MATEMATICKO-FYZIKÁLNY ČASOPIS SAV, 16, 3, 1966

НЕКОТОРЫЕ ЗАДАЧИ ТЕОРИИ ОЦЕНОК И ПРОГНОЗИРОВАНИЯ

ЮРИЙ АНАТОЛЬЕВИЧ РОЗАНОВ, Москва (СССР)

Введение. С точки зрения приложений весьма универсальной моделью случайного процесса $X(t)$ на некотором ограниченном интервале времени t является случайный процесс второго порядка вида

$$X(t) = \sum_1^n a_k A_k(t) + \xi(t), \tag{*}$$

где $a_1, a_2, \ldots, a_n$ — некоторые числовые величины, не меняющиеся с течением времени t на рассматриваемом интервале (скажем, $0 \leqslant t \leqslant T$); $A_1(t), \ldots, A_n(t)$ — некоторые известные функции, учитывающие характер ,,детерминированной'' компоненты $A(t) = \sum_1^n a_k A_k(t)$ случайного процесса $X(t)$; $\xi(t)$ — случайный процесс второго порядка с нулевым математическим ожиданием.

Рассмотрение нескольких процессов такого типа приводит в необходимости изучать векторные случайные процессы $X(t)$ вида (*), где $a_1, \ldots, a_n$ и $\xi(t)$ представляют собой векторы соответствующей размерности.

Существенным вопросом во многих задачах теории случайных процессов и ее различных приложений является прогнозирование (оценка) некоторого процесса $Y(t)$ по ,,наблюдаемому'' векторному процессу $X(t)$ вида (*) (наиболее часто в качестве $Y(t)$ выступает ,,детерминированная'' компонента самого процесса $X(t)$, т. е. $Y(t) = \sum_1^n a_k A_k(t)$).

Обсуждению этого вопроса и посвящена наша статья.

1. Оценки наименьших квадратов

Предположим, что нам ничего не известно о распределении вероятностей случайного процесса $X(t)$ вида (*), ,,наблюдаемого'' на интервале времени $0 \leqslant t \leqslant T$. Пусть параметр t меняется либо дискретно, пробегая целые значения, либо непрерывно. При непрерывном t пусть функции $A_1(t), \ldots, A_n(t)$ вместе с почти каждой траекторией случайного процесса $\xi(t)$ интегрируемы в квадрате на интервале $[0, T]$.

Введем векторное пространство R всех комплексных функций $x = \{x(t)\}$ на интервале $0 \leqslant t \leqslant T$ со скалярным произведением

$$(x_1, x_2) = \begin{cases} \sum\limits_0^T x_1(t)\overline{x_2(t)} \text{ — для дискретного } t, \\ \int\limits_0^T x_1(t)\overline{x_2(t)}\,dt \text{ — для непрерывного } t \end{cases}$$

и нормой $\|x\| = (x, x)^{\frac{1}{2}}$ и будем трактовать $X = \{X(t)\}$, $A = \{A(t)\}$ и $\xi = \{\xi(t)\}$ как элементы пространства R. Обозначим R' линейную оболочку заданных функций $A_1 = \{A_1(t)\}, \ldots, A_n = \{A_n(t)\}$. Известно, что функция $A = \sum\limits_1^n a_k A_k$ есть элемент подпространства R'. „Наблюдается" вектор $X = \{X(t)\}$.

Согласно хорошо известному методу наименьших квадратов в качестве оценки неизвестной функции $A = \{A(t)\}$ предлагается взять вектор $\hat{A} = \{\hat{A}(t)\}$ подпространства R', наименее удаленный от наблюдаемого вектора $X = \{X(t)\}$, т. е. основание перпендикуляра, опущенного из точки X пространства R на подпространство R' (рис. 1).

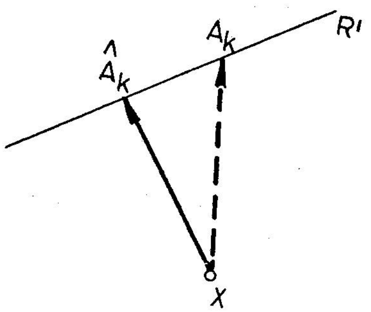

Рис. 1.

Известно, что оценка $\hat{A} = \{\hat{A}(t)\}$ всегда является несмещенной. Именно, как элемент подпространства R' вектор A может быть представлен в виде линейной комбинации заданных функций $A_1, \ldots, A_n$:

(1.1) $$\hat{A} = \sum_1^n \hat{a}_k A_k$$

и среднее значение $E\hat{A} = \sum\limits_1^n (E\hat{a}_k)A_k$ тождественно совпадает с неизвестной функцией $A = \sum\limits_1^n a_k A_k$:

(2.1) $$E\hat{A} = A$$

при любых значениях $a_1, \ldots, a_n$.

Естественно считать функции $A_1, \ldots, A_n$ линейно-независимыми. Тогда коэффициенты $\hat{a}_1, \ldots, \hat{a}_n$ разложения по векторам $A_1, \ldots, A_n$ определяются однозначно и представляют собой несмещенные оценки неизвестных коэффициентов $a_1, \ldots, a_n$:

(3.1) $$E\hat{a}_k = a_k, \quad k = \overline{1, n}$$

при любых значениях $a_1, \ldots, a_n$.

Отметим, что соотношение (3.1) сразу вытекает из условия ортогональности разности $X - \hat{A}$ к подпространству R':

$$(X - \hat{A}, A_k) = 0, \quad k = \overline{1, n}.$$

В самом деле, отсюда вытекает, что

$$E(X - \hat{A}, A_k) = (EX - E\hat{A}, A_k) = (A - E\hat{A}, A_k) = 0, \quad k = \overline{1, n}$$

т. е. вектор $A - E\hat{A}$ одновременно принадлежит подпространству R' и ортогонален ему, а значит $A - E\hat{A} = 0$.

Итак, если $A_1, \ldots, A_n$ — линейно независимы, оценки $\hat{a}_1, \ldots, \hat{a}_n$ по методу наименьших квадратов являются несмещенными оценками неизвестных коэффициентов $a_1, \ldots, a_n$:

(4.1) $$\hat{a}_k = \sum_1^n \frac{D_{kj}}{D}(X, A_j), \quad k = \overline{1, n},$$

где $D = \det\{(A_k, A_j)\}$ есть определитель Грамма векторов $A_1, \ldots, A_n$, а D_{kj} — алгебраическое дополнение к элементу (A_k, A_j).

