

Санкт-Петербургский государственный университет

В. Б. Мелас
П. В. Шпилев

ПЛАНИРОВАНИЕ И АНАЛИЗ ДЛЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ

У ч е б н о е п о с о б и е

Санкт-Петербург
2014

УДК 519.24

ББК 22.18

П25

Рецензенты:

проф. Е.В. Седунов (СЗИП СПГУТД)

проф. В. Б. Невзоров (С.-Петерб. гос. ун-т)

*Печатается по постановлению
Редакционно-издательского совета
С.-Петербургского государственного университета*

Мелас В.Б., Шпилев П.В.

П25 Планирование и анализ для регрессионных моделей: Учеб. пособие. — СПб., 2014. — 98 с.

В учебном пособии изложены основы регрессионного анализа и оптимального планирования эксперимента. В основу пособия положен курс лекций, читаемых на математико-механическом факультете СПбГУ.

Пособие рассчитано на студентов и аспирантов, специализирующихся в области математического моделирования.

ББК 22.18

© В. Б. Мелас,
П. В. Шпилев, 2014
© С.-Петербургский
государственный
университет, 2014

Предисловие

В настоящее время математические методы планирования эксперимента получили основательное теоретическое развитие и широко используются на практике. В частности, такие методы используются при анализе производственных процессов для изучения всевозможных зависимостей между величинами, влияющими на ход исследуемого процесса. Искомые зависимости могут быть выведены теоретически или же получены из экспериментальных данных. Раздел статистики, объединяющий практические методы исследования (регрессионной) зависимости между величинами по экспериментальным данным, получил название «регрессионный анализ». Целью регрессионного анализа является определение общего вида уравнения регрессии (модели), построение оценок неизвестных параметров, входящих в это уравнение, и проверка статистических гипотез об этих параметрах. При построении моделей по экспериментальным данным существенную роль играет оптимальный выбор условий проведения экспериментов. Этой задачей занимается теория оптимального планирования экспериментов, интенсивно развивающаяся с середины прошлого века и имеющая многочисленные приложения.

Цель данного учебного пособия — познакомить читателя с базовыми понятиями и результатами регрессионного анализа и математической теории планирования регрессионных экспериментов. Авторы постарались дать краткое, доступное и в то же время достаточно полное изложение материала по указанной теме.

В основе учебного пособия лежат курсы лекций, читаемых авторами на математико-механическом факультете СПбГУ.

Пособие состоит из двух частей. В первой части рассматриваются основные понятия и методы регрессионного анализа, подробно исследуются задачи построения оценок, в частности, для моделей неполных рангов, приводится ряд классических теоретических результатов. Вторая часть посвящена описанию результатов

теории оптимального планирования экспериментов. Подробно излагаются классические результаты, касающиеся свойств информационных матриц, теорема Каратеодори, теорема эквивалентности Кифера—Вольфовица и ее следствия. Также приводится ряд новых результатов, основанных на применении теорем эквивалентности.

1. Регрессионный анализ

1.1. Описательная регрессия

Рассмотрим следующую задачу. Пусть у нас есть некоторая система, работающая по принципу «черного ящика». На вход подается набор управляющих сигналов $x_1, \dots, x_m$ (на практике обычно $x_1, \dots, x_m$ — некоторые количественные величины). На выходе — некоторая скалярная величина y , зависящая от $x_1, \dots, x_m$. Мы хотим определить эту зависимость, т.е. найти функцию

$$y = \eta(x_1, \dots, x_m).$$

Для решения этой задачи на первом этапе необходимо получить экспериментальные данные, то есть результаты совместных измерений величин $y, x_1, \dots, x_m$. Допустим, мы провели N измерений. Результаты можно записать в виде матрицы эмпирических данных:

$$(Y, X) = \begin{bmatrix} y_1 & x_{11} & \dots & x_{1m} \\ y_2 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ y_N & x_{N1} & \dots & x_{Nm} \end{bmatrix}, \quad (1.1)$$

где

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots \\ x_{N1} & \dots & x_{Nm} \end{bmatrix}, \quad (1.2)$$

$y_j, x_{j1}, \dots, x_{jm}$, ($j = 1, \dots, N$) — значения величин $y, x_1, \dots, x_m$ в j -м измерении. Вектор Y называется *вектором результатов измерений*, а матрица X — *матрицей плана эксперимента* или просто

матрицей плана. Слово «план» отражает тот факт, что входные значения x_{ij} могут быть заданы или измерены экспериментатором до проведения эксперимента по измерению величин y_j .

В большинстве случаев функцию, точно выражающую искомую зависимость, найти не удастся, в связи с тем, что исследуемые зависимости оказываются весьма сложными. На практике используются упрощенные модели, аппроксимирующие предполагаемую статистическую зависимость с удовлетворительной точностью.

В общей постановке задача описания эмпирической зависимости с помощью параметрической регрессии предполагает, что задается функция, определенная с точностью до нескольких параметров, которые подбирают таким образом, чтобы получающаяся функция наилучшим образом описывала данные (1.1). Функция η называется эмпирической или описательной регрессией. Если η линейна как функция неизвестных параметров, то эмпирическая регрессия называется линейной, в противном случае — нелинейной. Простейшим примером линейной эмпирической регрессии является функция

$$\eta(x) = \eta(x_1, \dots, x_m, \theta_1, \dots, \theta_m) = \theta_1 x_1 + \dots + \theta_m x_m, \quad (1.3)$$

где $\theta_1, \dots, \theta_m$ — неизвестные параметры регрессии.

В частном случае, когда $x_1 \equiv 1$, функция регрессии (1.3) после перенумерации принимает вид

$$\eta(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_{m-1} x_{m-1}.$$

Если порядок модели равен 1, то есть имеются только две переменные y и x_1 , то эмпирические данные могут быть представлены графически.

Рассмотрим следующую задачу. Допустим, мы хотим найти зависимость между ростом (y) и весом (x_1) для определенной группы взрослых мужчин. На рис. 1 данные для 35 мужчин представлены в

виде точек (y_j, x_{1j}) . Отметим, что существуют такие значения роста, для которых встречаются разные значения веса, и наоборот. Таким образом, в данном случае мы не можем говорить о функциональной зависимости между y и x_1 . Однако функция, определяющая, как средний вес зависит от значений роста, существует и может быть аппроксимирована прямой $y = \theta_0 + \theta_1 x_1$. Подбор прямой линии к набору измерений — одна из наиболее часто встречающихся процедур, применяемых на практике. Двумерные графики, подобные рис. 1, широко используются для исследования статических зависимостей.

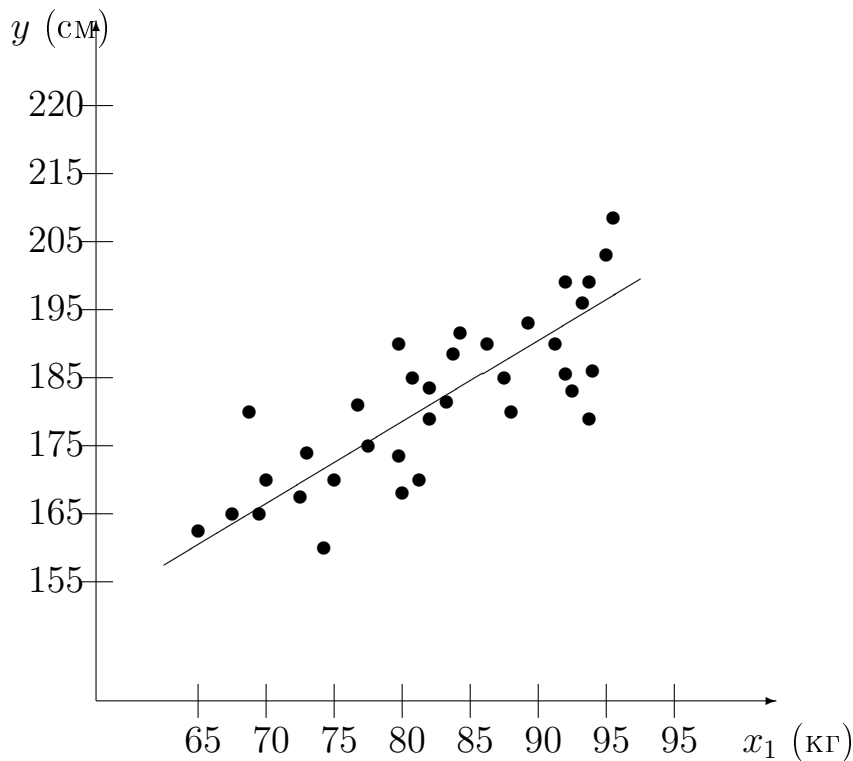


Рис. 1. Данные о соотношении веса и роста для 35 мужчин.

Следующей задачей после выбора типа функции η является задача нахождения значений неизвестных параметров $\theta_1, \dots, \theta_m$, при которых эмпирическая регрессия будет достаточно хорошо описывать

экспериментальные данные. Для этого необходимо выбрать критерий, позволяющий оценить степень соответствия эмпирической регрессии этим данным. Данный критерий должен учитывать отклонения между измеряемыми величинами $y_1, \dots, y_N$ и приближенными значениями $\tilde{y}_j$:

$$\epsilon_j = y_j - \tilde{y}_j, \quad j = 1, \dots, N, \quad (1.4)$$

где $\tilde{y}_j = \sum_{i=1}^m x_{ji}\theta_i$, $j = 1, \dots, N$.

Формулу (1.4) можно переписать следующим образом:

$$y_j = \sum_{i=1}^m x_{ji}\theta_i + \epsilon_j, \quad j = 1, \dots, N,$$

или же в матричной форме

$$y = X\theta + \epsilon, \quad (1.5)$$

где θ , $\theta = (\theta_1, \dots, \theta_m)^T$, — вектор неизвестных параметров, а ϵ , $\epsilon = (\epsilon_1, \dots, \epsilon_N)^T$, — вектор отклонений. Из (1.4) следует, что вектор ϵ и соответственно качество регрессионного описания при заданной матрице X зависит от значения вектора параметров θ . Существуют различные критерии, позволяющие найти «оптимальное» значение этого параметра. Например, наилучшими могут считаться те, для которых минимальна одна из величин

$$\max_{j=1, \dots, N} |\epsilon_j|, \quad \sum_{j=1}^N |\epsilon_j|, \quad \epsilon^T \epsilon.$$

Наиболее изучен и наибольшее распространение на практике получил третий критерий. Это связано с тем, что в этом случае формулы расчета для значений параметра θ относительно просты с вычислительной точки зрения, а сами значения обладают определенными свойствами оптимальности.

Определение 1.1. Вектор

$$\hat{\theta} = \arg \min_{\theta \in R^m} (Y - X\theta)^T (Y - X\theta) = \arg \min_{\theta \in R^m} \epsilon^T \epsilon \quad (1.6)$$

называется (эмпирической) оценкой метода наименьших квадратов (МНК-оценкой).

Справедлива следующая лемма.

Лемма 1.1. Для любой матрицы X и вектора Y соответствующих размерностей система уравнений

$$X^T X \theta = X^T Y \quad (1.7)$$

имеет решение. Причем вектор θ^* , удовлетворяющий системе, является МНК-оценкой.

Доказательство леммы 1.1.

Для произвольной матрицы A векторное пространство, порожденное ее столбцами, обозначим за $\mathcal{L}(A)$. Покажем, что

$$\mathcal{L}(A^T) = \mathcal{L}(A^T A). \quad (1.8)$$

Действительно, для любого вектора b , ортогонального пространству $\mathcal{L}(A^T)$, имеет место равенство $b^T A^T = \mathbf{0} \Rightarrow b^T A^T A = \mathbf{0}$.

В обратную сторону, если вектор b ортогонален пространству $\mathcal{L}(A^T A)$, т.е. $b^T A^T A = \mathbf{0}$, то выполнено $\mathbf{0} = b^T A^T A b = (A b)^T A b$, и следовательно, $b A^T = \mathbf{0}$.

Из равенства (1.8) автоматически следует, что для любого вектора $X^T Y \in \mathcal{L}(X^T)$ существует вектор $X^T X \theta^* \in \mathcal{L}(X^T X)$ такой, что $X^T X \theta^* = X^T Y$.

Покажем теперь, что любое решение уравнения (1.7) является МНК-оценкой, т.е. минимизирует функцию

$$(Y - X\theta)^T (Y - X\theta).$$

Обозначив за $\hat{\theta}$ произвольное решение (1.7), имеем

$$\begin{aligned} (Y - X\theta)^T (Y - X\theta) &= \\ &= (Y - X\hat{\theta} + X(\hat{\theta} - \theta))^T (Y - X\hat{\theta} + X(\hat{\theta} - \theta)) \stackrel{*}{=} \\ &\stackrel{*}{=} (Y - X\hat{\theta})^T (Y - X\hat{\theta}) + (\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta) \geq \\ &\geq (Y - X\hat{\theta})^T (Y - X\hat{\theta}). \end{aligned}$$

Равенство «*» выполнено в силу того, что $\hat{\theta}$ является решением уравнения (1.7), и следовательно,

$$(\hat{\theta} - \theta)^T X^T (Y - X\hat{\theta}) = (\hat{\theta} - \theta)^T (X^T Y - X^T X \hat{\theta}) = 0.$$

Таким образом, минимум $(Y - X\theta)^T (Y - X\theta)$ достигается при $\theta = \hat{\theta}$. Лемма доказана. □

Замечание 1.1. Если матрица $X^T X$ — невырожденная, то система (1.7) имеет единственное решение:

$$\hat{\theta} = (X^T X)^{-1} X^T Y.$$

1.2. Математическая модель и предпосылки регрессионного анализа

Математической моделью описательной регрессии является классическая регрессионная модель

$$y_j = \sum_{i=1}^m x_{ji}\theta_i + \epsilon_j, \quad j = 1, \dots, N, \quad (1.9)$$

или же в матричной форме

$$Y = X\theta + \epsilon, \quad (1.10)$$

где $Y = (y_1, \dots, y_N)^T$ — вектор результатов измерений,

$$X = (x_{ij})_{i=1, j=1}^{m, N} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{Nm} \end{pmatrix}$$

— матрица плана, $\epsilon = (\epsilon_1, \dots, \epsilon_N)^T$ — вектор случайных ошибок наблюдений, $\theta = (\theta_1, \dots, \theta_m)^T$ — вектор неизвестных коэффициентов.

В дальнейшем мы будем использовать для модели (1.10) следующее обозначение: $(Y, X\theta, \Sigma)$ где Σ — фиксированная, $N \times N$ -дисперсионная матрица случайных ошибок $\left(\Sigma = E\epsilon\epsilon^T = \|\epsilon_i\epsilon_j\|_{i,j=1}^N\right)$.

Для модели (1.10) имеют место следующие стандартные предположения об ошибках (см., например, [3], [38]):

- (1) несмещенность $E\epsilon_i = 0$;
- (2) равноточность $E\epsilon_i^2 = \sigma^2$;
- (3) некоррелированность ошибок $E\epsilon_i\epsilon_j = 0$ ($i \neq j$);

и требования к оценкам:

- (а) несмещенность $E\hat{\theta} = \theta$;
- (б) минимальность дисперсии: для любого вектора z соответствующей размерности и несмещенной оценки $\tilde{\theta}$ выполнено неравенство $D(z^T(\hat{\theta} - \theta)) \leq D(z^T(\tilde{\theta} - \theta))$;
- (с) линейность: $\hat{\theta} = SY$, где S некоторая матрица, не зависящая от Y .

Прокомментируем условие (б). Обозначим за $\tilde{D} = D_{\tilde{\theta}}$ дисперсию оценки $E(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T$. Имеем

$$D(z^T(\tilde{\theta} - \theta)) = E(z^T(\tilde{\theta} - \theta))^2 = Ez^T(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^Tz = z^T\tilde{D}z.$$

Последнее равенство выполняется в силу несмещенности оценки $\tilde{\theta}$. Таким образом, требование (б) означает, что выполнено неравенство $z^T\hat{D}z \leq z^T\tilde{D}z$, т.е. матрица $\tilde{D} - \hat{D}$ положительно определена. Отметим, что оценка, удовлетворяющая условию (б), не всегда существует.

Определение 1.2. Оценка, удовлетворяющая требованиям (а), (б) и (с), называется *наилучшей линейной несмещенной оценкой (НЛН-оценкой)*.

Пример 1.1. Рассмотрим линейную регрессионную модель. Пусть

$$\eta(t) = a + bt.$$

Эта модель описывает, в частности, изменение длины металлического стержня в зависимости от температуры.

Пусть измерения проводятся в точках $t_1, \dots, t_N$:

$$y_j = a + bt_j + \epsilon_j, \quad j = 1, \dots, N.$$

Представим это уравнение в матричной форме:

$$Y = X\theta + \epsilon, \quad \theta = (a, b)^T, \quad \epsilon = (\epsilon_1, \dots, \epsilon_N)^T,$$

$$X\theta = X \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a + bt_1 \\ \vdots \\ a + bt_N \end{pmatrix}.$$

1.3. Теорема Гаусса—Маркова

Пусть задана классическая регрессионная модель $(Y, X\theta, \Sigma)$ и матрица $X^T X$ — невырожденная.

Определение 1.3. Вектор

$$\hat{\theta} = (X^T X)^{-1} X^T Y \tag{1.11}$$

называется *(стандартной) оценкой метода наименьших квадратов (МНК-оценкой)*.

Слово «стандартная» в дальнейшем будет опускаться.

Пример 1.2. Рассмотрим для примера линейную по параметрам модель

$$y_j = \sum_{i=1}^m x_{ji} \theta_i + \epsilon_j = a + bt_j + \epsilon_j, \quad j = 1, \dots, N,$$

где $t_1, \dots, t_N$ — некоторые заданные вещественные числа.

$$X = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix}, \quad X^T X = \begin{pmatrix} N & \sum_i t_i \\ \sum_i t_i & \sum_i t_i^2 \end{pmatrix}.$$

Применяя известную формулу для обратной матрицы, получаем

$$(X^T X)^{-1} = \begin{pmatrix} \frac{\sum_i t_i^2}{\Delta} & -\frac{\sum_i t_i}{\Delta} \\ -\frac{\sum_i t_i}{\Delta} & \frac{N}{\Delta} \end{pmatrix}, \quad \Delta = N \sum_i t_i^2 - \left(\sum_i t_i \right)^2$$

и

$$X^T Y = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_N \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i y_i t_i \end{pmatrix}.$$

Искомая оценка

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} \frac{\sum_i t_i^2 \sum_i y_i - \sum_i t_i \sum_i y_i t_i}{\Delta} \\ \frac{N \sum_i y_i t_i - \sum_i t_i \sum_i y_i}{\Delta} \end{pmatrix}.$$

Пусть задана классическая модель $(Y, X\theta, \sigma^2 I_N)$, I_N — единичная матрица размера N и, кроме того, $E\epsilon = 0$ (т.е. ошибки ϵ_i удовлетворяют требованиям (1)–(3)). Предположим также, что матрица $X^T X$ — невырожденная. Тогда имеет место следующая теорема.

Теорема 1.1 (Гаусса—Маркова). *При сформулированных выше предположениях оценка метода наименьших квадратов $\hat{\theta} = (X^T X)^{-1} X^T Y$ является наилучшей линейной оценкой в классе несмещенных оценок. Ковариационная матрица этой оценки имеет вид*

$$D_{\hat{\theta}} = \sigma^2 (X^T X)^{-1}.$$

Для доказательства нам потребуются следующие вспомогательные утверждения.

Лемма 1.2. *МНК-оценка является линейно несмещенной оценкой (т.е. для нее выполнены требования (а) и (с)).*

Доказательство. Линейность МНК-оценки очевидна ($S = (X^T X)^{-1} X^T$), проверим несмещенность. Посчитаем математическое ожидание:

$$E\hat{\theta} = ESY = SEY = SE(X\theta + \epsilon) = SX\theta = (X^T X)^{-1} X^T X\theta = \theta.$$

Лемма доказана. □

Теперь сформулируем более общее утверждение о несмещенности оценки.

Лемма 1.3. *Для того чтобы оценка $\tilde{\theta} = AY$ была несмещенной, необходимо и достаточно, чтобы выполнялось равенство*

$$AX = I.$$

Эта лемма является основным инструментом доказательства теоремы Гаусса—Маркова.

Доказательство. Достаточность. Пусть $AX = I$, тогда

$$E\tilde{\theta} = EAY = AE(X\theta + \epsilon) = \underbrace{AX}_I \theta = \theta.$$

Таким образом, оценка $\tilde{\theta} = AY$ является несмещенной оценкой.

Необходимость. Пусть теперь выполнено $E\tilde{\theta} = \theta$, т.е. оценка — несмещенная. Получаем $E\tilde{\theta} = AX\theta = \theta$. Следовательно,

$$AX\theta = \theta.$$

Так как это равенство должно быть выполнено для любого вектора θ , можем в качестве него выбрать вектор $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ с единицей на i -м месте. Прделаав это для всех $i = 1, \dots, t$, получим систему уравнений $AX = I$, что завершает доказательство.

□

Лемма 1.4. *В предположениях теоремы 1.1 дисперсионная матрица любой линейной несмещенной оценки $\tilde{\theta} = AY$ имеет вид*

$$D_{\tilde{\theta}} = \sigma^2 AA^T.$$

В частном случае, если $\tilde{\theta}$ — МНК-оценка, то

$$D_{\tilde{\theta}} = \sigma^2 (X^T X)^{-1}.$$

Доказательство. Посчитаем дисперсионную матрицу оценки $\tilde{\theta}$:

$$D_{\tilde{\theta}} = E(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T = E(AY - \theta)(AY - \theta)^T = EAYY^T A^T - \theta\theta^T,$$

так как $EAY = (EY^T A^T)^T = \theta$. С другой стороны,

$$\begin{aligned} EAYY^T A^T &= AE(YY^T)A^T = AE(X\theta + \epsilon)(X\theta + \epsilon)^T A^T = \\ &= \underbrace{AX}_I \theta\theta^T \underbrace{X^T A^T}_I + \sigma^2 AA^T = \theta\theta^T + \sigma^2 AA^T. \end{aligned}$$

Таким образом, получаем, что $D_{\tilde{\theta}} = \sigma^2 AA^T$ и, в частности, в случае МНК-оценки (т.е. при $A = (X^T X)^{-1} X$)

$$D_{\hat{\theta}} = \sigma^2 AA^T = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.$$

Лемма доказана.

□

Вернемся теперь к доказательству теоремы.

Доказательство теоремы 1.1. Проверим, что выполняется неравенство

$$D_{\hat{\theta}} \leq D_{\tilde{\theta}}.$$

Обозначим матрицу $(X^T X)^{-1} X^T$ за S . Для произвольной несмещенной оценки $\tilde{\theta} = AY$ имеем

$$\begin{aligned} D_{\tilde{\theta}} = D(AY) &= D(AY - SY + SY) \stackrel{*}{=} D(AY - SY) + D(SY) = \\ &= D(AY - SY) + D_{\hat{\theta}} \geq D_{\hat{\theta}}; \end{aligned}$$

равенство «*» выполняется в силу того, что $(A - S)X = I - I = 0$, и следовательно,

$$\begin{aligned} E((A - S)YY^T S^T) &= (A - S)(X\theta\theta^T X^T + \sigma^2 I)S^T = (A - S)S^T \sigma^2 = \\ &= (A - S)X(X^T X)^{-1} \sigma^2 = 0. \end{aligned}$$

Теорема 1.1 доказана. □

Пример 1.3. Предположим, результаты измерений определяются следующей функцией:

$$y = \frac{2}{4 - 3x}.$$

Подчеркнем, что знание об этой функции будем использовать не при построении модели, а лишь для исследования ее свойств. Пусть проведены измерения в точках $x_1 = 0$, $x_2 = 2/3$, $x_3 = 1$, а модель, используемая для аппроксимации истинной зависимости, — квадратичная ($\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2$). При этом будем предполагать отсутствие ошибок измерений (что, конечно, является искусственным

допущением). Запишем, чему равны вектор Y и матрица X :

$$Y = \begin{pmatrix} 1/2 \\ 1 \\ 2 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \frac{2}{3} & \frac{4}{9} \\ 1 & 1 & 1 \end{pmatrix}.$$

Найдем МНК-оценку. Вычислим матрицу $X^T X$:

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ 0 & \frac{2}{3} & 1 \\ 0 & \frac{4}{9} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & \frac{2}{3} & \frac{4}{9} \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 5/3 & 13/9 \\ 5/3 & 13/9 & 35/27 \\ 13/9 & 35/27 & 97/81 \end{pmatrix},$$

и обратную к ней

$$(X^T X)^{-1} = \begin{pmatrix} 1 & -5/2 & 3/2 \\ -5/2 & 61/2 & -30 \\ 3/2 & -30 & 63/2 \end{pmatrix}.$$

По определению оценка $\hat{\theta}$, построенная по методу наименьших квадратов, равна

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 1/2 \\ -3/4 \\ 9/4 \end{pmatrix}.$$

Таким образом, мы получили приближение нашей исходной функции $\frac{2}{4-3x}$ квадратичной функцией, рассчитав ее коэффициенты по методу наименьших квадратов (см. рисунок 2). Отметим, что в этом случае полученная нами аппроксимирующая функция совпадает с интерполяционным полиномом Лагранжа, построенным по тем же узловым точкам.

Обобщим полученные результаты на случай векторной параметрической функции $\tau = T\theta$, где T есть $k \times m$ -матрица, $k \in \{1, \dots, m\}$.

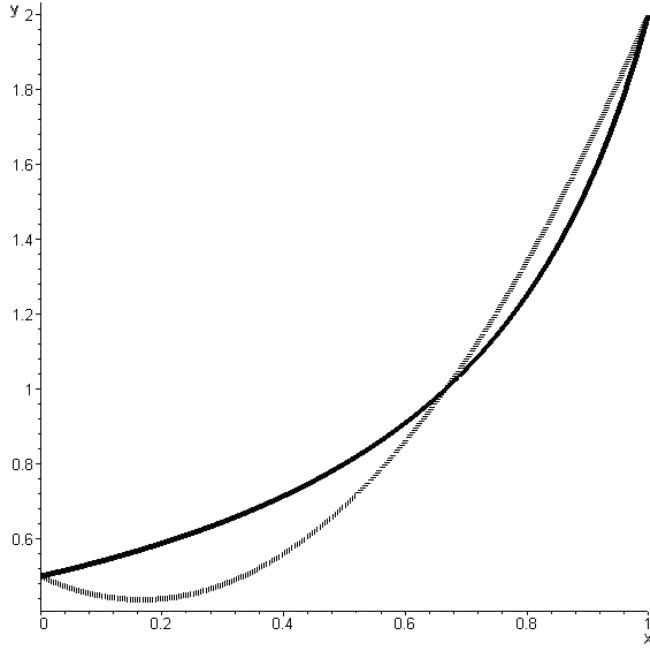


Рис. 2. Приближение функции $y = \frac{2}{4-3x}$ (сплошная линия) функцией $\bar{y} = 1/2 - 3/4x + 9/4x^2$ (штриховая линия), построенной по методу наименьших квадратов.

Определение 1.4. Векторная параметрическая функция $\tau = T\theta$ ($T \in R^{k \times m}$, $k \in \{1, \dots, m\}$) называется *оцениваемой*, если для нее существует несмещенная оценка вида $\hat{\tau} = AY$ ($A \in R^{k \times N}$).

Справедливо следующее утверждение.

Лемма 1.5. Для классической регрессионной модели (1.10), вектора ошибок, удовлетворяющего условию $E\epsilon = 0$, и некоторой фиксированной величины $k \in \{1, \dots, m\}$ следующие утверждения эквивалентны:

- (а) строки матрицы T принадлежат линейному пространству $\mathcal{L}(X^T) = \mathcal{L}(X^T X)$;

(b) вектор $T\hat{\theta}$ один и тот же для всех $\hat{\theta}$ решений нормального уравнения (1.7);

(c) параметрическая функция $\tau = T\theta$ ($T \in R^{k \times m}$) оцениваема.

Замечание 1.2. Отметим, что мы не делаем никаких предположений о ранге матрицы X . Это означает, что лемма 1.5 верна и для вырожденной матрицы $X^T X$.

