Введение

Статистический анализ данных (САД) – это научная дисциплина, основанная на методах теории вероятностей и математической статистики, целью которой является формирование научно обоснованных выводов и принятие решений относительно сложной системы (объекта или явления) на основе статистических данных о ней.

Введем следующие обозначения:

S – иследуемая сложная система;

X – имеющиеся статистические данные об S, причем $X \in \mathcal{X}$, где $\mathcal{X}$ – множество всех возможных данных об S;

Y – результаты применения методов САД к X;

$d(S)$ – выводы или решения относительно S.

Тогда САД можно представить в виде схемы на рисунке 1:

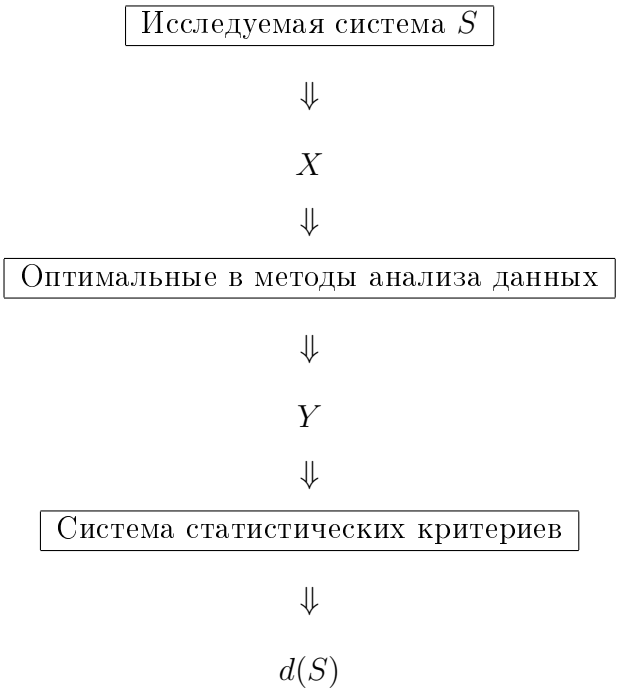

Рисунок 1. Схема статистического анализа данных.

В рамках статистического анализа данных предполагается, что данные о системе S имеют стохастическую природу, а для их описания и анализа используются вероятностно-статистические модели и методы.

В основе САД лежат следующие принципы:

– переход от анализа всех мыслимых значений случайной величины ξ, описывающей функционирование и свойства S (введенное выше множество $\mathcal{X}$), к анализу случайной выборки X значений из распределения вероятностей ξ;

– замена случайной выборки X некоторой совокупностью Y вычисленных на ее основе выборочных характеристик (функций от выборки), называемых *статистиками*;

– принятие решений (или статистических выводов) на основе анализа статистик с помощью статистических критериев, справедливых с некоторой заданной вероятностью ошибки.

Статистический анализ данных можно условно разделить на следующие этапы:

Этап 1. Исследование сложной системы и сбор априорной информации.

На этом этапе выявляются основные факторы, определяющие функционирование системы, а также изучаются ее свойства.

Э т а п 2. Формирование предположений относительно системы и построение ее математической модели.

С учетом результатов, полученных на первом этапе, строится математическая модель системы S и проверяется адекватность этой модели. При этом применяются методы теории вероятностей и математической статистики.

Э т а п 3. Планирование и проведение экспериментов с системой и сбор статистических данных.

Разрабатывается план экспериментов, в соответствии с которым проводятся эксперименты над математической моделью, полученной на втором этапе. Целью этого этапа является получение статистических данных.

Э т а п 4. Первичный (разведочный) анализ и сжатие данных.

Проводится предварительных анализ и, если это необходимо и возможно, сжатие данных.

Э т а п 5. Подбор методов и непосредственное проведение статистического анализа.

Э т а п 6. Анализ полученных результатов и проверка их адекватности.

Э т а п 7. Принятие решения об удовлетворительности полученных результатов или необходимости возврата на предыдущие этапы.

Глубина и достоверность статистических выводов и решений зависит от математической модели системы S и методов, применяемых при исследовании этой системы, а также от того, насколько корректно эти методы и модель применяются.

В различных приложениях статистического анализа данных существуют три основные типа данных:

– *пространственные данные* – это совокупность значений $x_1, x_2, \ldots, x_n \in \mathbb{R}^N$ анализируемых переменных, полученных для некоторого множества из n $(n > 1)$ объектов исследования в фиксированный момент (период) времени.

Пространственные данные характеризуют зависимости между объектами исследования, описываемые выбранными переменными.

– *временной ряд* – это последовательность значений $x_1, x_2, \ldots, x_T \in \mathbb{R}^N$ анализируемых переменных, которые соответствуют T $(T > 1)$ последовательным моментам (периодам) времени.

Временные ряды характеризуют динамику изменения анализируемых переменных во времени.

– *панельные данные* – это совокупность значений $x_{1t}, x_{2t}, \ldots, x_{nt} \in \mathbb{R}^N$ анализируемой переменной, полученных для некоторого множества из n $(n > 1)$ объектов исследования в последовательные моменты (периоды) времени $t = 1, 2, \ldots, T$.

Панельные данные описывают динамику изменения показателей, характеризующих состояние некоторой группы исследуемых объектов с учетом взаимосвязей между ними.

По характеру компонент (признаков), из которых образованы векторы-наблюдения $x_i = (x_{i1}, \ldots, x_{iN})^{\mathbf{T}}$, $i = 1, \ldots, n$, выделяют следующие модели данных[11]:

– *непрерывные* – это данные, когда все N признаков имеют совместое абсолютно-непрерывное распределение вероятностей с некоторой плотностью $p(z)$, $z \in R^N$, которая однозначно определяет их совместную функцию распределения $F(z)$, $z \in R^N$;

– *дискретные* – данные, когда наблюдения являются дискретными случайными векторами;

– *комбинированные* – данные, когда часть признаков являются непрерывными случайными величинами, а остальные – дискретными.

Статистический анализ данных обычно проводится с использованием компьютеров и различных пакетов прикладных программ [**?**], в основу которых положены теоретически обоснованные статистические методы.

1 Предварительный (первичный) статистический анализ данных

Предварительный анализ данных проводится с целью выявления особенностей и свойств исследуемой системы, определения структуры данных и формулировки предположений относительно их модели.

1.1 Математические модели пространственных данных

Случайная выборка из распределения вероятностей

Основные методы САД ориентированы на пространственные данные, адекватной математической моделью которых является случайная выборка.

Пусть состояние исследуемой системы характеризуется N-мерным вектором показателей $\xi \in \mathbb{R}^N,\ N \geq 1$. Предположим, что $x_1, \ldots, x_n$ — совокупность (выборка) наблюдаемых значений $\xi \in \mathbb{R}^N$, полученных в экспериментах по наблюдению за n объектами.

Математической моделью i-го наблюдения является случайная величина $x_i = (x_{i1}, \ldots, x_{iN})^{\mathbf{T}} \in \mathbb{R}^N$ (скалярная при $N = 1$ или векторная при $N > 1$) с некоторой функцией распределения

$$F_i(z_i) = F_i(z_{i1}, \ldots, z_{iN}) = \mathbf{P}\{x_{i1} < z_{i1}, \ldots, x_{iN} < z_{iN}\} \equiv \mathbf{P}\left\{x_i^{\mathbf{T}} < z_i^{\mathbf{T}}\right\}, \tag{1.1}$$

где $z_i = (z_{i1}, \ldots, z_{iN})^{\mathbf{T}} \in \mathbb{R}^N,\ i = 1, \ldots, n,\ N \geq 1$. В (1.1) и далее, если не оговорено особо, при $N > 1$ используется формальное обозначение:

$$\mathbf{P}\left\{x_i^{\mathbf{T}} < z_i^{\mathbf{T}}\right\} ::= \mathbf{P}\{x_{i1} < z_{i1}, \ldots, x_{iN} < z_{iN}\},\ i = 1, \ldots, n. \tag{1.2}$$

Математической моделью данных, представляющих собой выборку $x_1, \ldots, x_n$, является случайный вектор $X \in \mathbb{R}^{Nn}$ с Nn-мерной функцией распределения

$$\begin{aligned} F_X(z_1, \ldots, z_n) :: &= F_X(z_{11}, \ldots, z_{1N}; z_{21}, \ldots, z_{2N}; \ldots; z_{n1}, \ldots, z_{nN}) = \\ &= \mathbf{P}\left\{x_1^{\mathbf{T}} < z_1^{\mathbf{T}}, \ldots, x_n^{\mathbf{T}} < z_n^{\mathbf{T}}\right\}. \end{aligned} \tag{1.3}$$

Определение 1.1. *Данные* $X = \left(x_1^{\mathbf{T}}, \ldots, x_n^{\mathbf{T}}\right)^{\mathbf{T}} \in \mathbb{R}^{Nn}$ *(эквивалентная форма записи* $X = \{x_1, \ldots, x_n\}$*) принято называть* случайной выборкой *объема* n *из* N*-мерного распределения вероятностей* $\mathcal{L}\{\xi\}$ *с функцией распределения* $F(z)$ *(или просто – случайной выборкой из распределения* $\mathcal{L}\{\xi\}$*), если* $x_1, \ldots, x_n$ *являются независимыми одинаково распределенными случайными величинами (н. о. р. с. в.), то есть если для них выполняются предположения:*

P_1. *Случайные величины* $x_1, \ldots, x_n$ *одинаково распределены:*

$$F_{x_1}(z_1) = \ldots = F_{x_n}(z_n) \equiv F(z),\ z \in \mathbb{R}^N;$$

P_2. *Случайные величины* $x_1, \ldots, x_n$ *независимы в совокупности, при этом:*

$$F_X(z_1, \ldots, z_n) = \prod_{i=1}^{n} F_{x_i}(z_i) = (F(z))^n,\ z \in \mathbb{R}^N.$$

Математические модели данных, альтернативные случайной выборке

Во многих практических задачах предположения P_1 и P_2 о данных, сфомулированные в определении 1.1, могут нарушаться. Это приводит к моделям, альтернативным случайным выборкам. В связи с тем, что основная часть методов САД ориентирована на модели данных типа случайная выборка, на этапе предварителього анализа данных необходимо выявить факторы, приводящие к нарушениям предположений P_1 и P_2. В тех случаях, когда эти факторы могут быть вывявлены и устранены, такие модели могут быть представлены в виде случайной выборки из некоторого распределения вероятностей. Для анализа полученных данных могут применяться классические методы САД.

При анализе пространственных данных встречаются следующие альтернативные случайной выборке модели данных:

– преднамеренная (неслучайная) выборка;
– неоднородная выборка из смеси распределений;
– случайная выборка с засорениями.

Преднамеренная выборка – выборка, полученная в результате пристрастного отбора данных.

Такие выборки, очевидно, нельзя считать случайными, поскольку при их формировании нарушаются принципы случайности и независимости экспериментов, необходимых для получения статистических данных. Кроме того, преднамеренные выборки могут быть следствием и непреднамеренных действий. Например, при недостаточной точности измерительного инструмента.

Неоднородная выборка – случайная выборка из смеси нескольких распределений вероятностей.

Неоднородная выборка может быть следствием включения в одну группу объектов, которые существенно отличаются своими функциональными характеристиками. Например: выборка значений показателей единиц продукции, произведенной в различных условиях; выборка значений показателей физического состояния пациентов до и после лечения и т. д.

Такая выборка может быть описана следующий моделью. Пусть в пространстве $\mathbb{R}^N$, $N \geq 1$, с априорными вероятностями $\pi_0, \ldots, \pi_{L-1}$,

$$\pi_0 + \ldots + \pi_{L-1} = 1,$$

наблюдаются объекты из L классов, $L \geq 2$. Случайный вектор наблюдения $x \in \mathbb{R}^N$ для объектов из класса l имеет условную плотность распределения $p_l(z)$, $l = 0, \ldots, L-1$. Тогда выборочные наблюдения $x_1, \ldots, x_n$ можно рассматривать как случайную выборку из некоторого распределения, которое является смесью распределений и имеет плотность распределения вероятностей вида:

$$p(z_i) = \sum_{l=0}^{L-1} \pi_l p_l(z_i),\ i = 1, \ldots, n. \tag{1.4}$$

Случайная выборка с засорениями – частный случай неоднородной выборки.

Наиболее известная модель данного класса – модель засорений Тьюки. Она может быть описана формулой (1.4) при $L = 2$, $\pi_0 = 1 - \varepsilon$, $\pi_1 = \varepsilon$, где ε – близкий к нулю уровень засорения (доля аномальных наблюдений) $0 < \varepsilon < 0,5$. Выборку с моделью засорений Тьюки можно рассматривать как случайную выборку из распределения вероятностей с плотностью

$$p(z_i) = (1 - \varepsilon)p_0(z_i) + \varepsilon p_1(z_i),$$

где $p_0(z_i)$ – плотность распределения гипотетического распределения основной массы наблюдений, а $p_1(z_i)$ – плотность «засоряющего» распределения, $i = 1, \ldots, n$.

1.2 Математические модели статистических наблюдений в динамике

Во многих задачах САД существенным является порядок поступления наблюдений за исследуемой системой. Математическими моделями таких данных являются случайные функции (случайные процессы, временные ряды).

Определение 1.2. *Случайной функцией (случайным процессом) называется параметрическое семейство случайных векторов:*

$$\xi = \xi(t) = \xi(\omega, t) \in \mathbb{R}^N,$$

определенных на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$, $\omega \in \Omega$, *где* t *– параметр, изменяющийся на некотором множестве* $\mathcal{T}$*:* $t \in \mathcal{T} \subseteq \mathbb{R}^m$.

В зависимости от значений N и m размерностей пространства наблюдений $\mathbb{R}^N$ и пространства значений параметра $\mathbb{R}^m$ принята классификация случайных функций, приведенная в таблице 1 [12].

Таблица 1 – Классификация случайных функций

Размерность пространства наблюдений, N	Размерность пространства значений параметра, m	
	$m = 1$	$m > 1$
$N = 1$	*Случайный процесс* (Цена акций [3]; показатель солнечной активности [2])	*Случайное поле* (Экран монитора в черно-белом режиме, $m = 2$)
$N > 1$	*Векторный случайный процесс* (Несколько отведений электрокардиограммы, обычно $N = 8$)	*Векторное случайное поле* (Радиационная карта местности с несколькими радиоизотопами, $m = 2$)

Если множество $\mathcal{T}$ значений параметра t дискретно (конечно или счетно), то такой случайный процесс обычно называют *временным рядом* или *случайной последовательностью*, а одномерный параметр понимают как *время*. Векторный случайный процесс при этом называют *векторным временным рядом*.

2 Функциональные и числовые характеристики вероятностных моделей данных

2.1 Функциональные характеристики вероятностных моделей данных

На этапе разведочного анализа, как правило, исследуются одномерные (маргинальные) распределения вероятностей каждого признака по отдельности. Поэтому, вначале предположим, что наблюдается некоторый одномерный признак (случайная величина) ξ.

Вероятностный закон распределения $\mathcal{L}\{\xi\}$ одномерной случайной величины $\xi \in \mathbb{R}$ описывается следущими функциональными характеристиками [12].

Функция распределения вероятностей:

$$F_\xi(z) = \mathbf{P}\left\{\xi < z\right\},\ z \in \mathbb{R}. \tag{2.1}$$

Если ξ – дискретная случайная величина, принимающая значения из множества $C = \{c_1, \ldots, c_K\}$ с вероятностями

$$p_k = \mathbf{P}\left\{x = c_k\right\},\ k = 1, \ldots, K,\ K \leq +\infty,$$

то

$$F_\xi(z) = \sum_{k:c_k<z} \mathbf{P}\{\xi = c_k\} = \sum_{k:c_i<z} p_k,\ z \in \mathbb{R}. \tag{2.2}$$

Если ξ – непрерывная случайная величина, то

$$F_\xi(z) = \int_{-\infty}^{z} p_\xi(y)\ dy, z \in \mathbb{R}, \tag{2.3}$$

где $p_\xi(y)$ – плотность распределения вероятности $\xi,\ y \in \mathbb{R}$.

Распределение вероятностей дискретной случайной величины:

$$p_k = \mathbf{P}\left\{\xi = c_k\right\},\ k = 1, \ldots, K,\ K \leq +\infty, \tag{2.4}$$

где $C = \{c_1, \ldots, c_K\}$ – конечное $(K < +\infty)$ или счетное $(K = +\infty)$ множество вожможных значений ξ и

$$\sum_{k=1}^{K} p_k = 1.$$

Плотность распределения вероятностей непрерывной случайной величины:

$$p_\xi(z) = F'_\xi(z), z \in \mathbb{R}, \tag{2.5}$$

где $F_\xi(z)$ – функция распределения $\xi,\ z \in \mathbb{R}$.

Характеристическая функция:

$$f_\xi(t) = \mathbf{E}\left\{e^{it\xi}\right\} = \int_{-\infty}^{+\infty} e^{itz} dF_\xi(z), t \in \mathbb{R}, f_\xi(t) \in \mathbb{C}. \tag{2.6}$$

Если ξ – дискретная случайная величина, принимающая значения из множества $C = \{c_1, \ldots, c_K\}$ с вероятностями

$$p_k = \mathbf{P}\left\{\xi = c_k\right\},\ k = 1, \ldots, K,\ K \leq +\infty,$$

то

$$f_\xi(t) = \sum_{k=1}^{K} e^{itc_k} \cdot p_k,\ t \in \mathbb{R}. \tag{2.7}$$

Если ξ – непрерывная случайная величина, то

$$f_\xi(t) = \int_{-\infty}^{+\infty} e^{ity} p_x(y)\ dy, t \in \mathbb{R}, \tag{2.8}$$

где $p_\xi(y)$ – плотность распределения вероятностей $\xi,\ y \in \mathbb{R}$.

Если в эксперименте наблюдается N-мерный случайный вектор

$$\xi = (\xi_1, \xi_2, \ldots, \xi_N)^{\mathbf{T}} \in \mathbb{R}^N,\ N \geq 2,$$

то закон распределения вероятностей $\mathcal{L}(\xi)$ описывается следующими функциональными характеристиками.

N-мерная функция распределения:

$$F_\xi(z) = F_{\xi_1,\ldots,\xi_N}(z_1, \ldots, z_N) = \mathbf{P}\left\{\xi_1 < z_1, \ldots, \xi_n < z_N\right\}, \tag{2.9}$$

где $\xi,\ z \in \mathbb{R}^N$.

Распределение вероятностей дискретного случайного N-вектора:

$$\begin{gathered} p_{k_1,\ldots,k_N} = \mathbf{P}\left\{\xi = c_{k_1,\ldots,k_N}\right\} = \mathbf{P}\left\{\xi_1 = c_1^{(k_1)}, \ldots, \xi_N = c_N^{(k_N)}\right\}, \\ k_j = 1, 2, \ldots, K_j, j = 1, 2, \ldots, N, \end{gathered} \tag{2.10}$$

где конечное или счетное множество возможных значений случайного N-вектора ξ определено соотношениями:

$$\begin{gathered} C = \left\{c_{k_1,\ldots,k_N} :\ k_j = 1, 2, \ldots, K_j, j = 1, 2, \ldots, N\right\}, \\ c_{k_1,\ldots,k_N} = \left(c_1^{(k_1)}, \ldots, c_N^{(k_N)}\right)^{\mathbf{T}} \in \mathbb{R}^N,\ k_j = 1, 2, \ldots, K_j, j = 1, 2, \ldots, N, \end{gathered}$$

$K_1 \leq \infty, \ldots, K_N \leq \infty$, а

$$\sum_{k_1=1}^{K_1} \cdots \sum_{k_N=1}^{K_N} p_{k_1,\ldots,k_N} = 1.$$

N-мерная плотность распределения вероятностей:

$$p_\xi(z) = p_\xi(z_1, \ldots, z_N) = \frac{\partial^N F_\xi(z_1, \ldots, z_N)}{\partial z_1 \ldots \partial z_N},$$

где $F_\xi(z)$ – N-мерная функция распределения случайного N-вектора $\xi,\ z \in \mathbb{R}^N$.

Характеристическая функция N-мерного распределения вероятностей:

$$\begin{gathered} f_\xi(t) = \mathbf{E}\left\{e^{it^{\mathbf{T}}\xi}\right\} = \mathbf{E}\left\{e^{i<t,\xi>}\right\} = \mathbf{E}\exp\left\{i\sum_{j=1}^{N} t_j\xi_j\right\} = \\ = \int_{\mathbb{R}^N} \exp\left\{i\sum_{j=1}^{N} t_j z_j\right\} dF_\xi(z),\ t, z \in \mathbb{R}^N,\ f_\xi(t) \in \mathbb{C}, \end{gathered} \tag{2.11}$$

где $< t, \xi >$ — скалярное произведение векторов t и ξ.

2.2 Числовые характеристики вероятностных моделей данных

Числовые характеристики распределения вероятностей $\mathcal{L}\{\xi\}$ одномерной случайной величины $\xi \in \mathbb{R}$ позволяют составить наглядное представление об этом распределении. При разведочном анализе данных обычно вычисляют следующие числовые характеристики [12].

Характеристики положения

Математическое ожидание:

$$\mu = \mathbf{E}\{\xi\}.$$

Если ξ – дискретная случайная величина, принимающая значения из множества $C = \{c_1, \ldots, c_K\}$ с вероятностями

$$p_k = \mathbf{P}\{\xi = c_k\},\ k = 1, \ldots, K,\ K \leq +\infty,$$

то

$$\mu = \sum_{k=1}^{K} c_k \cdot p_k. \tag{2.12}$$

Если ξ – непрерывная случайная величина, то

$$\mu = \int_{-\infty}^{+\infty} y p_\xi(y)\ dy, \tag{2.13}$$

где $p_\xi(y)$ – плотность распределения вероятности ξ $(y \in \mathbb{R})$.

Наибольшее и наименьшее значения случайной величины ξ:

$$c_+ = \max\{\xi\},\ c_- = \min\{\xi\}. \tag{2.14}$$

Если ξ – дискретная случайная величина, принимающая значения из множества $C = \{c_1, \ldots, c_K\}$

$$c_+ = \max\{c_1, \ldots, c_K\},\ c_- = \min\{c_1, \ldots, c_K\}.$$

Если ξ – непрерывная случайная величина, то

$$c_+ = \max\{z\ :\ p_\xi(z) \neq 0\},\ c_- = \min\{z\ :\ p_\xi(z) \neq 0\},$$

где $p_\xi(z)$ – плотность распределения вероятности $\xi,\ y \in \mathbb{R}$.

Мода распределения вероятноостей $\mathcal{L}\{\xi\}$:

Если ξ – дискретная случайная величина, принимающая значения из множества $C = \{c_1, \ldots, c_K\}$ с вероятностями

$$p_k = \mathbf{P}\{x = c_k\},\ k = 1, \ldots, K,\ K \leq +\infty,$$

то

$$M = mod\,\xi = c_{k^*},\ \text{ где } k^* = arg\ \max\{p_1, \ldots, p_K\}. \tag{2.15}$$

Если ξ – непрерывная случайная величина с плотностью распределения вероятности $p_\xi(y),\ y \in \mathbb{R}$, то

$$M = mod\,\xi = arg \max_{y \in \mathbb{R}} p_\xi(y). \tag{2.16}$$

Медиана распределения вероятностей $\mathcal{L}\{\xi\}$:

$$m = med\,\xi = z_{0,5} = F_\xi^{-1}(0,5), \tag{2.17}$$

где $F_\xi(z)$ – функция распределения случайной величины ξ.

Характеристики рассеяния

Дисперсия:

$$\sigma^2 = \mathbf{D}\{\xi\} = \mathbf{E}\{(\xi - \mu)^2\}. \tag{2.18}$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\mathbf{D}\{\xi\}}. \tag{2.19}$$

Размах значений случайной величины ξ:

$$R = c_+ - c_-, \tag{2.20}$$

где c_+ и c_- определены в (2.14).

Интерквартильный размах:

$$IR = z_{\frac{3}{4}} - z_{\frac{1}{4}}, \tag{2.21}$$

где $z_{\frac{k}{4}}$ – k-я квартиль распределения вероятностей $\mathcal{L}\{\xi\}$, то есть

$$z_{\frac{k}{4}} = F_\xi^{-1}\left(\frac{k}{4}\right),\ k \in \{1, 3\}. \tag{2.22}$$

Характеристики формы

Асимметрия распределения вероятноостей $\mathcal{L}\{\xi\}$:

$$\beta_1 = \frac{\alpha_3}{\sigma^3}, \tag{2.23}$$

где α_3 – начальный момент 3-го порядка случайной величины ξ

$$\alpha_3 = \mathbf{E}\left\{\xi^3\right\},$$

а $\sigma = \sqrt{\mathbf{D}\{\xi\}}$.

Эксцесс распределения вероятностей $\mathcal{L}\{\xi\}$:

$$\beta_2 = \frac{\mu_4}{\sigma^4} - 3, \tag{2.24}$$

где μ_4 – центральный момент 4-го порядка случайной величины ξ

$$\mu_4 = \mathbf{E}\left\{(\xi - \mu)^4\right\},$$

а $\sigma = \sqrt{\mathbf{D}\{\xi\}}$.

Если наблюдаемая в эксперименте случайная величина ξ представляет собой случайный вектор размерности N, то есть

$$\xi = (\xi_1, \ldots, \xi_N)^{\mathbf{T}} \in \mathbb{R}^N,$$

то числовыми характеристиками этого случайного вектора, которые описывают N-мерный закон распределения $\mathcal{L}\{\xi\}$ являются

Вектор математического ожидания:

$$m = (m_1, \ldots, m_N)^{\mathbf{T}} \in \mathbb{R}^N, \tag{2.25}$$

где

$$m_j = \mathbf{E}\{\xi_j\},\ j = 1, \ldots, N. \tag{2.26}$$

Центральный смешанный момент порядка r:

$$\mu_{r_1,\ldots,r_N} = \mathbf{E}\{(\xi_1 - m_1)^{r_1} \cdot \ldots \cdot (\xi_N - m_N)^{r_N}\}, \tag{2.27}$$

где $r_1, \ldots, r_N$ — целые неотрицательные целые числа, задающие разбиение числа $r \in \mathbb{N}$, то есть

$$r = r_1 + \ldots + r_N.$$

Ковариационная матрица:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \ldots & \sigma_{NN} \end{pmatrix} \tag{2.28}$$

где σ_{jk} – ковариация случайных величин ξ_j и ξ_k (центральный смешанный момент второго порядка):

$$\sigma_{jk} = \mathbf{Cov}\{\xi_j, \xi_k\} = \mathbf{E}\left\{(\xi_j - m_j) \cdot (\xi_k - m_k)\right\},\ j,k = 1, \ldots, N. \tag{2.29}$$

Очевидно, что σ_{jj} – дисперсия случайной величины $\xi_j,\ j = 1, \ldots, N$:

$$\sigma_{jj} = \sigma_j^2.$$

Корреляционная матрица:

$$\mathcal{P} = \begin{pmatrix} \rho_{11} & \rho_{12} & \ldots & \rho_{1N} \\ \rho_{21} & \rho_{22} & \ldots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \ldots & \rho_{NN} \end{pmatrix} \tag{2.30}$$

где ρ_{jk} – коэффициент корреляции случайных величин ξ_j и ξ_k:

$$\rho_{jk} = \frac{\sigma_{jk}}{\sigma_{jj} \cdot \sigma_{kk}},\ j,k = 1, \ldots, N. \tag{2.31}$$

Очевидно, что $\rho_{jj} = 1,\ j = 1, \ldots, N.$

3 Статистическое оценивание функциональных и числовых характеристик вероятностных моделей данных

3.1 Основные понятия теории статистического оценивания параметров

Пусть в результате проведения экспериментов наблюдается случайная выборка $X = \{x_1, \ldots, x_n\}$ объема n из некоторого N-мерного распределения вероятностей $\mathcal{L}\{\xi\}$, $\xi \in \mathbb{R}^N$, информация о котором отсутствует полностью или частично. Основной задачей предварительного анализа является восстановление этого распределения по выборке случайных данных. Известны два основных подхода к решению данной задачи: параметрический и непараметрический.

Параметрический подход применяется тогда, когда распределение вероятностей $\mathcal{L}\{\xi\}$, $\xi \in \mathbb{R}^N$, известно с точностью до параметров, то есть известно с точностью до параметров аналитическое представление функции распределения

$$F_\xi(z) = F_0(z, \theta),$$

или других функциональных характеристик распределения вероятностей $\mathcal{L}\{\xi\}$, где $F_0(\cdot)$ – истинная (модельная) функция распределения, а $\theta = (\theta_1, \ldots, \theta_m)^{\mathbf{T}} \in \mathbb{R}^m$ – вектор неизвестных параметров, $m \geq 1$.