Возникает вопрос о *состоятельности* этих оценок, когда интервал наблюдения неограниченно увеличивается: при каких условиях

(5.1) $$\lim_{T\to\infty} E\,|\hat{a}_k - a_k|^2 = 0, \quad k = \overline{1, n}.$$

Отметим, что даже в случае независимых и одинаково распределенных величин $\xi(t)$ (время t меняется дискретно), одного условия линейной независимости функции $A_1, \ldots, A_n$ недостаточно для того, чтобы оценки $\hat{a}_1, \ldots, \hat{a}_n$ были состоятельны. Это условие нужно усилить, потребовав, например, следующее: угол $\alpha_k = \alpha_k(T)$ между вектором A_k и подпространством R_k — линейной оболочки всех остальных векторов $A_j, j \neq k$ — остается при $T \to \infty$ не меньше некоторого положительного α (рис. 2):

$$\alpha_k(T) \geqslant \alpha > 0, \qquad k = \overline{1, n}. \tag{6.1}$$

Конечно, состоятельность оценок $\hat{a}_1, \ldots, \hat{a}_n$ зависит не только от свойств функций $A_1(t), \ldots, A_n(t)$, но и от распределения вероятностей случайного процесса $\xi(t)$. Предположим, что выполнено условие (6.1) и

$$E\,|(\xi, A_k)|^2 = O\{\|A_k\|^2\}, \qquad k = \overline{1, n}. \tag{7.1}$$

Тогда

$$E\,|\hat{a}_k - a_k|^2 = O\{\|A_k\|^{-2}\}, \qquad k = \overline{1, n}. \tag{8.1}$$

Действительно

$$E\,|\hat{a}_k - a_k|^2 = E\left|\left(\xi, \sum_1^n \frac{D_{jk}}{D} A_j\right)\right|^2 \leqslant \sum_1^n \frac{D_{jk}^2}{D^2} E|(\xi, A_j)|^2 = O\left\{\sum_1^n \frac{D_{jk}^2}{D^2}\|A_j\|^2\right\}$$

и при условии (6.1)

$$\frac{D_{jk}}{D}\|A_j\| = O\{\|A_k\|^{-1}\}$$

для любого $j = \overline{1, n}$ (см., например, [3]).

Остановимся подробнее на рассмотрении случайных процессов $X(t)$ колебательного типа, когда функции $A_1(t), \ldots, A_n(t)$, отражающие характер средних колебаний процесса $X(t)$, представляются в виде

$$A_k(t) = \int e^{i\lambda\tau} m_k(d\lambda), \qquad k = \overline{1, n}, \tag{9.1}$$

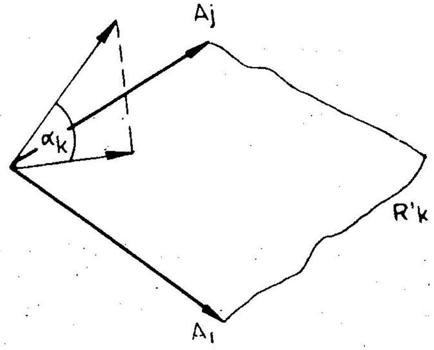

Рис. 2.

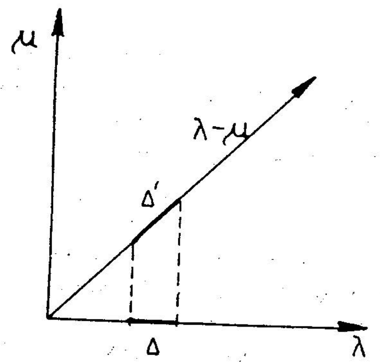

Рис. 3.

где $m_k(d\lambda)$ — некоторая комплексная мера ограниченной вариации (интегрирование ведется в пределах $-\pi \leqslant \lambda \leqslant \pi$ для дискретного t и $-\infty < \lambda < \infty$ для непрерывного t), а $\xi(t)$ является стационарным в широком смысле процессом.

Обозначим Λ_k — дискретный спектр колебательной функции $A_k(t)$, т. е. совокупность точек λ для которых $m_k(\lambda) \neq 0$ $(k = \overline{1, n})$ и Λ — объединение множеств Λ_k: $\Lambda = \bigcup_1^n \Lambda_k$.

Пусть $M_{kj}(d\lambda d\mu)$ — комплексная мера на плоскости (λ, μ), определяемая формулой

$$M_{kj}(d\lambda d\mu) = m_k(d\lambda)\overline{m_j(d\mu)}. \tag{10.1}$$

Обозначим $m_{kj}(d\lambda)$ комплексную меру на прямой λ получающуюся из рассматриваемой на диагонали $\lambda = \mu$ меры $M_{kj}(d\lambda d\mu)$ проектированием на прямую λ:

$$m_{kj}(\Delta) = M_{kj}(\Delta') \tag{11.1}$$

(здесь Δ есть проекция множества Δ', расположенного на диагонали $\lambda = \mu$ — рис. 3). Мера $m_{kj}(d\lambda)$ является чисто дискретной, сосредоточенной на общей части $\Lambda_k \cap \Lambda_j$ дискретных спектров Λ_k и Λ_j функций $A_k(t)$ и $A_j(t)$:

$$m_{kj}(\lambda) = m_k(\lambda)\overline{m_j(\lambda)}, \qquad \lambda \in \Lambda. \tag{12.1}$$

В случае непрерывного t имеем:

$$\lim_{T\to\infty} \frac{1}{T}(A_k, A_j) = \lim_{T\to\infty} \frac{1}{T}\int_0^T A_k(t)A_j(t)dt = \tag{13.1}$$

$$= \lim_{T\to\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{iT(\lambda-\mu)} - 1}{iT(\lambda - \mu)} M_{kj}(d\lambda d\mu) = \int m_{kj}(d\lambda).$$

То же соотношение справедливо и в случае дискретного t. Положим

$$m_{kj} = \int m_{kj}\,(d\lambda) = \sum_{\lambda\in\Lambda} m_k(\lambda)m_j(\lambda). \tag{14.1}$$

Очевидно, условие (6.1) ,,равномерной'' линейной независимости функций $A_1(t), \ldots, A_n(t)$ равносильно линейной независимости соответствующих мер $m_1(\lambda), \ldots, m_n(\lambda)$, рассматриваемых в точках дискретного спектра Λ:

$$d = \det\{m_{kj}\} > 0. \tag{15.1}$$

Пусть $\xi(t)$ — стационарный в широком смысле случайный процесс со спектральной мерой $F(\mathrm{d}\lambda)$ и d_{kj} означает алгебраическое дополнение к элементу m_{kj} в определителе d. Тогда

$$(16.1)\qquad \lim_{T\to\infty} E|\hat{a}_k - a_k|^2 = \sum_{\lambda\in\Lambda}\left|\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\lambda)\right|^2 F(\lambda), \qquad k = \overline{1, n}.$$

Соотношение (16.1) показывает, что *для состоятельности оценок $\hat{a}_1, \ldots, \hat{a}_k$ наименьших квадратов необходимо и достаточно, чтобы дискретный спектр Λ колебательных функций $A_1(t), \ldots, A_n(t)$ и дискретный спектр стационарного процесса $\xi(t)$ не имели бы общих точек*. Чтобы вывести это условие из соотношения (16.1), достаточно заметить, что для любой точки λ дисктетного спектра Λ хотя бы одна из мер $m_1(\lambda), \ldots, m_n(\lambda)$ отлична от нуля и в силу невырожденности определителя $\det\left\{\frac{d_{jk}}{d}\right\} = d^{-1}$ при каждом $\lambda \in \Lambda$ выражение $\sum_{j=1}^{n}\frac{d_{jk}}{d} m_j(\lambda)$ отлично от нуля хотя бы для одного значения $k = \overline{1, n}$.