Доказательство. Доказательство будем проводить для случая $k = 1$. В общем случае лемма доказывается аналогично. Схема доказательства следующая: $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

Пусть $T^T \in \mathcal{L}(X^T X)$. Это означает, что существует вектор p такой, что $T^T = X^T X p$. Пусть $\hat{\theta}$ — некоторое решение нормального уравнения. Тогда

$$T\hat{\theta} = p^T X^T X \hat{\theta} = p^T X^T Y,$$

т.е. оценка $T\hat{\theta}$ не зависит от $\hat{\theta}$. Кроме того,

$$ET\hat{\theta} = p^T X^T EY = p^T X^T X\theta = T\theta,$$

что означает, что функция $\tau = T\theta$ оцениваема. Осталось проверить, что из $(c) \Rightarrow (a)$. Пусть функция $\tau = T\theta$ оцениваема. По определению существует матрица A такая, что $EAY = T\theta$. Тогда $AX\theta = T\theta$, откуда следует, что $AX = T$ (см. доказательство леммы 1.3), и следовательно, $T^T \in \mathcal{L}(X^T)$. Лемма доказана.

□

Из этой леммы и теоремы Гаусса—Маркова вытекает следующая теорема:

Теорема 1.2. Если матрица $X^T X$ — невырожденная и выполнены предпосылки линейного регрессионного анализа, то для любой матрицы T соответствующей размерности функция $\tau = T\theta$ оценивается, и ее наилучшей линейной несмещенной оценкой является вектор

$$\hat{\tau} = T\hat{\theta} = T(X^T X)^{-1} X^T Y.$$

Ковариационная матрица оценки $\hat{\tau}$ имеет вид

$$D_{\hat{\tau}} = \sigma^2 T(X^T X)^{-1} T^T.$$

Далее, мы обобщим эту теорему на случай, когда матрица $X^T X$ вырожденная.

1.4. МНК-оценка в вырожденном случае

В предыдущем параграфе мы рассматривали задачу нахождения МНК-оценки в случае, когда матрица $X^T X$ — невырожденная. В данном параграфе мы обобщим полученные результаты для вырожденного случая, т.е. случая, когда $\text{rank } X^T X < m$. Заметим, что этот случай очень важен. В частности, матрица $X^T X$ является вырожденной в задачах дисперсионного анализа, которые будут рассмотрены в дальнейшем.

Итак, пусть $X^T X$ — вырожденная матрица. В этом случае под МНК-оценкой, по-прежнему, понимается вектор, минимизирующий $\|Y - X\theta\|^2$ и, согласно лемме 1.1, являющийся решением нормального уравнения

$$X^T X \theta = X^T Y.$$

Как уже было показано ранее, это уравнение совместно, однако в вырожденном случае оно имеет бесконечное множество решений, т.е. вектор $\hat{\theta}$, являющийся решением, определен не однозначно. Он выражается через так называемые обобщенно-обратные матрицы.

Пусть имеется произвольная система уравнений

$$Ax = y,$$

где $x \in R^n, y \in R^m$ — векторы, $A \in R^{n \times m}$ — матрица соответствующей размерности (не обязательно квадратная).

Определение 1.5. *Обобщенно-обратной* для матрицы A называется матрица порядка $n \times m$, обозначаемая A^- , такая, что для любого вектора y , для которого система $Ax = y$ совместна, вектор $x = A^-y$ является ее решением.

Для начала отметим одно важное свойство, которым обладает матрица A^- , а потом ответим на вопрос о ее существовании.

Лемма 1.6. *Для того чтобы матрица B являлась обобщенно-обратной матрицей для матрицы A , необходимо и достаточно, чтобы B удовлетворяла равенству*

$$ABA = A.$$

Доказательство. Пусть для матрицы A существует обобщенно-обратная матрица $B = A^-$. Выберем в качестве y i -й столбец матрицы A . Система

$$Ax = a_i,$$

очевидно, совместна. Ее решение существует. Например, $x = e_i$. Вектор $x = A^-a_i$ по определению является решением этой системы. Из этого следует, что

$$Ax = AA^-a_i = a_i, \quad \forall i,$$

что, в свою очередь, означает, что выполнено равенство $AA^-A = A$. Пусть теперь система $Ax = y$ совместна и матрица B удовлетворяет равенству

$$ABA = A.$$

Тогда, умножая обе части этого равенства справа на x , получим

$$AB \underbrace{Ax}_y = \underbrace{Ax}_y \Rightarrow ABx = x,$$

и следовательно, вектор $x = By$ является решением системы $Ax = y$. Это означает, по определению, что матрица B является обобщенно-обратной для A . То есть $B = A^-$. Лемма доказана.

□

Теперь рассмотрим вопрос существования обобщенно-обратной матрицы. Справедлива следующая лемма.

Лемма 1.7. *Для всякой матрицы A существует обобщенно-обратная матрица A^- .*

Доказательство. Сперва рассмотрим случай, когда A — симметричная квадратная матрица. По теореме о спектральном разложении в этом случае матрицу A можно представить в виде

$$A = P\Lambda P^T, \quad PP^T = P^TP = I,$$

Λ — матрица, на диагонали которой стоят собственные числа матрицы A , а внедиагональные элементы равны 0. Без ограничения общности можно считать, что матрица Λ имеет вид

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix},$$

где Λ_1 имеет ненулевые диагональные элементы. Положим,

$$A^- = P\Lambda^-P^T, \quad \Lambda^- = \begin{pmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Проверим, что $AA^-A = A$:

$$AA^-A = P\Lambda P^T P\Lambda^-P^T P\Lambda P^T = P\Lambda\Lambda^-\Lambda P^T = P\Lambda P^T = A.$$

Следовательно, по лемме 1.6 матрица A^- является обобщенно-обратной.

Для произвольной матрицы A существует представление $A = B\Lambda C$, где B и C — невырожденные матрицы, Λ — диагональная. Так же, как и в предыдущем случае, положим

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Lambda^- = \begin{pmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad A^- = C^{-1}\Lambda^-B^{-1}.$$

Несложно проверить, что для A^- выполнено равенство $AA^-A = A$:

$$AA^-A = B\Lambda C C^{-1}\Lambda^-B^{-1}B\Lambda C = B\Lambda\Lambda^-\Lambda C = B\Lambda C = A.$$

Следовательно, по лемме 1.6 матрица A^- является обобщенно-обратной. Лемма доказана. □

Справедлива следующая теорема.

Теорема 1.3. Пусть A^- — какая-либо обобщенно-обратная для A матрица. И пусть $H = A^-A$. Тогда

- (a) $H^2 = H$, т.е. матрица H идемпотентная;
- (b) $AH = A$ и $\text{rank}(A) = \text{rank}(H) = \text{tr}(H)$;
- (c) общее решение системы уравнений $Ax = 0$ имеет вид $x = (H - I)z$, где z — произвольный вектор;
- (d) общее решение совместной системы уравнений $Ax = y$ имеет вид $x = A^-y + (H - I)z$, где z — произвольный вектор;
- (e) произведение Tx принимает единственное значение для всех x , удовлетворяющих системе $Ax = y$, тогда и только тогда, когда $TH = T$.

Доказательство. Пункты (a)—(e) следуют из леммы 1.6. Прокомментируем вторую часть пункта (b). Пусть $\text{rank}(A) = k$,

$$\begin{aligned} \text{tr}A^-A &= \text{tr}C^{-1}\Lambda^{-}B^{-1}B\Lambda C = \text{tr}C^{-1}\Lambda^{-}\Lambda C = \text{tr}\Lambda^{-}\Lambda CC^{-1} = \\ &= \text{tr}\Lambda^{-}\Lambda = \text{tr}I_k = k = \text{rank}(A), \end{aligned}$$

где I_k — диагональная матрица с k единицами на главной диагонали ($k = \text{rank}(A)$). Остальные ее элементы — нули. Мы воспользовались известным результатом $\text{tr}AB = \text{tr}BA$ (для любых согласованных матриц A, B). Теорема доказана.

□

В случае, когда матрица X имеет ранг меньший m , невозможно несмещенно оценить все параметры регрессионной модели $(Y, X\theta, \Sigma)$. Однако можно оценить некоторые параметрические функции $\tau = T\theta$. Вид оценки и условия оцениваемости определяются в этом случае в следующей теореме, являющейся аналогом теоремы Гаусса—Маркова для вырожденной матрицы $X^T X$.

Теорема 1.4. Для классической регрессионной модели (1.10) и вектора ошибок, удовлетворяющего условию $E\epsilon = 0$,

(1) параметрическая функция $\tau = T\theta$ ($T \in R^{k \times m}$, $k \in \{1, \dots, m\}$) оценивается тогда и только тогда, когда

$$T(X^T X)^- X^T X = T;$$

(2) если выполнено условие (1), оценка для τ имеет вид

$$\hat{\tau} = T(X^T X)^- X^T Y,$$

определена однозначно и является наилучшей линейной несмещенной оценкой. Ковариационная матрица оценки $\hat{\tau}$ имеет вид

$$D_{\hat{\tau}} = \sigma^2 D, \quad D = T(X^T X)^- T^T.$$

Данная теорема следует непосредственно из леммы 1.5, теоремы 1.2 и теоремы 1.3.

Пример 1.4. Рассмотрим схему регрессии для задачи взвешивания трех предметов на двухчашечных весах. Предполагаем, что результат i -го взвешивания есть разность веса содержимого первой и второй чаш плюс некоторая случайная ошибка. Будем считать, что ошибка удовлетворяет предпосылкам (1)–(3). В матричной форме уравнение регрессии для этой задачи можно записать как

$$Y = X\theta + \epsilon,$$

где

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & x_{N3} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix},$$

при этом

$$x_{ij} = \begin{cases} 1, & \text{при } i\text{-м взвешивании} \\ & j\text{-й груз на первой чаше весов;} \\ 0, & \text{груз не участвует в } i\text{-м взвешивании;} \\ -1, & \text{при } i\text{-м взвешивании} \\ & j\text{-й груз на второй чаше весов.} \end{cases}$$

Пусть $N=3$ и измерения проводятся в следующих точках: $(0, 1, -1)$, $(-1, 1, 0)$, $(-1, 0, 1)$. Это означает, что матрица X имеет вид

$$X = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix},$$

$$X^T X = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}, \quad X^T Y = \begin{pmatrix} y_3 - y_2 \\ y_1 + y_2 \\ -y_1 - y_3 \end{pmatrix},$$

следовательно система нормальных уравнений $(X^T X \theta = X^T Y)$ имеет вид

$$\begin{aligned} 2\theta_1 - \theta_2 - \theta_3 &= y_3 - y_2, \\ -\theta_1 + 2\theta_2 - \theta_3 &= y_1 + y_2, \\ -\theta_1 - \theta_2 + 2\theta_3 &= -y_1 - y_3. \end{aligned}$$

Решая эту систему, получаем

$$\hat{\theta}_1 = \frac{2y_3 - y_2 + y_1}{3} + \theta_3, \quad \hat{\theta}_2 = \frac{y_2 + 2y_1 + y_3}{3} + \theta_3;$$

значение параметра θ_3 можно выбрать произвольно.

Теперь найдем оценку параметров θ_1, θ_2 с помощью теоремы 1.4. В качестве матриц $(X^T X)^{-}$ и T выберем соответственно матрицы

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Несложно проверить, что $(X^T X)^{-}$ удовлетворяет равенству из леммы 1.6, а T — условию (1) теоремы 1.4. Следовательно,

$$\hat{\tau} = T(X^T X)^{-} X^T Y = \begin{pmatrix} \frac{2y_3 - y_2 + y_1}{3} \\ \frac{y_2 + 2y_1 + y_3}{3} \\ 0 \end{pmatrix}$$

или, что то же самое,

$$\hat{\theta}_1 = \frac{y_2 - y_1 - 2y_3}{3}, \quad \hat{\theta}_2 = \frac{2y_1 + y_3 + y_2}{3}.$$

1.5. Оценка максимального правдоподобия

Рассмотрим классическую линейную регрессионную модель (1.9). Вектор Y представляет собой выборку из распределения с некоторой плотностью $L(Y, \theta)$, зависящей от неизвестных параметров. Эту функцию в математической статистике называют функцией максимального правдоподобия.

Определение 1.6. Величина $\hat{\theta}_{\text{ОМП}}$, максимизирующая значение функции $L(Y, \theta)$, называется *оценкой максимального правдоподобия*.

Теорема 1.5. Пусть дана классическая регрессионная модель:

$$Y = X\theta + \epsilon, \quad \epsilon \in N(0, \sigma^2 I_N). \quad (1.12)$$

Тогда оценка максимального правдоподобия также является МНК-оценкой, а величина

$$\hat{s}_{омп}^2 = \frac{1}{N} \left(Y - X\hat{\theta} \right)^T \left(Y - X\hat{\theta} \right)$$

— оценкой максимального правдоподобия для σ^2 .

Доказательство. Функция максимального правдоподобия $L(Y, \theta, \sigma^2)$ для нормально распределенного вектора Y имеет вид

$$L(Y, \theta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left\{ -\frac{1}{2\sigma^2} (Y - X\theta)^T (Y - X\theta) \right\}.$$

Задачу максимизации решать значительно проще, если перейти к логарифму этой функции:

$$\ln L(Y, \theta, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (Y - X\theta)^T (Y - X\theta).$$

Продифференцируем полученное выражение по θ и σ^2 и приравняем производные к нулю. Получаем

$$\sigma^2 \frac{\partial \ln L(Y, \theta, \sigma^2)}{\partial \theta} = X^T (Y - X\theta) = X^T Y - X^T X \theta = 0,$$

$$2\sigma^4 \frac{\partial \ln L(Y, \theta, \sigma^2)}{\partial (\sigma^2)} = -N\sigma^2 + (Y - X\theta)^T (Y - X\theta) = 0.$$

Так как первая система — это система нормальных уравнений, по лемме 1.1 вектор $\hat{\theta}$, являющийся ее решением, есть МНК-оценка. Решением второй системы является величина $\hat{s}_{омп}^2$. Теорема доказана.

□

Замечание 1.3. Оценка максимального правдоподобия $\hat{s}_{\text{ОМП}}^2$ является асимптотически несмещенной, т.е.

$$E\hat{s}_{\text{ОМП}}^2 \xrightarrow{N \rightarrow \infty} \sigma^2.$$

Из теоремы 1.5 следует известный результат, заключающийся в том, что для одномерной повторной выборки из нормального распределения с неизвестным средним θ и дисперсией σ^2 оценкой максимального правдоподобия для θ является среднее выборочное $\bar{y}$, а для σ^2 — выборочная дисперсия $\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$. Действительно, пусть дана модель

$$Y = X\theta + \epsilon, \quad \epsilon \in N(0, \sigma^2 I_N),$$

$Y = (y_1, y_2, \dots, y_N)^T$, $X = (1, 1, \dots, 1)^T$, θ — неизвестный параметр. Тогда $X^T X = N$, $(X^T X)^{-1} = 1/N$, $X^T Y = \sum_{i=1}^N y_i$ и, согласно теореме, оценка максимального правдоподобия будет

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y}.$$

Аналогично, оценка максимального правдоподобия для дисперсии есть

$$\hat{s}_{\text{ОМП}}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2.$$

1.6. Обобщенный метод наименьших квадратов

Определение 1.7. Модель $(Y, X\theta, \sigma^2 W)$, где W — известная положительно определенная $N \times N$ -матрица, σ^2 — параметр (не обязательно известный), называется *обобщенной линейной регрессионной моделью*.

На практике обобщенная регрессионная модель достаточно часто используется в тех случаях, когда проводятся повторные измерения в тех же точках. Допустим, имеются результаты измерений

$$y_i = x_{(i)}\theta + \epsilon_i,$$

где $x_{(i)} = (x_{i1}, \dots, x_{im})$ — точки, в которых проводятся измерения, ϵ_i — некоррелированные случайные ошибки с одинаковой дисперсией σ^2 .

Пусть среди N точек $x_{(i)}$ только M ($M \leq N$) точек различные. Обозначим эти точки через $t_{(1)}, \dots, t_{(M)}$, а через r_k — число измерений в точке $t_{(k)}$, $r_k \geq 1$, $\sum_{k=1}^M r_k = N$. Усреднив результаты различных измерений в точках $t_{(k)}$, получим модель

$$\tilde{y}_i = t_{(i)}\theta + \tilde{\epsilon}_i, \quad i = 1, \dots, M,$$

где $\tilde{y}_i$ — средние арифметические результатов наблюдений y_j в точках $t_{(i)}$, $\tilde{\epsilon}_i$ — средние арифметические ошибок измерений ϵ_i . При этом $E\tilde{\epsilon}_i = 0$, $E\tilde{\epsilon}_i\tilde{\epsilon}_j = 0$ ($i \neq j$), $\tilde{\epsilon}_i^2 = \frac{\sigma^2}{r_i}$.

Полученная модель является обобщенной линейной регрессионной моделью $(\tilde{Y}, \tilde{X}\theta, \sigma^2 W)$, W — диагональная $M \times M$ -матрица с элементами $\frac{1}{r_i}$ на главной диагонали.

Замечание 1.4. Обобщенная линейная регрессионная модель эквивалентна классической модели (1.10).

Действительно, пусть имеется обобщенная регрессионная модель $(Y, X\theta, \sigma^2 W)$. В силу теоремы о спектральном разложении матрицы и того, что матрица W положительно определена, существует такая невырожденная матрица $A \in R^{N \times N}$, что $W = AA^T$. Умножив обе части (1.10) на A^{-1} и обозначив за $\tilde{Y} = A^{-1}Y$, $\tilde{X} = A^{-1}X$, $\tilde{\epsilon} = A^{-1}\epsilon$,

получаем эквивалентную модель $(\tilde{Y}, \tilde{X}\theta, \sigma^2 I_N)$ в силу того, что

$$\begin{aligned} D\tilde{\epsilon} &= E\tilde{\epsilon}\tilde{\epsilon}^T = EA^{-1}\epsilon\epsilon^T(A^{-1})^T = A^{-1}\sigma^2 W(A^{-1})^T = \\ &= \sigma^2 A^{-1}AA^T(A^{-1})^T = \sigma^2 I. \end{aligned}$$

Если матрица $\tilde{X}^T\tilde{X}$ — невырожденная, то по определению МНК-оценка для этой модели есть

$$\begin{aligned} \hat{\theta} &= (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} = (X^T(AA^T)^{-1}X)^{-1}X^T(AA^T)^{-1}Y = \\ &= (X^TW^{-1}X)^{-1}X^TW^{-1}Y. \end{aligned}$$

Определение 1.8. Оценка

$$\hat{\theta} = (X^TW^{-1}X)^{-1}X^TW^{-1}Y$$

называется *обобщенной МНК-оценкой*.

Дисперсионная матрица этой оценки в силу леммы 1.4 имеет вид

$$D\hat{\theta} = \sigma^2(\tilde{X}^T\tilde{X})^{-1} = \sigma^2(X^TW^{-1}X)^{-1}.$$

В силу теоремы Гаусса—Маркова (теорема 1.1) обобщенная МНК-оценка является наилучшей линейной несмещенной оценкой.

1.7. Модели с априорной информацией о параметрах

На практике часто возникают ситуации, когда исследователь обладает некоторой априорной информацией о параметрах модели. Разумный учет такой информации позволяет улучшать качество регрессионного описания.

Существует несколько типов задач с априорной информацией о параметрах. Простейшим видом априорной информации являются линейные ограничения типа равенств. В этом случае задача нахождения оценки сводится к классической, рассмотренной ранее.

Бывают ситуации, когда исследователь обладает информацией о модели, полученной в ходе каких-либо предварительных измерений или, возможно, в его распоряжении имеются некоторые выборочные данные по какой-нибудь аналогичной модели. В таких случаях априорная информация, как правило, может быть представлена в виде

$$u = R\theta + \zeta, \quad \zeta \sim (0, D), \quad D > 0, \quad E\zeta\epsilon^T = 0, \quad (1.13)$$

где u — некоторый известный k -вектор, R — фиксированная $k \times m$ -матрица, D — известная, положительно определенная, дисперсионная матрица вектора ошибок ζ . Кроме того, вектора ϵ и ζ некоррелированы. Запись $\zeta \sim (0, D)$ означает, что вектор ζ имеет некоторое распределение с нулевым математическим ожиданием и дисперсионной матрицей D .

Достаточно часто встречаются задачи (см., например, [11], [2], [3], [38]), когда априорная информация задается в виде неравенств. Например, может быть дополнительно известно, что параметры модели принадлежат некоторому эллипсоиду

$$\Omega = \{\theta \in R^m | (\theta - \theta_0)^T A (\theta - \theta_0) \leq k\}, \quad (1.14)$$

где $\theta_0 \in R^m$ — центр эллипсоида, A — положительно определенная матрица, $k > 0$. При таких ограничениях на параметры используется минимаксный подход.

Иногда априорная информация выражается в том, что исследователю известно распределение вектора параметров в какой-нибудь заданной области, например Ω . В этом случае задача нахождения НЛН-оценок также сводится к классической.

Рассмотрим теперь более подробно каждый из этих типов априорной информации.

Линейные ограничения на параметры.

Пусть есть обобщенная регрессионная модель $(Y, X\theta, \sigma^2 W)$ и имеются ограничения на параметры θ вида $u = R\theta$, где R — известная $r \times m$ -матрица ранга $r < m$. Согласно теореме 1.3 общее решение системы уравнений $u = R\theta$ имеет вид $\theta = R^-u + (R^-R - I)\gamma$, γ — произвольный m -вектор. Положим $Z = Y - XR^-u$. Тогда математическое ожидание Z есть

$$EZ = X\theta - XR^-u = XR^-u + X(R^-R - I)\gamma - XR^-u = X(R^-R - I)\gamma,$$

следовательно от модели $(Y, X\theta, \sigma^2 W)$ мы можем перейти к модели $(Z, X(R^-R - I)\gamma, \sigma^2 W)$, в которой $Z = Y - XR^-u$ — новый вектор результатов измерений, а параметрами, которые мы хотим оценить, являются компоненты вектора γ . Несмещенная оценка этих параметров может быть найдена в том случае, когда матрица $X(R^-R - I)$ имеет полный ранг. Оценкой для параметров исходной модели является вектор $\hat{\theta} = R^-u + (R^-R - I)\hat{\gamma}$, где $\hat{\gamma}$ — оценка параметров γ .

Линейные ограничения со случайной ошибкой.

Пусть есть обобщенная регрессионная модель $(Y, X\theta, \sigma^2 W)$, и информация об этой модели может быть представлена в виде (1.13). В этом случае модель принято называть смешанной. Это название вытекает из того, что данная модель может быть записана в виде

$$(\tilde{Y}, \tilde{X}\theta, \sigma^2 \tilde{W}),$$

где

$$\tilde{Y} = \begin{bmatrix} Y \\ u \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} X \\ R \end{bmatrix}, \quad \tilde{W} = \begin{bmatrix} W & 0 \\ 0 & \frac{1}{\sigma^2} D \end{bmatrix}.$$

В соответствии с результатами раздела 1.6, обобщенная МНК-оценка

для этой модели имеет вид

$$\begin{aligned}\hat{\theta} &= (\tilde{X}^T \tilde{W}^{-1} \tilde{X})^{-1} \tilde{X}^T \tilde{W}^{-1} \tilde{Y} = \\ &= (X^T W^{-1} X + \sigma^2 R^T D^{-1} R^{-1})(X^T W^{-1} Y + \sigma^2 R^T D^{-1} u).\end{aligned}\quad (1.15)$$

На практике значение σ^2 обычно не известно, поэтому вместо σ^2 используется ее оценка (например, в разделе 1.8 в качестве такой оценки рассматривается s^2).

Ограничения по области.

Пусть есть обобщенная регрессионная модель $(Y, X\theta, \sigma^2 W)$, и известно, что вектор параметров θ принадлежит некоторому множеству Ω .

Определение 1.9. Минимаксной оценкой вектора параметров θ называется вектор

$$\begin{aligned}\hat{\theta} &= \arg \min_{\tilde{\theta} \in \Omega} \max_{\theta \in \Omega} g(\theta, \tilde{\theta}), \\ g(\theta, \tilde{\theta}) &= a^T \left(E((\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T) \right) a,\end{aligned}$$

где $\tilde{\theta}$ — некоторые линейные оценки θ , а $a \in R^n$ — ненулевой фиксированный вектор.

В том случае, когда $a = e_k$ (вектор с единицей на k -м месте и нулевыми остальными элементами), k -й элемент минимаксной оценки есть наилучшая оценка k -й компоненты вектора θ .

Имеет место следующая теорема.

Теорема 1.6. Пусть есть обобщенная регрессионная модель $(Y, X\theta, \sigma^2 W)$, и на параметры θ налагаются ограничения вида (1.14). Тогда минимаксная оценка может быть записана в виде

$$\hat{\theta}_{mM} = \left(\frac{\sigma^2}{k} A + X^T W^{-1} X \right)^{-1} X^T W^{-1} (Y - X\theta_0) + \theta_0. \quad (1.16)$$

Доказательство данной теоремы 1.6 может быть найдено в книге [3] (стр. 42).

Отметим, что результат, полученный в теореме, в частности, демонстрирует, что в случае, когда множество Ω имеет вид (1.14), минимаксная оценка $\hat{\theta}_{mM}$ не зависит от выбора вектора a , фигурирующего в определении.

В том случае, когда центр эллипсоида находится в начале координат, а модель является классической (т.е. $\theta_0=0$, $W = I_N$), минимаксная оценка принимает вид

$$\hat{\theta}_{mM} = \left(\frac{\sigma^2}{k} A + X^T X \right)^{-1} X^T Y.$$

В случае, когда Ω имеет вид $\Omega = \{\theta | \theta^T \theta \leq \sigma^2/k\}$, оценка $\hat{\theta}_{mM} = (kI_m + X^T X)^{-1} X^T Y$ называется гребневой.

Заметим также, что в пределе при $k \rightarrow \infty$ минимаксная оценка совпадает с обобщенной МНК-оценкой. Или, что то же самое, обобщенная МНК-оценка является минимаксной при отсутствии априорной информации. Кроме того, при $k < \infty$ минимаксная оценка, в отличие от стандартной МНК-оценки, всегда однозначно определена, вне зависимости от того, является матрица $X^T X$ вырожденной или нет.

Замечание 1.5. Минимаксная оценка, вообще говоря, является смещенной оценкой. В классе несмещенных оценок обобщенная МНК-оценка является минимаксной.

Последнее утверждение следует из того, что в классе несмещенных оценок в силу леммы 1.3 для любой оценки $\tilde{\theta} = CY$ имеем $CX = I_m$ и, соответственно, вектор

$$\tilde{\theta} - \theta = CY - \theta = C(X\theta + \epsilon) - \theta = \underbrace{(CX - I_m)}_{=0} \theta + C\epsilon = C\epsilon$$

не зависит от θ , а значит, и функция $g(\theta, \tilde{\theta}) = a^T C E(\epsilon \epsilon^T) C^T a = \sigma^2 a^T C W C^T a$ также не зависит от θ и, как было показано в 1.6, ее минимум достигается при $C = (X^T W^{-1} X)^{-1} X^T W^{-1}$ для любого $a \in R^m$.

Заданное распределение.

Предположим, что априорная информация о параметрах заключается в том, что на σ -алгебре подмножеств некоторого множества Ω задано независимое от распределения ошибок ϵ распределение $P(d\theta)$ со средним u и дисперсионной матрицей D . Или, что то же самое, на вектор параметров θ наложены ограничения вида (1.13) с матрицей $R = I_m$. Для такой матрицы оценка (1.15) принимает вид

$$\hat{\theta} = (X^T W^{-1} X + \sigma^2 D^{-1})(X^T W^{-1} Y + \sigma^2 D^{-1} u)$$

и называется *байесовской оценкой*.

1.8. Основы дисперсионного анализа

В данном разделе мы рассмотрим некоторые задачи, решаемые с помощью дисперсионного анализа (ДА). Идея ДА заключается в разложении общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак. Для этого, как правило, используется F -критерий (см., например, [1], [2], [13]).

