Московский государственный
университет имени М.В. Ломоносова

Механико-математический
факультет

Ю. Н. Тюрин

МНОГОМЕРНАЯ СТАТИСТИКА:

ГАУССОВСКИЕ ЛИНЕЙНЫЕ МОДЕЛИ

Издательство
Московского университета
2011

УДК 517.9575
ББК 22.16
Т98

Тюрин Ю.Н.
Т98 Многомерная статистика: гауссовские линейные модели. — М.: Издательство Московского университета, 2011. — 136 с.
ISBN 978-5-211-05915-3

Книга излагает наиболее разработанную к настоящему времени статистику гауссовских (т.е. нормально распределённых) случайных величин.

Ядро книги составляет общая теория многомерных линейных моделей, представленная геометрически. Она единым образом рассматривает до того изучавшиеся порознь их конкретные формы (дисперсионный анализ, регрессионный анализ). Математическим аппаратом служат модули над кольцами квадратных матриц, наделённые матрично-значным скалярным умножением. Для многомерных данных эта структура замещает векторную алгебру. На базе линейных моделей и нового понятия матричной корреляции изложена корреляционная теория.

От читателя ожидается владение математическим анализом, линейной алгеброй, а также основами теории вероятностей и математической статистики. Книга может быть полезна всем интересующимся математической статистикой, в особенности студентам и аспирантам математических и экономических факультетов. Книга может быть основой семестрового курса лекций.

Ключевые слова: Многомерное нормальное распределение. Таблицы многомерных данных. Модули таблиц над кольцами квадратных матриц. Подмодули. Матрично-значное скалярное умножение. Многомерные линейные модели. Линейные гипотезы. Проверка линейных гипотез. Матричный коэффициент корреляции. Проверка независимости многомерных случайных переменных.

УДК 517.9575
ББК ББК 22.16

ISBN 978-5-211-05915-3

ОГЛАВЛЕНИЕ

Предисловие

Многомерный статистический анализ – это часть математической статистики, которая изучает многомерные наблюдения. Наблюдение называют многомерным, если в этом наблюдении измеряют несколько показателей. Слово измерить здесь означает дать числовое значение. В традиционной математической статистике измеряемые показатели считаются случайными величинами; наблюдённые значения показателей суть реализации этих случайных величин. В книге мы будем держаться этой точки зрения.

Поскольку упомянутые случайные величины возникают в одном эксперименте, они имеют совместное распределение вероятностей. Совместное распределение вероятностей отражает взаимную зависимость показателей. Именно учёт этой взаимной зависимости выделяет многомерную статистику и отделяет её, к примеру, от изучения случайных показателей порознь.

Ядро многомерной статистики составляет так называемая гауссовская статистика, когда совместное распределение признаков является многомерным нормальным (гауссовским) распределением. Задачи о положении составляют наиболее важную практически и наиболее изученную теоретически часть многомерной статистики. Среди задач о положении наиболее востребованы и популярны линейные модели. Линейные статистические модели вместе с линейными гипотезами составляют основное содержание книги.

Для одномерных наблюдений теория линейных моделей использует геометрические понятия и средства линейной алгебры и евклидовых пространств. Для статистики многомерных наблюдений этих средств недостаточно.

В этой книге теория линейных статистических моделей основана на новом понятии таблицы (наблюдений) и на новой алгебраической структуре. Эта структура – модуль таблиц над кольцом квадратных матриц, наделённый матричнозначным скалярным (внутренним) умножением. С помощью этих средств оказалось возможным дать общее, геометрическое определение линейной статистической модели и линейной статистической гипотезы, дать многомерную (матричную) форму методу наименьших квадратов, ввести в пространство таблиц геометрические понятия ортогональности, проекции, длины и т.д. Всё это поставило теорию многомерной статистики в параллель классическому линейному статистическому анализу.

Корреляционный анализ составляет ещё одну важную часть многомерной статистики. Корреляционный анализ здесь тоже излагается по-новому, на основе нового понятия матричного коэффициента корреляции (введён и исследован в работах Е. Сухановой). Матричный коэффициент корреляции является многомерным аналогом классического коэффициента корреляции и сохраняет многие его черты и свойства. Его теоретические и выборочные свойства и упомянутые аналогии в книге подробно прослежены. Матричный коэффициент корреляции позволяет свести воедино многие известные понятия корреляционного анализа: канонические корреляции, канонические величины, множественную корреляцию и т.д. Благодаря тесной связи матричной корреляции с линейными статистическими моделями удалось ясно и лаконично рассказать о проверках по выборкам гипотез о статистической независимости многомерных признаков.

Теперь несколько подробнее о содержании книги. Многомерные нормальные распределения составляют вероятностную основу многомерного статистического анализа. Определению и свойствам многомерного нормального распределения посвящена гл. 1. Изложение здесь традиционно. Главы 2 и 3 посвящены изучению нормальной выборки: гл. 2 – оцениванию параметров, гл. 3 – проверке статистических гипотез (о математических ожиданиях). При проверке гипотез применены оба ныне известных метода для поиска статистических критериев – метод Роя (S.N. Roy) и метод отношения правдоподобий. Применительно к выборке оба метода дают один и тот же результат – статистику Хотеллинга. Вывод её распределения завершает гл. 3. В следующей главе (гл. 4) рассматривается более сложный вопрос – задача о нескольких выборках. Здесь при проверке гипотезы однородности упомянутые методы дают разные статистические правила. Так возникает вопрос об общем виде статистических критериев. Ответ на него будет получен позже, в гл. 6, в рамках проверки общих линейных гипотез.

Традиционное изложение следующих по сложности вопросов, например, линейной регрессии, на наш взгляд, становится неудовлетворительным. Поэтому следующая глава, гл. 5, посвящена построению подходящего к нашим статистическим данным математического аппарата – теории модулей над кольцом квадратных матриц, причём эти модули наделены матричнозначным скалярным умножением. Квадратные матрицы для этих модулей играют роль скаляров.

В гл. 6 на этой основе строится общая теория многомерных линейных моделей, использующая геометрические понятия координат, базисов, ортогональности, проецирования, ортогональных преобразований и прочего. В главе решаются основные для линейных статистических моделей задачи – оценивания параметров линейных моделей и проверки линейных гипотез. Для оценивания развит матричный метод наименьших квадратов. Как пример действия теории изучается задача многомерной множественной линейной регрессии, а также факторные модели (часто называемые многомерным дисперсионным анализом).

Последняя глава, гл. 7, посвящена корреляционной теории. Эта теория излагается с помощью уже упоминавшихся матричных коэффициентов корреляции. Матричная корреляция была введена и изучена Е.М. Сухановой в её кандидатской диссертации. В эту главу вошло многое из этой диссертации. Специальное внимание уделено статистической проверке гипотезы о независимости многомерных признаков. Изложение последнего вопроса опирается на свойства нормального распределения и ранее произведённый статистический анализ многомерной регрессии.

Книга сложилась на основе курсов лекций, прочитанных автором на механико-математическом факультете МГУ. Для чтения книги достаточно знания базовых курсов математического анализа, линейной алгебры, теории вероятностей и математической статистики, в особенности теории линейных статистических моделей.

Е.М. Суханова оказала автору очень большую помощь в работе над всей книгой. Приношу ей свою глубокую благодарность. Я также благодарен моим слушателям, студентам механико-математического факультета МГУ. Они с готовностью помогали мне в многочисленных переделках и оформлении рукописи. Я вспоминаю их с неизменной симпатией и жалею лишь о том, что молодые люди взрослеют и покидают нашу «волшебную гору».

Я признателен проф. Н.П. Трифонову и нс. В.И. Громыко за ценные советы, издательству МГУ за тщательную работу с рукописью.

Глава 1

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Все результаты этой книги опираются на понятие многомерного нормального распределения и на свойства многомерных случайных величин, подчиняющихся многомерному нормальному распределению. Нормальное, или гауссовское, распределение вероятностей в многомерном пространстве естественным образом обобщает нормальное распределение на числовой прямой (одномерное нормальное распределение вероятностей) и во многом повторяет его свойства. Предполагается, что с одномерными нормальными распределениями и случайными величинами читатель хорошо знаком.

§ 1. Определение

Скажем, что p-мерный случайный вектор $Y = (Y_1, \ldots, Y_p)^T$ имеет *p-мерное стандартное нормальное распределение*, если его координаты $Y_1, \ldots, Y_p$ – это независимые одномерные стандартные нормально распределенные случайные величины ($\sim N(0,1)$). Распределение случайного вектора Y (а иногда и сам вектор Y) будем обозначать $N_p(0, I)$, где I – единичная матрица; в данном случае это единичная $(p \times p)$-матрица. Основания для такого обозначения станут ясны вскоре.

Скажем, что случайный вектор X имеет *нормальное распределение*, если X можно представить в виде $X = a + BY$, где случайный вектор Y имеет стандартное нормальное распределение, a – неслучайный вектор, B – неслучайная матрица (размерности a, B и Y таковы, что указанные операции осуществимы). Легко видеть, что

$$\mathrm{E}\,X = a, \qquad \mathrm{Var}\,X = \mathrm{D}\,X = \mathrm{E}(X - \mathrm{E}\,X)(X - \mathrm{E}\,X)^T = BB^T \geqslant 0.$$

Положим $Q = BB^T$. Распределение p-мерного случайного нормального вектора $X = a + BY$ будем обозначать $N_p(a, Q)$.

Из этих определений следует, что аффинные преобразования нормально распределенных (синоним – гауссовских) случайных векторов тоже нормально распределенные векторы.

§ 2. Характеристическая функция распределения

По определению *характеристической функцией* случайного вектора X (для определенности – p-мерного вектора-столбца) называют функцию векторного аргумента u (тоже p-мерного вектора-столбца)

$$\varphi(u) = \varphi_X(u) = \mathrm{E} \exp\{iu^T X\}.$$

Свойства многомерных характеристических функций во многом повторяют свойства характеристических функций одномерных случайных величин (которые предполагаются известными), а потому в доказательном изложении не нуждаются. Все же ради некоторой систематичности рассказа укажем некоторые из них (те, которые мы будем использовать в дальнейшем).

Некоторые свойства характеристических функций.

1. Функция $\varphi_X(\cdot)$ существует для всякой p-мерной случайной величины X, непрерывна, причем $\varphi(0) = 1$, $|\varphi(u)| \leqslant 1$ для всех $u \in \mathbb{R}^p$.
2. Характеристические функции φ_X и φ_Y случайных векторов $X = a + BY$ и Y связаны соотношением

   $$\varphi_X(u) = \exp\{iu^T a\}\varphi_Y(u^T B).$$

   (Здесь a, B – неслучайные вектор и матрица; размерности a, B и Y таковы, что указанное аффинное преобразование Y в X возможно.)
3. Если X и Y – независимые p-мерные случайные величины, то

   $$\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u).$$

4. Различным распределениям вероятностей соответствуют различные характеристические функции.
5. Последовательность $F_n, n = 1, 2, \ldots$ распределений вероятностей в $\mathbb{R}^p$ сходится слабо к распределению вероятностей F тогда и только тогда, когда последовательность φ_n, $n = 1, 2, \ldots$ соответствующих характеристических функций сходится к непрерывной предельной функции φ. В этом случае φ есть характеристическая функция распределения F, а сходимость $\varphi_n(u) \to \varphi(u)$ равномерна на каждом компакте.

Следствие. *Непрерывная функция, являющаяся пределом (в смысле точечной сходимости) последовательности характеристических функций, сама является характеристической.*

Вычислим характеристическую функцию многомерного нормального закона. Вспомним, что характеристическая функция одномерного нормального распределения $N(a, \sigma^2)$ равна

$$\varphi(t) = \exp\{iat\} \exp\Big\{-\frac{t^2\sigma^2}{2}\Big\}, \quad t \in \mathbb{R}.$$

В частности, характеристическая функция стандартного нормального закона $N(0, 1)$ равна

$$\varphi(t) = \exp\Big\{-\frac{t^2}{2}\Big\}, \quad t \in \mathbb{R}.$$

Найдем характеристическую функцию произвольного многомерного нормального распределения. Как было определено в § 1, нормальный (нормально распределенный) случайный вектор X можно представить в виде линейной функции $X = a + BY$ стандартного нормального вектора Y. Ни размерность X, ни размерность Y обозначать на письме не будем – они могут быть произвольными. Но размерности a, B и Y должны быть согласованы. Характеристическая функция X:

$$\varphi_X(u) = \mathrm{E} \exp\{iu^T X\} = \exp\{iu^T a\} \,\mathrm{E} \exp\{i(u^T B)Y\}.$$

Координаты вектора $u^T B$ обозначим через $(u^T B)_1, (u^T B)_2, \ldots$; координаты вектора Y обозначим через $Y_1, Y_2, \ldots$:

$$u^T B = ((u^T B)_1, (u^T B)_2, \ldots), \qquad Y = (Y_1, Y_2, \ldots)^T B.$$

Согласно отмеченным выше свойствам случайных величин $Y_1, Y_2, \ldots$,

$$\begin{aligned}\mathrm{E} \exp\{i(u^T B)Y\} &= \prod_j \mathrm{E} \exp\{i(u^T B)_j Y_j\} = \prod_j \exp\Big\{-\frac{1}{2}\left[(u^T B)_j\right]^2\Big\} \\ &= \exp\Big\{-\frac{1}{2}\sum_j \left[(u^T B)_j\right]^2\Big\} = \exp\left\{-\frac{1}{2}(u^T B)(u^T B)^T\right\} \\ &= \exp\left\{-\frac{1}{2}u^T (BB^T) u\right\}.\end{aligned}$$

Мы уже отмечали, что $\mathrm{E}\, X = a$, $\mathrm{D}\, X = BB^T = Q$. Таким образом, доказана

Теорема 1.1. *Характеристическая функция многомерного нормального закона $N(a, Q)$ дается формулой*

$$\varphi(u) = \exp\{iu^T a\} \exp\left\{-\frac{1}{2}u^T Q u\right\}.$$

В силу взаимнооднозначного соответствия между распределениями и их характеристическими функциями мы можем сказать, что многомерные нормальные распределения полностью определяются первыми двумя моментами – вектором математического ожидания и матрицей ковариаций. (В точности так же, как в одномерном случае.)

§ 3. Распределение $N_p(a, Q)$

Пусть $a \in \mathbb{R}^p$ – произвольный вектор, $Q \geqslant 0$ – произвольная неотрицательно определенная матрица $p \times p$.

Теорема 1.2. *Существует p-мерное нормальное распределение, для которого вектор a служит математическим ожиданием, а Q – матрицей ковариаций.*

Доказательство. (Его первая часть будет использована далее, в § 4). Согласно теореме из линейной алгебры о приведении квадратичной формы к каноническому виду, существует невырожденная матрица P такая, что PQP^T – диагональная матрица, причем на диагонали стоят числа $+1, 0, -1$. Так как $Q \geqslant 0$, то -1 исключается. Следовательно,

$$PQP^T = I_{(m)},$$

где $I_{(m)}$ обозначает диагональную $p \times p$-матрицу, на главной диагонали которой стоят $m \leqslant p$ единиц, а прочие диагональные элементы (если есть) – нули. В частности, если $m = p$, то $I_{(m)}$ – единичная матрица. Легко видеть, что m равно рангу Q.

Из этой формулы находим, что

$$Q = P^{-1} I_{(m)} \left(P^T\right)^{-1} = \left(P^{-1} I_{(m)}\right) \left(P^{-1} I_{(m)}\right)^T.$$

Положим $B = P^{-1} I_{(m)}$ и введем случайный вектор $X = a + BY$, где $Y \sim \sim N_p(0, I)$. Ясно, что X – гауссовский вектор и что $\mathrm{E}\, X = a$, $\mathrm{D}\, X = Q$. Теорема доказана. $\square$

§ 4. Плотность распределения

Рассмотрим произвольное нормальное распределение $N_p(a, Q)$. Если ранг матрицы Q меньше, чем ее размерность, то распределение $N_p(a, Q)$ сосредоточено в некотором линейном многообразии, размерность которого равна рангу Q. По сути, доказательство этого факта содержится в § 3. (Этот факт верен не только для нормального распределения; см. также упражнение 2 к гл. 1.) Поэтому вырожденное (когда вырождена матрица Q) гауссовское распределение не имеет плотности относительно лебеговой меры в $\mathbb{R}^p$.

Теорема 1.3. *Пусть $X \sim N_p(a, Q)$, причем $Q > 0$ (т.е. матрица Q положительно определена и потому не вырождена). В таком случае распределение $N_p(a, Q)$ имеет плотность, и эта плотность в точке $y \in \mathbb{R}^p$ равна*

$$p_X(y; a, Q) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det Q}} \exp\left\{-\frac{1}{2}(y-a)^T Q^{-1}(y-a)\right\}.$$

Доказательство. Как следует из § 1, случайный вектор X в этом случае можно представить в виде

$$X = a + BY,$$

причем B – невырожденная квадратная матрица, $BB^T = Q$, $Y \sim N_p(0, I)$. Отметим, что в этом случае (когда $Q > 0$) данное преобразование обратимо:

$$Y = B^{-1}(X - a).$$

Из определения стандартного нормального распределения в пункте 1 следует, что $N_p(0, I)$ имеет плотность в точке $x \in \mathbb{R}^p$ (относительно меры Лебега в $\mathbb{R}^p$)

$$p_Y(x; 0, I) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \exp\left(-\frac{1}{2}x^T x\right).$$

Для вывода отсюда плотности случайной величины X нужна следующая

Лемма 1. *Пусть Y, X – случайные величины со значениями в $\mathbb{R}^p$, причем $X = f(Y)$, $f(\cdot)$ – дифференцируемое взаимнооднозначное преобразование $\mathbb{R}^p$ в $\mathbb{R}^p$. Допустим, что Y имеет плотность $p_Y(\cdot)$ относительно меры Лебега. В таком случае X тоже имеет плотность*

$$p_X(y) = p_Y(x)\left|\frac{df(x)}{dx}\right|,$$

где $y = f(x)$, $\left|\frac{df(x)}{dx}\right|$ обозначает абсолютное значение якобиана преобразования $y = f(x)$ в точке x.

Доказательство. Обозначим через U_x окрестность точки $x \in \mathbb{R}^p$. По определению плотности

$$P(Y \in U_x) = p_Y(x)\,|U_x| + o(|U_x|),$$

где $|U_x|$ обозначает объем (меру Лебега) области U_x, диаметр которой стремится к нулю. Так как $f(\cdot)$ – взаимнооднозначная функция, то $P(Y \in U_x) = P(f(Y) \in f(U_x))$, где $f(U_x)$ обозначает образ U_x при преобразовании $f(\cdot)$, т.е. некоторую окрестность точки $y = f(x)$. Еще раз обращаясь к определению плотности, запишем ту же вероятность в виде

$$P(Y \in U_x) = p_X(y)\,|f(U_x)| + o(|f(U_x)|).$$

Заметим, что $|f(U_x)| = |U_x| \cdot \left|\frac{df(x)}{dx}\right| + o(|U_x|)$. Сопоставляя два выражения для $P(Y \in U_x)$, получим утверждение леммы 1. □

Применим эту лемму к преобразованию $f(Y) = a + BY$. Это линейное преобразование, и потому его якобиан $\left|\frac{df}{dx}\right|$ для каждого x равен $\det B$. Так как $Q = BB^T$, то $\det B = \sqrt{\det Q}$. Остается применить результат леммы 1 к плотности $p_Y(x)$, полагая $x = B^{-1}(y-a)$. Это дает для плотности случайного вектора X выражение

$$p_X(y; a, Q) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det Q}} \exp\left\{-\frac{1}{2}\left[B^{-1}(y-a)\right]^T \left[B^{-1}(y-a)\right]\right\}.$$

Так как $\left(B^{-1}\right)^T B^{-1} = \left(BB^T\right)^{-1} = Q^{-1}$, получаем отсюда искомое выражение для плотности невырожденного многомерного нормального распределения. □

§ 5. Независимость гауссовских величин

Рассмотрим разбиение вектора X на подвекторы и соответствующие разбиения $\mathrm{E}\,X = a$ и $\mathrm{D}\,X = Q$:

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \qquad a = \begin{pmatrix} a^{(1)} \\ a^{(2)} \end{pmatrix}, \qquad Q = \begin{pmatrix} Q_{11} Q_{12} \\ Q_{21} Q_{22} \end{pmatrix}.$$

Уточняем: для $i, j = 1, 2$

$$a^{(i)} = \mathrm{E}\,X^{(i)}, \qquad Q_{ij} = \mathrm{E}(X^{(i)} - a^{(i)})(X^{(j)} - a^{(j)})^T.$$

Если $Q_{12} = 0$, то говорят, что $X^{(1)}$ и $X^{(2)}$ *не коррелированы*. Ясно, что независимые случайные величины всегда не коррелированы. Обратное, вообще говоря, неверно. Это справедливо тем не менее для подвекторов гауссовского вектора.

Теорема 1.4. *Пусть $X \sim N_p(a, Q)$. Если $Q_{12} = 0$, то $X^{(1)}$ и $X^{(2)}$ независимы.*

Доказательство этой теоремы будет очень коротким, если вспомнить одну теорему о характеристических функциях и одну лемму из алгебры матриц.

Теорема (см. [13]). *Для того чтобы компоненты случайного вектора были независимы, необходимо и достаточно, чтобы его характеристическая функция была произведением характеристических функций компонент.*

Лемма 2 (о перемножении блочных матриц).

$$\begin{pmatrix} A_{11}, A_{12} \\ A_{21}, A_{22} \end{pmatrix} \times \begin{pmatrix} B_{11}, B_{12} \\ B_{21}, B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21}, A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21}, A_{21}B_{12} + A_{22}B_{22} \end{pmatrix},$$

при условии, что размерности всех участвующих в формуле матриц согласованы.

Вольно выражаясь, можно передать содержание этой леммы так: блочные матрицы перемножают по правилу «строка на столбец», как и матрицы с числовыми элементами.

Доказательство теоремы 1.4. Заметим, что случайные векторы $X^{(1)}$ и $X^{(2)}$ распределены каждый по нормальному закону (как линейные функции от гауссовского вектора X):

$$X^{(1)} \sim N(a^{(1)}, Q_{11}), \qquad X^{(2)} \sim N(a^{(2)}, Q_{22}).$$

Рассмотрим характеристическую функцию случайного вектора X:

$$\varphi_X(u) = \exp\{iu^T a\} \exp\Big\{-\frac{1}{2}u^T Q u\Big\}.$$

Разобьем переменный вектор $u \in \mathbb{R}^p$ на два подвектора $u^{(1)}$ и $u^{(2)}$ (так же, как разделен вектор X). Заметим, что в силу леммы 2 и условия $Q_{12} = 0$,

$$\begin{aligned} u^T a &= u^{(1)T} a^{(1)} + u^{(2)T} a^{(2)}, \\ u^T Q u &= u^{(1)T} Q_{11} u^{(1)} + u^{(1)T} Q_{12} u^{(2)} + u^{(2)T} Q_{21} u^{(1)} + u^{(2)T} Q_{22} u^{(2)} \\ &= u^{(1)T} Q_{11} u^{(1)} + u^{(2)T} Q_{22} u^{(2)}. \end{aligned}$$

Следовательно,

$$\varphi_X(t) = \Big[\exp\{iu^{(1)T} a^{(1)}\} \exp\{-\frac{1}{2}u^{(1)T} Q_{11} u^{(1)}\}\Big]$$

$$\times \left[\exp\{iu^{(2)T}a^{(2)}\}\exp\{-\frac{1}{2}u^{(2)T}Q_{22}u^{(2)}\}\right].$$

Сомножителями здесь являются характеристические функции распределений $N(a^{(1)}, Q_{11})$ и $N(a^{(2)}, Q_{22})$. В силу цитированной теоремы получаем, что $X^{(1)}$ и $X^{(2)}$ независимы. □

§ 6. Условные распределения

Рассмотрим гауссовский вектор $X \sim N_p(a, Q)$, который разделим на подвекторы $X^{(1)}$ и $X^{(2)}$, как в § 5, и используем те же обозначения. Мы покажем, что условное распределение $X^{(1)}$ при данном значении $X^{(2)}$ гауссовское, и найдем параметры этого гауссовского распределения. Для этого мы вычислим условную характеристическую функцию $X^{(1)}$ при данном $X^{(2)}$. Мы будем использовать понятия условного математического ожидания. Нам понадобятся некоторые его свойства, из которых напомним два. Пусть Z_1, Z_2 – случайные величины, определенные на общем пространстве элементарных исходов.

(a) Если $g(\cdot)$ – измеримая функция, то

$$\mathrm{E}(Z_1 g(Z_2) \mid Z_2) = g(Z_2)\,\mathrm{E}(Z_1 \mid Z_2).$$

В частности,

$$\mathrm{E}(g(Z_2) \mid Z_2) = g(Z_2).$$

(b) Если Z_1 и Z_2 независимы, то

$$\mathrm{E}(Z_1 \mid Z_2) = \mathrm{E}(Z_1).$$

Начнем с того, что представим $X^{(1)}$ и $X^{(2)}$ в удобной форме. Ограничимся случаем $Q_{22} > 0$. Введем вектор

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

который получаем линейным преобразованием X по формулам

$$Y^{(1)} = X^{(1)} + TX^{(2)}, \qquad Y^{(2)} = X^{(2)}$$

(матрица T такова, что указанные преобразования возможны).

Выберем T так, чтобы векторы $Y^{(1)}$ и $Y^{(2)}$ оказались не коррелированы. Поскольку совместное распределение $Y^{(1)}$ и $Y^{(2)}$ гауссовское, то при таком выборе T эти векторы будут независимы.

Ковариация $Y^{(1)}$ и $Y^{(2)}$:

$$\begin{aligned}
&\mathrm{E}(Y^{(1)} - \mathrm{E}\,Y^{(1)})(Y^{(2)} - \mathrm{E}\,Y^{(2)})^T \\
&\qquad\qquad = \mathrm{E}[(X^{(1)} - a^{(1)}) + T(X^{(2)} - a^{(2)})](X^{(2)} - a^{(2)})^T \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = Q_{12} + TQ_{22}.
\end{aligned}$$

Приравняв это выражение нулю, найдем, что $Y^{(1)}$ и $Y^{(2)}$ независимы при

$$T = -Q_{12}Q_{22}^{-1}.$$

Обратите внимание, что число столбцов в матрице Q_{12} равно числу строк в матрице Q_{22}^{-1}, и потому их перемножение возможно. Теперь $X^{(1)}$ и $X^{(2)}$ линейно выражаются через независимые гауссовские векторы $Y^{(1)}$ и $Y^{(2)}$:

$$X^{(1)} = Y^{(1)} + Q_{12}Q_{22}^{-1}Y^{(2)}, \quad X^{(2)} = Y^{(2)}.$$

Далее, вектор $Y^{(1)}$ распределен по нормальному закону, а его параметры суть

$$\begin{aligned}
\mathrm{E}\,Y^{(1)} &= a^{(1)} - Q_{12}Q_{22}^{-1}a^{(2)}, \\
\mathrm{D}\,Y^{(1)} &= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}.
\end{aligned}$$

Последний результат получаем прямым вычислением. Действительно, так как матрица ковариаций не изменяется при изменениях вектора математических ожиданий a и, стало быть, не зависит от a, то в последующих вычислениях можно положить $a = 0$. В таком случае имеем

$$\begin{aligned}
\mathrm{D}\,Y^{(1)} &= \mathrm{E}(X^{(1)} - Q_{12}Q_{22}^{-1}X^{(2)})(X^{(1)} - Q_{12}Q_{22}^{-1}X^{(2)})^T \\
&= \mathrm{E}(X^{(1)}X^{(1)T}) - Q_{12}Q_{22}^{-1}\,\mathrm{E}(X^{(2)}X^{(1)T}) - \mathrm{E}\,X^{(1)}(X^{(2)T}Q_{12}^TQ_{22}^{-1}) \\
&\qquad\qquad + Q_{12}Q_{22}^{-1}(\mathrm{E}\,X^{(2)}X^{(2)T})Q_{22}^{-1}Q_{12}^T,
\end{aligned}$$

что после упрощений дает указанную выше формулу – если учесть, что $Q_{12} = Q_{21}^T$.