Пусть стационарный процесс $\xi(t)$ имеет непрерывную спектральную плотность $f(\lambda)$. Тогда (¹)

$$(17.1)\qquad \hat{\sigma}_k^2 = E|\hat{a}_k - a_k|^2 \sim \frac{2\pi}{T}\sum_{\lambda\in\Lambda}\left|\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\lambda)\right|^2 f(\lambda).$$

Для дискретного t это соотношение является частным случаем весьма общей формулы, предложенной Гренандером и Розенблатом [1], для непрерывного t оно содержится в работе Цзян-Цзе-пея [2], рассматривающего тригонометрические функции $A_k(t)$ вида $A_k(t) = e^{i\lambda_k t}$, $k = \overline{1, n}$. Доказательство предельных соотношений (16.1) и (17.1) совершенно аналогично. Останавливаясь лишь на случае непрерывного времени t, имеем:

$$E|\hat{a}_k - a_k|^2 = \int_{-\infty}^{\infty}\!\!\left|\int_{-\infty}^{\infty} \frac{e^{iT(\lambda-\mu)}-1}{i(\lambda-\mu)}\sum_{j=1}^{n}\frac{D_{jk}}{D}\, m_j(\mathrm{d}\mu)\right|^2 F(\mathrm{d}\lambda)$$

и

$$\lim_{T\to\infty} E|\hat{a}_k - a_k|^2 = \lim_{T\to\infty}\int_{-\infty}^{\infty}\!\!\left|\int_{-\infty}^{\infty} \frac{e^{iT(\lambda-\mu)}-1}{iT(\lambda-\mu)}\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\mathrm{d}\mu)\right|^2 F(\mathrm{d}\lambda) =$$

$$= \sum_{\lambda\in\Lambda}\left|\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\lambda)\right|^2 F(\lambda).$$

Если существует непрерывная спектральная плотность $f(\lambda)$ то

$$\lim_{T\to\infty} TE|\hat{a}_k - a_k|^2 = \lim_{T\to\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\frac{1}{T}\,\frac{e^{iT(\mu-\lambda)}-1}{i(\mu-\lambda)}\cdot \frac{e^{iT(\nu-\lambda)}-1}{i(\nu-\lambda)} f(\lambda)\mathrm{d}\lambda\right]\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\mathrm{d}\mu)\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\mathrm{d}\nu) =$$

$$= \lim_{T\to\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} T\,\frac{\sin^2\frac{\mu-\lambda}{2}T}{\left(\frac{\mu-\lambda}{2}T\right)^2} f(\lambda)\mathrm{d}\lambda\right]\cdot\sum_{i,j=1}^{n}\frac{d_{ik}}{d}\,\frac{\overline{d_{jk}}}{d}\, m_{ij}(\mathrm{d}\mu) =$$

$$= 2\pi\sum_{\lambda\in\Lambda} f(\lambda)\left|\sum_{j=1}^{n}\frac{d_{jk}}{d}\, m_j(\lambda)\right|^2.$$

Несомненно, соотношения типа (16.1), (17.1) могут быть получены и в обстановке, когда функции $A_1(t), \ldots, A_n(t)$ являются первообразными от колебательных функций вида (9.1), а случайный процесс $\xi(t)$ имеет лишь стационарные приращения какого-то порядка. Отметим, что большинство приложений с успехом может быть обслужено моделями $X(t) = \sum_{1}^{n} a_k A_k(t) + \xi(t)$ именно такого типа.

Простота и достаточная эффективность метода наименьших квадратов делает его основным орудием прикладных исследований. Поэтому весьма важным представляется дать подобный же метод и для многомерных процессов $X(t)$ вида (*), другими словами, когда имеется несколько случайных процессов $X_1(t), \ldots, X_m(t)$ вида

$$(18.1)\qquad X_j(t) = \sum_{k=1}^{n} a_{jk} A_k(t) + \xi_j(t), \qquad j = \overline{1, m}.$$

Отметим, что у разных процессов $X_j(t)$ могут быть разного типа „детерминированные компоненты" $A_k(t)$, так что некоторые коэффициенты a_{jk} просто равны нулю. Наличие нескольких процессов вида (18.1) вносит существенную новизну в вопрос об оценках коэффициентов a_{jk}, $j = \overline{1, m}$, $k = \overline{1, n}$.

(¹) означает эквивалентность переменных величин α и β, т. е. $\lim \alpha/\beta = 1$.

Пример. Пусть

$$X_1(t) = a_1A_1(t) + \xi(t),$$
$$X_2(t) = a_2A_2(t) + \xi(t),$$

где функции $A_1(t)$ и $A_2(t)$ линейно независимы, а случайный процесс $\xi(t)$ один и тот же как для $X_1(t)$ так и для $X_2(t)$. В этом случае рассмотрение разности

$$Y(t) = X_1(t) - X_2(t) = a_1A_1(t) + a_2A_2(t)$$

позволяет безошибочно определить неизветсные коэффициенты a_1 и a_2.

Пример. Пусть

$$X_1(t) = a + \xi_1(t),$$
$$X_2(t) = \xi_2(t),$$

где $\xi_1(t)$ и $\xi_2(t)$ — последовательности некоррелированных при разных t случайных величин с единичной дисперсией и коэффициентом корреляции $r = E\xi_1(t)\xi_2(t)$. Если от $X_1(t)$ и $X_2(t)$ перейти к процессу

$$Y(t) = X_1(t) - \varrho X_2(t),$$

$$\varrho = \frac{(X_1, X_2)}{(X_2, X_2)}$$

то в применении к процессу $Y(t) = a + \xi_1(t) - \varrho\xi_2(t)$ обычный метод наименьших квадратов дает оценку $\hat{a}$, такую, что

$$E\,|\hat{a} - a|^2 \sim \frac{1 - r^2}{T},$$

тогда как непосредственное его применение к процессу $X_1(t)$ давало оценку $\hat{a}$, для которой

$$E\,|\hat{a} - a|^3 \sim \frac{1}{T}.$$

Прежде чем предложить обобщение на многомерный случай метода наименьших квадратов в обстановке, когда нам ничего не известно о распределении вероятностей случайных процессов $X_j(t)$, остановимся на одном известном геометрическом факте. Пусть R' и R''-произвольные, но конечномерные подпространства некоторого гильбертова пространства R. Всегда можно так выбрать ортогональные базы векторов $x_1', \ldots, x_n'$ и $x_1'', \ldots, x_m''$ в R' и R'' соответственно, что $(x_k', x_j'') = 0$ при $k \neq j$.

Именно, если обозначить P'-оператор проектирования на подпространство R' и P''-оператор проектирования на подпространство R'', то в качестве элементов $x_1', \ldots, x_n'$ можно взять ортогональную базу собственных

векторов оператора $P = P'P''P'$, а в качестве $x_1'', \ldots, x_m''$ — векторы вида $x_k'' = Px_k'$ для тех k, при которых $x_k'' \neq 0$, произвольно дополнив их до ортогональной базы в R''.