Пусть дана классическая регрессионная модель (1.10). Будем считать, что вектор ошибок ϵ удовлетворяет предпосылкам (1)–(3) регрессионного анализа. Пусть, кроме того, вектор Y имеет нормальное распределение:

$$Y \sim N(X\theta, \sigma^2 I). \quad (1.17)$$

Как было показано в процессе доказательства леммы 1.1 дисперсия вектора Y представима в виде суммы

$$D(Y) = D(X\hat{\theta}) + D(Y - X\hat{\theta}). \quad (1.18)$$

Первая из этих дисперсий является дисперсией регрессии (объясненная доля дисперсии), вторая — дисперсией ошибки регрессии (необъясненная доля дисперсии). Чтобы оценить качество регрессионного описания, необходимо выяснить, каков вклад в общую сумму каждой из ее составляющих. Для решения этой задачи на практике используют выпуклые функционалы, заданные на множестве дисперсионных матриц. Мы остановимся на функционале $\Upsilon(A) = \text{tr}A$, благодаря своей простоте получившему широкое распространение. Для разложения (1.18) получаем

$$\begin{aligned} \text{tr}D(Y) &= \text{tr}D(X\hat{\theta}) + \text{tr}D(Y - X\hat{\theta}) = \\ &= E(X\hat{\theta} - X\theta)^T(X\hat{\theta} - X\theta) + E(Y - X\hat{\theta})^T(Y - X\hat{\theta}). \end{aligned}$$

Или, что то же самое,

$$\text{tr}D(Y) = \sum_{i=1}^N E(X_{(i)}\hat{\theta} - X_{(i)}\theta)^2 + \sum_{i=1}^N E(Y_i - X_{(i)}\hat{\theta})^2, \quad (1.19)$$

где $X_{(i)}$ — соответствующая строка матрицы X . Желательно, чтобы первая сумма была много больше, чем вторая. Другими словами, чтобы отношение $(\text{tr}D(X\hat{\theta})) / (\text{tr}D(Y))$ было близко к 1.

Замечание 1.6. Отметим, что матрица $X^T X$ может быть вырожденной, поэтому под оценкой $\hat{\theta}$, в данном случае, мы понимаем любое решение нормального уравнения $X^T X \theta = X^T Y$:

$$\hat{\theta} = (X^T X)^- X^T Y.$$

Теперь изучим более подробно компоненты разложения (1.19). Введем в рассмотрение следующую величину:

$$R_0^2 = (Y - X\hat{\theta})^T(Y - X\hat{\theta}).$$

В литературе (см, например, [1], [3]) данную величину обычно называют *остаточной суммой квадратов*, а вектор $(Y - X\hat{\theta})$, соответственно, остаточным вектором. Имеет место следующая лемма.

Лемма 1.8. *Несмещенной оценкой для σ^2 в модели $(Y, X\theta, \sigma^2 I)$ является статистика*

$$s^2 = R_0^2/(N - r),$$

где $r = \text{rank}(X)$.

Доказательство. Заметим, что

$$R_0^2 = (Y - X\hat{\theta})^T(Y - X\hat{\theta}) = Y^T Y - Y^T X(X^T X)^{-1} X^T Y = Y^T A Y,$$

где $A = I_N - X(X^T X)^{-1} X^T$. С другой стороны, имеют место равенства

$$X^T A = (AX)^T = X^T A X = 0.$$

Последнее равенство непосредственно вытекает из леммы 1.6 в силу того, что

$$X^T A X = X^T (I_N - X(X^T X)^{-1} X^T) X = X^T X - X^T X(X^T X)^{-1} X^T X.$$

А то, что матрица AX — нулевая, следует из равенств

$$\begin{aligned} (AX)^T A X &= X^T A A X = \\ &= (X^T - X^T X(X^T X)^{-1} X^T)(X - X(X^T X)^{-1} X^T X) = \\ &= X^T X - X^T X - X^T X + X^T X(X^T X)^{-1} \underbrace{X^T X(X^T X)^{-1} X^T X}_{X^T X} = 0. \end{aligned}$$

Таким образом,

$$Y^T AY = (Y - X\theta)^T A(Y - X\theta) = \epsilon^T A \epsilon.$$

Отсюда получаем

$$\begin{aligned} ER_0^2 &= E(Y^T AY) = E(\epsilon^T A \epsilon) = E(\text{tr} A \epsilon \epsilon^T) = \text{tr} A E(\epsilon \epsilon^T) = \\ &= \sigma^2 \text{tr} A = \sigma^2(N - r). \end{aligned}$$

Здесь мы использовали теорему 1.3 и известную формулу

$$\text{tr} A b b^T = \text{tr} b b^T A = b^T A b,$$

где A и b — любые согласованные матрица и вектор соответственно. Лемма доказана. □

Имеет место следующая теорема.

Теорема 1.7. Пусть вектор Y имеет распределение (1.17). Для того чтобы квадратичная форма $Y^T AY$ имела распределение $\sigma^2 \chi^2$ с числом степеней свободы $\text{tr} A$, необходимо и достаточно, чтобы матрица A была идемпотентной, т.е. $A^2 = A$.

Доказательство этой теоремы основано на использовании теоремы Фишера—Кочнера и имеется в книге [9] стр. 164.

Из этой теоремы непосредственно вытекает следующий результат.

Лемма 1.9. Для схемы регрессии (1.17) величина R_0^2 имеет распределение $\sigma^2 \chi^2$ с числом степеней свободы $N - r$, где $r = \text{rank}(X)$:

$$R_0^2 \sim \sigma^2 \chi^2(N - r).$$

Доказательство. В силу равенства $\text{tr}A = N - r$ и теорем 1.7 и 1.3 достаточно показать, что $A = A^2$:

$$AA = A(I_N - X(X^T X)^{-1}X^T) = A - \underbrace{AX(X^T X)^{-1}X^T}_{=0} = A.$$

Лемма доказана. □

Рассмотрим теперь задачу проверки линейных гипотез о параметрах линейной регрессии. Пусть имеется схема регрессии (1.17), $\text{rank}(X) = r \leq \min\{m, N\}$, и векторная параметрическая функция $\tau = T\theta$ (T — $k \times m$ -матрица ранга k) оценивается. Общую линейную гипотезу о параметрах модели (1.17) можно записать следующим образом:

$$H_0 : T\theta = \tau_0, \quad (1.20)$$

τ_0 — заданный вектор размерности k .

В силу теоремы 1.4 МНК-оценка $\hat{\tau}$ векторной параметрической функции $\tau = T\theta$ имеет распределение

$$\hat{\tau} \sim N(T\theta, \sigma^2 D), \quad D = T(X^T X)^{-1}T^T,$$

при этом $k \times k$ -матрица D — положительно определенная матрица. В силу теоремы о спектральном разложении матрицы (см. доказательство леммы 1.7) матрицу D можно представить в виде произведения двух невырожденных матриц:

$$D = V^T V.$$

Обозначим $v = V^{-1}(\hat{\tau} - \tau)$. Прямое вычисление дает

$$D(v) = E v v^T = V^{-1} \sigma^2 D (V^T)^{-1} = \sigma^2 I.$$

Следовательно, $v \sim N(0, \sigma^2 I)$, из чего получаем, что

$$(\hat{\tau} - \tau)^T D^{-1} (\hat{\tau} - \tau) = v^T v = \sum_{i=1}^k (v_i)^2 \sim \chi^2(k).$$

Имеет место следующая лемма.

Лемма 1.10. *Векторная функция $\hat{\tau} = T\hat{\theta}$ и статистика R_0^2 независимы.*

Доказательство. Как было показано при доказательстве леммы 1.8, $R_0^2 = Y^T A Y$. С другой стороны, $\hat{\tau} = T\hat{\theta} = T(X^T X)^{-1} X^T Y = B Y$ ($B = T(X^T X)^{-1} X^T$). Очевидно, что $BA = 0$ (в силу того, что $X^T A = 0$). По теореме о спектральном разложении имеем

$$A = \sum_{i=1}^N \lambda_i P_i P_i^T,$$

где λ_i — собственные числа, а P_i — ортонормированные собственные вектора матрицы A . Из того, что $BA = 0$, следует, что $B P_i = 0$, а значит, вектор $B Y$ некоррелирован с вектором $P_i Y$, $i = 1, \dots, N$, и следовательно, независим от него в силу нормальности. А это означает, что

$$B Y \text{ и } Y^T A Y = \sum_{i=1}^N \lambda_i (P_i^T Y)^2$$

независимы. Лемма доказана.

□

Теперь используем полученные результаты для проверки гипотез о параметрах модели.

Скалярная параметрическая функция. Сначала рассмотрим случай $\tau = T\theta$ ($T = 1 \times m$ -вектор), $\hat{\tau}$ — МНК-оценка параметра τ . Тогда величины

$$\hat{\tau} \sim N(T\theta\sigma^2D) \text{ и } R_0^2 \sim \sigma^2\chi^2(N-r)$$

в силу предыдущей леммы независимы. Мы хотим проверить гипотезу

$$H_0 : T\theta = \tau_0$$

против альтернативы

$$H_1 : T\theta \neq \tau_0.$$

Если величина σ^2 известна, то в случае, когда гипотеза H_0 верна, статистика

$$u = \frac{\hat{\tau} - \tau_0}{\sigma\sqrt{D}} \sim N(0, 1)$$

может быть выбрана в качестве тестовой статистики для проверки этой гипотезы.

Если величина σ^2 неизвестна, в качестве ее оценки можно использовать статистику $s^2 = R_0^2/(N-r)$. В этом случае, если гипотеза H_0 верна, получаем

$$t = \frac{\hat{\tau} - \tau_0}{\sigma\sqrt{D}} \Big/ \frac{s}{\sigma} = \frac{\hat{\tau} - \tau_0}{s\sqrt{D}} \sim t(N-r),$$

где $t(N-r)$ — стандартное распределение Стьюдента с $(N-r)$ степенями свободы. Если t_α — верхний квантиль порядка α величины $|t|$, т.е. $P(|t| > t_\alpha) = \alpha$, то

$$P(|t| \leq t_\alpha) = P\left(\frac{|\hat{\tau} - \tau_0|}{s\sqrt{D}} \leq t_\alpha\right) = 1 - \alpha,$$

откуда непосредственно получаем

$$P\left(\hat{\tau} - s\sqrt{D}t_\alpha \leq \tau_0 \leq \hat{\tau} + s\sqrt{D}t_\alpha\right) = 1 - \alpha.$$

Если для наблюдаемого значения $\hat{\tau}$

$$\frac{|\hat{\tau} - \tau_0|}{s\sqrt{D}} > t_\alpha,$$

гипотезу H_0 следует отвергнуть. В противном случае нет оснований отвергать гипотезу. Интервал $\left[\hat{\tau} - s\sqrt{D}t_\alpha, \hat{\tau} + s\sqrt{D}t_\alpha\right]$ является $(1 - \alpha)$ -доверительным интервалом для τ_0 (т.е. τ_0 попадает в этот интервал с вероятностью $1 - \alpha$).

Векторная параметрическая функция. Пусть теперь $\tau = T\theta$ ($T - k \times m$ -матрица ранга k), $\hat{\tau}$ — МНК-оценка вектора τ . Мы хотим проверить гипотезу

$$H_0 : T\theta = \tau_0,$$

τ_0 — заданный вектор размерности k , против альтернативы

$$H_1 : T\theta \neq \tau_0.$$

Как было показано ранее, случайные величины

$$(\hat{\tau} - \tau)^T D^{-1}(\hat{\tau} - \tau) \sim \chi^2(k) \text{ и } R_0^2 \sim \sigma^2 \chi^2(N - r)$$

независимы. Отношение этих величин есть статистика

$$F = \frac{(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0)}{k} \bigg/ \frac{R_0^2}{N - r} \sim F(k, N - r), \quad (1.21)$$

где $F(k, n - r)$ — распределение Фишера с k и $N - r$ степенями свободы соответственно. Рассмотрим величину отклонения

$$(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0).$$

Пользуясь теоремой 1.4, получаем

$$\begin{aligned} \frac{E(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0)}{k} &= \frac{1}{k} \text{tr} D^{-1} E(\hat{\tau} - \tau_0)(\hat{\tau} - \tau_0)^T = \\ &= \frac{1}{k} \text{tr} D^{-1} (E(\hat{\tau} - \tau)(\hat{\tau} - \tau)^T + (\tau - \tau_0)(\tau - \tau_0)^T) = \\ &= \frac{1}{k} \text{tr} D^{-1} (\sigma^2 D + (\tau - \tau_0)(\tau - \tau_0)^T) = \sigma^2 + \frac{(\tau - \tau_0)^T D^{-1}(\tau - \tau_0)}{k}. \end{aligned}$$

А это, означает, что

$$\frac{(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0)}{k} \geq \sigma^2.$$

Причем равенство достигается только в том случае, когда верна гипотеза H_0 , т.е. $\tau = \tau_0$.

С другой стороны, как было показано в лемме 1.8, статистика $R_0^2/(N - r)$ является несмещенной оценкой для σ^2 , т.е.

$$\frac{R_0^2}{N - r} = \sigma^2,$$

и следовательно, числитель статистики (1.21) в среднем превосходит знаменатель в случаях, когда гипотеза H_0 неверна. При этом, чем больше различие между векторами τ и τ_0 , тем большее значение принимает статистика F .

Если выбрать такой квантиль F_α , что

$$P(F > F_\alpha) = \alpha, \text{ где } F \sim F(k, N - r),$$

то критерием

$$F = \frac{(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0)}{k} \Bigg/ \frac{R_0^2}{N - r} > F_\alpha$$

можно пользоваться для проверки гипотезы H_0 с уровнем значимости α . Если α мало и данное неравенство не выполняется, то нет оснований отвергать гипотезу. В противном случае гипотезу H_0 следует отвергнуть.

Таблица 1

	Число степеней свободы	Сумма квадратов	Средний квадрат
Полная сумма	$N - r + k$	$R_1^2 = \min_{\theta: T\theta=\tau_0} (Y - X\theta)^T(Y - X\theta)$	—
Остаточная сумма	$N - r$	$R_0^2 = \min_{\theta} (Y - X\theta)^T(Y - X\theta)$	$\frac{R_0^2}{N-r}$
Отклонение от гипотезы	k	$R_1^2 - R_0^2$	$\frac{R_1^2 - R_0^2}{k}$

Замечание 1.7. Для вычисления статистики F можно воспользоваться упрощенной процедурой, которая не предполагает вычисления МНК-оценки вектора $\tau = T\theta$. Вместо этого вычисляется величина

$$R_1^2 = \min_{\theta: T\theta=\tau_0} (Y - X\theta)^T(Y - X\theta).$$

Можно показать, что

$$(\hat{\tau} - \tau_0)^T D^{-1}(\hat{\tau} - \tau_0) = R_1^2 - R_0^2.$$

Доказательство может быть найдено, например, в книге Ермаков С.М., Жиглявский А.А. Математическая теория оптимального эксперимента. М.: Наука, 1987. С учетом этого замечания статистику (1.21) можно переписать, как

$$F = \frac{R_1^2 - R_0^2}{k} \bigg/ \frac{R_0^2}{N - r}.$$

Вычисления, приводящие к статистике F , для удобства, обычно представляют в виде таблицы, которая называется таблицей дисперсионного анализа (см. таблицу 1)

Задачи проверки общих линейных гипотез вида (1.20) о параметрах регрессионной модели называют *задачами дисперсионного анализа*.

1.9. Задачи для самоконтроля

1. Рассмотрим регрессионную модель

$$y_j = \theta^T f(x_j) + \epsilon_j, \quad j = 1, \dots, N, \quad (1.22)$$

где $f(x) = (1, x, \dots, x^m)^T$ — вектор регрессионных функций. Предположим, что ошибки ϵ_j удовлетворяют условиям (1)–(3) т.е. $E\epsilon_i = 0$, $E\epsilon_i^2 = \sigma^2$, $E\epsilon_i\epsilon_j = 0$ ($i \neq j$). Пусть $m = 2$ и результаты измерений задаются таблицей

j	1	2	3
x_j	−1	0.5	1
y_j	0	1.5	2

Найти несмещенную оценку для σ^2 .

2. Рассмотрим регрессионную модель (1.22). Пусть $m = 2$, дисперсионная матрица ошибок имеет вид

$$\Sigma = \begin{pmatrix} 2/3 & 0 & w \\ 0 & 1 & 0 \\ w & 0 & 2/3 \end{pmatrix}$$

и результаты измерений задаются таблицей

j	1	2	3
x_j	−1	0	1
y_j	−1	0.5	3.5

Найти w , для которого оценка

$$\hat{\theta} = \begin{pmatrix} 1 \\ 9/4 \end{pmatrix}$$

является НЛН-оценкой.

3. Рассмотрим регрессионную модель (1.22). Пусть $m = 3$, дисперсионная матрица ошибок имеет вид

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}$$

и результаты измерений задаются таблицей

j	1	2	3	4
x_j	1	2	1	2
y_j	3.5	8	1.5	6

Найти МНК-оценку параметров модели.

4. Рассмотрим регрессионную модель (1.22). Будем считать, что ошибки удовлетворяют условиям (1)–(3). Предположим, что есть следующая априорная информация о параметрах:

$$\theta \in \Omega = \{\theta | \theta^T \theta \leq \sigma^2\}.$$

Пусть $m = 2$, и результаты измерений задаются таблицей

j	1	2	3
x_j	−1	0	1
y_j	−1	1.5	−0.5

Найти гребневую оценку вектора параметров θ .

5. Найти обобщенно-обратную матрицу для матрицы

$$A = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}.$$

6. В 1955 г. было проведено исследование под названием «Экономия стратегических материалов». Различные полиэфирные смолы подвергались испытаниям прочности на разрыв и удлинение с целью выявления условий, при которых металлы можно заменять пластиками и другими синтетическими материалами — полиэтиленом, полиэфиром, стекловолокном и т. п. В табл. 2 представлены данные для 19 образцов полиэфирной смолы толщиной 0,01 дюйма. Они разрывались при разных скоростях ползуна, измеряемых в дюймах в минуту. Предел прочности на разрыв дается в фунтах на квадратный дюйм. Предварительные исследования показали, что предел прочности на разрыв можно считать приблизительно линейной функцией логарифма скорости

$$E(Y|x) = a + bx,$$

где $x = \log$ (скорости), Y — предел прочности на разрыв. Используя эти данные,

(I) найти МНК-оценки параметров a и b ;

(II) построить график наблюдений (в виде точек) и регрессии (линия);

(III) предполагая, что ошибки наблюдений нормально распределены и удовлетворяют условиям (1)—(3), оценить дисперсию ошибки наблюдения.

7. При условии, что выполнены предположения пункта (III) задания 6

(I) найти мат ожидание и дисперсию оценок параметров a и b ;

(II) построить 95%-доверительные интервалы для оценок параметров a и b , предполагая, что ошибки наблюдений нормально распределены и удовлетворяют условиям (1)–(3).

8. Используя данные из задания 6 (табл. 2), проверить гипотезы $H_0 : b = b_0$ и $H_0 : a = a_0$ с уровнем значимости 1% при

(I) $b_0 = 1200$, $a_0 = 6200$;

(II) $b_0 = 1300$, $a_0 = 6250$;

(III) $b_0 = 1400$, $a_0 = 6300$.

Указание. В случае, если дисперсия ошибки наблюдений неизвестна, вместо σ^2 используется ее оценка s^2 , а в качестве тестовых статистик можно использовать

$$T_a = \frac{|a - a_0|}{D\hat{a}}, \quad T_b = \frac{|b - b_0|}{D\hat{b}},$$

имеющие распределение студента с $n - 2$ степенями свободы.

Таблица 2

Скорость ползуна X , дюйм/мин	$x = \log X$	Предел прочности на разрыв Y , фунт/кв. дюйм
0.25	-0.60206	5520
0.25	-0.60206	5390
0.25	-0.60206	5730
0.25	-0.60206	4940
0.25	-0.60206	5810
1	0.0	6840
1	0.0	5720
1	0.0	6120
1	0.0	6400
10	1.0	7100
10	1.0	7150
10	1.0	7260
10	1.0	7650
10	1.0	8210
10	1.0	7960
20	1.30103	7950
20	1.30103	6470
20	1.30103	8720
20	1.30103	8460

9. Выполнить задание 8, считая, что стандартное отклонение ошибок наблюдений известно: $\sigma = 550$. Как изменится в этом случае распределение тестовой статистики?

10. Вязкость в расплавленном состоянии — одна из главных характеристик полиэфиров. Она показывает степень полимеризации. С ее помощью можно установить, как далеко зашла реакция. При химическом исследовании расплавленного полиэфира при 200°F было выделено пять средневесовых степеней полимеризации (X). Вязкость Y измерялась при низком и высоком давлении. Результаты приведе-

ны в табл. 3. Используя данные из этой таблицы

- (I) постройте график зависимости Y от X ;
- (II) найдите методом наименьших квадратов линейную зависимость логарифма вязкости под высоким давлением $\log(Y_1)$ от средневесовой степени полимеризации;
- (III) найдите линейную зависимость $\log(Y_2)$ от X ;
- (IV) определите доверительные интервалы параметров зависимостей (II) и (III);
- (V) проверьте значимо ли отличаются коэффициенты регрессии (II) и (III).

Таблица 3

Средневесовые степени полимеризации (X)	Высокое давление Y_1	Низкое давление Y_2
4.05	28.7	27.1
4.05	29.7	27.4
4.05	31.9	31.8
5.08	60.1	65.3
5.08	62.0	61.1
5.08	63.8	65.8
6.34	123.2	121.6
6.34	120.6	122.4
6.34	116.8	116.9
7.37	182.9	168.5
7.37	179.9	188.0
7.37	178.6	180.8
7.84	228.0	256.5
7.84	232.6	256.8
7.84	230.6	266.5

2. Планирование экспериментов

2.1. Краткий обзор

В течение долгого времени обработка статистических данных проводилась без какого-либо предварительного планирования. Способ проведения эксперимента, т.е. где, когда и как осуществлять измерения, определялся, в основном, интуицией самого экспериментатора. Развитие науки и техники привело к увеличению стоимости экспериментальных исследований и, как следствие, к необходимости создания математического аппарата, позволяющего осуществлять рациональный выбор плана эксперимента. Таким аппаратом стала теория планирования эксперимента.

Основателем теории планирования эксперимента принято считать Р. Фишера, а основополагающим трудом — его книгу (Fisher, 1935). Ряд работ Р. Фишера посвящен разработке методов статистического анализа и планирования эксперимента в задачах агробиологии. Его подход основан на использовании комбинаторных принципов, в частности, латинских квадратов, и развивается по сей день.

В 50-х годах прошлого века американские ученые Дж. Бокс и Дж. Кифер опубликовали ряд статей, оказавших огромное влияние на развитие другого направления — теории оптимального планирования эксперимента. В рамках этой теории исследователю необходимо найти функцию, которая определяет связь между измеряемыми величинами и некоторыми условиями, влияющими на ход исследуемого процесса. Дж. Боксу принадлежат результаты по планированию экстремальных экспериментов и конструированию критериев оптимальности. Дж. Кифер является основателем классической теории планирования регрессионного эксперимента — наиболее развитого в настоящее время раздела теории оптимального планирования эксперимента. Этот раздел весьма привлекателен с вычислительной точки зрения. Теорема эквивалентности (Kiefer, Wolfowitz, 1960) и

ее многочисленные аналоги и обобщения дают необходимые и достаточные условия оптимальности и являются основным инструментом разработки аналитических и численных методов построения оптимальных планов.

Теория оптимального планирования интенсивно развивается и в настоящее время (см. [13]—[28]). Большой вклад в развитие внесли зарубежные и российские ученые Дж. Кифер, Дж. Вольфовиц, Дж. Бокс, В.В. Налимов, В.В. Федоров, Г.К. Круг, Е.В. Маркова, М.Б. Малютов, С.М. Ермаков и другие.

2.2. Основные понятия

В данном разделе рассматриваются некоторые основные понятия и результаты теории оптимального планирования эксперимента.

Рассмотрим набор вещественных величин $y_1, \dots, y_N$, являющихся результатами эксперимента. На практике, довольно часто (см., например, [11], [9], [38]) эти результаты могут быть представлены с помощью уравнения

$$y_j = \eta(t_j, \theta) + \epsilon_j, \quad j = 1, \dots, N, \quad (2.1)$$

где $\eta(t, \theta)$ — вещественная функция, известная с точностью до параметров $\theta = (\theta_0, \dots, \theta_m)^T$, $\epsilon_1, \dots, \epsilon_N$ — случайные величины, характеризующие ошибки наблюдений, $t_1, \dots, t_N$ — условия проведения эксперимента — элементы множества планирования χ .

В дальнейшем мы полагаем, что для уравнения (2.1) выполнены следующие стандартные допущения.

- (а) Несмещенность: $E\epsilon_j = 0$ ($j = 1, \dots, N$). Это означает, что $Ey_j = \eta(t_j, \theta)$, т.е. модель не имеет систематической ошибки.
- (б) Некоррелированность: $E\epsilon_i\epsilon_j = 0$ ($i \neq j$).
- (в) Равноточность: $E\epsilon_j^2 \equiv \sigma^2 > 0$ ($j = 1, \dots, N$).

- (г) Линейность параметризации: $\eta(t, \theta) = \theta^T f(t)$; $f(t) = (f_0(t), \dots, f_m(t))^T$, где $f_i(t)$, $i = 0, \dots, m$ — заданные базисные функции.
- (д) Функции $f_i(t)_{i=0}^m$ непрерывны и линейно независимы на χ .
- (е) χ — фиксированное множество, наделенное структурой компактного топологического пространства.

Эти допущения отражают особенности реальных экспериментов. В принципе, они могут быть ослаблены (см. [2]). Однако в данной работе мы эти случаи рассматривать не будем.

Целью эксперимента, как правило, является либо нахождение оценок параметров $\theta_0, \dots, \theta_m$, либо нахождение оценок некоторых функций от этих параметров, либо проверка гипотез относительно значений этих параметров. На точность полученных оценок влияет как выбор условий проведения экспериментов $(t_1, \dots, t_N)$, так и метод оценивания.

Оценка $\hat{\theta}$ вектора параметров θ является некоторой статистикой

$$\hat{\theta} = \hat{\theta}(t_1, \dots, t_N, y_1, \dots, y_N).$$

Выбор статистики $\hat{\theta}$ при фиксированных условиях проведения эксперимента составляет предмет отдельной теории. В предположениях (а)—(е) существует хорошо зарекомендовавший себя метод оценивания (метод наименьших квадратов), согласно которому в качестве оценки $\hat{\theta}$ выбирается решение экстремальной задачи

$$\sum_{j=1}^N (y_j - \eta(t_j, \theta))^2 \rightarrow \min_{\theta}.$$

Данная задача рассматривалась нами в первой части.

Как уже говорилось ранее, точность получаемых оценок можно улучшить, выбрав подходящие условия проведения экспериментов. Эти условия записываются в виде таблицы

$$\xi_N = \begin{pmatrix} t_1 & \cdots & t_N \\ 1/N & \cdots & 1/N \end{pmatrix}, \quad t_i \in \chi, \quad i = 1, 2, \dots, N, \quad (2.2)$$

где t_i , вообще говоря, могут равняться t_j при $i \neq j$, N — общее число измерений. Данную таблицу принято называть дискретным (нормированным) планом эксперимента. Дж. Кифер обобщил это понятие, что позволило существенно расширить класс исследуемых задач. Согласно его идеи под непрерывным (приближенным) планом эксперимента понимается вероятностная мера

$$\xi = \begin{pmatrix} t_1 & \cdots & t_n \\ \omega_1 & \cdots & \omega_n \end{pmatrix}, \quad t_i \in \chi, \quad i = 1, 2, \dots, n,$$

причем $t_i \neq t_j$ при $i \neq j$, $\omega_i \geq 0$, $\sum_{i=1}^n \omega_i = 1$, ω_i — весовые коэффициенты, а n , в данном случае, — число различных точек в плане. На практике достаточно часто могут быть реализованы только дискретные планы, и непрерывный план приходится заменять на близкий ему дискретный. В этом случае в точке t_i проводится, приблизительно, $N\omega_i$ измерений, $i = 1, 2, \dots, n$. Методы округления чисел $N\omega_i$ до целых можно найти в работах [11], [2].

Обозначим за Ξ_n множество непрерывных планов, сосредоточенных в n точках с ненулевыми коэффициентами.

Замечание 2.1. Множество Ξ_n является выпуклым, т.е. для $\forall \xi_1, \xi_2 \in \Xi_n$ верно, что $\xi = \alpha \xi_1 + (1 - \alpha) \xi_2 \in \Xi_n$, $\alpha \in [0, 1]$.

Поясним, что в данном случае понимается под планом ξ . Пусть

$$\xi_1 = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ \bar{\omega}_1 & \bar{\omega}_2 & \cdots & \bar{\omega}_n \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ \tilde{\omega}_1 & \tilde{\omega}_2 & \cdots & \tilde{\omega}_n \end{pmatrix}$$

(если точки носителей планов ξ_1 и ξ_2 не совпадают, дополняем их соответствующими точками с нулевыми весами). Тогда

$$\xi = \alpha\xi_1 + (1 - \alpha)\xi_2 = \begin{pmatrix} t_1 & t_2 & \dots & t_n \\ \omega_1 & \omega_2 & \dots & \omega_n \end{pmatrix},$$

где $\omega_i = \alpha\bar{\omega}_i + (1 - \alpha)\tilde{\omega}_i$, $i = 1, \dots, n$

Под информационной матрицей приближенного (непрерывного) плана мы будем понимать матрицу

$$M(\xi) = \int_{\chi} f(t)f^T(t)\xi(dt).$$

Пусть $\Xi = \bigcup_{n=1}^{\infty} \Xi_n$ — множество всех непрерывных планов, а

$$\mathcal{M} = \{M : M = M(\xi), \xi \in \Xi\}$$

есть множество всех информационных матриц. Основные свойства информационных матриц даны в следующей теореме.