Чтобы найти условное распределение $X^{(1)}$ при данном $X^{(2)}$ – далее это $\mathcal{L}(X^{(1)} \mid X^{(2)})$, обратимся к условной характеристической функции $X^{(1)}$ при данном $X^{(2)}$:

$$\varphi(u) = \mathrm{E}(\exp\{iu^TX^{(1)}\} \mid X^{(2)}).$$

Выразим $X^{(1)}$ и $X^{(2)}$ через $Y^{(1)}$ и $Y^{(2)}$. Тогда

$$\begin{aligned}\varphi(u) &= \mathrm{E}(\exp\{iu^T(Y^{(1)}+Q_{12}Q_{22}^{-1}Y^{(2)})\} \mid Y^{(2)}) \\ &= (\mathrm{E}\exp\{iu^TY^{(1)}\})\exp\{iu^TQ_{12}Q_{22}^{-1}Y^{(2)}\} \\ &= \exp\{iu^T(a^{(1)}-Q_{12}Q_{22}^{-1}a^{(2)})+iu^TQ_{12}Q_{22}^{-1}Y^{(2)}\} \\ &\qquad\times\exp\{-\frac{1}{2}u^T(Q_{11}-Q_{12}Q_{22}^{-1}Q_{21})u\} \\ &= \exp\{iu^T(a^{(1)}+Q_{12}Q_{22}^{-1}(X^{(2)}-a^{(2)}))\} \\ &\qquad\times\exp\Big\{-\frac{1}{2}u^T(Q_{11}-Q_{12}Q_{22}^{-1}Q_{21})u\Big\},\end{aligned}$$

если вспомнить, что $Y^{(2)}=X^{(2)}$. Из этой формулы следует, что $\mathcal{L}(X^{(1)}|X^{(2)})$ гауссовское, причем

$$\mathrm{E}(X^{(1)} \mid X^{(2)}) = \mathrm{E}\,X^{(1)} + Q_{12}Q_{22}^{-1}(X^{(2)}-\mathrm{E}\,X^{(2)})$$

– линейная функция от $X^{(2)}$,

$$\mathrm{D}(X^{(1)} \mid X^{(2)}) = Q_{11}-Q_{12}Q_{22}^{-1}Q_{21}$$

– неотрицательно определенная матрица, не зависящая от принятого $X^{(2)}$ значения.

§ 7. Линейная регрессия

Регрессией случайной переменной $X^{(1)}$ относительно переменной $X^{(2)}$ называют математическое ожидание $X^{(1)}$ при данном значении $X^{(2)}$. Переменная $X^{(2)}$ может быть как случайной, так и нет. Если переменная $X^{(2)}$ случайная (и обе переменные наблюдаются в одном эксперименте, т.е. обе они заданы на одном пространстве элементарных исходов), то регрессия $X^{(1)}$ по $X^{(2)}$ – это условное математическое ожидание $\mathrm{E}(X^{(1)} \mid X^{(2)})$.

Как мы только что видели, при совместном гауссовском распределении случайных переменных $X^{(1)}$ и $X^{(2)}$ регрессия $X^{(1)}$ по $X^{(2)}$ линейная. Более того, условное распределение $\mathcal{L}(X^{(1)} \mid X^{(2)})$ совпадает с распределением случайного вектора

$$a^{(1)}+Q_{12}Q_{22}^{-1}(X^{(2)}-a^{(2)})+Z,$$

где $X^{(2)}$ фиксирован, а вектор Z случаен, Z не зависит от $X^{(2)}$ и подчиняется гауссовскому распределению

$$Z \sim N(0, Q_{11}-Q_{12}Q_{22}^{-1}Q_{21}).$$

Другой способ передачи той же мысли: с вероятностью 1

$$X^{(1)} = a^{(1)} + Q_{12}Q_{22}^{-1}(X^{(2)} - a^{(2)}) + Z.$$

Вектор Z можно определить из этого выражения как

$$Z = (X^{(1)} - a^{(1)}) - Q_{12}Q_{22}^{-1}(X^{(2)} - a^{(2)}).$$

Выкладки § 6 показывают, что Z не зависит от $X^{(2)}$ и распределен, как было указано выше.

Эта форма записи условного распределения $X^{(1)}$ при данном $X^{(2)}$ – в виде линейной регрессии по $X^{(2)}$ с независимой от $X^{(2)}$ гауссовской «ошибкой» Z – соединяет условные гауссовские распределения с гауссовскими линейными моделями, которые будут изучены позже.

§ 8. Датчик нормальных чисел

Датчиком случайных чисел называют компьютерную программу, которая вырабатывает последовательность независимых случайных чисел $x_1, x_2, \ldots$, распределенных равномерно на отрезке $[0, 1]$. Так как вычислительная программа (алгоритм) может вырабатывать только строго закономерную последовательность, упомянутая последовательность $x_1, x_2, \ldots$ в точном смысле случайной не является. Она, однако, хорошо имитирует свойства случайных последовательностей, которые нужны для практических целей для вычислений методом Монте-Карло в первую очередь.

Для некоторых целей (для моделирования случайных процессов, например) бывают необходимы последовательности независимых гауссовских случайных величин. Поэтому далее мы поговорим о том, как последовательность независимых равномерно распределенных случайных чисел $x_1, x_2, \ldots$ превратить в последовательность независимых гауссовских случайных величин $y_1, y_2, \ldots$ Очевидный способ – переход от x_i к y_i с помощью функции квантилей $y_i = \Phi^{-1}(x_i)$ – не подходит, так как вычисление по этой формуле – процесс сложный и потому медленный. Между тем, компьютерная генерация последовательности $y_1, y_2, \ldots$ должна быть быстрой, так как для вычислений обычно требуются выборки большого объема. Широко распространенный сейчас метод генерации нормальных $N(0, 1)$-чисел основан на свойствах $N_2(0, I)$ (см. упражнение 3 к гл. 1) и возможности быстрого перехода от равномерно распределенной случайной величины к $\chi^2(2)$ – случайной величины хи-квадрат с двумя степенями свободы.

Известно, что плотность $\chi^2(2)$ как функция переменного $t \in \mathbb{R}$ равна $\exp\{-t/2\}/2$ для $t \geqslant 0$ (и равна 0 для $t < 0$). Эту формулу можно проверить прямым вычислением плотности $\chi^2(2) = y_1^2 + y_2^2$, где y_1, y_2 суть независимые $N(0,1)$. Функция дожития для $\chi^2(2)$ равна $P(\chi^2(2) \geqslant t) = \exp\{-t/2\}$.

Пусть теперь x – случайная величина, равномерно распределенная на $[0,1]$. Рассмотрим случайную величину $y = -2\ln x \in (0, \infty)$. Вычислим ее функцию дожития и убедимся, что она равна $\exp\{-t/2\}$ для $t \geqslant 0$. Действительно,

$$P(-2\ln x \geqslant t) = P\Big(\ln x \leqslant -\frac{t}{2}\Big) = P\Big(x \leqslant \exp\Big\{-\frac{t}{2}\Big\}\Big) = \exp\Big\{-\frac{t}{2}\Big\}.$$

Следовательно, $y = -2\ln x \stackrel{d}{=} \chi^2(2)$.

Теперь способ трансформации равномерных случайных чисел $x_1, x_2, \ldots$ в нормальные $y_1, y_2, \ldots$ очевиден. От пары x_{2k-1}, x_{2k}, где $k = 1, 2, \ldots$, переходим к паре $\varphi_k = 2\pi x_{2k-1}, r_k = \sqrt{-2\ln x_{2k}}$. Пару чисел φ_k, r_k рассматриваем как полярные координаты случайного вектора $N_2(0, I)$. Вычислив декартовы координаты этого вектора, получим пару y_{2k-1}, y_{2k}; $k = 1, 2, \ldots$. Это взаимно независимые случайные числа $N(0,1)$.

§ 9. Упражнения

1. Рассмотрим двумерный гауссовский вектор $X = (X_1, X_2)^T$. Положим $\mu_1 = \mathrm{E}\, X_1, \mu_2 = \mathrm{E}\, X_2, Q_1^2 = \mathrm{D}\, X_1, Q_2^2 = \mathrm{D}\, X_2$. Пусть $Q_1 > 0, Q_2 > 0$ и $\rho = \mathrm{Cov}(X_1, X_2)/(Q_1 Q_2)$.

   (a) Укажите формулу для плотности случайного вектора
   $$\left(\frac{X_1 - \mu_1}{Q_1}, \frac{X_2 - \mu_2}{Q_2}\right)^T.$$

   (b) Как выглядят линии уровня этой плотности при различных значениях ρ? В частности, при $\rho = 0$ и при приближении ρ к $+1$ и к -1.

2. Пусть X – p-мерный случайный вектор, причем ранг $\mathrm{D}\, X$ равен $m < p$. Покажите, что в этом случае вектор X с вероятностью 1 лежит в некотором линейном многообразии размерности m.
3. Пусть $X \sim N_p(0, I)$. Для определения X в сферических координатах надо указать направление φ (точку на единичной сфере) и длину r вектора X.

   Покажите, что:

   (a) случайные величины r и φ независимы;

(b) случайный вектор φ распределен на единичной сфере равномерно;

(c) случайная величина r^2 распределена по закону хи-квадрат с $(p-1)$ степенями свободы.

4. Пусть $X \sim N_p(a, Q)$.
Найти $\mathcal{L}(X \mid AX)$, где A – заданная матрица. В частности, найти условное распределение X при условии $\sum_{i=1}^{p} X_i = 0$.
5. Пусть $W(t), t \in [0,1]$ – винеровский процесс (т.е. гауссовский процесс с независимыми приращениями, $\mathrm{E}\,W(t) \equiv 0$, $\mathrm{D}\,W(t) = t$ для $t \in [0,1]$).
Покажите, что случайный процесс, получаемый из $W(\cdot)$ наложением условия $W(1) = 0$ (т.е. условный винеровский процесс, с условием $W(1) = 0$) это броуновский мост, т.е. такой гауссовский процесс $\beta(t), t \in \in [0,1]$, что $\mathrm{E}\,\beta(t) \equiv 0$, $\mathrm{E}\,\beta(t_1)\beta(t_2) = \min(t_1, t_2) - t_1 t_2$ для $t_1, t_2 \in [0,1]$.

Глава 2

НОРМАЛЬНАЯ ВЫБОРКА: ОЦЕНИВАНИЕ

В этой главе мы займёмся статистическим анализом на базе выборки из $N_p(a,Q)$. Выборка объема n — это совокупность $X_1, X_2, \ldots, X_n$ независимых одинаково распределенных p-мерных случайных величин; в нашем случае $X_i \sim N_p(a,Q)$, $i=\overline{1,n}$. Параметры a, Q нормального закона нам неизвестны. В этой главе мы найдём для a и Q разумные оценки.

§ 1. Достаточные статистики

Предположим сначала, что $N_p(a,Q)$ — невырожденное распределение, т. е. что матрица Q^{-1} существует. В этом случае распределение $N_p(a,Q)$ имеет плотность относительно меры Лебега в $\mathbb{R}^p$. Это дает нам возможность для поиска достаточных статистик воспользоваться **критерием факторизации**. Сформулируем его ради напоминания (через функции правдоподобия):

Пусть ξ — случайная величина, плотность распределения которой относительно некоторой меры есть $f(x,\vartheta)$, где $\vartheta \in \Theta$ — неизвестный параметр. Согласно критерию факторизации, статистика $T(\xi)$ достаточна для $\vartheta \in \Theta$, если для некоторых функций g, h

$$f(\xi,\vartheta) = g(T(\xi),\vartheta)h(\xi)$$

для всех $\vartheta \in \Theta$. Имеется в виду, что равенство для всякого $\vartheta \in \Theta$ выполняется с P_ϑ-вероятностью единица. Функция $g(T(\xi),\vartheta)$ зависит от ξ через посредство $T(\xi)$; функция $h(\xi)$ от ϑ не зависит.

В нашем случае ξ — это выборка $X_1, X_2, \ldots, X_n$; параметром ϑ служит пара (a,Q). Правдоподобие (a,Q) по выборке $X_1, \ldots, X_n$ из невырожденного нормального распределения $N_p(a,Q)$ равно

$$\left[\left(\frac{1}{\sqrt{2\pi}}\right)^p\right]^n \left(\frac{1}{\sqrt{\det Q}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i-a)^T Q^{-1}(X_i-a)\right\}.$$

Покажем, что это выражение можно преобразовать (тождественным образом) так, что оно будет зависеть от $X_1, \ldots, X_n$ через посредство

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{и} \qquad S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T.$$

Этим будет доказано, что пара $(\overline{X}, S)$ служит для (a, Q) достаточной статистикой.

Доказательство. Заметим, что стоящую в показателе экспоненты сумму можно записать как

$$\sum_{i=1}^{n}(X_i - a)^T Q^{-1}(X_i - a) = \sum_{i=1}^{n}(X_i - \overline{X})^T Q^{-1}(X_i - \overline{X}) + n(\overline{X} - a)^T Q^{-1}(\overline{X} - a).$$

Действительно,

$$\begin{aligned}\sum_{i=1}^{n}(X_i - a)^T Q^{-1}(X_i - a) &= \sum_{i=1}^{n}[(X_i - \overline{X}) + (\overline{X} - a)]^T Q^{-1}[(X_i - \overline{X}) + (\overline{X} - a)] \\ &= \sum_{i=1}^{n}(X_i - \overline{X})^T Q^{-1}(X_i - \overline{X}) + \sum_{i=1}^{n}(\overline{X} - a)^T Q^{-1}(\overline{X} - a) \\ &\quad + \sum_{i=1}^{n}(X_i - \overline{X})^T Q^{-1}(\overline{X} - a) + \sum_{i=1}^{n}(\overline{X} - a)^T Q^{-1}(X_i - \overline{X}).\end{aligned}$$

Оба последних слагаемых равны нулю; например,

$$\sum_{i=1}^{n}(X_i - \overline{X})^T Q^{-1}(\overline{X} - a) = 0, \quad \text{т.к.} \quad \sum_{i=1}^{n}(X_i - \overline{X}) = 0.$$

Для дальнейшего преобразования первой суммы в правой части воспользуемся свойством умножения матриц:

$$\operatorname{tr} AB = \operatorname{tr} BA,$$

если указанные операции с матрицами A и B возможны (а также тем, что след суммы равен сумме следов). Далее заметим, что след (tr) скалярной величины, которой является упомянутая сумма, равен ей самой. Теперь

$$\begin{aligned}\sum_{i=1}^{n}&(X_i-\overline{X})^T Q^{-1}(X_i-\overline{X})\\ &=\operatorname{tr}\Big\{\sum_{i=1}^{n}(X_i-\overline{X})^T Q^{-1}(X_i-\overline{X})\Big\}=\sum_{i=1}^{n}\operatorname{tr}\Big\{(X_i-\overline{X})^T Q^{-1}(X_i-\overline{X})\Big\}\\ &=\sum_{i=1}^{n}\operatorname{tr}\Big\{Q^{-1}(X_i-\overline{X})(X_i-\overline{X})^T\Big\}=\operatorname{tr}\Big\{\sum_{i=1}^{n}Q^{-1}(X_i-\overline{X})(X_i-\overline{X})^T\Big\}\\ &=\operatorname{tr}\Big\{Q^{-1}\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})^T\Big\}=(n-1)\operatorname{tr}(Q^{-1}S).\end{aligned}$$

Таким образом, правдоподобие оказывается равным

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{np}\left(\frac{1}{\sqrt{\det Q}}\right)^{n}\\ \times\exp\left\{-\frac{1}{2}[(\overline{X}-a)^T Q^{-1}(\overline{X}-a)+(n-1)\operatorname{tr}(Q^{-1}S)]\right\}.$$

Видно, что от исходной выборки правдоподобие зависит опосредованно, через статистики $\overline{X}$ и S. Следовательно, пара $(\overline{X}, S)$ достаточна для параметра (a, Q).

Статистика $(\overline{X}, S)$ достаточна для (a, Q) и в вырожденном случае, когда матрица Q^{-1} не существует. В вырожденном случае нормальное распределение не имеет плотности (относительно меры Лебега в $\mathbb{R}^p$), и потому приведённое выше рассуждение невозможно. Для доказательства достаточности $(\overline{X}, S)$ и в этом случае покажем, что условное распределение выборки $(X_1, \ldots, X_n)$ при фиксированном $(\overline{X}, S)$ не зависит от параметра (a, Q).

Рассмотрим распределение $N_p(a, Q+\varepsilon I)$, где число $\varepsilon > 0$, I — единичная матрица, и выборку $(X_1, \ldots, X_n)$ из этого распределения. Поскольку распределение $N_p(a, Q+\varepsilon I)$ не вырождено, условное распределение $(X_1, \ldots, X_n)$ при фиксированном $(\overline{X}, S)$ не зависит от параметра. Условное распределение при $\varepsilon = 0$, т.е. для параметра (a, Q), можно получить из условного распределения при $\varepsilon > 0$ предельным переходом $\varepsilon \to 0$. При этом предельном переходе независимость условного распределения от (a, Q) сохраняется. Поэтому и в вырожденном случае статистика $(\overline{X}, S)$ достаточна для (a, Q). $\square$

Впрочем, в теоретических исследованиях можно ограничиться рассмотрением только невырожденных распределений. Действительно, пусть наблюдаемый p-мерный случайный вектор имеет вырожденное нормальное распределение $N_p(a, Q)$ и ранг матрицы Q равен q, $q < p$. Тогда среди p координат этого вектора можно выбрать q таких, что прочие координаты являются определёнными аффинными функциями первых. Достаточно, следовательно, знать q случайных переменных, чтобы восстановить остальные. Поэтому разумно ограничить своё внимание именно ими. По другому говоря, от вырожденных многомерных наблюдений можно перейти к невырожденным, но меньшей размерности, исключив избыточные переменные. Поэтому в дальнейшем мы будем заниматься в основном невырожденными нормальными распределениями.

§ 2. Несмещённые оценки

Заметим, что
$$\mathrm{E}\,\overline{X} = a, \qquad \mathrm{E}\,S = Q.$$

Первое равенство очевидно. Для доказательства второго воспользуемся тождеством
$$\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T = \sum_{i=1}^{n}(X_i - a)(X_i - a)^T - n(\overline{X} - a)(\overline{X} - a)^T.$$
Так как $\mathrm{E}(\overline{X} - a)(\overline{X} - a)^T = \frac{1}{n}Q$, то
$$\mathrm{E}[(n-1)S] = nQ - Q = (n-1)Q,$$
что и требовалось доказать.

Таким образом, достаточные статистики $\overline{X}$ и S одновременно служат несмещёнными оценками параметров нормального распределения.

Попутно отметим, что $\overline{X}$ и S являются несмещёнными оценками математического ожидания и матрицы ковариаций для всякой выборки (не только нормальной).

§ 3. Оптимальность $\overline{X}$ и S

Несмещённые оценки одномерных параметров сравнивают по их дисперсиям: из двух несмещённых оценок параметра та лучше, чья дисперсия меньше.

Для многомерных параметров оценками служат многомерные случайные величины. Сравнивать многомерные оценки по какому-либо одному числовому показателю неразумно. Но многомерные несмещённые оценки можно сопоставлять по дисперсиям их линейных функционалов. Подробнее: пусть $\vartheta = (\vartheta_1, \dots, \vartheta_m)^T$ — неизвестный параметр, $\vartheta \in \Theta$. Пусть статистики $T(X) = (T_1(X), \dots, T_m(X))^T$ и $R(X) = (R_1(X), \dots, R_m(X))^T$ — несмещённые оценки параметра ϑ по наблюдению X. Это значит, что

$$\mathrm{E}\, T(X) = \vartheta,\ \mathrm{E}\, R(X) = \vartheta, \qquad \text{для любых } \vartheta \in \Theta.$$

Пусть $c_1, \dots, c_m$ — произвольные (неслучайные) переменные. При этом $\sum_{i=1}^m c_i T_i(X)$ и $\sum_{i=1}^m c_i R_i(X)$ — несмещённые оценки $\sum_{i=1}^m c_i \vartheta_i(X)$.

Мы скажем, что несмещённая оценка $T(X)$ *лучше*, чем несмещённая оценка $R(X)$, если

$$D(\sum_{i=1}^m c_i T_i(X)) \leqslant D(\sum_{i=1}^m c_i R_i(X))$$

для любых последовательностей $c_1, \dots, c_m$, причём для некоторых из них неравенство строгое.

Несмещённую оценку $T(X)$ естественно считать *наилучшей*, если $T(X)$ в указанном выше смысле превосходит любую другую несмещённую оценку, или хотя бы ни одной из них не уступает.

Несмещённые оценки мы сравнивали, сопоставляя дисперсии их линейных функционалов. Эти дисперсии можно выражать с помощью матриц ковариаций. Пусть

$$DT = \mathrm{E}(T - \mathrm{E}\, T)(T - \mathrm{E}\, T)^T, \qquad DR = \mathrm{E}(R - \mathrm{E}\, R)(R - \mathrm{E}\, R)^T.$$

Положим $c = (c_1, \dots, c_m)^T$. Тогда

$$D\Big(\sum_{i=1}^m c_i T_i(X)\Big) = c^T (DT) c, \qquad D\Big(\sum_{i=1}^m c_i R_i(X)\Big) = c^T (DR) c,$$

а если несмещённая оценка T не хуже, чем R, то

$$c^T (DT) c \leqslant c^T (DR) c, \quad \text{для всех } c = c_1, \dots, c_m.$$

Такое соотношение между квадратичными формами (от переменных $c_1, \dots, c_m$) в матричной алгебре толкуют как соотношение между (симметричными) матрицами DT и DR:

$$DT \leqslant DR.$$

Следовательно, мы можем сказать, что из двух несмещённых оценок многомерного параметра лучше та, которая имеет меньшую (в указанном смысле) матрицу ковариаций.

Наилучшие несмещённые оценки параметров существуют далеко не во всех задачах. Во-первых, далеко не всегда существуют несмещённые оценки. Затем, даже если несмещённые оценки существуют, далеко не всегда в совокупности ковариационных матриц, которую они порождают, существуют минимальный элемент или элементы.

Есть, однако, важный исключительный случай: если в задаче существует достаточная статистика для параметра, эта достаточная статистика полна и есть хоть одна несмещённая оценка параметра, то наилучшая несмещённая оценка существует, единственна и при этом является функцией упомянутой достаточной статистики.

В следующем разделе мы покажем, что для нормальной выборки достаточная статистика $(\overline{X}, S)$ полна. Тем самым будет доказано, что $\overline{X}$ и S как несмещённые оценки параметров a и Q являются наилучшими.

По отношению к $\overline{X}$ как к несмещённой оценке вектора-столбца a свойство оптимальности разъяснено выше: вектор-столбец $\overline{X}$ имеет наименьшую матрицу ковариаций среди всех несмещённых оценок a.

Для матрицы $S = \| \, s_{ij} \mid i, j = \overline{1, p} \, \|$ оптимальность её как несмещённой оценки матрицы $Q = \| \, q_{ij} \mid i, j = \overline{1, p} \, \|$ означает следующее. Пусть $Z = \| \, z_{ij} \mid i, j = \overline{1, p} \, \|$ означает произвольную переменную $(p \times p)$-матрицу. Оптимальность S означает, что статистика

$$\sum_{i=1}^{p} \sum_{i=1}^{p} z_{ij} s_{ij} = \operatorname{tr}(ZS)$$

среди всех несмещённых оценок параметра

$$\sum_{i=1}^{p} \sum_{i=1}^{p} z_{ij} s_{ij} = \operatorname{tr}(ZQ)$$

имеет наименьшую дисперсию для всякой матрицы Z.

К сожалению, это свойство не удаётся выразить стандартными средствами матричной алгебры. Поэтому иногда матрицы искусственно превращают в вектор-столбцы (либо в вектор-строки). Существует специальное обозначение такого преобразования.

Пусть A — некоторая матрица. Пусть $A_1, A_2, \ldots, A_m$ — её столбцы размерности, скажем, p. Символ $\operatorname{vec} A$ обозначает вектор-столбец размерности pm.

Этот $\operatorname{vec} A$ получают, последовательно размещая столбцы $A_1, A_2, \ldots$ один под другим: столбец A_2 помещают под столбцом A_1, далее столбец A_3 помещают под столбцом $\binom{A_1}{A_2}$ и т.д.

Для полученных таким способом векторов $\operatorname{vec} S$, $\operatorname{vec} Q$ и т.д. оптимальность $\operatorname{vec} S$ как несмещённой оценки $\operatorname{vec} Q$ означает минимальность матрицы $D(\operatorname{vec} S)$ в соответствующем классе $(pm \times pm)$-матриц.

В этой книге к векторизации матриц мы постараемся не прибегать.

§ 4. Полнота статистики $(\overline{X}, S)$

Напомним определение: статистика T называется *полной* для семейства распределений $\{P_\vartheta, \vartheta \in \Theta\}$, если уравнение (относительно функции $\psi(T)$)

$$\mathrm{E}_\vartheta \psi(T) = 0 \qquad \text{для всех } \vartheta \in \Theta$$

имеет лишь тривиальное решение: $\psi(T) = 0$ почти наверно P_ϑ, $\vartheta \in \Theta$.

Если статистика T достаточна для ϑ, $\vartheta \in \Theta$, и полна, то наилучшая несмещённая оценка ϑ, если существуют несмещённые оценки ϑ, является функцией статистики T. О том, что такое наилучшая несмещённая оценка для векторного или матричнозначного параметра мы говорили в § 3. Сейчас же покажем, что для семейства распределений $N_p(a, Q)$ и выборки из этого семейства статистика $(\overline{X}, S)$ является полной.

Объяснение. Пара статистик $\overline{X}$ и S есть статистка $(\overline{X}, S)$. Она достаточна для (a, Q) и является полной. Статистики $\overline{X}$ и S можно рассматривать порознь.

Распределения выборки из $N_p(a, Q)$ относительно параметра $\vartheta = (a, Q)$ образуют экспоненциальное семейство. При определённых условиях для экспоненциальных семейств существуют полные достаточные статистики. Напомним определения.

Семейство вероятностных распределений $\{P_\vartheta, \vartheta \in \Theta\}$ на множестве $\mathcal{X}$ называют экспоненциальным, если относительно некоторой меры эти распределения имеют плотность вида

$$p(x, \vartheta) = c(\vartheta) h(x) \exp\Big\{\sum_{i=1}^{m} c_i(\vartheta) \varphi_i(x)\Big\},$$

где $c(\vartheta) > 0$, $c_1(\vartheta), \ldots, c_m(\vartheta)$ — функции на Θ; $h(x)$, $\varphi_1(x), \ldots, \varphi_m(x)$ — функции на $\mathcal{X}$.

Предположим, что функции $\varphi_1(x), \ldots, \varphi_m(x)$ линейно независимы на $\mathcal{X}$, функции $c_1(\vartheta)$, $\ldots, c_m(\vartheta)$ линейно независимы на Θ, X — случайная переменная, распределение которой принадлежит этому семейству.