Перейдем к обобщению метода наименьших квадратов. Как и раньше, обозначим R' линейную оболочку векторов $A_1 = \{A_1(t)\}, \ldots, A_n = \{A_n(t)\}$ в гильбертовом пространстве R, и введем подпространство R''-линейную оболочку ,,наблюдаемых'' векторов $X_1 = \{X_1(t)\}, \ldots, X_m = \{X_m(t)\}$. Выберем в R' и R'' соответствующие ортогональные базы из векторов $B_1 = \{B_1(t)\}, \ldots, B_n = \{B_n(t)\}$ и $Y_1 = \{Y_1(t)\}, \ldots, Y_m = \{Y_m(t)\}$ обладающие тем свойством, что

(19.1) $$(B_k, Y_j) = 0 \quad \text{при} \quad k \neq j.$$

Пусть

(20.1) $$B_i(t) = \sum_{k=1}^{n} b_{ik}A_k(t), \quad i = \overline{1, n},$$

$$Y_i(t) = \sum_{j=1}^{n} c_{ij}X_j(t) = \sum_{k=1}^{n}\Big[\sum_{j=1}^{m} c_{ij}a_{jk}\Big]A_k(t) + \eta_i(t), \quad i = \overline{1, m}.$$

Положим

$$\varrho_i = \frac{(Y_i, B_i)}{(B_i, B_i)}.$$

Мы предлагаем определить оценки $\hat{a}_{jk}$ неизвестных коэффициентов a_{jk} из уравнений

(21.1) $$\sum_{j=1}^{m} c_{ij}\hat{a}_{jk} = \varrho_i b_{ik}, \quad k = \overline{1, n}$$

— для тех i, при которых $\varrho_i \neq 0$,

$$\sum_{j=1}^{n} c_{ij}a_{ik} = 0, \quad k = \overline{1, n}$$

— для остальных i, $i = \overline{1, m}$.

Если некоторые из коэффициентов a_{jk} являются известными, то следует в первую очередь исключить соответствующие $\hat{a}_{jk}$ из системы уравнений (21.1).

Нам представляется интересным исследовать свойства предлагаемых оценок $\hat{a}_{jk}$; например, сравнить их с оценками наименьших квадратов в одномерном процессе $X_j(t)$. Несомненно, что при некоторых обстоятельствах многомерные оценки могут оказаться лучше.

2. Наилучшие линейные оценки

Предположим, нам известна корреляционная функция $B(t, s)$ случайного процесса $\xi(t)$, фигурирующего в выражении (*) ,,наблюдаемого'' случайного процесса $X(t)$, $0 \leqslant t \leqslant T$.

Рассмотрим гильбертово пространство H всех случайных величин h, $E\,|h|^2 < \infty$, со скалярным произведением

$$\langle h_1, h_2 \rangle = Eh_1 \cdot \bar{h}_2. \tag{1.2}$$

Рассмотрим совокупность случайных величин $h \in H$ вида

$$h = (X, x), \quad x \in R, \tag{2.2}$$

где $X = \{X(t)\}$ — ,,наблюдаемый'' случайный процесс — и $x = \{x(t)\}$ трактуются как элементы введенного ранее пространства R, причем функция $x = \{x(t)\}$ удовлетворяет условию

$$(A_j, x) = \begin{cases} 1 & \text{при } j = k, \\ 0 & \text{при } j \neq k. \end{cases} \tag{3.2}$$

Замыкание множества всех величин h вида (2.2) обозначим H_k. В случае дискретного времени H_k целиком состоит лишь из величин вида (2.2); в случае непрерывного времени t оно пополняется предельными точками, например, величинами вида

$$h = \sum_{j=1}^{n} c(t_j) X(t_j),$$

где $t_1, \ldots, t_n$ — некоторые фиксированные моменты времени, для которых матрица $\{A_i(t_j)\}$, $i = \overline{1, n}$; $j = \overline{1, n}$ является невырожденной и коэффициенты $c(t_1), \ldots, c(t_n)$ определены из соотношений

$$\sum_{j=1}^{n} c(t_j) A_i(t_j) = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{при } i \neq k. \end{cases}$$

Множество H_k геометрически представляет собой плоскость в гильбертовом пространстве H (конечномерную — в случае дискретного t, и бесконечномерную — в случае непрерывного t), проходящую через точки h вида (2.2). С точки зрения статистики, H_k представляет собой совокупность всех линейных несмещённых оценок неизвестного параметра a_k:

$$Eh = a_k, \quad h \in H_k \tag{4.2}$$

какови бы ни были истинные значения $a_1, \ldots, a_n$. В частности, элементом плоскости H_k является оценка $\hat{a}_k$ наименьших квадратов. Наилучшей

среди всех оценок $h \in H_k$ естественно назвать элемент $\tilde{a}_k \in H_k$, наименее всех удалённый от истинного значения a_k:

$$E\,|\tilde{a}_k - a_k|^2 = \min_{h \in H_k} E\,|h - a_k|^2. \tag{5.2}$$

Геометрически наилучшая оценка $\tilde{a}_k$ представляет собой основание перпендикуляра, опущенного из точки a_k пространства H на плоскость H_k (рис. 4).

В случае дискретного времени t всякий элемент h плоскости H_k имеет вид (2.2); в частности,

$$\tilde{a}_k = (X, x_k). \tag{6.2}$$

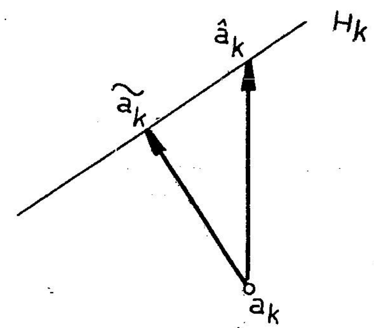

Рис. 4.

Функция $x_k = \{x_k(t)\}$, дающая выражение (6.2) для наилучшей оценки $\tilde{a}_k$, может быть найдена из (3.2) и уравнения

$$\sum_{0}^{T} B(t, s) x_k(s) = \sum_{1}^{n} \sigma_{kj} \overline{A_j(t)}, \quad 0 \leqslant t \leqslant T \tag{7.2}$$

вместе с некоторыми постоянными σ_{kj}, $j = \overline{1, n}$. В случае непрерывного t наилучшая оценка $\tilde{a}_k$ оказывается, как правило, лишь предельной точкой для величин вида (2.2) и сама, строго говоря, не может быть представлена в таком виде. Тем не менее, как известно, символическое выражение величины $\tilde{a}_k$ формулой (6.2) позволяет выписать интегральное уравнение для искомой функции $x_k = \{x_k(t)\}$:

$$\int_{0}^{T} B(t, s) x_k(s) ds = \sum_{1}^{n} \sigma_{kj} A_j(t), \quad 0 \leqslant t \leqslant T. \tag{8.2}$$

Функция $x_k = \{x_k(t)\}$, как правило, оказывается обобщённой. Отметим, что наилучшая оценка $\tilde{a}_k$ в случае непрывного времени t является пределом соответствующих оценок $\tilde{a}_{kN}$ в дискретной модели, когда t пробегает лишь значения $0, 1/N, 2/N, \ldots, T/N$:

$$\lim_{N \to \infty} E\,|\tilde{a}_{kN} - \tilde{a}_k|^2 = 0. \tag{9.2}$$

Отметим также, что в случае непрерывного времени t при некоторых обстоятельствах возможны безошибочные оценки неизвестных коэффициентов $a_1, \ldots, a_n$.