В этом случае по выборке $X = \{x_1, \ldots, x_n\}$ методами математической статистики (методом максимального правдоподобия, методом моментов, методом наименьших квадратов и т. д. [12]) вычисляются оценки $\hat{\theta}_1, \ldots, \hat{\theta}_m$ неизвестных параметров $\theta_1, \ldots, \theta_m$. Далее, в соответствии с так называемым «подстановочным принципом», полученные оценки $\hat{\theta}_1, \ldots, \hat{\theta}_m$ подставляются в соответствующее аналитическое выражение для функции распределения $F_0(z, \theta)$ (или другой известной функциональной характеристики), что приводит к ее статистической оценке $\hat{F}(z)$:

$$\hat{F}(z) \equiv \hat{F}_\xi(z) = F_0(z, \hat{\theta}) = F_0(z; \hat{\theta}_1, \ldots, \hat{\theta}_m) \; z \in \mathbb{R}^N.$$

Замечание 3.0.1. *Теоретической основой параметрического метода, а также и подстановочного принципа, являются статистические свойства оценок и предельные теоремы теории вероятностей [12].*

Непараметрический подход применяется тогда, когда распределение вероятностей $\mathcal{L}\{\xi\}$, $\xi \in \mathbb{R}^N$, неизвестно, то есть неизвестны его функциональные характеристики. В этом случае по выборке $X = \{x_1, \ldots, x_n\}$ вычисляют соответствующие выборочные функциональные характеристики.

При вычислении статистических оценок неизвестных функциональных и числовых характеристик распределения вероятностей $\mathcal{L}\{\xi\}$ по случайной выборке $X = \{x_1, \ldots, x_n\}$ объема n всегда возникает вопрос о требованиях, которые следует предъявить к этим статистикам, чтобы они были в каком-то определенном смысле близкими к истинному значению. Эти требования формулируются обычно с помощью следующих статистических свойств оценок: состоятельности, несмещенности и эффективности [12].

Пусть $\theta_n = \hat{\theta}(X)$ – оценка неизвестного параметра θ, вычисленная по случайной выборке $X = \{x_1, \ldots, x_n\}$ объема n из некоторого распределения вероятностей $\mathcal{L}\{\xi\}$. Тогда θ_n называется:

– *состоятельной* оценкой параметра θ, если

$$\theta_n \xrightarrow[n\to\infty]{\mathbf{P}} \theta;$$

– *строго состоятельной* оценкой параметра θ, если

$$\theta_n \xrightarrow[n\to\infty]{\mathbf{P}=1} \theta;$$

– *несмещенной* оценкой параметра θ, если для любого $n \geq 1$

$$\mathbf{E}\{\theta_n\} = \theta;$$

– *асимптотически несмещенной* оценкой параметра θ, если

$$\lim_{n\to\infty} \mathbf{E}\{\theta_n - \theta\} = 0;$$

– *эффективной* оценкой параметра θ, если θ_n – несмещенная оценка и для любых $\theta \in \mathbb{R}^m$ и $n \geq 1$ эффективность $\mathcal{E}_n$ равна единице:

$$\mathcal{E}_n = \frac{1}{|V||\mathcal{I}_n|} = 1,$$

где V – матрица вариаций оценки θ_n [12], а $\mathcal{I}_n$ – информационная матрица Фишера [12] для всей выборки X.

3.2 Выборочные функциональные характеристики и их статистические свойства

Пусть $X = \{x_1, \ldots, x_n\}$ – случайная выборка объема n из некоторого одномерного распределения вероятностей $\mathcal{L}\{\xi\}$. Рассмотрим задачу построения по выборке X статистических оценок функциональных характеристик этого распределения: функции распределения $F(\cdot)$, распределения вероятностей $\{p_k,\, k = 1, \ldots, K\}$, плотности распределения вероятностей $p(\cdot)$, характеристической функции $f(\cdot)$.

Вариационный ряд выборки и порядковые статистики

Определение 3.1. Вариационным рядом *выборки* $X = \{x_1, \ldots, x_n\}$ *объема* n, *называется последовательность* $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, *упорядоченная по неубыванию, то есть*

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}, \tag{3.1}$$

где $x_{(k)}$ – k*-я порядковая статистика [12],* $k \in \{1, \ldots, n\}$.

Вариационный ряд является одним из стандартных способов представления случайной выборки.

Выборочная функция распределения

Определение 3.2. Выборочной функцией распределения *или* эмпирической функцией распределения, *построенной по выборке* X *объема* n, *называется статистика*

$$\hat{F}(z) ::= F_n(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(z - x_i) = \frac{L_n(z)}{n},\ z \in \mathbb{R}, \tag{3.2}$$

где $L_n(z)$ – *число выборочных значений из множества* $\{x_1, \ldots, x_n\}$, *для которых* $x_i < z$, $i = 1, \ldots, n$:

$$L_n(z) = \sum_{i=1}^{n} \mathbf{I}(z - x_i),$$

а $\mathbf{I}(z)$ – *единичная функция Хевисайда:*

$$\mathbf{I}(z) = \begin{cases} 1 & \text{при } z > 0, \\ 0 & \text{при } z \leq 0. \end{cases} \tag{3.3}$$

Для построения графика выборочной функции распределения используют вариационный ряд (3.1) выборки X. Схематический график выборочной функции распределения представлен на рисунке 2 [12].

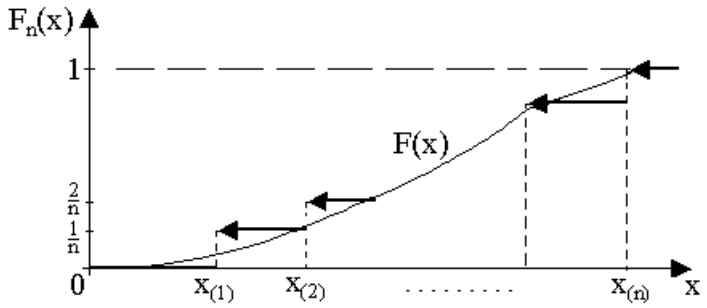

Рисунок 2. Выборочная функция распределения

Относительные частоты

Пусть $X = \{x_1, \ldots, x_n\}$ – случайная выборка объема n из некоторого одномерного дискретного распределения вероятностей $\mathcal{L}\{\xi\}$, а наблюдаемая в эксперименте случайная величина ξ принимает значения из множества $C = \{c_1, \ldots, c_K\}$, $K \leq \infty$. Тогда, согласно закону больших чисел [12], в качестве оценок $\{\hat{p}_k,\ k = 1, \ldots, K\}$ для неизвестных вероятностей

$$p_k = \mathbf{P}\{x = c_k\},\ k = 1, \ldots, K,$$

рассматривают относительные частоты

$$\hat{p}_k = \nu_k = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i, c_k},\ k = 1, \ldots, K, \tag{3.4}$$

где

$$\delta_{a,b} = \begin{cases} 1, & \text{если } a = b, \\ 0, & \text{если } a \neq b. \end{cases} \tag{3.5}$$

Гистограмма

Предположим, что плотность распределения вероятностей $p(z)$, $z \in \mathbb{R}$, наблюдаемой непрерывной случайной величины ξ сосредоточена на отрезке $\Gamma = [x_-, x_+]$, а вне этого отрезка она равна нулю. Зададим натуральное число M и осуществим разбиение Γ на M частей точками деления:

$$x_- = b_0 < b_1 < \ldots < b_M = x_+.$$

Обозначим m-ю ячейку этого разбиения через Γ_m:

$$\Gamma_m ::= [b_{m-1}, b_m),\ m = 1, \ldots, M,\ M \geq 2.$$

В дальнейшем Γ_m будем называть m-й ячейкой гистограммы, $m = 1, \ldots, M$.

Замечание 3.2.1. *При оценивании плотности распределения $p(z)$ по выборке $X = \{x_1, \ldots, x_n\}$ объема n как правило полагают:*

$$x_- = \min\{x_1, \ldots, x_n\},\ x_+ = \max\{x_1, \ldots, x_n\},$$

а

$$M ::= M_n = [\log_2 n] + 1.$$

Введем следующие обозначения:

– Δ_m– «размер» (мера Лебега) m-й ячейки гистограммы:

$$\Delta_m ::= \mu(\Gamma_m) = b_m - b_{m-1},\ m = 1, \ldots, M;$$

– ν_m– число выборочных значений, попавших в m-ю ячейку:

$$\nu_m = \sum_{i=1}^{n} \mathbf{I}_{\Gamma_m}(x_i),\ m = 1, \ldots, M,$$

где $\mathbf{I}_A(z)$ – индикатор множества A:

$$\mathbf{I}_A(z) = \begin{cases} 1, & \text{если } z \in A, \\ 0, & \text{если } z \notin A. \end{cases} \tag{3.6}$$

Определение 3.3. Гистограммой *или* гистограммной оценкой *плотности распределения* $p(z)$ *называется статистика*

$$\hat{p}(z) = \sum_{m=1}^{M} \frac{\nu_m}{n\Delta_m} \mathbf{I}_{\Gamma_m}(z),\ z \in \mathbb{R}.$$

График гистограммы с наложенным графиком истинной плотности распределения приведен на рисунке 3.

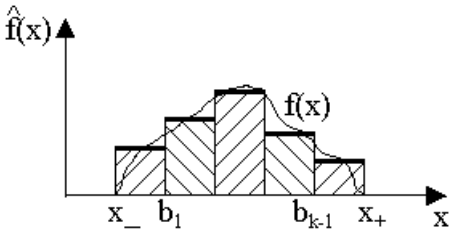

Рисунок 3. Гистограмма

Заметим, что гистограмма, как и плотность распределения, удовлетворяет условию нормировки:

$$\int_{-\infty}^{\infty} \hat{p}(z)dz = \int_{-\infty}^{\infty} p(z)dz = 1.$$

Отметим также, что гистограмма – смещенная и несостоятельная оценка плотности распределения [12]. Для ее состоятельности необходимо, чтобы при $n \to +\infty$ ячейки «измельчались» специальным образом [12]:

$$M = M_n \xrightarrow[n\to\infty]{} +\infty,\ \max_{1\le m\le M_n} \Delta_m \xrightarrow[n\to\infty]{} 0.$$

Выборочная характеристическая функция

Определение 3.4. Выборочной характеристической функцией, *построенной по выборке* X *объема* n, *называется комплекснозначная функция действительной переменной* t:

$$\hat{f}(t) ::= f_n(t) = \frac{1}{n} \sum_{k=1}^{n} e^{itx_k},\ t \in \mathbb{R}. \tag{3.7}$$

3.3 Выборочные числовые характеристики и их статистические свойства

Пусть $X = \{x_1, \ldots, x_n\}$ – выборка объема n из некоторого одномерного распределения вероятностей $\mathcal{L}\{\xi\}$, $F_\xi(z)$ и $F_n(z)$ – соответственно теоретическая и эмпирическая функции распределения. $F_n(z)$ можно рассматривать как функцию распределения некоторой дискретной случайной величины x, принимающей n значений $\{x_1, \ldots, x_n\}$ с вероятностями, равными $\dfrac{1}{n}$, причем, если какое-либо из этих значений встретится k раз в выборке X, то этому значению соответствует вероятность $\dfrac{k}{n}$ [7]. Так же как для теоретического распределения $\mathcal{L}\{\xi\}$, так и для эмпирического распределения, обозначим его $\mathcal{L}\{x\}$, связанного с выборкой X, вводятся аналогичные числовые характеристики, которые принято называть выборочными характеристиками.

Эмпирическая функция распределения $F_n(z)$ является сильно состоятельной несмещенной оценкой для теоретической функции распределения $F_\xi(z)$ [12]. Следовательно, $F_n(z)$ – статистический аналог $F_\xi(z)$ и выборочная характеристика – статистический аналог соответствующей теоретической характеристики [7]. В общем случае, если $g = \mathbf{E}\{g(\xi)\}$ – некоторая теоретическая характеристика наблюдаемой случайной величины ξ, то ее статистический аналог, то есть соответствующая выборочная характеристика, вычисляется по формуле [7]:

$$\hat{g} ::= g_n = \frac{1}{n}\sum_{i=1}^{n} g(x_i). \tag{3.8}$$

Таким образом, *выборочные начальные* и *выборочные центральные моменты* вычисляются, соответственно, по формулам:

$$\hat{\alpha}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r, \tag{3.9}$$

$$\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\alpha}_1)^r. \tag{3.10}$$

Причем выборочный начальный момент r-го порядка $\hat{\alpha}_r$ является сильно состоятельной оценкой для начального момента r-го порядка α_r [12].

На этапе предварительного анализа по случайной выборке $X = \{x_1, \ldots, x_n\}$ объема n обычно вычисляют следующие выборочные числовые характеристики:

Характеристики положения

Минимальное и максимальное значения:

$$\hat{c}_- = x_{(1)} = \min_{1\le i\le n} x_i, \quad \hat{c}_+ = x_{(n)} = \max_{1\le i\le n} x_i. \tag{3.11}$$

Выборочное среднее:

$$\overline{x} = \hat{\nu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.12}$$

$\overline{x}$ $(\hat{\nu}_1)$ – статистическая оценка математического ожидания $\mu = \mathbf{E}\{\xi\}$.

Выборочная мода:

$$\hat{M} = x_{k^*}, \text{ где } k^* = arg\,\max\{\hat{p}_1, \ldots, \hat{p}_K\}, \tag{3.13}$$

где $\hat{p}_k$, $k = 1, \ldots, K$, определены в (3.4).

Выборочная медиана:

$$\hat{m} = \begin{cases} x_{(k+1)}, & n = 2k+1; \\ \frac{x_{(k)}+x_{(k+1)}}{2}, & n = 2k, \end{cases} \tag{3.14}$$

где $x_{(k)}$ – k-я порядковая статистика.

$\hat{m}$ – статистическая оценка медианы m.

Характеристики рассеяния

Выборочная дисперсия:

$$\hat{\sigma}^2 ::= \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{3.15}$$

Выборочная дисперсия $\hat{\sigma}^2$ сходится по вероятности к теоретической дисперсии σ^2 [7]. Следовательно, $\hat{\sigma}^2$ – состоятельная оценка для σ^2. Но так как $\mathbf{E}\{\hat{\sigma}^2\} \neq \sigma^2$, то $\hat{\sigma}^2$ – смещенная оценка для σ^2.

Несмещенной состоятельной оценкой для σ^2 является статистика [7]:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{3.16}$$

При этом s – оценка среднеквадратического отклонения σ.

Размах выборки:

$$\hat{R} = \hat{c}_+ - \hat{c}_-, \tag{3.17}$$

где $\hat{c}_+$ и $\hat{c}_-$ определены в (3.11).

Интерквартильный размах выборки:

$$\hat{IR} = \hat{z}_{\frac{3}{4}} - \hat{z}_{\frac{1}{4}}, \tag{3.18}$$

где $\hat{z}_{\frac{k}{4}}$ – выборочная k-я квартиль:

$$\hat{z}_{\frac{k}{4}} = F_n^{-1}\left(\frac{k}{4}\right),\ k \in \{1,3\}. \tag{3.19}$$

Интервал концентрации:

$$[\overline{x} - 3s, \overline{x} + 3s] \quad \left(\mathbf{P}\{x_i \in [\mu - 3\sigma, \mu + 3\sigma]\} \geq \frac{8}{9}\right).$$

Характеристики формы

Выборочная асимметрия:

$$\hat{\beta}_1 = \frac{\hat{\alpha}_3}{s^3}, \tag{3.20}$$

где $\hat{\alpha}_3$ – выборочный начальный момент 3-го порядка случайной величины, а s – оценка среднеквадратического отклонения.

Выборочный эксцесс:

$$\hat{\beta}_2 = \frac{\hat{\mu}_4}{s^4} - 3, \tag{3.21}$$

где $\hat{\mu}_4$ – выборочный центральный момент 4-го порядка.

3.4 Проблема сжатия данных. Метод главных компонент

В многомерном статистическом анализе исследуемая система описывается вектором, размерность N которого в прикладных задачах может быть достаточно большой. Анализировать такие данные достаточно сложно. Кроме того, те факторы, от которых интересующий исследователя признак не зависит (или зависит в незначительной степени), ухудшают свойства статистических процедур. В частности, включние этих факторов в анализ увеличивает дисперсию оценок параметров и характеристик распределений. Поэтому вполне естественным является желание перейти от многомерной выборки к данным небольшой размерности. Возникает проблема *сжатия данных*.

Одним из наиболее часто используемых методов снижения размерности является метод главных компонент. Основная идея этого метода состоит в последовательном выявлении факторов, в которых данные имеют наибольший разброс.

Пусть в пространстве $\mathbb{R}^N$ зарегистрирована случайная выборка $X = \{x_1, \ldots, x_n\}$ объема n: $x_i \in \mathbb{R}^N$, $i = 1, \ldots, n$.

Необходимо преобразовать исходную выборку $X = \{x_1, \ldots, x_n\}$ в «сжатую» выборку $Y = \{y_1, \ldots, y_n\}$: $y_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, с меньшим числом признаков m ($m < N$), которые несут информацию об исходных наблюдениях с минимальными (наперед заданными) потерями.

Формально решение задачи сжатия данных сводится к поиску борелевского преобразования:

$$y = B(x) : \mathbb{R}^N \to \mathbb{R}^m, \tag{3.22}$$

переводящего исходное наблюдение $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ в «сжатое» наблюдение $y = (\tilde{y}_1, \ldots, \tilde{y}_m)^{\mathbf{T}} \in \mathbb{R}^m$. Выборка Y получается из выборки X преобразованием каждого из n наблюдений: $y_i = B(x_i)$, $i = 1, \ldots, n$.

Преобразование в (3.22), вообще говоря, нелинейное. Однако на практике обычно используют линейные преобразования:

$$y = Bx : \mathbb{R}^N \to \mathbb{R}^m,$$

где B – $(m \times N)$-матрица. Проблема заключается в выборе матрицы B.

Согласно методу главных компонент, случайный N-вектор-наблюдение $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ с ковариационной $(N \times N)$-матрицей

$$\Sigma = \mathbf{Cov}\{x, x\} = \mathbf{E}\{(x - \mu)(x - \mu)^{\mathbf{T}}\},$$

$\mu = \mathbf{E}\{x\} \in \mathbb{R}^N$ – вектор математического ожидания, подвергается следующему линейному преобразованию:

$$y = (\tilde{y}_1(x), \ldots, \tilde{y}_N(x))^{\mathbf{T}},$$

которое покомпонентно записывается в виде:

$$\tilde{y}_k = \tilde{y}_k(x) = \Psi_k^{\mathbf{T}} x, \quad k = 1, \ldots, N, \tag{3.23}$$

где $\{\Psi_k,\ k = 1, \ldots, N\}$ – ортонормированные собственные векторы ковариационной матрицы Σ, удовлетворяющие соотношениям

$$\Sigma\Psi_k = \lambda_k \Psi_k, \quad k = 1, \ldots, N; \tag{3.24}$$

$$\Psi_k^{\mathbf{T}} \Psi_l = \delta_{kl}, \quad k, l = 1, \ldots, N, \tag{3.25}$$

и соответствующие упорядоченным по убыванию собственным числам $\{\lambda_k,\ k = 1, \ldots, N\}$ матрицы Σ:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N.$$

Полученные таким образом случайные величины $\tilde{y}_1, \ldots, \tilde{y}_N$ называют *главными компонентами* для исходного наблюдения $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}}$. Исследуем вероятностные свойства главных компонент.

Теорема 3.1. *Главные компоненты $\tilde{y}_1, \ldots, \tilde{y}_N$, удовлетворяющие (3.23)–(3.25), некоррелированы:*

$$\mathbf{Cov}\{\tilde{y}_k, \tilde{y}_j\} = 0, \quad k \neq j,\ k, j = 1, \ldots, N,$$

а их дисперсии равны соответствующим собственным числам:

$$\mathbf{D}\{\tilde{y}_k\} = \lambda_k, \quad k = 1, \ldots, N. \tag{3.26}$$

Доказательство. С учетом (3.23)–(3.25) вычислим ковариацию случайных величин $\tilde{y}_k$ и $\tilde{y}_j$:

$$\mathbf{Cov}\{\tilde{y}_k, \tilde{y}_j\} = \mathbf{E}\{(\tilde{y}_k - \mathbf{E}\{\tilde{y}_k\})(\tilde{y}_j - \mathbf{E}\{\tilde{y}_j\})\} =$$

$$= \Psi_k^{\mathbf{T}} \mathbf{E}\{(x - \mu)(x - \mu)^{\mathbf{T}}\}\Psi_j = \Psi_k^{\mathbf{T}} \Sigma \Psi_j = \lambda_j \Psi_k^{\mathbf{T}} \Psi_j = \delta_{kj}\lambda_j,\ k, j = 1, \ldots, N.$$

□

Следствие 3.1. *Суммарная дисперсия исходных признаков равна суммарной дисперсии главных компонент:*

$$\sum_{k=1}^{N} \mathbf{D}\{\tilde{x}_k\} = \sum_{k=1}^{N} \mathbf{D}\{\tilde{y}_k\} = \sum_{k=1}^{N} \lambda_k. \tag{3.27}$$

Доказательство. очевидно и следует из (3.26) и известного свойства матрицы [9, 5]:

$$\sum_{k=1}^{N} \mathbf{D}\{\tilde{x}_k\} = \mathbf{tr}(\Sigma) = \sum_{k=1}^{N} \lambda_k.$$

□

Соотношения (3.26), (3.27) позволяют предложить критерий выбора *информативных признаков* в пространстве главных компонент, который состоит в том, что признаки, имеющие малые дисперсии, отбрасываются, а рассматриваются лишь m первых главных компонент из $\tilde{y}_1, \ldots, \tilde{y}_m, \ldots, \tilde{y}_N$ ($\lambda_1 \leq \ldots \leq \lambda_m \leq \ldots \leq \lambda_N$).

Число m обычно определяется по наперед заданной малой величине $\varepsilon \in [0, 1)$:

$$m = m(\varepsilon) = \min\{k :\ 1 - \varkappa(k) \leq \varepsilon,\ k = 1, \ldots, N\}; \tag{3.28}$$

$$\varkappa(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{N} \lambda_j},$$

где $0 < \varkappa(k) \leq 1$ в (3.28) – относительная доля суммарной дисперсии первых k главных компонент. Чем ближе $\varkappa(k)$ к единице (а $(1 - \varkappa(k))$ – к нулю), тем меньше потери информации при сжатии данных ($\varkappa(N) = 1$).

Замечание 3.4.1. *На практике обычно истинное значение ковариационной матрицы Σ неизвестно, и вместо нее в методе главных компонент используется оценка по исходной выборке $X = \{x_1, \ldots, x_n\}$:*

$$S = \frac{1}{n-1} \sum_{t=1}^{n} (x_t - \overline{x})(x_t - \overline{x})^{\mathbf{T}}, \quad \overline{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

Замечание 3.4.2. *Метод главных компонент применяется не только для сжатия данных. Он позволяет декоррелировать их – строить выборки из наблюдений с некоррелированными признаками. С его помощью также производят визуализацию многомерных данных: на плоскости ($m = 2$) и в трехмерном пространстве ($m = 3$). При этом учитывается информация, заключенная во всех исходных признаках, чего нельзя сказать о диаграммах рассеяния, на которых исходные признаки отображаются попарно (на плоскости) и тройками (в трехмерном пространстве).*

Замечание 3.4.3. *Недостаток метода главных компонент состоит в том, что главные компоненты, по сравнению с исходными признаками, не имеют на практике физической интерпретации. С этой точки зрения лучше напрямую выбирать информативные признаки из множества исходных признаков и не производить их функциональных преобразований, то есть осуществлять так называемый «прямой отбор» информативных признаков.*

4 Многомерное нормальное (гауссовское) распределение как модель многомерных данных и его свойства

Нормальное распределение играет особую роль в прикладных задачах теории вероятностей и математической статистики в следствие следующих причин:

– разнообразные статистические данные с хорошей степенью точности можно считать выборками из нормального распределения. Примерами могут служить помехи в электроаппаратуре, ошибки измерений, разброс попадания снарядов при стрельбе по заданной цели, рост наудачу взятого человека, скорость реакции на раздражитель и т.д.;

– если на отклонение исследуемой случайной величины ξ от некоторого заданного значения влияет множество различных факторов, причем влияние каждого из них вносит небольшой вклад в это отклонение, а их действия независимы или почти независимы, то можно предполагать, что ξ имеет нормальное распределение;

– в силу центральной предельной теоремы и ее формулировок для частных случаев ([4], [6], [10]) распределение целого ряда широко распространенных в статистике функций от случайных величин (статистик, оценок) хорошо аппроксимируется нормальным распределением.

4.1 Многомерное нормальное распределение и его свойства

Определение 4.1. *Непрерывный случайный N-мерный вектор $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ с конечными моментами второго порядка $\mathbf{E}\{\tilde{x}_j^2\} < +\infty$, $j = 1, \ldots, N$, имеет невырожденное многомерное (N-мерное) нормальное распределение:*

$$\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma),$$

с N-вектором математического ожидания

$$\mu = \mathbf{E}\{x\} = (\mu_1, \ldots, \mu_N)^{\mathbf{T}} \in \mathbb{R}^N$$

и невырожденной $(N \times N)$-ковариационной матрицей

$$\Sigma = \mathbf{E}\left\{(x-\mu)(x-\mu)^{\mathbf{T}}\right\}, \ |\Sigma| \neq 0,$$

если его плотность распределения вероятностей задается соотношением

$$p(z) = n_N(z|\mu, \Sigma) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(z-\mu)^{\mathbf{T}}\Sigma^{-1}(z-\mu)\Big). \tag{4.1}$$

Замечание 4.1.1. *В случае вырожденной ковариационной матрицы Σ ($|\Sigma| = 0$) иногда также определяют вырожденное многомерное нормальное распределение, которое, однако, не описывается плотностью (6.24). Считается, что случайный N-вектор $x \in \mathbb{R}^N$ имеет вырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| = 0$), если он может быть представлен в виде $x = By + b$, где случайный p-вектор $y \in \mathbb{R}^p$ ($p < N$) имеет невырожденное p-мерное нормальное распределение, $(p \times N)$-матрица линейного преобразования B – полный ранг:* $\mathbf{rank}(B) = p$, *а p-вектор сдвига $b = \mu - B\mathbf{E}\{y\}$.*

Многомерное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$ однозначно определяется своими параметрами:

– *вектором математического ожидания* $\mu = (\mu_1, \ldots, \mu_N)^{\mathbf{T}} \in \mathbb{R}^N$, где $\mu_j = \mathbf{E}\{\tilde{x}_j\}$ – математическое ожидание j-й компоненты $j = 1, \ldots, N$;

– *ковариационной матрицей*

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{pmatrix} \tag{4.2}$$

где σ_{jk} – ковариация случайных величин $\tilde{x}_j$ и $\tilde{x}_k$:

$$\sigma_{jk} = \mathbf{E}\left\{(\tilde{x}_j - m_j) \cdot (\tilde{x}_k - m_k)\right\},\ j,k = 1,\dots,N, \tag{4.3}$$

а σ_{jj} – дисперсия j-й компоненты:

$$\sigma_{jj} = \mathbf{D}\{\tilde{x}_j\} = \mathbf{E}\{(\tilde{x}_j - \mu_j)^2\},\ j = 1,\dots,N.$$

Все характеристики (функциональные и числовые) многомерного нормального распределения $\mathcal{N}_N(\mu,\Sigma)$ можно определить используя параметры μ и Σ в том числе:

– *коэффициент корреляции*:

$$\rho_{jk} = \frac{\mathbf{Cov}\{\tilde{x}_j,\tilde{x}_k\}}{\sqrt{\mathbf{D}\{\tilde{x}_j\}\mathbf{D}\{\tilde{x}_k\}}} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}} \in [-1,1],\ j,k = 1,\dots,N;$$

– *центральные моменты третьего порядка*:

$$\mathbf{E}\{(\tilde{x}_j - \mu_j)(\tilde{x}_k - \mu_k)(\tilde{x}_l - \mu_l)\} = 0,\ j,k,l = 1,\dots,N;$$

– *центральные моменты четвертого порядка*:

$$\mathbf{E}\{(\tilde{x}_j - \mu_j)(\tilde{x}_k - \mu_k)(\tilde{x}_l - \mu_l)(\tilde{x}_m - \mu_m)\} = \sigma_{jk}\sigma_{lm} + \sigma_{jl}\sigma_{km} + \sigma_{jm}\sigma_{lk}; \tag{4.4}$$

$$\mathbf{E}\{(\tilde{x}_j - \mu_j)^4\} = 3\sigma_{jj}^2,\ j,k,l,m = 1,\dots,N.$$

– *характеристическую функцию.*

Теорема 4.1 (О характеристической функции многомерного нормального распределения)**.** *Пусть случайный N-мерный вектор $x \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $\mathcal{N}_N(\mu,\Sigma)$ ($|\Sigma| \neq 0$), тогда его характеристическая функция*

$$f_x(t) = \mathbf{E}\{e^{it^{\mathbf{T}}x}\},\ t \in \mathbb{R}^N,$$

имеет вид

$$f_x(t) = e^{it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t}, \tag{4.5}$$

где i – мнимая единица.

Доказательство. Характеристическая функция непрерывного случайного N-мерного вектора x согласно формуле (2.11) определяется равенством:

$$f_x(t) = \int_{\mathbb{R}^N} \exp\left\{ i\sum_{j=1}^{N} t_j z_j \right\} p_x(z)dz,\ t,z \in \mathbb{R}^N, \tag{4.6}$$

где $p_x(z)$ — N-мерная плотность распределения вероятностей x.