Теорема. [1] [2] *В этих условиях семейство распределений является полным, если множество значений $c_1(\vartheta), \ldots, c_m(\vartheta)$ m-мерной функции параметра ϑ, $\vartheta \in \Theta$, содержит открытое множество в пространстве $\mathbb{R}^m$. Статистика $(\varphi_1(x), \ldots, \varphi_m(x))$ при этом является полной достаточной статистикой для параметра ϑ.*

При доказательстве на основании этой теоремы полноты статистики $(\overline{X}, S)$ нам будет удобнее обсуждать правдоподобие, чем плотность. Пусть $X = (X_1, \ldots, X_n)$ — выборка из $N_p(a, Q)$. Правдоподобие параметра (a, Q) по выборке равно

$$\left(\frac{1}{\sqrt{2}\,\pi}\right)^{pn}\left(\frac{1}{\sqrt{\det Q}}\right)^{n}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-a)^TQ^{-1}(X_i-a)\right\}.$$

Введём новый параметр $\tau = (b, R)$, где $b = Q^{-1}a$, $R = Q^{-1}$, взаимнооднозначно связанный с параметром $\vartheta = (a, Q)$. Стоящую в показателе экспоненты сумму преобразуем уже известными приёмами;

$$\begin{aligned}
&\sum_{i=1}^{n}(X_i-a)^TQ^{-1}(X_i-a)\\
&\qquad= \sum_{i=1}^{n}(X_i-\overline{X})^TQ^{-1}(X_i-\overline{X}) + n(\overline{X}-a)^TQ^{-1}(\overline{X}-a)\\
&\qquad= (n-1)\,\mathrm{tr}(Q^{-1}S) + n\overline{X}^TQ^{-1}\overline{X} - 2n\overline{X}^TQ^{-1}a + na^TQ^{-1}a\\
&\qquad= (n-1)\,\mathrm{tr}(RS) + n\,\mathrm{tr}\{R(\overline{X}\;\overline{X}^T)\} - 2n\overline{X}^Tb + nb^TR^{-1}b\\
&\qquad= \mathrm{tr}\{R[(n-1)S + n\overline{X}\;\overline{X}^T]\} - 2n\overline{X}^Tb + nb^TR^{-1}b.
\end{aligned}$$

В новой параметризации правдоподобие равно

$$\left(\frac{1}{\sqrt{2}\,\pi}\right)^{pn}(\det R)^{\frac{n}{2}} \times \exp\left\{-\frac{1}{2}\,\mathrm{tr}\{R[(n-1)S + n\overline{X}\;\overline{X}^T]\} + n\overline{X}^Tb - \frac{n}{2}b^TR^{-1}b\right\}.$$

[1] см. [3]: гл. 2, § 25, теорема 2, стр. 180.
[2] см. [7]: гл. 4, § 3, теорема 1, стр. 183.

Видно, что правдоподобие удовлетворяет условиям теоремы о полноте и что полной достаточной статистикой является $(\overline{X}, (n-1)S + n\overline{X}\overline{X}^T)$. В силу взаимнооднозначной связи её со статистикой $(\overline{X}, S)$ эта последняя тоже достаточна для (a, Q) и полна.

§ 5. Независимость $\overline{X}$ и S; распределение S

Вспомним, как решаются поставленные в заголовке и им подобные задачи для одномерных нормальных совокупностей. Основой служит **лемма об ортогональных разложениях**:

Пусть $x_1, \ldots, x_n$ суть независимые нормальные случайные величины с общей дисперсией, скажем, σ^2. Совокупность $x_1, \ldots, x_n$ рассмотрим как вектор $\mathbf{x} = (x_1, \ldots, x_n)^T$. Пусть C – ортогональная $(n \times n)$-матрица. Пусть $\mathbf{y} = C\mathbf{x}$ – результат ортогонального преобразования $\mathbf{x}$, $\mathbf{y} = (y_1, \ldots, y_n)^T$. Тогда случайные величины $y_1, \ldots, y_n$ независимы в совокупности и распределены нормально с той же общей дисперсией σ^2.

В гл. 6, которая развивает геометрический взгляд на совокупности многомерных наблюдений, будет сформулирован соответствующий многомерный вариант этой леммы. Но до той поры мы будем обходиться традиционными средствами. Традиционное обобщение упомянутой леммы проходит алгебраически.

Лемма. *Пусть $X_1, X_2, \ldots, X_n$ – независимые гауссовские векторы: $X_i \sim N_p(\cdot, Q), i = \overline{1, n}$. Пусть $C = \| c_{\alpha\beta} \|$ – ортогональная матрица размера $n \times n$. Рассмотрим случайные векторы $Y_1, Y_2, \ldots, Y_n$, где*

$$Y_\alpha = \sum_{\beta=1}^{n} c_{\alpha\beta} X_\beta, \qquad \alpha = 1, \ldots, n.$$

Тогда

(a) *совокупность $Y_1, Y_2, \ldots, Y_n$ имеет совместное нормальное распределение, так как является линейным преобразованием нормально распределенной совокупности X_1, $X_2, \ldots, X_n$;*

(b)
$$E\,Y_\alpha = \sum_{\beta=1}^{n} c_{\alpha\beta}\, E\, X_\beta\,; \tag{2.1}$$

(c)
$$\mathrm{Cov}(Y_\alpha, Y_\beta) = \delta_{\alpha\beta} Q, \tag{2.2}$$

так что векторы $Y_1, Y_2, \ldots, Y_n$ независимы в совокупности;

(d)
$$\sum_{\alpha=1}^{n} X_\alpha X_\alpha^T = \sum_{\alpha=1}^{n} Y_\alpha Y_\alpha^T. \tag{2.3}$$

Доказательство всех утверждений получается прямыми выкладками. Докажем, например, п.(c).

$$\begin{aligned}
\operatorname{Cov}(Y_\alpha, Y_\beta) &= \mathrm{E}(Y_\alpha - \mathrm{E}\, Y_\alpha)(Y_\beta - \mathrm{E}\, Y_\beta)^T \\
&= \mathrm{E}\Big[\sum_{i=1}^{n} c_{\alpha i}(X_i - \mathrm{E}\, X_i)\Big]\Big[\sum_{j=1}^{n} c_{\beta j}(X_j - \mathrm{E}\, X_j)\Big]^T \\
&= \sum_{i=1}^{n} c_{\alpha i} \sum_{j=1}^{n} c_{\beta j} E(X_i - \mathrm{E}\, X_i)(X_j - \mathrm{E}\, X_j)^T \\
&= \sum_{i=1}^{n} c_{\alpha i} \sum_{j=1}^{n} c_{\beta j}\delta_{ij} Q = \sum_{i=1}^{n} c_{\alpha i} c_{\beta i} Q = \delta_{\alpha\beta} Q.
\end{aligned}$$

$\square$

Вернёмся к выборке $X_1, \ldots, X_n$ из $N_p(a, Q)$. Теперь мы можем доказать независимость статистик $\overline{X}$ и S и указать распределение S.

Применим лемму к совокупности $X_1 - a, \ldots, X_n - a$. Пусть $C = \| c_{\alpha\beta} \|$ – ортогональная матрица, последняя строка которой есть $\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$, прочие строки произвольны. Пусть $Y_\alpha = \sum\limits_{\beta=1}^{n} c_{\alpha\beta}(X_\beta - a)$, $\alpha = 1, \ldots, n$. В силу леммы случайные векторы $Y_1, \ldots, Y_n$ независимы в совокупности и распределены по закону $N_p(0, Q)$. При этом $Y_n = \sqrt{n}\,(\overline{X} - a)$. Как мы уже видели,

$$(n-1)S = \sum_{i=1}^{n} (X_i - a)(X_i - a)^T - n(\overline{X} - a)(\overline{X} - a)^T.$$

Воспользуемся п.(d) леммы, согласно которому

$$\sum_{i=1}^{n} (X_i - a)(X_i - a)^T = \sum_{i=1}^{n} Y_i Y_i^T\,.$$

Поскольку $Y_n Y_n^T = n(\overline{X} - a)(\overline{X} - a)^T$, то заключаем, что

$$(n-1)S = \sum_{i=1}^{n-1} Y_i Y_i^T.$$

Теперь видно, что статистики $\overline{X}$ и S независимы, так как их формируют независимые случайные величины Y_n и $(Y_1, Y_2, \dots, Y_{n-1})$ соответственно.

Распределение S**.** Пусть $\xi_1, \dots, \xi_m$ – независимые случайные векторы, распределенные по закону $N_p(0, Q)$. Говорят, что случайная величина (матрица)

$$\sum_{i=1}^{m} \xi_i \xi_i^T$$

имеет (центральное) *распределение Уишарта* (Wishart, 1932), которое обозначают $W_p(m, Q)$. Теперь для распределения S есть название: $(n-1)S$ распределено, как $W_p(n-1, Q)$. При $p = 1$ распределение Уишарта $W_p(m, Q)$ превращается в распределение $\sigma^2\chi^2(m)$. Поэтому распределение Уишарта можно рассматривать как обобщение распределения хи-квадрат.

Случайная матрица, распределённая по Уишарту, – это совокупность случайных величин, ее элементов. Характеристическая функция их совместного распределения известна, так же как и их совместная плотность (плотность существует только при $m \geqslant p$). Но ни то ни другое нам не понадобится. Поэтому приводить их выражения мы не будем. Их можно найти, например в [1, гл. 7].

Распределение Уишарта зависит от матрицы Q. Эту зависимость можно сделать более явной, если ввести матрицу $Q^{1/2}$. Для произвольной неотрицательно определённой матрицы A матрицу $A^{1/2}$ (квадратичный корень из матрицы A) можно определять несколькими эквивалентными способами. Можно сказать так: $A^{1/2}$ – это (единственная) неотрицательно определённая матрица, которая удовлетворяет уравнению

$$X^2 = Q.$$

Можно также обратиться к спектральному представлению симметричных матриц. Если A – симметричная матрица порядка p, то A можно представить как

$$A = \lambda_1 U_1 U_1^T + \lambda_2 U_2 U_2^T + \dots,$$

где $U_1, U_2, \dots$ – это ортогональный базис $\mathbb{R}^p$, составленный из собственных векторов матрицы A; $\lambda_1, \lambda_2, \dots$ – соответствующие им собственные значения. Для функции числового аргумента $f(\lambda)$ можно определить функцию матричного аргумента $f(A)$, положив

$$f(A) = f(\lambda_1) U_1 U_1^T + f(\lambda_2) U_2 U_2^T + \dots.$$

Для неотрицательно определённых матриц их собственные значения неотрицательны, поэтому существуют величины $\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots$ и функция $A^{1/2}$:

$$A^{1/2} = \lambda_1^{1/2} U_1 U_1^T + \lambda_2^{1/2} U_2 U_2^T + \ldots.$$

Вернёмся к распределению Уишарта. Будем использовать символ $W_p(m, Q)$ и как обозначение этого распределения, и как случайную матрицу, следующую этому распределению. Тогда

$$W_p(m, Q) = \xi_1 \xi_1^T + \ldots + \xi_p \xi_p^T,$$

где $\xi_1, \xi_2, \ldots, \xi_p$ — независимые нормально распределённые векторы, $\xi_i \sim$ $\sim N_p(0, Q)$. Используя матрицу $Q^{1/2}$, случайные векторы $\xi_i, i = \overline{1, p}$, представим как $\xi_i = Q^{1/2}\eta_i$, где $\eta_i \sim N(0, I)$, I – единичная $(p \times p)$-матрица. Тогда

$$W_p(m, Q) = \sum_{i=1}^{p} \xi_i \xi_i^p = Q^{1/2} \Big(\sum_{i=1}^{p} \eta_i \eta_i^T\Big) Q^{1/2} = Q^{1/2} W_p(m, I) Q^{1/2}.$$

Этим представлением мы часто будем пользоваться в дальнейшем.

§ 6. Оценивание (a, Q) по выборке

Из многих возможных методов оценивания остановимся на двух наиболее известных — методе наименьших квадратов и методе наибольшего правдоподобия.

Метод наименьших квадратов по своему духу применим к оцениванию параметра сдвига, в данной задаче – к оцениванию a. Оценивание по методу наименьших квадратов состоит в выборе для неизвестного a такого значения, при котором сумма различий между a и наблюдениями $X_1, \ldots, X_n$ минимальна. Различие между X_i и, следовательно, сумму различий можно понимать как

(a) $\sum_{i=1}^n |X_i - a|^2 = \sum_{i=1}^n (X_i - a)^T (X_i - a)$,

(b) $\sum_{i=1}^n (X_i - a)(X_i - a)^T$.

В обоих случаях вектор a следует выбирать так, чтобы упомянутые суммы были минимальны. Существование решения не вызывает сомнения в случае (a).

В случае (b) предстоит сравнивать неотрицательно определённые матрицы. Напомним, что для симметричных матриц неравенство $A \geqslant B$ эквивалентно выполнению неравенства $z^T A z \geqslant z^T B z$ для всякого вектора z. В случае (b) предстоит найти минимальный элемент в некотором множестве неотрицательно определённых матриц указанного выше вида. Существование минимального элемента в этом множестве матриц не очевидно.

В обоих случаях к цели ведёт, по существу, одна и та же выкладка:

$$\sum_{i=1}^{n}(X_i-a)^T(X_i-a)=\sum_{i=1}^{n}(X_i-\overline{X})^T(X_i-\overline{X})+n(\overline{X}-a)^T(\overline{X}-a),$$

$$\sum_{i=1}^{n}(X_i-a)(X_i-a)^T=\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})^T+n(\overline{X}-a)(\overline{X}-a)^T.$$

Первое тождество нам уже известно; второе тождество легко проверить. Как в случае (a), так и в случае (b) исходная величина превращается в сумму двух слагаемых. Из них первое не зависит от переменного a. Второе слагаемое неотрицательно либо как число (случай (a)), либо как квадратная матрица (случай (b), второе слагаемое есть неотрицательно определённая матрица). В обоих случаях результат минимален, когда минимально второе слагаемое, т.е. когда оно обращается в нуль при выборе $\widehat{a}=\overline{X}$.

Существование решения в матричном понимании квадрата расстояний уже указывает на возможность использования этого вида расстояния в дальнейшем. Систематически эта мысль будет развита в гл. 5 и использована ради статистических целей в гл. 6. А сейчас отметим, что остаточная сумма квадратов в матричном методе $\sum(X_i-\overline{X})(X_i-\overline{X})$ даёт уже известную несмещённую оценку матрице ковариаций Q:

$$\widehat{Q}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X}).$$

Метод наибольшего правдоподобия можно применять, когда возможно выписать правдоподобие (a, Q). Предположим, что Q^{-1} существует. Как мы видели в п. 1, правдоподобие можно представить в виде

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{pn}\left(\frac{1}{\sqrt{\det Q}}\right)^{n}\exp\Big\{-\frac{1}{2}[n(\overline{X}-a)^TQ^{-1}(\overline{X}-a)+n\mathrm{tr}Q^{-1}S_n]\Big\},$$

если положить

$$S_n=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})^T.$$

В какой-то момент у нас появятся элементы матриц S_n и Q. Пусть $Q = \| q_{ij} \|$, $S_n = \| s_{ij} \|$, где $i, j = \overline{1,p}$. (Отметим, что S_n несколько отличается от S, несмещённой оценки матрицы Q, введённой ранее.)

Правдоподобию надо дать максимальное значение, изменяя $a \in \mathbb{R}^p$ и матрицу $Q \geqslant 0$. Ясно, что при фиксированном Q наибольшее значение правдоподобие получает при $a = \overline{X}$, и это при любом Q. Выбрав, таким образом, для a наилучшее значение $\widehat{a} = \overline{X}$, выберем затем максимизирующее значение для Q. При этом мы можем рассматривать логарифмическое правдоподобие

$$-\frac{n}{2}\log\det Q - \frac{n}{2}\mathrm{tr} Q^{-1}S_n + \text{ const}.$$

Здесь const обозначает не зависящее от Q выражение.

Приведем одно общее замечание об оценивании по методу наибольшего правдоподобия. Предположим, что, максимизируя правдоподобие, скажем, $f(X, \vartheta)$ по параметру $\vartheta \in \Theta$ (здесь X – наблюдение), мы оцениваем таким способом параметр $\vartheta \in \Theta$. Предположим далее, что переменная ϑ находится во взаимнооднозначном соответствии с другой переменной, скажем τ, $\tau = \tau(\vartheta)$ и $\vartheta = \vartheta(\tau)$. Когда ϑ пробегает множество Θ, τ пробегает множество, скажем, T. Вместо правдоподобия $f(X, \vartheta)$ по переменной ϑ мы можем максимизировать правдоподобие $f(X, \vartheta(\tau))$ по переменной τ, это дает оценку наибольшего правдоподобия $\widehat{\tau}$ для τ. Замечание состоит в том, что $\widehat{\vartheta} = \vartheta(\widehat{\tau})$, $\widehat{\tau} = \tau(\widehat{\vartheta})$. Из этого наблюдения следует, что оценки наибольшего правдоподобия эквивариантны относительно выбора параметризации и что эту параметризацию можно менять, руководствуясь соображениями простоты, удобства и прочими.

В согласии с этим замечанием на время возьмем другую параметризацию закона $N_p(a, Q)$: a и Q^{-1}. Максимизировать логарифмическое правдоподобие будем по (матричному) параметру Q^{-1}. Для удобства положим $Q^{-1} = R$, $R = \| r_{ij} \|$. Теперь логарифмическое правдоподобие R имеет вид

$$\frac{n}{2}\log\det R - \frac{n}{2}\mathrm{tr} RS_n + \text{ const}.$$

Необходимое условие экстремума – обращение в нуль всех частных производных. Надо, следовательно, рассмотреть и решить систему уравнений

$$\frac{\partial}{\partial r_{ij}}\left[\frac{n}{2}\log\det R - \frac{n}{2}\mathrm{tr} RS_n\right] = 0, \qquad i, j = \overline{1, n}.$$

Выпишем эти производные отдельно для каждого из двух слагаемых.

Для определителя R рассмотрим его разложение по строке:

$$\det R = \sum_{\alpha=1}^{p} r_{i\alpha} R_{i\alpha},$$

где $R_{i\alpha}$ обозначают алгебраические дополнения. Так как матрица R симметрична, то $R_{i\alpha} = R_{\alpha i}$. Теперь

$$\frac{\partial}{\partial r_{ij}} \log \det R = \frac{1}{\det R} \frac{\partial}{\partial r_{ij}} \sum_{\alpha=1}^{p} r_{i\alpha} R_{i\alpha} = \frac{R_{ij}}{\det R} = \frac{R_{ji}}{\det R}.$$

Как известно, $\frac{R_{ji}}{\det R}$ – это элемент (с номером (i, j)) матрицы, обратной R. В данном случае это элемент q_{ij} матрицы $Q = \| q_{ij} \|$.

Дифференцируя второе слагаемое, получаем

$$\frac{\partial}{\partial r_{ij}} \mathrm{tr} R S_n = \mathrm{tr} \left(\frac{\partial}{\partial r_{ij}} R S_n \right) = \mathrm{tr}(I(i, j) S_n),$$

где $I(i, j)$ обозначает матрицу, все элементы которой равны 0, исключая элемент с номером (i, j), который равен 1. Произведение $I(i, j) S_n$ дает матрицу, все строки которой, исключая строку с номером i, равны нулю, а строка с номером i такова: $s_{j1}, s_{j2}, \ldots, s_{jp}$. На диагональ матрицы $I(i, j) S_n$ попадает элемент s_{ji}. Поэтому

$$\mathrm{tr}\Big(I(i, j) S_n \Big) = s_{ji}.$$

Таким образом, система уравнений правдоподобия – это система

$$q_{ij} - s_{ji} = 0, \qquad 1 \leqslant i,\ j \leqslant p.$$

Следовательно, оценкой наибольшего правдоподобия для (a, Q) по выборке из $N_p(a, Q)$ служит пара

$$\widehat{a} = \overline{X} \qquad \text{и} \qquad \widehat{Q} = S_n.$$

Отметим сходство полученного результата с известными оценками наибольшего правдоподобия для параметров одномерного нормального распределения $N(a, \sigma^2)$ по выборке $x_1, \ldots, x_n$ объема n:

$$\widehat{a} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Метод Роя (S.N. Roy). Этот метод состоит в трансформации многомерной задачи в множество одномерных задач, их решении и последующем конструировании решения многомерной задачи из одномерных решений.

Изначально метод был применён С.Н. Роем (1952) к проверке гипотез. В этом качестве мы используем его позже. Сейчас мы применим этот метод к оцениванию.

Пусть $u \in \mathbb{R}^p$ – переменный p-мерный вектор. От случайных p-мерных величин $X_1, X_2, \ldots, X_n$ перейдём к случайным величинам $x_1 = u^T X_1, x_2 = u^T X_2, \ldots, x_n = u^T X_n$. Случайные величины $x_1, \ldots, x_n$ образуют выборку из нормального распределения (одномерного) $N(u^T a, u^T Q u)$. Оценки наибольшего правдоподобия по выборке для параметров нормального закона хорошо известны:

$$\widehat{(u^T a)} = \overline{x}, \qquad \widehat{(u^T Q u)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

или

$$\widehat{(u^T a)} = u^T \overline{X}, \qquad \widehat{(u^T Q u)} = \frac{1}{n} u^T \left[\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T\right] u.$$

Отсюда следует, что в качестве оценок неизвестных a, Q можно взять

$$\widehat{a} = \overline{X}, \qquad \widehat{Q} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T.$$

Мы пришли к известным нам оценкам наибольшего правдоподобия $\overline{X}$ и S_n.

Можно ли доказать, что метод Роя (как он применён выше) даёт именно оценки наибольшего правдоподобия?

§ 7. Упражнения

1. Дана выборка из $N_p(a, Q)$, причём матрица Q известна, $Q > 0$. Разобьем вектор $a \in \mathbb{R}^p$ на подвекторы a_1, a_2 так, что $a^T = \left(a_1^T, a_2^T\right)^T$.
   В предположении, что $a_2 = 0$, найдите для a_1 оценку наибольшего правдоподобия. (Необходимые обозначения введите сами).
2. Пусть $X_1, \ldots, X_n$ независимы, $X_k \sim N_p(c_k a, Q)$, $k = \overline{1, n}$, где $c_1, \ldots, c_n$ – заданные числа, параметры a, Q неизвестны.
   Найдите для a, Q оценки наибольшего правдоподобия.
3. Дана выборка из распределения $N_p(A\vartheta, Q)$, ϑ и Q – неизвестные параметры, A – заданная $(q \times p)$-матрица ранга q, $q < p$.
   (a) Укажите для ϑ несмещённую оценку, скажем $\widehat{\vartheta}$.
   (b) Как доказать, что из двух несмещенных оценок $A\vartheta$, а именно $\overline{X}$ и $A\widehat{\vartheta}$, вторая имеет меньшую матрицу ковариаций? (Этот результат следует из развитой в гл. 6 общей теории.)
4. Дана выборка объема n из распределения $N_p(a, Q)$, где параметры $a^T = (a_1, a_2, \ldots, a_p)$ и Q неизвестны.

Найдите для a, Q оценки наименьших квадратов, когда

(a) Известно, что $a_1 = 0$;

(b) Известно, что $a_1 + \ldots + a_p = 0$;

(c) Известна линейная комбинация $c_1 a_1 + \ldots + c_p a_p$ с заданными коэффициентами $c_1, \ldots, c_p$;

(d) Известен вектор Ba, где B – заданная матрица размера $(q \times p)$ ранга q, $q < p$.

(e) Дана выборка из $N_p\,(a, Q)$, причём вектор a известен. Найдите для неизвестной матрицы Q оценку наибольшего правдоподобия при этом условии.

Глава 3

НОРМАЛЬНАЯ ВЫБОРКА: ПРОВЕРКА ГИПОТЕЗ

В этой главе, как преимущественно и во всей книге, мы будем обсуждать только гипотезы о параметре положения. Это гипотезы о параметре a распределения $N_p(a, Q)$. Начнём с проверки по выборке $X_1, \ldots, X_n$ из $N_p(a, Q)$ гипотезы

$$H : a = a_0,$$

где a_0 – заданный p-мерный вектор.

Это простейшая задача о параметре положения. Мы начинаем с её решения не по её практическому значению. На практике гипотезы о точном значении параметров вероятностных распределений встречаются редко. Но эта задача открывает путь к более реалистичным и более сложным статистическим гипотезам.

Надо сказать, что предположение о точном равенстве неизвестного параметра a некоторому определенному a_0 имеет характер условный. Обычно имеют в виду, что a мало отличается от названного a_0. Настолько мало, что этим различием можно пренебречь на фоне неизбежной случайной изменчивости наблюдений. Иначе говоря, отличие подлинного значения параметра от его предложенного гипотетического значения не может быть обнаружено при том объёме наблюдений, которым мы располагаем. О том, как в точном смысле можно проверить приближенную гипотезу $H^* : a \approx a_0$, мы тоже поговорим, но позже, в § 4 этой главы.

В последующих разделах будет рассказано о двух способах, с помощью которых можно строить статистические критерии для проверки гипотез о средних (о математических ожиданиях) по многомерным гауссовским наблюдениям: это метод Роя (S.N. Roy) и метод отношения правдоподобий. (Критерий отношения правдоподобий читателю может быть знаком, ибо для указанных целей он применяется и в одномерных гауссовских моделях.) Для одной выборки при проверке упомянутой гипотезы H оба метода приводят к одному и тому же статистическому критерию (критерию Хотеллинга), но для более сложных задач статистические правила получаются разными. Об этом будет рассказано в последующих главах. Сопоставляя эти конкретные критерии, мы придём к пониманию, как в многомерном линейном анализе устроены разумные критерии для проверки линейных гипотез и как распределены их статистики при гипотезах и альтернативах.

§ 1. Метод Роя

Пусть u обозначает переменный p-мерный вектор, $u \in \mathbb{R}^p$. От выборки $X_1, \ldots, X_n$ из $N_p(a, Q)$ перейдем к скалярным величинам $x_1, \ldots, x_n$, положив $x_i = u^T X_i$, $i = \overline{1, n}$. Случайные величины $x_1, \ldots, x_n$ образуют выборку из $N(u^T a, u^T Q u)$. Гипотеза H для данного u переходит в гипотезу

$$H^{(u)} : u^T a = u^T a_0.$$

Ясно, что гипотеза $H : a = a_0$ верна тогда (и только тогда), когда верны все гипотезы $H^{(u)}, u \in \mathbb{R}^p$. Поэтому гипотезу H следует отвергнуть, если отвергнутой окажется хотя бы одна из гипотез $H^{(u)}$.

Методы проверки $H^{(u)}$ по выборкам из нормальной совокупности хорошо известны. В зависимости от того, задана наблюдателю матрица Q или нет, надо пользоваться либо гауссовским, либо стьюдентовским распределением. Мы рассмотрим случай, когда матрица Q неизвестна. Случай известной матрицы Q вынесен в упражнения к главе.

Поскольку матрица Q неизвестна, то неизвестна и дисперсия $\sigma^2 = u^T Q u$ наблюдений $x_1, \ldots, x_n$. В этом случае для проверки $H^{(u)} : u^T a = u^T a_0$ следует использовать «стьюдентово отношение»

$$\sqrt{n}\, \frac{u^T(\overline{X} - a_0)}{s},$$

где

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u^T X_i - u^T X)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n} [u^T(X_i - \overline{X})][u^T(X_i - \overline{X})]^T = u^T S u\,. \end{aligned}$$

Полученная статистика распределена по Стьюденту с $(n-1)$ степенями свободы и параметром нецентральности:

$$\delta = \frac{\sqrt{n}\, u^T(a - a_0)}{\sqrt{u^T Q u}}.$$

Гипотезу $H^{(u)} : a = a_0$, или $\delta = 0$, следует отвергнуть (на уровне 2α, $0 < \alpha < 0.5$), если

$$\left| \sqrt{n}\, \frac{u^T(\overline{X} - a_0)}{s} \right| > t_{1-\alpha},$$

где $t_{1-\alpha}$ обозначает $(1-\alpha)$-квантиль распределения Стьюдента с $(n-1)$ степенями свободы.

Поскольку критическое значение $t_{1-\alpha}$ (при заданном α уровне значимости) одно и то же для всех $u \in \mathbb{R}^p$, представляется разумным отвергать гипотезу H, когда «слишком велика» статистика

$$\max_{u \in \mathbb{R}^p} \sqrt{n} \left| \frac{u^T(\overline{X} - a_0)}{s} \right| = \max_{u \in \mathbb{R}^p} \sqrt{n}\, \frac{u^T(\overline{X} - a_0)}{s}.$$

При этом ясно, что критические значения надо выбирать, руководствуясь её распределением при гипотезе.

Вычисление полученной статистики опирается на следующую ниже лемму.

