричный анализ. М.: Мир, 1989.

27. Хьюбер П. Робастность в статистике. М.: Мир, 1984.

28. Энджел Й. Практическое введение в машинную графику. М.: Радио и связь, 1984.

Оглавление

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ	3
ОСНОВНЫЕ СОКРАЩЕНИЯ	4
ПРЕДИСЛОВИЕ	5
1 СТАЦИОНАРНЫЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ И ВРЕМЕННЫЕ РЯДЫ	6
1.1. Основные понятия и числовые характеристики	6
1.2. Стационарность и эргодичность	8
1.3. Спектральные свойства	9
1.4. Статистические критерии	11
Лабораторная работа 1. Исследование стационарного случайного про- цесса	15
2 ОЦЕНИВАНИЕ ФУНКЦИИ РЕГРЕССИИ. ВЫДЕЛЕНИЕ ТРЕНДОВ	18
2.1. Оценивание полиномиального тренда методом наименьших квадратов	18
Лабораторная работа 2. Выделение полиномиального тренда	24
2.2. Оценивание функции регрессии. Общий случай	26
Лабораторная работа 3. Изучение функций регрессии	31
2.3. Оценивание функции регрессии. Рекуррентные методы	33
Лабораторная работа 4. Рекуррентные методы оценивания функции ре- грессии	40
2.4. Метод максимального правдоподобия. М-оценки	41
Лабораторная работа 5. Оценивание функции регрессии. Метод макси- мального правдоподобия	46
2.5. Устойчивые процедуры оценивания параметров регрессии	48
Лабораторная работа 6. Устойчивые методы оценивания параметров регрессии	54
2.6. Аппроксимация функции регрессии сплайнами	55
Лабораторная работа 7. Приближение функции регрессии сплайнами	59
3 СТАЦИОНАРНЫЕ ПРОЦЕССЫ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО	61
3.1. Основные понятия	61
Лабораторная работа 8. Вычисление ковариационных и частных кова- риационных функций	65

3.2. Оценивание параметров процесса авторегрессии и процессов скользящего среднего	66
Лабораторная работа 9. Исследование методов оценивания процессов АР (p) и СС (q)	72
3.3. Оценивание параметров процесса АРСС	73
Лабораторная работа 10. Оценивание параметров процесса АРСС (p, q)	82
3.4. Обнаружение разладки процессов АРСС	82
Лабораторная работа 11. Обнаружение разладки процессов АР (p)	89
4 ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ	90
4.1. Прогнозирование тренда	90
Лабораторная работа 12. Прогнозирование нормально распределенных временных рядов	93
Лабораторная работа 13. Прогнозирование значений тренда	94
4.2. Предсказание ССП	94
Лабораторная работа 14. Предсказание стационарных временных рядов	100
4.3. Построение моделей на основе предсказания	101
Лабораторная работа 15. Определение параметров моделей АРСС (p, q) на основе предсказания	107
Лабораторная работа 16. Исследование остаточных разностей в процессе построения моделей временного ряда	108
4.4. Прогнозирование нестационарных случайных процессов, построенных на основе АРСС	108
Лабораторная работа 17. Прогнозирование процессов, порождаемых моделью ПАРСС (p, d, q)	116
Лабораторная работа 18. Исследование процессов, порождаемых моделью СПАРСС (p, d, q) \times (P, D, Q) _s	117
4.5. Калмановская фильтрация и прогнозирование	118
Лабораторная работа 19. Калмановская фильтрация и прогнозирование	122
5 СПЕКТРАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	123
5.1. Спектральные представления СП и оценивание спектральных плотностей	123
Лабораторная работа 20. Оценивание спектральной плотности временного ряда	135
5.2. Выделение скрытых периодичностей	135
Лабораторная работа 21. Выявление скрытых периодичностей	139
6 МНОГОМЕРНЫЕ ВРЕМЕННЫЕ РЯДЫ	140
6.1. Оценивание числовых характеристик многомерных числовых рядов	140
Лабораторная работа 22. Оценивание числовых характеристик многомерных временных рядов	148

6.2. Прогнозирование многомерных временных рядов и оценка их параметров на основе предсказания	150
Лабораторная работа 23. Прогнозирование многомерных временных рядов	155
6.3. Взаимные спектры и их оценивание	156
Лабораторная работа 24. Оценивание взаимных спектров и коэффициентов когерентности	161
7 ОБРАБОТКА ИЗОБРАЖЕНИЙ	162
7.1. Основные понятия	162
Лабораторная работа 25. Моделирование цифровых изображений	164
7.2. Геометрия изображений	165
Лабораторная работа 26. Геометрические преобразования изображений	167
7.3. Улучшение изображений	168
Лабораторная работа 27. Улучшение цифровых изображений	173
7.4. Выделение границ (кромки, края) изображений	174
Лабораторная работа 28. Выделение границ на изображении	179
7.5. Сегментация изображений	179
Лабораторная работа 29. Сегментация изображений	182
7.6. Специальные модели изображений	182
Лабораторная работа 30. Оценивание параметров МСП, порождаемых уравнениями АР	188
ЛИТЕРАТУРА	190