Следовательно, если $\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma)$, то, заменив в (4.20) $p_x(z)$ на $n_N(z|\mu, \Sigma)$ из (6.24), получаем:

$$\begin{aligned} f_x(t) &= \int_{\mathbb{R}^N} \exp\left\{it^{\mathbf{T}}z\right\} n_N(z|\mu,\Sigma)dz = \\ &= \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \int_{\mathbb{R}^N} \exp\left\{it^{\mathbf{T}}z - \frac{1}{2}(z-\mu)^{\mathbf{T}}\Sigma^{-1}(z-\mu)\right\},\ t, z \in \mathbb{R}^N. \end{aligned} \tag{4.7}$$

Выделим полный квадрат в выражении, стоящем под экспонентой:

$$\begin{aligned} & it^{\mathbf{T}}z - \frac{1}{2}(z-\mu)^{\mathbf{T}}\Sigma^{-1}(z-\mu) = \\ &= -\frac{1}{2}(z-(\mu+i\Sigma t)+i\Sigma t)^{\mathbf{T}}\Sigma^{-1}(z-(\mu+i\Sigma t)+i\Sigma t) + it^{\mathbf{T}}z = \\ &= -\frac{1}{2}(z-(\mu+i\Sigma t))^{\mathbf{T}}\Sigma^{-1}(z-(\mu+i\Sigma t)) + it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t. \end{aligned} \tag{4.8}$$

Подставим (4.22) в (4.21):

$$\begin{aligned} f_x(t) &= \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \int_{\mathbb{R}^N} \exp\left\{-\frac{1}{2}(z-(\mu+i\Sigma t))^{\mathbf{T}}\Sigma^{-1}(z-(\mu+i\Sigma t)) + it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t\right\} = \\ &= \exp\left\{it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t\right\} \int_{\mathbb{R}^N} \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(z-(\mu+i\Sigma t))^{\mathbf{T}}\Sigma^{-1}(z-(\mu+i\Sigma t))\right\} = \\ &= \exp\left\{it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t\right\} \int_{\mathbb{R}^N} n_N(z|\mu+i\Sigma t,\Sigma)dz = \\ &= \exp\left\{it^{\mathbf{T}}\mu - \frac{1}{2}t^{\mathbf{T}}\Sigma t\right\},\ t, z \in \mathbb{R}^N, \end{aligned} \tag{4.9}$$

так как в выполняется условие нормировки:

$$\int_{\mathbb{R}^N} n_N(z|\mu+i\Sigma t,\Sigma)dz = 1.$$

□

4.2 Линейные преобразования гауссовских случайных векторов

Теорема 4.2 (О линейном преобразовании гауссовского случайного вектора)**.** *Пусть $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ – случайный вектор, имеющий невырожденное нормальное распределение:*

$$\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma),$$

C – неслучайная $(m \times N)$-матрица

$$C = \begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1N} \\ c_{21} & c_{22} & \ldots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mN} \end{pmatrix},$$

$m \leq N$, полного ранга ($rank(C) = m$), $a \in \mathbf{R}^m$ – фиксированный произвольный m-вектор. Тогда случайный вектор

$$y = Cx + a \tag{4.10}$$

является гауссовским случайным вектором, имеющим распределение $\mathcal{N}_N(\nu, \Xi)$, *где*

$$\begin{aligned} \nu &= C\mu + a, \\ \Xi &= C\Sigma C^{\mathbf{T}}. \end{aligned} \tag{4.11}$$

Доказательство. Найдем характеристическую футкцию $f_y(t)$ случайного вектора y, определенного линейным преобразванием (4.24) гауссовского случайного вектора x:

$$\begin{aligned} f_y(t) = \mathbf{E}\left\{e^{it^{\mathbf{T}}y}\right\} = \mathbf{E}\left\{e^{it^{\mathbf{T}}(Cx+a)}\right\} = e^{it^{\mathbf{T}}a}\mathbf{E}\left\{e^{i(C^{\mathbf{T}}t)^{\mathbf{T}}x}\right\} = f_x\left(C^{\mathbf{T}}t\right) = \\ = e^{i(C^{\mathbf{T}}t)^{\mathbf{T}}\mu - \frac{1}{2}(C^{\mathbf{T}}t)^{\mathbf{T}}\Sigma C^{\mathbf{T}}t}e^{it^{\mathbf{T}}a} = e^{it^{\mathbf{T}}(C\mu+a) - \frac{1}{2}t^{\mathbf{T}}(C\Sigma C^{\mathbf{T}})t}. \end{aligned} \tag{4.12}$$

В правой части соотношения (4.25) получено выражение, которое представляет собой характеристическую функцию гауссовского m-мерного случайного вектора с распределением $\mathcal{N}_N(\nu, \Xi)$, где ν и Ξ определены в (4.23). Распределение вероятностей $\mathcal{N}_N(\nu, \Xi)$ невырождено, так как ковариационная $(m \times m)$-матрица $\Xi = C\Sigma C^{\mathbf{T}}$ имеет ранг m, в силу невырожденности исходной ковариационной $(N \times N)$-матрицы Σ и полного ранга матрицы C $(rank(C) = m)$ имеет ранг m и, следовательно, невырождена. □

Теорема 4.3 (О маргинальных распределениях гауссовского случайного вектора)**.** *Пусть случайный* N*-вектор* $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ *имеет невырожденное нормальное распределение* $\mathcal{N}_N(\mu, \Sigma)$ $(|\Sigma| \neq 0)$, *и пусть* $x^* = (\tilde{x}_{j_1}, \ldots, \tilde{x}_{j_m})^{\mathbf{T}} \in \mathbb{R}^m$ *–* m*-вектор, образованный из каких-либо* m $(1 \leq m \leq N)$ *компонент вектора* x $(j_k \neq j_l \in \{1, \ldots, N\},\ k \neq l,\ k, l = 1, \ldots, m)$. *Тогда случайный* m*-вектор* x^* *имеет невырожденное* m*-мерное маргинальное нормальное распределение с математическим ожиданием* $\mu^* = (\tilde{\mu}_{j_k})_{k=1}^m \in \mathbb{R}^m$ *и ковариационной* $(m \times m)$*-матрицей* $\Sigma^* = (\sigma_{j_k,j_l})_{k,l=1}^m$ $(|\Sigma^*| \neq 0)$, *образованными из соответствующих компонент исходных математического ожидания* $\mu = (\tilde{\mu}_k)_{k=1}^N \in \mathbb{R}^N$ *и ковариационной* $(N \times N)$*-матрицы* $\Sigma = (\sigma_{kl})_{k,l=1}^N$*:*

$$\mathcal{L}\{x^*\} = \mathcal{N}_m(\mu^*, \Sigma^*).$$

Доказательство. Представим m-вектор $x^* = (\tilde{x}_{j_1}, \ldots, \tilde{x}_{j_m})^{\mathbf{T}}$ как линейное преобразование исходного N-вектора $x \in \mathbb{R}^N$:

$$x^* = Cx,$$

где $(m \times N)$-матрица C содержит единицы на пересечении k-й строки и j_k-го столбца $(j_k \in \{1, \ldots, N\},\ k = 1, \ldots, m)$, а остальные ее элементы равны нулю. Очевидно, что $rank(C) = m$, и из теоремы 4.2, с учетом вида матрицы C, получаем доказываемое. □

Следствие 4.1. *Компоненты случайного* N*-мерного вектора* $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$, *распределенного по невырожденному многомерному нормальному закону* $\mathcal{N}_N(\mu, \Sigma)$ $(|\Sigma| \neq 0)$, *имеют одномерные маргинальные нормальные распределения:*

$$\mathcal{L}\{\tilde{x}_j\} = \mathcal{N}_1(\tilde{\mu}_j, \sigma_{jj}),\ \sigma_{jj} > 0,\ j = 1, \ldots, N.$$

Замечание 4.1.2. *Если компоненты случайного вектора имеют одномерные нормальные распределения, то их совместное распределение, вообще говоря, не является многомерным нормальным. Оно будет таковым в случае их независимости (при этом некоррелированности компонент не достаточно!).*

4.3 Независимость компонент гауссовских случайных векторов

Рассмотрим гауссовский случайный вектор $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$, распределенный по многомерному нормальному закону $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Предположим, что x разбит на K векторов-блоков:

$$x = \left((x^{(1)})^{\mathbf{T}} \vdots \ldots \vdots (x^{(K)})^{\mathbf{T}}\right)^{\mathbf{T}} = \\ = (\underbrace{\tilde{x}_1, \ldots, \tilde{x}_{N_1}}_{(x^{(1)})^{\mathbf{T}}}, \underbrace{\tilde{x}_{N_1+1}, \ldots, \tilde{x}_{N_1+N_2}}_{(x^{(2)})^{\mathbf{T}}}, \ldots, \underbrace{\tilde{x}_{N_1+N_2+\ldots+N_{K-1}+1}, \ldots, \tilde{x}_N}_{(x^{(K)})^{\mathbf{T}}})^{\mathbf{T}}, \tag{4.13}$$

Очевидно, что размерность подвектора $x^{(k)}$ равна N_k, $k = 1, \ldots, K$, и $N_1+N_2+\ldots+N_K = N$.

Часто в прикладных задачах требуется установить зависимость или независимость векторов-блоков этого разбиения.

Теорема 4.4. *Векторы-блоки $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ разбиения (4.26) гауссовского случайного N-вектора x:*

$$\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma),$$

независимы тогда и только тогда, когда они некоррелированы.

Доказательство. Так как случайные векторы $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ составлены из компонент гауссовского случайного вектора x, то согласно теореме 4.3, они также являются гауссовскими, причем:

$$\mathcal{L}\{x^{(k)}\} = \mathcal{N}_{N^{(k)}}(\mu^{(k)}, \Sigma^{(k)}),\ |\Sigma^{(k)}| \neq 0,\ k = 1, \ldots, K,$$

Но по известному критерию независимости [11] $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ независимы тогда и только тогда, когда их совместная плотность распределения представима в виде произведения их плотностей, то есть

$$n_N(x|\mu, \Sigma) = \prod_{k=1}^{K} n_{N^{(k)}}(x^{(k)}|\mu^{(k)}, \Sigma^{(k)}),\ x = \left((x^{(1)})^{\mathbf{T}} \vdots \ldots \vdots (x^{(K)})^{\mathbf{T}}\right)^{\mathbf{T}} \in \mathbb{R}^N.$$

Исходя из вида (6.24) плотности многомерного нормального распределения, заключаем, что последнее соотношение выполняется тогда и только тогда, когда случайные векторы $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ некоррелированы: $\Sigma = diag\{\Sigma^{(1)}, \ldots, \Sigma^{(K)}\}$. □

Следствие 4.2. *Пусть случайные векторы $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$ независимы в совокупности и имеют невырожденные нормальные распределения:*

$$\mathcal{L}\{x^{(k)}\} = \mathcal{N}_{N^{(k)}}(\mu^{(k)}, \Sigma^{(k)}),\ |\Sigma^{(k)}| \neq 0,\ k = 1, \ldots, K.$$

Тогда составной вектор $x = \left((x^{(1)})^{\mathbf{T}} \vdots \ldots \vdots (x^{(K)})^{\mathbf{T}}\right)^{\mathbf{T}} \in \mathbb{R}^N$, $N_1 + N_2 + \ldots + N_K = N$, также имеет невырожденное нормальное распределение:

$$\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma),$$

с вектором математического ожидания

$$\mu = \left((\mu^{(1)})^{\mathbf{T}} \vdots \ldots \vdots (\mu^{(K)})^{\mathbf{T}}\right)^{\mathbf{T}}$$

и ковариационной матрицей

$$\Sigma = diag\{\Sigma^{(1)}, \ldots, \Sigma^{(K)}\},\ |\Sigma| \neq 0.$$

4.4 Эллипсоид рассеяния

С невырожденным многомерным нормальным распределением $\mathcal{N}_N(\mu, \Sigma)$ в силу вида его плотности (6.24) связывают *метрику Махаланобиса*:

$$\rho(x, z) = \sqrt{(x - z)^{\mathbf{T}} \Sigma^{-1} (x - z)},\ x, z \in \mathbb{R}^N,$$

которая позволяет построить в $\mathbb{R}^N$ так называемый *эллипсоид рассеяния* с центром в точке $\mu \in \mathbb{R}^N$:

$$V_r = \{x : \rho(x, \mu) \le r,\ r > 0\}.$$

Воспользуемся известным в матричном анализе [9, 5] представлением для невырожденной ковариационной $(N \times N)$-матрицы Σ, которая при этом в силу известного свойства неотрицательной определенности ковариационной матрицы положительно определена ($\Sigma \succ 0$):

$$\Sigma = \Sigma^{1/2} (\Sigma^{1/2})^{\mathbf{T}},$$

где невырожденная $(N \times N)$-матрица $\Sigma^{1/2}$ определяется неоднозначно и является решением по Y матричного уравнения

$$Y^{\mathbf{T}} \Sigma^{-1} Y = \mathbf{I}_N.$$

В частности, можно выбрать

$$\Sigma^{1/2} = (\sqrt{\lambda_1}\Psi_1 \vdots \ldots \vdots \sqrt{\lambda_N}\Psi_N),$$

где $\{\lambda_k, \Psi_k\}_{k=1}^N$ – собственные числа и соответствующие им ортонормированные собственные векторы матрицы Σ, определенные в (3.24), (3.25).

Введем в рассмотрение случайный N-вектор

$$y = (\Sigma^{1/2})^{-1} (x - \mu).$$

Согласно теореме 4.1,

$$\mathcal{L}\{y\} = \mathcal{N}_N(\mathbf{0}_N, \mathbf{I}_N),$$

и случайная величина

$$\rho^2(x, \mu) = y^{\mathbf{T}} y$$

является суммой N независимых в совокупности стандартных нормальных случайных величин и имеет χ^2-распределение с N степенями свободы. Поэтому вероятность попадания случайного вектора x в эллипсоид рассеяния V_r радиуса r равна

$$\mathbf{P}\{x \in V_r\} = F_{\chi^2_N}(\mathbb{R}^2).$$

Этот факт позволяет по заданному уровню вероятности

$$\mathbf{P}\{x \in V_r\} = 1 - \alpha,$$

где $\alpha \in (0, 1)$ – мало, определить радиус эллипсоида рассеяния:

$$r = r(\alpha) = \sqrt{F^{-1}_{\chi^2_N}(1 - \alpha)},$$

где $F^{-1}_{\chi^2_N}(1 - \alpha)$ – квантиль уровня $1 - \alpha$ от χ^2-распределения с N степенями свободы.

Эллипсоид рассеяния V_r, $r = r(\alpha)$, является областью концентрации наблюдений вокруг своего математического ожидания, а область в $\mathbb{R}^N$ вне эллипсоида рассеяния содержит наблюдения, описываемые «хвостом» распределения вероятностей: вероятность попадания в эту область мала и равна α.

4.5 Условные распределения гауссовских случайных векторов. Частный коэффициент корреляции

Пусть случайный N-мерный вектор $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$. Иногда необходимо исследовать зависимость определенного подмножества из m ($m < N$) компонент вектора x между собой при фиксированных (выбранных) значениях остальных $N - m$ компонент.

Предположим, не ограничивая общности, что исследуемыми m компонентами являются $\tilde{x}_1, \ldots, \tilde{x}_m$. Это предположение порождает разбиение вектора $x \in \mathbb{R}^N$ его на два подвектора:

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix},$$

где

$$x^{(1)} = (\tilde{x}_1, \ldots, \tilde{x}_m)^{\mathbf{T}} \in \mathbb{R}^m, \quad x^{(2)} = (\tilde{x}_{m+1}, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^{N-m},$$

а соответствующего ему N-вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ – на соответствующие блоки:

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \ \Sigma_{21} = \Sigma_{12}^{\mathbf{T}}. \tag{4.14}$$

Согласно теореме 4.3 случайные векторы $x^{(1)}$ и $x^{(2)}$ являются гауссовскими с невырожденными нормальными распределениями:

$$\begin{aligned} \mathcal{L}\left\{x^{(1)}\right\} &= \mathcal{N}_m(\mu^{(1)}, \Sigma_{11}), \\ \mathcal{L}\left\{x^{(2)}\right\} &= \mathcal{N}_{N-m}(\mu^{(2)}, \Sigma_{22}). \end{aligned} \tag{4.15}$$

Вычислим условное распределение вероятностей случайного m-вектора $x^{(1)}$ при фиксированном значении компонент $(N - m)$-вектора $x^{(2)}$.

Теорема 4.5 (Об условном распределении вероятностей гауссовских случайных векторов)**.** *Пусть случайный N-вектор $x \in \mathbb{R}^N$, имеющий невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$:*

$$\mathcal{L}\{x\} = \mathcal{N}_N(\mu, \Sigma),$$

разбит на два подвектора $x^{(1)} \in \mathbb{R}^m$ и $x^{(2)} \in \mathbb{R}^{N-m}$:

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}, \ 1 \le m < N.$$

Тогда условное распределение случайного вектора $x^{(1)}$ при фиксированном значении вектора $x^{(2)}$ – m-мерное нормальное со следующими значениями параметров:

$$\mathcal{L}\left\{x^{(1)} | x^{(2)}\right\} = \mathcal{N}_m(\mu_{1|2}, \Sigma_{11|2}), \tag{4.16}$$

$$\begin{aligned} \mu_{1|2} &= \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \\ \Sigma_{11|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^{\mathbf{T}}. \end{aligned} \tag{4.17}$$

Доказательство. Перейдем от случайного N-вектора-наблюдения

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}, \ x^{(1)} \in \mathbb{R}^m, \ x^{(2)} \in \mathbb{R}^{N-m}, \ 1 \le m < N,$$

используя невырожденное линейное функциональное преобразование

$$y = y(x) : \mathbb{R}^N \to \mathbb{R}^N$$

следующего вида:

$$\begin{aligned} y^{(1)} &= x^{(1)} + Bx^{(2)}, \\ y^{(2)} &= x^{(2)}, \end{aligned} \tag{4.18}$$

к вспомогательному вектору

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}$$

где $(m \times (N-m))$-матрица преобразования B выбирается так, чтобы случайные векторы $y^{(1)}$ и $y^{(2)}$ были некоррелированными, то есть, чтобы выполнялось равенство:

$$\mathbf{Cov}\{y^{(1)}, y^{(2)}\} = \mathbf{0}_{m\times(N-m)}, \tag{4.19}$$

где $\mathbf{0}_{m\times(N-m)}$ – нулевая $(m \times (N-m))$-матрица.

Чтобы выбрать матрицу B, вычислим ковариацию между $y^{(1)}$ и $y^{(2)}$:

$$\mathbf{Cov}\left\{y^{(1)}, y^{(2)}\right\} = \mathbf{E}\left\{\left(y^{(1)} - \mathbf{E}\left\{y^{(1)}\right\}\right)\left(y^{(2)} - \mathbf{E}\left\{y^{(2)}\right\}\right)^{\mathbf{T}}\right\}, \tag{4.20}$$

Учитывая равенства (4.18) получаем:

$$\begin{aligned} \mathbf{E}\left\{y^{(1)}\right\} &= \mu^{(1)} + B\mu^{(2)}, \\ \mathbf{E}\left\{y^{(2)}\right\} &= \mu^{(2)}. \end{aligned}$$

Следовательно,

$$\begin{aligned} y^{(1)} - \mathbf{E}\left\{y^{(1)}\right\} &= x^{(1)} + Bx^{(2)} - \mu^{(1)} - B\mu^{(2)} = \dot{x}^{(1)} + B\dot{x}^{(2)}, \\ y^{(2)} - \mathbf{E}\left\{y^{(2)}\right\} &= x^{(2)} - \mu^{(2)} = \dot{x}^{(2)}, \end{aligned}$$

где

$$\begin{aligned} \dot{x}^{(1)} &::= x^{(1)} - \mu^{(1)}, \\ \dot{x}^{(2)} &::= x^{(2)} - \mu^{(2)}. \end{aligned}$$

Таким образом,

$$\begin{aligned} \mathbf{Cov}\left\{y^{(1)}, y^{(2)}\right\} &= \mathbf{E}\left\{\left(\dot{x}^{(1)} + B\dot{x}^{(2)}\right)\left(\dot{x}^{(2)}\right)^{\mathbf{T}}\right\} = \mathbf{E}\left\{\dot{x}^{(1)}\left(\dot{x}^{(2)}\right)^{\mathbf{T}}\right\} + B\mathbf{E}\left\{\dot{x}^{(2)}\left(\dot{x}^{(2)}\right)^{\mathbf{T}}\right\}, \\ &= \Sigma_{12} + B\Sigma_{22}. \end{aligned} \tag{4.21}$$

Приравнивая правую часть соотношения (4.21) к $\mathbf{0}_{m\times(N-m)}$, получаем

$$B = -\Sigma_{12}\Sigma_{22}^{-1}, \tag{4.22}$$

где в силу невырожденности исходной ковариационной матрицы Σ ее диагональный блок Σ_{22} также невырожден.

Так как случайный вектор

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}$$

является линейным преобразованием случайного вектора x, имеющего невырожденное нормальное распределение, то в силу теоремы 4.2, y также имеет нормальное распределение, а, согласно теореме 4.3, его подвекторы $y^{(1)}$ и $y^{(2)}$ – маргинальные нормальные распределения:

$$\begin{aligned} \mathcal{L}\left\{y^{(1)}\right\} &= \mathcal{N}_m\left(\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2}\right), \\ \mathcal{L}\left\{y^{(2)}\right\} &= \mathcal{N}_{N-m}(\mu^{(2)}, \Sigma_{22}), \end{aligned} \tag{4.23}$$

где $\Sigma_{11|2}$ – определена в (4.17) и является ковариационной матрицей вектора

$$y^{(1)} = x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}x^{(2)}.$$

Из некоррелированности случайных векторов $y^{(1)}$ и $y^{(2)}$ в силу их совместного нормального распределения следует их независимость. Поэтому плотность распределения вероятностей вектора y равна произведению плотностей распределения векторов $y^{(1)}$ и $y^{(2)}$:

$$p_y(u) = p_{y^{(1)}}(u^{(1)}) \cdot p_{y^{(2)}}(u^{(2)}) = n_m(u^{(1)}|\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2}) \cdot n_{N-m}(u^{(2)}|\mu^{(2)}, \Sigma_{22}), \tag{4.24}$$

где

$$u = \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix} \in \mathbf{R}^N,\ u^{(1)} \in \mathbf{R}^m,\ u^{(2)} \in \mathbf{R}^{N-m}.$$

Запишем теперь плотность распределения вектора x как преобразования y в x ($x = x(y)$), обратного пребразованию (4.18):

$$\begin{aligned} x^{(1)} &= y^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}y^{(2)}, \\ x^{(2)} &= y^{(2)}. \end{aligned}$$

Отметим, что преобразования здесь имеют единичный якобиан, и из известного соотношения для плотности при невырожденном преобразовании [11] имеем:

$$\begin{aligned} p_x(v) = p_y(u)|_{u=y(v)} &= n_m(v^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}v^{(2)}|\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2}) \cdot n_{N-m}(v^{(2)}|\mu^{(2)}, \Sigma_{22}) = \\ &= n_m(v^{(1)}|\mu_{1|2}, \Sigma_{11|2}) \cdot n_{N-m}(v^{(2)}|\mu^{(2)}, \Sigma_{22}), \end{aligned} \tag{4.25}$$

где

$$v = \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} \in \mathbf{R}^N,\ v^{(1)} \in \mathbf{R}^m,\ v^{(2)} \in \mathbf{R}^{N-m}.$$

а m-вектор $\mu_{1|2}$ определен в (4.17).

С другой стороны, случайный вектор

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}$$

имеет N-мерное распределение вероятностей с плотностью

$$p_x(v) = n_N(v|\mu, \Sigma),$$

а его подвектор $x^{(2)}$ – маргинальное нормальное распределение с плотностью

$$p_{x^{(2)}}(v^{(2)}) = n_{N-m}(v^{(2)}|\mu^{(2)}, \Sigma_{22}).$$

По известной формуле произведения плотностей [11] имеем

$$\begin{aligned} p_x(v) &= p_{x^{(1)}|x^{(2)}}(v^{(1)}|x^{(2)} = v^{(2)}) \cdot p_{x^{(2)}}(v^{(2)}) = \\ &= p_{x^{(1)}|x^{(2)}}(v^{(1)}|x^{(2)} = v^{(2)}) \cdot n_{N-m}(v^{(2)}|\mu^{(2)}, \Sigma_{22}). \end{aligned} \tag{4.26}$$

Приравнивая правые части соотношений (4.25) и (4.26), находим условную плотность:

$$p_{x^{(1)}|x^{(2)}}(v^{(1)}|x^{(2)}=v^{(2)}) = n_m(v^{(1)}|\mu_{1|2},\Sigma_{11|2}).$$

Таким образом,

$$\mathcal{L}\left\{x^{(1)}|x^{(2)}\right\} = \mathcal{N}_m(\mu_{1|2},\Sigma_{11|2}).$$

□

Определение 4.2. $(m\times m)$-*матрицу* $\Sigma_{11|2}$, *определенную в (4.17), иногда называют* условной ковариационной матрицей, *ее внедиагональные элементы* – условными ковариациями, *а диагональные элементы* – условными дисперсиями.

В условиях теоремы 4.1 ковариационная $(m\times m)$-матрица

$$\Sigma_{11|2} = \mathbf{Cov}\{x^{(1)}, x^{(1)}|x^{(2)}\}$$

из (4.17) может быть записана поэлементно:

$$\Sigma_{11|2} = \begin{pmatrix} \sigma_{11|m+1,\dots,N} & \sigma_{12|m+1,\dots,N} & \cdots & \sigma_{1m|m+1,\dots,N} \\ \sigma_{21|m+1,\dots,N} & \sigma_{22|m+1,\dots,N} & \cdots & \sigma_{2m|m+1,\dots,N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1|m+1,\dots,N} & \sigma_{m2|m+1,\dots,N} & \cdots & \sigma_{mm|m+1,\dots,N} \end{pmatrix},$$

где $\sigma_{kl|m+1,\dots,N}$ – условная ковариация k-й и l-й компонент при фиксированных значениях остальных компонент $x^{(2)} = (\tilde{x}_{m+1},\dots,\tilde{x}_N)^{\mathbf{T}}$:

$$\sigma_{kl|m+1,\dots,N} = \mathbf{Cov}\{\tilde{x}_k,\tilde{x}_l|x^{(2)}\} = \sigma_{kl} - \sigma^{\mathbf{T}}_{(k)}\Sigma^{-1}_{22}\sigma_{(l)},\ k,l=1,\dots,m, \tag{4.27}$$

а $\sigma^{\mathbf{T}}_{(k)}$ – k-я строка матрицы Σ_{12}:

$$\sigma^{\mathbf{T}}_{(k)} = \mathbf{Cov}\{\tilde{x}_k, x_{(2)}\},\ k=1,\dots,m. \tag{4.28}$$

Определение 4.3. Частным коэффициентом корреляции *между* k*-й и* l*-й компонентами* $\tilde{x}_k$ *и* $\tilde{x}_l$ $(k,l=1,\dots,m)$ *случайного* N*-вектора* x *при фиксированных значениях компонент*

$$x^{(2)} = (\tilde{x}_{m+1},\dots,\tilde{x}_N)^{\mathbf{T}}$$

называется величина

$$\rho_{kl|m+1,\dots,N} = \frac{\sigma_{kl|m+1,\dots,N}}{\sqrt{\sigma_{kk|m+1,\dots,N}\sigma_{ll|m+1,\dots,N}}}, \tag{4.29}$$

где условная ковариация $\sigma_{kl|m+1,\dots,N}$ *называется* частной ковариацией, *а условная дисперсия* $\sigma_{kk|m+1,\dots,N}$ – частной дисперсией.

Частный коэффициент корреляции является характеристикой парной взаимосвязи на множестве выбранных компонент при фиксированных значениях остальных компонент и не зависит от фиксируемых значений компонент.

Отметим, что частные корреляция, ковариация и дисперсия обладают всеми свойствами своих обычных (безусловных) аналогов.