Лемма 3. *Пусть R – произвольная положительно-определенная $(p \times p)$-матрица, Y – произвольный p-мерный вектор. Тогда*

$$\max_{u \in \mathbb{R}^p} \frac{u^T Y}{\sqrt{u^T R u}} = \sqrt{Y^T R^{-1} Y}\,.$$

Доказательство. Рассматриваемую экстремальную задачу можно заменить задачей о вычислении условного экстремума

$$\max_{u : u^T R u = 1} u^T Y.$$

Функция Лагранжа в этой задаче равна

$$\mathcal{L}(u, \lambda) = u^T Y - \lambda(u^T R u - 1).$$

Приравнивая частные производные функции Лагранжа по $u \in \mathbb{R}^p$ и $\lambda \in \mathbb{R}^1$ нулю, получаем уравнения:

$$\begin{cases} Y = 2\lambda R u, \\ u^T R u = 1, \end{cases}$$

или

$$\begin{cases} u = \frac{1}{2\lambda} R^{-1} Y, \\ u^T R u = 1. \end{cases}$$

Подставив в последнее уравнение значение u, данное первым уравнением, получим $(2\lambda)^2 = Y^T R^{-1} Y$. Так как $u^T Y = \frac{1}{2\lambda} Y^T R^{-1} Y$, то $\pm u^T Y = \sqrt{Y^T R^{-1} Y}$, что и требовалось доказать. □

Критическая статистика, найденная методом Роя, равна

$$\sqrt{n(\overline{X} - a_0)^T S^{-1} (\overline{X} - a_0)}\,.$$

Эквивалентная статистика

$$T^2 \underset{df}{=} n(\overline{X} - a_0)^T S^{-1} (\overline{X} - a_0)$$

более удобна.

Статистику T^2 называют статистикой Хотеллинга (Hotelling, Harolds, 1931). Она является очевидным многомерным обобщением статистики Стьюдента $t = \sqrt{n}\,\frac{\overline{x}-a}{s}$. Хотеллинг нашёл распределение T^2 как при гипотезе H, так и при альтернативе, когда $a \neq a_0$. Мы сделаем то же в § 3.

Полученные результаты сформулируем как теорему.

Теорема 3.1. *Пусть $X_1, \ldots, X_n$ – независимые случайные векторы $N_p(a, Q)$. Другими словами, $X_1, \ldots, X_n$ – выборка из распределения $N_p(a, Q)$. Для проверки по этой выборке гипотезы*

$$H : a = a_0,$$

где вектор a_0 фиксирован, применение метода Роя даёт критическую статистику

$$T^2 \underset{df}{=} n(\overline{X} - a_0)^T S^{-1} (\overline{X} - a_0),$$

известную как статистика Хотеллинга. Проверяемую гипотезу следует отвергать при больших значениях статистики T^2.

§ 2. Метод отношения правдоподобий

Напомним, в чём состоит этот общий метод, часто применяемый для построения статистических критериев. Пусть ξ – случайная величина (неважно, в каком пространстве ξ принимает значения), имеющая плотность $p(x, \vartheta)$ относительно некоторой меры; ϑ – параметр, $\vartheta \in \Theta$, Θ – заданное множество.

По наблюдению ξ подлежит проверке гипотеза $H : \vartheta \in \mathcal{K}$, где $\mathcal{K}$ – заданное множество, $\mathcal{K} \subset \Theta$.

По аналогии со статистикой Неймана–Пирсона (для проверки простой гипотезы против простой альтернативы) предлагается следующее семейство критических множеств:

$$Q_z \underset{df}{=} \left\{ x : \frac{p(x, \widehat{\vartheta})}{p(x, \widehat{\vartheta}_H)} > z \right\},$$

где $\widehat{\vartheta}$ – оценка наибольшего правдоподобия для ϑ при условии $\vartheta \in \Theta$; $\widehat{\vartheta}_H$ – оценка наибольшего правдоподобия для ϑ при условии $\vartheta \in \mathcal{K}$, т. е. при гипотезе H.

Статистику этого критерия можно записать как отношение двух правдоподобий:

$$\lambda \underset{df}{=} \frac{p(\xi, \widehat{\vartheta})}{p(\xi, \widehat{\vartheta}_H)}, \qquad \lambda \underset{df}{=} \frac{\max\limits_{\vartheta \in \Theta} p(\xi, \vartheta)}{\max\limits_{\vartheta \in \mathcal{K}} p(\xi, \vartheta)}.$$

Гипотезу $H : \vartheta \in \mathcal{K}$ предлагается отвергать, если наблюдённое λ чрезмерно велико, т.е. если λ превосходит критическое значение. Это критическое значение следует вычислять, имея распределение статистики λ при гипотезе $\vartheta \in \mathcal{K}$.

Не исключается, вообще говоря, что распределение статистики λ окажется зависящим от истинного значения параметра ϑ, даже если $\vartheta \in \mathcal{K}$, т.е. даже если проверяемая гипотеза верна. К счастью, в этой и последующих гауссовских задачах распределение отношения правдоподобий λ оказывается свободным от влияния ϑ, если $\vartheta \in \mathcal{K}$ (одним и тем же для всех $\vartheta \in \mathcal{K}$). Как говорят, при гипотезе статистика λ оказывается свободной (distribution free). Именно этот результат оправдывает применение здесь метода отношения правдоподобий.

В нашей задаче параметром ϑ служит пара (a, Q), гипотеза H относится к вектору a. Множества $\mathcal{K}$ и Θ определяются в зависимости от того, известна наблюдателю матрица Q или нет. Мы рассмотрим основной для приложения случай неизвестной матрицы Q. Случай известной Q вынесен в упражнения к главе.

Здесь Θ – это множество всех пар (a, Q), где $a \in \mathbb{R}^p$, Q – произвольная положительно определенная $(p \times p)$-матрица. Множество $\mathcal{K}$ состоит из пар (a_0, Q). Как мы установили в гл. 2,

$$\widehat{\vartheta} = (\overline{X}, S_n),$$

где

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad S_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^T.$$

Оценкой наибольшего правдоподобия для неизвестной матрицы Q по выборке из $N_p(a_0, Q)$, т.е. при $H : \vartheta \in \mathcal{K}$, служит

$$\widetilde{Q} = \frac{1}{n} \sum_{i=1}^{n} (X_i - a_0)(X_i - a_0)^T.$$

(Последнее утверждение составляет результат Упражнения 1 к Гл. 2). Таким образом,

$$\widehat{\vartheta}_H = (a_0, \widetilde{Q}).$$

Отсюда статистика λ равна

$$\lambda = \left(\frac{\sqrt{\det \widetilde{Q}}}{\sqrt{\det S_n}} \right)^n \times \exp\left\{ -\frac{1}{2} \Big[\sum_{i=1}^n (X_i - \overline{X})^T S_n^{-1} (X_i - \overline{X}) - \sum_{i=1}^n (X_i - a_0)^T \widetilde{Q}^{-1} (X_i - a_0) \Big] \right\}.$$

Уже известным приёмом преобразуем

$$\sum_{i=1}^n (X_i - a_0)^T \widetilde{Q}^{-1} (X_i - a_0) = \mathrm{tr}\Big[\widetilde{Q}^{-1} \sum_{i=1}^n (X_i - a_0)(X_i - a_0)^T \Big] = \mathrm{tr}[nI] = np.$$

Аналогично

$$\sum_{i=1}^n (X_i - \overline{X})^T S_n^{-1} (X_i - \overline{X}) = \mathrm{tr}\Big[S_n^{-1} \sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})^T \Big] = \mathrm{tr}[nI] = np.$$

Поэтому статистика λ равна

$$\lambda = \left(\frac{\det\left[\sum\limits_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})^T + n(\overline{X} - a_0)(\overline{X} - a_0)^T \right]}{\det\left[\sum\limits_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})^T \right]} \right)^{\frac{n}{2}}.$$

Согласно критерию отношения правдоподобий, проверяемую гипотезу H следует отвергать при больших значениях статистики λ. Эквивалентное правило — отвергать H при больших значениях статистики $\lambda^{2/n}$. Это позволяет сформулировать промежуточное правило для проверки гипотезы H — отвергать гипотезу $H : a = a_0$ при больших значениях статистики

$$\widetilde{\lambda} \underset{df}{=} \frac{\det\left[\sum\limits_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T + n(\overline{X} - a_0)(\overline{X} - a_0)^T\right]}{\det\left[\sum\limits_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T\right]}.$$

Эту статистику можно упростить, если воспользоваться следующей леммой об определителях блочных матриц.

Лемма 4. *Пусть квадратная матрица Q разбита на блоки, причем Q_{11}, Q_{22} – квадратные невырожденные матрицы:*

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}.$$

Тогда

$$\det Q = \det Q_{11} \det(Q_{22} - Q_{21}Q_{11}^{-1}Q_{12}) = \det Q_{22} \det(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}).$$

Доказательство. При доказательстве надо вспомнить, как перемножаются блочные матрицы (размеры их блоков согласованы так, чтобы их перемножение было возможно) и что

$$\det\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \det A \det C,$$

если A, B – квадратные матрицы.

Смотри выкладку:

$$\begin{aligned}
\det\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} &= \det\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \det\begin{pmatrix} I, & -Q_{11}^{-1}Q_{12} \\ 0, & I \end{pmatrix} \\
&= \det\begin{pmatrix} Q_{11}\cdot I + Q_{12}\cdot 0, & -Q_{11}Q_{11}^{-1}Q_{12} + Q_{12}\cdot I \\ Q_{21}\cdot I + Q_{22}\cdot 0, & -Q_{21}Q_{11}^{-1}Q_{12} + Q_{22}\cdot I \end{pmatrix} \\
&= \det Q_{11} \det(Q_{22} - Q_{21}Q_{11}^{-1}Q_{12}).
\end{aligned}$$

Вторая формула доказывается аналогично:

$$\begin{aligned}
\det\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} &= \det\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \det\begin{pmatrix} I & 0 \\ -Q_{22}^{-1}Q_{21} & I \end{pmatrix} \\
&= \det\begin{pmatrix} Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}, & -Q_{11}\cdot 0 + Q_{12}\cdot I \\ Q_{21} - Q_{22}Q_{22}^{-1}Q_{21}, & -Q_{21}\cdot 0 + Q_{22}\cdot I \end{pmatrix} \\
&= \det(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}) \det Q_{11}.
\end{aligned}$$

Доказательство леммы окончено. □

Вернемся к статистике λ, к её полученной выше промежуточной форме $\widetilde{\lambda}$. Определитель, который стоит в числителе этой формулы, представим как

$$D \underset{df}{=} \det \begin{pmatrix} \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T, & -\sqrt{n}\,(\overline{X} - a_0) \\ \sqrt{n}\,(\overline{X} - a_0)^T, & 1 \end{pmatrix}.$$

В том, что это верно, нас убеждает вторая из формул леммы 4. Применим теперь для вычисления D первую из формул леммы 4. Попутно выражение $\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T$ заменим на $(n-1)S$. Получим, что D, т.е. числитель $\widetilde{\lambda}$, равен

$$D = \det\Big[\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T\Big]$$
$$\times \det\left\{1 + n(\overline{X} - a_0)^T\Big[\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T\Big]^{-1}(\overline{X} - a_0)\right\}.$$

Вспомним, что в § 1 данной главы по методу Роя мы получили для проверки гипотезы H статистику Хотеллинга

$$T^2 = n(\overline{X} - a_0)^T S^{-1}(\overline{X} - a_0).$$

После упрощений и с использованием T^2 статистика $\widetilde{\lambda}$ превращается в

$$\det\Big(1 + \frac{1}{n-1}T^2\Big).$$

Впрочем, символ $\det$ следует опустить, так как он относится к скалярной величине. В итоге получаем, что гипотезу H следует отвергать, если

$$T^2 \geqslant \text{ const},$$

что совпадает с критерием Роя.

Полученный результат сформулируем в виде теоремы.

Теорема 3.2. *Пусть $X_1, \ldots, X_n$ – выборка из $N_p(a, Q)$. Для проверки гипотезы*

$$H : a = a_0,$$

где вектор a_0 фиксирован, метод отношения правдоподобий даёт критическую статистику Хотеллинга T^2. Отвергать гипотезу H следует при больших значениях T^2.

Мы видим, что в данной задаче оба известных метода построения статистических критериев привели к одному результату — статистике Хотеллинга.

Это укрепляет как сам результат, так и каждый из методов. Далее в подобных, но более сложных задачах мы будем действовать так же.

§ 3. Распределение статистики Хотеллинга

Цель этого параграфа — доказательство следующей теоремы.

Теорема 3.3. *Если $n > p$, то*

$$T^2 = n(\overline{X} - a_0)^T S^{-1}(\overline{X} - a_0) \stackrel{d}{=} \frac{(n-1)p}{n-p} F(p, n-p, \Delta),$$

где параметр нецентральности Δ равен

$$\Delta = n(a - a_0)^T Q^{-1}(a - a_0).$$

Напомним некоторые определения.

Говорят, что случайная величина $\chi^2(m, \Delta)$ следует распределению хи-квадрат с $m \geqslant 1$ степенями свободы и параметром нецентральности $\Delta \geqslant 0$, если

$$\chi^2(m, \Delta) = (\eta_1 + a_1)^2 + \ldots + (\eta_m + a_m)^2,$$

где случайные величины $\eta_1, \ldots, \eta_m$ — независимые $N(0, 1)$, числа $a_1, \ldots, a_m$ произвольны, но таковы, что

$$a_1^2 + \ldots + a_m^2 = \Delta.$$

Если $\Delta = 0$, то распределение хи-квадрат называют центральным и пишут $\chi^2(m)$; при $\Delta > 0$ распределение называют нецентральным.

Говорят, что случайная величина $F(m_1, m_2, \Delta)$ следует эф-распределению с $m_1 \geqslant 1$, $m_2 \geqslant 1$ степенями свободы и параметром нецентральности $\Delta \geqslant 0$, если

$$F(m_1, m_2, \Delta) = \frac{\frac{1}{m_1}\chi^2(m_1, \Delta)}{\frac{1}{m_2}\chi^2(m_2)},$$

причём стоящие в числителе и знаменателе случайные величины независимы. В случае $\Delta = 0$ распределение называют центральным и пишут $F(m_1, m_2)$; в случае $\Delta > 0$ – нецентральным.

Теорема 3.3 вытекает из следующей теоремы.

Теорема 3.4. *Пусть $Y, Y_1, \ldots, Y_k$ – независимые случайные векторы $N_p(0, Q)$, причём матрица Q не вырождена. Положим*

$$W = \sum_{i=1}^{k} Y_i Y_i^T.$$

Заметим, что при $k \geqslant p$ матрица W не вырождена (с вероятностью 1). Пусть

$$\mathcal{T}^2 = (Y + \nu)^T W^{-1} (Y + \nu),$$

где $\nu \in \mathbb{R}^p$ – произвольный вектор. Тогда при $k \geqslant p$

$$\mathcal{T}^2 = \frac{\chi^2(p, \Delta)}{\chi^2(k - p + 1)},$$

причём случайные величины в числителе и знаменателе независимы. Отсюда

$$\mathcal{T}^2 = \frac{p}{k - p + 1} F(p, k - p + 1, \Delta).$$

Доказательство теоремы 3.3. Покажем, что теорема 3.3 следует из теоремы 3.4. Напомним, что $\overline{X}$ и $(n-1)S = \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^T$ независимы, причём

$$\overline{X} \stackrel{d}{=} N_p\Big(a, \frac{1}{n} Q\Big), \qquad (n-1)S \stackrel{d}{=} W_p(n-1, Q) \stackrel{d}{=} Q^{1/2} W_p(n-1, I) Q^{1/2}.$$

Положим $Y = \sqrt{n}\, Q^{-1/2}(\overline{X} - a) \stackrel{d}{=} N_p(0, I)$. Тогда $\sqrt{n}\,(\overline{X} - a_0) = Q^{1/2}(Y + \nu)$, где $\nu = \sqrt{n}\, Q^{-1/2}(a - a_0)$. Отсюда

$$T^2 = [Q^{1/2}(Y + \nu)]^T [Q^{1/2} W_p(n-1, I) Q^{1/2}] [Q^{1/2}(Y + \nu)].$$

Поэтому статистика Хотеллинга T^2 распределена так же, как случайная величина $\mathcal{T}^2$. $\square$

Доказательство теоремы 3.4 существенно опирается на следующую далее лемму. Чтобы её сформулировать, введём обозначения для элементов матриц Q, Q^{-1}, W, W^{-1}. Уже отмечалось, что при $k \geqslant p$ матрица W не вырождена с вероятностью 1. Положим

$$Q = \|q_{ij}\|, \quad Q^{-1} = \|q^{ij}\|, \quad W = \|w_{ij}\|, \quad W^{-1} = \|w^{ij}\|.$$

Лемма 5. *В указанных выше обозначениях,*

(a) $\dfrac{q^{11}}{w^{11}} \stackrel{d}{=} \chi^2(k - p + 1)$,

(b) $\dfrac{l^T Q^{-1} l}{l^T W^{-1} l} \stackrel{d}{=} \chi^2(k-p+1)$ *для любого* $l \in \mathbb{R}^p,\ l \neq 0$.

Доказательство леммы 5, п. (а). Рассмотрим случайный вектор $X \sim N_p(0, Q)$, $X = (x_1, \dots, x_p)^T$. Разобьем X на подвекторы $X^{(1)} = x_1, X^{(2)} = (x_2, \dots, x_p)^T$. Соответственно разобьем на блоки матрицы Q и W:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \quad W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}.$$

Здесь $Q_{11} = q_{11}, W_{11} = w_{11}$, матрицы Q_{22} и W_{22} – квадратные матрицы размера $p-1$.

Как мы отмечали в § 6,

$$\mathcal{L}(X^{(1)} \mid X^{(2)}) = N(Q_{12}Q_{22}^{-1}X^{(2)}, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}).$$

Заметим, что в данном случае, когда размерность $X^{(1)}$ равна единице,

$$Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} = (q^{11})^{-1}.$$

Действительно, по определению

$$q^{11} = \frac{\det Q_{22}}{Q},$$

а в силу леммы 3.2

$$\det Q = \det Q_{22} \cdot \det(Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}).$$

Аналогично

$$(w^{11})^{-1} = W_{12} - W_{12}W_{22}^{-1}W_{21}. \tag{3.1}$$

Как было отмечено в § 7,

$$X^{(1)} = Q_{12}Q_{22}^{-1}X^{(2)} + \xi,$$

где ξ – независимая от $X^{(2)}$ случайная величина, распределённая по гауссовскому закону $N(0, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})$, т. е. в силу полученного выше по закону $N(0, (q^{11})^{-1})$.

Для случайных векторов $Y_1, Y_2, \dots, Y_k$, фигурирующих в условии теоремы, сказанное означает, что

$$Y_i^{(1)} = Q_{12}Q_{22}^{-1}Y_i^{(2)} + \xi_i \quad \text{для} \quad i = \overline{1,k},$$

где $\xi_1, \dots, \xi_k$ — независимые (и не зависящие от $Y_1^{(2)}, \dots, Y_k^{(2)}$) гауссовские $N\left(0, (q^{11})^{-1}\right)$.

Положим
$$R_0^2 = \min_{\beta\in\mathbb{R}^{p-1}} \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^2 .$$

При фиксированных значениях $Y_1^{(2)}, \ldots, Y_k^{(2)}$ случайная величина R_0^2 распределена как $(q^{11})^{-1}\chi^2(k-p+1)$. Это следует из одномерной теории гауссовской линейной регрессии. Поскольку $Y_1^{(2)}, \ldots, Y_k^{(2)}$ – случайные величины, сказанное выше относится к условному распределению R_0^2 при фиксированных $Y_1^{(2)}, \ldots, Y_k^{(2)}$. Поскольку при любых значениях $Y_1^{(2)}, \ldots, Y_k^{(2)}$ условное распределение R_0^2 одно и то же, оно не зависит от условия и поэтому совпадает с (безусловным) распределением R_0^2. Итак, установлено, что случайная величина R_0^2
$$R_0^2 = (q^{11})^{-1}\chi^2(k-p+1).$$

Остаётся лишь указать для R_0^2 более явное выражение. Для этого можно воспользоваться готовыми формулами из линейной регрессии, но проще проделать необходимые выкладки непосредственно. Найдем сначала значение $\widehat{\beta}$, при котором $\sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^2$ достигает минимума. Введём функцию

$$Q(\beta) \underset{df}{=} \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^2 = \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)\left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^T .$$

Пусть $d\beta$ обозначает приращение β. Вычислим приращение $Q(\cdot)$, представив его в виде
$$Q(\beta + d\beta) - Q(\beta) = \nabla Q(\beta) d\beta + o(d\beta).$$

Здесь $\nabla Q(\beta)$ — градиент функции $Q(\cdot)$ в точке β, ∇Q – вектор-строка. Прямым вычислением находим, что

$$Q(\beta + d\beta) - Q(\beta) = -2\sum_{i=1}^{k}(Y_i^{(1)} - \beta^T Y_i^{(2)})(Y_i^{(2)})^T d\beta + o(d\beta).$$

Приравняв нулю полученное для ∇Q выражение, находим, что градиент обращается в нуль (а функция $Q(\beta)$ соответственно достигает минимального значения) при

$$\widehat{\beta} = \left[\sum_{i=1}^{k} Y_i^{(2)}\left(Y_i^{(2)}\right)^T\right]^{-1}\left[\sum_{i=1}^{k} Y_i^{(1)}\left(Y_i^{(2)}\right)^T\right].$$

Подставив полученное значение $\widehat{\beta}$ в $Q(\beta)$, найдем, что

$$R_0^2 = \sum_{i=1}^{k} Y_i^{(1)} Y_i^{(1)} - \left(\sum_{i=1}^{k} Y_i^{(1)} Y_i^{(2)}\right)^T \left(\sum_{i=1}^{k} Y_i^{(2)} Y_i^{(2)}\right)^{-1} \left(\sum_{i=1}^{k} Y_i^{(1)} Y_i^{(2)}\right)^T.$$

Теперь заметим, что это выражение составлено из блоков матрицы $W = \sum_{i=1}^{k} Y_i Y_i^T$, которые мы указали ранее. Поэтому

$$R_0^2 = W_{11} - W_{12} {W_{22}}^{-1} W_{21} = \left(w^{11}\right)^{-1}.$$

Выше мы показали, что $R_0^2 = \left(q^{11}\right)^{-1} \chi^2(k-p+1)$. Отсюда

$$\left(w^{11}\right)^{-1} \stackrel{d}{=} \left(q^{11}\right)^{-1} \chi^2(k-p+1),$$

что и составляет п. (a) доказываемой леммы. □

Доказательство леммы 5, п. (b). Выберем вектор l таким, что $|l| = 1$. Рассмотрим ортогональную матрицу B, первая строка которой l^T, в остальном произвольную. Введём случайные векторы $\widetilde{Y}_1, \ldots, \widetilde{Y}_k$, положив $\widetilde{Y}_i = BY_i, i = \overline{1,k}$. Ясно, что $\widetilde{Y}_1, \ldots, \widetilde{Y}_k$ — независимые $N_p(0, \widetilde{Q})$, где $\widetilde{Q} = BQB^T$. Введём также $\widetilde{W} = \sum_{i=1}^{k} \widetilde{Y}_i \widetilde{Y}_i^T = BWB^T$. Элементы матриц $\widetilde{Q}, \widetilde{W}, \widetilde{Q}^{-1}, \widetilde{W}^{-1}$ будем обозначать так же, как и прежде элементы матриц Q, W, Q^{-1}, W^{-1}, но снабжать дополнительной волной сверху. Например, $\widetilde{q}^{11}, \widetilde{w}^{11}$ — элементы с номерами (1, 1) матриц $\widetilde{Q}^{-1}, \widetilde{W}^{-1}$ соответственно.

Заметим, что в силу выбора матрицы B

$$\begin{aligned} l^T Q^{-1} l &= (Bl)^T \left[\left(B^T\right)^{-1} Q^{-1} B^{-1}\right] (Bl) \\ &= (1, 0, \ldots, 0)^T \left(\widetilde{Q}\right)^{-1} (1, 0, \ldots, 0) = \widetilde{q}^{11}. \end{aligned}$$

Аналогично

$$l^T W^{-1} l = \widetilde{w}^{11}.$$

Для векторов $\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_k$, матриц $\widetilde{Q}, \widetilde{W}$ и их элементов верно всё то, что было установлено в п. (a) леммы 5 для векторов $Y_1, Y_2, \ldots, Y_k$ и матриц Q, W. Поэтому при любом $l \neq 0$

$$\frac{l^T Q^{-1} l}{l^T W^{-1} l} = \frac{\widetilde{q}^{11}}{\widetilde{w}^{11}} \stackrel{d}{=} \chi^2(k-p+1)$$

в силу п. (a) доказываемой леммы. Доказательство леммы 5 окончено. □

Доказательство теоремы 3.4. Представим
$$\mathcal{T}^2 = (Y+\nu)^T W^{-1}(Y+\nu)$$
в виде
$$\mathcal{T}^2 = \frac{(Y+\nu)^T W^{-1}(Y+\nu)}{(Y+\nu)^T Q^{-1}(Y+\nu)}(Y+\nu)^T Q^{-1}(Y+\nu).$$
В силу леммы 5, п. (b), условное распределение при фиксированном Y случайной величины
$$\frac{(Y+\nu)^T Q^{-1}(Y+\nu)}{(Y+\nu)^T W^{-1}(Y+\nu)}$$
это распределение $\chi^2(k-p+1)$. Заметим, что это условное распределение одно и то же при всех значениях Y. Следовательно, «безусловное» или просто распределение этой дроби — тоже распределение $\chi^2(k-p+1)$. При этом, как случайная величина, эта дробь не зависит от Y. Следовательно, она не зависит и от
$$(Y+\nu)^T Q^{-1}(Y+\nu).$$

Как известно, квадратичная форма $(Y+\nu)^T Q^{-1}(Y+\nu)$ распределена по закону $\chi^2(p,\Delta)$ с параметром нецентральности $\Delta = \nu^T Q^{-1}\nu$ (как в утверждениях теоремы).

В результате получаем, что
$$\mathcal{T}^2 = \frac{\chi^2(p,\Delta)}{\chi^2(k-p+1)},$$
что и требовалось доказать. □

Проверка гипотезы $H : a = a_0$. На основании доказанной выше теоремы 3.3 мы теперь можем предложить статистическое правило для проверки гипотезы $H : a = a_0$. Так как распределения $F(p, n-p, \Delta)$ при фиксированных n и p стохастически упорядочены относительно параметра $\Delta \geqslant 0$, то гипотезу H следует отвергать при «слишком больших» значениях T^2. Отсюда следует

Статистическое правило. *По выборке $X_1, \ldots, X_n$ из $N_p(a, Q)$ отвергать гипотезу $H : a = a_0$ на уровне α, если*
$$n(\overline{X} - a_0)^T S^{-1}(\overline{X} - a_0) > \frac{(n-1)p}{n-p} F_{1-\alpha}(p, n-p),$$
где $F_{1-\alpha}(p, n-p)$ обозначает $(1-\alpha)$-квантиль $F(p, n-p)$.

Правило действует при $n > p$ и предназначено для случая неизвестной матрицы Q.

§ 4. Проверка приближённой гипотезы $H^* : a \approx a_0$

Статистические гипотезы часто высказывают как утверждения о точном значении того или иного функционала от распределения вероятностей или о точных значениях нескольких функционалов. Типичный пример — рассматриваемая в этой главе гипотеза $H : a = a_0$, где a_0 — заданный вектор; эту гипотезу предлагается проверить по выборке из $N_p(a, Q)$. Задача статистики в таком случае состоит в том, чтобы выяснить, совместимо или нет утверждение гипотезы с имеющимися наблюдениями.

Никто, впрочем, не принимает столь точно высказанного предположения буквально. Все понимают, что истинное значение обсуждаемого функционала может несколько отличаться от гипотетического: $a \approx a_0$. Гипотеза $H_0 : a = a_0$ принимается (и считается истинной), если различием между истинным a и его гипотетическим значением a_0 можно пренебречь. Точнее, гипотезу мы принимаем, если это различие мало по сравнению со случайным разбросом критической статистики.

Однако лишь приблизительная правильность статистической гипотезы, приемлемая при малом или умеренном объёме наблюдений, непременно обнаруживается, когда объём наблюдений возрастает, что и приводит к явлению, хорошо известному специалистам: при большом объёме наблюдений всякая статистическая гипотеза отвергается. Чтобы этого избежать, придадим предположению

$$H^* : a \approx a_0$$

точный смысл.