Пример. Пусть

$$X(t) = aA(t) + \xi(t),$$

где $\xi(t)$ — процесс с непрерывными траекториями, а функция $A(t)$ имеет вид

$$A(t) = \begin{cases} 0 & \text{при } 0 \leqslant t \leqslant t_0, \\ 1 & \text{при } t > t_0. \end{cases}$$

В этом случае неизвестное значение a определяется без ошибки:

$$a = X(t+0) - X(t-0).$$

Естественно попытаться описать все случаи, когда сами коэффициенты $a_1, \ldots, a_n$ или их некоторые линейные комбинации могут быть определены безошибочно, другими словами, когда распределение вероятностей наилучших оценок $\tilde{a}_1, \ldots, \tilde{a}_n$ является вырожденным.

Пусть $\xi(t)$ — стационарный в широком смысле случайный процесс со спектральной мерой $F(d\lambda)$. Введем класс функций $\varphi = \varphi(\lambda)$, являющихся среднеквадратичным пределом тригонометрических функций вида $\sum_{t \in T} c(t) e^{i\lambda t}$:

(10.2)
$$\inf \int |\varphi(\lambda) - \sum_{t \in T} c(t) e^{i\lambda t}|^2 F(d\lambda) = 0.$$

Обозначим этот класс функций $L_T^2(F)$. Оказывается, *чтобы при любых $a_1, \ldots, a_n$ совместное распределение вероятностей наилучших оценок $\tilde{a}_1, \ldots, \tilde{a}_n$ было невырожденным, необходимо и достаточно, чтобы каждая из функций $A_k(t)$, $k = \overline{1, n}$, могла быть представлена в спектральном виде:*

(11.2)
$$A_k(t) = \int e^{-i\lambda t} \psi_k(\lambda) F(d\lambda), \quad k = \overline{1, n},$$

где $\psi_k(\lambda)$ — некоторые функции класса $L_T^2(F)$.

При этом, решение $x_k = \{x_k(t)\}$ интегрального уравнения (8.2), задающее наилучшую оценку $\tilde{a}_k$, является преобразованием Фурье функции $\varphi_k(\lambda)$ вида

(12.2)
$$\varphi_k(\lambda) = \sum_{j=1}^{n} \sigma_{kj} \psi_j(\lambda):$$

$$x_k(t) = \frac{1}{2\pi} \int e^{i\lambda t} \varphi_k(\lambda) d\lambda,$$

где коэффициенты σ_{kj} определяются из условий несмещённости (3.2):

(13.2)
$$\sum_{i=1}^{n} \sigma_{ki} \int \psi_i(\lambda) \overline{\psi_i(\lambda)} F(d\lambda) = \begin{cases} 1 & \text{при } j = k, \\ 0 & \text{при } j \neq k, \end{cases}$$

(см. Писаренко и Розанов [3]).

Отметим, что матрица $\sigma^2 = \{\sigma_{kj}\}$ представляет собой корреляционную матрицу наилучших оценок $\tilde{a}_1, \ldots, \tilde{a}_n$:

(14.2)
$$\sigma_{kj} = E(\tilde{a}_k - a_k)\overline{(\tilde{a}_j - a_j)}, \quad k, j = \overline{1, n}.$$

Соотношение (11.2) представляет собой интегральное уравнение типа Винера-Хопфа относительно неизвестных функций $\psi_k(\lambda)$, $k = \overline{1, n}$, разные методы решения которого хорошо известны. (см., например, Яглом [4], Писаренко и Розанов [3]).

С нашей точки зрения рассмотренные выше вопросы являются весьма важными и решение их для многомерной модели типа (*) заслуживает всеобщего внимания.

Остановимся подробнее снова на колебательных случайных процессах $X(t)$, когда функции $A_1(t), \ldots, A_n(t)$ представляются в виде (9.1), а $\xi(t)$ является стационарным в широком смысле процессом. Мы уже рассмотрели вопрос об асимптотике ошибок в оценках наименьших квадратов и установили, что

$$\hat{\sigma}_k^2 = E\,|\hat{a}_k - a_k|^2 \sim \frac{2\pi}{T} \sum_{\lambda \in \Lambda} f(\lambda) \left| \sum_{j=1}^{n} \frac{d_{jk}}{d} m_j(\lambda) \right|^2, \quad k = \overline{1, n},$$

где $f(\lambda)$ — спектральная плотность стационарного процесса $\xi(t)$, предполагающаяся непрерывной функцией λ. Аналогичный вопрос возникает и для наилучших оценок $\tilde{a}_k$. Известно (см. Гренандер и Розенблат [1]), что в случае дискретного времени t и непрерывной положительной спектральной плотности $f(\lambda)$

(15.2)
$$\tilde{\sigma}_k^2 = E\,|\hat{a}_k - a_k|^2 \sim \frac{2\pi}{T} \alpha_{kk},$$

где α_{kk} есть диагональный элемент матрицы $\{\alpha_{kj}\}$, обратный к матрице вида $\sum_{\lambda \in \Lambda} f^{-1}(\lambda) m_{kj}(\lambda)$.

Для непрерывного t аналогичный результат для частного случая тригонометрических функций $A_k(t) = e^{i\lambda_k t}$, $k = \overline{1, n}$, был получен Цзян-Цзепеем [2].

Мы укажем прием, позволяющий получить асимптотическую оценку типа (15.2) в случае непрерывного времени, опираясь на соответствующие формулы дискретного случая.

Именно, рассмотрим случайный процесс $X(t)$ в дискретные моменты времени $t = 0, 1/N, 2/N, \ldots, T$ (считаем, что T кратно $1/N$). В такой дискретной модели нужно перейти к соответствующим спектральным мерам $m_k^*(d\lambda)$ и плотности $f^*(\lambda)$ на отрезке $-\pi N \leqslant \lambda \leqslant \pi N$, отождествив точки всей прямой $-\infty < \lambda < \infty$, совпадающие mod $2\pi N$:

$$m_k^*(\lambda) = \sum_{j=-\infty}^{\infty} m_k(\lambda + 2\pi N_j), \tag{16.2}$$

$$f^*(\lambda) = \sum_{j=-\infty}^{\infty} f(\lambda + 2\pi N_j).$$

Предположим, что для дискретной модели $X(t)$ имеют место асимптотические формулы вида

$$\tilde{\sigma}_k^2(N) = E\,|\tilde{a}_{kN} - a_k|^2 \sim \frac{2\pi}{T}\,\alpha_{kk}(N), \tag{17.2}$$

где $\alpha_{kk}(N)$ — некоторые постоянные. Ясно, что при каждом k последовательность $\alpha_{kk}(N)$ монотонно убывает при $N \to \infty$. Предположим, что

$$\alpha_{kk} = \lim_{N\to\infty} \alpha_{kk}(N) > 0. \tag{18.2}$$

Тогда в случае непрерывного времени t для величины

$$\tilde{\sigma}_k^2 = E\,|\tilde{a}_k^2 - a_k|^2$$

имеет место следующая асимптотическая оценка:

$$\overline{\lim_{T\to\infty}}\, T\tilde{\sigma}_k^2 \leqslant 2\pi\alpha_k. \tag{19.2}$$

В самом деле, при любом фиксированном T последовательность $\tilde{\sigma}_{kk}^2(N)$, $N = 1, 2, \ldots$, монотонно убывает и $\tilde{\sigma}_k^2 = \lim_{N\to\infty} \tilde{\sigma}_k^2(N)$. Поэтому, $\overline{\lim_{N\to\infty}}\, T\tilde{\sigma}_k^2 \leqslant$ $\leqslant 2\pi\alpha_k$.