Теорема 2.1. *Следующие утверждения верны:*

- (1) *любая информационная матрица является неотрицательно определенной;*
- (2) *если $n < m + 1$, то $\det M(\xi) = 0$ при $\xi \in \Xi_n$;*
- (3) *множество $\mathcal{M}$ является выпуклым;*
- (4) *при выполнении условий (a)–(e) множество $\mathcal{M}$, рассматриваемое как множество векторов, состоящих из диагональных и наддиагональных элементов матриц, является компактным подмножеством $R^{(m+2)(m+1)/2}$;*

(5) при выполнении условий (a)–(e) для любого плана $\xi \in \Xi$ найдется план $\tilde{\xi} \in \Xi_n$, где $n \leq (m+2)(m+1)/2 + 1$, такой, что

$$M(\tilde{\xi}) = M(\xi).$$

Доказательство. Докажем последовательно все пункты теоремы.

1. Пункт (1) непосредственно следует из определения неотрицательно определенной матрицы. В самом деле, для любого вектора $l \in R^m$ имеют место соотношения

$$l^T M(\xi) l = \int_{\chi} l^T f(t) f^T(t) l \xi(dt) = \int_{\chi} (l^T f(t))^2 \xi(dt) \geq 0.$$

2. Для доказательства пункта (2) воспользуемся известным неравенством $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$, в силу которого

$$\text{rank}(M(\xi)) \leq \sum_{i=1}^n \text{rank}(f(x_i) f^T(x_i)),$$

а матрица $f(t) f^T(t)$ имеет единичный ранг, т.к.

$$f(t) f^T(t) = \begin{pmatrix} f_0(t) f(t)^T \\ f_1(t) f(t)^T \\ \vdots \\ f_m(t) f(t)^T \end{pmatrix},$$

т.е., все строки матрицы $f(t) f^T(t)$ линейно зависимы от вектора $f^T(t)$. Таким образом, $\text{rank}(M(\xi)) \leq n$ и пункт (2) доказан.

3. Нам требуется показать, что для любых согласованных матриц $M_1, M_2 \in \mathcal{M}$ матрица $M = \alpha M_1 + (1 - \alpha) M_2$ также принадлежит

$\mathcal{M}$. По определению

$$\begin{aligned}\alpha M_1 + (1 - \alpha)M_2 &= \alpha \int f(t)f^T(t)\xi_1(dt) + (1 - \alpha) \int f(t)f^T(t)\xi_2(dt) = \\ &= \int f(t)f^T(t)\alpha\xi_1(dt) + (1 - \alpha)\xi_2(dt) = \int f(t)f^T(t)\xi(dt)\end{aligned}$$

и, согласно замечанию 2.1, $\xi \in \Xi_n$. Из этого следует, что $M = M(\xi) \in \mathcal{M}$. Для доказательства пунктов (4) и (5) нам потребуются несколько вспомогательных утверждений и понятий. В связи с этим, доказательство приводится в отдельном разделе (см. лемму 2.3 и следствие 2.1).

Теорема доказана.

□

2.3. Теорема Каратеодори

В данном разделе мы будем рассматривать множество векторов в конечномерном евклидовом пространстве R^k . Нам понадобятся следующие вспомогательные понятия.

Определение 2.1. *Выпуклой комбинацией* векторов v_1 и v_2 называется вектор $\alpha v_1 + (1 - \alpha)v_2$, где $0 < \alpha < 1$.

Определение 2.2. Множество называется *выпуклым*, если оно содержит выпуклую комбинацию любых двух своих векторов:

$$v_1, v_2 \in V \Rightarrow \alpha v_1 + (1 - \alpha)v_2 \in V.$$

Определение 2.3. *Выпуклой оболочкой* множества V называется пересечение всех выпуклых множеств, содержащих это множество. Обозначается $\text{conv}V$.

Рассмотрим следующие множества:

$$\widehat{V}_n = \left\{ v \in R^k; v = \sum_{i=1}^n \alpha_i v_i, \alpha_i > 0, \sum \alpha_i = 1, v_i \in V \right\}, \quad \widehat{V} = \bigcup_{n=1}^{\infty} \widehat{V}_n$$

Имеет место следующая лемма.

Лемма 2.1. *Множество $\widehat{V}$ есть выпуклая оболочка множества V , то есть $\widehat{V} = \text{conv}V$.*

Доказательство. Разобьем доказательство на две части. Вначале докажем, что $\widehat{V} \subset \text{conv}V$, а потом докажем, что $\text{conv}V \subset \widehat{V}$.

1. $\widehat{V} \subset \text{conv}V$. Доказательство будем проводить по методу математической индукции. Очевидно, что $\widehat{V}_1 \subset \text{conv}V$. Пусть верно, что $\widehat{V}_n \subset \text{conv}V$, докажем, что $\widehat{V}_{n+1} \subset \text{conv}V$. Любой элемент $v \in \widehat{V}_{(n+1)}$ имеет вид

$$v = \sum_{i=1}^{n+1} \alpha_i v_i = \sum_{i=1}^n \alpha_i v_i + \alpha_{n+1} v_{n+1}.$$

Без потери общности можем считать, что $\alpha_{n+1} > 0$. Пусть $\alpha = 1 - \alpha_{n+1} = \sum_{i=1}^n \alpha_i$. Рассмотрим вектора

$$\bar{v}_1 = v_{n+1}, \quad \bar{v}_2 = \sum_{i=1}^n \frac{\alpha_i}{\sum_{i=1}^n \alpha_i} v_i = \sum_{i=1}^n \alpha'_i v_i, \quad \sum_{i=1}^n \alpha'_i = 1.$$

Тогда $\bar{v}_1 \in V \subset \text{conv}V$ и $\bar{v}_2 \in \widehat{V}_n \subset \text{conv}V$ (по индукционному предположению). И, так как множество $\text{conv}V$ выпукло, $(1 - \alpha)\bar{v}_1 + \alpha\bar{v}_2 \in \text{conv}V$. Таким образом, по методу математической индукции мы доказали, что $\widehat{V} \subset \text{conv}V$. Докажем теперь, что $\text{conv}V \subset \widehat{V}$.

2. $\text{conv}V \subset \widehat{V}$. Очевидно, что $V \subset \widehat{V}$ (т.к. $\widehat{V}_1 = V$). Кроме того, несложно заметить, что множество $\widehat{V}$ — выпуклое. Действительно,

пусть $\widehat{v}_1, \widehat{v}_2 \in \widehat{V}$, для определенности $\widehat{v}_1 = \sum_{i=1}^{n_1} \alpha_i v_i$, $\widehat{v}_2 = \sum_{i=1}^{n_2} \alpha'_i v'_i$. Тогда $v = \alpha \widehat{v}_1 + (1 - \alpha) \widehat{v}_2 \in \widehat{V}$. В самом деле,

$$v = \alpha \sum_{i=1}^{n_1} \alpha_i v_i + (1 - \alpha) \sum_{i=1}^{n_2} \alpha'_i v'_i = \sum_{i=1}^{n_1+n_2} \tilde{\alpha}_i \tilde{v}_i,$$

где $\tilde{v}_i = v_i$, $\tilde{\alpha}_i = \alpha \alpha_i$, $i \leq n_1$ и $\tilde{v}_i = v'_{i-n_1}$, $\tilde{\alpha}_i = (1 - \alpha) \alpha'_{i-n_1}$, $i > n_1$. Кроме того, очевидно, что $\sum_{i=1}^{n_1+n_2} \tilde{\alpha}_i = 1$ и $\tilde{\alpha}_i \geq 0$. В силу того, что по определению $\text{conv} V$ содержится в любом выпуклом множестве, содержащем V , имеем $\text{conv} V \subset \widehat{V}$.

Лемма доказана. □

Лемма 2.2. Пусть имеется набор векторов $v_1, \dots, v_{k+2} \in R^k$. Тогда существуют такие числа $\beta_1, \dots, \beta_{k+2}$, что

$$\sum_{i=1}^{k+2} \beta_i v_i = 0, \quad \sum_{i=1}^{k+2} \beta_i = 0,$$

причем не все β_i равны 0.

Доказательство. Рассмотрим векторы $\bar{v}_i = (v_i^T, 1)^T$. Это есть векторы в $k+1$ -мерном пространстве, и так как в $k+1$ -мерном пространстве существует не более, чем $k+1$ линейнонезависимых векторов, существуют не равные одновременно нулю числа $\beta_1, \dots, \beta_{k+2}$ такие, что

$$\sum_{i=1}^{k+2} \beta_i \bar{v}_i = 0,$$

из чего следует утверждение леммы.

Лемма доказана. □

Теперь сформулируем и докажем теорему Каратеодори.

Теорема 2.2 (Каратеодори). *Любой элемент выпуклой оболочки $\text{conv} V$ произвольного компактного множества V в k -мерном евклидовом пространстве представим в виде выпуклой комбинации с положительными коэффициентами не более, чем $k+1$ элементов множества V .*

Доказательство. Пусть $v \in \text{conv} V$ — произвольный элемент. В силу леммы 2.1 всякий элемент множества $\text{conv} V$ представим в виде выпуклой комбинации точек множества V . Если v представима в виде $v = \sum_{i=1}^n \alpha_i v_i$, $v_i \in V$, $\alpha_i > 0$, $\sum_i \alpha_i = 1$, $n < k+2$, то доказывать нечего. Предположим теперь, что $n \geq k+2$. В силу леммы 2.2 существуют такие не равные одновременно нулю числа β_i , $\sum \beta_i = 0$, что

$$\sum_{i=1}^{k+2} \beta_i v_i = 0.$$

Обозначим за $i_0 = \arg(\min_i \{\alpha_i / \beta_i : \beta_i > 0\})$ и $\bar{\alpha}_i = \alpha_i - \alpha_{i_0} \beta_i / \beta_{i_0}$, $i \leq k+2$, $\bar{\alpha}_i = \alpha_i$, $i > k+2$. Тогда, очевидно, что v представима в виде

$$v = \sum_{i=1}^n \bar{\alpha}_i v_i, \quad \sum_i \bar{\alpha}_i = 1, \quad \bar{\alpha}_i \geq 0,$$

причем по крайней мере одно из $\bar{\alpha}_i$ обращается в 0 (при $i = i_0$). Таким образом, нам удалось уменьшить число ненулевых компонент в выпуклой комбинации как минимум на 1. Продолжая данную процедуру, получаем

$$v = \sum_{i=1}^{k+1} \tilde{\alpha}_i v_i, \quad \sum_i \bar{\alpha}_i = 1, \quad \bar{\alpha}_i \geq 0,$$

Теорема доказана. □

Чтобы мы могли применить эту теорему к множеству информационных матриц, нам потребуется еще один вспомогательный результат. Пусть

$$\tilde{V} = \left\{ \tilde{v} : \tilde{v} = \int v(x) \xi(dx), v \in V \right\},$$

где ξ — произвольная вероятностная мера на $(X, \mathcal{B})$, $\mathcal{B}$ — σ -алгебра подмножеств множества X . Множество $V = V(x)$ есть множество непрерывных вектор-функций на X .

Лемма 2.3. *Если множество X компактно, то $\tilde{V}$ совпадает с выпуклой оболочкой множества V . То есть $\tilde{V} = \text{conv} V$.*

Доказательство. Отметим, что в силу непрерывности функций $v(x) \in V$ и компактности X множество $\text{conv} V$ является компактным. Как и в лемме 2.1, легко убедиться, что $\text{conv} V \subset \tilde{V}$ (т.к. $\tilde{V}$ выпуклое и содержит V). Предположим, что $\tilde{V} \neq \text{conv} V$, тогда существует план ξ^* :

$$v^* = \int v \xi^*(dv), \quad v^* \notin \text{conv} V.$$

Воспользуемся теоремой отделимости, согласно которой между точкой v^* и компактным множеством $\text{conv} V$ можно провести плоскость так, чтобы множество $\text{conv} V$ и точка v^* принадлежали разным полупространствам, порождаемым этой плоскостью. Другими словами, это означает, что для любой точки множества $\text{conv} V$ существует вектор α такой, что

$$\alpha^T v = \sum \alpha_i v_i \leq C < \alpha^T v^* = \sum \alpha_i v_i^*.$$

С другой стороны, проинтегрировав левую часть неравенства, имеем

$$\alpha^T v^* = \int \alpha^T v \xi^*(dv) \leq \int C \xi^*(dv) = C.$$

Полученное противоречие завершает доказательство леммы.
Лемма доказана. □

Из данной леммы и теоремы 2.2 непосредственно вытекает следующий результат (пункт (5) теоремы 2.1).

Следствие 2.1. *Если выполнены условия (a)–(e), то информационная матрица любого плана ξ может быть представлена в виде выпуклой комбинации не более чем $(m+1)(m+2)/2+1$ точек $f(x_i)f^T(x_i)$. То есть,*

$$M(\xi) = \sum_{i=1}^n \omega_i f(x_i) f^T(x_i), \quad \omega_i \geq 0, \quad \sum \omega_i = 1,$$

где $n \leq (m+1)(m+2)/2+1$.

Замечание 2.2. В силу того, что информационная матрица симметрична, она определяется своими $(m+1)(m+2)/2$ элементами. Другими словами, ей можно сопоставить вектор размерности $(m+1)(m+2)/2$.

Замечание 2.3. Мы получили следующий важный результат. Непрерывные планы с требуемыми свойствами оптимальности можно строить на множестве планов вида

$$\xi = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ \omega_1 & \omega_2 & \cdots & \omega_n \end{pmatrix},$$

где $n \leq (m+1)(m+2)/2+1$.

2.4. Критерии оптимальности

Планы, информационная матрица которых невырожденная, называются невырожденными планами.

Дисперсионной матрицей невырожденного плана мы будем называть матрицу

$$D(\xi) = M(\xi)^{-1}.$$

Как правило, не существует такого невырожденного плана ξ^* , что

$$D(\xi^*) \leq D(\xi)$$

для любого невырожденного плана $\xi \in \Xi$. Поэтому в качестве критериев оптимальности используются некоторые вогнутые функции, заданные на множестве информационных матриц, или выпуклые функции, заданные на множестве дисперсионных матриц, имеющие строгий статистический смысл.

Рассмотрим ряд наиболее распространенных критериев.

D-критерий. Название этого критерия использует первую букву слова determinant. Критерий D -оптимальности имеет вид

$$\log \det M(\xi) \rightarrow \sup_{\xi \in \Xi}$$

или

$$\log \det D(\xi) \rightarrow \inf_{\xi \in \Xi}.$$

Этот критерий соответствует требованию минимизации объема доверительного эллипсоида, который имеет вид

$$\{\tilde{\theta} : (\tilde{\theta} - \hat{\theta})^T D \hat{\theta} (\tilde{\theta} - \hat{\theta}) \leq \alpha\},$$

где α — некоторая постоянная, зависящая только от доверительного уровня.

L-критерий. Критерий L -оптимальности имеет вид

$$\text{tr} LD(\xi) \rightarrow \inf_{\xi \in \Xi_H},$$

где $\Xi_H = \{\xi \in \Xi : \det M(\xi) \neq 0\}$, а $L \in R^{(d+1) \times (d+1)}$ — фиксированная неотрицательно определенная матрица.

Статистический смысл L -оптимального плана заключается в том, что для него обобщенные квадратичные потери

$$E(\hat{\theta} - \theta)^T L(\hat{\theta} - \theta)$$

минимальны.

e_k -критерий. Критерий e_k -оптимальности имеет вид

$$e_k^T M^-(\xi) e_k \rightarrow \inf_{\xi \in \Xi_{e_k}},$$

где $e_k = (0, \dots, 0, 1, 0, \dots, 0)^T$ — $k + 1$ -й орт евклидова пространства R^{d+1} , а $\Xi_{e_k} = \{\xi \in \Xi : e_k^T M^-(\xi) M(\xi) = e_k^T\}$.

Статистический смысл этого критерия заключается в минимизации дисперсии $(M^-(\xi))_{kk}$ оценки параметра θ_k .

E -критерий. Критерий E -оптимальности имеет вид

$$\lambda_{\min}(M(\xi)) \rightarrow \sup_{\xi \in \Xi_H}$$

или

$$\lambda_{\max}(D(\xi)) \rightarrow \inf_{\xi \in \Xi_H}.$$

Этот критерий означает минимизацию максимальной длины оси доверительного эллипсоида МНК-оценок параметров.

G -критерий. Критерий G -оптимальности имеет вид

$$\max_{t \in \chi} d(t, \xi) \rightarrow \inf_{\xi \in \Xi_H},$$

где $d(t, \xi) = f^T(t)D(\xi)f(t)$.

Этот критерий означает минимизацию максимальной по t дисперсии поверхности отклика (функции регрессии).

Отметим, что все перечисленные критерии могут быть записаны в виде

$$\Psi(M^-(\xi)) \rightarrow \inf_{\xi}$$

или

$$\Phi(M(\xi)) \rightarrow \sup_{\xi}.$$

Причем функции Ψ и Φ удовлетворяют следующим свойствам:

- (a) монотонность: $[\Phi(M(\xi_1)) \leq \Phi(M(\xi_2))]$, если $M(\xi_1) \leq M(\xi_2)$;
- (b) однородность: $[\Phi(\theta M(\xi)) \leq \gamma(\theta)\Phi(M(\xi))]$, где $\gamma(\theta)$ — возрастающая функция;
- (c) вогнутость: (для Ψ выпуклость)

$$\Phi(M((1 - \alpha)\xi_1 + \alpha\xi_2)) \geq (1 - \alpha)\Phi(M(\xi_1)) + \alpha\Phi(M(\xi_2)).$$

Более полный список критериев можно найти в [2], [3], [38].

2.5. Теоремы эквивалентности

Важным инструментом для построения D -оптимальных планов является следующая теорема.

Теорема 2.3 (Кифер—Вольфовиц). *Если множество информационных матриц компактно, то следующие условия эквивалентны:*

- (a) план ξ^* — D -оптимальный;

(b) план ξ^* — G -оптимальный;

(c) $\max_{x \in \chi} d(x, \xi^*) = m$.

Причем, если план ξ^* сосредоточен в конечном числе точек, то последнее равенство достигается в точках x_i^* оптимального плана ξ^* . Кроме того, информационные матрицы всех оптимальных планов совпадают.

Напомним, что $d(x, \xi) = f^T(x)D(\xi)f(x)$.

Для доказательства теоремы Кифера—Вольфовица, которая также известна как теорема эквивалентности, нам потребуется несколько вспомогательных утверждений.

Лемма 2.4. Пусть A — произвольная положительно определенная матрица размерности $m \times m$. Справедливо следующее интегральное представление для ее определителя:

$$(\det A)^{-1/2} = \frac{1}{\pi^{m/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-x^T A x} dx_1 \dots dx_m$$

Доказательство. Будем использовать известное равенство, которое доказывается в стандартном курсе математического анализа:

$$\int_{-\infty}^{\infty} e^{-z^2} dz = \sqrt{\pi}.$$

Пусть $A = \Lambda$, Λ — диагональная матрица (на главной диагонали ненулевые элементы λ_i , остальные элементы равны нулю); тогда

$$\begin{aligned} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-x^T A x} dx_1 \dots dx_m &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\sum_{i=1}^m \lambda_i x_i^2} dx_1 \dots dx_m = \\ &= \prod_{i=1}^m \int_{-\infty}^{\infty} e^{-\lambda_i x^2} dx = \prod_{i=1}^m \sqrt{\frac{\pi}{\lambda_i}} = \pi^{m/2} (\det A)^{-1/2}. \end{aligned}$$

Если A — произвольная положительно определенная матрица, то, воспользовавшись теоремой о спектральном разложении, получаем

$$A = P^T \Lambda P, \quad P^T P = I, \quad \Lambda — \text{диагональная}.$$

Введем замену переменных $x = Py$. Получаем

$$x^T A x = (Py)^T A Py = y^T \Lambda y = \sum_{i=1}^m \lambda_i y_i^2.$$

В силу того, что столбцы матрицы P — ортонормированные векторы, якобиан замены равен 1. И следовательно,

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-x^T A x} dx_1 \dots dx_m = \pi^{m/2} (\det A)^{-1/2}.$$

Лемма доказана. □

Лемма 2.5. Пусть A и B — произвольные положительно определенные матрицы размерности $m \times m$. Справедливо следующее неравенство:

$$\det(\alpha A + (1 - \alpha)B) \geq (\det A)^\alpha (\det B)^{1-\alpha}, \quad (2.3)$$

причем равенство достигается, если и только если $A = B$.

Доказательство. Доказательство неравенства (2.3) основано на использовании неравенства Гельдера. Пусть $p > 1$, $q = p/(p - 1)$, $f \in L^p$, $g \in L^q$ — функции n переменных, $x \in X \subset R^n$; тогда

$$\int_X f(x)g(x)dx \leq \left(\int_X f(x)^p \right)^{1/p} \left(\int_X g(x)^q \right)^{1/q},$$

причем равенство достигается, если и только если

$$\frac{f(x)^p}{\int_X f(x)^p dx} = \frac{g(x)^q}{\int_X g(x)^q dx}.$$

Положим $p = 1/\alpha$, $q = 1/(1 - \alpha)$, $f(x) = \exp(-\alpha x^T A x)$, $g(x) = \exp(-(1 - \alpha)x^T B x)$. Получаем

$$\begin{aligned} \frac{\pi^{n/2}}{\sqrt{\det(\alpha A + (1 - \alpha)B)}} &= \int \dots \int e^{-z^T(\alpha A + (1 - \alpha)B)z} dz = \\ &= \int f(z)g(z)dz \leq \left(\int f^{1/\alpha}(z)dz \right)^\alpha \left(\int g^{1/(1-\alpha)}(z)dz \right)^{1-\alpha} = \\ &= \left(\frac{\pi^{n/2}}{\sqrt{\det(A)}} \right)^\alpha \left(\frac{\pi^{n/2}}{\sqrt{\det(B)}} \right)^{1-\alpha}. \end{aligned}$$

Отсюда следует, с учетом интегрального представления из леммы 2.4, что

$$\det(\alpha A + (1 - \alpha)B) \geq (\det A)^\alpha (\det B)^{1-\alpha},$$

причем равенство достигается, если и только если $A = B$. Лемма доказана. □

Лемма 2.6. *Функция $f(A) = \ln \det A$ является строго вогнутой на классе положительно определенных матриц.*

Доказательство. По определению достаточно доказать, что для любого $0 < \alpha < 1$ при $A \neq B$ должно быть выполнено неравенство

$$f(\alpha A + (1 - \alpha)B) > \alpha f(A) + (1 - \alpha)f(B).$$

Прологарифмируем неравенство (2.3) из предыдущей леммы и получим то, что требуется доказать. Лемма доказана. □

Лемма 2.7. Для произвольной матрицы $A > 0$ имеет место следующая формула:

$$\frac{\partial}{\partial \alpha} \ln \det A(\alpha) = \operatorname{tr} A^{-1}(\alpha) \frac{\partial}{\partial \alpha} A(\alpha).$$

Доказательство. Применяя формулу для вычисления производной логарифма функции, получаем

$$\frac{\partial}{\partial \alpha} \ln \det A(\alpha) = \frac{1}{\det A(\alpha)} \frac{\partial}{\partial \alpha} \det A(\alpha).$$

Заметим, что

$$\det A = \sum_{j=1}^m (-1)^{i+j} a_{ij} A_{ij},$$

где A_{ij} — алгебраические дополнения, $i = 1, \dots, m$. В силу того, что A_{ij} не зависят от элементов a_{ij} , имеет место равенство

$$\frac{\partial \det A}{\partial a_{ij}} = (-1)^{i+j} A_{ij}.$$

Обозначим за a^{ij} элементы матрицы A^{-1} . Воспользовавшись известной формулой $a^{ij} = (-1)^{i+j} A_{ji} / \det A$, получаем

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln \det A(\alpha) &= \frac{1}{\det A(\alpha)} \frac{\partial}{\partial \alpha} \det A(\alpha) = \frac{1}{\det A} \sum_{i,j=1}^m \frac{\partial \det A}{\partial a_{ij}} \frac{\partial a_{ij}(\alpha)}{\partial \alpha} = \\ &= \sum_{i,j=1}^m \underbrace{\frac{(-1)^{i+j} A_{ij}}{\det A}}_{a^{ji}} \frac{\partial a_{ij}(\alpha)}{\partial \alpha} = \operatorname{tr} A^{-1} \frac{\partial A}{\partial \alpha}. \end{aligned}$$

Лемма доказана. □

Лемма 2.8. Для любого невырожденного плана ξ имеет место неравенство $\sup_x d(x, \xi) \geq m$.

Доказательство. Прямое вычисление дает

$$\begin{aligned} \sup_x d(x, \xi) &\geq \int d(x, \xi) \xi(dx) = \int f^T(x) M^{-1}(\xi) f(x) \xi(dx) = \\ &= \text{tr} M^{-1}(\xi) \int f(x) f^T(x) \xi(dx) = \text{tr} M^{-1}(\xi) M(\xi) = \text{tr} I_m = m. \end{aligned}$$

Лемма доказана. □

Теперь докажем теорему 2.3.

Доказательство теоремы 2.3. Схема доказательства будет следующей: $(a) \Rightarrow (c) \Rightarrow (b) \Rightarrow (a)$.

1. (a) $\Rightarrow$ (c). Пусть ξ_D — D -оптимальный план. Рассмотрим план $\xi_\alpha = (1 - \alpha)\xi_D + \alpha\xi_x$, где $\xi_x = \{x, 1\}$ — план, сосредоточенный в точке x , $0 < \alpha < 1$. В силу D -оптимальности плана ξ_D имеет место следующее неравенство:

$$\ln \det M(\xi_\alpha) \leq \ln \det M(\xi_D) \Leftrightarrow \frac{\ln \det M(\xi_\alpha) - \ln \det M(\xi_D)}{\alpha} \leq 0.$$

Переходя к пределу при $\alpha \rightarrow 0+$, получаем

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln \det M(\xi_\alpha) \big|_{\alpha=0+} &= \text{tr} M^{-1}(\xi_\alpha) \frac{\partial M(\xi_\alpha)}{\partial \alpha} \big|_{\alpha=0+} = \\ &= \text{tr} M^{-1}(\xi_\alpha) \frac{\partial(1 - \alpha)M(\xi_D) + \alpha M(\xi_x)}{\partial \alpha} \big|_{\alpha=0+} = \\ &= \text{tr} M^{-1}(\xi_D) (f(x) f^T(x) - M(\xi_D)) = d(x, \xi_D) - m \leq 0, \end{aligned}$$

т.е. мы получили, что $d(x, \xi_D) \leq m$ для $\forall x$, и следовательно, $\sup_x d(x, \xi_D) = m$.

2. (c) $\Rightarrow$ (b). Пусть для плана ξ^* выполнено $\sup_x d(x, \xi^*) = m$, тогда

по лемме 2.8 план ξ^* — G -оптимальный.

3. (b) $\Rightarrow$ (a). Доказательство будем проводить от противного. Предположим, что существует G -оптимальный план, который не является D -оптимальным. Обозначим этот план через ξ^* . Пусть план ξ_D — некоторый D -оптимальный план. Для плана $\xi_\alpha = (1 - \alpha)\xi^* + \alpha\xi_D$ в силу леммы 2.6 имеем

$$\ln \det M(\xi_\alpha) \geq (1 - \alpha) \ln \det M(\xi^*) + \alpha \ln \det M(\xi_D).$$

Вычитая из обеих частей $\ln \det M(\xi^*)$, получаем

$$\ln \det M(\xi_\alpha) - \ln \det M(\xi^*) \geq \alpha (\ln \det M(\xi_D) - \ln \det M(\xi^*)) > 0,$$

так как план ξ^* не является D -оптимальным. Из полученного неравенства следует, что

$$\frac{\ln \det M(\xi_\alpha) - \ln \det M(\xi^*)}{\alpha} > 0,$$

и значит,

$$\frac{\partial}{\partial \alpha} \ln \det M(\xi_\alpha) \big|_{\alpha=0+} = \operatorname{tr} M^{-1}(\xi^*) M(\xi_D) - m > 0,$$

т.е. мы получили, что $\operatorname{tr} M^{-1}(\xi^*) M(\xi_D) > m$. С другой стороны, так как план ξ^* — G -оптимальный, имеет место неравенство

$$f^T(x) M^{-1}(\xi^*) f(x) \leq m \quad \forall x,$$

но это, в свою очередь, означает, что

$$\operatorname{tr} M^{-1}(\xi^*) M(\xi_D) = \int f^T(x) M^{-1}(\xi^*) f(x) \xi_D(dx) \leq m.$$

Получили противоречие. Следовательно, G -оптимальный план ξ^* также является D -оптимальным. Таким образом, мы доказали эквивалентность трех условий (a), (b), (c). То, что максимум функции

$d(x, \xi)$ достигается в точках оптимального плана (если план сосредоточен в конечном числе точек), следует из леммы 2.8 (для плана $\xi = \xi^*$ неравенство в лемме превращается в равенство). То, что информационные матрицы всех оптимальных планов совпадают, следует из леммы 2.6 и замечания 2.1 для непрерывных планов.