Начнем с одномерного случая — с проверки по выборке $x_1, \ldots, x_n$ из $N(a, \sigma^2)$ гипотезы

$$H^* : a \approx a_0,$$

где a_0 – заданное число (используем в одномерном случае для ожидаемых значений те же обозначения, что и в многомерном). Ясно, что следует говорить не об абсолютной, а об относительной близости истинного a к гипотетическому значению a_0, т. е. о малости $a - a_0$ по сравнению с σ, мерой случайного разброса наблюдений. С этим легко согласиться, если вспомнить, что мощность разумных статистических критериев, применяемых для проверки гипотезы $H : a = a_0$, зависит от $\frac{a-a_0}{\sigma}$ (критерий Стьюдента, например). Поэтому точной формой для приближенной гипотезы $a \approx a_0$ может служить утверждение

$$H^*_\varepsilon : \left| \frac{a - a_0}{\sigma} \right| \leqslant \varepsilon,$$

где $\varepsilon > 0$ – некоторое заданное (малое) число.

В многомерном случае мысль об относительной малости вектора $a - a_0$ можно высказать так: для вектора $u \in \mathbb{R}^p$ математическое ожидание случайной величины $u^T(X - a_0)$ мало́ по сравнению с ее стандартным отклонением, т. е. с величиной $\sqrt{u^T Q u}$:

$$\left| \frac{u^T(a - a_0)}{\sqrt{u^T Q u}} \right| \leqslant \varepsilon.$$

В эквивалентной форме

$$\max_{u: u^T Q u = 1} u^T(a - a_0) \leqslant \varepsilon.$$

Согласно лемме 3, эта величина равна $\sqrt{(a - a_0)^T Q^{-1}(a - a_0)}$. Поэтому приближённая гипотеза $a \approx a_0$ в точной форме (при заданном $\varepsilon > 0$) может быть высказана как

$$H_\varepsilon^* : (a - a_0)^T Q^{-1}(a - a_0) \leqslant \varepsilon^2.$$

Для проверки «точной» гипотезы $a = a_0$ по выборке $X_1, ..., X_n$ из $N_p(a, Q)$ была выведена статистика Хотеллинга

$$T^2 = n(\overline{X} - a_0)^T S^{-1}(\overline{X} - a_0).$$

При $n > p$

$$T^2 = \frac{(n-1)p}{n-p} F(p, n-p, \Delta),$$

где параметр нецентральности

$$\Delta = (a - a_0)^T Q^{-1}(a - a_0).$$

Мы видим, что гипотеза H_ε^* для статистики T^2 означает, что ее параметр нецентральности не превосходит $n\varepsilon^2$:

$$H_\varepsilon^* : \Delta \leqslant n\varepsilon^2.$$

Воспользуемся для проверки H_ε^* статистикой T^2 и станем отвергать H_ε^* при больших значениях T^2. Критическое значение уровня α, скажем $t(\alpha)$, определяем как решение относительно t экстремальной задачи:

$$\max_{\Delta: \Delta \leqslant n\varepsilon^2} P\left\{ \frac{(n-1)p}{n-p} F(p, n-p, \Delta) \leqslant t \right\} = 1 - \alpha.$$

Так как случайные величины $F(\nu_1, \nu_2, \Delta)$ стохастически упорядочены относительно $\Delta \geqslant 0$, то максимальное значение указанной вероятности достигается при $\Delta = n\varepsilon^2$. Поэтому критическое значение $t(\alpha)$ есть решение уравнения

$$P\left\{F(p, n-p, n\varepsilon^2) \leqslant \frac{n-p}{(n-1)p} t(\alpha)\right\} = 1 - \alpha.$$

Статистическое правило. *Гипотезу $H^*_\varepsilon : (a - a_0)^T Q^{-1}(a - a_0) \leqslant \varepsilon^2$ по выборке объёма n из распределения $N_p(a, Q)$ следует отвергнуть на гарантированном уровне α, если*

$$T^2 \geqslant t(\alpha),$$

где критическое значение $t(\alpha)$ задано выше.

§ 5. Доверительные выводы

Как обычно, обладание статистическим критерием для проверки гипотезы о точном значении параметра позволяет нам указать и правила для построения доверительных множеств для этого параметра. Доверительное множество (с доверительной вероятностью $1 - \alpha$) для неизвестного истинного значения параметра ϑ_0 из параметрического множества Θ составляют те значения параметра $\vartheta \in \Theta$, для которых гипотеза $H : \vartheta = \vartheta_0$ не отвергается (на уровне α) упомянутым выше статистическим критерием.

Доверительные эллипсоиды. В этой главе мы обсуждаем выборку $X_1, ..., X_n$ из нормального распределения $N_p(a, Q)$. Для проверки гипотезы $H : \vartheta = \vartheta_0$ мы можем применить статистику Хотеллинга T^2. В этом случае в качестве доверительного множества для неизвестного $a \in \mathbb{R}^p$ (при доверительной вероятности $1 - \alpha$) мы получаем эллипсоид с центром $\overline{X}$:

$$\left\{a : n(\overline{X} - a)^T S^{-1}(\overline{X} - a) < \frac{(n-1)p}{n-p} F_{1-\alpha}(p, n-p)\right\},$$

если $n > p$. (Понятно, что это правило дает доверительное множество для a_0 и при известной матрице Q.) Здесь $F_{1-\alpha}(p, n-p)$ – это $(1-\alpha)$-квантиль эф-распределения (центрального) с $p, n-p$ степенями свободы.

Множественные сравнения. Полагаясь на доверительные эллипсоиды указанного вида, можно предложить метод для одновременного доверительного оценивания линейных форм от вектора a, т. е. величин вида $c^T a$, где c – заданный вектор, $c \in \mathbb{R}^p$. Одновременное доверительное оценивание нескольких разных линейных форм называют также *множественным сравнением* (компонент вектора a). Сравнение такого вида бывает необходимо, в частности, в дисперсионном анализе. Оттуда еще одно название: формы вида $c^T a$ называют *контрастами*.

В данной задаче мы сможем получать доверительные интервалы вида $\left|c^T(a - \overline{X})\right| < d$. Величина d находится в зависимости от c и доверительной вероятности. Если задан конечный набор векторов $c_1, c_2, ..., c_k \in \mathbb{R}^p$ и доверительная вероятность $1 - \alpha$, мы найдём такие числа $d_1, ..., d_k$, что

$$P\left\{\left|c_1^T(a - \overline{X})\right| \leqslant d_1, \ldots, \left|c_k^T(a - \overline{X})\right| \leqslant d_k\right\} \geqslant 1 - \alpha.$$

Если векторы $c \in \mathbb{R}^p$ образуют некоторое заданное (произвольное) множество C, то мы найдём такую функцию $d(c)$, что

$$P\left\{\left|c^T(a - \overline{X})\right| < d(c), \quad \text{для любых } c \in C\right\} \geqslant 1 - \alpha.$$

Пусть $Э$ обозначает эллипсоид, который накрывает истинное значение параметра a с вероятностью $1 - \alpha$. Для вектора $c \in C$ рассмотрим пару параллельных гиперплоскостей вида

$$\left\{\vartheta : c^T(\vartheta - \overline{X}) = \text{ const}\right\},$$

касающихся эллипсоида $Э$. Две эти гиперплоскости «вырезают» в пространстве $\mathbb{R}^p$ слой, содержащий $Э$. Этот слой – множество

$$\left\{\vartheta : \left|c^T(\vartheta - \overline{X})\right| \leqslant d(c)\right\},$$

где

$$d(c) = \max_{\vartheta \in Э} \left|c^T(\vartheta - \overline{X})\right|.$$

Если C – заданное множество векторов c, то пересечение слоев, каждый их которых определен некоторым $c \in C$, образует некоторое выпуклое множество. Назовем его G. Ясно, что $Э \subset G$. Поэтому множество G накрывает истинное значение a с вероятностью не меньшей, чем это делает $Э$:

$$P(\vartheta \in G) \geqslant P(\vartheta \in Э) = 1 - \alpha.$$

Это означает, что

$$P\left(-d(c) + c^T\overline{X} \leqslant c^T\vartheta \leqslant d(c) + c^T\overline{X}, \quad \text{для любых } c \in C\right) \geqslant 1 - \alpha.$$

Чтобы получить окончательный ответ, остается лишь вычислить $d(c)$. Ясно, что

$$d(c) = \max_{\vartheta : (\vartheta - \overline{X})^T S^{-1}(\vartheta - \overline{X}) = k^2} c^T(\vartheta - \overline{X}),$$

где
$$k^2 = \frac{p(n-1)}{n(n-p)} F_{1-\alpha}(p, n-p).$$

Если положить $u = \vartheta - \overline{X}, R = k^{-2} S^{-1}$, то ответ даёт лемма 3:
$$d(c) = k\sqrt{c^T S c}\,.$$

С этим значением $d(c)$ получаем явную форму множественных сравнений [H. Sheffe, 1952]:
$$P\left(-k\sqrt{c^T S c} + c^T\overline{X} < c^T\vartheta > k\sqrt{c^T S c} + c^T\overline{X}, \quad \text{для любых } c \in C\right) \geqslant (1-\alpha),$$

где k было указано выше:
$$k = \sqrt{\frac{p(n-1)}{(n-p)n}}\,\sqrt{F_{1-\alpha}(p, n-p)}\,.$$

§ 6. Две нормальные выборки

Применим полученные выше результаты к анализу двух нормальных выборок. Пусть $X_1, \ldots, X_m$ и $Y_1, \ldots, Y_n$ – это независимые выборки из $N_p(a, Q)$ и $N_p(b, Q)$ соответственно. Наши цели:

1. указать способ для проверки гипотезы
$$H : a = b;$$
2. указать доверительные множества для $(a-b)$.

Рассмотрим статистики $\overline{X}, \overline{Y}$ и S:
$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \qquad \overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j,$$
$$S = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i - \overline{X})(X_i - \overline{X})^T + \sum_{j=1}^{n}(Y_j - \overline{Y})(Y_j - \overline{Y})^T\right].$$

Заметим, что статистики $\overline{X}, \overline{Y}$ и S независимы.

Очевидно, что
$$\sqrt{\frac{mn}{m+n}}\,(\overline{X} - \overline{Y}) \sim N_p\Big(\sqrt{\frac{mn}{m+n}}\,(a-b), Q\Big).$$

Рассмотрим статистику

$$T^2 = \frac{mn}{m+n}(\overline{X} - \overline{Y})^T S^{-1}(\overline{X} - \overline{Y}).$$

В силу теоремы 3.3

$$T^2 = \frac{(m+n-2)p}{m+n-2-p+1} F(p, m+n-p-1, \Delta),$$

где

$$\Delta = \frac{mn}{m+n-2}(a-b)^T Q^{-1}(a-b).$$

Гипотеза $H : a = b$ эквивалентна гипотезе $\Delta = 0$.

Для проверки гипотезы $H : a = b$ теперь можно применить правило из § 3 данной главы с очевидными изменениями.

Для построения доверительных множеств и множественных сравнений для $(a - b)$ можно применить результаты § 5 данной главы с очевидными изменениями.

§ 7. Упражнения

1. Пусть Q – (симметричная) положительно определенная матрица, a – вектор. Разобьем их на блоки:

   $$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \qquad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

   (Размерности блоков согласованы).
   Покажите, что

   $$a^T Q^{-1} a \geqslant a_1^T Q_{11}^{-1} a_1.$$

   (Смысл задачи в том, что $\sqrt{a^T Q^{-1} a}$ — «правильная» длина вектора a — уменьшается, если вместо полного вектора оставить лишь его часть.)
2. Пусть $Q = \|q_{ij}\|$ — неотрицательно определенная $(p \times p)$-матрица.
   Покажите, что $\det Q \leqslant \prod_{i=1}^{p} q_{ii}$.
3. По выборке $x_1, \ldots, x_n$ из $N_p(a, Q)$, где a и Q неизвестны, укажите область, в которую с заданной доверительной вероятностью попадёт будущее независимое наблюдение x_{n+1} из $N_p(a, Q)$.
4. Даны две независимые выборки: $X_1, \ldots, X_m$ из $N_p(a, Q)$ и $Y_1, \ldots, Y_n$ из $N_p(b, Q)$; векторы a, b и матрица Q неизвестны.

   (a) Как проверить гипотезу $H : a = 2b$?

   (b) Укажите доверительное множество для $(2a + b)$.

(c) Найдите для Q оценку наибольшего правдоподобия.

5. Пусть $X_1, \dots, X_n$ – выборка из $N_p(a, Q)$, причём матрица Q известна и обратима.
   Примените метод Роя нахождения критической статистики для проверки гипотезы $H_0 : a = a_0$. Найдите распределение этой статистики.
6. Пусть $X_1, \dots, X_n$ – выборка из $N_p(a, Q)$, причём матрица Q известна и обратима.
   Вычислите критерий отношения правдоподобий для проверки гипотезы $H_0 : a = a_0$. Найдите распределение статистики этого критерия.
7. Обозначим $W_p(m, \Sigma)$ случайную матрицу, распределённую по Уишарту.
   (a) Покажите, что $z^T W_p(m, \Sigma) z = (z^T \Sigma z)\chi^2(m)$, где $z \in \mathbb{R}^p$, $\chi^2(m)$ – случайная величина, следующая распределению хи-квадрат с m степенями свободы.

   (b) Используя лемму 5, покажите для $m - p \geqslant 2$, что
   $$\mathrm{E}\, W_p^{-1}(m, \Sigma) = \frac{1}{m - p - 1}\Sigma^{-1}.$$
   (Может быть полезен тот факт, что $\mathrm{E}[\chi^2(m)]^{-1} = \frac{1}{m-2}$ для $m \geqslant 3$.)

Глава 4

НЕСКОЛЬКО НОРМАЛЬНЫХ ВЫБОРОК: ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

В этой главе мы исследуем несколько независимых нормальных выборок, возможно отличающихся параметрами положения. Параметры масштаба, т.е. матрицы ковариаций, будем считать одинаковыми. Ни математические ожидания, ни общая матрица ковариаций представленных этими выборками нормальных распределений не известны.

Ради возможности дальнейшего обсуждения введём обозначения для параметров и наблюдений и их нумерацию. Пусть Q обозначает упомянутую общую $(p \times p)$-матрицу ковариаций, причём $Q > 0$ (не вырождена). Пусть $k \geqslant 2$ – число выборок. Выборкам и представленным ими нормальным распределениям дадим номера $1, 2, \ldots, k$. Пусть $a_1, a_2, \ldots, a_k$ – их математические ожидания; $a_1, a_2, \ldots, a_k \in \mathbb{R}^p$. Выборки, следовательно, представляют нормальные распределения $N_p(a_j, Q), j = \overline{1, k}$. Пусть $n_1, n_2, \ldots, n_k$ – объемы соответствующих выборок. Элементам выборок тоже дадим номера: элементам выборки j, где $j = \overline{1, k}$, дадим номера от 1 до n_j. Пусть X_{ij} – наблюдение i из выборки j, $X_{ij} \in \mathbb{R}^p$; здесь $i = \overline{1, n_j}, j = \overline{1, k}$. Общее число наблюдений обозначим N, $N = n_1 + n_2 + \ldots + n_k$. Предполагаем, что $N \geqslant k + p - 1$.

Всю совокупность наблюдений представляем в виде таблицы

Номер наблюдения╲Номер выборки	1	2	…	k
1	X_{11}	X_{12}	…	X_{1k}
2	X_{21}	X_{22}	…	X_{2k}
⋮	⋮	⋮	⋮	⋮
	$X_{n_1 1}$	$X_{n_2 2}$	…	$X_{n_k k}$
Средние	$X_{.1}$	$X_{.2}$	…	$X_{.k}$

Статистическая модель: наблюдения — независимые случайные векторы, причём $X_{ij} \sim N_p(a_j, Q), j = \overline{1, k}$, где векторы математических ожиданий $a_1, \ldots, a_k$ и матрица Q (общая для всех наблюдений) неизвестны. Считаем, что матрица Q не вырождена.

Основная (исходная для всякого анализа) задача – проверка «нулевой» гипотезы
$$H : a_1 = a_2 = \ldots = a_k.$$
Если эта гипотеза не отвергается, то для различения и сравнения параметров $a_1, a_2, \ldots, a_k$ нет статистических оснований.

Так же, как мы поступали при анализе одной (и двух) нормальных выборок, мы рассмотрим два пути к статистической проверке H — метод Роя и метод отношения правдоподобий. В отличие от упомянутой одновыборочной задачи, здесь эти методы приводят к разным статистическим критериям. Анализируя и сравнивая их, мы сможем увидеть и другие разумные статистические критерии и даже указать их общий вид.

Мы также обсудим законы распределения критических статистик (при гипотезе и альтернативах).

§ 1. Метод Роя

Пусть $u \in \mathbb{R}^p, u \neq 0$ – произвольный вектор. От векторных наблюдений X_{ij} (из таблицы, приведенной в начале главы) переходим к скалярным случайным величинам $x_{ij} = u^T X_{ij}$. По отношению к набору $\{x_{ij}\}$ получаем известную задачу о сравнении k одномерных нормальных выборок из $N(\mu_1, \sigma^2), \ldots, N(\mu_k, \sigma^2)$, где $\mu_j = u^T a_j$ для $j = \overline{1, k}$, $\sigma^2 = u^T Q u$.

Гипотеза $H : a_1 = a_2 = \ldots = a_k$ для данного $u \in \mathbb{R}^p$ превращается в гипотезу
$$H^{(u)} : \mu_1 = \mu_2 = \ldots = \mu_k.$$
Для проверки $H^{(u)}$ можно применить классическую эф-статистику дисперсионного анализа
$$F^{(u)} = \frac{\frac{1}{k-1} \sum\limits_{j=1}^{k} n_j (x_{.j} - \overline{x})^2}{\frac{1}{N-k} \sum\limits_{j=1}^{k} \sum\limits_{i=1}^{n_j} (x_{ij} - x_{.j})^2}.$$
Здесь
$$x_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} = \frac{1}{n_j} u^T X_{.j}, \quad j = \overline{1, k},$$
$$\overline{x} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij} = \frac{1}{N} \sum_{j=1}^{k} n_j x_{.j} = u^T \overline{X} = u^T \Big(\frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij} \Big).$$

Таким образом,

$$F^{(u)} = \frac{u^T \left[\frac{1}{k-1} \sum\limits_{j=1}^{k} n_j (X_{.j} - \overline{X})(X_{.j} - \overline{X})^T \right] u}{u^T \left[\frac{1}{N-k} \sum\limits_{j=1}^{k} \sum\limits_{i=1}^{n_j} (X_{ij} - X_{.j})(X_{ij} - X_{.j})^T \right] u}.$$

Введём матрицы S_1 и S_2, положив

$$S_1 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - X.j)(X_{ij} - X.j)^T$$

и

$$S_2 = \sum_{j=1}^{k} n_j (X_{.j} - \overline{X})(X_{.j} - \overline{X})^T.$$

В принятом нами предположении $N \geqslant p + k - 1$ матрица S_1 не вырождена (с вероятностью 1). Это можно доказать непосредственно, но простое доказательство будет следовать из общих результатов о гауссовских линейных моделях, которые будут даны в гл. 6.

Матрица $\frac{1}{(N-k)} S_1$ – это несмещённая оценка матрицы ковариаций Q, выведенная без обращения к гипотезе H; потому она несмещённо оценивает Q вне зависимости от того, верна или нет гипотеза H. Распределение матрицы S_1 также не связано с гипотезой H. Оно одно и то же как при гипотезе H, так и при ее нарушении.

Матрица $\frac{1}{(k-1)} S_2$ представляет собой оценку матрицы Q, вычисленную в предположении, что гипотеза H верна. Эта оценка основывается на случайных векторах $X_{.1}, \ldots, X_{.k}$, которые распределены как $N(a_1, \frac{1}{n_1} Q), \ldots, N(a_k, \frac{1}{n_k} Q)$. При гипотезе H матрица $\frac{1}{k-1} S_2$ несмещённо оценивает Q. Если же гипотеза H неверна (при альтернативе), то $\frac{1}{k-1} S_2$ как оценка Q приобретает смещение.

С употреблением S_1 и S_2 статистика $F^{(u)}$ обретает вид

$$F^{(u)} = \frac{\frac{1}{k-1} u^T S_2 u}{\frac{1}{N-k} u^T S_1 u}.$$

При гипотезе $H^{(u)}$ статистика $F^{(u)}$ подчиняется центральному эф-распределению с $(k-1,\ N-k)$ степенями свободы. При альтернативе эф-распределение становится нецентральным.

Гипотезу $H^{(u)}$ следует поэтому отвергнуть (на уровне $\alpha > 0$), если

$$F^{(u)} > F_{1-\alpha}(k-1, N-k).$$

Здесь $F_{1-\alpha}(k-1, N-k)$ обозначает $(1-\alpha)$-квантиль центрального эф-распределения с соответствующими числами степеней свободы.

Гипотеза H верна, если верна каждая гипотеза $H^{(u)}$. Против гипотезы $H^{(u)}$ говорят большие значения статистики $F^{(u)}$. Поэтому против H говорит большое значение статистики $\max_u F^{(u)}$. Заметим, что здесь важно не то, каково конкретное выражение для критического уровня при данном $u \in \mathbb{R}^p$, а то, что этот критический уровень общий для всех $u \in \mathbb{R}^p$. Рой предложил для проверки H (против $\overline{H}$) статистику

$$V \underset{df}{=} \max_{u \in \mathbb{R}^p, u \neq 0} \frac{u^T S_2 u}{u^T S_1 u} = \max_{|u|=1} \frac{u^T S_2 u}{u^T S_1 u}.$$

Статистику V называют статистикой Роя. Гипотезу H следует отвергать при больших значениях V.

Для расчета критических уровней статистики V нам придётся найти распределение статистики V при гипотезе. Но сначала надо упростить выражение для V.

Теорема из линейной алгебры: если A – симметричная квадратная матрица, то

$$\sup_{|u|=1} u^T A u = \text{максимальное собственное значение } A.$$

Доказательство этого известного результата, которое мы приводим в порядке напоминания, будет основано на спектральном представлении матрицы A.

Теорема 4.1 (спектральное представление симметричной матрицы)**.** *Пусть A – симметричная матрица порядка $p \times p$, пусть $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$ – ее собственные значения; $U_1, U_2, \ldots, U_p$ – соответствующие им собственные векторы, при этом $U_i^T U_j = \delta_{ij}$. Тогда*

$$A = \lambda_1 U_1 U_1^T + \lambda_2 U_2 U_2^T + \ldots + \lambda_p U_p U_p^T.$$

Доказательство. Эта теорема является одной из форм известного результата «о приведении квадратичной формы к главным осям», а точнее, такого факта: если A – симметричная $(p \times p)$-матрица, то в $\mathbb{R}^p$ существует ортогональный и нормированный базис из собственных векторов матрицы A.

Так как $U_1, \ldots, U_p$ – ортонормированный базис в $\mathbb{R}^p$, то всякий вектор $X \in \mathbb{R}^p$ можно представить как $X = \sum\limits_{k=1}^{p} \xi_k U_k$, где $\xi_1, \ldots, \xi_p$ – некоторые числа, координаты X в базисе $U_1, \ldots, U_p$. Теперь $AX = \sum\limits_{k=1}^{p} \xi_k A U_k = \sum\limits_{k=1}^{p} \xi_k \lambda_k U_k$. Сравним это выражение с таким:

$$\Big(\sum_{j=1}^{p} \lambda_j U_j U_j^T\Big) X = \Big(\sum_{j=1}^{p} \lambda_j U_j U_j^T\Big)\Big(\sum_{k=1}^{p} \xi_k U_k\Big) = \sum_{k=1}^{p}\sum_{j=1}^{p} \xi_k \lambda_j U_j (U_j^T U_k) = \sum_{k=1}^{p} \xi_k \lambda_k U_k.$$

Убеждаемся, что эти выражения совпадают. □

Следствие 4.1.

$$\sup_{|u|=1} u^T A u = \lambda_1.$$

Доказательство. Для произвольного вектора $X \in \mathbb{R}^p$, $|X| = 1$, и квадратичной формы $X^T A X$ напишем их представления в координатной форме:

$$X = \sum_{k=1}^{p} \xi_k U_k, \quad \text{причём} \ \sum_{k=1}^{p} \xi_k^2 = 1;$$

$$X^T A X = \sum_{j=1}^{k} \lambda_j \xi_j^2.$$

Следствие очевидно вытекает из этих формул. □

Теорема 4.2. *Если A, B – симметричные матрицы, причём $A > 0$, то*

$$\max_{u \in \mathbb{R}^p} \frac{u^T B u}{u^T A u} = \textit{максимальный корень уравнения } \det(B - \lambda A) = 0.$$

Доказательство. Существует невырожденная матрица, скажем, F, такая, что $F^T A F = I$. При этом $A = (F^T)^{-1} F^{-1}$. Положим $u = Fv$. Получаем, что

$$\max_{u \in \mathbb{R}^p} \frac{u^T B u}{u^T A u} = \max_{v \in \mathbb{R}^p} \frac{v^T (F^T B F) v}{v^T v} = \max_{|v|=1} v^T (F^T B F) v.$$

В силу следствия 1 это выражение равно максимальному корню уравнения

$$\det(F^T B F - \lambda I) = 0$$

или максимальному корню уравнения

$$\det(B - \lambda (F^T)^{-1} F^{-1}) = 0,$$

т. е. максимальному корню уравнения

$$\det(B - \lambda A) = 0,$$

что и требовалось доказать. $\square$

Возвращаясь к статистике V, заключаем, что

$$V = \text{максимальный корень уравнения } \det(S_2 - \lambda S_1) = 0.$$

Можно также сказать, что

$$V = \text{максимальное собственное значение матрицы } S_2 S_1^{-1}.$$

Полученный результат сформулируем как теорему.

Теорема 4.3. *Пусть $X_{1j}, X_{2j}, \ldots, X_{n_j j}$ – выборка из $N_p(a_j, Q)$, $j = \overline{1,k}$. Все k выборок взаимно независимы. Матрица $Q > 0$ и неизвестна.*

Положим

$$X_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}, \quad \overline{X} = \frac{1}{n_1 + \ldots + n_k}\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij},$$

$$S_1 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - X\cdot j)(X_{ij} - X\cdot j)^T,$$

$$S_2 = \sum_{j=1}^{k} n_j (X_{\cdot j} - \overline{X})(X_{\cdot j} - \overline{X})^T.$$

Применение метода Роя к проверке гипотезы

$$H : a_1 = a_2 = \ldots = a_k$$

приводит к критической статистике

$$V = \max\{\lambda_1, \lambda_2, \ldots, \lambda_p\} = \lambda_1,$$

где $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0$ – корни характеристического уравнения

$$\det(S_2 - \lambda S_1) = 0.$$

Отвергать гипотезу H следует при больших значениях V.

О распределении корней этого и схожих с ним уравнений будем говорить позже. Также позже мы обсудим смысл этого и других (появятся ниже) статистических критериев для проверки гипотезы H.

§ 2. Критерий отношения правдоподобий

О критерии отношения правдоподобий см. начало § 2 главы 3. Критерий отношения правдоподобий имеет вид

$$\lambda = \frac{p(X, \widehat{\vartheta})}{p(X, \widehat{\vartheta}_H)} > \text{ const},$$

где X – наблюдение (в данном случае таблица с элементами X_{ij}, см. начало главы), $p(x, \vartheta)$ – плотность вероятности X в точке x, зависящая от параметра ϑ. В данной модели

$$\vartheta = (a_1, \ldots, a_k, Q).$$

Через $\widehat{\vartheta}$ обозначена оценка наибольшего правдоподобия для ϑ, полученная без гипотезы $H : a_1 = a_2 = \ldots = a_k$; через $\widehat{\vartheta}_H$ обозначена оценка ϑ при этой гипотезе, т. е. оценка параметра

$$(a, \ldots, a, Q).$$

Правдоподобие $p(X, \vartheta)$ в данной модели равно

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{Np} \left(\frac{1}{\sqrt{\det Q}}\right)^{N} \exp\left\{-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - a_j)^T Q^{-1}(X_{ij} - a_j)\right\}.$$

При гипотезе H правдоподобие получается заменой параметра $(a_1, \ldots, a_k, Q)$ на $(a, \ldots, a, Q)$, т. е. замещением параметров $a_1, \ldots, a_k$ их общим (неизвестным) значением a.