Для колебательных функций предельное значение α_{kk} легко получается из выражений для $\alpha_{kk}(N)$. Именно, $\alpha_{kk} = \lim_{N\to\infty} \alpha_{kk}(N)$ есть диагональный элемент матрицы $\{\alpha_{kj}\}$ вида

$$\{\alpha_{kj}\} = \{\sum_{\lambda\in\Lambda} f^{-1}(\lambda) m_{kj}(\lambda)\}^{-1}. \tag{20.2}$$

Сделаем несколько замечаний.

При выводе формулы (15.2) Гренандер и Розенблат [1] предполагали строгую положительность спектральной плотности $f(\lambda)$ не только в точках λ дискретного спектра Λ колебательных функций $A_1(t), \ldots, A_n(t)$ но и во всём диапозоне спектра стационарного процесса $\xi(t)$. Повидимому, доста-

точно потребовать положительность $f(\lambda)$ лишь в ε-окрестности множества $[\Lambda]$ — замыкания дискретного спектра колебательных функций.

Несомненно, что асимптотические формулы типа (15.2) и (19.2) могут быть получены, когда функции $A_1(t), \ldots, A_n(t)$ являются первообразными колебательных функций вида (9.1), а случайный процесс $\xi(t)$ имеет лишь стационарные приращения определённого порядка.

Сравнение асимптотики ошибок оценок наименьших квадратов $\hat{a}_k$ и наилучших линейных оценок $\tilde{a}_k$ показывает, что во многих случаях они грубо эквивалентны друг другу при $T \to \infty$:

$$\hat{\sigma}_k \asymp \tilde{\sigma}_k, \quad k = \overline{1, n}, \tag{21.2}$$

т.е.

$$0 < \lim \frac{\tilde{\sigma}_k}{\hat{\sigma}_k} \leqslant \overline{\lim}\, \frac{\tilde{\sigma}_k}{\hat{\sigma}_k} < \infty,$$

а в некоторых случаях и точно эквивалентны:

$$\hat{\sigma}_k \sim \tilde{\sigma}_k, \quad k = \overline{1, n}. \tag{22.2}$$

С нашей точки зрения заслуживает всяческого интереса вопрос о том, при каких условиях имеет место соотношение (21.2) или (22.2). Необходимым условием для этого нам кажется следующее:

$$E\,|(\xi, A_k)|^2 \asymp \|A_k\|^2, \quad k = \overline{1, n}. \tag{23.2}$$

В случае колебательных процессов $X(t)$, у которых дискретный спектр Λ колебательных функций $A_1(t), \ldots, A_n(t)$ ограничен, условие (23.2) является, повидимому, и достаточным. Нарушение условия (23.2), как правило, приводит к тому, что оценки наименьших квадратов оказываются значительно хуже наилучших оценок.

Пример. Пусть

$$X(t) = a + \xi(t),$$

где $\xi(t) = \eta(t) - \eta(t-1)$ есть разность некоррелированных величин $\eta(t)$ с единичной дисперсией. Применение метода наименьших квадратов непосредственно к процессу $X(t)$ даёт ошибку $\hat{\sigma}_k$:

$$\hat{\sigma}_k^2 \asymp \frac{1}{T^2}.$$

Если перейти к процессу $Y(t)$ вида

$$Y(t) = \sum_0^t X(s) = at + \eta(t) - \eta(0),$$

то применение лишь метода наименьших квадратов к процессу $Y(t)$ даёт ошибку $\hat{\sigma}_k$:

$$\hat{\sigma}_k^2 \asymp \frac{1}{T^4}$$

(наилучшая оценка $\tilde{\sigma}_k^2$ имеет ошибку того же порядка). Отметим, что для $X(t)$ условие (23.2) нарушено, а для процесса $Y(t)$ выполнено.

Далее, оценки $\hat{a}_1, \ldots, \hat{a}_n$ наименьших квадратов совпадают с наилучшими оценками, когда соответствующий процесс $\xi(t)$ является ,,белым шумом''. Другими словами, в обстановке, когда нам ничего не изветсно о распределении вероятностей процесса $\xi(t)$ мы считаем его ,,белым шумом'' и находим наилучшие оценки — они то и есть оценки наименьших квадратов. Естественно ожидать, что если учитывать некоторые свойства имеющегося процесса $\xi(t)$, то наилучшие оценки с учётом этих свойств окажутся выгоднее оценок наименьших квадратов. Это следует ожидать в обстановке, когда оценки наименьших квадратов значительно хуже наилучших оценок, в частности, когда нарушаются условия типа (23.2). Исследование в этом направлении нам кажется весьма актуальным.

Вот одна из конкретных задач. Пусть стационарный псроцесс $\xi(t)$ образуется на выходе линейного устройства, описывающегося дифференциальным уравнением

$$\sum_k c_k \xi^{(k)}(t) = \eta(t).$$

Относительно входного сигнала $\eta(t)$ ничего не изветсно. Какие оценки неизвестных коэффициентов $a_1 \ldots, a_n$ следует дать в этой обстановке Выгодно ли ,,усложнять'' оценки наименьших квадратов

Подробного исследования с разных точек зрения ждёт многомерный случай процессов $X(t)$ вида (*).

Наряду с вопросом об асимптотическом поведении ошибок $\hat{\sigma}_k$, $\tilde{\sigma}_k$ оценок коэффициентов a_k $(k = \overline{1, n})$ важное значение имеет вопрос об асимптотическом распределении вероятностей этих оценок. Несомненно, при весьма общих предположениях относительно функций $A_1(t), \ldots, A_n(t)$ и процесса $\xi(t)$ нормированные величины $\dfrac{\hat{a}_k - a_k}{\hat{\sigma}_k}$, $\dfrac{\tilde{a}_k - a_k}{\tilde{\sigma}_k}$ будут асимптотически нормальны при $T \to \infty$.