4.6 Функция регрессии. Множественный коэффициент корреляции

Пусть случайный N-мерный вектор $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}}$ разбит на два подвектора:

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix},$$

где

$$x^{(1)} = (\tilde{x}_1, \ldots, \tilde{x}_m)^{\mathbf{T}} \in \mathbb{R}^m, \quad x^{(2)} = (\tilde{x}_{m+1}, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^{N-m}.$$

Определение 4.4. Функцией регрессии *подвектора* $x^{(1)} \in \mathbb{R}^m$ *на* $x^{(2)} \in \mathbb{R}^{N-m}$ *называется условное математическое ожидание*

$$\mu_{1|2} = \mu_{1|2}\left(x^{(2)}\right) = \mathbf{E}\left\{x^{(1)}|x^{(2)}\right\}. \tag{4.30}$$

Предположим, что N-вектор $x \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$. В этом случае функция регрессии $x^{(1)}$ на $x^{(2)}$, согласно (4.17), линейна и имеет вид:

$$\mu_{1|2} = \mu^{(1)} + \Sigma_{12} \cdot \Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}),$$

где $\mu^{(1)}$, $\mu^{(2)}$ Σ_{12} и Σ_{22} определены в (6.24). Матрица $B = \Sigma_{12} \cdot \Sigma_{22}^{-1}$ называется *матрицей коэффициентов регрессии*:

$$B = \Sigma_{12} \cdot \Sigma_{22}^{-1} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \ldots & \beta_{1,N-m} \\ \beta_{2,1} & \beta_{2,2} & \ldots & \beta_{2,N-m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \ldots & \beta_{m,N-m} \end{pmatrix} = \begin{pmatrix} \beta_1^{\mathbf{T}} \\ \beta_2^{\mathbf{T}} \\ \vdots \\ \beta_m^{\mathbf{T}} \end{pmatrix}, \tag{4.31}$$

где $\beta_k^{\mathbf{T}}$ – k-я строка матрицы $B = \Sigma_{12} \cdot \Sigma_{22}^{-1}$:

$$\beta_k^{\mathbf{T}} = \sigma_{(k)}^{\mathbf{T}} \cdot \Sigma_{22}^{-1} \;\; k = 1, \ldots, m, \tag{4.32}$$

а $\sigma_{(k)}^{\mathbf{T}}$ – k-я строка матрицы Σ_{12}:

$$\sigma_{(k)}^{\mathbf{T}} = (\sigma_{k,m+1}\sigma_{k,m+2}\ldots\sigma_{k,N}) \;\; k = 1, \ldots, m. \tag{4.33}$$

Учитывая (4.32) и (4.33) функцию регрессии k й компоненты $\tilde{x}_k$ вектора $x^{(1)}$ на $x^{(2)}$ можно записать в виде:

$$\begin{aligned} \mathbf{E}\left\{\tilde{x}_k|x^{(2)}\right\} &= \mu_k + \beta_k^{\mathbf{T}} \cdot \left(x^{(2)} - \mu^{(2)}\right) = \mu_k - \beta_k^{\mathbf{T}} \cdot \left(\mu^{(2)} - x^{(2)}\right) = \\ &= \mu_k - \sum_{j=m+1}^{N} \beta_{k,j}\left(\mu_j - \tilde{x}_j\right) \;\; k = 1, \ldots, m. \end{aligned} \tag{4.34}$$

Замечание 4.4.1. *Из формулы (4.34) видно, что коэффициент регрессии* $\beta_{k,j}$ *можно рассматривать как коэффициент влияния компоненты* $\tilde{x}_j$ *на* $\mathbf{E}\left\{\tilde{x}_k|x^{(2)}\right\}$*:*

– *чем больше значение* $|\beta_{k,j}|$*, тем сильнее влияет* $\tilde{x}_j$ *на* $\mathbf{E}\left\{\tilde{x}_k|x^{(2)}\right\}$, $k = 1, \ldots, m$, $j = m+1, \ldots, N$*;*

– *если* $\beta_{k,j} = 0$*, то* $\tilde{x}_j$ *не влияет на* $\mathbf{E}\left\{\tilde{x}_k|x^{(2)}\right\}$, $k = 1, \ldots, m$, $j = m+1, \ldots, N$.

Предположим, что необходимо оценить неизвестные значения компонент из $x^{(1)} \in \mathbb{R}^m$ по наблюдаемым компонентам $x^{(2)} \in \mathbb{R}^{N-m}$ $(m < N)$. Эта задача имеет большое прикладное значение при прогнозировании и оценивании («заполнении») пропущенных значений

компонент в выборке. Математически задача прогнозирования (или оценивания пропущенных значений компонент в выборке) заключается в построении функционального преобразования $g(\cdot) : \mathbb{R}^{N-m} \to \mathbb{R}^m$

$$\hat{x}^{(1)} = g\left(x^{(2)}\right),\ g(\cdot) \in \mathcal{G} \subset \mathbb{R}^m, \tag{4.35}$$

где $\hat{x}^{(1)}$ – вычисленное предсказанние (оценка) для $x^{(1)}$, называемое *прогнозом* или *значением прогноза*, а $g(\cdot)$ – *функция прогнозирования* (*предиктор*), $\mathcal{G}$ – некоторый допустимый класс функций.

Теорема 4.6 (Об оптимальных свойствах функции регрессии)**.** *В условиях теоремы 4.5 среди всех прогнозов*

$$\hat{x}^{(1)} = \left(\hat{\tilde{x}}_1, \ldots, \hat{\tilde{x}}_m\right)^{\mathbf{T}}$$

для $x^{(1)}$ по $x^{(2)}$ прогноз

$$\hat{x}^{(1)}_* = \left(\hat{\tilde{x}}^*_1, \ldots, \hat{\tilde{x}}^*_m\right)^{\mathbf{T}} = \mu_{1|2}(x^{(2)}),$$

где $\mu_{1|2} = \mu_{1|2}(x^{(2)})$ – функция регрессии, определенная в (4.30)*, является оптимальным прогнозом в смысле среднеквадратической ошибки прогнозирования, то есть:*

$$\hat{\tilde{x}}^*_k = \arg\min_{\hat{\tilde{x}}_k} \mathbf{E}\{(\hat{\tilde{x}}_k - \tilde{x}_k)^2\},\ k = 1, \ldots, m, \tag{4.36}$$

и максимально коррелирован с предсказываемым значением $\tilde{x}_k$

$$\hat{\tilde{x}}^*_k = \arg\max_{\hat{\tilde{x}}_k} \mathbf{Corr}\{\hat{\tilde{x}}_k, \tilde{x}_k\},\ k = 1, \ldots, m. \tag{4.37}$$

Доказательство. Пусть $\hat{\tilde{x}}_k$ – любой допустимый прогноз компоненты $\tilde{x}_k$ подвектора $x^{(1)}$. Запишем для этого прогноза среднеквадратическую ошибку прогнозирования $\Delta^2(\hat{\tilde{x}}_k)$:

$$\Delta^2(\hat{\tilde{x}}_k) = \mathbf{E}\{(\hat{\tilde{x}}_k - \tilde{x}_k)^2\} = \mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k + \hat{\tilde{x}}^*_k - \tilde{x}_k)^2\} =$$

$$= \mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)^2\} + 2\mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)(\hat{\tilde{x}}^*_k - \tilde{x}_k)\} + \mathbf{E}\{(\hat{\tilde{x}}^*_k - \tilde{x}_k)^2\},\ k = 1, \ldots, m, \tag{4.38}$$

где $\hat{\tilde{x}}^*_k$ – условное математическое ожидание $\tilde{x}_k$ при фиксированном значении вектора $x^{(1)}$:

$$\hat{\tilde{x}}^*_k = \mathbf{E}\{\tilde{x}_k | x^{(2)}\}\ k = 1, \ldots, m.$$

По формуле полного математического ожидания [11] имеем для любого допустимого прогноза $\hat{\tilde{x}}_k$:

$$\mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)(\hat{\tilde{x}}^*_k - \tilde{x}_k)\} = \mathbf{E}\left\{\mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)(\hat{\tilde{x}}^*_k - \tilde{x}_k)\}|x^{(2)}\right\} =$$

$$= \mathbf{E}\left\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)\mathbf{E}\{(\hat{\tilde{x}}^*_k - \tilde{x}_k)\}|x^{(2)}\right\} = \mathbf{E}\left\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)(\hat{\tilde{x}}^*_k - \hat{\tilde{x}}^*_k)\right\} = 0\ k = 1, \ldots, m.$$

Следовательно, так как

$$\mathbf{E}\{(\hat{\tilde{x}}^*_k - \tilde{x}_k)^2\} = \Delta^2(\hat{\tilde{x}}^*_k)$$

формулу (4.40) можно записать в виде:

$$\Delta^2(\hat{\tilde{x}}_k) = \mathbf{E}\{(\hat{\tilde{x}}_k - \hat{\tilde{x}}^*_k)^2\} + \Delta^2(\hat{\tilde{x}}^*_k) \geq \Delta^2(\hat{\tilde{x}}^*_k),\ k = 1, \ldots, m. \tag{4.39}$$

Равенство будет в том случае, когда $\hat{\tilde{x}}_k \equiv \hat{\tilde{x}}^*_k,\ k = 1, \ldots, m$. Таким образом, $\hat{\tilde{x}}^*_k$ – оптимальный прогноз в смысле среднеквадратической ошибки прогнозирования компоненты $\tilde{x}_k$ подвектора $x^{(1)}$ при фиксированном $x^{(2)},\ k = 1, \ldots, m$.

Для доказательства второго утверждения теоремы вычислим ковариацию произвольного прогноза $\hat{\tilde{x}}_k$ с прогнозируемой компонентой $\tilde{x}_k$, учитывая, что

$$\mathbf{E}\{\tilde{x}_k\} = \mathbf{E}\left\{\mathbf{E}\{\tilde{x}_k|x^{(2)}\}\right\} = \mathbf{E}\{\hat{\tilde{x}}_k^*\},\ k = 1, \ldots, m:$$

$$\mathbf{Cov}\{\hat{\tilde{x}}_k, \tilde{x}_k\} = \mathbf{E}\{(\hat{\tilde{x}}_k - \mathbf{E}\{\hat{\tilde{x}}_k\})(\tilde{x}_k - \mathbf{E}\{\tilde{x}_k\})\} == \mathbf{E}\left\{(\hat{\tilde{x}}_k - \mathbf{E}\{\hat{\tilde{x}}_k\})\mathbf{E}\{(\tilde{x}_k - \mathbf{E}\{\tilde{x}_k\})|x^{(2)}\}\right\} =$$

$$= \mathbf{E}\{(\hat{\tilde{x}}_k - \mathbf{E}\{\hat{\tilde{x}}_k\})(\hat{\tilde{x}}_k^* - \mathbf{E}\{\hat{\tilde{x}}_k^*\})\} = \mathbf{Cov}\{\hat{\tilde{x}}_k, \hat{\tilde{x}}_k^*\},\ k = 1, \ldots, m.$$

Следовательно,

$$\mathbf{Corr}^2\{\hat{\tilde{x}}_k, \tilde{x}_k\} = \frac{\mathbf{Cov}^2\{\hat{\tilde{x}}_k, \tilde{x}_k\}}{\mathbf{D}\{\hat{\tilde{x}}_k\}\mathbf{D}\{\tilde{x}_k\}} = \frac{\mathbf{Cov}^2\{\hat{\tilde{x}}_k, \hat{\tilde{x}}_k^*\}}{\mathbf{D}\{\hat{\tilde{x}}_k\}\mathbf{D}\{\hat{\tilde{x}}_k^*\}}\frac{\mathbf{D}\{\hat{\tilde{x}}_k^*\}}{\mathbf{D}\{\tilde{x}_k\}} =$$

$$= \mathbf{Corr}^2\{\hat{\tilde{x}}_k, \hat{\tilde{x}}_k^*\}\mathbf{Corr}^2\{\hat{\tilde{x}}_k^*, \tilde{x}_k\},\ k = 1, \ldots, m.$$

Поскольку $\mathbf{Corr}^2\{\hat{\tilde{x}}_k, \hat{\tilde{x}}_k^*\} \le 1$, то получаем

$$|\mathbf{Corr}\{\hat{\tilde{x}}_k, \tilde{x}_k\}| \le |\mathbf{Corr}\{\hat{\tilde{x}}_k^*, \tilde{x}_k\}|,\ k = 1, \ldots, m.$$

□

Следствие 4.3.

$$\mathbf{E}\{(\hat{\tilde{x}}_k^* - \tilde{x}_k)^2\} = \sigma_{kk|m+1,\ldots,N}; \tag{4.40}$$

$$\mathbf{Corr}\{\hat{\tilde{x}}_k^*, \tilde{x}_k\} = \sqrt{\frac{\sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}\sigma_{(k)}}{\sigma_{kk}}}, \tag{4.41}$$

где Σ_{22} – ковариационная матрица вектора $x^{(2)}$, σ_{kk} – дисперсия k-й компоненты, а $(N-m)$-вектор $\sigma_{(k)}$ – определен в (4.27), $k = 1, \ldots, m$.

Доказательство. Для доказательства формул (4.40) и (4.41) воспользуемся соотношением (4.30) и запишем оптимальный прогноз $\hat{x}_*^{(1)} = (\hat{\tilde{x}}_1^*, \ldots, \hat{\tilde{x}}_k^*)^{\mathbf{T}}$ покомпонентно:

$$\hat{\tilde{x}}_k^* = \tilde{\mu}_k + \sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}),\ k = 1, \ldots, m, \tag{4.42}$$

где $\tilde{\mu}_k = \mathbf{E}\{\tilde{x}_k\}$, а $\mu^{(2)} = \mathbf{E}\{x^{(2)}\}$. Учтем, что

$$\mathbf{E}\{\hat{\tilde{x}}_k^*\} = \mathbf{E}\{\mathbf{E}\{\tilde{x}_k|x^{(2)}\}\} = \mathbf{E}\{\tilde{x}_k\} = \tilde{\mu}_k,$$

и вычислим ковариацию:

$$\mathbf{Cov}\{\hat{\tilde{x}}_k^*, \tilde{x}_k\} = \mathbf{E}\{(\hat{\tilde{x}}_k^* - \tilde{\mu}_k)(\tilde{x}_k - \tilde{\mu}_k)\} =$$

$$= \mathbf{E}\{\sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})(\tilde{x}_k - \tilde{\mu}_k)\} = \sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}\sigma_{(k)}.$$

Отметим, что $\mathbf{D}\{\tilde{x}_k\} = \sigma_{kk}$, а для $\mathbf{D}\{\hat{\tilde{x}}_k^*\}$ имеем

$$\mathbf{D}\{\hat{\tilde{x}}_k^*\} = \mathbf{E}\{(\hat{\tilde{x}}_k^* - \tilde{\mu}_k)^2\} = \mathbf{E}\{(\sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}))^2\} =$$

$$= \sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}\mathbf{Cov}\{x^{(2)}, x^{(2)}\}\Sigma_{22}^{-1}\sigma_{(k)} = \sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}\sigma_{(k)}.$$

Полученные соотношения и позволяют вычислить корреляцию в (4.41), а также доказать (4.40):

$$\mathbf{E}\{(\hat{\tilde{x}}_k^* - \tilde{x}_k)^2\} = \mathbf{D}\{\hat{\tilde{x}}_k^* - \tilde{x}_k\} = \mathbf{D}\{\hat{\tilde{x}}_k^*\} - 2\mathbf{Cov}\{\hat{\tilde{x}}_k^*, \tilde{x}_k\} + \mathbf{D}\{\tilde{x}_k\} =$$

$$= \sigma_{kk} - \sigma_{(k)}^{\mathbf{T}}\Sigma_{22}^{-1}\sigma_{(k)} = \sigma_{kk|m+1,\ldots,N},$$

где учтено (4.27). □

Определение 4.5. Множественным коэффициентом корреляции *между компонентой* $\tilde{x}_k$ $(k = 1, \ldots, m)$ *и компонентами подвектора* $x^{(2)} = (\tilde{x}_{m+1}, \ldots, \tilde{x}_N)^{\mathbf{T}}$ *называется величина*

$$\overline{R}_{k,m+1,\ldots,N} = \sqrt{\frac{\sigma_{(k)}^{\mathbf{T}} \Sigma_{22}^{-1} \sigma_{(k)}}{\sigma_{kk}}}. \tag{4.43}$$

Свойства множественного коэффициента корреляции

Свойство 1. $0 \leq \overline{R}_{k,m+1,\ldots,N} \leq 1,\ k = 1, \ldots, m.$

Доказательство. Доказательство следует из определения (4.43) и соотношения (4.27):

$$\mathbf{D}\{\tilde{x}_k | x^{(2)}\} = \sigma_{kk|m+1,\ldots,N} = \sigma_{kk} - \sigma_{(k)}^{\mathbf{T}} \Sigma_{22}^{-1} \sigma_{(k)} \geq 0,\ k = 1, \ldots, m.$$

□

Свойство 2. $\overline{R}_{k,m+1,\ldots,N} = 0$ тогда и только тогда, когда k-я компонента $\tilde{x}_k$ $(k = 1, \ldots, m)$ некоррелирована с компонентами из множества $\{\tilde{x}_{m+1}, \ldots, \tilde{x}_N\}$.

Доказательство. Из (4.43) видно, что $\overline{R}_{k,m+1,\ldots,N} = 0$ тогда и только тогда, когда

$$\sigma_{(k)} = \mathbf{Cov}\{x^{(2)}, \tilde{x}_k\} = \mathbf{0}_{N-m},\ k = 1, \ldots, m.$$

□

Свойство 3. $\overline{R}_{k,m+1,\ldots,N} = 1$ тогда и только тогда, когда k-я компонента $\tilde{x}_k$ представима в виде

$$\tilde{x}_k \overset{\mathrm{P}=1}{=} b_k^{\mathbf{T}} x^{(2)} + d_k,$$

где $b_k \in \mathbb{R}^{N-m}$, $d_k \in \mathbb{R}^1$ – детерминированы, $k = 1, \ldots, m$.

Доказательство. Из (4.43), (4.41) и свойств обычного коэффициета корреляции следует, что

$$\overline{R}_{k,m+1,\ldots,N} = \mathbf{Corr}\{\hat{\tilde{x}}_k^*, \tilde{x}_k\} = 1$$

тогда и только тогда, когда

$$\tilde{x}_k \overset{\mathrm{P}=1}{=} \tilde{b}_k^{\mathbf{T}} \hat{\tilde{x}}_k^* + \tilde{d}_k,$$

где $\tilde{b}_k \in \mathbb{R}^{N-m}$, $\tilde{d}_k \in \mathbb{R}^1$ – детерминированы, $k = 1, \ldots, m$. Подставим вместо $\hat{\tilde{x}}_k^*$ его выражение через $x^{(2)}$ из (4.42) и получим доказываемое. □

Свойство 4. Справедливо следующее соотношение, связывающее частную дисперсию $\sigma_{kk|m+1,\ldots,N}$ из (4.27) и множественный коэффициент корреляции $\overline{R}_{k,m+1,\ldots,N}$ $(k = 1, \ldots, m)$:

$$\sigma_{kk|m+1,\ldots,N} = (1 - \overline{R}_{k,m+1,\ldots,N}^2)\sigma_{kk},$$

и частная дисперсия $\sigma_{kk|m+1,\ldots,N}$ никогда не превосходит соответствующей безусловной дисперсии σ_{kk}: $\sigma_{kk|m+1,\ldots,N} \leq \sigma_{kk}$.

Доказательство. непосредственно следует из соотношения (4.27) и свойства 1. □

Свойство 5. Если $N = 2$, $m = 1$, то множественный коэффициент корреляции (4.43) совпадает с точностью до знака с парным коэффициентом корреляции:

$$\overline{R}_{1,2} = |\rho_{12}|.$$

Замечание 4.5.1. *В теории множественной регрессии множественный коэффициент корреляции служит мерой оптимальности прогноза: чем он ближе к единице, тем точнее прогноз.*

5 Статистическое оценивание параметров многомерной гауссовской модели

5.1 Оценки максимального правдоподобия параметров многомерного нормального распределения

Пусть имеется случайная выборка $X = \{x_1, \ldots, x_n\}$ объема n из невырожденного N-мерного нормального распределения $\mathcal{N}_N(\mu, \Sigma)$ с неизвестными истинными значениями параметров: математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ ($|\Sigma| \neq 0$), для построения статистических оценок которых ($\hat{\mu}$ и $\hat{\Sigma}$) по выборке X воспользуемся методом максимального правдоподобия [12].

Теорема 5.1 (Об оценках максимального правдоподобия параметров многомерного нормального распределения)**.** *Пусть наблюдения $x_1, \ldots, x_n \in \mathbb{R}^N$, образующие выборку X объема $n > N$, являются независимыми в совокупности, одинаково распределенными случайными N-векторами с невырожденным нормальным распределением $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Тогда единственными оценками максимального правдоподобия (МП-оценками) для вектора математического ожидания μ и ковариационной матрицы Σ являются соответственно выборочное среднее $\hat{\mu}$:*

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \in \mathbb{R}^N, \tag{5.1}$$

и выборочная ковариационная $(N \times N)$-матрица $\hat{\Sigma}$:

$$\hat{\Sigma} = \frac{1}{n}A, \tag{5.2}$$

где

$$A = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathbf{T}}. \tag{5.3}$$

Доказательство. Плотность распределения вероятностей $n_N(\cdot|\mu, \Sigma)$ многомерного нормального закона распределения $\mathcal{N}_N(\mu, \Sigma)$, согласно определению (6.24) имеет вид:

$$n_N(z|\mu, \Sigma) = (2\pi)^{-\frac{N}{2}}|\Sigma|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}(z-\mu)^{\mathbf{T}}\Sigma^{-1}(z-\mu)\Big).$$

Следовательно, логарифмическая функциия прадоподобия $l(\mu, \Sigma)$ [12] для неизвестных значений параметров μ и Σ, построенная по случайной выборке по выборке $X = \{x_1, \ldots, x_n\}$, записывается следующим образом:

$$l(\mu, \Sigma) = \ln\left(\prod_{i=1}^{n} n_N(x_i|\mu, \Sigma)\right) = \ln\left(\prod_{i=1}^{n}(2\pi)^{-\frac{N}{2}}|\Sigma|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}(x_i-\mu)^{\mathbf{T}}\Sigma^{-1}(x_i-\mu)\Big)\right) =$$

$$= \sum_{i=1}^{n}\left(-\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_i-\mu)^{\mathbf{T}}\Sigma^{-1}(x_i-\mu)\right) =$$

$$= -\frac{nN}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^{\mathbf{T}}\Sigma^{-1}(x_i-\mu) =$$

$$= -\frac{nN}{2}\ln(2\pi) + \frac{n}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{n} tr\big(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^{\mathbf{T}}\big) =$$

$$= -\frac{nN}{2}\ln(2\pi) + \frac{n}{2}\ln|\Sigma^{-1}| - \frac{1}{2}tr\big(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^{\mathbf{T}}\big), \tag{5.4}$$

где использованы следующие свойства следа матрицы $tr(\cdot)$:

- $c = tr(c),\ c \in \mathbb{R}$;
- $tr(BC) = tr(CB)$;
- $tr(B+D) = tr(B) + tr(D)$.

В формуле (5.4) преобразуем сумму

$$\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^{\mathbf{T}},$$

учитывая очевидное тождество

$$\sum_{t=1}^{n}(x_t-\overline{x}) \equiv \mathbf{0}_N. \tag{5.5}$$

Имеем:

$$\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^{\mathbf{T}} = \sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\mu)(x_i-\overline{x}+\overline{x}-\mu)^{\mathbf{T}} =$$

$$= \sum_{i=1}^{n}[(x_i-\overline{x})(x_i-\overline{x})^{\mathbf{T}} + (x_i-\overline{x})(\overline{x}-\mu)^{\mathbf{T}} + (\overline{x}-\mu)(x_i-\overline{x})^{\mathbf{T}} + (\overline{x}-\mu)(\overline{x}-\mu)^{\mathbf{T}}] =$$

$$= A + n(\overline{x}-\mu)(\overline{x}-\mu)^{\mathbf{T}}, \tag{5.6}$$

где $\overline{x}$ и A определены соответственно в (5.1) и (5.3).

Подставив (5.6) в (5.4) и используя свойства следа матрицы, получаем:

$$l(\mu,\Sigma) = -\frac{nN}{2}\ln(2\pi) + \frac{n}{2}\ln|\Sigma^{-1}| - \frac{1}{2}tr(\Sigma^{-1}A) - \frac{n}{2}(\overline{x}-\mu)^{\mathbf{T}}\Sigma^{-1}(\overline{x}-\mu). \tag{5.7}$$

Из вида (5.7) логарифмической функции правдоподобия $l(\mu,\Sigma)$ заключаем, что максимум по $\mu \in \mathbb{R}^N$ в ней достигается лишь на выборочном среднем $\hat{\mu} = \overline{x}$, которое и является МП-оценкой для вектора математического ожидания μ.

Из (5.7) также видно, что вместо МП-оценки $\hat{\Sigma}$ для ковариационной матрицы Σ целесообразно строить МП-оценку

$$\widehat{\Sigma^{-1}} = (\sigma^*_{kj})_{k,j=1}^{N}$$

для обратной ковариационной матрицы Σ^{-1}. В силу взаимно-однозначного соответствия между Σ и Σ^{-1} МП-оценкой для Σ будет

$$\hat{\Sigma} = (\widehat{\Sigma^{-1}})^{-1},$$

которая находится из решения задачи:

$$l^*(\Sigma^{-1}) = l(\overline{x},\Sigma) = -\frac{nN}{2}\ln(2\pi) + \frac{n}{2}\ln|\Sigma^{-1}| - \frac{1}{2}tr(\Sigma^{-1}A) \to \max_{\Sigma^{-1}}.$$

Вычислим частные производные от $l^*(\Sigma^{-1})$ по σ^*_{kj}, используя поэлементное представление матрицы A:

$$A = (a_{kj})_{k,j=1}^{N},$$

разложение определителя по строке и известное в матричном анализе представление Крамера для элементов обратной матрицы [9, 5]:

$$\frac{\partial l^*(\Sigma^{-1})}{\partial \sigma^*_{kj}} = \frac{n}{2}\frac{1}{|\Sigma^{-1}|}\frac{\partial|\Sigma^{-1}|}{\partial\sigma^*_{kj}} - \frac{1}{2}\frac{\partial}{\partial\sigma^*_{kj}}\sum_{i,l=1}^{N}\sigma^*_{il}a_{li} = \frac{n}{2}\sigma_{kj} - \frac{1}{2}a_{jk},\ k,j = 1,\ldots,N.$$

Согласно необходимому условию максимума, полученные выражения для производных приравниваем к нулю и получаем МП-оценку $\hat{\Sigma} = (\hat{\sigma}_{kj})_{k,j=1}^{N}$ для ковариационной матрицы $\Sigma = (\sigma_{kj})_{k,j=1}^{N}$:

$$\hat{\sigma}_{kj} = \frac{1}{n} a_{jk},\ k, j = 1, \ldots, N,$$

откуда в силу симметричности матрицы A:

$$a_{jk} = a_{kj},\ k, j = 1, \ldots, N,$$

окончательно имеем

$$\hat{\Sigma} = \frac{1}{n} A,$$

что совпадает с (5.2). Условие $n > N$ обеспечивает невырожденность полученной оценки ковариационной матрицы: $\mathbf{P}\{|A| = 0\} = 0$. □

Выборочное среднее

$$\hat{\mu} = \overline{x}$$

и выборочная ковариационная матрица

$$\hat{\Sigma} = \frac{1}{n} A,$$

полученные в теореме 5.1, как МП-оценки обладают всеми их свойствами: сильная состоятельность, асимптотическая несмещенность и асимптотическая нормальность.

5.2 Распределение вероятностей оценок максимального правдоподобия параметров многомерного нормального распределения

Пусть C – ортогональная $(n \times n)$-матрица:

$$C = (c_{ij})_{i,j=1}^{n}, \tag{5.8}$$

обладающую свойством [9], [5]:

$$\begin{aligned} &CC^{\mathbf{T}} = C^{\mathbf{T}}C = \mathbf{I}_n,\ (C^{-1} = C^{\mathbf{T}}), \\ &\sum_{m=1}^{n} c_{km} \cdot c_{ml} = \delta_{kl} = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases}\ k, l = 1, \ldots, n. \end{aligned} \tag{5.9}$$

Лемма 5.1. *Пусть $x_1, \ldots, x_n$ – независимые в совокупности N-мерные гауссовские случайные векторы:*

$$\mathcal{L}\{x_i\} = \mathcal{N}_N(\mu_i, \Sigma),\ i = 1, \ldots, n.$$

Тогда случайные векторы $y_1, \ldots, y_n$:

$$y_i = \sum_{j=1}^{n} c_{ij} \cdot x_j\ i = 1, \ldots, n, \tag{5.10}$$

где $C = (c_{ij})_{i,j=1}^{n}$ – ортогональная матрица, определеннная в (5.8) и (5.9), также являются независимыми и имеют нормальные распределения:

$$\mathcal{L}\{y_i\} = \mathcal{N}_N(\nu_i, \Sigma),\ i = 1, \ldots, n, \tag{5.11}$$

причем

$$\nu_i = \sum_{j=1}^{n} c_{ij} \cdot \mu_j \; i = 1, \ldots, n. \tag{5.12}$$

Кроме того, остается неизменной матричная сумма квадратов:

$$\sum_{j=1}^{n} x_j \cdot (x_j)^{\mathbf{T}} = \sum_{j=1}^{n} y_j \cdot (y_j)^{\mathbf{T}}. \tag{5.13}$$

Доказательство. Случайные векторы $y_1, \ldots, y_n$ являются линейными преобразованиями нормально распределенных независимых в совокупности случайных векторов $x_1, \ldots, x_n$, имеющих, согласно следствию 4.2, совместное нормальное распределение. По теореме 4.2 $y_1, \ldots, y_n$ также имеют совместное нормальное распределение и по теореме 4.3 нормально распределены:

$$\mathcal{L}\{y_i\} = \mathcal{N}_N(\nu_i, \Sigma_i), \; i = 1, \ldots, n.$$

Найдем параметры ν_i и Σ_i, $i = 1, \ldots, n$, этих распределений.