(а) Вычисление $\widehat{\vartheta}$. При фиксированном (произвольном) $Q > 0$ правдоподобие достигает наибольшего значения при

$$\widehat{a_1} = X_{\cdot 1}, \ldots, \widehat{a_k} = X_{\cdot k},$$

где $X_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}, j = 1, \ldots, k$.

После такого выбора значений $a_1, \ldots, a_k$ правдоподобие для Q превращается в

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{Np} \left(\frac{1}{\sqrt{\det Q}}\right)^{N} \exp\left\{-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - X_{\cdot j})^T Q^{-1}(X_{ij} - X_{\cdot j})\right\}.$$

Этому выражению надо доставить максимальное значение, выбирая $Q > 0$.

Такая задача была уже (фактически) решена в Главе 2 при отыскании оценок наибольшего правдоподобия для многомерной нормальной выборки (см. § 2). Тогда пришлось решать задачу

$$\left(\frac{1}{\sqrt{\det Q}}\right)^n \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{n} Y_\alpha^T Q^{-1} Y_\alpha\right\} \to \max_Q .$$

Ответ:

$$\widehat{Q} = \frac{1}{n}\sum_{\alpha=1}^{n} Y_\alpha Y_\alpha^T .$$

Прилагая этот ответ к нашей задаче, получаем для Q оценку

$$\widehat{Q} = \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - X_{\cdot j})(X_{ij} - X_{\cdot j})^T.$$

Заметим, что $\widehat{Q}$ очень близка к уже известной нам несмещённой оценке $\frac{1}{N-k}S_1$, так что

$$\widehat{Q} = \frac{1}{N}S_1.$$

Остается вычислить $p(X, \widehat{\vartheta})$. Для этого показатель экспоненты преобразуем уже известным приёмом, используя равенство $\operatorname{tr} AB = \operatorname{tr} BA$ (если матрицы A и B можно перемножать указанным способом и при этом AB и BA – матрицы квадратные). Показатель экспоненты при $\vartheta = \widehat{\vartheta}$ есть

$$\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n_j} \operatorname{tr}(X_{ij} - X_{\cdot j})^T \widehat{Q}^{-1}(X_{ij} - X_{\cdot j}) \\
&= -\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n_j} \operatorname{tr}[\widehat{Q}^{-1}(X_{ij} - X_{\cdot j})(X_{ij} - X_{\cdot j})^T] \\
&= -\frac{1}{2}\operatorname{tr}[\widehat{Q}^{-1}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - X_{\cdot j})(X_{ij} - X_{\cdot j})^T] \\
&= -\frac{N}{2}\widehat{Q}^{-1}\widehat{Q} = -\frac{1}{2}Np,
\end{aligned}$$

ибо $\operatorname{tr} I$ здесь равен p. Поэтому

$$p(X, \widehat{\vartheta}) = \left(\frac{1}{\sqrt{2\pi}}\right)^{Np}\left(\frac{1}{\sqrt{\det \widehat{Q}}}\right)^{N} \exp\Big\{-\frac{Np}{2}\Big\}.$$

(b) Вычисление $\widehat{\vartheta}_H$. При гипотезе H все X_{ij} распределены одинаково, так что мы имеем дело с одной выборкой из $N(a, Q)$, объем этой выборки равен $N = \sum_{j=1}^{k} n_j$. Ответ, т. е. оценки наибольшего правдоподобия для a и Q, уже известен из главы 2 (см. 6):

$$\widehat{a}_H = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij} \qquad (\text{далее } \overline{X}),$$

$$\widehat{Q}_H = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X})(X_{ij} - \overline{X})^T.$$

Для вычисления статистики λ остаётся вычислить $p(X, \widehat{\vartheta}_H)$. Как и при вычислении $p(X, \widehat{\vartheta})$, найдём, что показатель экспоненты равен $-Np/2$. Поэтому

$$p(X, \widehat{\vartheta}_H) = \left(\frac{1}{\sqrt{2\pi}}\right)^{Np} \left(\frac{1}{\sqrt{\det \widehat{Q}_H}}\right)^{N} \exp\left\{-\frac{Np}{2}\right\}.$$

Сопоставляя результаты из (a) и (b), найдем, что критериальная статистика

$$\lambda = \left(\frac{\det \widehat{Q}_H}{\det \widehat{Q}}\right)^{N/2} \qquad \text{эквивалентна} \qquad \widetilde{\lambda} = \frac{\det \widehat{Q}_H}{\det \widehat{Q}}.$$

Гипотезу H следует отвергать при больших наблюдённых значениях $\widetilde{\lambda}$. Впрочем, об определении критических значений, т. е. о распределении $\widetilde{\lambda}$ при гипотезе, еще придется говорить особо.

Толкование. Матрицы $\widehat{Q}$ и $\widehat{Q}_H$ измеряют разброс (дисперсию) исходных наблюдений. $\widehat{Q}_H$ – это разброс относительно общего (единого) центра $\overline{X}$; $\widehat{Q}$ – это суммарный разброс k выборок относительно их центров. Эти центры $X_{\cdot 1}, \ldots, X_{\cdot k}$. Ясно, что $\widehat{Q}_H \geqslant \widehat{Q}$, что, впрочем, ниже будет доказано. Чтобы увидеть, насколько значительно (статистики говорят «значимо») уменьшается разброс при переходе от одного центра к нескольким, надо сопоставить какие-либо скалярные характеристики матриц $\widehat{Q}_H$ и $\widehat{Q}$. В данном случае это их определители. Значимое уменьшение разброса при переходе от одного центра к нескольким противоречит предположению о том, что центр един, т.е. гипотезе о равенстве средних.

Ради дальнейшего упрощения и прояснения преобразуем $\widehat{Q}_H$:

$$N\widehat{Q}_H = \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(X_{ij} - X_{\cdot j}) + (X_{\cdot j} - \overline{X})][(X_{ij} - X_{\cdot j}) + (X_{\cdot j} - \overline{X})]^T$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(X_{ij} - X_{\cdot j})(X_{ij} - X_{\cdot j})^T$$

$$+ (X_{\cdot j} - \overline{X})(X_{\cdot j} - \overline{X})^T (X_{ij} - X_{\cdot j})(X_{\cdot j} - \overline{X})^T$$

$$+ (X_{\cdot j} - \overline{X})(X_{ij} - X_{\cdot j})^T] = (1) + (2) + (3) + (4).$$

Суммы (3) и (4), отличающиеся одна от другой транспонированием, равны нулю. Например,

$$(3) = \sum_{j=1}^{k}\left[\sum_{i=1}^{n_j}(X_{ij} - X_{\cdot j})\right](X_{\cdot j} - \overline{X}) = 0,$$

ибо $\sum\limits_{i=1}^{n_j}(X_{ij} - X_{\cdot j}) = 0$. Суммы (1) и (2) нам уже знакомы по § 1 этой главы:

$$(1) = S_1, \qquad (2) = S_2.$$

Поэтому

$$N\widehat{Q}_H = (S_1 + S_2).$$

Отсюда

$$\widetilde{\lambda} = \frac{\det \widehat{Q}_H}{\det \widehat{Q}} = \frac{\det(S_1 + S_2)}{\det S_1}.$$

Как уже говорилось, для выбора критических значений надо знать распределение статистики $\widetilde{\lambda}$ при гипотезе $H : a_1 = a_2 = \ldots = a_k$. Более простым оказывается распределение обратной величины. По этой причине на практике используют статистику

$$\Lambda \underset{df}{=} \frac{\det S_1}{\det(S_1 + S_2)}.$$

Гипотезу H отвергают при *малых* значениях Λ, полученных в опыте. Статистика Λ была предложена Уилксом (S. S. Wilks, 1932).

Легко видеть, что статистика Λ равна произведению корней уравнения

$$\det(S_1 - \beta(S_1 + S_2)) = 0.$$

Полученные результаты сформулируем в виде теоремы.

Теорема 4.4. *Пусть $X_{1j}, X_{2j}, \ldots, X_{n_j j}$ – выборка из $N_p(a_j, Q)$, $j = \overline{1,k}$. Все k выборок взаимно независимы. Матрица $Q > 0$ и неизвестна.*

Положим

$$X_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}, \quad \overline{X} = \frac{1}{n_1 + \ldots + n_k}\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij},$$

$$S_1 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - X{\cdot}j)(X_{ij} - X{\cdot}j)^T,$$

$$S_2 = \sum_{j=1}^{k} n_j(X_{\cdot j} - \overline{X})(X_{\cdot j} - \overline{X})^T.$$

Применение метода отношения правдоподобий к проверке гипотезы

$$H : a_1 = a_2 = \ldots = a_k$$

приводит к критической статистике

$$\Lambda = \det S_1^{-1}(S_1 + S_2) = \prod_i \beta_i,$$

где $\beta_1, \beta_2, \ldots$ — корни характеристического уравнения

$$\det[S_1 - \beta(S_1 + S_2)] = 0.$$

Гипотезу H следует отвергать при малых значениях статистики Λ.

Корни обсуждаемого характеристического уравнения, которое можно преобразовать в уравнение

$$\det\left(S_2 - \frac{1-\beta}{\beta}S_1\right) = 0,$$

взаимнооднозначно связаны с корнями уже известного нам из § 1 уравнения

$$\det(S_2 - \lambda S_1) = 0.$$

Получим, что если λ – корень уравнения $\det(S_2 - \lambda S_1) = 0$, то $\beta = \frac{1}{\lambda+1}$ – это корень уравнения $\det[S_1 - \beta(S_1 + S_2)] = 0$. Ранее, см. § 1, для корней уравнения $\det(S_2 - \lambda S_1) = 0$ мы уже ввели обозначения $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant \geqslant 0$. Отсюда

$$\Lambda = \prod_i \frac{1}{1 + \lambda_i} = \prod_i \beta_i.$$

§ 3. Критические статистики

Помимо статистик Λ и V для проверки H можно применять и другие критерии. Употребительна, в частности, статистика T^2 Лоули и Хотеллинга (D. N. Lawley, 1938; H. Hotelling, 1951)

$$T^2 = \sum_i \lambda_i, \quad \text{или} \quad T^2 = \operatorname{tr}(S_2 S_1^{-1}).$$

В качестве статистических критериев можно использовать и другие функции от корней характеристических уравнений. Функция корней может служить статистическим критерием для проверки H, если она монотонно возрастает (либо монотонно убывает) при изменении этих корней. Статистики Λ и T^2 к тому же строго монотонны.

Критерии, основанные на статистиках V, Λ, T^2 и др., неэквивалентны друг другу, вообще говоря. Их функции мощности не находятся в отношении «больше-меньше». К сожалению, наилучшего статистического критерия для проверки H не существует даже в локальном смысле.

Стохастические свойства статистик S_1, S_2 будут выявлены при изучении общих линейных моделей в гл. 6. Сейчас мы без доказательства сообщим, что статистики S_1 и S_2

1. стохастически независимы;
2. распределены по Уишарту

   $$S_1 = W_p(N-k, Q), \;\; S_2 = W_p(k-1, Q) \quad \text{при гипотезе } H;$$

   при альтернативе уишартовское распределение статистики S_2 становится нецентральным.

Теорема 4.5. *Корни характеристических уравнений инвариантны относительно невырожденных аффинных преобразований исходных наблюдений.*

Доказательство. Аффинное преобразование – это замена X_{ij} на $X'_{ij} = b + BX_{ij}$ для $j = \overline{1,k}, i = \overline{1,n_j}$. Здесь b – вектор, $b \in \mathbb{R}^p$, B – невырожденная $(p \times p)$-матрица. Далее предполагаем, что b, B не зависят от $\{X_{ij} \mid j = \overline{1,k}, i = \overline{1,n_j}\}$. Скажем, b и B постоянны.

Пусть S'_1 и S'_2 — матрицы S_1 и S_2, вычисленные по наблюдениям X'_{ij}. Очевидно, что

$$S'_1 = BS_1B^T, S'_2 = BS_2B^T.$$

Отсюда следует, что уравнение

$$\det(S_2' - \lambda S_1') = 0$$

эквивалентно уравнению $\det(S_2 - \lambda S_1) = 0$. То же верно и для уравнения $\det[S_1 - \beta(S_1 + S_2)] = 0$ и его варианта, относящегося к матрицам S_1', S_2'. $\square$

Следствие 4.2. *При гипотезе $H : a_1 = a_2 = \ldots = a_k$ можно считать, что исходные p-мерные наблюдения распределены по стандартному нормальному закону $N_p(0, I)$.*

С учётом сказанного выше о свойствах S_1 и S_2 из этого следствия вытекает, что при гипотезе H корни характеристического уравнения

$$\det(S_2 - \lambda S_1) = 0$$

распределены так же, как корни уравнения

$$\det[W_p(k-1, I) - \lambda W_p(N-k, I)] = 0,$$

где уишартовские матрицы $W_p(k-1, I)$ и $W_p(N-k, I)$ стохастически независимы.

Аналогичное утверждение верно и для корней характеристического уравнения

$$\det[S_1 - \beta(S_1 + S_2)] = 0.$$

§ 4. Упражнения

1. Рассмотрим характеристическое уравнение

$$\det[W_p(m, Q) - \lambda W_p(n, Q)] = 0$$

для $m = 1, n \geqslant p$. В этом случае отличный от нуля корень λ_1 всего один. Найдите его распределение.
Указание: $\lambda_1 = \mathrm{tr}[YY^T W_p^{-1}(n, I)] = Y^T W_p^{-1}(n, I) Y$. Распределение последней случайной величины дано в теореме 3.4 главы 3.

2. Рассмотрим m независимых выборок объёмов $n_1, \ldots, n_m$ из распределений $N_p(a_1, Q)$, $\ldots, N_p(a_m, Q)$ соответственно. Параметры $a_1, \ldots, a_m, Q$ неизвестны.
Предложите статистический критерий для проверки гипотезы

$$H : a_k = c a_{k-1}, k = 2, \ldots, m,$$

считая число c известным,

(a) методом Роя;

(b) методом отношения правдоподобий.

3. В условия предыдущей задачи и считая, что гипотеза H верна, укажите оценки для a_1, Q

   (a) методом Роя;

   (b) методом отношения правдоподобий.

   Каковы статистические свойства этих оценок?

Глава 5
МОДУЛИ НАД КОЛЬЦАМИ МАТРИЦ

Линейный статистический анализ и метод наименьших квадратов как его часть получили свою современную завершенную форму на языке линейной алгебры, т.е. на языке геометрии. *Многомерный* линейный статистический анализ на геометрическом языке может быть изложен столь же красиво и ясно. Для этого стандартные средства линейной алгебры должны быть расширены.

Многомерное наблюдение – это наблюдаемые в одном случайном эксперименте несколько случайных величин. Многомерные наблюдения мы записываем в виде столбцов. Многомерные наблюдения (как и одномерные) мы обычно снабжаем индексами. В простейшем случае индексами служат натуральные числа. Для независимых одинаково распределенных наблюдений это подходящий способ организации данных. Если распределения наблюдений зависят от одного или нескольких факторов, индексами могут служить значения или комбинации значений этих факторов. Впрочем, обычно уровни факторов тоже снабжают номерами. В этом случае индекс – это набор чисел. Так, в двухфакторной схеме (классификация по двум признакам) индексами служат пары натуральных чисел.

Совокупность наблюдений, снабженных индексами, таким способом организованную, мы будем называть *таблицей*.

Для теоретического анализа наиболее удобна линейная нумерация данных. Далее мы будем придерживаться этой системы. При рассмотрении примеров мы будем возвращаться, если потребуется, к естественной индексации данных.

В одномерном статистическом анализе нумерация данных предлагает их запись в виде строки. В многомерном случае совокупность нумерованных данных, т.е. таблицу, мы также будем воспринимать как строку, позиции в которой заняты столбцами. Во многих случаях (но не всегда!) такую таблицу можно воспринимать как матрицу.

Таблицы одной формы естественно образуют векторное пространство относительно операций сложения и умножения на числа. Для целей статистического анализа это векторное пространство наделяют скалярным произведением. В одномерном анализе если наблюдения независимы и имеют одинаковые дисперсии, то наиболее подходящим скалярным произведением является евклидово произведение. Более подробно: пусть наблюдения снабжены индексом α; пусть таблицы $\mathbf{T}_X$ и $\mathbf{T}_Y$ состоят из одномерных элементов X_α, Y_α соответственно. Тогда евклидово скалярное произведение таблиц $\mathbf{T}_X$ и $\mathbf{T}_Y$ есть

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \sum_\alpha X_\alpha Y_\alpha,$$

где индекс суммирования пробегает все значения.

Многомерные наблюдения мы записываем в виде столбцов. В многомерном случае элементы X_α, Y_α являются столбцами. Для таблиц, элементами которых являются столбцы, примем такое определение скалярного произведения таблиц $\mathbf{T}_X$ и $\mathbf{T}_Y$:

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \sum_\alpha X_\alpha Y_\alpha^T.$$

Скалярное произведение в нашем случае – это квадратная матрица. Следовательно, для таблиц, элементами которых являются столбцы, квадратные матрицы соответствующей размерности должны играть роль скаляров. С помощью введённого скалярного произведения будет построена теория многомерного статистического анализа, в основных чертах параллельная сложившейся одномерной теории.

§ 1. Пространство таблиц

Во введении мы условились ради упрощения и единообразия теоретического анализа придерживаться линейного порядка индексации (т.е. нумерации) наблюдений. Однако все вводимые нами понятия и связывающие их теоремы с очевидными изменениями остаются верными и для таблиц с иной индексацией.

Рассмотрим таблицу с n элементами

$$\boldsymbol{T} \underset{df}{=} \{X_i \mid i = \overline{1,n}\},$$

где $X_1, X_2, \ldots, X_n$ — p-мерные векторы-столбцы. Совокупность таких таблиц обозначим T_n^p Таблицы одной формы естественно образуют линейное пространство относительно операций сложения и умножения на числа.

Сложение:

$$\{X_i \mid i = \overline{1,n}\} + \{Y_i \mid i = \overline{1,n}\} = \{X_i + Y_i \mid i = \overline{1,n}\}.$$

Умножение на числа: пусть λ – число; тогда

$$\lambda\{X_i \mid i = \overline{1,n}\} = \{\lambda X_i \mid i = \overline{1,n}\}.$$

Помимо умножения на числа, мы будем рассматривать поэлементное умножение таблиц на квадратные матрицы соответствующих размеров.

Умножение на матрицы слева: пусть K – квадратная матрица размерности $p \times p$. Положим

$$K\{X_i \mid i = \overline{1,n}\} = \{KX_i \mid i = \overline{1,n}\}.$$

Отметим, что умножение таблицы на число можно рассматривать как частный случай умножения на квадратную матрицу, а именно: умножение на число λ есть умножение на матрицу λI, где I – единичная матрица размера $p \times p$.

Умножение на матрицы справа: пусть $Q = \| q_{ij} \mid i = \overline{1,n}, j = \overline{1,n} \|$ – квадратная $(n \times n)$-матрица. Определим умножение таблицы $\boldsymbol{T} = \{X_i \mid i = \overline{1,n}\}$ на матрицу Q справа, положив

$$\{X_i \mid i = \overline{1,n}\}Q = \Big\{\sum_{j=1}^{n} X_j q_{ji} \mid i = \overline{1,n}\Big\}.$$

Видно, что произведение $\mathbf{T}Q$ определено обычным способом по правилу «строка на столбец» с той лишь разницей, что элементами строки $\mathbf{T}$ (таблицы $\mathbf{T}$) служат не числа, а столбцы $X_1, \ldots, X_n$.

Определим в пространстве таблиц **внутреннее произведение.** За его свойства и последующее употребление будем называть его *скалярным произведением.* Более подробно: пусть

$$\mathbf{T} = \{X_i \mid i = \overline{1,n}\}, \quad \mathbf{R} = \{Y_i \mid i = \overline{1,n}\}.$$

Скалярным произведением таблиц $\mathbf{T}$ и $\mathbf{R}$ назовем

$$\langle \mathbf{T},\ \mathbf{R} \rangle = \sum_{i=1}^{n} X_i Y_i^T.$$

Результатом произведения является квадратная $(p \times p)$-матрица. Скалярное произведение некоммутативно:

$$\langle \mathbf{R},\ \mathbf{T} \rangle = \langle \mathbf{T},\ \mathbf{R} \rangle^T.$$

Скалярный квадрат таблицы $\mathbf{T}$ есть симметричная неотрицательно определенная $(p \times p)$-матрица:

$$\langle \mathbf{T},\ \mathbf{T} \rangle = \sum_{i=1}^{n} X_i X_i^T.$$

Для обозначения скалярного квадрата будем употреблять привычный символ абсолютного значения $\langle \mathbf{T},\ \mathbf{T} \rangle = |\mathbf{T}|^2$. В нашем случае $|\mathbf{T}|$ – это так называемый матричный модуль.

Свойства скалярного произведения в пространствах таблиц напоминают свойства традиционного скалярного произведения в евклидовых векторных пространствах: если $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ – это таблицы общей формы, то

$\langle \mathbf{T}_1 + \mathbf{T}_2,\ \mathbf{T}_3 \rangle = \langle \mathbf{T}_1,\ \mathbf{T}_3 \rangle + \langle \mathbf{T}_2,\ \mathbf{T}_3 \rangle;$

$\langle K\mathbf{T}_1,\ \mathbf{T}_2 \rangle = K\langle \mathbf{T}_1,\ \mathbf{T}_2 \rangle,$ если K – квадратная $(p \times p)$-матрица;

$\langle \mathbf{T}_1,\ \mathbf{T}_1 \rangle \geqslant 0$ в смысле сравнения квадратных симметричных матриц;

$\langle \mathbf{T}_1,\ \mathbf{T}_1 \rangle = 0$ тогда и только тогда, когда $\mathbf{T}_1 = 0$.

Мы скажем, что таблица $\mathbf{T}$ *ортогональна* таблице $\mathbf{R}$, если

$$\langle \mathbf{T},\ \mathbf{R} \rangle = 0.$$

Заметим, что если $\langle \mathbf{T},\ \mathbf{R} \rangle = 0$, то и $\langle \mathbf{R},\ \mathbf{T} \rangle = 0$. Поэтому свойство ортогональности таблиц является взаимным. Ортогональность таблиц $\mathbf{T}$ и $\mathbf{R}$ будем обозначать $\mathbf{T} \perp \mathbf{R}$.

Отметим форму теоремы Пифагора: если таблицы $\mathbf{T}$ и $\mathbf{R}$ ортогональны, то

$$\langle \mathbf{T} + \mathbf{R},\ \mathbf{T} + \mathbf{R} \rangle = \langle \mathbf{T},\ \mathbf{T} \rangle + \langle \mathbf{R},\ \mathbf{R} \rangle.$$

Еще раз отметим, что результатом скалярного произведения двух таблиц является $(p \times p)$-матрица. Поэтому в пространствах таблиц квадратные матрицы соответствующих размеров должны играть роль скаляров. В частности, умножение на $(p \times p)$-матрицу слева мы будем рассматривать как умножение на скаляр и таблицу $K\mathbf{T}$ будем считать пропорциональной таблице $\mathbf{T}$.

С алгебраической точки зрения пространство таблиц с указанными выше операциями – это модуль над кольцом квадратных матриц соответствующего размера (в данном случае над кольцом $(p \times p)$-матриц). В этом модуле определена бинарная операция со значениями в упомянутом кольце. Эту операцию из-за ее свойств мы назвали скалярным произведением, ибо элементы кольца для модуля таблиц играют роль скаляров.

Таблицы и матрицы. Вместе с таблицами из $\mathbb{T}_n^p$ вида $\boldsymbol{T}_X = \{X_i \mid i = \overline{1, n}\}$ будем рассматривать взаимнооднозначно соответствующие им матрицы

$$\mathcal{X} = \| \, X_1, X_2, \ldots, X_n \, \|.$$

Матрица $\mathcal{X}$ – это матрица с p строками и n столбцами.

Многие операции с таблицами можно проводить в их матричных формах. В частности, сложению таблиц соответствует сложение их матриц; левому умножению на квадратную $(p \times p)$-матрицу K соответствует матричное произведение $K\mathcal{X}$; умножению (справа) на матрицу Q соответствует матричное произведение $\mathcal{X}Q$; скалярное произведение таблиц $\boldsymbol{T}_X$ и

$$\mathbf{T}_Y = \{Y_i \mid i = \overline{1, n}\}$$

равно произведению соответствующих им матриц $\mathcal{X}$ и $\mathcal{Y}$:

$$\langle \mathbf{T}_X, \; \mathbf{T}_Y \rangle = \mathcal{X}\mathcal{Y}^T.$$

Все эти соотношения проверяются прямыми выкладками.

Линейные преобразования. Многие понятия классической линейной алгебры почти автоматически переносятся на пространства таблиц с естественным расширением запаса скаляров до кольца квадратных матриц.

В частности, преобразование $f(\cdot)$ пространства таблиц $\mathbb{T}_n^p$ в себя естественно назвать *линейным*, если для любых таблиц $\mathbf{T}_1$ и $\mathbf{T}_2$ и для любых $(p \times p)$-матриц K_1 и K_2

$$f(K_1\mathbf{T}_1 + K_2\mathbf{T}_2) = K_1 f(\mathbf{T}_1) + K_2 f(\mathbf{T}_2).$$

Линейные преобразования в пространствах таблиц осуществляются умножением справа на квадратные матрицы. Пусть $Q = \| \, q_{ij} \, \|$ – произвольная $n \times n$-матрица, $\mathbf{T}$ – произвольная таблица из $\mathbb{T}_n^p$. То, что преобразования

$$f(\mathbf{T}) = \mathbf{T}Q$$

являются линейными в указанном выше смысле, непосредственно следует из определения операции умножения таблицы на матрицу справа. То, что других линейных преобразований нет, следует из того, что их нет даже в случае $p = 1$. (Как известно, все линейные преобразования в векторных пространствах строк осуществляются умножением справа на квадратные $(n \times n)$-матрицы.)

Отметим, что матричная форма представления таблиц подходит и для описания линейных преобразований: матрица $\mathcal{X}Q$ (произведение матриц $\mathcal{X}$ и Q) совпадает с матричной формой таблицы

$$\mathbf{T}_X Q = \{X_i \mid i = \overline{1,n}\}Q = \Big\{\sum_{j=1}^{n} X_j q_{ji} \mid i = \overline{1,n}\Big\}.$$

Линейное преобразование (пространства таблиц в себя) мы будем называть *ортогональным*, если это преобразование сохраняет скалярное произведение, а именно если для таблиц $\mathbf{T}$ и $\mathbf{R}$

$$\langle \mathbf{T}Q,\ \mathbf{R}Q\rangle = \langle \mathbf{T},\ \mathbf{R}\rangle.$$

Легко видеть, что ортогональное преобразование осуществляется умножением справа на ортогональную матрицу. Действительно,

$$\langle \mathbf{T}Q,\ \mathbf{R}Q\rangle = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} X_j q_{ji}\right)\left(\sum_{l=1}^{n} Y_l q_{li}\right)^T = \sum_{j=1}^{n}\sum_{q=1}^{n} X_j Y_l^T \left(\sum_{i=1}^{n} q_{ji} q_{li}\right) = \sum_{j=1}^{n} X_j Y_j^T,$$

поскольку матрица Q ортогональна, и потому

$$\sum_{i=1}^{n} q_{ji} q_{li} = \delta_{jl} \quad \text{(символ Кронекера)}.$$

Обозначения. Матрицы с числовыми элементами, у которых p строк и n столбцов, будем называть $(p \times n)$-матрицами. Совокупность $(p \times n)$-матриц обозначаем $\mathbb{R}_n^p$. Матрицы размера $p \times 1$ будем называть p-столбцами, или просто столбцами. Совокупность p-столбцов обозначим $\mathbb{R}_1^p = \mathbb{R}^p$. Матрицы $1 \times n$ называем n-строками, или просто строками. Совокупность n-строк обозначаем $\mathbb{R}_n^1 = \mathbb{R}_n$.

§ 2. Таблицы со случайными элементами

Матрицы почти всегда могут замещать таблицы. Но в некоторых случаях таблицы необходимы.