Теорема доказана. □

Рассмотрим пример, иллюстрирующий, как теорему 2.3 можно использовать для проверки D -оптимальности плана.

Пример 2.1. Пусть вектор регрессионных функций имеет вид $f(x) = (1, x, x^2)^T$. В качестве множества планирования рассмотрим интервал $\chi = [-1, 1]$. Предположим, мы хотим проверить на D -оптимальность следующий план:

$$\xi^* = \begin{pmatrix} x_1 & x_2 & x_3 \\ \omega_1 & \omega_2 & \omega_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Найдем информационную матрицу этого плана и обратную к ней:

$$M(\xi^*) = \sum_{i=1}^3 f(x_i) f^T(x_i) \omega_i = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{pmatrix},$$

$$M^{-1}(\xi^*) = D(\xi^*) = \begin{pmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{pmatrix}.$$

Для проверки D -оптимальности плана ξ^* согласно теореме 2.3 достаточно убедиться, что функция $d(x, \xi^*) = f(x) D(\xi^*) f^T(x)$ не превосходит значение $m = 3$ для любых $x \in \chi = [-1, 1]$:

$$d(x, \xi^*) = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} 3 & 0 & -3 \\ 0 & 1.5 & 0 \\ -3 & 0 & 4.5 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} = 3 + \frac{9}{2} x^2 (x^2 - 1).$$

Заметим, что второе слагаемое неположительно. Поэтому максимум нашей функции на интервале $x \in [-1, 1]$ равен 3 и достигается в точках $\{-1\}, \{0\}, \{1\}$. Согласно теореме эквивалентности это означает, что план ξ^* — D -оптимальный.

Замечание 2.4. Отметим, что для дискретных планов (2.2) утверждение теоремы, вообще говоря, не верно. То есть дискретные D -оптимальные планы могут отличаться от G -оптимальных.

Рассмотрим следующий пример.

Пример 2.2. Пусть вектор регрессионных функций имеет вид

$$f(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}.$$

В качестве множества планирования рассмотрим интервал $\chi = [-1, 1]$. Предположим, что имеется возможность провести 3 измерения ($N=3$). Таким образом, D -оптимальный план мы ищем в классе планов

$$\xi^* = \begin{pmatrix} x_1 & x_2 & x_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix},$$

причем по крайней мере две точки из трех различны (в противном случае, план будет вырожденным по теореме 2.1). Запишем информационную матрицу этого плана:

$$M(\xi^*) = \sum_{i=1}^3 \frac{1}{3} f(x_i) f^T(x_i) = \begin{pmatrix} 1 & \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} & \frac{x_1^2+x_2^2+x_3^2}{3} \end{pmatrix}.$$

Найдем определитель этой матрицы:

$$\begin{aligned} \det(M(\xi^*)) &= \frac{2}{9}(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3 - x_1x_3) = \\ &= \frac{1}{9}((x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2). \end{aligned}$$

Без ограничения общности можно положить $x_1 \leq x_2 \leq x_3$. Анализ производной функции по x_1 и x_3 показывает, что ее максимум по x_1 достигается при $x_1 = -1$, а по x_3 — при $x_3 = 1$. Подставляя эти значения, получаем квадратичную функцию x_2 , которая достигает максимума при $x_2 = -1$ и $x_2 = 1$. Таким образом, имеется два эквивалентных D -оптимальных плана

$$\xi_1^* = \begin{pmatrix} -1 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}, \quad \xi_2^* = \begin{pmatrix} -1 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix},$$

причем $\det(M(\xi_1^*)) = \det(M(\xi_2^*)) = 8/9$. Прямыми вычислениями получаем

$$M(\xi_1^*) = \begin{pmatrix} 1 & -1/3 \\ -1/3 & 1 \end{pmatrix}, \quad D(\xi_1^*) = \begin{pmatrix} 9/8 & 3/8 \\ 3/8 & 9/8 \end{pmatrix}$$

и

$$d(x, \xi_1^*) = f^T(x)D(\xi_1^*)f(x) = \frac{1}{8}(3x+1)^2 + 1.$$

Анализ этой квадратичной функции показывает, что

$$\max_{x \in [-1,1]} d(x, \xi_1^*) = 3;$$

аналогично получаем, что

$$\max_{x \in [-1,1]} d(x, \xi_2^*) = \max_{x \in [-1,1]} \frac{1}{8}(3x-1)^2 + 1 = 3.$$

С другой стороны, для плана

$$\bar{\xi} = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

имеем

$$M(\bar{\xi}) = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}, \quad D(\bar{\xi}) = \begin{pmatrix} 1 & 0 \\ 0 & 3/2 \end{pmatrix}$$

и $d(x, \bar{\xi}) = 1 + 3/2x^2$, причем

$$\max_{x \in [-1, 1]} d(x, \bar{\xi}) = 2.5 < 3.$$

Следовательно, D -оптимальные планы ξ_1^* , ξ_2^* не являются G -оптимальными.

Существует несколько разновидностей теоремы эквивалентности для разных критериев оптимальности. Мы приведем некоторые из них без доказательств.

Теорема эквивалентности для E -критерия. Класс всех неотрицательно определенных матриц A таких, что $\text{tr} A = 1$ обозначим за $\mathcal{A}$.

Теорема 2.4. *Для модели (2.1) план $\xi^* \in \Xi_n$ является E -оптимальным тогда и только тогда, когда существует матрица $A^* \in \mathcal{A}$ такая, что для всех $x \in \chi$*

$$f^T(x) A^* f(x) \leq \lambda_{\min}(M(\xi^*)).$$

Кроме того,

$$\begin{aligned} \min_{A \in \mathcal{A}} \max_{x \in \chi} f^T(x) A f(x) &= \max_{\xi} \lambda_{\min}(M(\xi)), \\ f^T(x_i^*) A^* f(x_i^*) &= \lambda_{\min}(M(\xi^*)), \end{aligned}$$

где $x_i^*, i = 1, \dots, n$ — опорные точки E -оптимального плана.

Доказательство и примеры использования этой теоремы можно найти в [6].

Теорема эквивалентности для L -критерия. Для матрицы $L = \sum_{i=1}^k l_i l_i^T$ с заданными векторами $l_i \in R^m$ определим множество Ξ_L как множество всех непрерывных планов, для которых линейные комбинации параметров $l_i^T \theta, i = 1, \dots, k$, оцениваемы.

Определение 2.4. Для произвольной квадратной вырожденной матрицы A *обратной матрицей Мура* будем называть матрицу A^+ , удовлетворяющую условиям

$$\begin{aligned} (a) \quad & AA^+A = A, & (b) \quad & A^+AA^+ = A^+, \\ (c) \quad & (AA^+)^T = AA^+, & (d) \quad & (A^+A)^T = A^+A. \end{aligned}$$

Определение 2.5. Будем говорить, что непрерывный план η принадлежит классу Ξ_L^* , если $\eta \in \Xi_L$ и для любого непрерывного плана ξ существует предел

$$\lim_{\alpha \rightarrow 0} f^T(t)M^+(\xi_\alpha)LM^+(\xi_\alpha)f(t) = f^T(t)M^+(\eta)LM^+(\eta)f(t),$$

где $\xi_\alpha = (1 - \alpha)\eta + \alpha\xi$, $\alpha \in [0, 1]$.

Определение 2.6. Вырожденный план $\xi^* \in \Xi_L^*$ будем называть *L-оптимальным*, если

$$\xi^* = \arg \min_{\xi \in \Xi_L^*} \text{tr} LM^+(\xi),$$

где L — фиксированная, неотрицательно определенная матрица.

Имеет место следующая теорема, которая является модификацией известного результата для невырожденных планов (см. [3], теорема 2.4, стр. 112).

Теорема 2.5. Рассмотрим уравнение регрессии (2.1). Пусть матрица L имеет вид $L = \sum_{i=1}^k l_i l_i^T$ с заданными векторами $l_i \in R^m$. Предположим, что существует оптимальный план $\xi^* \in \Xi_L^*$. Тогда

1) план ξ принадлежит классу Ξ_L , если и только если для всех векторов l_i выполнено

$$l_i^T M^-(\xi)M(\xi) = l_i^T, \quad i = 1, \dots, k;$$

2) план $\xi^* \in \Xi_L^*$ является L -оптимальным, если и только если выполнено

$$a) \max_{t \in \chi} \varphi(t, \xi^*) = \text{tr} LM^+(\xi^*),$$

$$\text{где } \varphi(t, \xi) = f^T(t) M^+(\xi) L M^+(\xi) f(t),$$

при этом в точках $t_i \in \text{supp}(\xi^*)$ имеет место равенство

$$b) \varphi(t_i, \xi^*) = \text{tr} LM^+(\xi^*).$$

Доказательство этой теоремы можно найти в [27].

Теорема эквивалентности для e_k -критерия. Для e_k -критерия справедлива следующая теорема эквивалентности.

Теорема 2.6. Обозначим за $\bar{f}_k(t)$ вектор, полученный удалением компоненты $f_k(t)$ из вектора $f(t) = (f_1(t), \dots, f_m(t))^T$, $k = 1, \dots, m$. План ξ^* оптимален для оценивания параметра θ_k в модели (2.1) (в предположениях (a)–(e)) тогда и только тогда, когда $\exists h > 0$, $h \in R$ и вектор $q \in R^{m-1}$ такой, что функция

$$\varphi(t) = f_k(t) - q^T \bar{f}_k(t)$$

удовлетворяет следующим условиям:

- 1) $h\varphi^2(t) \leq 1$ для любых $t \in \chi$;
- 2) $\text{supp}(\xi^*) \subset \{t \in \chi \mid h\varphi^2(t) = 1\}$;
- 3) $\int_{\chi} \varphi(t) \bar{f}_k(t) \xi^*(dt) = 0 \in R^{m-1}$.

Кроме того, в этом случае $h = e_k^T M^-(\xi) e_k$ и функция φ называется экстремальным полиномом.

Доказательство теоремы можно найти в [24]

2.6. Применение теорем эквивалентности

В данном параграфе, мы рассмотрим несколько конкретных задач построения оптимальных планов на примере тригонометрической и полиномиальной моделей. Эти модели представляют значительный теоретический и практический интерес в связи с их универсальностью. Теоремы эквивалентности, рассмотренные нами ранее, являются полезными инструментами для решения подобных задач. С их помощью, в ряде случаев, удастся найти оптимальные планы в явном виде. Начнем с D -критерия оптимальности и теоремы Кифера—Вольфовица.

2.6.1. Полиномиальная модель

Рассмотрим полиномиальную регрессионную модель

$$y_j = \theta^T f(x) + \epsilon_j, \quad j = 1, \dots, N, \quad (2.4)$$

со стандартными предположениями об ошибках ϵ_j и вектором регрессионных функций

$$f(x) = (f_1(x), f_2(x), \dots, f_m(x))^T = (1, x, \dots, x^{m-1})^T.$$

Имеет место следующая теорема.

Теорема 2.7. *Для полиномиальной регрессии (2.4) на отрезке $[-1, 1]$ непрерывный D -оптимальный план существует, единственен и сосредоточен с равными весами в m точках, которые являются корнями $(x^2 - 1)P'_{m-1}(x)$, где $P_{m-1}(x)$ — полином Лежандра порядка $m - 1$.*

Доказательство. Существование D -оптимального плана вытекает из непрерывности регрессионных функций и компактности множества информационных матриц. Для доказательства остальных

утверждений воспользуемся теоремой Кифера—Вольфовица. Согласно этой теореме для D -оптимального плана ξ^*

$$\max_x d(x, \xi^*) = \max_x f^T(x) D(\xi^*) f(x) = \max_x \sum_{i,j=1}^m d_{i,j}(\xi^*) x^{i+j-2} = m,$$

где $d_{i,j}(\xi^*)$ — элементы матрицы $D(\xi^*)$. Из этого представления вытекает, что функция $d(x, \xi^*)$ является полиномом степени $2m - 2$ (в силу положительной определенности матрицы $D(\xi^*)$, ее диагональные элементы положительны, в частности, $d_{m,m} > 0$), и значит, у этой функции может быть не больше, чем m точек локального максимума, включая концы отрезка $[-1, 1]$. По теореме Кифера—Вольфовица максимум достигается в точках оптимального плана. С другой стороны, число точек оптимального плана не может быть меньше m . В противном случае оптимальный план будет вырожденным. Следовательно, план ξ^* имеет вид

$$\xi^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{m-1} & x_m \\ \omega_1 & \omega_2 & \dots & \omega_{m-1} & \omega_m \end{pmatrix},$$

$$\text{где } -1 = x_1 < x_2 < \dots < x_{m-1} < x_m = 1.$$

Информационную матрицу плана ξ^* мы можем записать в виде произведения трех матриц

$$\begin{aligned} M(\xi^*) &= \sum_{i=1}^m f(x_i) f^T(x_i) \omega_i = \\ &= \underbrace{\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \\ \vdots & & \vdots \\ x_1^{m-1} & \dots & x_m^{m-1} \end{pmatrix}}_F \underbrace{\begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_m \end{pmatrix}}_W \underbrace{\begin{pmatrix} 1 & x_1 & \dots & x_1^{m-1} \\ 1 & x_2 & \dots & x_2^{m-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^{m-1} \end{pmatrix}}_{F^T}. \end{aligned}$$

Из этого следует, что

$$\det(M(\xi^*)) = \det(FWF^T) = \det(F^T)^2 \prod_{i=1}^m \omega_i.$$

Определитель $\det(F^T)$ называется определителем Вандермонда и вычисляется рекурсивно. Обозначим за $\Delta_{m-1} = \det(F^T)$ определитель порядка m . Тогда

$$\Delta_{m-1} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 1 & x_2 - x_1 & \dots & x_2^{m-1} - x_2^{m-2}x_1 \\ \vdots & \vdots & & \vdots \\ 1 & x_m - x_1 & \dots & x_m^{m-1} - x_m^{m-2}x_1 \end{vmatrix} = \prod_{i=2}^m (x_i - x_1) \Delta_{m-2}.$$

Таким образом, получаем

$$\Delta_{m-1} = \prod_{1 \leq i < j \leq m} (x_i - x_j).$$

Найдем x_i -е, доставляющие максимальное значение этой функции. Продифференцируем по x_i , $i = 2, \dots, m-1$, и получим систему уравнений

$$\begin{aligned} \frac{\partial \Delta_{m-1}}{\partial x_i} &= 0 \Leftrightarrow \\ \Leftrightarrow \frac{1}{x_i - x_1} + \dots + \frac{1}{x_i - x_{i-1}} + \frac{1}{x_i - x_{i+1}} + \dots + \frac{1}{x_i - x_m} &= 0. \end{aligned}$$

Рассмотрим функцию $\varphi(x) = (x - x_2) \dots (x - x_{m-1})$. Нашу систему можно переписать следующим образом:

$$\begin{aligned} \frac{1}{x_i + 1} + \frac{1}{x_i - 1} + \frac{\varphi''(x_i)}{2\varphi'(x_i)} &= 0, \quad i = 2, \dots, m-1 \Leftrightarrow \\ (x_i^2 - 1)\varphi''(x_i) + 4x_i\varphi'(x_i) &= 0, \quad i = 2, \dots, m-1 \end{aligned}$$

Данный многочлен имеет ту же степень и те же корни, что и многочлен $\varphi(x)$, следовательно,

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) = \text{const}\varphi(x).$$

Приравнивая коэффициенты при x^{m-2} , находим $\text{const} = m(m-1) - 2$ и получаем дифференциальное уравнение

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) - (m(m-1) - 2)\varphi(x) = 0. \quad (2.5)$$

Рассмотрим уравнение

$$(x^2 - 1)P''(x) + 2xP'(x) - m(m-1)P(x) = 0,$$

где $P(x)$ — многочлен степени $m-1$. Из теории ортогональных многочленов известно (см. [10]), что единственным, с точностью до постоянного множителя, решением этого уравнения является полином Лежандра $m-1$ -го порядка:

$$P_{m-1}(x) = \frac{1}{2^{m-1}(m-1)!} \frac{\partial^{m-1}}{\partial x^{m-1}} (x^2 - 1)^{m-1}.$$

Очевидно, что P_{m-1} также является решением уравнения

$$\begin{aligned} ((x^2 - 1)P''(x) + 2xP'(x) - m(m-1)P(x))' &= 0 \Leftrightarrow \\ (x^2 - 1)P'''(x) + 4xP''(x) - (m(m-1) - 2)P'(x) &= 0, \end{aligned}$$

а это означает, что $\overline{\varphi}(x) = P'_{m-1}(x)$ является решением уравнения (2.5).

Решая систему

$$\begin{cases} \frac{\partial \prod_{i=1}^m \omega_i}{\partial \omega_i} = (\omega_m - \omega_i) \prod_{j=1, j \neq i}^m \omega_j = 0, & i = 1, \dots, m-1, \\ \omega_m = 1 - \sum_{i=1}^{m-1} \omega_i, \end{cases}$$

получаем, что $\omega_1 = \omega_2 = \dots = \omega_m = 1/m$.

Теорема доказана. □

Рассмотрим пример.

Пример 2.3. Пусть $m = 4$, т.е. вектор регрессионных функций модели (2.4) имеет вид $f(x) = (1, x, x^2, x^3)^T$. В качестве интервала планирования возьмем $\chi = [-1, 1]$. В этом случае, согласно доказанной теореме, D -оптимальный план имеет вид

$$\xi^* = \begin{pmatrix} -1 & x_2^* & x_3^* & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix},$$

где $x_2^* = -1/\sqrt{5}$, $x_3^* = 1/\sqrt{5}$ являются корнями производной полинома Лежандра третьего порядка

$$P_3'(x) = \frac{3}{2}(5x^2 - 1).$$

2.6.2. Тригонометрическая модель

Рассмотрим регрессионную модель (2.1) с функцией регрессии

$$\eta(x, \theta) = \theta_0 + \sum_{j=1}^m \theta_{2j-1} \sin(jx) + \sum_{j=1}^m \theta_{2j} \cos(jx), \quad (2.6)$$

θ — вектор неизвестных параметров: $\theta = (\theta_0, \theta_1, \dots, \theta_{2m})^T$, а $f(t)$ — вектор регрессионных функций: $f(t) = (1, \sin t, \cos t, \dots, \sin(mt), \cos(mt))^T$. В качестве интервала планирования будем рассматривать интервал $\chi = [-\pi, \pi]$. Приведем без доказательства несколько теорем, определяющих вид оптимальных планов для этой модели для различных критериев оптимальности.

D -оптимальные планы. Для тригонометрической модели (2.6) вид D -оптимального плана определяет следующая теорема.

Теорема 2.8. Пусть $\chi = [-\pi, \pi]$. Непрерывным D -оптимальным планом для тригонометрической регрессионной модели является любой план

$$\xi_N^* = \begin{pmatrix} t_1^* & \dots & t_N^* \\ 1/N & \dots & 1/N \end{pmatrix},$$

$$t_i^* = \frac{i-1}{N}2\pi - \pi, \quad i = 1, \dots, N, N \geq 2m+1,$$

где m — порядок регрессионной модели.

Также D -оптимальным является равномерный план

$$\xi^* = \frac{1}{2\pi} dx.$$

Доказательство. В силу того, что для любых натуральных i, j выполнены равенства

$$\begin{aligned} \int_{-\pi}^{\pi} \sin ix \, dx &= \int_{-\pi}^{\pi} \cos jx \, dx = \int_{-\pi}^{\pi} \cos jx \sin ix \, dx = 0, \\ \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin^2 ix \, dx &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 - \cos 2ix}{2} dx = \frac{1}{2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos^2 jx \, dx, \\ \sin ix \cos jx &= \frac{1}{2} (\sin(i-j)x + \sin(i+j)x), \\ \sin ix \sin jx &= \frac{1}{2} (\cos(i-j)x - \cos(i+j)x), \\ \cos ix \cos jx &= \frac{1}{2} (\cos(i-j)x + \cos(i+j)x), \end{aligned}$$

информационная матрица равномерного плана имеет следующий вид:

$$M(\xi^*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) f^T(x) \xi(dx) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \frac{1}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{2} \end{pmatrix}.$$

В силу теоремы 2.3 для проверки оптимальности плана ξ^* достаточно показать, что $d(x, \xi^*) = f^T(x) D(\xi^*) f(x) = 1 + 2m$. Действительно, обратная матрица есть

$$M^{-1}(\xi^*) = D(\xi^*) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 \end{pmatrix},$$

и следовательно, $d(x, \xi) = 1 + 2(\sin^2 x + \cos^2 x + \dots + \sin^2 mx + \cos^2 mx) = 1 + 2m$. Оптимальность плана ξ_N^* следует из того, что для любого $j = 1, 2, \dots, 2m$ имеют место равенства

$$\begin{aligned} \sum_{i=1}^N \sin jt_i^* &= \frac{1}{2} \frac{\cos \frac{j\pi}{N} - \cos(2j\pi - \frac{j\pi}{N})}{\sin(\frac{j\pi}{N})} = 0, \\ \sum_{i=1}^N \cos jt_i^* &= \frac{1}{2} \frac{\sin(2j\pi - \frac{j\pi}{N}) + \sin(\frac{j\pi}{N})}{\sin(\frac{j\pi}{N})} = 0. \end{aligned}$$

Теорема доказана. □

Замечание 2.5. Планы, удовлетворяющие условию теоремы 2.8, являются также E -оптимальными планами для модели (2.6) на интервале планирования $[-\pi, \pi]$ (см. [6], теорема 3.1).

e_k -оптимальные планы. Для тригонометрической модели (2.6) вид e_k -оптимального плана на полном интервале планирования определяют следующие две теоремы.

Теорема 2.9. *Рассмотрим тригонометрическую регрессионную модель (2.6) на интервале планирования $[-\pi, \pi]$.*

(a) *Для любого l , такого, что $m/3 < l \leq m$, и любого $\beta \in [0, \frac{1}{2l}]$ план*

$$\xi_{2l}^* = \begin{pmatrix} -\pi & -\pi + \frac{\pi}{l} & \dots & -\pi + \frac{2l-1}{l}\pi & \pi \\ \frac{1}{2l} - \beta & \frac{1}{2l} & \dots & \frac{1}{2l} & \beta \end{pmatrix} \quad (2.7)$$

оптимален для оценивания параметра θ_{2l} . Более того, в этом случае $\Phi_{2l}(\xi_{2l}^*) = 1$.

(b) *Для любого l , такого, что $m/2 < l \leq m$, план ξ_{2l}^* , определенный в (2.7), оптимален для оценки β_0 .*

(c) *Для любого l , такого, что $m/3 < l \leq m$, план*

$$\xi_{2l-1}^* = \begin{pmatrix} -\pi + \frac{\pi}{2l} & -\pi + \frac{3\pi}{2l} & \dots & -\pi + \frac{4l-3}{2l}\pi & \frac{-\pi}{2l} + \pi \\ \frac{1}{2l} & \frac{1}{2l} & \dots & \frac{1}{2l} & \frac{1}{2l} \end{pmatrix}$$

оптимален для оценивания параметра θ_{2l-1} . Более того, в этом случае $\Phi_{2l-1}(\xi_{2l-1}^*) = 1$.

Доказательство данной теоремы можно найти в работе [23].

Обозначим через $\lfloor N \rfloor$ целую часть произвольного положительного числа N и рассмотрим случай $1 \leq l \leq m/3$. Для этого случая e_k -оптимальные планы определены в следующей теореме.

Теорема 2.10. *Рассмотрим тригонометрическую регрессионную модель (2.6). Пусть $m \geq 3$ и $1 \leq l \leq m/3$.*

(a) *Пусть $p = \lfloor \frac{m+3l}{2l} \rfloor$. План*

$$\xi_{2l-1}^* = \begin{pmatrix} -t_n & \dots & -t_1 & t_1 & \dots & t_n \\ \omega_n & \dots & \omega_1 & \omega_1 & \dots & \omega_n \end{pmatrix}, \text{ где } n = l(p-1),$$

$$t_i = \left(i + \left\lfloor \frac{i-1}{p-1} \right\rfloor \right) \cdot \frac{\pi}{lp}, \quad i = 1, \dots, n,$$

$$\omega_i = \frac{|\sin(lt_i)|}{2 \sum_{j=1}^n |\sin(lt_j)|}, \quad i = 1, \dots, n,$$

является оптимальным для оценивания параметра θ_{2l-1} . При этом значение критерия оптимальности определяется по формуле

$$\Phi_k(\xi_{2l-1}^*) = \left(\frac{2}{lp} \sum_{i=1}^n |\sin(lt_i)| \right)^2 = \left(\frac{2}{p} \cot \left(\frac{\pi}{2p} \right) \right)^2. \quad (2.8)$$

(b) Если $p = \lfloor \frac{m+3l}{2l} \rfloor$ – нечетное, то план

$$\xi_{2l}^* = \begin{pmatrix} -t_n & \dots & -t_1 & t_1 & \dots & t_n \\ \omega_n & \dots & \omega_1 & \omega_1 & \dots & \omega_n \end{pmatrix}, \quad \text{где } n = l(p-1),$$

$$t_i = \left(2i - 1 + 2 \left\lfloor \frac{i-1}{p-1} + \frac{1}{2} \right\rfloor \right) \cdot \frac{\pi}{2lp}, \quad i = 1, \dots, n,$$

$$\omega_i = \frac{|\cos(lt_i)|}{2 \sum_{j=1}^n |\cos(lt_j)|}, \quad i = 1, \dots, n,$$

является оптимальным для оценивания параметра θ_{2l} . При этом значение критерия оптимальности определяется по формуле

$$\Phi_k(\xi_{2l}^*) = \left(\frac{2}{lp} \sum_{i=1}^n |\cos(lt_i)| \right)^2 = \left(\frac{2}{p} \cot \left(\frac{\pi}{2p} \right) \right)^2. \quad (2.9)$$

(c) Если $p = \lfloor \frac{m+3l}{2l} \rfloor$ – четное, то план

$$\xi_{2l}^* = \begin{pmatrix} -\pi & -t_n & \dots & -t_2 & 0 & t_2 & \dots & t_n & \pi \\ \mu & \omega_n & \dots & \omega_2 & \omega_1 & \omega_2 & \dots & \omega_n & \omega_1 - \mu \end{pmatrix}, \quad \text{где}$$

$$\mu \in [0, \omega_1], \quad n = l(p-1),$$

$$t_i = \left(2(i-1) + 2 \left\lfloor \frac{i-1}{p-1} + \frac{1}{2} \right\rfloor \right) \cdot \frac{\pi}{2lp}, \quad i = 1, \dots, n,$$

$$\omega_i = \frac{|\cos(lt_i)|}{2 \sum_{j=1}^n |\cos(lt_j)|}, \quad i = 1, \dots, n,$$

является оптимальным для оценивания параметра θ_{2l} . При этом значение критерия оптимальности определяется по формуле (2.9).

Доказательство. Доказательство основано на применении теоремы 2.6, но требует нескольких вспомогательных результатов.