Для математических ожиданий имеем

$$\nu_i = \mathbf{E}\{y_i\} = \mathbf{E}\left\{\sum_{j=1}^{n} c_{ij} x_j\right\} = \sum_{j=1}^{n} c_{ij} E\{x_j\} = \sum_{j=1}^{n} c_{ij}\mu_j, \; i = 1, \ldots, n.$$

Вместо вычисления ковариационных матриц Σ_i, $i = 1, \ldots, n$, решим более общую задачу – найдем ковариации векторов $y_1, \ldots, y_n$:

$$\mathbf{Cov}\{y_i, y_l\} = \mathbf{E}\{(y_i - \nu_i)(y_l - \nu_l)^{\mathbf{T}}\} =$$

$$= \mathbf{E}\left\{\left(\sum_{j=1}^{n} c_{ij} x_j - \sum_{j=1}^{n} c_{ij}\mu_j\right)\left(\sum_{k=1}^{n} c_{lk} x_k - \sum_{k=1}^{n} c_{lk}\mu_k\right)^{\mathbf{T}}\right\} =$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} c_{ij} c_{lk} \mathbf{E}\{(x_j - \mu_j)(x_k - \mu_k)^{\mathbf{T}}\} = \sum_{j=1}^{n}\sum_{k=1}^{n} c_{ij} c_{lk} \mathbf{Cov}\{x_j, x_k\}. \tag{5.14}$$

Так как случайные векторы $x_1, \ldots, x_n$ – независимы в совокупности и

$$\mathbf{Cov}\{x_j, x_j\} = \Sigma, \; j = 1, \ldots, n,$$

то

$$\mathbf{Cov}\{x_j, x_k\} = \delta_{jk}\Sigma = \begin{cases} \Sigma, & j = k, \\ 0, & j \neq k, \end{cases} \; j, k = 1, \ldots, n. \tag{5.15}$$

Подставив (5.15) в соотношение (5.14) получаем, учитывая (5.9):

$$\mathbf{Cov}\{y_i, y_l\} = \Sigma \sum_{j=1}^{n} c_{ij} c_{lj} = \delta_{il}\Sigma = \begin{cases} \Sigma, & i = l, \\ 0, & i \neq l, \end{cases} \; i, l = 1, \ldots, n. \tag{5.16}$$

Таким образом, случайные векторы $y_1, \ldots, y_n$ имеют совместное нормальное распределение, некоррелированы, имеют нормальные распределения (5.11) и по теореме 4.4 независимы в совокупности.

Докажем равенство (5.13).

$$\sum_{i=1}^{n} y_i y_i^{\mathbf{T}} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_j \left(\sum_{k=1}^{n} c_{ik} x_k\right)^{\mathbf{T}} =$$

$$= \sum_{j,k=1}^{n} \sum_{i=1}^{n} c_{ij} c_{ik} x_j x_k^{\mathbf{T}} = \sum_{j,k=1}^{n} \delta_{jk} x_j x_k^{\mathbf{T}} = \sum_{i=1}^{n} x_i x_i^{\mathbf{T}}.$$

□

Определение 5.1. *Говорят, что случайная* $(N \times N)$*-матрица* A *имеет* распределение Уишарта $W_N(\Sigma, m)$*, если она распределена как матрица*

$$\mathcal{A} = \sum_{j=1}^{m} y_j y_j^{\mathbf{T}},$$

где случайные N*-векторы* $y_1, \ldots, y_n$ *независимы в совокупности и одинаково распределены по невырожденному нормальному закону* $\mathcal{N}_N(\mathbf{0}_N, \Sigma)$ $(|\Sigma| \neq 0)$*.*

Распределение Уишарта обладает следующими свойствами:

Свойство 1. Если случайные матрицы $A_1, \ldots, A_K$ независимы в совокупности и имеют распределения Уишарта:

$$\mathcal{L}\{A_k\} = W_N(\Sigma, m_k),\ k = 1, \ldots, K,$$

то их сумма также имеет распределение Уишарта:

$$\mathcal{L}\left\{\sum_{k=1}^{K} A_k\right\} = W_N(\Sigma, \sum_{k=1}^{K} m_k);$$

Свойство 2. Если случайная матрица A имеет распределение Уишарта:

$$\mathcal{L}\{A\} = W_N(\Sigma, m),$$

то $\mathbf{E}\{A\} = m\Sigma$, и

$$\mathcal{L}\{cA\} = W_N(c\Sigma, m),$$

где $c > 0$ – некоторая постоянная величина.

Теорема 5.2 (О распределении оценок максимального правдоподобия параметров многомерного нормального распределения)**.** *В условиях теоремы 5.1 МП-оценки*

$$\hat{\mu} = \overline{x} \quad и \quad \hat{\Sigma} = \frac{1}{n} A,$$

определенные в (5.1) *и* (5.2)*, независимы и имеют следующие распределения вероятностей:*

$$\mathcal{L}\{\overline{x}\} = \mathcal{N}_N\left(\mu, \frac{1}{n}\Sigma\right), \tag{5.17}$$

$$\mathcal{L}\{n\hat{\Sigma}\} = \mathcal{L}\{A\} = W_N(\Sigma, n-1). \tag{5.18}$$

Доказательство. Введем в рассмотрение ортогональную $(n \times n)$-матрицу C:

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \ldots & c_{1,n} \\ c_{2,1} & c_{2,2} & \ldots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n-1,1} & c_{n-1,2} & \ldots & c_{n-1,n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \ldots & \frac{1}{\sqrt{n}} \end{pmatrix} \tag{5.19}$$

и рассмотрим ортогональные преобразования $y_1, \ldots, y_n$ элементов случайной выборки $X = \{x_1, \ldots, x_n\}$:

$$y_i = \sum_{j=1}^{n} c_{ij} \cdot x_j \; i = 1, \ldots, n.$$

Выразим оценки $\overline{x}$ и $\hat{\Sigma}$ через полученные случайные векторы $y_1, \ldots, y_n$.

Согласно лемме 5.1 случайные векторы $y_1, \ldots, y_n$ имеют совместное нормальное распределение, некоррелированы, имеют следующие нормальные распределения:

$$\mathcal{L}\{y_i\} = \mathcal{N}_N(\nu_i, \Sigma), \; i = 1, \ldots, n, \tag{5.20}$$

где, с учетом (5.19), ν_i определяются следующим образом:

$$\nu_i = \sum_{j=1}^{n} c_{ij}\mu = \mu \sum_{j=1}^{n} c_{ij} \frac{1}{\sqrt{n}} \sqrt{n} = \sqrt{n}\mu \sum_{j=1}^{n} c_{ij}c_{jn} = \begin{cases} \sqrt{n}\mu, & i = n, \\ 0, & i \neq n, \end{cases}, \; i = 1, \ldots, n. \tag{5.21}$$

Таким образом,

$$\begin{aligned} \mathcal{L}\{y_i\} &= \mathcal{N}_N(\mathbf{0}_N, \Sigma), \; i = 1, \ldots, n-1, \\ \mathcal{L}\{y_n\} &= \mathcal{N}_N(\sqrt{n}\mu, \Sigma). \end{aligned} \tag{5.22}$$

Из (5.21) следует:

$$y_n = \sum_{j=1}^{n} c_{nj}x_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j = \sqrt{n}\overline{x} \tag{5.23}$$

и

$$\overline{x} = \frac{1}{\sqrt{n}} y_n. \tag{5.24}$$

Из последнего равенства следует, что $\overline{x}$ – линейное преобразование случайного вектора y_n, имеющего невырожденное нормальное распределение $\mathcal{N}_N(\sqrt{n}\mu, \Sigma)$. Следовательно,

$$\mathcal{L}\{\overline{x}\} = \mathcal{N}_N(\mu, \frac{1}{\sqrt{n}}\Sigma). \tag{5.25}$$

Определим закон распределения $\hat{\Sigma}$.

Так как

$$\hat{\Sigma} = \frac{1}{n}A = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^{\mathbf{T}} = \frac{1}{n} \left(\sum_{i=1}^{n} x_i x_i^{\mathbf{T}} - n\overline{x}\,\overline{x}^{\mathbf{T}} \right),$$

то с учетом равенства (5.13), получим:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n-1} y_i y_i^{\mathbf{T}}. \tag{5.26}$$

Согласно (5.23) и определению распределения Уишарта:

$$\mathcal{L}\{n\hat{\Sigma}\} = \mathcal{L}\{A\} = W_N(\Sigma, n-1).$$

Независимость статистических оценок $\overline{x}$ и $\hat{\Sigma}$ непосредственно следует из представлений (5.24) и (5.26), согласно которым $\overline{x}$ и $\hat{\Sigma}$ не содержат общих случайных векторов из $y_1, \ldots, y_n$, так как $\overline{x}$ определяется y_n, а $\hat{\Sigma}$ строится на основе $y_1, \ldots, y_{n-1}$.

□

5.3 Вероятностные свойства выборочного среднего и выборочной ковариационной матрицы. Несмещенная выборочная ковариационная матрица

Теорема 5.3. *В условиях теоремы 5.1 выборочное среднее $\overline{x}$ является несмещенной оценкой для математического ожидания μ:*

$$\mathbf{E}\{\overline{x}\} = \mu,$$

а выборочная ковариационная матрица $\hat{\Sigma}$ – асимптотически несмещенной:

$$\mathbf{E}\{\hat{\Sigma}\} \xrightarrow[n\to\infty]{} \Sigma.$$

Несмещенной оценкой ковариационной матрицы является стаистика

$$S = \frac{1}{n-1}A, \tag{5.27}$$

где

$$A = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathbf{T}}.$$

Все оценки $\overline{x}$, $\hat{\Sigma}$ и S сильно состоятельны:

$$\overline{x} \xrightarrow[n\to\infty]{\mathbf{P}=1} \mu,\ \hat{\Sigma} \xrightarrow[n\to\infty]{\mathbf{P}=1} \Sigma,\ S \xrightarrow[n\to\infty]{\mathbf{P}=1} \Sigma, \quad n \to +\infty,$$

Доказательство. Несмещенность выборочного среднего $\overline{x}$ следует из рассуждений:

$$\mathbf{E}\{\overline{x}\} = \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\{x_i\} = \mu.$$

Вычислим $\mathbf{E}\left\{\hat{\Sigma}\right\}$ используя равенство (5.26):

$$\mathbf{E}\left\{\hat{\Sigma}\right\} = \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n-1} y_i y_i^{\mathbf{T}}\right\} = \frac{1}{n}\sum_{i=1}^{n-1}\mathbf{E}\left\{y_i y_i^{\mathbf{T}}\right\} = \frac{n-1}{n}\Sigma. \tag{5.28}$$

Так как

$$\mathbf{E}\left\{\hat{\Sigma}\right\} \neq \Sigma,$$

то $\hat{\Sigma}$ – смещенная оценка для Σ; смещение этой оценки – $b = \frac{1}{n}\Sigma$.

Сильная состоятельность оценок $\overline{x}$, $\hat{\Sigma}$ и S вытекает из усиленного закона больших чисел для независимых в совокупности одинаково распределенных случайных векторов [11].

Действительно, так как $x_1, \ldots, x_n$ – независимы в совокупности и одинаково распределены с математическим ожиданием:

$$\mathbf{E}\{x_k\} = \mu,\ k = 1, \ldots, n,$$

то

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow[n\to\infty]{\mathbf{P}=1} \mu,$$

а $y_1 y_1^{\mathbf{T}}, \ldots, y_{n-1} y_{n-1}^{\mathbf{T}}$ – независимые в совокупности, одинаково распределенные случайные матрицы с математическим ожиданием

$$\mathbf{E}\{y_i y_i^{\mathbf{T}}\} = \Sigma,$$

то

$$\hat{\Sigma} = \frac{n-1}{n} \frac{1}{n-1} \sum_{i=1}^{n-1} y_i y_i^{\mathbf{T}} \xrightarrow[n\to\infty]{\mathbf{P}=1} \Sigma,$$

и

$$S = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i y_i^{\mathbf{T}} \xrightarrow[n\to\infty]{\mathbf{P}=1} \Sigma.$$

□

Теорема 5.4. *Элементы выборочной ковариационной матрицы $\hat{\Sigma}$ и несмещенной выборочной ковариационной матрицы S*

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{1,1} & \hat{\sigma}_{1,2} & \ldots & \hat{\sigma}_{1,N} \\ \hat{\sigma}_{2,1} & \hat{\sigma}_{2,2} & \ldots & \hat{\sigma}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{N,1} & \hat{\sigma}_{N,2} & \ldots & \hat{\sigma}_{N,N} \end{pmatrix} \quad \textit{и} \quad S = \begin{pmatrix} s_{1,1} & s_{1,2} & \ldots & s_{1,N} \\ s_{2,1} & s_{2,2} & \ldots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \ldots & s_{N,N} \end{pmatrix} \tag{5.29}$$

имеют совместные асимптотически нормальные распределения с ковариациями, определяемыми элементами истинной ковариационной матрицы Σ:

$$\lim_{n\to+\infty} n\mathbf{Cov}\{\hat{\sigma}_{kl}, \hat{\sigma}_{rq}\} = \lim_{n\to+\infty} n\mathbf{Cov}\{s_{kl}, s_{rq}\} = \sigma_{kr}\sigma_{lq} + \sigma_{kq}\sigma_{lr},\ k,l,r,q = 1,\ldots,N. \tag{5.30}$$

Доказательство. Доказательство теоремы основано на использовании центральной предельной теоремы для одинаково распределенных случайных векторов [1], согласно которой, если $v_1, \ldots, v_m, v_{m+1}, \ldots \in \mathbb{R}^p$ – последовательность независимых в совокупности, одинаково распределенных случайных p-мерных векторов с математическим ожиданием μ_v и ковариационной матрицей Σ_v:

$$\mu_v = \mathbf{E}\{v_i\},\ \Sigma_v = \mathbf{Cov}\{v_i, v_i\},\ i = 1,\ldots,m,m+1,\ldots,$$

то

$$\mathcal{L}\left\{\frac{1}{\sqrt{m}} \sum_{i=1}^{m} (v_i - \mu_v)\right\} \xrightarrow[m\to\infty]{} \mathcal{N}_p(\mathbf{0}_p, \Sigma_v).$$

Согласно (5.26):

$$n\hat{\Sigma} = (n-1)S = A = \sum_{i=1}^{n-1} y_i y_i^{\mathbf{T}}, \tag{5.31}$$

где случайные N-векторы $y_i = (y_{i1}, \ldots, y_{iN})$, независимы в совокупности и, согласно (5.22), распределены по закону $\mathcal{N}_N(\mathbf{0}_N, \Sigma)$.

Из (5.31) следует

$$\mathbf{E}\left\{\frac{1}{n-1} a_{kl}\right\} = \sigma_{kl},\ k,l = 1,\ldots,N. \tag{5.32}$$

Введем в рассмотрение нормированные элементы a°_{kl} матрицы A:

$$a^{\circ}_{kl} = \frac{1}{\sqrt{n-1}} \left(a_{kl} - (n-1)\sigma_{kl}\right),\ k,l = 1,\ldots,N, \tag{5.33}$$

из которых, учитывая, что по построению $a^{\circ}_{kl} = a^{\circ}_{lk},\ k, l = 1, \ldots, N$, составим $\frac{N(N+1)}{2}$-вектор

$$a^{\circ} = (a^{\circ}_{11}, a^{\circ}_{21}, a^{\circ}_{22}, a^{\circ}_{31}, a^{\circ}_{32}, a^{\circ}_{33}, a^{\circ}_{41}, \ldots, a^{\circ}_{NN})^{\mathbf{T}}$$

который имеет асимптотически нормальное $(n \to +\infty)$ $\frac{N(N+1)}{2}$-мерное распределение с нулевыми математическими ожиданиями и ковариациями:

$$\mathbf{Cov}\{a^{\circ}_{kl}, a^{\circ}_{rq}\} = \mathbf{Cov}\left\{\frac{1}{\sqrt{n-1}}(a_{kl} - (n-1)\sigma_{kl}), \frac{1}{\sqrt{n-1}}(a_{rq} - (n-1)\sigma_{rq})\right\} =$$

$$= \mathbf{Cov}\{y_{ik}y_{il}, y_{ir}y_{iq}\} = \mathbf{E}\{y_{ik}y_{il}y_{ir}y_{iq}\} - \mathbf{E}\{y_{ik}y_{il}\}\mathbf{E}\{y_{ir}y_{iq}\} =$$

$$= \sigma_{kl}\sigma_{rq} + \sigma_{kr}\sigma_{lq} + \sigma_{kq}\sigma_{lr} - \sigma_{kl}\sigma_{rq} = \sigma_{kr}\sigma_{lq} + \sigma_{kq}\sigma_{lr}, \quad k, l, r, q = 1, \ldots, N,$$

где использована известная формула (4.4) для моментов четвертого порядка многомерного нормального распределения.

Очевидное соотношение:

$$\lim_{n\to+\infty} n\mathbf{Cov}\{\hat{\sigma}_{kl}, \hat{\sigma}_{rq}\} = \lim_{n\to+\infty} n\mathbf{Cov}\{s_{kl}, s_{rq}\} =$$

$$= \lim_{n\to+\infty} \mathbf{Cov}\left\{\frac{1}{\sqrt{n-1}}(a_{kl} - (n-1)\sigma_{kl}), \frac{1}{\sqrt{n-1}}(a_{rq} - (n-1)\sigma_{rq})\right\},\ k, l, r, q = 1, \ldots, N,$$

завершает доказательство. □

6 Корреляционный анализ (статистическое исследование зависимостей)

6.1 Выборочный коэффициент корреляции и его свойства

6.1.1 Определение и основные свойства выборочного коэффициента корреляции

Пусть в пространстве из N, $(N \geq 1)$ признаков наблюдается выборка $X = \{x_1, \ldots, x_n\}$ объема n, образованная наблюдениями с вектором математического ожидания μ и ковариационной $(N \times N)$-матрицей Σ:

$$\mu = \mathbf{E}\{x_i\} \in \mathbb{R}^N \text{ и } \Sigma = \mathbf{Cov}\{x_i, x_i\},\ i = 1, \ldots, n.$$

Элементы σ_{kl}, $k, l = 1, \ldots, N$, ковариационной матрицы Σ, описывают *парные зависимости* признаков между собой. Однако на практике для исследования парной зависимости признаков чаще используется коэффициент корреляции, называемый еще *парной корреляцией*:

$$\rho_{kl} = \frac{\sigma_{kl}}{\sqrt{\sigma_{kk}\sigma_{ll}}},\ k, l = 1, \ldots, N, \tag{6.1}$$

В тех случаях, когда вектор математического ожидания μ и ковариационная матрица Σ неизвестны, для исследования зависимостей используют статистическую оценку парного коэффициента корреляции, вычисленную по выборке X объема n – *выборочный коэффициент корреляции.*

Определение 6.1. Выборочным коэффициентом корреляции $\hat{\rho}_{kl}$ *между компонентами* $\tilde{x}_k$ *и* $\tilde{x}_l$ *случайного* N*-мерного вектора* $x = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^N$*, вычисленным по случайной выборке* $X = \{x_1, \ldots, x_n\}$ *объема* n*, называется статистика*

$$\hat{\rho}_{kl} = r_{kl} = \frac{\hat{\sigma}_{kl}}{\sqrt{\hat{\sigma}_{kk}\hat{\sigma}_{ll}}},\ k, l = 1, \ldots, N, \tag{6.2}$$

где $\hat{\sigma_{kl}}$, $\hat{\sigma_{kk}}$, $\hat{\sigma_{ll}}$ *– элементы выборочной ковариационной матрицы* $\hat{\Sigma}$, $k, l = 1, \ldots, N$.

Так как

$$\hat{\sigma}_{kl} = \frac{1}{n} a_{kl},\ k, l = 1, \ldots, N,$$

где статистики a_{kl}, определены равенствами

$$a_{kl} = \sum_{i=1}^{n} (x_{ik} - \overline{x}_k)(x_{il} - \overline{x}_l),\ k, l = 1, \ldots, N, \tag{6.3}$$

и являются элементами матрицы A:

$$A = \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^{\mathbf{T}},$$

то эквивалентным соотношению (6.2) является равенство

$$\hat{\rho}_{kl} = r_{kl} = \frac{a_{kl}}{\sqrt{a_{kk}a_{ll}}},\ k, l = 1, \ldots, N. \tag{6.4}$$

Свойства выборочного коэффициента корреляции $\hat{\rho}_{kl}$

Свойство 1. $-1 \leq \hat{\rho}_{kl} \leq 1$.

Свойство 2. $\hat{\rho}_{kl}$ инвариантен относительно масштабного преобразования признаков в выборке, то есть не изменяют своих значений при умножении признаков в наблюдениях $x_i = (x_{i1}, \ldots, x_{iN})^{\mathbf{T}}$ на соответствующие константы: $x_i := (b_1 x_{i1}, \ldots, b_N x_{in})^{\mathbf{T}}$, $i = 1, \ldots, n$.

Свойство 3. Вычисленный по случайной выборке X объема n выборочный коэффициент корреляции $\hat{\rho}_{kl}$ – строго состоятельная оценка парного коэффициента корреляции ρ_{kl}:

$$\hat{\rho}_{kl} \xrightarrow[n\to\infty]{\mathbf{P}=1} \rho_{kl}. \tag{6.5}$$

Предположим теперь, что наблюдения из выборки $X = \{x_1, \ldots, x_n\}$ независимы в совокупности и имеют невырожденное многомерное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Исследуем дополнительные свойства выборочного коэффициента корреляции (6.2), который в этом случае является оценкой максимального правдоподобия для коэффициента корреляции (6.1) (поскольку (6.2) – функциональное преобразование выборочной ковариационной матрицы, являющейся оценкой максимального правдоподобия).

Теорема 6.1 (О распределении выборочного коэффициента корреляции)**.** *[1] Пусть $\tilde{x}_k$ и $\tilde{x}_l$ – компоненты случайного вектора x, распределенного по закону $\mathcal{N}_N(\mu, \Sigma)$, причем $\rho_{ij} = \mathbf{Corr}\{x_i, x_j\} = 0$. Тогда выборочный коэффициент корреляции, вычисленный по выборке $X = \{x_1, \ldots, x_n\}$, имеет плотность распределения вероятностей $p_0(r)$ вида*

$$p_0(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1 - r^2)^{\frac{n-4}{2}}, \tag{6.6}$$

где

$$\Gamma(a) = \int_0^\infty e^{-u} u^{a-1} du,\ \Gamma(0,5) = \sqrt{\pi},\ |r| < 1.$$

Очевидно, что функция $p_0(r)$ четная, то есть

$$p_0(r) = p_0(-r). \tag{6.7}$$

Из (6.7) следует, что для функции распределения вероятностей $F_0(r)$:

$$F_0(r) = \int_{-\infty}^{r} p_0(u)\, du, \tag{6.8}$$

будет выполняться равенство

$$F_0(r) = 1 - F_0(-r). \tag{6.9}$$

6.1.2 Проверка гипотезы о значимости коэффициента корреляции

Пусть имеется выборка $X = \{x_1, \ldots, x_n\}$ из распределения вероятностей $\mathcal{N}_N(\mu, \Sigma)$. Необходимо призаданном уровне значимости α проверить гипотезу о равенстве нулю коэффициента корреляции ρ_{ij}:

$$\begin{array}{lcl} H_0 & : & \rho_{ij} = 0; \\ H_1 = \overline{H_0} & : & \rho_{ij} \neq 0. \end{array} \tag{6.10}$$

Решение задачи (6.10) формулируется при помощи статистики r_{ij} следующим образом:

$$\begin{array}{lcl} H_0 & : & |r_{ij}| < \Delta; \\ H_1 = \overline{H_0} & : & |r_{ij}| \geq \Delta, \end{array} \tag{6.11}$$

где Δ – порог критерия, который определяется следующим образом:

$$\begin{aligned}\alpha = \mathbf{P}\{H_1 \,|\, H_0\} = \mathbf{P}\{|r_{ij}| \geq \Delta \,|\, H_0\} = \mathbf{P}\{r_{ij} \leq -\Delta \,|\, H_0\} + \mathbf{P}\{r_{ij} \geq \Delta \,|\, H_0\} = \\ = \mathbf{P}\{r_{ij} \leq -\Delta \,|\, H_0\} + 1 - \mathbf{P}\{r_{ij} < \Delta \,|\, H_0\} = F_0(-\Delta) + 1 - F_0(\Delta),\end{aligned} \tag{6.12}$$

где $F_0(\Delta)$ определена в (6.8). Учитывая равенство (6.9) можно записать

$$\alpha = 2 \cdot (1 - F_0(\Delta)),$$

из которого следует, что порог критерия (6.12) определяется равенством

$$\Delta = F_0^{-1}\left(1 - \frac{\alpha}{2}\right).$$

6.1.3 Z-преобразование и Z-статистика Фишера

Теорема 6.2 (Об асимптотической нормальности выборочного коэффициента корреляции)**.** *[12] В условиях теоремы 5.1 и при условии, что соответствующее истинное абсолютное значение коэффициента корреляции ρ_{ij} не равно единице:*

$$|\rho_{ij}| \neq 1,$$

выборочный коэффициент корреляции r_{ij} из (6.2) *имеет асимптотически нормальное распределение:*

$$\mathcal{L}\{\sqrt{n-1}(r_{ij} - \rho_{ij})\} \xrightarrow[n\to\infty]{} \mathcal{N}_1(0, (1-\rho_{ij}^2)^2), \,; i, j = 1, \ldots, N. \tag{6.13}$$

Теорема 6.3 (Об асимптотической нормальности функционально преобразованной последовательности)**.** *[1] Пусть $u_1, \ldots, u_m$ – асимптотически нормальная сходящаяся по вероятности случайная последовательность:*

$$u_m \xrightarrow[m\to\infty]{\mathbf{P}} b,$$

$$\mathcal{L}\{\sqrt{m}(u_m - b)\} \xrightarrow[m\to\infty]{} \mathcal{N}_1(0, \sigma^2),$$

где $b \in \mathbb{R}$, а $\sigma^2 > 0$.

Если функция

$$f = f(z) : \mathbb{R}^1 \to \mathbb{R}^1$$

имеет в окрестности точки $z = b$ первую и вторую производные и

$$\phi_b = f'(z)|_{z=b} \neq 0, \tag{6.14}$$

то

$$\mathcal{L}\{\sqrt{m}(f(u_m) - f(b))\} \xrightarrow[m\to\infty]{} \mathcal{N}_1(0, \beta), \tag{6.15}$$

где

$$\beta = \sigma^2 \cdot \phi_b^2. \tag{6.16}$$

Доказательство. Доказательство основано на разложении функции $f(z)$ в точке $z = b$ в ряд Тейлора с остаточным членом в форме Лагранжа и на вычислении соответствующих моментов. □

Следствие 6.1. *Пусть* $\Psi = \Psi(v) : \mathbb{R}^1 \to \mathbb{R}^1$ *– дважды дифференцируема в точке* $v = \rho_{ij}$, *и такая, что*

$$\frac{d}{dv}\Psi(v)|_{v=\rho_{ij}} \neq 0,$$

тогда в условиях теоремы 6.2 для выборочного коэффициента корреляции r_{ij} *выполняется* $(n \to +\infty)$*:*

$$\mathcal{L}\{\sqrt{n-1}(\Psi(r_{ij}) - \Psi(\rho_{ij}))\} \to \mathcal{N}_1\left(0, \left(\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{ij}}\right)^2 (1-\rho_{ij}^2)^2\right). \tag{6.17}$$

Доказательство. Доказательство непосредственно следует из результата теоремы 6.2 и теоремы 6.3. □

Замечание 6.1.1. *В соотношениях (6.13) и (6.17) вместо* $n-1$ *можно использовать* n *(поскольку* $n \to +\infty$*), однако считается, что* $n-1$ *«точнее».*

Выберем преобразование $\Psi = \Psi(v) : \mathbb{R}^1 \to \mathbb{R}^1$ так, чтобы дисперсия предельного нормального распределения в (6.17) была единичной:

$$\left(\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{ij}}\right)^2 (1-\rho_{ij}^2)^2 = 1.$$

Получим дифференциальное уравнение:

$$\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{ij}} = \frac{1}{1-\rho_{ij}^2} = \frac{1}{(1-\rho_{ij})(1+\rho_{ij})} = \frac{1}{2}\left(\frac{1}{1+\rho_{ij}} + \frac{1}{1-\rho_{ij}}\right),$$

решая которое, найдем

$$\Psi(v) = \frac{1}{2}\ln\frac{1+v}{1-v}.$$

Определение 6.2. *Пусть* r_{ij} *– выборочный коэффициент корреляции* (6.2)*, тогда статистика*

$$Z = Z(r_{ij}) = \Psi(r_{ij}) = \frac{1}{2}\ln\frac{1+r_{ij}}{1-r_{ij}} \tag{6.18}$$

называется Z*-статистикой Фишера, а преобразование* $\Psi(\cdot)$ *–* Z*-преобразованием Фишера.*

Следствие 6.2. *В условиях теоремы 6.2 случайная величина* $\sqrt{n-1}(Z(r_{ij}) - Z(\rho_{ij}))$ *имеет при* $n \to +\infty$ *стандартное нормальное распределение* $\mathcal{N}_1(0,1)$.