Рассмотрим таблицу, элементы которой представляют собой p-мерные случайные величины, записанные в виде p-столбцов:

$$\mathbf{T} = \{X_i \mid i = \overline{1, n}\}.$$

Для таких таблиц (со случайными элементами) определим математические ожидания и ковариации, т.е. первые и вторые моменты. *Математическим ожиданием* $\mathbf{T}$ назовем таблицу

$$\mathrm{E}\,\mathbf{T} = \{\mathrm{E}\,X_i \mid i = \overline{1, n}\}. \tag{5.1}$$

Матрицу ковариаций для таблицы $\boldsymbol{T}$ определим, следуя образцу, который даёт матрица ковариаций случайного вектора. Пусть

$$t = (x_1, \ldots, x_n)$$

– n-строка, составленная из случайных величин $x_1, x_2, \ldots, x_n$. Как известно, матрицей ковариаций случайного вектора t называют $(n \times n)$-матрицу с элементами

$$\sigma_{ij} = \mathrm{Cov}(x_i, x_j), \quad \text{где } i, j = \overline{1, n}.$$

Алгебраически, с помощью матричных операций, матрицу ковариаций вектора t можно определить как

$$\mathrm{Var}\, t = \mathrm{E}(t - \mathrm{E}\, t)^T (t - \mathrm{E}\, t) = \{\mathrm{Cov}(X_i, X_j) \mid i, j = \overline{1, n}\}.$$

Следуя этому определению, *таблицей ковариаций* случайной таблицы $\boldsymbol{T}$ мы назовем таблицу

$$\mathrm{Var}\,\mathbf{T} \underset{df}{=} \mathrm{E}(\mathbf{T} - \mathrm{E}\,\mathbf{T})^T (\mathbf{T} - \mathrm{E}\,\mathbf{T}) = \{\mathrm{Cov}(X_i, X_j) \mid i, j = \overline{1, n}\}. \tag{5.2}$$

Здесь $\mathrm{Cov}(X_i, X_j)$ – матрица ковариаций случайных векторов-столбцов X_i и X_j:

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i - \mathrm{E}\,X_i)(X_j - \mathrm{E}\,X_j)^T. \tag{5.3}$$

Отметим, что таблицу ковариаций $\mathrm{Var}\,\mathbf{T}$ мы воспринимаем как квадратную таблицу размера $n \times n$, элементы которой – это $(p \times p)$-матрицы $\mathrm{Cov}(X_i, X_j)$. Именно такого взгляда на $\mathrm{Var}\,\mathbf{T}$ мы будем придерживаться, когда будем обсуждать преобразования таблиц ковариаций при линейных преобразованиях случайных таблиц.

Рассмотрим случайную таблицу $\mathbf{R}$, полученную линейным преобразованием таблицы $\mathbf{T}$:

$$\mathbf{R} = \mathbf{T}Q,$$

где Q – некоторая $(n \times n)$-матрица. Очевидно, что для математических ожиданий таблиц действует обычное правило линейности:

$$\mathrm{E}\,\mathbf{R} = (\mathrm{E}\,\mathbf{T})Q.$$

Для $\mathrm{Var}\,\mathbf{R}$, выполняя перемножения таблиц (и матриц) по правилу «строка на столбец», получим

$$\mathrm{Var}\,\mathbf{R} = \mathrm{E}[((\mathbf{T} - \mathrm{E}\,\mathbf{T})Q)^T((\mathbf{T} - \mathrm{E}\,\mathbf{T})Q)] = Q^T(\mathrm{Var}\,\mathbf{T})Q.$$

Для математической статистики особый интерес представляют таблицы со статистически независимыми случайными элементами, когда матрицы ковариаций этих элементов одинаковы. Пусть $\mathbf{T}$ – такая таблица, так что

$$\mathrm{Cov}(X_i, X_j) = \delta_{ij}Q \quad \text{для} \quad i, j = \overline{1, n}.$$

Здесь Q – некоторая неотрицательно определенная $(p \times p)$-матрица, δ_{ij} – символ Кронекера. Рассмотрим ортогональное преобразование таблицы $\mathbf{T}$:

$$\mathbf{R} = \mathbf{T}C,$$

где C – некоторая ортогональная $(n \times n)$-матрица. Справедлива простая, но важная лемма.

Лемма 6.

$$\mathrm{Var}\,\mathbf{R} = \mathrm{Var}\mathbf{T} = \{\delta_{ij}Q \mid i, j = \overline{1, n}\}.$$

Смысл леммы заключается в том, что при ортогональных преобразованиях таблиц с некоррелированными элементами с одинаковыми ковариациями появляются случайные таблицы, элементы которых тоже попарно не коррелированы, ковариации которых одинаковы и совпадают с ковариацией элементов исходной таблицы.

Доказательство. Доказательство леммы – простая выкладка. Чтобы упростить формулы, предположим, что $\mathrm{E}\,\mathbf{T} = 0$. Тогда, выполняя перемножения таблиц (и матриц) по правилу «строка на столбец», получим

$$\begin{aligned}\mathrm{Var}(\mathbf{T}C) &= \mathrm{E}[(\mathbf{T}C)^T(\mathbf{T}C)] \\ &= C^T(\mathrm{Var}\,\mathbf{T})C = C^T\{\delta_{ij}Q \mid i, j = \overline{1, n}\}C \\ &= \{\delta_{ij}Q \mid i, j = \overline{1, n}\}.\end{aligned}$$

$\square$

К сожалению, у нас в запасе нет алгебраической операции, которая при умножении единичной $(n \times n)$-матрицы, *воспринимаемой как таблица*, на матрицу Q давала бы сразу таблицу $\{\delta_{ij}Q \mid i,j=\overline{1,n}\}$. С использованием такой операции выкладка была бы особенно выразительна: была бы полная аналогия в формулировке и доказательстве с аналогичным утверждением о случайных векторах.

§ 3. Базисы и координаты

Порождающие базисы. Особую роль в пространстве таблиц $\mathbb{T}^p_n$, когда они представлены в виде $(p\times n)$-матриц, играют $(p\times n)$-матрицы вида αx, где α – это p-столбец, x – это n-строка, символ αx обозначает произведение матриц $\alpha\in\mathbb{R}^p_1, x\in\mathbb{R}^1_n$.

Пусть n-строки $e_1,e_2,\ldots,e_n\in\mathbb{R}^1_n$ образуют базис пространства $\mathbb{R}^1_n$. Пусть $\alpha_1,\alpha_2,\ldots,\alpha_n\in\mathbb{R}^p_1$ – произвольные p-столбцы. Рассмотрим $(p\times n)$-матрицы $\alpha_1e_1,\alpha_2e_2,\ldots,\alpha_ne_n$.

Теорема 5.1. *Всякую $(p\times n)$-матрицу $\mathcal{X}$ можно представить как*

$$\mathcal{X}=\sum_{i=1}^n\alpha_ie_i$$

при некотором выборе $\alpha_1,\alpha_2,\ldots,\alpha_n\in\mathbb{R}^p_1$, и притом единственным образом.

Доказательство. Определим матрицу E размера $n\times n$, образованную n-строками $e_1,e_2,\ \ldots,e_n$. Введем также переменную матрицу $\mathcal{A}$ размера $p\times n$, образованную p-столбцами $\alpha_1,\alpha_2,\ldots,\alpha_n$. С помощью матриц $\mathcal{A}$ и E матрица $\mathcal{X}$ может быть записана как

$$\mathcal{X}=\sum_{i=1}^n\alpha_ie_i=\mathcal{A}E.$$

Вот вычисления, подтверждающие это утверждение. Пусть $\alpha_i= =(\alpha_{1i},\alpha_{2i},\ldots,\alpha_{pi})^T,\ e_i=(e_{i1},e_{i2},\ldots,e_{in})$,

$$\sum_{i=1}^n\alpha_ie_i=\sum_{i=1}^n(\alpha_{1i},\alpha_{2i},\ldots,\alpha_{pi})^T(e_{i1},e_{i2},\ldots,e_{in}).$$

Элементы каждого из n произведений α_ie_i, стоящие на месте (k,l), суть $\alpha_{ki}e_{il}$. Их общая сумма, т. е. элемент матрицы $\sum_{i=1}^n\alpha_ie_i$, равна $\sum_{i=1}^n\alpha_{ki}e_{il}$.

Элемент матрицы $\mathcal{A}E$, который стоит на месте (k, l) (вычисляемый по правилу «строка на столбец»), равен

$$\sum_{i=1}^{n} \alpha_{ki} e_{il}.$$

Результаты вычислений совпали.

Теорема будет доказана, если мы покажем, что уравнение

$$\mathcal{X} = \mathcal{A}E$$

имеет решение относительно $(p \times n)$-матрицы $\mathcal{A}$, и притом единственное. Поскольку $(n \times n)$-матрица E не вырождена, это решение очевидно:

$$\mathcal{A} = \mathcal{X}E^{-1}.$$

□

Теорема 5.1 позволяет говорить, что базис в $\mathbb{R}^1_n$ порождает (указанным выше способом) все пространство $\mathbb{R}^p_n$. Поэтому базисы в $\mathbb{R}^1_n$ мы будем применительно к $\mathbb{R}^p_n$ называть *порождающими базисами*. При этом p-столбцы $\alpha_1, \alpha_2, \ldots, \alpha_n$ из разложения $(p \times n)$-матрицы $\mathcal{X}$ по порождающему базису $e_1, e_2, \ldots, e_n$ можно рассматривать как координаты $\mathcal{X}$ в этом базисе. Для канонического базиса пространства $\mathbb{R}^1_n$ (когда e_i – это n-строка, в которой единица стоит на месте i, а прочие элементы равны нулю) координаты $\mathcal{X}$ относительно этого базиса совпадают с p-столбцами $X_1, \ldots, X_n \in \mathbb{R}^p_1$, образующими матрицу $\mathcal{X}$.

Координаты в разных базисах. Координаты $(p \times n)$-матрицы $\mathcal{X}$ относительно разных порождающих базисов связаны линейным преобразованием. Пусть, например, n-строки $f_1, \ldots, f_n$ образуют базис $\mathbb{R}^1_n$. Пусть F – $(n \times n)$-матрица, составленная из этих n-строк. Согласно теореме 5.1, существует, и притом единственная, совокупность p-столбцов $\beta_1, \beta_2, \ldots, \beta_n$ – координат $\mathcal{X}$ относительно порождающего базиса $f_1, \ldots, f_n$. Матрицы $\mathcal{B} = \| \beta_1, \ldots, \beta_n \|$ и F связаны с $(p \times n)$-матрицей $\mathcal{X}$ соотношением

$$\mathcal{X} = \mathcal{B}F.$$

Вместе с разложением $\mathcal{X} = \sum_{i=1} \alpha_i e_i = \mathcal{A}E$ по порождающему базису $e_1, \ldots$ $\ldots, e_n$ это дает

$$\mathcal{B}F = \mathcal{A}E.$$

Отсюда

$$\mathcal{B} = \mathcal{A}EF^{-1}, \qquad \mathcal{A} = \mathcal{B}FE^{-1}.$$

Следствие 5.1. *Если порождающие базисы $e_1, \ldots, e_n$ и $f_1, \ldots, f_n$ ортогональные, то переход от координат относительно одного базиса к координатам относительно другого базиса осуществляется умножением на ортогональную матрицу.*

Скалярные произведения в координатной форме. Рассмотрим в $\mathbb{R}^1_n$ произвольный ортогональный базис $e_1, \ldots, e_n$. Для произвольных $(p \times n)$-матриц $\mathcal{X}$ и $\mathcal{Y}$, соответствующих таблицам $\boldsymbol{T}_X$ и $\boldsymbol{T}_Y$, рассмотрим их разложения по этому базису:

$$\mathcal{X} = \sum_{i=1}^{n} \alpha_i e_i, \quad \mathcal{Y} = \sum_{i=1}^{n} \gamma_i e_i.$$

Легко видеть, что

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \mathcal{X}\mathcal{Y}^T = \sum_{i=1}^{n} \alpha_i \gamma_i^T.$$

Данное равенство означает, что в ортогональном базисе скалярное произведение равно попарному произведению координат. Действительно,

$$\mathcal{X}\mathcal{Y}^T = \left\langle \sum_{i=1}^{n} \alpha_i e_i \sum_{j=1}^{n} \gamma_j e_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i e_i e_j^T \gamma_j^T = \sum_{i=1}^{n} \alpha_i \gamma_i^T,$$

так как для ортогонального базиса $e_i e_j^T = \delta_{ij}$. Отсюда скалярный квадрат $\mathbf{T}_X$ равен

$$|\mathbf{T}_X|^2 = \mathcal{X}\mathcal{X}^T = \sum_{i=1}^{n} \alpha_i \alpha_i^T. \tag{5.4}$$

Получили, что квадрат длины (матричный модуль в квадрате) произвольной таблицы равен сумме квадратов ее координат, совсем как квадрат длины вектора.

§ 4. Подмодули и линейные подпространства

Подмодулем в пространстве таблиц $\mathbb{T}^p_n$ или соответствующих им матриц $\mathbb{R}^p_n$ назовем такое множество, которое замкнуто относительно линейных операций — сложения и умножения на скаляры (т.е. умножения слева на $(p \times p)$-матрицы). Для определенности далее будем говорить о таблицах в их матричных формах. Итак:

Определение 1. *Множество $\mathcal{L} \subset \mathbb{R}_n^p$ назовем подмодулем пространства $\mathbb{R}_n^p$, если для любых $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{L}$*

$$K_1\mathcal{X}_1 + K_2\mathcal{X}_2 \in \mathcal{L}$$

при произвольных $(p \times p)$-матрицах K_1, K_2.

Теорема 5.2. *Всякий подмодуль $\mathcal{L}$, $\mathcal{L} \subset \mathbb{R}_n^p$, порождается некоторой линейно независимой системой n-строк. Число элементов в этой системе определяется $\mathcal{L}$ единственным образом. Это число может быть названо размерностью подмодуля $\mathcal{L}$.*

Доказательство. Пусть $\mathcal{X} \in \mathcal{L}$. Множество $(p \times n)$-матриц вида $K\mathcal{X}$, где K – произвольная $(p \times p)$-матрица, образует подмодуль. Обозначим его $\mathcal{L}(\mathcal{X})$. Пусть $x_1, \ldots, x_p$ – это n-строки $(p \times n)$-матрицы $\mathcal{X}$. Выберем среди этих n-строк максимальную линейно независимую подсистему, скажем $y_1, \ldots, y_k$. Очевидно, что

$$\mathcal{L}(\mathcal{X}) = \Big\{\mathcal{Y} \mid \mathcal{Y} = \sum_{i=1}^{k} \beta_i y_i,\ \beta_1, \ldots, \beta_k \in \mathbb{R}_1^p\Big\}.$$

Если $\mathcal{L}(\mathcal{X}) = \mathcal{L}$, то $y_1, \ldots, y_k$ составляют для $\mathcal{L} \subset \mathbb{R}_n^p$ образующий базис. Если $\mathcal{L}(\mathcal{X}) \neq \mathcal{L}$, то находим в $\mathcal{L}$ элемент, скажем, $\mathcal{Z}$, не принадлежащий $\mathcal{L}(\mathcal{X})$. Пополняем систему $y_1, \ldots, y_k$ n-строками $z_1, \ldots, z_p$ $(p \times n)$-матрицы $\mathcal{Z}$, находим в этой совокупности n-строк максимальную линейно независимую подсистему и т.д. На некотором шаге процесс остановится. $\square$

Следствие 5.2. *Всякий подмодуль $\mathcal{L} \subset \mathbb{R}_n^p$ может быть представлен как прямая сумма некоторых одномерных подмодулей $\mathcal{L}_i \subset \mathbb{R}_n^p$:*

$$\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \ldots \oplus \mathcal{L}_l,$$

где

$$\mathcal{L}_i = \{\mathcal{X} \mid \mathcal{X} = \alpha y_i, \alpha \in \mathbb{R}_1^p\}$$

для некоторого $y_i \in \mathbb{R}_n^1$. Число l – одно и то же в любом представлении подмодуля $\mathcal{L}$ в виде прямой суммы одномерных подмодулей. Это число можно называть размерностью подмодуля $\mathcal{L}$: $l = \dim \mathcal{L}$. Линейно независимую систему n-строк из L для подмодуля $\mathcal{L}$ можно назвать его порождающим базисом.

Замечание. Линейно независимую систему n-строк, которая порождает $\mathcal{L}$, можно выбрать ортогональной. Для доказательства достаточно заметить, что порождающую систему можно превратить в ортогональную процессом ортогонализации.

Теорема 5.2 устанавливает взаимнооднозначное соответствие между подмодулями векторного пространства $\mathbb{R}_n^1$ и пространства $(p \times n)$-матриц $\mathbb{R}_n^p$. Выделим этот факт как элемент теории в следующем виде;

Следствие 5.3. *Каждому линейному пространству L в пространстве n-строк $\mathbb{R}_n^1$ соответствует некоторый подмодуль $\mathcal{L}$ в пространстве $(p \times n)$-матриц $\mathbb{R}_n^p$ и обратно. При этом размерности линейных подпространств L и подмодуля $\mathcal{L}$ совпадают.*

Таким образом, пространство $\mathbb{R}_n^p$ (и соответствующее пространство таблиц) и пространство $\mathbb{R}_n^1$ располагают одинаковым «запасом» подмодулей и линейных подпространств. Это замечание имеет важные последствия для многомерного статистического анализа.

Важную для дальнейшего форму связи $\mathcal{L}$ и L выделим в виде теоремы.

Теорема 5.3. *Пусть $\lambda \in \mathbb{R}^p$ пробегает $\mathbb{R}^p$. Тогда*

$$\lambda^T \mathcal{L} = L.$$

Определение 2. *Ортогональным дополнением к подмодулю $\mathcal{L}$ назовем множество*

$$\mathcal{L}^\perp = \{\mathcal{X} \mid \mathcal{X} \in \mathbb{R}_n^p,\ \langle \mathcal{X},\ \mathcal{Y} \rangle = 0,\ \textit{для любых}\ \ \mathcal{Y} \in \mathcal{L}\}.$$

Легко видеть, что $\mathcal{L}^\perp$ является подмодулем и что

$$\mathcal{L} \oplus \mathcal{L}^\perp = \mathbb{R}_n^p, \qquad \dim \mathcal{L}^\perp = n - \dim \mathcal{L}.$$

§ 5. Ортогональные проекции

Рассмотрим пространство таблиц $\mathbb{T}_n^p$ с введённым в нем скалярным произведением $\langle\ \cdot,\ \ \cdot \rangle$. Напомним, что для произвольных таблиц $\boldsymbol{T}$ и $\boldsymbol{R}$ и их матричных форм $\mathcal{X}$ и $\mathcal{Y}$

$$\langle \boldsymbol{T},\ \boldsymbol{R} \rangle = \mathcal{X}\mathcal{Y}^T \in \mathbb{R}_p^p.$$

Пусть $\mathcal{L}$ – некоторый подмодуль. Назовем *проекцией таблицы* $\mathbf{T}$ *на подмодуль* $\mathcal{L}$ ту точку из $\mathcal{L}$, которая является ближайшей к $\mathbf{T}$ в смысле сравнения скалярных квадратов. Скажем подробнее. Пусть таблица $\mathbf{R}$ пробегает множество $\mathcal{L}$. Мы называем точку $\mathbf{R}^0 \in \mathcal{L}$ *ближайшей к* $\mathbf{T}$, если для всякого $\mathbf{R} \in \mathcal{L}$

$$\langle \mathbf{T} - \mathbf{R}^0,\ \mathbf{T} - \mathbf{R}^0 \rangle \leqslant \langle \mathbf{T} - \mathbf{R},\ \mathbf{T} - \mathbf{R} \rangle.$$

Обратим внимание, что $\langle \mathbf{T}-\mathbf{R},\ \mathbf{T}-\mathbf{R}\rangle$ – это функция $\mathbf{R}$ со значениями в множестве матриц. Существование в множестве матриц минимального элемента не очевидно и не обеспечивается автоматически. Поэтому само существование $\operatorname{proj}_{\mathcal{L}} \mathbf{T}$ нуждается в доказательстве. Мы докажем следующую теорему.

Теорема 5.4. *Проекция $\mathbf{T}$ на $\mathcal{L}$ существует, единственна и обладает привычными (по евклидовым пространствам) свойствами.*

1. *Для любой таблицы $\mathbf{R} \in \mathcal{L}$ справедливо*
$$|\mathbf{T}-\mathbf{R}|^2 \geqslant |\mathbf{T} - \mathop{\mathit{proj}}_{\mathcal{L}} \mathbf{T}|^2,$$
причем равенство достигается только при $\mathbf{R} = \mathit{proj}_{\mathcal{L}} \mathbf{T}$;
2. $(\mathbf{T} - \mathit{proj}_{\mathcal{L}} \mathbf{T}) \perp \mathcal{L}$;
3. $\mathit{proj}_{\mathcal{L}}(K_1\mathbf{T}_1 + K_2\mathbf{T}_2) = K_1\,\mathit{proj}_{\mathcal{L}} \mathbf{T}_1 + K_2\,\mathit{proj}_{\mathcal{L}} \mathbf{T}_2$.

Теорема 5.4 — следствие теоремы 5.5 (см. ниже). Доказательства этих теорем мы дадим для матричного представления таблиц.

Теорема 5.5. *Пусть $\mathcal{L} \subset \mathbb{R}^p_n$ – подмодуль в пространстве $(p \times n)$-матриц, пусть $L \subset \mathbb{R}^1_n$ – порождающее $\mathcal{L}$ линейное подпространство в пространстве n-строк. Пусть Π – $(n \times n)$-матрица оператора проектирования на L в пространстве $\mathbb{R}^1_n$: для любого $x \in \mathbb{R}^1_n$*
$$\mathop{\mathit{proj}}_{L} x = x\Pi.$$
Тогда для любого $\mathcal{X} \in \mathbb{R}^p_n$
$$\mathop{\mathit{proj}}_{\mathcal{L}} \mathcal{X} = \mathcal{X}\Pi.$$

Доказательство. Покажем, что для любого $\mathcal{X} \in \mathbb{R}^p_n$
$$\mathcal{X}\Pi \in \mathcal{L}, \qquad \mathcal{X} - \mathcal{X}\Pi \in \mathcal{L}^{\perp}.$$

Пусть $r = \dim \mathcal{L} = \dim L$. Выберем в L ортогональный базис $e_1, \ldots, e_r$; в $L^{\perp}$ выберем ортогональный базис $e_{r+1}, \ldots, e_n$. Тогда $e_1, \ldots, e_n$ – ортогональный базис $\mathbb{R}^1_n$ и порождающий базис $\mathbb{R}^p_n$. Пусть Π – матрица оператора проектирования на L в этом базисе. Тогда $e_i\Pi = e_i$ для $i = \overline{1, r}$, $e_i\Pi = 0$ для $i = \overline{r+1, n}$.

Представим $\mathcal{X}$ в виде $\mathcal{X} = \sum_{i=1}^{n} \alpha_i e_i$, где $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^p$ – некоторые p-столбцы. Теперь
$$\mathcal{X}\Pi = \left(\sum_{i=1}^{n} \alpha_i e_i\right)\Pi = \sum_{i=1}^{r} \alpha_i e_i \in \mathcal{L},$$

$$\mathcal{X} - \mathcal{X}\Pi = \sum_{i=r+1}^{n} \alpha_i e_i \in \mathcal{L}^{\perp}.$$

Остается показать, что

$$\langle \mathcal{X} - \mathcal{Y},\ \mathcal{X} - \mathcal{Y} \rangle \geqslant \langle \mathcal{X} - \mathcal{X}\Pi,\ \mathcal{X} - \mathcal{X}\Pi \rangle$$

для любого $\mathcal{Y} \in \mathcal{L}$, причем равенство достигается только при $\mathcal{Y} = \mathcal{X}\Pi$. Это соотношение между двумя симметричными $(p \times p)$-матрицами означает, что для любого $\lambda \in \mathbb{R}^p$

$$\lambda^T \langle \mathcal{X} - \mathcal{Y},\ \mathcal{X} - \mathcal{Y} \rangle \lambda \ \geqslant \lambda^T \langle \mathcal{X} - \mathcal{X}\Pi,\ \mathcal{X} - \mathcal{X}\Pi \rangle \lambda,$$

или

$$|\lambda^T (\mathcal{X} - \mathcal{Y})|^2 \geqslant |\lambda^T (\mathcal{X} - \mathcal{X}\Pi)|^2,$$

или

$$|\lambda^T \mathcal{X} - \lambda^T \mathcal{Y}|^2 \geqslant |\lambda^T \mathcal{X} - (\lambda^T \mathcal{X})\Pi|^2.$$

Как было отмечено выше, n-строка $y = \lambda^T \mathcal{Y}$ принадлежит L, причем $\lambda^T \mathcal{X}\Pi = x\Pi$ – это проекция $\lambda^T \mathcal{X}$ на L. В силу свойств евклидовых проекций

$$|x - y|^2 \geqslant |x - x\Pi|^2$$

для всякого $y \in L$, причем равенство достигается только при $y = x\Pi$. Таким образом, $\mathcal{X}\Pi$ действительно является ближайшей к $\mathcal{X}$ точкой из $\mathcal{L}$. Теорема 5.5 доказана. $\square$

§ 6. Матричный метод наименьших квадратов

Вычисления проекций на подмодуль $\mathcal{L} \subset \mathbb{R}_n^p$ облегчаются, если известен вид проекций на порождающее $\mathcal{L}$ линейное подпространство L. Согласно теореме 5.3 из § 4, для любого $\lambda \in \mathbb{R}_1^p$

$$\lambda^T \operatorname*{proj}_{\mathcal{L}} \mathcal{X} = \operatorname*{proj}_{L} (\lambda^T \mathcal{X}).$$

Предположим, что для правой части данного тождества мы располагаем явной формулой $y = \operatorname{proj}_L x$. Тогда в силу линейности операций это даст нам для $\operatorname{proj}_{\mathcal{L}} (\lambda^T \mathcal{X})$ явную формулу вида $\lambda^T \mathcal{Y}$:

$$\lambda^T \operatorname*{proj}_{\mathcal{L}} \mathcal{X} = \lambda^T \mathcal{Y}.$$

Так мы получаем явное выражение для $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$. Можно сказать, что это – вычисление $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$ методом Роя (S.N. Roy).

Пример (вычисление среднего арифметического). Пусть $X_1, X_2, \ldots, X_n \in \in \mathbb{R}_1^p$ – совокупность p-столбцов. Рассмотрим таблицу $\mathbf{T} = \{X_i \mid i = \overline{1,n}\}$ и представим ее в матричной форме;

$$\mathcal{X} = \| \, X_1, X_2, \ldots, X_n \, \|.$$

Наша задача – найти таблицу $\mathcal{Y}$ с одинаковыми столбцами, т.е. таблицу вида

$$\mathcal{Y} = \| \, Y, Y, \ldots, Y \, \|, \quad Y \in \mathbb{R}_1^p,$$

ближайшую к $\mathcal{X}$.

Таблицы с одинаковыми столбцами образуют одномерный подмодуль. Обозначим его $\mathcal{L}$, $\mathcal{L} \subset \mathbb{R}_n^p$. Надо, следовательно, найти $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$. Подмодулю $\mathcal{L}$ соответствует одномерное линейное подпространство L, $L \subset \mathbb{R}_n^1$, натянутое на n-строку $e = (1, 1, \ldots, 1)$; L порождает $\mathcal{L}$. Пусть x – произвольная n-строка, $x = (x_1, \ldots, x_n)$. Вид проекции x на L хорошо известен:

$$\operatorname*{proj}_{L} x = (\overline{x}, \ldots, \overline{x}).$$

Применяя метод Роя, от матрицы $\mathcal{X}$ перейдем к n-строке $x = \lambda^T \mathcal{X}$, при этом $x_i = \lambda^T X_i$, $i = \overline{1,n}$. Получим, что

$$\operatorname*{proj}_{L} x = (\lambda^T \overline{X}, \ldots, \lambda^T \overline{X}).$$

Следовательно,

$$\operatorname*{proj}_{\mathcal{L}} \mathcal{X} = (\overline{X}, \ldots, \overline{X}).$$

Разумеется, это не единственный и не всегда наиболее экономный метод. В этом примере, как и в других случаях, можно применить матричный метод наименьших квадратов непосредственно и найти

$$\widehat{Y} = \arg \min_{Y \in \mathbb{R}_1^p} \sum_{i=1}^{n} (X_i - Y)(X_i - Y)^T.$$

Преобразуем целевую функцию: для произвольного $Y \in \mathbb{R}_1^p$

$$\sum_{i=1}^{n} (X_i - Y)(X_i - Y)^T$$

$$= \sum_{i=1}^{n} [(X_i - \overline{X}) + (\overline{X} - Y)][(X_i - \overline{X}) + (\overline{X} - Y)]^T$$

$$= \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T + n\sum_{i=1}^{n}(\overline{X} - Y)(\overline{X} - Y)^T$$

$$= (1) + (2),$$

так как «попарные произведения» обращаются в нуль:

$$\sum_{i=1}^{n}(X_i - \overline{X})(\overline{X} - Y)^T = 0, \qquad \sum_{i=1}^{n}(\overline{X} - Y)(X_i - \overline{X})^T = 0.$$

Теперь целевая функция представляет собой сумму двух неотрицательно определенных матриц, причем первая из них не зависит от изменений Y. Минимальный результат мы получим, положив $Y = \overline{X}$: это обращает в нуль неотрицательно определенную матрицу (2).