Докажем пункт (а). Нам необходимо проверить то, что план, определенный в пункте (а), является оптимальным для оценивания коэффициента β_{2l-1} при $1 \leq l \leq m/3$. Доказательство будет основано на явном построении экстремального полинома φ из теоремы 2.6. Отметим, что $p \geq 3$, т.к. $l \leq m/3$. Вначале мы докажем пункт (а) для случая $l = 1$. Рассмотрим тригонометрический полином

$$Q(t, a) = \sum_{j=1}^{p-1} a_j \sin((2j-1)t), \quad (2.10)$$

где $(a_1, \dots, a_{p-1})^T \in R^{p-1}$ — заданный вектор. Мы определим этот вектор так, чтобы функция $Q(t, a)$ удовлетворяла следующим условиям:

$$\begin{aligned} Q(t_i, a) &= \frac{\sin(t_i)}{|\sin(t_i)|}, & i &= 1, \dots, \left\lfloor \frac{p}{2} \right\rfloor, \\ Q'(t_i, a) &= 0, & i &= 1, \dots, \left\lfloor \frac{p-1}{2} \right\rfloor, \end{aligned} \quad (2.11)$$

где точки $t_1, \dots, t_{\lfloor \frac{p}{2} \rfloor}$ определены в пункте (а). Заметим, что система уравнений (2.11) эквивалентна системе

$$Ba = e, \quad (2.12)$$

где вектор $e \in R^{p-1}$ определяется как

$$e = \left(\frac{\sin(t_1)}{|\sin(t_1)|}, \dots, \frac{\sin(t_{\lfloor \frac{p}{2} \rfloor})}{|\sin(t_{\lfloor \frac{p}{2} \rfloor})|}, 0, \dots, 0 \right)^T,$$

а матрица B имеет вид

$$B = \begin{pmatrix} \sin t_1 & \sin 3t_1 & \dots & \sin((2p-3)t_1) \\ \vdots & \vdots & & \vdots \\ \sin t_{\lfloor \frac{p}{2} \rfloor} & \sin 3t_{\lfloor \frac{p}{2} \rfloor} & \dots & \sin((2p-3)t_{\lfloor \frac{p}{2} \rfloor}) \\ \cos t_1 & 3 \cos 3t_1 & \dots & (2p-3) \cos((2p-3)t_1) \\ \vdots & \vdots & & \vdots \\ \cos t_{\lfloor \frac{p-1}{2} \rfloor} & 3 \cos 3t_{\lfloor \frac{p-1}{2} \rfloor} & \dots & (2p-3) \cos((2p-3)t_{\lfloor \frac{p-1}{2} \rfloor}) \end{pmatrix}.$$

Докажем, что $\det B \neq 0$. Предположим, напротив, что определитель матрицы B равен нулю. Тогда B — вырожденная матрица. Следовательно, существует ненулевой вектор $\tilde{a} \in R^{p-1}$ такой, что

$$B\tilde{a} = 0.$$

Это равносильно тому, что многочлен

$$\tilde{Q}(t, a) = \sum_{j=1}^{p-1} \tilde{a}_j \sin((2j-1)t)$$

степени $(2p-3)$ имеет $2p$ корней на интервале $[0, \pi]$. Но это невозможно, следовательно $\det B \neq 0$.

Отсюда следует, что система уравнений (2.12) имеет единственное решение, скажем $\hat{a} = (\hat{a}_1, \dots, \hat{a}_{p-1})^T$. Далее мы исследуем свойства функции $Q(t, \hat{a})$, соответствующей этому решению. Из условия ортонормированности

$$\frac{2}{\pi} \int_0^\pi \sin(it) \sin(jt) dt = \begin{cases} 1, & i = j \neq 0, \\ 0, & i \neq j, \end{cases}$$

получаем, что коэффициент при $\sin t$ функции $Q(t, \widehat{a})$ представим в виде

$$\widehat{a}_1 = \frac{2}{\pi} \int_0^\pi Q(t, \widehat{a}) \sin t dt. \quad (2.13)$$

Используя тригонометрические формулы для синусов и косинусов суммы двух углов, можно по индукции проверить, что функция $Q(t, \widehat{a}) \sin t$ может быть представлена в виде многочлена

$$Q(t, \widehat{a}) \sin t = q_1 \cos^{2p-2} t + \dots + q_{p-1} \cos^2 t, \quad (2.14)$$

где $q = (q_1, \dots, q_{p-1}) \in R^{p-1}$ — вектор коэффициентов, полученный в результате соответствующих преобразований. Обозначим этот многочлен через $P_{2p-2}(t, q)$.

Лемма 2.9. *Верна следующая формула:*

$$\frac{1}{\pi} \int_0^\pi \cos^{2j} t dt = \frac{1}{p} \sum_{k=1}^p \cos^{2j} \left(\frac{k\pi}{p} \right) = \frac{C_{2j}^j}{2^{2j}},$$

где $C_{2j}^j = \frac{(2j)!}{j!j!}$, $j = 1, \dots, p-1$.

Доказательство. Непосредственное вычисление дает

$$\begin{aligned} \frac{1}{\pi} \int_0^\pi \cos^{2j} t dt &= \frac{(2j-1)}{\pi} \left(\int_0^\pi \cos^{2j-2} t dt - \int_0^\pi \cos^{2j} t dt \right) \Rightarrow \\ &\Rightarrow \frac{1}{\pi} \int_0^\pi \cos^{2j} t dt = \frac{2j-1}{2j} \int_0^\pi \cos^{2(j-1)} t dt = \\ &= \frac{(2j-1) \cdot (2j-3) \cdot \dots \cdot 3}{(2j) \cdot (2j-2) \cdot \dots \cdot 2} = \frac{C_{2j}^j}{2^{2j}}. \end{aligned}$$

Далее, с помощью формулы Эйлера $\left(\cos(t) = \frac{e^{it} + e^{-it}}{2}\right)$ найдем сумму

$$\begin{aligned} & \frac{1}{p} \sum_{k=1}^p \cos^{2j} \left(\frac{k\pi}{p} \right) = \\ &= \frac{1}{p \cdot 2^{2j-1}} \sum_{k=1}^p \left(C_{2j}^0 \cos \left(\frac{2jk\pi}{p} \right) + C_{2j}^1 \cos \left(\frac{2(j-1)k\pi}{p} \right) + \dots + \frac{C_{2j}^j}{2} \right). \end{aligned}$$

Заметим, что для любого $j = 1, \dots, p-1$ имеет место равенство

$$\sum_{k=1}^p \cos \left(\frac{2jk\pi}{p} \right) = \frac{1 - \cos(\frac{2j\pi}{p}) + \cos(\frac{2j\pi}{p}) - \cos(2j\pi)}{2(\cos(\frac{2j\pi}{p}) - 1)} = 0.$$

Это можно проверить, умножая обе части первого равенства на $2(\cos(\frac{2j\pi}{p}) - 1)$ и проводя вычисления по тригонометрическим формулам из доказательства теоремы 2.8.

Отсюда получаем

$$\frac{1}{p} \sum_{k=1}^p \cos^{2j} \left(\frac{k\pi}{p} \right) = \frac{1}{p \cdot 2^{2j-1}} \sum_{k=1}^p \frac{C_{2j}^j}{2} = \frac{C_{2j}^j}{2^{2j}} \quad \text{при } j = 1, 2, \dots, p-1.$$

Лемма доказана. □

Из доказанной леммы непосредственно вытекает

Следствие 2.2. *Имеет место следующее равенство:*

$$\begin{aligned} \frac{1}{\pi} \int_0^\pi P_{2p-2}(t, q) \, dt &= \frac{1}{\pi} \int_0^\pi (q_1 \cos^{2p-2} t + \dots + q_{p-1} \cos^2 t) \, dt = \\ &= \frac{1}{p} \sum_{k=1}^p P_{2p-2} \left(\frac{k\pi}{p}, q \right). \end{aligned}$$

Теперь, с учетом (2.10), (2.13), (2.14) и доказанного следствия, получаем

$$\begin{aligned}
\widehat{a}_1 &= \frac{2}{\pi} \int_0^\pi Q(t, \widehat{a}) \sin t \, dt = \frac{2}{p} \sum_{k=1}^{p-1} Q\left(\frac{k\pi}{p}, \widehat{a}\right) \sin \frac{k\pi}{p} = \\
&= \frac{2}{p} \sum_{k=1}^{p-1} Q(t_k, \widehat{a}) \sin t_k = \frac{2}{p} \sum_{k=1}^{p-1} \frac{\sin t_k}{|\sin t_k|} \sin t_k = \\
&= \frac{2}{p} \sum_{k=1}^{p-1} \sin t_k = \frac{\cos(\frac{\pi}{2p}) - \cos(\pi - \frac{\pi}{2p})}{p \cdot \sin(\frac{\pi}{2p})} = \frac{2}{p} \cot\left(\frac{\pi}{2p}\right). \quad (2.15)
\end{aligned}$$

Перед тем, как продолжать доказательство теоремы 2.10, сформулируем еще одну лемму, которая понадобится нам в дальнейшем.

Лемма 2.10. *Полином $Q(t, \widehat{a})$, определенный в (2.10), удовлетворяет неравенству*

$$|Q(t, \widehat{a})| \leq 1, \quad \forall t \in [-\pi, \pi]. \quad (2.16)$$

Доказательство. Отметим, что тригонометрический многочлен $Q(t, \widehat{a})$ имеет степень $2p - 3$ и симметричен относительно $\frac{\pi}{2}$. На интервале $[0, \pi]$ данный многочлен имеет не больше, чем $2p - 3$ точек экстремума. Из условия (2.11) и того, что $Q(0, \widehat{a}) = 0$, следует, что точки t_i , $i = 1, \dots, p - 1$ являются точками максимума многочлена $Q(t, \widehat{a})$, и на каждом из интервалов $[t_i, t_{i+1}]$, $i = 1, \dots, p - 2$ данный многочлен имеет ровно один локальный минимум.

Сравним функцию $Q(t, \widehat{a})$ с функцией $\cos(2pt)$. Для этой цели введем тригонометрический полином

$$U(t) = Q(t, \widehat{a}) - \cos(2pt). \quad (2.17)$$

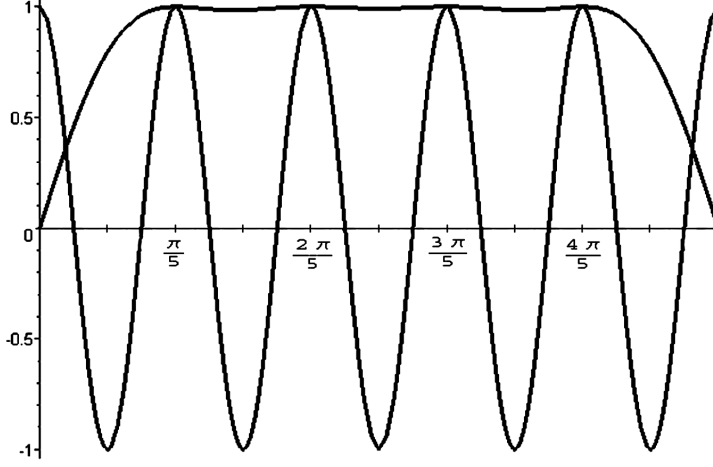


Рис. 3. Функция $Q(t, \hat{a})$ и функция $\cos(2pt)$ в случае $p = 5$. Полином $U(t)$, определенный в (2.17), имеет четыре корня второй кратности в точках $t_1 = \frac{\pi}{5}$, $t_2 = \frac{2\pi}{5}$, $t_3 = \frac{3\pi}{5}$ и $t_4 = \frac{4\pi}{5}$. Оба графика пересекаются на интервалах $(0, \frac{\pi}{5})$ и $(\frac{4\pi}{5}, \pi)$

Из определения точек t_i и (2.11) следует, что

$$U(t_i) = 0 \text{ и } U'(t_i) = 0, \quad i = 1, \dots, p-1.$$

Таким образом, на интервале $[0, \pi]$ полином $U(t)$ имеет $2p - 2$ корня с учетом кратности. Кроме того, $U(t)$ имеет корень на интервале $[0, \frac{\pi}{p}]$, так как функция $Q(t, \hat{a})$ изменяется на интервале $[0, \frac{\pi}{p}]$ от 0 до 1, а функция $\cos(2pt)$ изменяется на интервалах $[0, \frac{\pi}{2p}]$ и $[\frac{\pi}{2p}, \frac{\pi}{p}]$ от 1 до -1 и от -1 до 1. Также в силу симметрии $U(t)$ имеет корень на интервале $[\frac{(p-1)\pi}{p}, \pi]$. В результате мы получили, что на интервале $[0, \pi]$ функция $U(t)$ имеет $2p$ корней. Так как $U(t)$ — тригонометрический многочлен $2p$ -й степени и $U(0) = U(\pi) = -1 \neq 0$, на интервале $[0, \pi]$ других корней у этого многочлена быть не может.

Теперь предположим, что существует точка t^* такая, что

$$Q(t^*, \hat{a}) < -1.$$

В силу доводов, приведенных выше, точка t^* принадлежит одному из интервалов $(t_i, t_i + 1)$, $i = 1, \dots, p - 2$. Получается, что на этом интервале многочлен $U(t)$ имеет два дополнительных корня, а это невозможно. Следовательно, мы доказали

$$|Q(t, \hat{a})| \leq 1 \quad \forall t \in [0, \pi].$$

Неравенство (2.16) следует из того, что $|Q(t, \hat{a})| = |Q(-t, \hat{a})|$. Лемма доказана. $\square$

Поведение функций $Q(t, \hat{a})$ и $\cos(2pt)$ для случая $p = 5$ показано на рис. 3.

Теперь вернемся к доказательству теоремы. Рассмотрим функцию

$$\varphi(t) = Q(t, \hat{a})/\hat{a}_1 \quad (2.18)$$

и отметим, что коэффициент при $\sin t$ равен 1. То есть функция $\varphi(t)$ может быть представлена в виде

$$\varphi(t) = \sin t - \tilde{a}^T \bar{f}_1(t)$$

с соответствующим вектором $\tilde{a}$ и

$$\bar{f}_1(t) = (1, \cos t, \sin 2t, \cos 2t, \dots, \sin mt, \cos mt)^T.$$

Докажем, что функция $\varphi(t)$ удовлетворяет условиям (1)–(3) теоремы 2.6. Определим $h = \hat{a}_1^2$. Из (2.16) и (2.18) получаем

$$h\varphi^2(t) \leq 1 \quad \text{при } \forall t \in [-\pi, \pi], \quad (2.19)$$

причем в точках $\pm t_i$, $i = 1, \dots, n$ достигается равенство. Таким образом, функция $\varphi(t)$ удовлетворяет условиям (1) и (2) теоремы 2.6.

Осталось доказать, что $\varphi(t)$ удовлетворяет условию (3) этой теоремы. Отметим, что $\varphi(-t_i) = -\varphi(t_i)$, $i = 1, \dots, n$. Из этого получаем

$$\begin{aligned} & \int_{-\pi}^{\pi} \varphi(t) \cos(jt) d\xi_1^*(t) = \\ & = \sum_{i=1}^n (\varphi(t_i) \cos(jt_i) - \varphi(t_i) \cos(-jt_i)) \omega_i = 0 \quad \text{при } j = 0, \dots, m. \end{aligned}$$

Необходимо проверить, что

$$\int_{-\pi}^{\pi} \varphi(t) \sin(jt) d\xi_1^*(t) = 0, \quad j = 2, 3, \dots, m.$$

Имеем

$$\begin{aligned} \int_{-\pi}^{\pi} \varphi(t) \sin(jt) d\xi_1^*(t) &= 2 \sum_{i=1}^n \varphi(t_i) \sin(jt_i) \omega_i = \\ &= 2 \sum_{i=1}^n \frac{\sin(t_i)}{|\sin(t_i)|} \sin(jt_i) \frac{\sin(t_i)}{2 \sum_{k=1}^n |\sin(t_k)|} = \\ &= 2 \sum_{i=1}^n \sin(jt_i) \frac{\sin(t_i)}{2 \sum_{k=1}^n \sin(t_k)} = \operatorname{tg} \left(\frac{\pi}{2p} \right) \cdot \sum_{i=1}^{p-1} \sin(jt_i) \sin(t_i) = \\ &= \operatorname{tg} \left(\frac{\pi}{2p} \right) \frac{(-1)^j (\cos(\frac{\pi}{p}) - \cos(\frac{j\pi}{p}))}{4(\cos(\frac{\pi}{p}) - \cos(\frac{j\pi}{p}))} = 0, \quad j = 2, \dots, 2p-2. \end{aligned}$$

Так как $m \leq 2 \lfloor \frac{m+3}{2} \rfloor - 2 \leq 2p-2$, получаем

$$\int_{-\pi}^{\pi} \varphi(t) \bar{f}_1(t) d\xi_1^*(t) = 0 \in R^{2m}.$$

Таким образом, мы доказали, что функция $\varphi(t)$ удовлетворяет условию (3) теоремы 2.6. Следовательно, план ξ_1^* является оптимальным для оценивания параметра β_1 в тригонометрической регрессионной модели (2.6). Равенство в (2.8) следует из (2.15). Пункт (a) теоремы 2.10 в случае $l = 1$ доказан.

Докажем пункт (a) в общем случае, т.е. при $2 \leq l \leq m/3$. Рассмотрим функцию

$$\tilde{\varphi}(t) = \varphi(lt),$$

где $\varphi(t)$ определена в (2.18). Очевидно, в силу (2.19) и определения точек t_i функция $\tilde{\varphi}(t)$ удовлетворяет условиям (1) и (2) теоремы 2.6. Докажем, что эта функция удовлетворяет условию (3). Необходимо проверить, что

$$\int_{-\pi}^{\pi} \varphi(lt) \sin(jt) d\xi_{2l-1}^*(t) = 0, \quad j = 1, 2, 3, \dots, l-1, l+1, \dots, m.$$

Имеем

$$\begin{aligned} \int_{-\pi}^{\pi} \varphi(lt) \sin(jt) d\xi_{2l-1}^*(t) &= 2 \sum_{i=1}^n \varphi(lt_i) \sin(jt_i) \omega_i = \\ &= 2 \sum_{i=1}^n \frac{\sin(lt_i)}{|\sin(lt_i)|} \sin(jt_i) \frac{|\sin(lt_i)|}{2 \sum_{k=1}^n |\sin(lt_k)|} = \\ &= \frac{1}{l} \operatorname{tg} \left(\frac{\pi}{2p} \right) \cdot \left(\sum_{i=1}^{l(p-1)} \sin(jt_i) \sin(lt_i) \right) = \\ &= \frac{1}{l} \operatorname{tg} \left(\frac{\pi}{2p} \right) \sum_{i=1}^{lp} \sin \left(j \frac{i\pi}{lp} \right) \sin \left(l \frac{i\pi}{lp} \right) = \\ &= \frac{1}{l} \operatorname{tg} \left(\frac{\pi}{2p} \right) \frac{(-1)^{j-l} - (-1)^{j+l} (\cos(\frac{\pi}{p}) - \cos(\frac{j\pi}{lp}))}{4(\cos(\frac{\pi}{p}) - \cos(\frac{j\pi}{lp}))} = 0. \end{aligned}$$

Последнее равенство выполнено при любых $j = 1, \dots, 2lp - 2$, $j \neq l$.

Так как $m \leq 2l \left\lfloor \frac{m+3l}{2l} \right\rfloor - 2 \leq 2lp - 2$, получаем

$$\int_{-\pi}^{\pi} \varphi(t) \bar{f}_l(t) d\xi_{2l-1}^*(t) = 0 \in R^{2m},$$

где $\bar{f}_l(t) = f(t) \setminus \{\sin(lt)\}$.

Таким образом, мы доказали, что функция $\tilde{\varphi}(t)$ удовлетворяет условию (3) теоремы 2.6. Следовательно, план ξ_{2l-1}^* является оптимальным для оценивания параметра β_{2l-1} в тригонометрической регрессионной модели (2.6) при $1 \leq l \leq m/3$. Пункт (a) теоремы 2.10 доказан. Пункты (b) и (c) доказываются аналогично.

Теорема 2.10 доказана. □

Пример 2.4. Рассмотрим тригонометрические регрессионные модели порядков $m = 3$ и $m = 4$. Планы, оптимальные для оценивания коэффициентов θ_0 и θ_i ($i \geq 3$), могут быть непосредственно получены с помощью теоремы 2.9. Нас интересуют два оставшихся оптимальных плана. Начнем с плана, оптимального для оценивания коэффициента θ_1 . Заметим, что в этом случае

$$p = \left\lfloor \frac{m+3}{2} \right\rfloor = 3 \text{ при } m = 3, 4.$$

В соответствии с пунктом (a) теоремы 2.10 план, оптимальный для оценивания коэффициента θ_1 в тригонометрической регрессионной модели (2.6) порядка m ($m = 3, 4$), имеет вид

$$\xi_1^* = \begin{pmatrix} -\frac{2\pi}{3} & -\frac{\pi}{3} & \frac{\pi}{3} & \frac{2\pi}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Критерий оптимальности принимает значение $\Phi_1(\xi_1^*) = \frac{4}{3}$. Экстремальный полином $\varphi(t)$, удовлетворяющий условиям (1)–(3) теоремы 2.6, имеет вид

$$\varphi(t) = \sin(t) + \frac{1}{6} \sin(3t).$$

Согласно пункту (с) теоремы 2.10 план, оптимальный для оценивания коэффициента β_2 в тригонометрической регрессионной модели (2.6) порядка m ($m = 3, 4$), имеет вид

$$\xi_2^* = \begin{pmatrix} -\frac{5\pi}{6} & -\frac{\pi}{6} & \frac{\pi}{6} & \frac{5\pi}{6} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Критерий оптимальности принимает значение $\Phi_2(\xi_2^*) = \frac{4}{3}$. Экстремальный полином $\varphi(t)$, удовлетворяющий условиям (1)–(3) теоремы 2.6, имеет вид

$$\varphi(t) = \cos(t) - \frac{1}{6} \cos(3t).$$

Точки носителей двух оптимальных планов изображены на рис. 4. Точки носителей планов ξ_1^* и ξ_2^* обозначены, как a_j и b_j соответственно.

Пример 2.5. Рассмотрим тригонометрические регрессионные модели порядков $m = 5$ и $m = 6$. Если $m = 5$, то только оптимальные планы для оценивания коэффициентов β_1 и β_2 не могут быть получены с помощью теоремы 2.9.

В соответствии с пунктом (а) теоремы 2.10 план, оптимальный для оценивания коэффициента β_1 тригонометрической регрессионной модели (2.6) порядка $m = 5$, имеет вид

$$(2.3.1) \quad \xi_1^* = \begin{pmatrix} -\frac{3\pi}{4} & -\frac{\pi}{2} & -\frac{\pi}{4} & \frac{\pi}{4} & \frac{\pi}{2} & \frac{3\pi}{4} \\ \frac{1}{4+2\sqrt{2}} & \frac{\sqrt{2}}{4+2\sqrt{2}} & \frac{1}{4+2\sqrt{2}} & \frac{1}{4+2\sqrt{2}} & \frac{\sqrt{2}}{4+2\sqrt{2}} & \frac{1}{4+2\sqrt{2}} \end{pmatrix}.$$

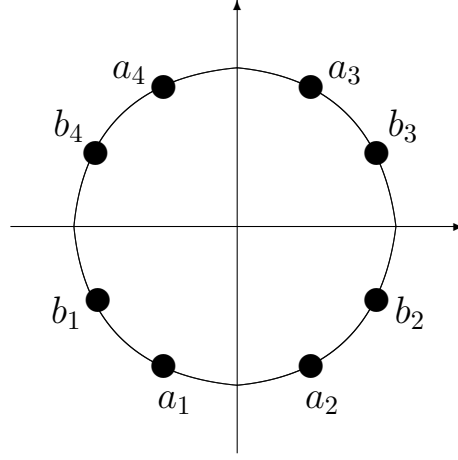


Рис. 4. Точки носителей оптимальных планов ξ_1^* и ξ_2^* в тригонометрических регрессионных моделях порядка $m = 3$ и $m = 4$. Точки планов ξ_1^* и ξ_2^* обозначены как a_j и b_j соответственно.

Критерий оптимальности принимает значение $\Phi_1(\xi_1^*) = \frac{3+2\sqrt{2}}{4}$. Экстремальный полином $\varphi(t)$, удовлетворяющий условиям (1)-(3) теоремы 2.6, имеет вид

$$(2.3.2) \quad \varphi(t) = \sin(t) + \left(\frac{8-5\sqrt{2}}{4}\right) \sin(3t) + \left(\frac{3\sqrt{2}-4}{4}\right) \sin(5t).$$

Согласно пункту (b) теоремы 2.10 план, оптимальный для оценивания коэффициента β_2 тригонометрической регрессионной модели (2.1.1) порядка $m = 5$, имеет вид

$$(2.3.3) \quad \xi_2^* = \begin{pmatrix} -\pi & -\frac{3\pi}{4} & -\frac{\pi}{4} & 0 & \frac{\pi}{4} & \frac{3\pi}{4} & \pi \\ \mu & \frac{\sqrt{2}}{4+4\sqrt{2}} & \frac{\sqrt{2}}{4+4\sqrt{2}} & \frac{2}{4+4\sqrt{2}} & \frac{\sqrt{2}}{4+4\sqrt{2}} & \frac{\sqrt{2}}{4+4\sqrt{2}} & \frac{2}{4+4\sqrt{2}} - \mu \end{pmatrix},$$

при $\forall \mu \in [0, \frac{2}{(4+4\sqrt{2})}]$. Критерий оптимальности принимает значение $\Phi_2(\xi_2^*) = \frac{3+2\sqrt{2}}{4}$. Экстремальный полином $\varphi(t)$, удовлетворяющий

условиям (1)-(3) теоремы 2.6, имеет вид

$$(2.3.4) \quad \varphi(t) = \cos(t) - \left(\frac{8 - 5\sqrt{2}}{4}\right) \cos(3t) + \left(\frac{3\sqrt{2} - 4}{4}\right) \cos(5t).$$

Графики полиномов (2.3.2) и (2.3.4) изображены на рисунке 2.3, положение на круге точек носителей оптимальных планов показано на рисунке 2.4.

В тригонометрической регрессионной модели порядка $m = 6$ планы, оптимальные для оценивания коэффициентов $\beta_1, \beta_2, \beta_3, \beta_4$, не могут быть получены с помощью теоремы 2.1. Для того, чтобы найти эти планы, необходимо воспользоваться теоремой 2.2. Планы, оптимальные для оценивания коэффициентов β_1 и β_2 , построены в (2.3.1) и (2.3.3) соответственно. План, оптимальный для оценивания коэффициента β_3 , сосредоточен в восьми точках

$$\pm\frac{\pi}{6}, \pm\frac{2\pi}{6}, \pm\frac{4\pi}{6}, \pm\frac{5\pi}{6}$$

с одинаковыми весами.

Критерий оптимальности принимает значение $\Phi_3(\xi_3^*) = \frac{4}{3}$. Экстремальный полином $\varphi(t)$, удовлетворяющий условиям (1)-(3) теоремы 2.6, имеет вид

$$\varphi(t) = \sin(2t) + \frac{1}{6} \sin(6t).$$

И, наконец, план, оптимальный для оценивания коэффициента β_4 , сосредоточен в восьми точках

$$\pm\frac{\pi}{12}, \pm\frac{5\pi}{12}, \pm\frac{7\pi}{12}, \pm\frac{11\pi}{12}$$

с одинаковыми весами.

Критерий оптимальности принимает значение $\Phi_4(\xi_4^*) = \frac{4}{3}$. Экстремальный полином $\varphi(t)$, удовлетворяющий условиям (1)-(3) теоремы 2.6, имеет вид

$$\varphi(t) = \cos(2t) - \frac{1}{6} \cos(6t).$$

Точки носителей планов ξ_3^* и ξ_4^* изображены на рисунке 2.5.

Пример 2.3. Обычно для тригонометрических регрессионных моделей используют планы, сосредоточенные в равноотстоящих точках с равными весами. Причем число различных точек носителя полагается не меньше, чем $2m+1$. Как отмечено в работах (Федоров, 1971; Ермаков, Жиглявский, 1987), эти планы оптимальны в смысле D -критерия для МНК-оценки всего вектора параметров. Интересно исследовать эффективность этих планов относительно оценивания младших коэффициентов тригонометрической регрессионной модели (2.1.1). Напомним, что дисперсионная матрица равномерного плана ξ_u , т.е. плана, сосредоточенного в равноотстоящих точках с равными весами, имеет вид

$$D(\xi_u) = \text{diag}(1, 2, \dots, 2) \in R^{2m+1 \times 2m+1},$$

а e_k -эффективность равномерного плана в данном случае имеет вид

$$eff_{e_k}(\xi_u) = \frac{\Phi_k(\xi_k^*)}{\Phi_k(\xi_u)} = (1 - \delta_{k0})^{-1} \left(\frac{2}{p} \cot\left(\frac{\pi}{2p}\right) \right)^2,$$

где $k = 2l, 2l - 1$, $1 \leq l \leq m/3$ и $p = \lfloor \frac{m+3l}{2l} \rfloor$. Например, если $m = 6$ и $k = 4(l = 2)$ или $k = 3(l = 2)$, мы получим

$$eff_{e_3}(\xi_u) = eff_{e_4}(\xi_u) = \frac{2}{3},$$

это означает, что потеря эффективности для оценивания коэффициентов β_3 и β_4 в тригонометрической регрессионной модели порядка

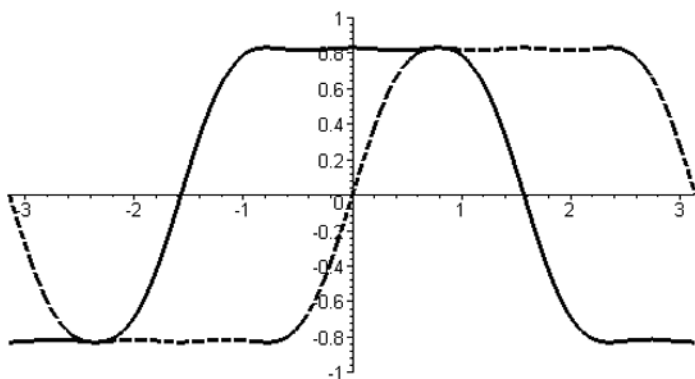


Рис. 5. Экстремальные функции из теоремы 1.6, соответствующие планам, оптимальным для оценивания коэффициентов β_1 и β_2 в тригонометрической регрессионной модели порядка $m = 5$. Функции определены в явном виде в (2.3.2) (пунктирная линия) и в (2.3.4) (непрерывная линия).