6.1.4 Проверка гипотезы о значении коэффициента корреляции

Z-статистика Фишера (6.18) имеет большое прикладное значение – с ее помощью проверяются гипотезы о значении коэффициента корреляции:

$$\begin{array}{lcl} H_0 & : & \rho_{ij} = \rho_{ij}^o; \\ H_1 = \overline{H_0} & : & \rho_{ij} \neq \rho_{ij}^o, \end{array} \tag{6.19}$$

где ρ_{ij}^o – предполагаемое значение коэффициента корреляции. Очевидно, что задача (6.19) эквивалентна следующей задаче

$$\begin{array}{lcl} H_0 & : & \sqrt{n-1}(\rho_{ij} - \rho_{ij}^o) = 0; \\ H_1 = \overline{H_0} & : & \sqrt{n-1}(\rho_{ij} - \rho_{ij}^o) \neq 0. \end{array} \tag{6.20}$$

Учитывая следствие 6.2, а именно тот факт, что при $n \to +\infty$ и верной гипотезе H_0

$$\mathcal{L}\{\sqrt{n-1}(Z(r_{ij}) - Z(\rho_{ij}))\} = \mathcal{N}_1(0,1),$$

получим при следующий критерий для проверки гипотез H_0, H_1:

$$\begin{cases} H_0 & : \ |\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij}))| \le \Delta; \\ H_1 = \overline{H_0} & : \ |\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij}))| > \Delta, \end{cases} \tag{6.21}$$

где порог критерия Δ определяется по наперед заданному малому значению уровня значимости α (вероятность принять гипотезу H_1 при условии, что верна гипотеза H_0):

$$\alpha = \mathbf{P}\{H_1|H_0\} = 1 - \mathbf{P}\{H_0|H_0\} \in (0,1),$$

$$\begin{aligned} \alpha &= \mathbf{P}\{H_1 \,|\, H_0\} = \mathbf{P}\{|\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij}))| > \Delta \,|\, H_0\} = \\ &= \mathbf{P}\{\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij})) \le -\Delta \,|\, H_0\} + \mathbf{P}\{\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij})) \ge \Delta \,|\, H_0\} = \\ &= \mathbf{P}\{\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij})) \le -\Delta \,|\, H_0\} + 1 - \mathbf{P}\{\sqrt{n-1}(Z(r_{ij}) - Z(\rho^o_{ij})) < \Delta \,|\, H_0\} = \\ &= \Phi(-\Delta) + 1 - \Phi(\Delta) = 2(1 - \Phi(\Delta)), \end{aligned} \tag{6.22}$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона распределения. Таким образом:

$$\Delta = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{6.23}$$

то есть Δ – квантиль уровня $1 - \dfrac{\alpha}{2}$ стандартного нормального закона $\mathcal{N}_1(0,1)$.

Замечание 6.2.1. *Построенный критерий (6.21), (6.23) позволяет также определить доверительный интервал [12] уровня* $1-\alpha$ *для* ρ_{ij}*:*

$$\mathbf{P}\left\{\mathrm{th}\left(Z(r_{ij}) - \frac{\Delta}{\sqrt{n-1}}\right) < \rho_{ij} < \mathrm{th}\left(Z(r_{ij}) + \frac{\Delta}{\sqrt{n-1}}\right)\right\} = 1 - \alpha,$$

где $\Psi^{-1}(v)$*– преобразование, обратное к преобразованию Фишера* $\Psi(v)$*:*

$$\Psi^{-1}(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}} = \mathrm{th}(v),$$

$$\Psi(v) = \frac{1}{2}\ln\frac{1+v}{1-v}.$$

Замечание 6.2.2. *На практике чаще всего проверяются гипотезы:*

$$\begin{array}{lll} H_0 & : & \rho_{ij} = 0; \\ H_1 = \overline{H_0} & : & \rho_{ij} \ne 0. \end{array}$$

Если принимается H_0*, то* i*-й и* j*-й признаки* ($k \ne l = \overline{1,N}$) *считаются некоррелированными, а в силу предположения их совместной нормальности – и независимыми. Поэтому гипотезу* H_0 *иногда называют гипотезой независимости. Проверяются гипотезы* H_0, H_1 *при помощи критерия (6.21), (6.23) (*$\rho^o_{ij} = 0$*, и* $Z(\rho^o_{ij}) = Z(0) = 0$*).*

В случае выборки «малого объема» (асимптотика $n \to +\infty$ *не предполагается) для проверки гипотез* H_0, H_1 *считается целесообразным использование другого критерия:*

$$\begin{cases} H_0 & : \ \sqrt{n-2}\frac{|r_{ij}|}{\sqrt{1-r_{ij}}} \le \Delta; \\ H_1 = \overline{H_0} & : \ \sqrt{n-2}\frac{|r_{ij}|}{\sqrt{1-r_{ij}}} > \Delta, \end{cases}$$

где относительно статистики критерия в пункте 6.2 (теорема 6.5, соотношение (6.29)) будет установлено, что при истинной гипотезе H_0 она является модулем случайной величины, имеющей t-распределение Стьюдента с $n-2$ степенями свободы, и порог критерия $\Delta = F^{-1}_{t_{n-2}}(1-\frac{\alpha}{2})$ – соответствующая квантиль уровня $1-\frac{\alpha}{2}$ ($\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$). Однако данный критерий «более чувствителен» к отклонениям распределения вероятностей выборочных значений компонент от нормального закона, и при «достаточно большом» объеме выборки n лучше использовать критерий (6.21), (6.23), основанный на Z-статистике Фишера (6.18).

6.2 Выборочный частный коэффициент корреляции и его свойства

6.2.1 Определение и свойства выборочного частного коэффициента корреляции

Пусть случайный N-мерный вектор признаков $x = (\tilde{x}_1, \ldots, \tilde{x}_N) \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$ с неизвестным математическим ожиданием μ и неизвестной ковариационной матрицей Σ, а $X = \{x_1, \ldots, x_n\}$ – случайная выборка объема n, $n > N$ наблюдений за x.

Оценим по выборке X частный коэффициент корреляции (4.29). Согласно разбиению случайного вектора

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix},$$

на подвекторы $x^{(1)} \in \mathbb{R}^m$ и $x^{(2)} \in \mathbb{R}^{N-m}$ $(m < N)$, где

$$x^{(1)} = (\tilde{x}_1, \ldots, \tilde{x}_m)^{\mathbf{T}} \in \mathbb{R}^m, \quad x^{(2)} = (\tilde{x}_{m+1}, \ldots, \tilde{x}_N)^{\mathbf{T}} \in \mathbb{R}^{N-m},$$

порождающего разбиения N-вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ – на соответствующие блоки:

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \ \Sigma_{21} = \Sigma_{12}^{\mathbf{T}}. \tag{6.24}$$

построим соответствующие разбиения их статистических оценок – арифметического среднего $\overline{x}$ и выборочной ковариационной матрицы $\hat{\Sigma}$:

$$\overline{x} = \begin{pmatrix} \overline{x}^{(1)} \\ \overline{x}^{(2)} \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \ A_{21} = A_{12}^{\mathbf{T}}, \tag{6.25}$$

где $(N \times N)$-матрица A определена в (5.3).

Введем в рассмотрение $(m \times m)$-матрицу $A_{11|2}$:

$$A_{11|2} = A_{11} - A_{12}A_{22}^{-1}A_{12}^{\mathbf{T}},$$

и обозначим (k,l)-й элемент $A_{11|2}$ через $a_{kl|m+1,\ldots,N}$:

$$a_{kl|m+1,\ldots,N} = a_{kl} - a_{(k)}^{\mathbf{T}}A_{22}^{-1}a_{(l)}, \ k,l = 1,\ldots,m,$$

где $a_{(k)}^{\mathbf{T}}$ – k-я строка матрицы A_{12}, $k = 1,\ldots,m$.

Теорема 6.4. *Пусть элементы случайной выборки $X = \{x_1, \ldots, x_n\}$ объема $n > N$ независимы в совокупности и имеют невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$, тогда оценкой максимального правдоподобия для частного коэффициента корреляции* (4.29) *является статистика:*

$$r_{kl|m+1,\ldots,N} = \frac{a_{kl|m+1,\ldots,N}}{\sqrt{a_{kk|m+1,\ldots,N} a_{ll|m+1,\ldots,N}}},\ k, l = 1, \ldots, m. \tag{6.26}$$

Доказательство. Доказательство очевидно и основано на подстановке в (4.29) с учетом обозначений (6.25) вместо неизвестной ковариационной матрицы Σ ее МП-оценки

$$\hat{\Sigma} = \frac{1}{n} A$$

по выборке X, полученной в теореме 5.1, условия которой здесь выполняются. □

Определение 6.3. Выборочным частным коффициентом корреляции *между k-й и l-й компонентами N-мерного случайного вектора $x \in \mathbb{R}^N$ при условии, что компоненты $\tilde{x}_{m+1}, \ldots, \tilde{x}_N$ фиксированы, вычисленным по случайной выборке $X = \{x_1, \ldots, x_n\}$ объема $n > N$ называется статистика*

$$\hat{\rho}_{kl|m+1,\ldots,N} = r_{kl|m+1,\ldots,N} = \frac{a_{kl|m+1,\ldots,N}}{\sqrt{a_{kk|m+1,\ldots,N} a_{ll|m+1,\ldots,N}}},\ k, l = 1, \ldots, m. \tag{6.27}$$

Свойства выборочного частного коэффициента корреляции

Свойство 1. $-1 \leq \hat{\rho}_{kl|m+1,\ldots,N} \leq 1$.

Свойство 2. Вычисленный по случайной выборке X объема n выборочный частный коэффициент корреляции $\hat{\rho}_{kl|m+1,\ldots,N}$ – состоятельная оценка парного коэффициента кореляции $\rho_{kl|m+1,\ldots,N}$:

$$\hat{\rho}_{kl|m+1,\ldots,N} \xrightarrow[n\to\infty]{\mathbf{P}} \rho_{kl|m+1,\ldots,N}. \tag{6.28}$$

Найдем распределение вероятностей случайной матрицы

$$A_{11|2} = A_{11} - A_{12} A_{22}^{-1} A_{12}^{\mathbf{T}},$$

элементы которой служат для определения выборочного частного (6.27) коэффициента корреляции.

Теорема 6.5. *В условиях теоремы 6.4 справедливы следующие утверждения:*

1) *случайная матрица $A_{11|2}$:*

$$A_{11|2} = A_{11} - A_{12} A_{22}^{-1} A_{12}^{\mathbf{T}}$$

имеет распределение Уишарта

$$\mathcal{L}\{A_{11|2}\} = W_m(\Sigma_{11|2}, n - (N - m) - 1)$$

и не зависит от матрицы $A_{12} A_{22}^{-1} A_{12}^{\mathbf{T}}$, которая при $\Sigma_{12} = \mathbf{0}_{m\times(N-m)}$ также имеет распределение Уишарта:

$$\mathcal{L}\{A_{12} A_{22}^{-1} A_{12}^{\mathbf{T}} | \Sigma_{12} = \mathbf{0}_{m\times(N-m)}\} = W_m(\Sigma_{11|2}, N - m).$$

2) *выборочный частный коэффициент корреляции $r_{kl|m+1,\ldots,N}$ $(k, l = \overline{1, m})$, вычисленный по выборке объема n, распределен так же, как соответствующий ему обычный выборочный коэффициент корреляции r_{kl}, подсчитанный по выборке объема $n - (N - m)$ с истинным значением коэффициента корреляции $\rho_{kl} := \rho_{kl|m+1,\ldots,N}$, $k, l = \overline{1, m}$.*

3) *для выборочного коэффициента корреляции* r_{kl} $(k,l=\overline{1,N})$, *определенного по выборке объема* n, *при нулевом истинном значении:* $\rho_{kl}=0$, *статистика*

$$\sqrt{n-2}\frac{|r_{kl}|}{\sqrt{1-r_{kl}^2}} \tag{6.29}$$

распределена как модуль случайной величины, имеющей t*-распределение Стьюдента с* $n-2$ *степенями свободы.*

6.2.2 Проверка гипотезы о значении частного коэффициента корреляции

Второе утверждение теоремы 6.5 позволяет использовать для проверки гипотез о значении частного коэффициента корреляции

$$\begin{array}{lcl} H_0 & : & \rho_{kl|m+1,\ldots,N}=\rho^o_{kl|m+1,\ldots,N}; \\ H_1=\overline{H_0} & : & \rho_{kl|m+1,\ldots,N}\neq\rho^o_{kl|m+1,\ldots,N}, \end{array} \tag{6.30}$$

где $\rho^o_{kl|m+1,\ldots,N}$ – предполагаемое значение частного коэффициента корреляции, все критерии из параграфа 6.1, заменив в них n на $n-(N-m)$.

Гипотеза независимости:

$$\rho^o_{kl|m+1,\ldots,N}=0,$$

в данном случае называется *гипотезой условной независимости*, $k,l=1,\ldots,m$.

6.3 Выборочный множественный коэффициент корреляции и его свойства

Определение 6.4. Выборочным множественным коффициентом корреляции *между между компонентой* $\tilde{x}_k$ *и компонентами подвектора* $x^{(2)}=(\tilde{x}_{m+1},\ldots,\tilde{x}_N)^{\mathbf{T}}$, *вычисленным по случайной выборке* $X=\{x_1,\ldots,x_n\}$ *объема* $n>N$ *назавается статистика*

$$R_{k,m+1,\ldots,N}=\sqrt{\frac{a^{\mathbf{T}}_{(k)}A^{-1}_{22}a_{(k)}}{a_{kk}}}\ k=1,\ldots,m, \tag{6.31}$$

где $a^{\mathbf{T}}_{(k)}$ – k*-я строка матрицы* A_{12}, $k=1,\ldots,m$, *являющейся блоком разбиения определенной в (5.3)* $(N\times N)$*-матрицы* A*:*

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},\ A_{21}=A^{\mathbf{T}}_{12}.$$

Теорема 6.6. *В условиях теоремы 6.4 статистика*

$$\frac{n-(N-m)-1}{N-m}\frac{R^2_{k,m+1,\ldots,N}}{1-R^2_{k,m+1,\ldots,N}}$$

при нулевом истинном значении соответствующего множественного коэффициента корреляции:

$$\overline{R}_{k,m+1,\ldots,N}=0,$$

имеет F*-распределение Фишера с* $N-m$ *и* $n-(N-m)-1$ *степенями свободы:*

$$\mathcal{L}\left\{\frac{n-(N-m)-1}{N-m}\frac{R^2_{k,m+1,\ldots,N}}{1-R^2_{k,m+1,\ldots,N}}\Big|\overline{R}_{k,m+1,\ldots,N}=0\right\}=F_{N-m,n-(N-m)-1},\ k=1,\ldots,m. \tag{6.32}$$

Воспользуемся выборочным множественным коэффициентом корреляции (6.31) для проверки гипотез о некоррелированности k-го признака ($1 \le k \le m$) с группой из $N-m$ выбранных признаков:

$$\begin{array}{lcl} H_0 & : & \overline{R}_{k,m+1,\ldots,N} = 0; \\ H_1 = \overline{H_0} & : & \overline{R}_{k,m+1,\ldots,N} \neq 0. \end{array}$$

Гипотеза H_0 в предположении нормальности также называется *гипотезой независимости*.

Учтем, что статистика в (6.32) монотонно убывает с уменьшением $R^2_{k,m+1,\ldots,N}$ (приближением его значения к нулю), и построим следующий критерий для проверки гипотез H_0, H_1:

$$\left\{ \begin{array}{lcl} H_0 & : & \frac{n-(N-m)-1}{N-m} \frac{R^2_{k,m+1,\ldots,N}}{1-R^2_{k,m+1,\ldots,N}} \le \Delta; \\ H_1 = \overline{H_0} & : & \frac{n-(N-m)-1}{N-m} \frac{R^2_{k,m+1,\ldots,N}}{1-R^2_{k,m+1,\ldots,N}} > \Delta, \end{array} \right. \tag{6.33}$$

где порог критерия Δ определяется с учетом (6.32) по заданному малому значению уровня значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$:

$$\Delta = F^{-1}_{N-m,n-(N-m)-1}(1-\alpha) \tag{6.34}$$

квантиль уровня $1-\alpha$ от F-распределения с $N-m$ и $n-(N-m)-1$ степенями свободы.

Замечание 6.4.1. *Критерий (6.33), (6.34) в множественном регрессионном анализе называется проверкой регрессии на значимость. Если гипотеза H_0 отвергается, то регрессия k-го признака ($1 \le k \le m$) на выбранные $N-m$ признаков считается значимой, и возможно прогнозирование значения $\tilde{x}_k$ этого признака по $x^{(2)} \in \mathbb{R}^{N-m}$ при помощи подстановочной статистической оценки функции регрессии (4.17), (4.42):*

$$\hat{\tilde{x}}_k = \overline{x}_k + a^{\mathbf{T}}_{(k)} A^{-1}_{22} (x^{(2)} - \overline{x}^{(2)}). \tag{6.35}$$

6.4 Проверка общих гипотез о независимости

Пусть случайный N-мерный вектор $x \in \mathbb{R}^N$, имеющий невырожденное нормальное распределение $\mathcal{N}_N(\mu, \Sigma)$, разбит на $2 \le K \le N$ подвекторов-блоков:

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(K)} \end{pmatrix}; \qquad x^{(k)} \in \mathbb{R}^{N^{(k)}},\ k = 1, \ldots, K; \qquad N^{(1)} + \ldots + N^{(K)} = N.$$

Данное разбиение инициирует соответствующие разбиения N-мерного вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ:

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \vdots \\ \mu^{(K)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \ldots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \ldots & \Sigma_{KK} \end{pmatrix}.$$

По случайной выборке $X = \{x_1, \ldots, x_n\}$ объема $n > N$, образованной независимыми в совокупности наблюдениями над случайным вектором $x \in \mathbb{R}^N$, необходимо проверить гипотезу о том, что подвекторы (группы признаков) $\{x^{(k)} \in \mathbb{R}^{N^{(k)}},\ k = 1, \ldots, K\}$ независимы в совокупности между собой.

В силу того что подвекторы $\{x^{(k)} \in \mathbb{R}^{N^{(k)}},\ k = 1, \ldots, K\}$ имеют совместное нормальное распределение, их независимость:

$$n_N(x|\mu, \Sigma) = \prod_{k=1}^{K} n_{N^{(k)}}(x^{(k)}|\mu^{(k)}, \Sigma_{kk}),$$

согласно теореме 4.4 эквивалентна их некоррелированности:

$$\mathbf{Cov}\{x^{(k)}, x^{(j)}\} = \Sigma_{kj} = \mathbf{0}_{N^{(k)} \times N^{(j)}},\ k \neq j,\ k, j = 1, \ldots, K,$$

и *гипотеза независимости* сводится к гипотезе

$$H_0 \ : \ \Sigma_{kj} = \delta_{kj}\Sigma_{kk},\ k, j = 1, \ldots, K, \tag{6.36}$$

то есть к гипотезе о том, что Σ имеет вид:

$$\Sigma_0 = \begin{pmatrix} \Sigma_{11} & 0 & \ldots & 0 \\ 0 & \Sigma_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Sigma_{KK} \end{pmatrix},$$

при альтернативе общего вида: $H_1 = \overline{H_0}$.

Воспользуемся для проверки сложных гипотез H_0, H_1 критерием отношения правдоподобия [12]. Определим обобщенную статистику отношения правдоподобия:

$$\lambda = \frac{\max\limits_{(\mu,\Sigma)\in\Theta_0} L_N(\mu, \Sigma)}{\max\limits_{(\mu,\Sigma)\in\Theta} L_N(\mu, \Sigma)} \in [0, 1], \tag{6.37}$$

где $L_N(\mu, \Sigma)$ – функция правдоподобия , вычисленная по выборке X объема n:

$$L_N(\mu, \Sigma) = \prod_{i=1}^{n} n_N(x_i|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}nN}|\Sigma|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^{\mathbf{T}}\Sigma^{-1}(x_i - \mu)\right); \tag{6.38}$$

где Θ – множество всех допустимых значений параметров:

$$\Theta = \{(\mu, \Sigma) : \quad \mu \in \mathbb{R}^N,\ \Sigma = \Sigma^{\mathbf{T}} \succ 0\},$$

а Θ_0 – множество значений параметров, соответствующее гипотезе H_0:

$$\Theta_0 = \big\{(\mu, \Sigma) :\ \mu \in \mathbb{R}^N, \Sigma = diag\{\Sigma_{11}, \ldots, \Sigma_{KK}\},\ \Sigma_{kk} = \Sigma_{kk}^{\mathbf{T}} \succ 0,\ k = 1, \ldots, K\}.$$

Критерий отношения правдоподобия, основанный на статистике (6.37), имеет вид

$$\begin{cases} H_0 & : \ \lambda \geq \lambda_o(\alpha); \\ H_1 = \overline{H_0} & : \ \lambda < \lambda_o(\alpha), \end{cases} \tag{6.39}$$

где порог критерия $\lambda_o = \lambda_o(\alpha)$ определяется по наперед заданному малому значению уровня значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$.

Разобьем вычисленные по выборке X объема $n > N$ выборочное среднее (5.1):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

выборочную ковариационную матрицу (5.2):

$$\hat{\Sigma} = \frac{1}{n}A, \quad A = \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^{\mathbf{T}} = (a_{kl})_{k,l=1}^{N},$$

а также выборочную корреляционную матрицу:

$$\mathcal{R} = (r_{kl})_{k,l=1}^{N}, \quad r_{kl} = \frac{a_{kl}}{\sqrt{a_{kk}a_{ll}}},\ k,l = \overline{1,N},$$

на соответствующие блоки:

$$\overline{x} = \begin{pmatrix} \overline{x}^{(1)} \\ \overline{x}^{(2)} \\ \vdots \\ \overline{x}^{(K)} \end{pmatrix},\ \hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \dots & \hat{\Sigma}_{1K} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & \dots & \hat{\Sigma}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Sigma}_{K1} & \hat{\Sigma}_{K2} & \dots & \hat{\Sigma}_{KK} \end{pmatrix};\ \Sigma_{kj} = \frac{1}{n}A_{kj},\ k,j = 1,\dots,K,$$

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1K} \\ A_{21} & A_{22} & \dots & A_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ A_{K1} & A_{K2} & \dots & A_{KK} \end{pmatrix};\ \mathcal{R} = \begin{pmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} & \dots & \mathcal{R}_{1K} \\ \mathcal{R}_{21} & \mathcal{R}_{22} & \dots & \mathcal{R}_{2K} \\ \vdots & \vdots & & \vdots \\ \mathcal{R}_{K1} & \mathcal{R}_{K2} & \dots & \mathcal{R}_{KK} \end{pmatrix}.$$

Теорема 6.7. *Пусть выборка $X = \{x_1,\dots,x_i\}$ объема $n > N$ образована независимыми в совокупности, одинаково распределенными нормальными случайными N-векторами с невырожденным распределением $\mathcal{N}_N(\mu,\Sigma)$, тогда критерий отношения правдоподобия для проверки гипотезы независимости H_0 из (6.36) против альтернативы $H_1 = \overline{H_0}$ может быть записан в виде*

$$\begin{cases} H_0 & : \ V \geq V_o(\alpha); \\ H_1 = \overline{H_0} & : \ V < V_o(\alpha), \end{cases} \tag{6.40}$$

где статистика критерия является V-статистикой [1]:

$$V = \frac{|\hat{\Sigma}|}{\prod_{k=1}^{K}|\hat{\Sigma}_{kk}|} = \frac{|A|}{\prod_{k=1}^{K}|A_{kk}|} = \frac{|\mathcal{R}|}{\prod_{k=1}^{K}|\mathcal{R}_{kk}|} \in [0,1], \tag{6.41}$$

а порог $V_o = V_o(\alpha)$ при $n \to +\infty$ определяется по уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$ из асимптотического соотношения:

$$V_o(\alpha) = \exp\left(-\frac{F^{-1}_{\chi^2_f}(1-\alpha)}{g}\right), \quad f = \frac{1}{2}\left(N^2 - \sum_{k=1}^{K}\left(N^{(k)}\right)^2\right), \tag{6.42}$$

$$g = g(n) = n - \frac{3}{2} - \frac{1}{3}\frac{N^3 - \sum_{k=1}^{K}\left(N^{(k)}\right)^3}{N^2 - \sum_{k=1}^{K}\left(N^{(k)}\right)^2},$$

$F^{-1}_{\chi^2_f}(\cdot)$– квантиль χ^2-распределения с f степенями свободы.

Доказательство. Воспользуемся теоремой 5.1, согласно доказательству которой выборочные среднее $\overline{x}$ и ковариационная матрица $\hat{\Sigma}$ являются оценками максимального правдоподобия и максимизируют функцию правдоподобия $L_N(\mu,\Sigma)$ (логарифмическую функцию правдоподобия $l(\mu,\Sigma) = \ln L_N(\mu,\Sigma)$ – в доказательстве теоремы 5.1) по $(\mu,\Sigma) \in \Theta$:

$$\max_{(\mu,\Sigma)\in\Theta} L_N(\mu,\Sigma) = L_N(\overline{x},\hat{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}nN}|\hat{\Sigma}|^{\frac{1}{2}n}}e^{-\frac{1}{2}nN}.$$

При истинной гипотезе H_0 получим

$$\max_{(\mu,\Sigma)\in\Theta_0} L_N(\mu,\Sigma) = \prod_{k=1}^{K} \max_{\mu^{(k)},\Sigma_{kk}} L_{N^{(k)}}(\mu^{(k)},\Sigma_{kk}) =$$

$$= \prod_{k=1}^{K} L_{N^{(k)}}(\overline{x}^{(k)},\hat{\Sigma}_{kk}) = \prod_{k=1}^{K} \frac{1}{(2\pi)^{\frac{1}{2}nN^{(k)}}|\hat{\Sigma}_{kk}|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN^{(k)}} =$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}nN}(\prod_{k=1}^{K}|\hat{\Sigma}_{kk}|)^{\frac{1}{2}n}} e^{-\frac{1}{2}nN},$$

где учтено, что $\overline{x}^{(k)}$ и $\hat{\Sigma}_{kk}$ являются оценками максимального правдоподобия для математического ожидания $\mu^{(k)}$ и ковариационной матрицы Σ_{kk} маргинального нормального распределения $\mathcal{N}_{N^{(k)}}(\mu^{(k)},\Sigma_{kk})$ и максимизируют соответствующую функцию правдоподобия $L_{N^{(k)}}(\mu^{(k)},\Sigma_{kk}),\ k=1,\ldots,K$.

Из последних двух соотношений для статистики отношения правдоподобия имеем

$$\lambda = \frac{\max\limits_{(\mu,\Sigma)\in\Theta_0} L_N(\mu,\Sigma)}{\max\limits_{(\mu,\Sigma)\in\Theta} L_N(\mu,\Sigma)} = \left(\frac{|\hat{\Sigma}|}{\prod_{k=1}^{K}|\hat{\Sigma}_{kk}|}\right)^{\frac{n}{2}} = V^{\frac{n}{2}} -$$

монотонно возрастающая функция от V-статистики

$$V = \frac{|\hat{\Sigma}|}{\prod_{k=1}^{K}|\hat{\Sigma}_{kk}|},$$

что приводит к эквивалентной записи критерия отношения правдоподобия (6.39), (6.37) в виде (6.40), (6.41).

Покажем, что для V также справедливы представления

$$V = \frac{|A|}{\prod_{k=1}^{K}|A_{kk}|} = \frac{|\mathcal{R}|}{\prod_{k=1}^{K}|\mathcal{R}_{kk}|}.$$

Первое из них очевидно и следует из того, что $\hat{\Sigma} = \frac{1}{n}A$. Докажем второе представление. Воспользуемся соотношением, связывающим корреляционную матрицу $\mathcal{R}$ с матрицей $A = (a_{kl})_{k,l=1}^{N}$:

$$\mathcal{R} = CAC, \quad C = diag\left\{\frac{1}{\sqrt{a_{11}}},\ldots,\frac{1}{\sqrt{a_{NN}}}\right\},$$

и разобьем диагональную матрицу C на соответствующие блоки:

$$C = diag\{C_{11},\ldots,C_{KK}\}, \quad |C| = \prod_{k=1}^{K}|C_{kk}|,$$

тогда

$$\frac{|\mathcal{R}|}{\prod_{k=1}^{K}|\mathcal{R}_{kk}|} = \frac{|CAC|}{\prod_{k=1}^{K}|C_{kk}A_{kk}C_{kk}|} = \frac{|C|^2|A|}{\prod_{k=1}^{K}|C_{kk}|^2|A_{kk}|} =$$

$$= \frac{|A|}{\prod_{k=1}^{K}|A_{kk}|} = V.$$

Осталось определить порог критерия $V_o = V_o(\alpha)$. В [1] установлено, что при $n \to +\infty$ в условиях гипотезы H_0 для распределения вероятностей V-статистики справедливо асимптотическое разложение ($g = g(n) \to +\infty$):

$$\mathbf{P}_{H_0}\{-g\ln(V) \le z\} = F_{\chi^2_f}(z) + O(g^{-2}),\ z \in \mathbb{R},$$

из которого по заданному уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$ и определяется порог (6.42). □

Следствие 6.3. *В условиях теоремы 6.7 критерий отношения правдоподобия для проверки гипотезы о независимости всех N компонент между собой ($K = N$; $N^{(k)} = 1,\ k = 1,\dots,N$) против общей альтернативы имеет вид*

$$\left\{ \begin{array}{lcl} H_0 & : & |\mathcal{R}| \ge V_o(\alpha); \\ H_1 = \overline{H_0} & : & |\mathcal{R}| < V_o(\alpha), \end{array} \right.$$

где $|\mathcal{R}|$ – определитель выборочной корреляционной матрицы, а для порога критерия $V_o = V_o(\alpha)$, $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$, при $n \to +\infty$ справедливо соотношение

$$V_o(\alpha) = \exp\left(-\left(n - \frac{3}{2} - \frac{N+1}{3}\right)^{-1} F^{-1}_{\chi^2_{\frac{N(N-1)}{2}}}(1-\alpha)\right).$$

7 Проверка гипотез в многомерном статистическом анализе

7.1 T^2-статистика Хотеллинга: ее свойства и распределение вероятностей

7.1.1 T^2-статистика Хотеллинга и ее свойства

Пусть в пространстве $\mathbb{R}^N$ регистрируются наблюдения $x_1, \ldots, x_n$ ($x_j \in \mathbb{R}^N$, $j = \overline{1,n}$), являющиеся независимыми одинаково распределенными нормальными случайными векторами

$$\mathcal{L}\{x_j\} = \mathcal{N}_N(\mu, \Sigma),\ |\Sigma| \neq 0,\ j = \overline{1,n}.$$

Пусть также зафиксирована некоторая точка $\mu^0 \in \mathbb{R}^N$.