3. Оценки максимума правдоподобия

Предположим, что случайный процесс $\xi(t)$, фигурирующий в представлении (*) рассматриваемого процесса, является стационарным гауссовским

процессом со спектральной мерой $F(d\lambda)$. Как известно, при условии (11.2) распределение вероятностей $P_{a_1, \ldots, a_n}(dX)$ случайного процесса $X = \{X(t)\}$ с параметрами $a_1, \ldots, a_n$ будет абсолютно непрерывно относительно распределения вероятностей $P(dX)$, отвечающее нулевым значениям параматров $a_1 = 0, \ldots, a_n = 0$, так что существует плотность

$$\varrho_{a_1, \ldots, a_n}(X) = \frac{P_{a_1, \ldots, a_n}(dX)}{P(dX)}, \tag{1.3}$$

которая выражается по формуле

$$\log \varrho_{a_1, \ldots, a_n}(X) = (X, x) - \tfrac{1}{2} E|(X, x)|^2 \tag{2.3}$$

(см. [3], [7]). Фигурирующая в скалярном произведении (X, x) функция $x = \{x(t)\}$ есть

$$x(t) = \frac{1}{2\pi} \sum_{k=1}^{n} a_k \int e^{i\lambda t} \psi_k(\lambda) d\lambda, \tag{3.3}$$

где $\psi_1(\lambda), \ldots, \psi_n(\lambda)$ — функции, дающие спектральное представление (11.2) для $A_1(t), \ldots, A_n(t)$; при этом

$$E|(X, x)|^2 = \sum_{k,j=1}^{n} a_k a_j \int \psi_k(\lambda) \psi_j(\lambda) F(d\lambda). \tag{4.3}$$

Согласно известному методу максимума правдоподобия, в качестве оценок $\tilde{a}_1, \ldots, \tilde{a}_n$ неизвестных параметров $a_1, \ldots, a_n$ можно взять точки максимума ,,наблюдаемой'' функции правдоподобия

$$L(a_1, \ldots, a_n) = \log \varrho_{a_1, \ldots, a_n}(X). \tag{5.3}$$

Эти оценки находятся из уравнений

$$\frac{\partial}{\partial a_k} L(\tilde{a}_1, \ldots, \tilde{a}_n) = 0, \qquad k = \overline{1, n} \tag{6.3}$$

и имеют вид

$$\tilde{a}_k = (X, x_k), \qquad k = \overline{1, n}, \tag{7.3}$$

где

$$x_k = \frac{1}{2\pi} \sum_{j=1}^{n} \frac{c_{kj}}{c} \int e^{i\lambda t} \psi_j(\lambda) d\lambda, \tag{8.3}$$

$c = \det \left\{ \int \psi_k \psi_j F(d\lambda) \right\}$ а c_{kj} — алгебраическое дополнение к элементу $\int \psi_k(\lambda) \psi_j(\lambda) F(d\lambda)$ определителя c.

Нетрудно заметить, что полученные оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ максимума правдоподобия в точности совпадают с наилучшими линейными оценками, определяемыми формулами (6.2).

Напомним, что несмещенные оценки действительных параметров $a_1, \ldots, a_n$ называются **эффективными**, если их корреляционная матрица $\sigma^2 = \{\sigma_{kj}\}$:

$$\sigma_{kj} = E(\tilde{a}_k - a_k)(\tilde{a}_j - a_j), \qquad k, j = \overline{1, n} \tag{9.3}$$

является обратной к так называемой информационной матрице Фишера $r^2 = \{r_{kj}\}$:

$$r_{kj} = E\frac{\partial L}{\partial a_k}\cdot\frac{\partial L}{\partial a_j}, \qquad k, j = \overline{1, n}. \tag{10.3}$$

Как известно, во всех регулярных случаях для корреляционной матрицы σ^2 любых несмещенных оценок имеет место неравенство Крамера-Рао-Фреше:

$$\sigma^2 \leqslant r^{-2} \tag{11.3}$$

т. е. разность $\sigma^2 - r^{-2}$ является положительно определенной матрицей (см. [5], [6]).

В нашем случае легко подсчитать, что

$$r_{kj} = \int \psi_k(\lambda)\overline{\psi_j(\lambda)}F(d\lambda), \qquad k, j = \overline{1, n}, \tag{12.3}$$

откуда видно, что оценки максимума правдоподобия (наилучшие линейные оценки) являются эффективными.

Нам представляется интересным установить аналогичные свойства оценок максимума правдоподобия в многомерной модели случайного процесса $X(t)$.

4. Байессовские оценки. Прогнозирование

В большинстве задач регулирования систем со случайными возмущениями краеугольным камнем лежит вопрос о прогнозе (оценке) некоторой величины Y по известным данным о поведении некоторого случайного процесса $X(t)$ в определнном промежутке времени; чаще всего величина Y представляет собой неизвестное значение какого-либо случайного процесса $Y(\tau)$ в текущий или относящийся к будущему момент времени τ.

Ясно, что рецепт удовлетворительного прогноза зависит от вероятностных закономерностей процесса $X(t)$ и связи его с величиной Y. Очень часто все это остается неизвестным, и дающая прогноз система должна в каждой конкретной обстановке в той или иной степени „настроиться"

на соответствующие закономерности течения процесса $X(t)$. Ниже для решения подобного рода задач предлагается использовать байессовский подход, который в наиболее простых и важных случаях дает вполне удовлетворительное с практической точки зрения решение.

Легко представить себе, что случайность может проявлять себя двояко. Во-первых, от случая могут зависеть условия применения данной системы, характеризуемые некоторым параметром Θ (термин „условия применения" должен пониматься достаточно широко). Во-вторых, при соответствующих условиях Θ величина Y и наблюдаемый случайный процесс $X(t)$ характеризуются совместным распределение вероятноятей P_Θ, зависящим от параметра Θ. Распределение вероятностей $\pi(d\Theta)$ случайного параметра Θ, характеризующего „условия применения" системы, предположительно может быть найдено на основе статистических данных при многократном их рассмотрении, и считается известным. Задача заключается в том, чтобы по случайному процессу $X(t)$, „наблюдаемого" на некотором ограниченном интервале времени (скажем, $0 \leqslant t \leqslant T$) дать прогноз $\hat{Y}$ неизветсной величины Y:

$$\hat{Y} = \hat{Y}[X(t),\ 0 \leqslant t \leqslant T] \tag{1.4}$$

наилучший в том смысле, что

$$E|\hat{Y} - Y|^2 = \min, \tag{2.4}$$

где E означает математическое ожидание, соответствующее вероятностной мере $P = P_\Theta\pi(d\Theta)$.

Общий вид наилучшего прогноза $\hat{Y}$ хорошо известен и дается формулой условного математического ожидания:

$$\hat{Y} = E[Y|X(t),\ 0 \leqslant t \leqslant T] \tag{3.4}$$

так что задача состоит в нахождении эффективных методов вычисления условного математического ожидания типа (3.4). С практической точки зрения наиболее интересными являются случайные процессы $X(t)$ рассматриваемого ранее вида (*); при этом коэффициенты $a_1, \ldots, a_n$ остаются постоянными с течением времени t на интервале $0 \leqslant t \leqslant T$, но могут зависеть от случая (от случайного параметра Θ с распределением вероятностей $\pi(d\Theta)$), а случайный процесс $\xi(t)$ является гауссовским.