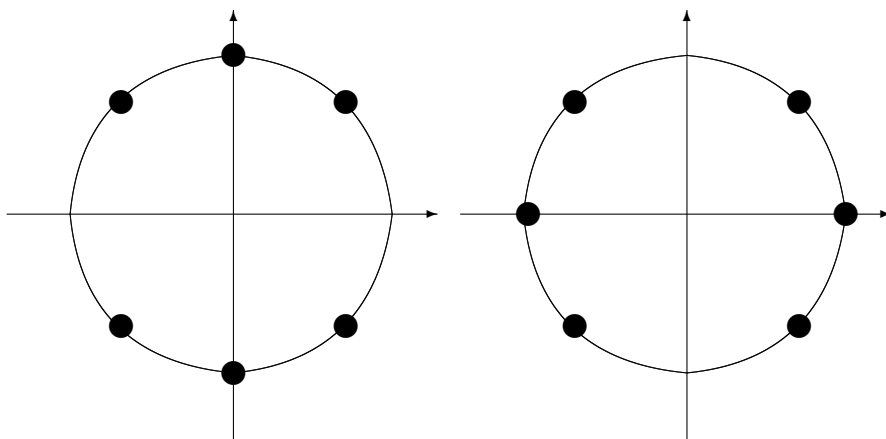


Рис. 6. Точки носителей оптимальных планов ξ_1^* (слева) и ξ_2^* (справа) в тригонометрической регрессионной модели порядка $m = 5$.

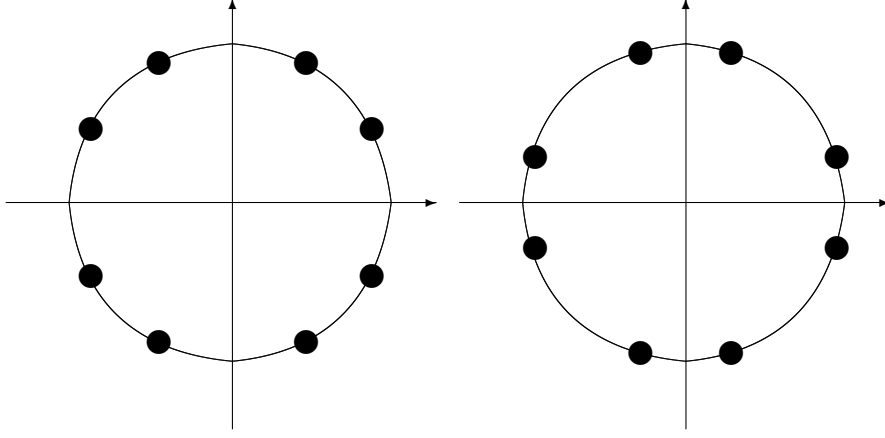


Рис. 7. Точки носителей оптимальных планов ξ_3^* (слева) и ξ_4^* (справа) в тригонометрической регрессионной модели порядка $m = 6$.

$m = 6$ будет примерно 33% при использовании равномерного плана. Однако при увеличении порядка модели, эффективность равномерного плана также увеличивается. В частности, если $m \rightarrow \infty$, мы получаем

$$\lim_{p \rightarrow \infty} \frac{2}{p} \cot \left(\frac{\pi}{2p} \right) = \frac{4}{\pi},$$

а предел функции эффективности равномерного плана равен

$$\lim_{m \rightarrow \infty} \text{eff}_{e_k}(\xi_u) = \frac{8}{\pi^2} \approx 0.8097, \quad k \geq 1$$

(отметим, что предел не зависит от k). Таким образом, если равномерный план используется в тригонометрической регрессии большого порядка (как, например, в Eubank, 1999), потеря эффективности для оценивания младших коэффициентов составляет 20%. Однако, если равномерный план используется в тригонометрических регрессионных моделях небольших порядков (Young, Ehrlich, 1977), потеря эффективности более существенная.

L -оптимальные планы. Рассмотрим тригонометрическую модель (2.6) степени $m = Nk$ и задачу построения L -оптимального плана для неотрицательно определенных матриц $L^{(k)} \in R^{(2m+1) \times (2m+1)}$ вида

$$\tilde{L}^{(k)} = P L^{(k)} P = \begin{pmatrix} L_{\cos}^{(k)} & 0 \\ 0 & L_{\sin}^{(k)} \end{pmatrix}, \quad (2.20)$$

где P — невырожденная матрица, удовлетворяющая уравнению

$$\widetilde{M}(\xi) = P M(\xi) P = \begin{pmatrix} M_c(\xi) & 0 \\ 0 & M_s(\xi) \end{pmatrix},$$

$$\begin{aligned} M_c(\xi) &= \left(\int_{-\pi}^{\pi} \cos(it) \cos(jt) d\xi(t) \right)_{i,j=0}^m, \\ M_s(\xi) &= \left(\int_{-\pi}^{\pi} \sin(it) \sin(jt) d\xi(t) \right)_{i,j=1}^m. \end{aligned}$$

Если план ξ является симметричным, то такая матрица всегда существует.

Под матрицами $L_{\cos}^{(k)} \in \mathbb{R}^{m+1 \times m+1}$ и $L_{\sin}^{(k)} \in \mathbb{R}^{m \times m}$ понимаются матрицы с элементами $L_{\cos,ij}^{(k)}$ и $L_{\sin,ij}^{(k)}$, определенными как

$$L_{\cos,uv}^{(k)} = \begin{cases} L_{\cos,ij}^{(1)}, & \text{если } u = (i-1)k+1, v = (j-1)k+1, \\ 0, & \text{в противном случае} \end{cases} \quad (2.21)$$

$(u, v = 1, \dots, Nk+1; \quad i, j = 1, \dots, N+1)$ и

$$L_{\sin,uv}^{(k)} = \begin{cases} L_{\sin,ij}^{(1)}, & \text{если } u = ik; v = jk, \\ 0, & \text{в противном случае} \end{cases} \quad (2.22)$$

$(u, v = 1, \dots, Nk)$. При этом $L_{\sin}^{(1)} = (L_{\sin, u, v}^{(1)})_{u, v=1}^N$ и $L_{\cos}^{(1)} = (L_{\cos, u, v}^{(1)})_{u, v=1}^{N+1}$ — заданные $N \times N$ - и $(N+1) \times (N+1)$ -матрицы.

В качестве иллюстрации рассмотрим случай $N = 2$ и матрицу $L_{\sin}^{(1)} \in \mathbb{R}^{2 \times 2}$ вида

$$L_{\sin}^{(1)} = \begin{pmatrix} L_{\sin \ 11}^{(1)} & L_{\sin \ 12}^{(1)} \\ L_{\sin \ 12}^{(1)} & L_{\sin \ 22}^{(1)} \end{pmatrix};$$

тогда для $k = 2$ получаем матрицу $L_{\sin}^{(2)} \in \mathbb{R}^{4 \times 4}$:

$$L_{\sin}^{(2)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & L_{\sin \ 11}^{(1)} & 0 & L_{\sin \ 12}^{(1)} \\ 0 & 0 & 0 & 0 \\ 0 & L_{\sin \ 21}^{(1)} & 0 & L_{\sin \ 22}^{(1)} \end{pmatrix}.$$

Аналогично, если матрица

$$L_{\cos}^{(1)} = \begin{pmatrix} L_{\cos \ 11}^{(1)} & L_{\cos \ 12}^{(1)} & L_{\cos \ 13}^{(1)} \\ L_{\cos \ 12}^{(1)} & L_{\cos \ 22}^{(1)} & L_{\cos \ 23}^{(1)} \\ L_{\cos \ 13}^{(1)} & L_{\cos \ 23}^{(1)} & L_{\cos \ 33}^{(1)} \end{pmatrix} \in \mathbb{R}^{3 \times 3},$$

то матрица $L_{\cos}^{(2)}$ имеет вид

$$L_{\cos}^{(2)} = \begin{pmatrix} L_{\cos \ 11}^{(1)} & 0 & L_{\cos \ 12}^{(1)} & 0 & L_{\cos \ 13}^{(1)} \\ 0 & 0 & 0 & 0 & 0 \\ L_{\cos \ 12}^{(1)} & 0 & L_{\cos \ 22}^{(1)} & 0 & L_{\cos \ 23}^{(1)} \\ 0 & 0 & 0 & 0 & 0 \\ L_{\cos \ 13}^{(1)} & 0 & L_{\cos \ 23}^{(1)} & 0 & L_{\cos \ 33}^{(1)} \end{pmatrix}.$$

Теперь сформулируем теорему, определяющую структуру оптимального плана для матриц L такого вида.

Теорема 2.11. *Рассмотрим тригонометрическую модель (2.6) степени $m = Nk$, $N, k \in \mathbb{N}$, $N \geq 2$, и пусть L – заданная матрица (2.20) с блоками $L_{\cos}^{(k)}$ и $L_{\sin}^{(k)}$, определенными в (2.21) и в (2.22) соответственно.*

1. Определим план $\xi_m^{\sin}$ как

$$\xi_m^{\sin} = \begin{pmatrix} -t_m & -t_{m-1} & \dots & -t_1 & t_1 & \dots & t_m \\ \omega_m & \omega_{m-1} & \dots & \omega_1 & \omega_1 & \dots & \omega_m \end{pmatrix}, \quad (2.23)$$

где

$$\begin{aligned} t_1 &= \frac{x_1}{k}, \dots, \quad t_{\frac{N}{2}} = \frac{x_{\frac{N}{2}}}{k}, \\ t_{\frac{N}{2}+1} &= \frac{\pi - x_{\frac{N}{2}}}{k}, \dots, \quad t_N = \frac{\pi - x_1}{k}, \\ t_i &= t_{i-N} + \frac{\pi}{k}, \quad i = N+1, \dots, m, \\ \omega_1 &= \frac{z_1}{k}, \dots, \quad \omega_{\frac{N}{2}-1} = \frac{z_{\frac{N}{2}-1}}{k}, \quad \omega_{\frac{N}{2}} = \frac{z_{\frac{N}{2}}}{k}, \\ \omega_{\frac{N}{2}+1} &= \frac{z_{\frac{N}{2}}}{k}, \dots, \quad \omega_N = \frac{z_1}{k}, \quad \sum_{j=1}^{\frac{N}{2}} z_j = \frac{1}{4}, \\ \omega_i &= \omega_{i-N}, \quad i = N+1, \dots, m, \end{aligned}$$

если N четное и

$$\begin{aligned}
t_1 &= \frac{x_1}{k}, \dots, \quad t_{\frac{N-1}{2}} = \frac{x_{\frac{N-1}{2}}}{k}, \quad t_{\frac{N+1}{2}} = \frac{\pi}{2k}, \\
t_{\frac{N+3}{2}} &= \frac{\pi - x_{\frac{N-1}{2}}}{k}, \dots, \quad t_N = \frac{\pi - x_1}{k}, \\
t_i &= t_{i-N} + \frac{\pi}{k}, \quad i = N+1, \dots, m, \\
\omega_1 &= \frac{z_1}{k}, \dots, \quad \omega_{\frac{N-1}{2}} = \frac{z_{\frac{N-1}{2}}}{k}, \\
\omega_{\frac{N+1}{2}} &= \frac{1}{2k} - 2 \sum_{j=1}^{\frac{N-1}{2}} \frac{z_j}{k}, \quad \omega_{\frac{N+3}{2}} = \frac{z_{\frac{N-1}{2}}}{k}, \dots, \quad \omega_N = \frac{z_1}{k}, \\
\omega_i &= \omega_{i-N}, \quad i = N+1, \dots, m,
\end{aligned}$$

если N — нечетное. Если матрица L — матрица вида (2.20) с $L_{\cos}^{(k)} = 0$, тогда для плана $\xi_n^{\sin}$ величины $\text{tr} L_{\sin}^{(k)} M_s^+(\xi_n^{\sin})$ и коэффициенты функции

$$\varphi(t, \xi_n^{\sin}) = f_s^T(t) M_s^+(\xi_n^{\sin}) L_{\sin}^{(k)} M_s^+(\xi_n^{\sin}) f_s(t)$$

не зависят от значений k для любой матрицы $L_{\sin}^{(k)} \in \mathbb{R}^{m \times m}$.

2. Определим план $\xi_n^{\cos}$ как

$$\xi_n^{\cos} = \begin{pmatrix} -\pi & -t_{n-1} & \dots & -t_1 & 0 & t_1 & \dots & t_{n-1} & \pi \\ \omega_n - \alpha & \omega_{n-1} & \dots & \omega_1 & \omega_0 & \omega_1 & \dots & \omega_{n-1} & \alpha \end{pmatrix},$$

где $\alpha \in [0, \omega_n]$, $n = (N+1)k$ и

$$\begin{aligned}
t_0 &= 0, \quad t_1 = \frac{x_1}{k}, \dots, \quad t_{\frac{N}{2}} = \frac{x_{\frac{N}{2}}}{k}, \\
t_{\frac{N}{2}+1} &= \frac{\pi - x_{\frac{N}{2}}}{k}, \dots, \quad t_N = \frac{\pi - x_1}{k},
\end{aligned}$$

$$\begin{aligned}
t_{i+1} &= t_{i-N} + \frac{\pi}{k}, \quad i = N, \dots, n-1, \\
\omega_0 &= \frac{1 - 4 \sum_{i=1}^{\frac{N}{2}} z_i}{2k}, \quad \omega_1 = \frac{z_1}{k}, \dots, \quad \omega_{\frac{N}{2}-1} = \frac{z_{\frac{N}{2}-1}}{k}, \\
\omega_{\frac{N}{2}} &= \frac{z_{\frac{N}{2}}}{k}, \quad \omega_{\frac{N}{2}+1} = \frac{z_{\frac{N}{2}}}{k}, \dots, \quad \omega_N = \frac{z_1}{k}, \\
\omega_{i+1} &= \omega_{i-N}, \quad i = N, \dots, n-1,
\end{aligned}$$

если N — четное, и

$$\begin{aligned}
t_0 &= 0, \quad t_1 = \frac{x_1}{k}, \dots, \quad t_{\frac{N-1}{2}} = \frac{x_{\frac{N-1}{2}}}{k}, \quad t_{\frac{N+1}{2}} = \frac{\pi}{2k}, \\
t_{\frac{N+3}{2}} &= \frac{\pi - x_{\frac{N-1}{2}}}{k}, \dots, \quad t_N = \frac{\pi - x_1}{k}, \\
t_{i+1} &= t_{i-N} + \frac{\pi}{k}, \quad i = N, \dots, n-1, \\
\omega_0 &= z_0, \quad \omega_1 = z_1, \dots, \quad \omega_{\frac{N-1}{2}} = z_{\frac{N-1}{2}}, \\
\omega_{\frac{N+1}{2}} &= \frac{1}{2k} - \sum_{j=0}^{\frac{N-1}{2}} \frac{z_j}{k}, \quad \omega_{\frac{N+3}{2}} = z_{\frac{N-1}{2}}, \dots, \quad \omega_N = z_1, \\
\omega_{i+1} &= \omega_{i-N}, \quad i = N, \dots, m,
\end{aligned}$$

если N — нечетное. Если матрица L — матрица вида (2.20) с $L_{\sin}^{(k)} = 0$, тогда для плана $\xi_n^{\cos}$ величины $\text{tr} L_{\cos}^{(k)} M_c^+(\xi_n^{\cos})$ и коэффициенты функции

$$\varphi(t, \xi_n^{\cos}) = f_c^T(t) M_c^+(\xi_n^{\cos}) L_{\cos}^{(k)} M_c^+(\xi_n^{\cos}) f_c(t)$$

не зависят от k для любой матрицы $L_{\cos}^{(k)} \in \mathbb{R}^{m+1 \times m+1}$.

Доказательство данной теоремы может быть найдено в [28].

Следствие 2.3. Рассмотрим тригонометрическую модель (2.6), $m > 3$.

1. План

$$\xi_{(2\lfloor \frac{m}{2} \rfloor - 1, 4\lfloor \frac{m}{2} \rfloor - 1)}^* = \begin{pmatrix} -t_n & -t_{n-1} & \dots & -t_1 & t_1 & \dots & t_n \\ \omega_n & \omega_{n-1} & \dots & \omega_1 & \omega_1 & \dots & \omega_n \end{pmatrix}, \text{ где}$$

$$n = 2 \left\lfloor \frac{m}{2} \right\rfloor, \quad t_i = 2 \left\lfloor \frac{i}{2} \right\rfloor \frac{\pi}{n} + (-1)^{(i-1)}x, \quad \omega_i = \frac{1}{2n}, \quad x = \frac{2 \arctg(\sqrt[4]{5})}{n}$$

является L -оптимальным планом, минимизирующим сумму дисперсий оценок параметров $\beta_{2\lfloor \frac{m}{2} \rfloor - 1}$ и $\beta_{4\lfloor \frac{m}{2} \rfloor - 1}$.

Причем

$$\text{tr} LM^+(\xi_{(2\lfloor \frac{m}{2} \rfloor - 1, 4\lfloor \frac{m}{2} \rfloor - 1)}^*) = \frac{\sqrt{5}}{2} + \frac{3}{2} \approx 2.618034 \dots$$

2. Для любого $\alpha \in [0, \omega_n]$ план

$$\xi_{(2\lfloor \frac{m}{2} \rfloor, 4\lfloor \frac{m}{2} \rfloor)}^* = \begin{pmatrix} -\pi & -t_{n-1} & \dots & -t_1 & 0 & t_1 & \dots & t_{n-1} & \pi \\ \omega_n - \alpha & \omega_{n-1} & \dots & \omega_1 & \omega_0 & \omega_1 & \dots & \omega_{n-1} & \alpha \end{pmatrix}, \text{ где}$$

$$n = 2 \left\lfloor \frac{m}{2} \right\rfloor, \quad t_i = \frac{(i-1)\pi}{n}, \quad \omega_0 = \sqrt{5}\omega_1, \quad \omega_1 = \frac{\sqrt{5}-1}{4n}, \quad \omega_i = \omega_{i-2}, \quad i = 2, \dots, n$$

является L -оптимальным планом, минимизирующим сумму дисперсий оценок параметров $\beta_{2\lfloor \frac{m}{2} \rfloor}$ и $\beta_{4\lfloor \frac{m}{2} \rfloor}$. Кроме того, план $\xi_{(0, 2\lfloor \frac{m}{2} \rfloor)}^*$, минимизирующий сумму дисперсий оценок параметров β_0 и $\beta_{2\lfloor \frac{m}{2} \rfloor}$, совпадает с планом $\xi_{(2\lfloor \frac{m}{2} \rfloor, 4\lfloor \frac{m}{2} \rfloor)}^*$.

Причем

$$\text{tr} LM^+(\xi_{(2\lfloor \frac{m}{2} \rfloor, 4\lfloor \frac{m}{2} \rfloor)}^*) = \text{tr} LM^+(\xi_{(0, 2\lfloor \frac{m}{2} \rfloor)}^*) = \frac{\sqrt{5}}{2} + \frac{3}{2} \approx 2.618034 \dots$$

Доказательство.

Докажем первый пункт Следствия 2.3. Второй доказывается аналогично. Для удобства обозначений будем считать, что m — четное. В этом случае $\lfloor \frac{m}{2} \rfloor = \frac{m}{2}$ и первый пункт теоремы можно переписать так:

$$\xi_{(\frac{m}{2}, m)}^* = \begin{pmatrix} -t_m & -t_{m-1} & \dots & -t_1 & t_1 & \dots & t_m \\ \frac{1}{2m} & \frac{1}{2m} & \dots & \frac{1}{2m} & \frac{1}{2m} & \dots & \frac{1}{2m} \end{pmatrix},$$

$$t_i = \left\lfloor \frac{i}{2} \right\rfloor \frac{2\pi}{m} + (-1)^{(i-1)}x, \quad x = \frac{2}{m} \arctg(\sqrt[4]{5}).$$

Идея доказательства состоит в следующем. Мы покажем, что

$$\begin{cases} m_{s[j, \frac{m}{2}]} = 0, & j = 1, 2, \dots, m, \quad j \neq \frac{m}{2} \\ m_{s[j, m]} = 0, & j = 1, 2, \dots, m-1 \end{cases}, \begin{cases} m_{s[\frac{m}{2}, \frac{m}{2}]} = \sin^2(\frac{m}{2}x) \\ m_{s[m, m]} = \sin^2(mx) \end{cases}.$$

где $m_{s[i, j]}$ — элемент матрицы $M_s \in R^{m \times m}$, стоящий на пересечении i -ой строки и j -ого столбца.

В результате мы получим функцию $\varphi(t, \xi^*)$ в явном виде

$$\varphi(t, \xi^*) = f^T(t) M^+(\xi) L M^+(\xi) f(t) = \frac{\sin^2(\frac{m}{2}t)}{\sin^4(\frac{m}{2}x)} + \frac{\sin^2(mt)}{\sin^4(mx)}.$$

После этого, решая уравнение $\frac{\partial \varphi(t, \xi^*)}{\partial t} \big|_{t=x} = 0$, найдем нужное значение x , а затем выведем, что

$$tr L M^+(\xi^*) = \varphi(t_i, \xi^*) = \frac{1}{\sin^2(\frac{m}{2}x)} + \frac{1}{\sin^2(mx)} = \frac{\sqrt{5}}{2} + \frac{3}{2}.$$

Итак, нужно лишь доказать, что

$$\begin{cases} m_{s[j, \frac{m}{2}]} = 0 & j = 1, 2, \dots, m, \quad j \neq \frac{m}{2} \\ m_{s[j, m]} = 0, & j = 1, 2, \dots, m-1 \end{cases}, \begin{cases} m_{s[\frac{m}{2}, \frac{m}{2}]} = \sin^2(\frac{m}{2}x) \\ m_{s[m, m]} = \sin^2(mx) \end{cases}.$$

С учетом вышесказанного получаем:

$$m_{s[j, \frac{m}{2}]} = 2 \sum_{k=1}^m \sin(jt_k) \sin(\frac{m}{2}t_k) \omega_k = \frac{1}{m} \sum_{k=1}^m \sin(jt_k) \sin(\frac{m}{2}t_k).$$

Отметим, что $\sin(\frac{m}{2}t_k) = (-1)^{\lfloor \frac{k-1}{2} \rfloor} \sin(\frac{m}{2}x)$. Следовательно, получаем:

$$m_{s[j, \frac{m}{2}]} = \frac{\sin(\frac{m}{2}x)}{m} \sum_{k=1}^m (-1)^{\lfloor \frac{k-1}{2} \rfloor} \sin(jt_k).$$

Докажем, что $\sum_{k=1}^m (-1)^{\lfloor \frac{k-1}{2} \rfloor} \sin(jt_k) = 0$ при $j \neq \frac{m}{2}$. В этом случае

$$\begin{aligned} \sum_{k=1}^m (-1)^{\lfloor \frac{k-1}{2} \rfloor} \sin(jt_k) &= \sin(jx) + \sin\left(\frac{2\pi j}{m}\right) \cos(jx) - \cos\left(\frac{2\pi j}{m}\right) \sin(jx) + \\ &+ (-1) \left(\sin\left(\frac{2\pi j}{m}\right) \cos(jx) + \cos\left(\frac{2\pi j}{m}\right) \sin(jx) \right) + \dots \\ &+ (-1)^{(\frac{m}{2}-1)} \left(\sin\left(\left(\frac{m}{2}-1\right)\frac{2\pi j}{m}\right) \cos(jx) + \cos\left(\left(\frac{m}{2}-1\right)\frac{2\pi j}{m}\right) \sin(jx) \right) + \\ &+ (-1)^{(j+1)+(\frac{m}{2}-1)} \sin(jx) = \sin(jx) \left(2 \sum_{k=0}^{\frac{m}{2}-1} (-1)^k \cos\left(\frac{2\pi k j}{m}\right) + (-1)^{j+\frac{m}{2}} - 1 \right). \end{aligned}$$

Как несложно видеть, при $j \neq \frac{m}{2}$ справедливо равенство

$$\begin{aligned} 2 \sum_{k=0}^{\frac{m}{2}-1} (-1)^k \cos\left(\frac{2\pi k j}{m}\right) &= \frac{1}{\cos(\frac{j\pi}{m})} \left(\cos\left(\frac{j\pi}{m}\right) + \cos\left(\frac{3j\pi}{m}\right) + \dots + \right. \\ &\left. + (-1)^{\frac{m}{2}-1} \left(\cos\left(\frac{(m-3)j\pi}{m}\right) + \cos\left(\frac{(m-1)j\pi}{m}\right) \right) \right) = 1 + (-1)^{j+\frac{m}{2}-1}. \end{aligned}$$

Следовательно, искомая сумма равна

$$\sum_{k=1}^m (-1)^{\lfloor \frac{k-1}{2} \rfloor} \sin(jt_k) = 2 \sin(jx) \left(1 + (-1)^{\frac{m}{2}-1+j} + (-1)^{j+\frac{m}{2}} - 1 \right) = 0.$$

Аналогичным образом посчитаем, чему равен элемент $m_{s[j,m]}$ при $j \neq m$

$$\begin{aligned} m_{s[j,m]} &= \frac{1}{m} \sum_{k=1}^m \sin(jt_k) \sin(mt_k) = \frac{\sin(mx)}{m} \sum_{k=1}^m (-1)^k \sin(jt_k) = \\ &= \sin(jx) \left(2 \sum_{k=0}^{\frac{m}{2}-1} \cos\left(\frac{2\pi k j}{m}\right) + (-1)^j - 1 \right) = \sin(jx) (1 - (-1)^j + (-1)^j - 1) = 0. \end{aligned}$$

Из вышесказанного следует, что

$$\begin{cases} m_{s[\frac{m}{2}, \frac{m}{2}]} = \sin^2(\frac{m}{2}x) \frac{1}{m} \sum_{k=1}^m (-1)^{2\lfloor \frac{k-1}{2} \rfloor} = \sin^2(\frac{m}{2}x) \\ m_{s[m,m]} = \sin^2(mx) \frac{1}{m} \sum_{k=1}^m (-1)^{2k} = \sin^2(mx) \end{cases}.$$

Теорема доказана.

□

Замечание 2.6. Отметим, что Следствие 2.3 верно для тригонометрических моделей порядка $m > 3$. При $m = 2$ оптимальный план $\xi_{(1,3)}^*$ имеет вид

$$\xi_{(1,3)}^* = \begin{pmatrix} -\pi + x & -x & x & \pi - x \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad x = \arctg(\sqrt[4]{5}),$$

$$tr L M_s^{-1}(\xi^*) = \frac{\sqrt{5}}{2} + \frac{3}{2}.$$

Для любого $\alpha \in \left[0, \frac{5-\sqrt{5}}{8}\right]$ оптимальный план $\xi_{(0,2)}^*$ равен

$$\xi_{(0,2)}^* = \begin{pmatrix} -\pi & -\frac{\pi}{2} & 0 & \frac{\pi}{2} & \pi \\ \frac{5-\sqrt{5}}{8} - \alpha & \frac{\sqrt{5}-1}{8} & \frac{5-\sqrt{5}}{8} & \frac{\sqrt{5}-1}{8} & \alpha \end{pmatrix}, \quad \text{tr}LM_c^{-1}(\xi^*) = \frac{\sqrt{5}}{2} + \frac{3}{2}.$$

При $m = 3$ оптимальный план $\xi_{(0,2)}^*$ равен

$$\xi_{(0,2)}^* = \begin{pmatrix} -\pi & -\pi + x & -x & 0 & x & \pi - x & \pi \\ \frac{1}{4} - z & z & z & \frac{1}{4} - z & z & z & \alpha \end{pmatrix}, \quad \alpha \in \left[0, \frac{1}{4} - z\right],$$

где $z \approx 0.151950668\dots$, $x \approx 0.9329288036\dots$, $\text{tr}LM_c^{-1}(\xi^*) \approx 2.770045647\dots$.