Введем в рассмотрение статистики:

$$\overline{x} = \frac{1}{n}\sum_{j=1} nx_j,$$

$$S = \frac{1}{n-1}A,\ \text{где}\ A = \sum_{j=1}^{n}(x_j - \overline{x})(x_j - \overline{x})^{\mathbf{T}},$$

$\overline{x}$ – выборочное среднее, S – несмещенная оценка ковариационной матрицы ковариации Σ. Везде далее считаем, что $n > N$ и $|S| \neq 0$.

Определение 7.1. *T^2-статистикой Хотеллинга называется следующая статистика:*

$$T^2 = T^2(\mu) = n(\overline{x} - \mu)^{\mathbf{T}} S^{-1}(\overline{x} - \mu), \tag{7.1}$$

где $\mu \in \mathbb{R}^N$ – фиксированный N-вектор.

Свойства T^2-статистики Хотеллига

Свойство 1. $T^2 \geq 0$ и $T^2 = 0 \iff \overline{x} = \mu$.

Доказательство. Обозначим

$$u ::= \sqrt{n}(\overline{x} - \mu).$$

Тогда T^2-статистика Хотеллинга записывается в виде

$$T^2 = T^2(\mu) = u^{\mathbf{T}} S^{-1} u.$$

Так как $S \succ 0$, $|S| \neq 0$, то $S^{-1} \succ 0$, $|S^{-1}| \neq 0$. Следовательно,

$$T^2 \geq 0 \text{ и } T^2 = 0 \iff \overline{x} = \mu.$$

□

Свойство 2. При $N > 1$ T^2-статистика пропорциональна оценке квадрата расстояния Махаланобиса.

При $N = 1$ T^2-статистика пропорциональна квадрату отклонения $(\overline{x} - \mu)^2$.

Доказательство. Функция

$$\rho(\mu, \mu^o) = \sqrt{(\mu - \mu^o)^{\mathbf{T}} \Sigma^{-1} (\mu - \mu^o)},$$

которая носит название расстояние Махаланобиса, определяет расстояние между истинным значением вектора математического ожидания μ и его предполагаемым значением μ^o.

Статистика, вычисленная по случайной выборке $X = \{x_1, \ldots, x_n\}$ ($x_j \in \mathbb{R}^N$, $j = \overline{1,n}$),

$$\hat{\rho}(\overline{x}, \mu^o) = \sqrt{(\overline{x} - \mu^o)^{\mathbf{T}} S^{-1} (\overline{x} - \mu^o)} - \tag{7.2}$$

является оценкой расстояния Махаланобиса $\rho(\mu, \mu^o)$.

Сравнивая равенства

$$\hat{\rho}^2(\overline{x}, \mu^o) = (\overline{x} - \mu^o)^{\mathbf{T}} S^{-1} (\overline{x} - \mu^o)$$

и

$$T^2 = T^2(\mu^o) = n(\overline{x} - \mu^o)^{\mathbf{T}} S^{-1} (\overline{x} - \mu^o),$$

видим, что T^2-статистика пропорциональна оценке квадрата расстояния Махаланобиса.

Для случайной выборки $X = \{x_1, \ldots, x_n\}$ из некоторого одномерного распределения вероятностей получаем, полагая в равенстве (7.1) $N = 1$, что T^2-статистика пропорциональна квадрату отклонения $(\overline{x} - \mu)^2$. □

Свойство 3. При $\mu^0 = \mathbf{0_N}$ T^2-статистика инвариантна относительно невырожденного линейного преобразования

$$y_j = Cx_j,\ C - (N \times N)\text{-матрица},\ ,\ |C| \neq 0,\ j = \overline{1,n}.$$

Доказательство. Введем обозначение

$$T_y^2 = T_y^2(\mu^0)\Big|_{\mu^0 = \mathbf{0_N}} = n\overline{y}^{\mathbf{T}} S_y^{-1} \overline{y}, \tag{7.3}$$

где

$$\overline{y} = \frac{1}{n} \sum_{j=1} n y_j = \frac{1}{n} \sum_{j=1} n C x_j = C\overline{x}, \tag{7.4}$$

$$\begin{aligned} S_y &= \frac{1}{n-1} \sum_{j=1}^{n} (y_j - \overline{y}) (y_j - \overline{y})^{\mathbf{T}} = \frac{1}{n-1} \sum_{j=1}^{n} (Cx_j - C\overline{x}) (Cx_j - C\overline{x})^{\mathbf{T}} = \\ &= \frac{1}{n-1} \sum_{j=1}^{n} C (x_j - \overline{x}) (x_j - \overline{x})^{\mathbf{T}} C^{\mathbf{T}} = C \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \overline{x}) (x_j - \overline{x})^{\mathbf{T}} C^{\mathbf{T}} = \\ &= C S_x C^{\mathbf{T}}. \end{aligned} \tag{7.5}$$

Подстановкой (7.4) и (7.5) в (7.3) получаем

$$T_y^2 = n\overline{y}^{\mathbf{T}} S_y^{-1} \overline{y} = n\overline{x}^{\mathbf{T}} C^{\mathbf{T}} \left(C^{\mathbf{T}}\right)^{-1} S_x^{-1} C^{-1} C\overline{x} = T_x^2.$$

□

7.1.2 Распределение вероятностей T^2-статистики Хотеллинга

Определение 7.2. *Пусть $\xi_1, \ldots, \xi_m$ – независимые в совокупности гауссовские случайные величины с законами распределения вероятностей*

$$\mathcal{N}_1(\mu_1, 1), \ldots, \mathcal{N}_1(\mu_m, 1).$$

Тогда распределение вероятностей случайной величины

$$\eta = \xi_1^2 + \ldots + \xi_m^2$$

называется нецентральным χ^2-распределением с m степенями свободы и параметром нецентральности τ^2:

$$\mathcal{L}(\eta) = \chi^2_{m,\tau^2},$$

где

$$\tau^2 = \mu_1^2 + \ldots + \mu_m^2.$$

При $\mu_1 = \ldots = \mu_m = 0$ нецентральное χ^2-распределение называется просто χ^2-распределением:

$$\chi^2_{m,\tau^2}\Big|_{\tau^2=0} = \chi^2_m.$$

Определение 7.3. *Пусть случайная величина η_1 имеет нецентральное χ^2-распределение с m степенями свободы и параметром нецентральности τ^2:*

$$\mathcal{L}(\eta_1) = \chi^2_{m,\tau^2},$$

а случайная величина η_2 не зависит от η_1 и имеет распределение χ^2_n:

$$\mathcal{L}(\eta_2) = \chi^2_n.$$

тогда распределение вероятностей случайной величины $\eta = \dfrac{\frac{1}{m}\eta_1}{\frac{1}{n}\eta_2}$ называется нецентральным F-распределением с (m, n) степенями свободы и параметром нецентральности τ^2:

$$\mathcal{L}(\eta) = F_{m,n,\tau^2}.$$

При $\tau^2 = 0$ имеем центральное F-распределение:

$$\mathcal{L}(\eta) = F_{m,n,0} =:: F_{m,n}.$$

Теорема 7.1 (о распределении T^2-статистики)**.** *Пусть $x_1, \ldots, x_n$ ($x_j \in \mathbb{R}^N$, $j = \overline{1,n}$) – независимые в совокупности одинаково распределенные по закону $\mathcal{N}_N(\mu, \Sigma)$, ($|\Sigma| \neq 0$). Пусть $n > N$, а выборочная ковариационная матрица S невырожденная ($|S| \neq 0$). Тогда T^2-статистика, определенная в (7.1), с условием нормировки имеет следующее нецентральное F-распределение:*

$$\mathcal{L}\left\{\frac{n-N}{N} \cdot \frac{T^2}{n-1}\right\} = F_{N,n-N,\tau^2}, \tag{7.6}$$

$$\tau^2 = n\left(\mu - \mu^0\right)^{\mathbf{T}} \Sigma^{-1} \left(\mu - \mu^0\right).$$

Следствие 7.1. *При $\mu = \mu^0$ в условиях теоремы $\tau^2 = 0$ и*

$$\mathcal{L}\left\{\frac{n-N}{N} \cdot \frac{T^2}{n-1}\right\} = F_{N,n-N}. \tag{7.7}$$

7.2 Проверка гипотезы о значении вектора математического ожидания. T^2-Стьюдента и его оптимальность

Пусть в пространстве $\mathbb{R}^N$ наблюдается случайная выборка $X = \{x_1, \ldots, x_n\}$ объема n из невырожденного N-мерного нормального распределения $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Истинные значения N-вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ неизвестны. Необходимо проверить гипотезы о значении вектора математического ожидания:

$$\begin{array}{lcl} H_0 & : & \mu = \mu^0; \\ H_1 = \overline{H_0} & : & \mu \neq \mu^0, \end{array} \tag{7.8}$$

где $\mu^o \in \mathbb{R}^N$ – предполагаемое (гипотетическое) значение математического ожидания.

Соласно свойству 2: чем меньше значение статистики $T^2(\mu^o)$, тем «ближе» истинное значение математического ожидания μ (выборочное среднее $\overline{x}$ является оценкой μ) к гипотетическому значению μ^o. Эти соображения позволяют предложить следующий критерий для проверки гипотез (7.8):

$$\left\{ \begin{array}{lcl} H_0 & : & T^2(\mu^o) \leq T^2_o(\alpha), \\ H_1 & : & T^2(\mu^o) > T^2_o(\alpha). \end{array} \right. \tag{7.9}$$

где $T^2_o(\alpha)$ — пороговое значение критерия, подлежащее определению по наперед заданному уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$.

Соотношение (7.7) позволяет найти порог критерия (7.9), вычислив вероятность ошибки первого рода:

$$\mathbf{P}\{H_1|H_0\} = \mathbf{P}\{T^2(\mu^o) > T^2_o|\mu = \mu^o\} = 1 - \mathbf{P}\{T^2(\mu^o) \leq T^2_o|\mu = \mu^o\} =$$

$$= 1 - F_{N,n-N}\left(\frac{n-N}{N(n-1)}T^2_o\right) := \alpha,$$

откуда

$$T^2_o = T^2_o(\alpha) = \frac{N(n-1)}{n-N}F^{-1}_{N,n-N}(1-\alpha), \tag{7.10}$$

где $F^{-1}_{N,n-N}(1-\alpha)$ — квантиль уровня $1-\alpha$ от F-распределения Фишера с $(N, n-N)$ степенями свободы.

Следует отметить, что порог $T^2_o = T^2_o(\alpha)$ позволяет также построить доверительную область уровня $1-\alpha$ для вектора математического ожидания:

$$\mathcal{M}_\alpha = \left\{\mu \in R^N : \ (\overline{x} - \mu)^{\mathbf{T}} S^{-1} (\overline{x} - \mu) \leq \frac{T^2_o(\alpha)}{n}\right\},$$

являющуюся случайным N-мерным эллипсоидом в $\mathbb{R}^N$ с центром в точке $\overline{x}$, форма и «объем» (мера Лебега) которого определяются матрицей S^{-1} и вероятностью $1-\alpha$, с которой в него «попадает» истинное значение вектора математического ожидания μ:

$$\mathbf{P}\{\mu \in \mathcal{M}_\alpha\} = 1 - \alpha.$$

Исследуем вероятностные свойства критерия (7.9), (7.10), называемого еще T^2-критерием Стьюдента.

Теорема 7.2. *T^2-критерий* (7.9), (7.10), *построенный по выборке*

$$X = \{x_1, \ldots, x_n\}$$

объема $n > N$, образованной независимыми в совокупности, одинаково распределенными N-векторами-наблюдениями с невырожденным нормальным распределением $\mathcal{N}_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), является критерием отношения правдоподобия для проверки гипотез (7.8) *о значении вектора математического ожидания*

$$\begin{array}{rcl} H_0 & : & \mu = \mu_0, \\ H_1 & : & \overline{H_0}, \end{array} \tag{7.11}$$

при неизвестной ковариационной матрице Σ, и при любом фиксированном уровне значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$ является равномерно наиболее мощным критерием среди всех критериев для проверки гипотез (7.8)*, мощность которых $w = \mathbf{P}\{H_1|H_1\}$ зависит лишь от величины*

$$n(\mu - \mu^o)^{\mathbf{T}}\Sigma^{-1}(\mu - \mu^o)$$

.

Доказательство. Построим критерий отношения правдоподобия для проверки гипотез (7.8):

$$\begin{array}{rcl} H_0 & : & \lambda \geq \lambda_o(\alpha), \\ H_1 & : & \lambda < \lambda_o(\alpha). \end{array} \tag{7.12}$$

где статистика отношения правдоподобия

$$\lambda = \frac{\max\limits_{(\mu,\Sigma)\in\Theta_0} L_N(\mu, \Sigma)}{\max\limits_{(\mu,\Sigma)\in\Theta} L_N(\mu, \Sigma)}$$

определяется функцией правдоподобия $L_N(\mu, \Sigma)$, максимальное значение которой на множестве всех допустимых значений параметров $\Theta = \{(\mu, \Sigma) : \quad \mu \in R^N,\ \Sigma = \Sigma^{\mathbf{T}} \succ 0\}$ равно:

$$\max_{(\mu,\Sigma)\in\Theta} L_N(\mu, \Sigma) = L_N(\overline{x}, \hat{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}nN}|\hat{\Sigma}|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN}, \quad \hat{\Sigma} = \frac{1}{n}A.$$

На множестве значений параметров Θ_0, которому соответствует гипотеза H_0 из (7.8): $\Theta_0 = \{(\mu^o, \Sigma) : \quad \Sigma = \Sigma^{\mathbf{T}} \succ 0\}$, имеем

$$\max_{(\mu,\Sigma)\in\Theta_0} L_N(\mu, \Sigma) = \max_{\Sigma=\Sigma^{\mathbf{T}}\succ 0} L_N(\mu^o, \Sigma) = L_N(\mu^o, \hat{\Sigma}_0) =$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}nN}|\hat{\Sigma}_0|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN},$$

где

$$\hat{\Sigma}_0 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \mu^o)(x_t - \mu^o)^{\mathbf{T}} = \frac{1}{n}A + (\overline{x} - \mu^o)(\overline{x} - \mu^o)^{\mathbf{T}}.$$

С учетом полученного соотношения преобразуем статистику отношения правдоподобия:

$$\lambda = \frac{L(\mu^o, \hat{\Sigma}_0)}{L(\overline{x}, \hat{\Sigma})} = \left(\frac{\left|\hat{\Sigma}\right|}{\left|\hat{\Sigma}_0\right|}\right)^{\frac{n}{2}} = \left(\frac{|A|}{|A + n(\overline{x} - \mu^o)(\overline{x} - \mu^o)^{\mathbf{T}}|}\right)^{\frac{n}{2}}.$$

Далее воспользуемся известными формулами для вычисления определителя блочной матрицы:

$$\begin{vmatrix} E & F \\ G & H \end{vmatrix} = |E||H - GE^{-1}F|, \quad |E| \neq 0;$$

$$\begin{vmatrix} E & F \\ G & H \end{vmatrix} = |H||E - FH^{-1}G|, \quad |H| \neq 0.$$

Положим $E ::= 1$, $F ::= \sqrt{n}(\overline{x}-\mu^o)^{\mathbf{T}}$, $G ::= -\sqrt{n}(\overline{x}-\mu^o)$, $H = A$ и получим соотношение

$$\begin{vmatrix} 1 & \sqrt{n}(\overline{x}-\mu^o)^{\mathbf{T}} \\ -\sqrt{n}(\overline{x}-\mu^o) & A \end{vmatrix} = |1||A - \big(-\sqrt{n}(\overline{x}-\mu_0)\big)1^{-1}\sqrt{n}(\overline{x}-\mu_0)^{\mathbf{T}}| =$$

$$= |A + n(\overline{x}-\mu^o)(\overline{x}-\mu^o)^{\mathbf{T}}| = |A||1 + n(\overline{x}-\mu^o)^{\mathbf{T}}A^{-1}(\overline{x}-\mu^o)|,$$

с учетом которого статистика отношения правдоподобия

$$\lambda = \left(\frac{1}{1+\frac{T^2(\mu^o)}{n-1}}\right)^{\frac{n}{2}} —$$

строго убывающая функция от T^2-статистики $T^2(\mu^o)$ из (7.1), поэтому критерий отношения правдоподобия на ее основе эквивалентен T^2-критерию Стьюдента (7.9), (7.10).

Доказательство того, что T^2-критерий (7.9), (7.10) является равномерно наиболее мощным, носит технический характер и основано на анализе распределения вероятностей T^2-статистики в случае, когда гипотеза H_0 неверна. □

На практике иногда по выборке $X = \{x_1, \ldots, x_n\}$ объема n из невырожденного нормального распределения $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$) необходимо проверить еще одну гипотезу о значении вектора математического ожидания — *гипотезу симметрии*, которая состоит в предположении о том, что значения компонент вектора математического ожидания $\mu = (\tilde{\mu}_1, \ldots, \tilde{\mu}_N)^{\mathbf{T}}$ равны между собой при альтернативе общего вида:

$$\begin{array}{lcl} H_0 & : & \tilde{\mu}_1 = \ldots = \tilde{\mu}_N, \\ H_1 & : & \exists i \neq j \in \{1, \ldots, N\},\ \tilde{\mu}_i \neq \tilde{\mu}_j. \end{array} \tag{7.13}$$

Очевидно, что гипотеза симметрии H_0 из (7.13) может быть записана в виде $\mu = \tilde{\mu}\mathbf{1}_N$, где $\tilde{\mu} \in \mathbb{R}$ — произвольное общее значение компонент, а $\mathbf{1}_N$ — N-вектор, составленный из единиц.

Сведем гипотезы из (7.13) к гипотезам о значении вектора математического ожидания. Для этого введем в рассмотрение произвольную $((N-1) \times N)$-матрицу C полного ранга: $rank(C) = N - 1$, строки которой линейно независимы между собой и ортогональны к единичному N-вектору $\mathbf{1}_N$, то есть

$$C\mathbf{1}_N = \mathbf{0}_{N-1}.$$

Осуществим линейное преобразование выборки $X = \{x_1, \ldots, x_n\}$ в выборку $Y = \{y_1, \ldots, y_n\}$:

$$y_j = Cx_j,\ j = \overline{1,n}.$$

Наблюдения из выборки Y имеют невырожденное нормальное распределение:

$$\mathcal{L}\{y_j\} = N_{N-1}(C\mu, C\Sigma C^{\mathbf{T}}),\ |C\Sigma C^{\mathbf{T}}| \neq 0,$$

и при истинной гипотезе H_0, по построению, их математическое ожидание – нулевой $(N-1)$-вектор:

$$C\mu = \tilde{\mu}C\mathbf{1}_N = \tilde{\mu}\mathbf{0}_{N-1} = \mathbf{0}_{N-1}.$$

Поэтому проверка гипотез (7.13) эквивалентна проверке гипотезы о том, что математическое ожидание наблюдений из выборки $Y = \{y_1, \ldots, y_n\}$ – нулевое, против альтернативы

общего вида. Остается лишь воспользоваться T^2-критерием (7.1), (7.10), (7.9), положив в нем $N ::= N-1$; $x_t ::= Cx_t$, $t = \overline{1,n}$; $\mu^o ::= \mathbf{0}_{N-1}$.

Получим следующий критерий для проверки гипотез (7.13):

$$\begin{array}{rcl} H_0 & : & n\overline{x}^{\mathbf{T}}C^{\mathbf{T}}(CSC^{\mathbf{T}})^{-1}C\overline{x} \le \frac{(N-1)(n-1)}{n-(N-1)}F^{-1}_{N-1,n-(N-1)}(1-\alpha), \\ H_1 & : & n\overline{x}^{\mathbf{T}}C^{\mathbf{T}}(CSC^{\mathbf{T}})^{-1}C\overline{x} > \frac{(N-1)(n-1)}{n-(N-1)}F^{-1}_{N-1,n-(N-1)}(1-\alpha). \end{array} \tag{7.14}$$

где $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$ — наперед заданный уровень значимости. Мощность полученного критерия зависит от матрицы линейного преобразования C: $w = w(C) = \mathbf{P}\{H_1|H_1\}$, и его можно оптимизировать по мощности, выбирая соответствующим образом матрицу C.

7.3 Сравнение векторов математических ожиданий по двум выборкам. Многомерная проблема Беренса – Фишера

7.3.1 Сравнение векторов математических ожиданий при неизвестной одинаковой ковариационной матрице

Пусть в пространстве наблюдений $\mathbb{R}^N$ имеются две случайные выборки

$$X^{(1)} = \{x_1^{(1)}, \ldots, x_{N_1}^{(1)}\} \text{ и } X^{(2)} = \{x_1^{(2)}, \ldots, x_{N_2}^{(2)}\}$$

соответственно объемов $N_1 > N$ и $N_2 > N$. Наблюдения $\{x_j^{(i)}\}_{j=1}^{N_i}$ из выборки $X^{(i)}$ $(i = 1,2)$ независимы в совокупности и одинаково распределены по нормальному закону $\mathcal{N}_N(\mu_i, \Sigma)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и невырожденной ковариационной $(N \times N)$-матрицей Σ $(|\Sigma| \neq 0)$. Выборка $X^{(1)}$ не зависит от выборки $X^{(2)}$. Необходимо проверить гипотезу о совпадении векторов математических ожиданий μ_1 и μ_2 против альтернативы общего вида:

$$\begin{array}{rcl} H_0 & : & \mu_1 = \mu_2, \\ H_1 & : & \mu_1 \neq \mu_2. \end{array} \tag{7.15}$$

при неизвестном значении одинаковой для обеих выборок ковариационной матрицы Σ.

Для проверки гипотез (7.15) воспользуемся предложенной в предыдущем парарафе T^2-статистикой, а точнее — леммой. По выборкам $X^{(1)}$ и $X^{(2)}$ определим статистику:

$$y = \frac{1}{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}(\overline{x}_{(1)} - \overline{x}_{(2)}), \tag{7.16}$$

а также построим оценку для общей ковариационной матрицы Σ, которая учитывает все $N_1 + N_2$ наблюдений из выборок $X^{(1)}$ и $X^{(2)}$:

$$S = \frac{1}{N_1 + N_2 - 2}\left((N_1 - 1)S_{(1)} + (N_2 - 1)S_{(2)}\right), \tag{7.17}$$

где

$$\overline{x}_{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j^{(i)}, \quad S_{(i)} = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(x_j^{(i)} - \overline{x}_{(i)})(x_j^{(i)} - \overline{x}_{(i)})^{\mathbf{T}} -$$

выборочные среднее и ковариационная матрица для выборки $X^{(i)}$ $(i = 1,2)$.

Ранее было доказано, что $\overline{x}_{(i)}$ и $S_{(i)}$ независимы и имеют распределения:

$$\mathcal{L}\{\overline{x}_{(i)}\} = \mathcal{N}_N\left(\mu_i, \frac{1}{N_i}\Sigma\right);$$

$$\mathcal{L}\{(N_i-1)S_{(i)}\} = W_N(\Sigma, N_i-1), \quad i=1,2,$$

поэтому в силу независимости выборок $X^{(1)}$ и $X^{(2)}$ статистики y и S также независимы, а по свойствам многомерного нормального распределения и распределения Уишарта:

$$\mathcal{L}\{y\} = \mathcal{N}_N\left(\frac{1}{\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}}(\mu_1-\mu_2), \Sigma\right);$$

$$\mathcal{L}\{(N_1+N_2-2)S\} = W_N(\Sigma, N_1+N_2-2).$$

Если через μ_y обозначить

$$\mu_y = \mathbf{E}\{y\} = \frac{1}{\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}}(\mu_1-\mu_2),$$

то гипотезы (7.15) примут вид

$$\begin{array}{rcl} H_0 & : & \mu_y = \mathbf{0}_N, \\ H_1 & : & \mu_y \neq \mathbf{0}_N. \end{array} \tag{7.18}$$

Введем в рассмотрение *обобщенную* $\overline{T}^2$*-статистику*:

$$\overline{T}^2 = y^{\mathbf{T}} S^{-1} y, \tag{7.19}$$

основанный на ней критерий для проверки гипотез (7.15) имеет вид

$$\left\{\begin{array}{rcl} H_0 & : & \overline{T}^2 \leq \overline{T}^2_o, \\ H_1 & : & \overline{T}^2 > \overline{T}^2_o. \end{array}\right. \tag{7.20}$$

где порог $\overline{T}^2_o = \overline{T}^2_o(\alpha)$ определяется по наперед заданному малому значению уровня значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$.

Найдем пороговое значение $\overline{T}^2_o = \overline{T}^2_o(\alpha)$, применив лемму к $\overline{T}^2$-статистике, задаваемой соотношениями (7.19), (7.16), (7.17):

$$\mathbf{P}\{H_1|H_0\} = 1 - \mathbf{P}\{H_0|H_0\} = 1 - \mathbf{P}\{\overline{T}^2 \leq \overline{T}^2_o|\ H_0\} =$$

$$= 1 - \mathbf{P}\left\{\frac{N_1+N_2-N-1}{N}\frac{\overline{T}^2}{N_1+N_2-2} \leq \frac{N_1+N_2-N-1}{N}\frac{\overline{T}^2_o}{N_1+N_2-2}\right\} =$$

$$= 1 - F_{N,N_1+N_2-N-1}\left(\frac{N_1+N_2-N-1}{N}\frac{\overline{T}^2_o}{N_1+N_2-2}\right) := \alpha,$$

откуда получаем

$$\overline{T}^2_o = \overline{T}^2_o(\alpha) = \frac{N(N_1+N_2-2)}{N_1+N_2-N-1} F^{-1}_{N,N_1+N_2-N-1}(1-\alpha). \tag{7.21}$$

Следует заметить, что критерий (7.20), (7.21) является критерием отношения правдоподобия.

7.3.2 Сравнение векторов математических ожиданий при различных ковариационных матрицах. Многомерная проблема Беренца – Фишера

Обобщим задачу из предыдущего пункта. Пусть в пространстве наблюдений $\mathbb{R}^N$ имеются две случайные выборки

$$X^{(1)} = \{x_1^{(1)}, \ldots, x_{N_1}^{(1)}\} \text{ и } X^{(2)} = \{x_1^{(2)}, \ldots, x_{N_2}^{(2)}\}$$

соответственно объемов $N_1 > N$ и $N_2 > N$. Наблюдения $\{x_j^{(i)}\}_{j=1}^{N_i}$ из выборки $X^{(i)}$ $(i = 1, 2)$ независимы в совокупности и одинаково распределены по нормальному закону $\mathcal{N}_N(\mu_i, \Sigma_i)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и невырожденной ковариационной $(N \times N)$-матрицей Σ_i $(|\Sigma_i| \neq 0)$, $i = 1, 2$. Выборка $X^{(1)}$ не зависит от выборки $X^{(2)}$. Необходимо проверить гипотезы о совпадении векторов математических ожиданий при условии, что ковариационные матрицы Σ_1 и Σ_2, вообще говоря, различны, а их значения неизвестны:

$$\begin{array}{rcl} H_0 & : & \mu_1 = \mu_2, \\ H_1 & : & \mu_1 \neq \mu_2. \end{array} \tag{7.22}$$

Рассмотрим два случая.

Выборки равного объема $(N_1 = N_2 = n)$. Образуем новую выборку

$$X = \{x_1, \ldots, x_n\}$$

:

$$x_j = x_j^{(1)} - x_j^{(2)}, \ j = \overline{1, n},$$

где $\{x_j\}_{j=1}^n$ независимы в совокупности и одинаково распределены $(j = \overline{1, n})$:

$$\mathcal{L}\{x_t\} = \mathcal{N}_N(\mu_1 - \mu_2, \Sigma_1 + \Sigma_2).$$

Обозначим: $\mu = \mathbf{E}\{x_t\} = \mu_1 - \mu_2$, тогда гипотезы (7.22) примут эквивалентный вид:

$$\begin{array}{rcl} H_0 & : & \mu = \mathbf{0}_N, \\ H_1 & : & \mu \neq \mathbf{0}_N. \end{array}$$

Построим выборочную ковариационную матрицу по выборке X:

$$S = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \overline{x})(x_j - \overline{x})^{\mathbf{T}} = S_{(1)} + S_{(2)},$$

где

$$\overline{x} = \overline{x}_{(1)} - \overline{x}_{(2)},$$

а $\overline{x}_{(1)}$, $\overline{x}_{(2)}$ и $S_{(1)}$, $S_{(1)}$ — выборочные средние и выборочные ковариационные матрицы из (7.17), вычисленные по выборкам $X^{(1)}$, $X^{(2)}$.