Предположим временно, что нам известно истинное значение введенного случайного параметра Θ, так что коэффициенты $a_1, \ldots, a_n$ будет просто известными постоянными. Для каждого фиксированного значения Θ наилучший прогноз неизвестсной величины Y дается формулой

$$\hat{Y}_\Theta = E_\Theta[Y|X(t),\ 0 \leqslant t \leqslant T], \tag{4.4}$$

где E_Θ означает условное математическое ожидание, соответствующее вероятностной мере P_Θ. Если совместное распределение вероятностей величины Y и процесса $X(t)$ (при фиксированном Θ) являются гауссовскими, то величина $\tilde{Y}_\Theta$ может быть представлена в виде

$$Y_\Theta = b_\Theta + (\xi, y_\Theta) = (X, y_\Theta) - \sum_1^n a_k(A_k, y_\Theta), \tag{5.4}$$

где

$$b_\Theta = E_\Theta Y$$

а функция $y_\Theta = \{y_\Theta(t)\}$ может быть найдена из уравнений типа Винера-Хопфа:

$$\sum_0^T B_\Theta(t, s) y_\Theta(s) = B_\Theta(t) \quad \text{— для дискретного } t, \tag{6.4}$$

$$\int_0^T B_\Theta(t, s) y_\Theta(s) \mathrm{d}s = B_\Theta(t) \quad \text{— для непрерывного } t,$$

где

$$B_\Theta(t) = E_\Theta Y \xi(t); \qquad B_\Theta(t, s) = E_\Theta \xi(t)\xi(s). \tag{7.4}$$

В случае стационарности процесса $\xi(t)$ функция $y_\Theta(t)$ является преобразованием Фурье обычной функции $\varphi = \varphi(\lambda)$ класса $L_T^2(F)$, удовлетворяющей спектральному аналогу уравнения (6.4):

$$\int \mathrm{e}^{-\mathrm{i}\lambda t} \varphi_\Theta(\lambda) F_\Theta(\mathrm{d}\lambda) = B_\Theta(t), \qquad 0 \leqslant t \leqslant T. \tag{8.4}$$

Методы решения подобных уравнений вполне эффективны в наиболее важном случае рационального спектра, когда спектральная мера $F_\Theta(\mathrm{d}\lambda)$ стационарного процесса $\xi(t)$ абсолютно непрерывна и спектральная плотность $f_\Theta(\lambda)$ является рациональной (см. например, [3], [4]).

После того, как найдено выражение (5.4), можно перейти к самой величине $\tilde{Y}$, выражающейся через $\tilde{Y}_\Theta$ и ,,апостериорное'' распределение вероятностей $\pi(\mathrm{d}\Theta/X(t),\ 0 \leqslant t \leqslant T)$ случайного параметра Θ следующим образом:

$$\hat{Y} = \int_{\{\Theta\}} \tilde{Y}_\Theta \pi(\mathrm{d}\Theta/X(t),\ 0 \leqslant t \leqslant T); \tag{9.4}$$

здесь $\pi(\mathrm{d}\Theta/X(t), 0 \leqslant t \leqslant T)$ есть условное распределение вероятностей параметра Θ, когда фиксируется траектория наблюдаемого процесса $X(t)$.

Не ограничивая общности можно считать, что при любых значениях параметра Θ все вероятностные распределения $P_\Theta(\mathrm{d}X)$ взаимно абсолютно

непрерывны, и тогда условное распределение $\pi(\mathrm{d}\Theta/X(t),\ 0 \leqslant t \leqslant T)$ можно вычислить по формуле:

$$\pi(\mathrm{d}\Theta/X(t),\quad 0 \leqslant t \leqslant T) = \frac{p_\Theta(X)\pi(\mathrm{d}\Theta)}{\int_{\{\Theta\}} p_\Theta(X)\pi(\mathrm{d}\Theta)}, \tag{10.4}$$

где $p_\Theta(X)$ плотность распределения вероятностей $P_\Theta(\mathrm{d}X)$ случайного процесса $X = \{X(t)\}$ при фиксированном значении параметра Θ относительно меры $P_{\Theta_0}(\mathrm{d}X)$, отвечающей некоторому значению параметра $\Theta = \Theta_0$.

Если считать случайный гауссовский процесс $\xi(t)$ стационарным, и причем его спектральную меру $F(\mathrm{d}\lambda)$ считать одной и той же при любых значениях Θ (при любых ,,условиях применения'' рассматриваемой модели (*)), то плотность вероятности $p_\Theta(X)$ выражается формулой (2.3). Именно,

$$p_\Theta(X) = \exp\{(X, x) - \tfrac{1}{2} \sum_{k, j=1}^n a_k a_j \int \psi_k(\lambda)\psi_j(\lambda) F(\mathrm{d}\lambda)\}, \tag{11.4}$$

где

$$x(t) = \frac{1}{2\pi} \sum_1^n a_k \int \mathrm{e}^{\mathrm{i}\lambda t} \psi_k(\lambda) \mathrm{d}\lambda,$$

Например, наилучшая ,,байессовская оценка'' коэффициентов $a_1, \ldots, a_n$ будет даваться выражением:

$$\tilde{a}_k = \int_{\{\Theta\}} a_k \pi(\mathrm{d}\Theta/X(t),\ 0 \leqslant t \leqslant T), \qquad k = \overline{1, n}. \tag{12.4}$$

Зная ,,байессовские оценки'' $\tilde{a}_1, \ldots, \tilde{a}_n$ нетрудно дать наилучший прогноз значения $Y = X(\tau)$ самого процесса X вне интервала наблюдения $0 \leqslant t \leqslant T$. Именно,

$$\tilde{Y} = \sum_1^n \tilde{a}_k A_k(\tau) + (X - \sum_1^n \tilde{a}_k A_k, y), \tag{13.4}$$

где функция $y = \{y(t)\}$ находится из уравнения (6.4) Винера-Хопфа с правой частью

$$B_\Theta(t) = B(\tau, t), \tag{14.4}$$

где $B(\tau, t)$ — корреляционная функция процесса $\xi(t)$.

ЛИТЕРАТУРА

[1] Grenander U., Rosenblatt M., *Statistical analysis of stationary time series*, New York 1957.

[2] Chiang Tse-pei, *On the estimation of regression coefficients*, Теория вероятностей и ее применения *4* (1959), 405—423.

[3] Писаренко В. Ф., Розанов Ю. А., *Некоторые задачи для стационарных процессов, приводящих к интегральным уравнениям типа Винера-Хопфа*, Теория оптимального кодирования *14* (1963), 113—135.

[4] Hájek J., *On linear estimation theory for an infinite number of observations*, Теория вероятностей и ее применения *6* (1961), 182—193.

[5] Cramer H., *Mathematical methods of statistics*, Princeton, N. J. 1946.

[6] Grenander U., *Stochastic processes and statistical inference*, Arkiv mat. *1* (1950), 195—276.

[7] Hájek J., *On linear statistical problems in stochastic processes*, Чехосл. матем. ж. *12* (1962), 404—444.

Поступило 12. 5. 1965.

Математический ин-т им. В. А. Стеклова
Академии наук СССР,
Москва, СССР