Численные значения z и x найдены путем непосредственного решения, соответствующих систем уравнений.

Замечание 2.7. Отметим, что в условиях Следствия 2.3 сумма дисперсий оценок соответствующих коэффициентов равна $\frac{\sqrt{5}}{2} + \frac{3}{2}$. При D -оптимальном планировании сумма дисперсий - равна 4. Таким образом, сумма дисперсий при L -оптимальном планировании в данном случае будет меньше суммы дисперсий, полученной при D -оптимальном планировании примерно в 1.5 раза.

В качестве иллюстрации того, как теорему 2.11 можно применять при построении оптимальных планов, рассмотрим небольшой пример.

Пример 2.5. L -оптимальный план для оценивания коэффициентов при $\sin(40t)$ и $\sin(50t)$ в тригонометрической регрессионной модели порядка $m = 50$.

Рассмотрим тригонометрическую модель (2.6) степени $m = Nk$ и воспользуемся теоремой 2.11 для нахождения L -оптимального плана для оценивания пары коэффициентов $\theta_{2(N-1)k-1}$ и θ_{2Nk-1} , которые

соответствуют компонентам регрессии $\sin((N-1)kt)$ и $\sin(Nkt)$. Пусть, для определенности, $m = 50$, $N = 5$, $k = 10$. Теорема 2.11 позволяет свести задачу построения L -оптимального плана для оценивания коэффициентов β_{79} и β_{99} для модели порядка $m = 50$ к задаче построения L -оптимального плана для оценивания коэффициентов β_7 и β_9 для модели порядка $m = 5$, для которой мы можем найти оптимальный план в явном виде непосредственно из системы уравнений

$$\begin{cases} \frac{\partial}{\partial x_i} \text{tr} L_{\sin}^{(1)} M_s^+(\xi_5^{\sin}) = 0, \\ \frac{\partial}{\partial z_i} \text{tr} L_{\sin}^{(1)} M_s^+(\xi_5^{\sin}) = 0, \end{cases}$$

где $\xi_5^{\sin}$ план со структурой, определенной в теореме 2.11. Таким образом, мы получаем

$$\xi_5^{\sin} = \begin{pmatrix} -\pi + x_1 & \dots & -x_1 & x_1 & x_2 & \frac{\pi}{2} & \pi - x_2 & \pi - x_1 \\ z_1 & \dots & z_1 & z_1 & z_2 & \frac{1-4z_1-4z_2}{2} & z_2 & z_1 \end{pmatrix},$$

где $x_1 = 0.3519978036$, $x_2 = 1.020599385$, $z_1 = 0.1112542423$, $z_2 = 0.09865639359$.

Из теоремы 2.11 непосредственно следует, что L -оптимальный план для оценивания коэффициентов $\beta_{2(N-1)k-1}$ и β_{2Nk-1} (т.е. β_{79} и β_{99} в нашем случае) имеет вид

$$\xi_{Nk}^{\sin} = \xi_{50}^{\sin} = \begin{pmatrix} -t_{50} & -t_{49} & \dots & -t_1 & t_1 & \dots & t_{50} \\ \omega_{50} & \omega_{49} & \dots & \omega_1 & \omega_1 & \dots & \omega_{50} \end{pmatrix},$$

первые 5 точек и весов определяются в явном виде:

$$\begin{aligned} t_1 &= \frac{x_1}{10}, \quad t_2 = \frac{x_2}{10}, \quad t_3 = \frac{\pi}{20}, \quad t_4 = \frac{\pi - x_2}{10}, \quad t_5 = \frac{\pi - x_1}{10}, \\ \omega_1 &= \frac{z_1}{10}, \quad \omega_2 = \frac{z_2}{10}, \quad \omega_3 = \frac{1}{20} - \frac{z_1}{5} - \frac{z_2}{5}, \quad \omega_4 = \frac{z_2}{10}, \quad \omega_5 = \frac{z_1}{10}, \end{aligned}$$

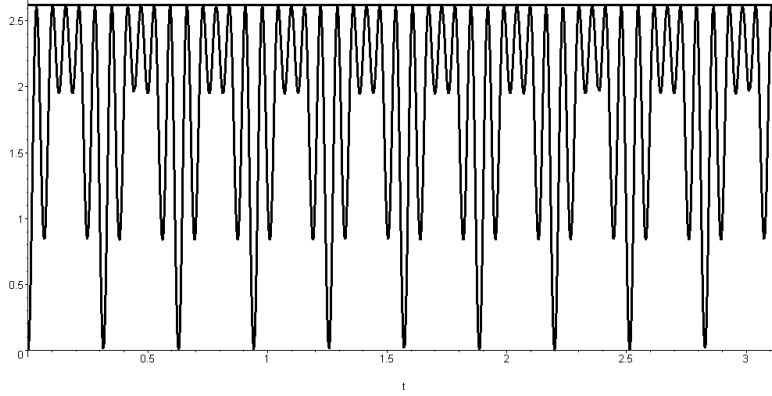


Рис. 8. Функция $\varphi(t, \xi_{79,99}^*)$ определенная в теореме эквивалентности 2.5 для L -оптимального плана рассмотренного в примере 2.5

а остальные — выражаются с помощью рекурсивных формул:

$$t_i = t_{i-5} + \frac{\pi}{10}, \quad \omega_i = \omega_{i-5}, \quad i = 6, \dots, 50,$$

где x_1, x_2, z_1, z_2 определены ранее. В заключение отметим, что экстремальная функция $\varphi(t, \xi_{79,99}^*)$ из теоремы эквивалентности 2.5 имеет вид

$$\begin{aligned} \varphi(t, \xi_{79,99}^*) &= f^T(t) M^+(\xi_{79,99}^*) L M^+(\xi_{79,99}^*) f(t) = \\ &= 0.2 \sin^2(30t) - 1.05 \sin(50t) \sin(30t) + 1.37 \sin^2(50t) + 2.09 \sin^2(40t). \end{aligned}$$

Поведение этой функции отображено на рис. 8.

2.7. Факторные планы

В сегодняшней лекции мы рассмотрим некоторые понятия и результаты, связанные с факторными экспериментами, с которых начала свое развитие теория планирования экспериментов. Впервые факторные эксперименты были предложены Фишером для решения агробиологических задач.

Рассмотрим серию регрессионных экспериментов, результаты измерений в которых зависят от значений k переменных (факторов) x_i , $i = 1, \dots, k$. До появления методов факторного планирования в каждом эксперименте подвергалось изменению значение только одного фактора. При факторном планировании изменяются уровни всех факторов. Факторные эксперименты часто проводятся в интересах дисперсионного анализа и служат для того, чтобы сократить, по возможности, число необходимых экспериментов. Статистический анализ выполняется единым образом, независимо от структуры (качественной или количественной) используемых факторов. Под качественным фактором в данном случае понимается переменная, принимающая конечное множество значений (называемых также уровнями факторов). Количественный фактор принимает значения из некоторого, определенного исследователем, интервала. Идея подхода заключается в том, что экспериментатор для каждого фактора x_i определяет некоторое конечное число ($s_i \geq 2$) значений (уровней), которые может принимать этот фактор. Каждому из этих уровней ставятся в соответствие символы $0, 1, \dots, s_i - 1$, независимо от того, является переменная x_i количественной или качественной, и вместо фактора x_i рассматривают некоторую абстракцию — фактор F_i , принимающий соответствующие переменной x_i значения $0, 1, \dots, s_i - 1$.

Факторным планом эксперимента называется совокупность всевозможных комбинаций уровней факторов, по которым могут проводиться измерения. Факторный план называется полным, если согласно этому плану измерения проводятся по одному для каждой возможной комбинации уровней. Полный факторный план требует $s_1 s_2 \dots s_k$ числа измерений. Факторный план называется дробным, если он предназначен для меньшего числа измерений. Дробные планы подробно рассматриваются в монографии [3] (глава 2).

Большое распространение на практике получили так называемые двухуровневые планы, т.е. такие планы, факторы которых принима-

ют значения только на двух уровнях. Это произошло по ряду причин. В частности, даже если фактор может принимать значения на многих уровнях, часто выбирают только два крайних, так как, как правило, чем больше различие в значениях факторов, тем точнее МНК-оценки параметров функции регрессии. Кроме того, в реальных задачах достаточно типична ситуация, когда факторы могут либо присутствовать, либо отсутствовать, т.е. имеют только два значения. При рассмотрении двухуровневых планов уровни факторов обычно кодируются числами -1 и 1 и условно называются верхним и нижним соответственно. Полный двухфакторный план обозначается так 2^k .

Схема регрессии, в рамках которой проводится факторный эксперимент, называется факторной моделью. Модели могут быть разными. Наиболее общая факторная модель имеет вид

$$\begin{aligned} \eta(x_1, \dots, x_k) = & \theta_0 + \sum_{i=1}^k \theta_i x_i + \sum_{i_1 < i_2} \theta_{i_1 i_2} x_{i_1} x_{i_2} + \dots + \\ & + \sum_{i_1 < i_2 < \dots < i_k} \theta_{i_1 i_2 \dots i_k} x_{i_1} x_{i_2} \dots x_{i_k}. \end{aligned} \quad (2.24)$$

Параметр θ_0 называется общим средним, параметры θ_i ($i = 1, \dots, k$) — главными эффектами (взаимодействиями нулевого порядка), $\theta_{i_1 i_2}$ — эффектами взаимодействий первого порядка (эффектами двухфакторных взаимодействий), $\theta_{i_1 i_2 \dots i_k}$ — эффектами взаимодействий порядка $k - 1$ (эффектами k -факторных взаимодействий).

Наиболее часто используются два частных случая функции регрессии (2.24)

$$\eta(x_1, \dots, x_k) = \theta_0 + \sum_{i=1}^k \theta_i x_i \quad (\text{линейная}) \quad (2.25)$$

и

$$\eta(x_1, \dots, x_k) = \theta_0 + \sum_{i=1}^k \theta_i x_i + \sum_{i_1 < i_2} \theta_{i_1 i_2} x_{i_1} x_{i_2} \text{ (квадратичная)}. \quad (2.26)$$

Последняя также называется неполной квадратичной в силу того, что коэффициенты при x_i^2 равны 0.

2.7.1. МНК-оценка

Рассмотрим метод вычисления МНК-оценок для функции регрессии (2.24) для измерений, проводимых по полному двухуровневому факторному плану.

Докажем следующую лемму.

Лемма 2.11. *Для любой функции регрессии вида (2.24) матрица X (см. раздел 1.2) для полного факторного плана 2^k обладает свойством*

$$X^T X = N I_s, \quad (2.27)$$

где $N = 2^k$, I_s — единичная матрица порядка s (s — количество параметров регрессии).

Доказательство. Обозначим столбцы матрицы X через

$$x_{(i)} = (x_{(i)1}, \dots, x_{(i)N})^T, \quad i = 0, 1, \dots, s-1.$$

В силу того, что все элементы матрицы X равны либо -1 , либо 1 , диагональные элементы матрицы $X^T X$ равны $x_{(i)}^T x_{(i)} = N$. Осталось проверить, что внедиагональные равны 0.

Рассмотрим случай, когда $s=N$, который соответствует полной факторной модели (2.24). Остальные случаи рассматриваются аналогично.

Будем использовать метод математической индукции. Для случая $k = 1$ матрица X имеет вид

$$X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

и $X^T X = 2I_2$. Пусть утверждение доказано для $k = m$. Докажем его для $k = m + 1$. Для определенности обозначим матрицу плана 2^m через X_m . Тогда матрицу плана 2^{m+1} можно записать в виде

$$X_{m+1} = \begin{pmatrix} X_m & X_m \\ X_m & -X_m \end{pmatrix},$$

где запись « $-X_m$ » означает, что все элементы матрицы X_m домножены на -1 . Проверим, что $X_{m+1}^T X_{m+1} = 2^{m+1} I_{2^{m+1}}$. Действительно,

$$\begin{aligned} X_{m+1}^T X_{m+1} &= \begin{pmatrix} X_m^T & X_m^T \\ X_m^T & -X_m^T \end{pmatrix} \begin{pmatrix} X_m & X_m \\ X_m & -X_m \end{pmatrix} = \\ &= \begin{pmatrix} 2X_m^T X_m & 0 \\ 0 & 2X_m^T X_m \end{pmatrix} = \begin{pmatrix} 2 \cdot 2^m I_{2^m} & 0 \\ 0 & 2 \cdot 2^m I_{2^m} \end{pmatrix} = 2^{m+1} I_{2^{m+1}}. \end{aligned}$$

Лемма доказана. □

Из этой леммы и из (1.11) следует, что МНК-оценка параметров модели (2.24) имеет вид

$$\hat{\theta}_i = \frac{1}{N} \sum_{j=1}^N x_{(ij)} y_j, \quad i = 1, \dots, k.$$

2.7.2. *D-оптимальные планы*

Пусть функция регрессии имеет вид (2.26), а множество планирования есть m -мерный куб ($\chi = [-1, 1]^m$). Рассмотрим дискретный

план

$$\xi_N = \begin{pmatrix} x_1 & \dots & x_N \\ 1/N & \dots & 1/N \end{pmatrix}, \quad (2.28)$$

где точка $x_i = (x_{i1}, \dots, x_{im})$ — m -мерная точка, координаты которой соответствуют значению факторов при i -м измерении. Матрица X плана ξ имеет вид

$$X = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1m} \\ x_{20} & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{N0} & x_{N1} & \dots & x_{Nm} \end{pmatrix},$$

где $x_{j0} \equiv 1$, $j = 0, 1, \dots, N$ — фиктивная $m + 1$ -я переменная, соответствующая параметру θ_0 . Информационная матрица плана ξ , по определению, есть матрица

$$M(\xi) = \frac{1}{N} X^T X.$$

Замечание 2.8. В силу леммы 2.11 информационная матрица полного факторного плана 2^m (т.е. плана ξ^* вида (2.28), точки которого сосредоточены во всех вершинах m -мерного куба) является единичной

$$M(\xi^*) = \frac{1}{N} X^T X = I_{m+1}.$$

Из этого следует, в частности, что максимум функции

$$d(x, \xi^*) = 1 + \sum_{i=1}^m x^2(i),$$

где $x(i)$ — координаты точки $x \in \chi$, достигается в точках плана ξ^* и равен $m + 1$. По теореме Кифера—Вольфовица (см. теорему 2.3) получаем, что план ξ^* является D -оптимальным.

Несмотря на ряд полезных свойств (ортогональность и D -оптимальность) полного факторного плана, при больших m на практике он используется редко. Причина этого в слишком большом числе измерений — 2^m .

Для практического использования значительный интерес представляют, в частности, планы, сосредоточенные в минимально возможном числе точек, равном $m + 1$. Такие планы принято называть насыщенными.

Имеет место следующее утверждение.

Лемма 2.12. *Насыщенный план ξ вида (2.28) имеет информационную матрицу, пропорциональную единичной тогда и только тогда, когда он сосредоточен в вершинах правильного m -мерного симплекса (т.е. правильного m -мерного многогранника с количеством вершин $m+1$).*

Доказательство. Обозначим строки матрицы X через $\bar{z}_{(i)}^T$, $i = 1, \dots, N$. По определению

$$\bar{z}_{(i)}^T = (1 \ \bar{x}_i^T),$$

где $\bar{x}_i$ — вектор, соответствующий m -мерной точке плана ξ .

Обозначим через φ_{ij} угол между векторами $\bar{x}_i$ и $\bar{x}_j$. По определению

$$\cos \varphi_{ij} = \frac{(\bar{x}_i, \bar{x}_j)}{\|\bar{x}_i\| \cdot \|\bar{x}_j\|}, \text{ где } (\bar{x}_i, \bar{x}_j) = \sum_{l=1}^m x_{il}x_{jl}, \quad \|\bar{x}_i\| = (\bar{x}_i, \bar{x}_i).$$

То, что информационная матрица плана ξ пропорциональна единичной, можно записать следующим образом:

$$(\bar{z}_i, \bar{z}_j) = \begin{cases} 0, & \text{если } i \neq j, \\ C, & \text{если } i = j, \end{cases} \quad (2.29)$$

где $C > 1$ — некоторая константа. В силу определения вектора $\bar{z}_i$ условие (2.29) можно переписать как

$$(\bar{x}_i, \bar{x}_j) = \begin{cases} -1, & \text{если } i \neq j, \\ C - 1, & \text{если } i = j, \end{cases}$$

что равным счетом означает, что

$$\cos \varphi_{ij} = \frac{1}{1 - C}, \quad j \neq i, \quad i, j = 1, 2, \dots, m.$$

Полученное условие эквивалентно тому, что точки x_i ($i = 1, \dots, N$) плана ξ лежат в вершинах правильного m -мерного симплекса. Планы, точки которых сосредоточены в вершинах правильного симплекса, называются симплекс-планами.

Некоторые дальнейшие сведения о факторных планах можно найти в монографии [3] (глава 2).

2.8. Задачи для самоконтроля

1. Найти D -оптимальный план:

(I) для модели

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3, \quad x \in [-1, 1];$$

(II) для модели

$$\eta(x, \theta) = \theta_1 + \theta_2 \sin x + \theta_3 \cos x + \theta_4 \sin^2 x + \theta_4 \cos^2 x, \quad x \in [-\pi, \pi].$$

2. Найти L -оптимальный план (в качестве матрицы L выбрать диагональную, единичную):

(I) для модели

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2, \quad x \in [-1, 1];$$

(II) для модели

$$\eta(x, \theta) = \theta_1 + \theta_2 \sin x + \theta_3 \cos x + \theta_4 \sin^2 x + \theta_4 \cos^2 x, \quad x \in [-\pi, \pi].$$

3. Найти e_2 -оптимальный план для модели

$$\eta(x, \theta) = \theta_0 + \sum_{j=1}^{20} \theta_{2j-1} \sin(jx) + \sum_{j=1}^{20} \theta_{2j} \cos(jx), \quad x \in [-\pi, \pi].$$

4. Пусть множество планирования имеет вид

$$\chi = \{x = (x_1, \dots, x_m))^T, \sum_{i=1}^m x_i^2 \leq 1\},$$

а функция регрессии равна

$$\eta(x_1, \dots, x_m) = \theta_0 + \sum_{i=1}^m \theta_i x_i + \sum_{i_1 < i_2} \theta_{i_1 i_2} x_{i_1} x_{i_2}.$$

Доказать, что D -оптимальным является план $\xi^* = \alpha \xi_1 + (1 - \alpha) \xi_2$, где $\alpha = 2/((m+1)(m+2))$, план ξ_1 сосредоточен в нуле, а план ξ_2 представляет собой равномерную меру на поверхности сферы

$$\left\{ x = (x_1, \dots, x_m)^T, \sum_{i=1}^m x_i^2 = 1 \right\}.$$

5. Для полиномиальной регрессионной модели n -го порядка на интервале $[-1, 1]$ показать, что план

$$\xi = \begin{pmatrix} -1 & \cos\left(\frac{n-1}{n}\pi\right) & \cos\left(\frac{n-2}{n}\pi\right) & \dots & \cos\left(\frac{1}{n}\pi\right) & 1 \\ \frac{1}{2n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{2n} \end{pmatrix}$$

является e_n -оптимальным.

6. Найти в явном виде E -оптимальный план для полиномиальной регрессионной модели 2-го порядка на интервале $[a, b]$, где a и b — произвольные вещественные числа ($a < b$).

Указание: оптимальный план искать на множестве двухточечных планов вида

$$\xi = \begin{pmatrix} x_1 & x_2 \\ \omega_1 & \omega_2 \end{pmatrix}.$$

7. Показать, что

- (I) $B(B^T B)^{-2} B^T$ является обобщенно обратной для $B B^T$,
- (II) $B^T (B B^T)^{-1} B = I_s$ для любой матрицы $B \in R^{k \times s}$, у которой $\text{rank}(B) = s$.

8. Найти матрицу L , для которой для полиномиальной регрессионной модели 3-го порядка на интервале $[-1, 1]$ планы вида

$$\xi^* = \begin{pmatrix} -1 & -x^* & x^* & 1 \\ 1/2 - \omega^* & \omega^* & \omega^* & 1/2 - \omega^* \end{pmatrix}$$

являются L -оптимальными. Найти множество допустимых значений x^* .

9. Привести пример, показывающий, что если L -оптимальный план не принадлежит классу Ξ_L^* , то условие (2) теоремы (2.5) может быть нарушено.

Указание: для тригонометрической модели третьего порядка рассмотреть план, оптимальный для оценивания коэффициентов при $\sin(t)$, $\sin(2t)$. Изучить поведение экстремального полинома.

Литература

1. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. М.: Вильямс, 2007.
2. *Математическая теория планирования эксперимента* / Под ред. С.М. Ермакова. М.: Наука, 1983.
3. *Ермаков С.М., Жиглявский А.А.* Математическая теория оптимального эксперимента. М.: Наука, 1987.
4. *Карлин С., Стадден В.* Чебышёвские системы и их применение в анализе и статистике. М.: Наука, 1976.
5. *Круг, Г.К.* Планирование экспериментов. М.: Наука, 1966.
6. *Мелас В.Б.* E -оптимальные планы эксперимента. СПб.: Изд-во С.-Петербур. ун-та, 1997.
7. *Мелас В.Б.* Локально-оптимальные планы эксперимента. СПб.: Изд-во СПбГТУ, 1999.
8. *Мелас В.Б.* Общая теория функционального подхода к оптимальному планированию эксперимента. СПб.: Издательство С.-Петербур. ун-та, 1999.
9. *Rao С.Р.* Линейные статистические методы и их применение. М.: Наука, 1968.
10. *Сеге Г.* Ортогональные многочлены. М.: Наука, 1962.
11. *Федоров В.В.* Теория оптимального эксперимента. М.: Наука, 1971.

12. *Федоров В.В.* Планирование при нелинейной зависимости поверхности отклика от оцениваемых параметров. Препринт 21. М.: Изд-во Моск. ун-та, 1971.
13. *Шеффе Г.* Дисперсионный анализ. — М.: Наука, 1980
14. *Abt M., Liski E.P., Mandal N.K., Sinha B.K.* Optimal designs in growth curve models. Part I: Correlated model for linear growth: Optimal designs for slope parameter estimation and growth prediction // Journal of Statistical Planning and Inference. 1997. N 64. P. 141–150.
15. *Abt M., Liski E.P., Mandal N.K., Sinha B.K.* Optimal designs in growth curve models. Part II: Correlated model for quadratic growth: Optimum designs for slope parameter estimation and growth prediction // Journal of Statistical Planning and Inference. 1998. N 67. P. 287–296.
16. *Bickel, P.J.; Herzberg, A.M.* Robustness of design against autocorrelation in time. I: Asymptotic theory, optimality for location and linear regression // Ann. Stat. 1979. N 7. P. 77–95.
17. *Bickel P.J., Herzberg A.M., Schilling M.F.* Robustness of design against autocorrelation in time. II: Optimality, theoretical and numerical results for the first-order autoregressive process // Journal of the American Statistical Association. 1981. N 76. P. 870–877.
18. *Box G.E.P., Hunter W.G.* Sequential design of experiments for nonlinear models // Proc. IBM Scient. Comp. Symposium in Statistics. 1965. P. 113–137.
19. *Chernoff H.* Locally optimal design for estimating parameters // Ann. Math. Stat. 1953. Vol. 24. P. 586–602.

20. *Detle H., Haines L.M., Imhof L.* Maximin and Bayesian optimal designs for regression models. Preprint Bochum: Ruhr-Universität Bochum, 2003.
21. *Detle H., Kunert J., Pepelyshev A.* Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica*. 2008. 18(1), P. 135-154.
22. *Detle H., Melas V.B.* *E*-optimal designs in Fourier regression model. *Math. Methods Statist.* 2002. N 11, P. 259-296.
23. *Detle H., Melas V.B.* Optimal designs for estimating individual coefficients in Fourier regression models // *Ann. Stat.* 2003. Vol. 3. N 5. P. 1669–1692.
24. *Detle H., Melas V.B., Pepelyshev A.* Optimal designs for estimating individual coefficients in polynomial regression – a functional approach // *J. of Statis. Plan. and Inference*. 2004. N 118. P. 201–219.
25. *Detle H., Melas V.B., Pepelyshev A.N.* *D*-optimal designs for trigonometric regression models on a partial circle // *Ann. Inst. Stat. Math.* 2002. Vol. 54. N 4. P. 945–959.
26. *Detle H., Melas V.B., Shpilev P.V.* Optimal designs for estimating the coefficients of the lower frequencies in trigonometric regression models // *Ann. Inst. Stat. Math.* 2007. Vol. 59. N 4. P. 655-673.
27. *Detle H., Melas V.B., Shpilev P.V.* Optimal designs for estimating pairs of coefficients in Fourier regression models. // *Statistica Sinica*. 2009. 19. P. 1587-1601
28. *Detle H., Melas V.B., Shpilev P.V.* Optimal designs for trigonometric regression models. 2011. // *Journal of Statistical Planning and Inference*, Vol. 141, No 3, P. 1343-1353

29. *Fedorov, V.V., Hackl P.* Model-oriented design of Experiments. Springer. 1997.
30. *Gaffke N., Krafft O.* Exact D -optimum designs for quadratic regression // J. R. Stat. Soc. 1982. Ser. B. N 44. P. 394–397.
31. *Hoel P.G.* Efficiency Problems in Polynomial Estimation // Annals of Mathematical Statistics 1958. N 29. P. 1134–1145,
32. *Hohmann G., Jung W.* On sequential and nonsequential D -optimal experimental design // Biometr. Z. 17, P. 329–336. 1975.
33. *Imhof L., Krafft O., Schaefer M.* D -optimal exact designs for parameter estimation in a quadratic model // Sankhya. 2000. Ser. B. Vol. 62. N 2. P. 266–275.
34. *Jennrich R.J.* Asymptotic properties of non-linear least squares estimators // Ann. Math. Stat. 1969. N. 40. P. 633–643.
35. *Kiefer J., Wolfowitz J.* The equivalence of two extremum problems // Canad. J. Math. 1960. Vol. 14, P. 363–366.
36. *Kitsos C.P., Titterington D.M., Torsney B.* An optimal design problem in rhythmometry // Biometrics. 1998. N. 44. P. 657–671,
37. *Melas V.B.* Optimal designs for exponential regression // Math. Operations forsh. Statist. 1978. Ser. Statistics. N 9. P. 45–59.
38. *Pukelsheim F.* Optimal design of experiments. N.Y.: Wiley, 1993.

Содержание

Предисловие	3
1 Регрессионный анализ	5
1.1 Описательная регрессия	5
1.2 Математическая модель и предпосылки регрессионно- го анализа	11
1.3 Теорема Гаусса—Маркова	13
1.4 МНК-оценка в вырожденном случае	21
1.5 Оценка максимального правдоподобия	28
1.6 Обобщенный метод наименьших квадратов	30
1.7 Модели с априорной информацией о параметрах	32
1.8 Основы дисперсионного анализа	37
1.9 Задачи для самоконтроля	47
2 Планирование экспериментов	53
2.1 Краткий обзор	53
2.2 Основные понятия	54
2.3 Теорема Каратеодори	59
2.4 Критерии оптимальности	65
2.5 Теоремы эквивалентности	67
2.6 Применение теорем эквивалентности	80
2.6.1 Полиномиальная модель	80
2.6.2 Тригонометрическая модель	84
2.7 Факторные планы	116
2.7.1 МНК-оценка	119
2.7.2 D -оптимальные планы	120
2.8 Задачи для самоконтроля	124
Литература	127

Учебное издание

Вячеслав Борисович Мелас

Петр Валерьевич Шпилев

Планирование и анализ для регрессионных моделей

Учебное пособие

Зав. редакцией *Г.И. Чередниченко*

Редактор *Ф.С. Бастиан*

Обложка *А.В. Калининой*

Техн. редактор *Л.Н. Иванова*

Подписано в печать 2014.

Ф-т 60×84/16. Усл. печ. л. 6,375.

Уч.-изд. л. 5,64. Тираж ... экз. Заказ N

Редакция оперативной подготовки изданий
С.-Петербургского государственного университета.
199034, С.-Петербург, Университетская наб., 7/9.

Типография Издательства СПбГУ.
199061, С.-Петербург, Средний пр., 41.