Далее применяем к выборке $X = \{x_j\}_{j=1}^n$ T^2-критерий Стьюдента, полаая в нем $\mu^o ::= \mathbf{0}_N$, для T^2-статистики получим соотношение

$$T^2 = n\overline{x}^{\mathbf{T}} S^{-1} \overline{x} = n(\overline{x}_{(1)} - \overline{x}_{(2)})^{\mathbf{T}} (S_{(1)} + S_{(2)})^{-1} (\overline{x}_{(1)} - \overline{x}_{(2)}),$$

а сам T^2-критерий имеет вид

$$\left\{ \begin{array}{rcl} H_0 & : & T^2 \leq T_o^2(\alpha), \\ H_1 & : & T^2 > T_o^2(\alpha). \end{array} \right.$$

где пороговое значение $T_o^2(\alpha)$ вычисляется по уровню значимости $\alpha = \mathbf{P}\{H_1 | H_0\} \in (0, 1)$ из соотношения:

$$T_o^2(\alpha) = \frac{N(n-1)}{n-N} F_{N,n-N}^{-1}(1 - \alpha).$$

Выборки различного объема ($N_1 \neq N_2$; пусть $N_1 < N_2$). В этом случае можно из выборки большего объема отбросить наблюдения, чтобы ее объем стал равен объему меньшей выборки, и воспользоваться критерием, полученным выше для случая равных объемов.

Однако чтобы учесть все наблюдения, целесообразнее поступить следующим образом [1]. Ввести в рассмотрение новую выборку $X = \{x_t\}_{t=1}^{N_1}$:

$$x_t = x_t^{(1)} - \frac{1}{N_2}\sum_{j=1}^{N_2} x_j^{(2)} - \sqrt{\frac{N_1}{N_2}}\left(x_t^{(2)} - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^{(2)}\right),\ t = \overline{1, N_1},$$

где $\{x_t\}_{t=1}^{N_1}$ некоррелированы:

$$\mathbf{Cov}\{x_t, x_l\} =$$

$$= \mathbf{Cov}\left\{x_t^{(1)} - \frac{1}{N_2}\sum_{j=1}^{N_2} x_j^{(2)} - \sqrt{\frac{N_1}{N_2}}\left(x_t^{(2)} - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^{(2)}\right),\right.$$

$$\left. x_l^{(1)} - \frac{1}{N_2}\sum_{j=1}^{N_2} x_j^{(2)} - \sqrt{\frac{N_1}{N_2}}\left(x_l^{(2)} - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^{(2)}\right)\right\} =$$

$$= \delta_{tl}\Sigma_1 + \mathbf{Cov}\left\{\frac{1}{N_2}\sum_{j=1}^{N_2} x_j^{(2)} + \sqrt{\frac{N_1}{N_2}}\left(x_t^{(2)} - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^{(2)}\right),\right.$$

$$\left.\frac{1}{N_2}\sum_{j=1}^{N_2} x_j^{(2)} + \sqrt{\frac{N_1}{N_2}}\left(x_l^{(2)} - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i^{(2)}\right)\right\} =$$

$$= \delta_{tl}\Sigma_1 + \mathbf{Cov}\left\{\sqrt{\frac{N_1}{N_2}}x_t^{(2)} + \left(\frac{1}{N_2} - \frac{1}{\sqrt{N_1 N_2}}\right)\sum_{j=1}^{N_1} x_j^{(2)} + \frac{1}{N_2}\sum_{j=N_1+1}^{N_2} x_j^{(2)},\right.$$

$$\left.\sqrt{\frac{N_1}{N_2}}x_l^{(2)} + \left(\frac{1}{N_2} - \frac{1}{\sqrt{N_1 N_2}}\right)\sum_{j=1}^{N_1} x_j^{(2)} + \frac{1}{N_2}\sum_{j=N_1+1}^{N_2} x_j^{(2)}\right\} =$$

$$= \delta_{tl}\Sigma_1 + \delta_{tl}\frac{N_1}{N_2}\Sigma_2 + 2\left(\frac{1}{N_2} - \frac{1}{\sqrt{N_1 N_2}}\right)\sqrt{\frac{N_1}{N_2}}\Sigma_2 +$$

$$+ \left(\frac{1}{N_2} - \frac{1}{\sqrt{N_1 N_2}}\right)^2 N_1\Sigma_2 + \left(\frac{1}{N_2}\right)^2 (N_2 - N_1)\Sigma_2 =$$

$$= \delta_{tl}\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right), \quad t, l = \overline{1, N_1},$$

имеют нормальное распределение:

$$\mathcal{L}\{x_t\} = \mathcal{N}_N\left(\mu_1 - \mu_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right),\ t = \overline{1, N_1},$$

и поэтому независимы в совокупности.

По построенной выборке $X = \{x_t\}_{t=1}^{N_1}$ объема N_1 вычислим T^2-статистику из ($\mu^o ::= \mathbf{0}_N$, $n ::= N_1$):

$$T^2 = N_1 \overline{x}^{\mathbf{T}} S^{-1}\overline{x},$$

где

$$\overline{x} = \frac{1}{N_1}\sum_{t=1}^{N_1} x_t, \quad S = \frac{1}{N_1 - 1}\sum_{t=1}^{N_1}(x_t - \overline{x})(x_t - \overline{x})^{\mathbf{T}} -$$

выборочные среднее и ковариационная матрица по $X = \{x_t\}_{t=1}^{N_1}$.

Далее для проверки гипотез H_0, H_1 применяем к выборке $X = \{x_t\}_{t=1}^{N_1}$ T^2-критерий Стьюдента:

$$\begin{array}{rcl} H_0 & : & T^2 \le T_o^2(\alpha), \\ H_1 & : & T^2 > T_o^2(\alpha). \end{array}$$

$$T_o^2(\alpha) = \frac{N(N_1-1)}{N_1-N} F^{-1}_{N,N_1-N}(1-\alpha), \quad \alpha = \mathbf{P}\{H_1|H_0\} \in (0,1).$$

Замечание 7.3.1. *Для построенных в критериев не показано, что они являются оптимальными по мощности. В этом и состоит многомерная проблема Беренца – Фишера. Поэтому если заранее известно, что ковариационные матрицы в выборках совпадают, то лучше использовать критерий из, являющийся критерием отношения правдоподобия.*

7.4 Проверка гипотез относительно параметров многомерного нормального распределения

В пространстве $\mathbb{R}^N$ наблюдается случайная выборка $X = \{x_1, \ldots, x_n\}$ объема $n > N$ из невырожденного нормального распределения $\mathcal{N}_N(\mu, \Sigma)$ с неизвестными значениями параметров: вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицы Σ ($|\Sigma| \neq 0$). Рассмотрим всевозможные гипотезы относительно μ и Σ.

7.4.1 Проверка гипотез о значении вектора математического ожидания

Пусть ковариационная матрица Σ ($|\Sigma| \neq 0$) неизвестна. Необходимо проверить гипотезы о том, что математическое ожидание μ совпадает с наперед заданным вектором $\mu^o \in \mathbb{R}^N$:

$$\begin{array}{rcl} H_0 & : & \mu = \mu^0; \\ H_1 = \overline{H_0} & : & \mu \neq \mu^0, \end{array} \tag{7.23}$$

где $\mu^o \in \mathbb{R}^N$ – предполагаемое (гипотетическое) значение математического ожидания.

Данная задача была решена в параграфе 7.2.

7.4.2 Проверка гипотез о значении ковариационной матрицы

Пусть вектор математического ожидания $\mu \in \mathbb{R}^N$ неизвестен. Необходимо проверить гипотезы:

$$\begin{array}{rcl} H_0 & : & \Sigma = \Sigma^o; \\ H_1 = \overline{H_0} & : & \Sigma \neq \Sigma^o, \end{array} \tag{7.24}$$

где Σ^o – наперед заданная невырожденная ковариационная $(N \times N)$-матрица $(\Sigma^o = (\Sigma^o)^{\mathbf{T}} \succ 0)$.

Воспользуемся критерием отношения правдоподобия:

$$\left\{ \begin{array}{rcl} H_0 & : & \lambda \ge \lambda_o(\alpha), \\ H_1 & : & \lambda < \lambda_o(\alpha). \end{array} \right.$$

где статистика отношения правдоподобия вычисляется по функции правдоподобия $L_N(\mu, \Sigma)$

$$L_N(\mu, \Sigma) = \prod_{i=1}^{n} n_N(x_i|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}nN}|\Sigma|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^{\mathbf{T}}\Sigma^{-1}(x_i-\mu)\right); \tag{7.25}$$

аналогично доказательству теоремы 7.2:

$$\lambda = \frac{\max\limits_{\mu} L_N(\mu, \Sigma^o)}{\max\limits_{\mu,\Sigma} L_N(\mu, \Sigma)} = \frac{L_N(\overline{x}, \Sigma^o)}{L_N(\overline{x}, \hat{\Sigma})} =$$

$$= \frac{(2\pi)^{-\frac{nN}{2}} |\Sigma^o|^{-n/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(x_j - \overline{x})^{\mathbf{T}}(\Sigma^o)^{-1}(x_j - \overline{x})\right)}{(2\pi)^{-\frac{nN}{2}} |\hat{\Sigma}|^{-n/2} e^{-\frac{nN}{2}}} =$$

$$= \left(\frac{e}{n}\right)^{\frac{nN}{2}} |\Sigma^o|^{-n/2} |A|^{n/2} \exp\left(-\frac{1}{2} tr\left((\Sigma^o)^{-1} A\right)\right),$$

где $\overline{x}$ – выборочное среднее:

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j,$$

$\hat{\Sigma}$ – выборочная ковариационная матрица:

$$\hat{\Sigma} = \frac{1}{n} A,\ A = \sum_{j=1}^{n}(x_j - \overline{x})(x_j - \overline{x})^{\mathbf{T}}.$$

Построим критерий, эквивалентный критерию отношения правдоподобия:

$$\left\{ \begin{array}{lcl} H_0 & : & -2\ln\lambda \le \Delta, \\ H_1 & : & -2\ln\lambda > \Delta. \end{array} \right.$$

где относительно статистики $-2\ln\lambda$ установлено [1]:

$$\mathcal{L}_{H_0}\{-2\ln\lambda\} \to \chi^2_f, \quad n \to +\infty,$$

где χ^2_f – χ^2-распределение с $f = \frac{N(N+1)}{2}$ степенями свободы. Этот факт позволяет при «большом» объеме выборки n ($n \to +\infty$) по наперед заданному уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$ найти пороговое значение $\Delta = \Delta(\alpha)$:

$$\Delta(\alpha) = F^{-1}_{\chi^2_f}(1-\alpha), \quad f = \frac{N(N+1)}{2}.$$

7.4.3 Проверка гипотез о совпадении многомерного нормального распределения с наперед заданным многомерным нормальным распределением

Необходимо проверить гипотезы:

$$\begin{array}{lcl} H_0 & : & \mu = \mu^o,\ \Sigma = \Sigma^o, \\ H_1 & : & \mu \ne \mu^o \text{ и/или } \Sigma \ne \Sigma^o. \end{array}$$

где μ^o и Σ^o – наперед заданные гипотетические значения N-вектора математического ожидания ($\mu^o \in^{\mathbb{R}} N$) и ковариационной $(N \times N)$-матрицы ($\Sigma^o = (\Sigma^o)^{\mathbf{T}} \succ 0$).

Критерий отношения правдоподобия имеет вид

$$\left\{ \begin{array}{lcl} H_0 & : & \lambda \ge \lambda_o(\alpha), \\ H_1 & : & \lambda < \lambda_o(\alpha). \end{array} \right.$$

где

$$\lambda = \frac{L_N(\mu^o, \Sigma^o)}{\max\limits_{\mu,\Sigma} L_N(\mu, \Sigma)} = \frac{L_N(\mu^o, \Sigma^o)}{L_N(\overline{x}, \hat{\Sigma})} =$$

$$= \frac{(2\pi)^{-\frac{nN}{2}}|\Sigma^o|^{-n/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{n}(x_t-\mu^o)^{\mathbf{T}}(\Sigma^o)^{-1}(x_t-\mu^o)\right)}{(2\pi)^{-\frac{nN}{2}}|\frac{1}{n}A|^{-n/2}e^{-\frac{nN}{2}}} =$$

$$= \left(\frac{e}{n}\right)^{\frac{nN}{2}}|\Sigma^o|^{-n/2}|A|^{n/2}\exp\left(-\frac{1}{2}(x_t-\mu^o)^{\mathbf{T}}(\Sigma^o)^{-1}(x_t-\mu^o)\right) =$$

$$= \left(\frac{e}{n}\right)^{\frac{nN}{2}}|\Sigma^o|^{-n/2}|A|^{n/2}\exp\left(-\frac{1}{2}tr\left((\Sigma^o)^{-1}A\right) -\right.$$

$$\left. -\frac{n}{2}(\overline{x}-\mu^o)^{\mathbf{T}}(\Sigma^o)^{-1}(\overline{x}-\mu^o)\right).$$

Используем эквивалентный критерий:

$$\left\{\begin{array}{lcl} H_0 & : & -2\ln\lambda \le \Delta, \\ H_1 & : & -2\ln\lambda > \Delta. \end{array}\right.$$

где для статистики $-2\ln\lambda$ также установлено предельное распределение [1]:

$$\mathcal{L}_{H_0}\{-2\ln\lambda\} \to \chi^2_f,\ f = \frac{N(N+1)}{2} + N, \quad n \to +\infty,$$

что позволяет определить при $n \to +\infty$ пороговое значение $\Delta = \Delta(\alpha)$, $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$:

$$\Delta(\alpha) = F^{-1}_{\chi^2_f}(1-\alpha), \quad f = \frac{N(N+1)}{2} + N.$$

7.5 Проверка гипотез относительно нескольких выборок из многомерных нормальных распределений

Пусть в пространстве $\mathbb{R}^N$ наблюдается $L \ge 2$ независимых в совокупности случайных выборок $X^{(i)} = \{x_1^{(i)}, \ldots, x_{n_i}^{(i)}\}$, $i = \overline{1,L}$. Выборка $X^{(i)}$ объема $n_i > N$ $(i = \overline{1,L})$ образована независимыми в совокупности, одинаково распределенными N-векторами-наблюдениями, имеющими невырожденное нормальное распределение $\mathcal{N}_N(\mu_i, \Sigma_i)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и ковариационной $(N \times N)$-матрицей Σ_i $(|\Sigma_i| \ne 0)$.

7.5.1 Гипотеза о совпадении векторов математических ожиданий при неизвестной одинаковой ковариационной матрице

Проверим гипотезу о равенстве между собой векторов математических ожиданий $\mu_1, \ldots, \mu_L$ против альтернативы общего вида:

$$\begin{array}{lcl} H_0 & : & \mu_1 = \ldots = \mu_L, \\ H_1 & : & \overline{H_0}. \end{array}$$

при неизвестном значении одинаковой для всех выборок ковариационной матрицы:

$$\Sigma_1 = \ldots = \Sigma_L = \Sigma \quad (|\Sigma| \ne 0).$$

Запишем функцию правдоподобия для объединенной выборки

$$X = \bigcup_{i=1}^{L} X^{(i)}$$

объема

$$n = \sum_{i=1}^{L} n_i :$$

$$L\left(\mu_1, \Sigma_1; \ldots; \mu_L, \Sigma_L\right) = \prod_{i=1}^{L} L_N^{(i)}(\mu_i, \Sigma_i) =$$

$$= (2\pi)^{-\frac{nN}{2}} \prod_{i=1}^{L} |\Sigma_i|^{-n_i/2} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{n_i} (x_j^{(i)} - \mu_i)^{\mathbf{T}} \Sigma_i^{-1} (x_j^{(i)} - \mu_i) \right),$$

где учтено, что выборки $X^{(1)}, \ldots, X^{(L)}$ независимы между собой, а функция правдоподобия для i-й выборки имеет вид

$$L_N^{(i)}(\mu_i, \Sigma_i) = \prod_{j=1}^{n_i} n_N(x_j^{(i)} | \mu_i, \Sigma_i) =$$

$$= \frac{1}{(2\pi)^{\frac{1}{2} n_i N} |\Sigma_i|^{\frac{1}{2} n_i}} \exp\left(-\frac{1}{2} \sum_{j=1}^{n_i} (x_j^{(i)} - \mu_i)^{\mathbf{T}} \Sigma_i^{-1} (x_j^{(i)} - \mu_i) \right).$$

Воспользуемся критерием отношения правдоподобия и определим статистику отношения правдоподобия:

$$\lambda = \frac{\max\limits_{\mu, \Sigma} L\left(\mu_1, \Sigma_1; \ldots; \mu_L, \Sigma_L\right)}{\max\limits_{\mu_1, \ldots, \mu_L, \Sigma} L\left(\mu_1, \Sigma; \ldots; \mu_L, \Sigma\right)} =$$

$$= \frac{(2\pi)^{-\frac{nN}{2}} |\frac{1}{n} A|^{-n/2} e^{-\frac{nN}{2}}}{(2\pi)^{-\frac{nN}{2}} |\frac{1}{n} \sum_{i=1}^{L} A_i|^{-n/2} e^{-\frac{nN}{2}}} = \left(\frac{|\sum_{i=1}^{L} A_i|}{|A|} \right)^{\frac{n}{2}},$$

где

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{L} \sum_{j=1}^{n_i} x_j^{(i)} -$$

выборочное среднее,

$$\hat{\Sigma}_0 = \frac{1}{n} A, \quad A = \sum_{i=1}^{L} \sum_{j=1}^{n_i} (x_j^{(i)} - \overline{x})(x_j^{(i)} - \overline{x})^{\mathbf{T}} -$$

выборочная ковариационная матрица, вычисленные по объединенной выборке X объема n и являющиеся при истинной гипотезе H_0 оценками максимального правдоподобия для общих вектора математического ожидания μ и ковариационной матрицы Σ;

$$\overline{x}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^{(i)},\ i = \overline{1, L},\ -$$

выборочные средние,

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} A_i, \quad A_i = \sum_{j=1}^{n_i}(x_j^{(i)} - \overline{x}_{(i)})(x_j^{(i)} - \overline{x}_{(i)})^{\mathbf{T}},\ i = \overline{1,L},\ -$$

выборочная ковариационная матрица, являющиеся оценками максимального правдоподобия для математических ожиданий $\mu_1, \ldots, \mu_L$ и общей для всех выборок $X^{(1)}, \ldots, X^{(L)}$ ковариационной матрицы Σ в общем случае (когда гипотеза H_0, вообще говоря, неверна).

Вместо статистики λ воспользуемся так называемой U-статистикой:

$$U = \lambda^{\frac{2}{n}} = \frac{|\sum_{i=1}^{L} A_i|}{|A|},$$

для которой известно [1]:

$$\mathcal{L}_{H_0}\{-g \ln U\} \to \chi^2_f, \quad n_i \to \infty,\ i = \overline{1,L}\ (n \to +\infty),$$

где $g = n - L - \frac{N-L}{2} - 1$, $f = N(L-1)$. Тогда критерий отношения правдоподобия

$$\left\{ \begin{array}{lcl} H_0 & : & \lambda \geq \lambda_o(\alpha), \\ H_1 & : & \lambda < \lambda_o(\alpha). \end{array} \right.$$

запишется в эквивалентном виде:

$$\left\{ \begin{array}{lcl} H_0 & : & -g \ln U \leq \Delta(\alpha), \\ H_1 & : & -g \ln U > \Delta(\alpha). \end{array} \right.$$

где порог критерия $\Delta = \Delta(\alpha)$ определяется по наперед заданному уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0,1)$:

$$\Delta(\alpha) = F^{-1}_{\chi^2_f}(1-\alpha), \quad f = N(L-1).$$

В случае двух выборок ($L = 2$), воспользовавшись соотношениями

$$A_1 + A_2 = A + n_1(\overline{x} - \overline{x}_{(1)})(\overline{x} - \overline{x}_{(1)})^{\mathbf{T}} + n_2(\overline{x} - \overline{x}_{(2)})(\overline{x} - \overline{x}_{(2)})^{\mathbf{T}};$$

$$n\overline{x} = n_1\overline{x}_{(1)} + n_2\overline{x}_{(2)},$$

и формулой блочного определителя, по аналогии с доказательством теоремы 7.2 для U-статистики получим

$$U = \frac{|A_1 + A_2|}{|A|} =$$

$$= \frac{|A_1 + A_2|}{|A_1 + A_2 - n_1(\overline{x} - \overline{x}_{(1)})(\overline{x} - \overline{x}_{(1)})^{\mathbf{T}} - n_2(\overline{x} - \overline{x}_{(2)})(\overline{x} - \overline{x}_{(2)})^{\mathbf{T}}|} =$$

$$= \frac{|A_1 + A_2|}{|A_1 + A_2 - \frac{n_1 n_2}{n_1+n_2}(\overline{x}_{(1)} - \overline{x}_{(2)})(\overline{x}_{(1)} - \overline{x}_{(2)})^{\mathbf{T}}|} =$$

$$= \frac{|A_1 + A_2|}{|A_1 + A_2||1 + \frac{n_1 n_2}{n_1+n_2}(\overline{x}_{(1)} - \overline{x}_{(2)})^{\mathbf{T}}(A_1 + A_2)^{-1}(\overline{x}_{(1)} - \overline{x}_{(2)})|} =$$

$$= \frac{1}{1 + \frac{\overline{T}^2}{n_1+n_2-2}} -$$

выражается через обобщенную $\overline{T}^2$-статистику из (7.19), и построенный ранее в п. 7.3.1 на ее основе критерий (7.20), (7.21) также является критерием отношения правдоподобия. Более того, он обладает очевидным преимуществом перед критерием, использующем U-статистику: пороговое значение (7.21) в нем не требует асимптотики по объемам выборок. Поэтому для проверки гипотезы о совпадении векторов математических ожиданий в случае двух выборок целесообразнее использовать критерий из п. 7.3.1.

7.5.2 Проверка гипотезы о равенстве ковариационных матриц

Проверим гипотезы о равенстве ковариационных матриц:

$$\begin{array}{rcl} H_0 & : & \Sigma_1 = \ldots = \Sigma_L, \\ H_1 & : & \overline{H_0}. \end{array}$$

Статистика отношения правдоподобия в этом случае имеет вид

$$\lambda = \frac{\max\limits_{\mu_1,\ldots,\mu_L,\Sigma} L(\{\mu_i, \Sigma_i = \Sigma\}_{i=1}^L)}{\max\limits_{\{\mu_i,\Sigma_i\}_{i=1}^L} L(\{\mu_i, \Sigma_i\}_{i=1}^L)} =$$

$$= \frac{\max\limits_{\mu_1,\ldots,\mu_L,\Sigma} L(\{\mu_i, \Sigma_i = \Sigma\}_{i=1}^L)}{\prod_{i=1}^L \max\limits_{\mu_i,\Sigma_i} L_N^{(i)}(\mu_i, \Sigma_i)} = \frac{(2\pi)^{-\frac{nN}{2}} |\frac{1}{n}\sum_{i=1}^L A_i|^{-n/2} e^{-\frac{nN}{2}}}{\prod_{i=1}^L (2\pi)^{-\frac{n_i N}{2}} |\frac{1}{n_i} A_i|^{-n_i/2} e^{-\frac{n_i N}{2}}} =$$

$$= \frac{n^{\frac{nN}{2}}}{\prod_{i=1}^L n_i^{\frac{n_i N}{2}}} \frac{\prod_{i=1}^L |A_i|^{n_i/2}}{|\sum_{i=1}^L A_i|^{n/2}}.$$

Введем в рассмотрение «подправленную» статистику, предложенную Бартлеттом [1]:

$$\lambda^* = \frac{m^{\frac{mN}{2}}}{\prod_{i=1}^L m_i^{\frac{m_i N}{2}}} \frac{\prod_{i=1}^L |A_i|^{m_i/2}}{|\sum_{i=1}^L A_i|^{m/2}},$$

где $m_i = n_i - 1$ $(i = \overline{1, L})$, $m = \sum_{i=1}^L m_i = n - L$, и построим критерий, асимптотически эквивалентный критерию отношения правдоподобия:

$$\left\{ \begin{array}{rcl} H_0 & : & -2\rho \ln \lambda^* \le \Delta(\alpha), \\ H_1 & : & -2\rho \ln \lambda^* > \Delta(\alpha). \end{array} \right.$$

где

$$\rho = 1 - \left(\sum_{i=1}^L \frac{1}{m_i} - \frac{1}{m} \right) \frac{2N^2 + 3N - 1}{6(N+1)(L-1)}$$

и установлено [1]:

$$\mathcal{L}_{H_0}\{-2\rho \ln \lambda^*\} \to \chi^2_f, \ f = \frac{L-1}{2} N(N+1), \quad n_i \to \infty, \ i = \overline{1, L},$$

что позволяет по уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$ определить порог:

$$\Delta(\alpha) = F^{-1}_{\chi^2_f}(1 - \alpha), \quad f = \frac{L-1}{2} N(N+1).$$

7.5.3 Гипотеза об эквивалентности нормальных распределений (гипотеза однородности)

Проверим гипотезы о том, что наблюдения из всех выборок имеют одно и то же нормальное распределение:

$$\begin{array}{rcl} H_0 & : & \mu_1 = \ldots = \mu_L, \quad \Sigma_1 = \ldots = \Sigma_L, \\ H_1 & : & \overline{H_0}. \end{array}$$

Запишем статистику отношения правдоподобия:

$$\lambda = \frac{\max\limits_{\mu,\Sigma} L(\{\mu_i = \mu, \Sigma_i = \Sigma\}_{i=1}^{L})}{\max\limits_{\{\mu_i,\Sigma_i\}_{i=1}^{L}} L(\{\mu_i, \Sigma_i\}_{i=1}^{L})} =$$

$$= \frac{(2\pi)^{-\frac{nN}{2}} |\frac{1}{n}A|^{-n/2} e^{-\frac{nN}{2}}}{\prod_{i=1}^{L}(2\pi)^{-\frac{n_i N}{2}} |\frac{1}{n_i}A_i|^{-n_i/2} e^{-\frac{n_i N}{2}}} = \frac{n^{\frac{nN}{2}}}{\prod_{i=1}^{L} n_i^{\frac{n_i N}{2}}} \frac{\prod_{i=1}^{L} |A_i|^{n_i/2}}{|A|^{n/2}}.$$

Определим соответствующую «подправленную» статистику:

$$\overline{\lambda}^* = \frac{m^{\frac{mN}{2}}}{\prod_{i=1}^{L} m_i^{\frac{m_i N}{2}}} \frac{\prod_{i=1}^{L} |A_i|^{m_i/2}}{|A|^{m/2}},$$

и построим критерий, асимптотически эквивалентный критерию отношения правдоподобия и имеющий вид

$$\begin{cases} H_0 & : \quad -2\rho^* \ln \overline{\lambda}^* \le \Delta(\alpha), \\ H_1 & : \quad -2\rho^* \ln \overline{\lambda}^* > \Delta(\alpha). \end{cases}$$

где

$$\rho^* = 1 - \left(\sum_{i=1}^{L} \frac{1}{m_i} - \frac{1}{m} \right) \frac{2N^2 + 3N - 1}{6(N+3)(L-1)} - \frac{1}{m} \frac{N - L + 2}{N + 3}$$

и также установлено [1]:

$$\mathcal{L}_{H_0}\{-2\rho^* \ln \overline{\lambda}^*\} \to \chi^2_f, \ f = \frac{L-1}{2} N(N+3), \quad n_i \to \infty, \ i = \overline{1, L},$$

что позволяет по $\alpha = \mathbf{P}\{H_1 | H_0\} \in (0, 1)$ определить порог:

$$\Delta(\alpha) = F^{-1}_{\chi^2_f}(1 - \alpha), \quad f = \frac{L-1}{2} N(N+3).$$

Список литературы

[1] *Андерсон, Т.* Введение в многомерный статистический анализ /Т. Андерсон. М. : Физматгиз, 1963.

[2] *Андерсон, Т.* Статистический анализ временных рядов /Т. Андерсон. М. : Мир, 1976.

[3] *Бокс, Дж.* Анализ временных рядов. Прогноз и управление /Дж. Бокс, Г. Дженкинс. М. : Мир, 1974. Вып. 1.

[4] *Боровков, А.А.* Теория вероятностей /А. А. Боровков. М. : Наука, 1998.

[5] *Гантмахер, Ф.Р.* Теория матриц /Ф.Р. Гантмахер. М. : Наука, 1967.

[6] *Гнеденко, Б.В.* Курс теории вероятностей /Б. В. Гнеденко М.: Физматгиз, 1988.

[7] *Ивченко Г. И.* Введение с математическую статистику / Г. И. Ивченко, Ю. И. Медведев М.: Издательство ЛКИ, 2014.

[8] *Харин, Ю. С.* Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. — Минск : БГУ, 2011.

[9] *Хорн, Р.* Матричный анализ / Р. Хорн, Ч. Джонсон. М. : Мир, 1989.

[10] *Ширяев, А.Н.* Вероятность /А.Н. Ширяев. М. : Наука, 1989.

[11] *Харин, Ю.С.* Теория вероятностей /Ю.С. Харин, Н.М. Зуев. Мн. : БГУ, 2004.

[12] Харин, Ю. С. Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. – Минск : БГУ, 2011.