Этот ответ, среднее арифметическое, $\widehat{Y} = \overline{X}$, конечно, хорошо известен. Его получают, применяя не матричный, а **обычный метод наименьших квадратов**, т.е. решая экстремальную задачу

$$\widehat{Y} = \arg\min_{Y \in \mathbb{R}_1^p} \sum_{i=1}^{n}(X_i - Y)^T(X_i - Y).$$

Матричный и скалярный методы наименьших квадратов. Так же соотносятся результаты матричного метода наименьших квадратов и традиционного в случае проектирования на иные подмодули $\mathcal{L} \subset \mathbb{R}_n^p$. Причина проста: если таблица $\mathcal{Y}$ — решение матричной задачи

$$\sum_{i=1}^{n}(X_i - Z_i)(X_i - Z_i)^T = (\mathcal{X} - \mathcal{Z})(\mathcal{X} - \mathcal{Z})^T \to \min_{\mathcal{Z} \in \mathcal{L}},$$

то $\mathcal{Y}$ дает решение и скалярной задачи

$$\operatorname{tr}\Big\{\sum_{i=1}^{n}(X_i - Z_i)(X_i - Z_i)^T\Big\} = \sum_{i=1}^{n}(X_i - Z_i)^T(X_i - Z_i) \to \min_{\mathcal{Z} \in \mathcal{L}}.$$

Поэтому, в частности, при вычислении проекций на подмодули можно обходиться традиционным скалярным методом наименьших квадратов. Оба метода наименьших квадратов – и матричный, и скалярный – в линейных моделях дают одинаковые ответы (одинаковые оценки параметров).

Необходимость матричных скалярных произведений и матричные понятия ортогональности, проектирования, подмодулей и прочего обнаруживается при статистической проверке линейных гипотез. Об этом – в следующей главе.

Глава 6

ЛИНЕЙНЫЕ МОДЕЛИ

§ 1. Линейные модели многомерного статистического анализа

Рассмотрим таблицу $\mathbf{T}$ со случайными элементами $X_1 \ldots, X_n \in \mathbb{R}^p$

$$\mathbf{T} = \{X_i \mid i = \overline{1, n}\}.$$

Определение 3. *Говорят, что таблица* $\mathbf{T}$ *со случайными элементами следует линейной модели, если:*

(a) *для некоторого заданного подмодуля* $\mathcal{L}$

$$E\,\mathbf{T} \in \mathcal{L};$$

(b) *элементы* $X_1, \ldots, X_n$ *таблицы* $\mathbf{T}$ *независимы и одинаково распределены.*

Если это общее для всех $X_i, i = \overline{1, n}$, распределение гауссовское, то о таблице $\mathbf{T}$ говорят, что она следует линейной гауссовской модели. Далее мы будем изучать линейные гауссовские модели.

Общую для всех p-столбцов матрицу ковариаций обозначим Q. Таблица $\mathrm{E}\,\mathbf{T}$ и матрица Q являются параметрами модели. Они обычно неизвестны; матрица Q предполагается невырожденной.

Для случайных таблиц, следующих гауссовской модели, часто обсуждают особый тип гипотез, называемых линейными гипотезами. В рамках линейной модели определения 3 линейная гипотеза имеет вид

$$\mathrm{E}\,\mathbf{T} \in \mathcal{L}_1,$$

где $\mathcal{L}_1$ — некоторый заданный подмодуль, причем $\mathcal{L}_1 \subset \mathcal{L}$.

Покажем, что обсуждаемые в многомерном статистическом анализе линейные модели и линейные гипотезы действительно имеют указанную структуру. Основные линейные модели — это факторные и регрессионные. Как примеры мы рассмотрим однофакторную и регрессионную модели многомерного статистического анализа.

Однофакторная модель — простейшая среди факторных моделей. Это задача о нескольких (скажем, m) независимых нормальных выборках, в которой матрицы ковариаций отдельных наблюдений одинаковы. Эта задача была рассмотрена в гл. 4. Таблица данных в этом случае имеет двойную нумерацию: $\mathbf{T} = \{X_{ij} \mid j = \overline{1,m},\ i = \overline{1,n_j}\}$. Здесь m — число различных уровней некоего фактора, оказывающего влияние на ожидаемые значения откликов; n_j — количество независимых повторных наблюдений на уровне фактора $j = \overline{1,m}$. Наконец, многомерные случайные переменные $X_{ij} \in \mathbb{R}_1^p$ — это независимые (при разных значениях индекса) значения p-мерного отклика. Положим $N = n_1 + \ldots + n_m$. Основное предположение модели —

$$X_{ij} \sim N_p(a_j, Q).$$

Наблюдения, составляющие таблицу, расположим линейно и далее будем представлять совокупность данных в виде $(p \times N)$-матрицы:

$$\mathcal{X} = \| X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}, \ldots, X_{m1}, \ldots X_{mn_m} \|.$$

При этом

$$\mathrm{E}\,\mathcal{X} = \| \underbrace{a_1, \ldots, a_1}_{n_1 \text{ раз}}, \underbrace{a_2, \ldots, a_2}_{n_2 \text{ раз}}, \ldots, \underbrace{a_m, \ldots, a_m}_{n_m \text{ раз}} \|.$$

Введем N-строки

$$e_1 = (\underbrace{1, \ldots, 1}_{n_1 \text{ раз}}, 0, \ldots, 0), \quad \ldots, \quad e_m = (\underbrace{0, \ldots, 0}_{n_1 \text{ раз}}, \underbrace{0, \ldots, 0}_{n_2 \text{ раз}}, \ldots, \underbrace{1, \ldots, 1}_{n_m \text{ раз}}).$$

Очевидно, что $\mathrm{E}\,\mathcal{X} = \sum_{i=1}^m a_i e_i$. Тем самым $\mathrm{E}\,\mathcal{X}$ принадлежит m-мерному подмодулю пространства $\mathbb{R}_N^p$, которое порождено N-строками $e_1, \ldots, e_m$.

Гипотеза $H: a_1 = a_2 = \cdots = a_m$, с проверки которой обычно начинают статистический анализ m выборок, очевидно является линейной гипотезой вида $H: \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1$, где $\mathcal{L}_1$ — одномерный подмодуль, порожденный единственной N-строкой $e = e_1 + \ldots + e_m$.

Многомерная множественная регрессия. В одномерном варианте множественной линейной регрессии наблюдение (отклик) y_i, где i — номер наблюдения, $i = 1, \ldots, n$, можно выразить через влияние факторов и ошибок следующим образом:

$$y_i = (a_1, a_2, \ldots, a_m) \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{pmatrix} + \varepsilon_i,$$

где $a_1, a_2, \ldots, a_m$ — неизвестные постоянные коэффициенты; $X_i = (x_{1i}, x_{2i}, \ldots$ $\ldots, x_{mi})^T$ — известные (заданные) значения факторов в опыте i; $\varepsilon_1, \ldots, \varepsilon_n$ — независимые случайные ошибки. Наблюдения $y_1, \ldots, y_n$ известны (после опыта), план $\{x_{\alpha i} \mid \ \alpha = \overline{1, m};\ i = \overline{1, n}\} = \{X_i \mid i = \overline{1, n}\}$ задан. В гауссовской постановке: $\varepsilon_i \sim N(0, \sigma^2)$, σ^2 неизвестна.

В многомерном варианте отклик многомерный, все прочее — как в одномерном случае. Получается, что для каждой координаты многомерного отклика есть своя линейная регрессия. В опыте i

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{pi} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{p1} & a_{p2} & \ldots & a_{pm} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{pi} \end{pmatrix}.$$

В матричной форме можно записать так: для $i = 1, \ldots, n$

$$Y_i = AX_i + E_i, \ \text{где } E_i \sim N_p(0, Q),$$

причем случайные векторы $E_1, \ldots, E_n$ — независимые $N_p(0, Q)$, где Q — неизвестная невырожденная матрица.

Задачи статистики: по наблюдениям оценить A, Q; проверить гипотезы относительно A (все, как в одномерном случае).

Многомерная множественная регрессия в своей матричной форме имеет вид

$$\mathcal{Y} = \mathcal{A}\mathcal{X} + \mathcal{E},$$

где $\mathcal{Y}$ — $(p \times n)$-матрицы p-мерных откликов, где $\mathcal{X}$ — заданная $(m \times n)$-матрица плана, $\mathcal{A}$ — $(p \times n)$-матрица неизвестных коэффициентов регрессии, $\mathcal{E} = \| E_1, E_2, \ldots, E_n \|$ — $(p \times n)$-матрица, составленная из p-столбцов $E_1, E_2, \ldots$ $\ldots, E_n$ независимых случайных ошибок. В гауссовской модели

$$E_i \sim N_p(0, Q),$$

где $(p \times p)$-матрица Q предполагается невырожденной. Обычно Q считается неизвестной.

Пусть $A_1, A_2, \dots, A_m$ — p-столбцы, составляющие матрицу $\mathcal{A}$; пусть $x_1, x_2, \dots, x_m$ — n-строки, составляющие матрицу плана $\mathfrak{X}$. Тогда

$$\mathcal{A}\mathfrak{X} = \sum_{i=1}^{m} A_i x_i.$$

Полученное выражение показывает, что $\mathrm{E}\,\mathcal{Y} = \mathcal{A}\mathfrak{X}$ принадлежит m-мерному линейному подпространству пространства $\mathbb{R}_n^p$, порожденному линейной системой n-строк $x_1, x_2, \dots, x_m$.

§ 2. Теорема об ортогональном разложении

Теорема 6.1. *Пусть $\mathfrak{X} = \| X_1, X_2, \dots, X_n \|$ — гауссовская $(p \times n)$-матрица с независимыми p-столбцами $X_1, X_2, \dots, X_n \in \mathbb{R}_1^p$, причем $\mathit{Var}\, X_i = Q$ для всех $i = \overline{1, n}$. Пусть $\mathcal{L}_1, \mathcal{L}_2, \dots$ — попарно ортогональные подмодули $\mathbb{R}_n^p$, прямая сумма которых составляет $\mathbb{R}_n^p$:*

$$\mathbb{R}_n^p = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \dots .$$

Рассмотрим разложение $(p \times n)$-матрицы $\mathfrak{X}$ на попарно ортогональные проекции $\mathfrak{X}$ на подмодули $\mathcal{L}_1, \mathcal{L}_2, \dots$:

$$\mathfrak{X} = \mathrm{proj}_{\mathcal{L}_1}\mathfrak{X} + \mathrm{proj}_{\mathcal{L}_2}\mathfrak{X} + \dots .$$

Тогда:

(a) *случайные $(p \times n)$-матрицы $\mathrm{proj}_{\mathcal{L}_1}\mathfrak{X}, \mathrm{proj}_{\mathcal{L}_2}\mathfrak{X}, \dots$ независимы, распределены нормально и $E\,\mathrm{proj}_{\mathcal{L}_i}\mathfrak{X} = \mathrm{proj}_{\mathcal{L}_i} E\,\mathfrak{X}$;*

(b) *$\langle \mathrm{proj}_{\mathcal{L}_i}\mathfrak{X},\ \mathrm{proj}_{\mathcal{L}_i}\mathfrak{X} \rangle = W_p(\dim L_i, Q, \Delta_i)$, где $W_p(\nu, Q, \Delta)$ обозначает случайную матрицу (размера $p \times p$), распределенную по Уишарту с ν степенями свободы и параметром нецентральности Δ. В данном случае*

$$\Delta_i = \langle \mathrm{proj}_{\mathcal{L}_i} E\,\mathfrak{X},\ \mathrm{proj}_{\mathcal{L}_i} E\,\mathfrak{X} \rangle.$$

Доказательство. Каждому подмодулю $\mathcal{L} \subset \mathbb{R}_n^p$ в пространстве $\mathbb{R}_n^1$ взаимнооднозначно соответствует некоторое порождающее его линейное подпространство $L \subset \mathbb{R}_n^1$, причем $\dim \mathcal{L} = \dim L$. Пусть подмодулям $\mathcal{L}_1, \mathcal{L}_2, \dots \subset \mathbb{R}_n^p$ соответствуют подпространства $L_1, L_2, \dots \subset \mathbb{R}_n^1$. При этом подпространства $L_1, L_2, \dots \subset \mathbb{R}_n^1$ попарно ортогональны и их прямая сумма образует все пространство $\mathbb{R}_n^1$. Обозначим размерности подмодулей $\mathcal{L}_1, \mathcal{L}_2, \dots \subset \mathbb{R}_n^p$ (и подпространств $L_1, L_2, \dots \subset \mathbb{R}_n^1$) через $m_1, m_2, \dots$.

Выберем в каждом подпространстве $L_1, L_2, \ldots$ ортогональный базис. Пусть для $\mathcal{L}_1$ это n-строки $f_1, \ldots, f_{m_1}$; для $\mathcal{L}_2$ это n-строки $f_{m_1+1}, \ldots, f_{m_1+m_2}$ и т.д. С помощью этих n-строк каждое из подпространств $\mathcal{L}_1, \mathcal{L}_2, \ldots$ можно представить как прямую сумму одномерных подмодулей из $\mathbb{R}_n^p$. Например, $\mathcal{L}_1 = \mathcal{F}_1 \oplus \mathcal{F}_2 \oplus \ldots \oplus \mathcal{F}_{m_1}$, где

$$\begin{aligned}
&\mathcal{F}_1 = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_1,\ \alpha \text{ пробегает } \mathbb{R}_1^p\},\\
&\mathcal{F}_2 = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_2,\ \alpha \text{ пробегает } \mathbb{R}_1^p\},\\
&\ldots\\
&\mathcal{F}_{m_1} = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_{m_1},\ \alpha \text{ пробегает } \mathbb{R}_1^p\}.
\end{aligned}$$

Совокупность всех n-строк $f_1, f_2, \ldots, f_n$ образует ортогональный базис $\mathbb{R}_n^1$ и одновременно образующий базис пространства $\mathbb{R}_n^p$. Поэтому всякую $(p \times n)$-матрицу $\mathcal{X} \in \mathbb{R}_n^p$ можно представить в виде

$$\mathcal{X} = \sum_{i=1}^{n} Y_i f_i,$$

где $Y_1, \ldots, Y_n$ — некоторые p-столбцы, т.е. $Y_1, \ldots, Y_n \in \mathbb{R}_1^p$. При этом

$$\begin{aligned}
\operatorname{proj}_{\mathcal{L}_1} \mathcal{X} &= \sum_{i=1}^{m_1} Y_i f_i,\\
\operatorname{proj}_{\mathcal{L}_2} \mathcal{X} &= \sum_{i=m_1+1}^{m_2} Y_i f_i \qquad \text{и т.д.}
\end{aligned}$$

Здесь p-столбцы $Y_1, Y_2, \ldots, Y_n$ — это координаты $(p \times n)$-матрицы $\mathcal{X}$ относительно порождающего базиса $f_1, \ldots, f_n$. При этом p-столбцы $X_1, X_2, \ldots, X_n$ — это координаты той же $(p \times n)$-матрицы $\mathcal{X}$ относительно ортогонального канонического базиса $\mathbb{R}_n^1$: $e_1 = (1, 0, \ldots)$, $e_2 = (0, 1, 0, \ldots)$ и т.д. Как было отмечено ранее, переход от одних координат к другим осуществляется умножением $(p \times n)$-матрицы $\mathcal{X}$ справа на $(n \times n)$-матрицу перехода, в данном случае — на некоторую ортогональную $(n \times n)$-матрицу C:

$$\| Y_1, Y_2, \ldots, Y_n \| = \| X_1, X_2, \ldots, X_n \| C, \qquad \text{или} \qquad \mathcal{Y} = \mathcal{X} C.$$

Из этого следует, что p-столбцы $Y_1, \ldots, Y_n$ в совокупности распределены нормально. Затем, согласно лемме 6,

$$\operatorname{Var} \mathcal{Y} = \operatorname{Var} \mathcal{X} = \{\delta_{ij} Q \mid i, j = \overline{1, n}\}.$$

Это значит, что p-столбцы $Y_1, \ldots, Y_n$ статистически независимы и имеют $(p \times \times p)$-матрицу ковариаций Q — такую же, как p-столбцы $X_1, \ldots, X_n$.

Рассмотрим случайные $(p \times p)$-матрицы

$$\langle \mathrm{proj}_{\mathcal{L}_1} \mathcal{X}\, \mathrm{proj}_{\mathcal{L}_1} \mathcal{X} \rangle, \langle \mathrm{proj}_{\mathcal{L}_2} \mathcal{X}\, \mathrm{proj}_{\mathcal{L}_2} \mathcal{X} \rangle, \ldots.$$

Например,

$$\langle \mathrm{proj}_{\mathcal{L}_1} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_1} \mathcal{X} \rangle = \sum_{i=1}^{m_1} Y_i Y_i^T.$$

Распределения таких случайных матриц называют распределениями Уишарта. Если при этом $\mathrm{E}\, Y_1 = \mathrm{E}\, Y_2 = \ldots = \mathrm{E}\, Y_{m_1} = 0$, мы получаем так называемое центральное распределение Уишарта $W_p(m_1, Q)$. Как мы уже отмечали ранее, если использовать обозначение $W_p(m, Q)$ не только для закона распределения, но и для случайной матрицы с таким распределением, то можно сказать, что

$$W_p(m, Q) = Q^{\frac{1}{2}} W_p(m, I) Q^{\frac{1}{2}},$$

где $Q^{1/2}$ — единственное симметричное решение матричного уравнения: $Z^2 = = Q$.

Нецентральное распределение Уишарта. Говорят, что случайная $(p \times \times p)$-матрица W имеет нецентральное распределение Уишарта, если

$$W = \sum_{i=1}^{m} (\xi_i + a_i)(\xi_i + a_i)^T,$$

где p-столбцы $\xi_1, \xi_2, \ldots, \xi_m$ — независимые $N_p(0, Q)$, $a_1, a_2, \ldots, a_m$ — некоторые неслучайные p-столбцы, вообще говоря отличные от нуля. Распределение W каким-то образом зависит от p-столбцов $a_1, a_2, \ldots, a_m$. Покажем, что распределение W зависит от упомянутых p-столбцов через так называемый параметр нецентральности — $(p \times p)$-матрицу

$$\Delta = \sum_{i=1}^{m} a_i a_i^T.$$

Введем $(p \times m)$-матрицы

$$\xi = \|\, \xi_1, \xi_2, \ldots, \xi_m \,\|,$$
$$\mathcal{A} = \|\, a_1, a_2, \ldots, a_m \,\|.$$

В этих обозначениях

$$W = \langle \xi + \mathcal{A},\ \xi + \mathcal{A} \rangle.$$

Пусть C — произвольная ортогональная $(m \times m)$-матрица. Положим $\eta = \xi C$. При этом $\eta \stackrel{d}{=} \xi$, и

$$W \stackrel{d}{=} \langle \eta + AC,\ \eta + AC \rangle.$$

Мы видим, что нецентральное распределение Уишарта зависит от $\mathcal{A} = \| a_1, \ldots \ldots, a_m \|$ не непосредственно, а через максимальный инвариант $\mathcal{A}$ при ортогональных преобразованиях, т.е. через $\langle \mathcal{A},\ \mathcal{A} \rangle = \sum_{i=1}^{m} a_i a_i^T$.

Следовательно, в общем случае

$$\langle \mathrm{proj}_{\mathcal{L}_i} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i} \mathcal{X} \rangle = W_p(m_i, Q, \Delta_i),$$

где $\Delta_i = \langle \mathrm{proj}_{\mathcal{L}_i} \mathrm{E}\, \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i} \mathrm{E}\, \mathcal{X} \rangle$. Доказательство теоремы закончено. □

§ 3. Достаточные статистики, наилучшие несмещённые оценки

Рассмотрим линейную гауссовскую модель (см. определение 3). Представим её в матричной форме:

$$\mathcal{X} = \mathcal{M} + \mathcal{E},$$

где $\mathcal{M} = \mathrm{E}\, \mathcal{X}$ — неизвестная $(p \times n)$-матрица;

$$\mathcal{M} = \| M_1, M_2, \ldots, M_n \| \in \mathcal{L},$$

где $\mathcal{L}$ — некоторый заданный подмодуль $\mathbb{R}_n^p$;

$$\mathcal{E} = \| E_1, E_2, \ldots, E_n \|$$

— $(p \times n)$-матрица, p-столбцы которой $E_1, E_2, \ldots, E_n$ — независимые $N_p(0, Q)$ p-мерные случайные величины.

Параметром (неизвестным) этой гауссовской модели служит пара $(\mathcal{M}, Q)$. Найдем для этого параметра достаточную статистику (с помощью теоремы факторизации).

Правдоподобие пары $\mathcal{M}, Q$ по наблюдению $\mathcal{X}$ равно

$$\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} \right)^p \frac{1}{\sqrt{\det Q}} \exp\left\{ -\frac{1}{2} (X_i - M_i)^T Q^{-1} (X_i - M_i) \right\}$$
$$= \left(\frac{1}{\sqrt{2\pi}} \right)^{np} \left(\frac{1}{\sqrt{\det Q}} \right)^n \exp\left\{ -\frac{1}{2} \operatorname{tr} Q^{-1} [\sum_{i=1}^{n} (X_i - M_i)(X_i - M_i)^T] \right\}.$$

Сумма в квадратных скобках — это $\langle \mathcal{X} - \mathcal{M},\ \mathcal{X} - \mathcal{M} \rangle$. Представим $\mathcal{X} - \mathcal{M}$ как

$$\mathcal{X} - \mathcal{M} = (\mathcal{X} - \mathrm{proj_L X}) + (\mathrm{proj_L X} - \mathcal{M}) = (1) + (2)$$

и заметим, что
$$(1) = \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X} \in \mathcal{L}^\perp, \qquad (2) \in \mathcal{L}.$$

Поэтому (теорема Пифагора)
$$\langle \mathcal{X} - \mathcal{M},\ \mathcal{X} - \mathcal{M}\rangle = \langle \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X},\ \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X}\rangle + \langle \mathrm{proj}_{\mathrm{L}}\mathrm{X} - \mathcal{M},\ \mathrm{proj}_{\mathrm{L}}\mathrm{X} - \mathcal{M}\rangle.$$

Заключаем отсюда, что правдоподобие пары $\mathcal{M}, Q$ по наблюдению $\mathcal{X}$ выражается через статистики $\mathrm{proj}_{\mathrm{L}}\mathrm{X}$ и $\langle \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X},\ \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X}\rangle$. Поэтому вместе они достаточны для $\mathcal{M}, Q$.

Статистика $\mathrm{proj}_{\mathrm{L}}\mathrm{X}$, очевидно, является несмещенной оценкой $\mathcal{M}$. Как функция достаточной статистики, она является наилучшей несмещенной оценкой $\mathcal{M}$.

Статистика
$$\frac{1}{\dim \mathcal{L}^\perp}\langle \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X},\ \mathrm{proj}_{\mathrm{L}^\perp}\mathrm{X}\rangle = \frac{1}{n-m} Q^{\frac{1}{2}} W_p(n-m, I) Q^{\frac{1}{2}},$$
очевидно, является несмещенной оценкой Q. Как функция достаточной статистики, она является наилучшей несмещенной оценкой параметра Q.

§ 4. Проверка линейных гипотез

Копируя одномерную линейную модель, линейной гипотезой в многомерной линейной модели
$$\mathrm{E}\,\mathcal{X} = \mathcal{M} \in \mathcal{L},$$
где $\mathcal{L}$ — некоторый заданный подмодуль $\mathbb{R}^p_n$ и p-столбцы матрицы $\mathcal{X}$ — независимые $N_p(0, Q)$ p-мерные случайные величины, мы назвали гипотезу
$$H : \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1,$$
где $\mathcal{L}_1$ — заданный подмодуль, причем $\mathcal{L}_1 \subset \mathcal{L}$.

В этом разделе мы предложим статистики, которые могут служить основой для построения статистических критериев для проверки H, распределённых свободно (при H) от параметров $\mathcal{M}$, Q.

Введём подмодуль $\mathcal{L}_2$, ортогонально дополняющий $\mathcal{L}_1$ до $\mathcal{L}$:
$$\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2.$$

Рассмотрим разложение пространства $\mathbb{R}^p_n$ на три попарно ортогональных подпространства:
$$\mathbb{R}^p_n = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \mathcal{L}^\perp.$$

В силу теоремы 6.1 случайные матрицы

$$S_1 \underset{df}{=} \langle \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X} \rangle \qquad \text{и} \qquad S_2 \underset{df}{=} \langle \mathrm{proj}_{\mathcal{L}_2} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2} \mathcal{X} \rangle$$

статистически независимы и распределены по Уишарту. При этом

$$S_1 = \langle \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X} \rangle = W_p(n-m, Q).$$

Если гипотеза $H : \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1$ верна, то

$$S_2 = \langle \mathrm{proj}_{\mathcal{L}_2} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2} \mathcal{X} \rangle = W_p(m_2, Q).$$

(Здесь и далее приняты обозначения $m = \dim \mathcal{L}$, $m_1 = \dim \mathcal{L}_1$, $m_2 = \dim \mathcal{L}_2$.)

При нарушении гипотезы H распределение Уишарта статистики S_2 становится нецентральным, приобретая параметр нецентральности

$$\Delta = \langle \mathrm{proj}_{\mathcal{L}_2} \mathrm{E}\,\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2} \mathrm{E}\,\mathcal{X} \rangle,$$

который тем больше, чем «сильнее» нарушена гипотеза $H : \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1$.

В одномерном случае (когда $p = 1$) статистики S_1, S_2 превращаются в случайные величины, распределённые как $\sigma^2\chi^2(n-m)$ и $\sigma^2\chi^2(m_2)$. Их отношение (при гипотезе) распределено свободно и потому может быть использовано как статистический критерий для проверки H. Это F-отношение Фишера.

В многомерном случае аналог эф-отношения — это отношение $(p \times p)$-матриц S_2 и S_1. При $n - m \geqslant p$ матрица S_1 не вырождена, и потому существует статистика ($(p \times p)$-матрица)

$$S_2 S_1^{-1} = \langle \mathrm{proj}_{\mathcal{L}_2} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2} \mathcal{X} \rangle\ \langle \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp} \mathcal{X} \rangle^{-1}.$$

В отличие от одномерного случая ($p = 1$) предложенная матричная статистика не распределена свободно: по распределению эта статистика равна

$$Q^{\frac{1}{2}}\ W_p(m_2, I)\ W_p^{-1}(n-m, I)\ Q^{-\frac{1}{2}}.$$

Однако собственные значения этой матрицы при гипотезе H распределены свободно от $\mathcal{M}$, Q. Эти собственные значения совпадают с корнями уравнения (относительно λ)

$$\det(W_p(m_2, I) - \lambda W_p(n-m, I)) = 0.$$

Ранее мы отмечали (теорема 4.5), что эти корни и, следовательно, упомянутые собственные значения инвариантны относительно аффинных преобразований исходных наблюдений. Именно поэтому некоторые функции от собственных значений матрицы $S_2 S_1^{-1}$, т.е. от корней уравнения $\det(S_2 - \lambda S_1) = 0$, традиционно используют в качестве критических статистик для проверки линейных гипотез.

§ 5. Распределение характеристических корней

Приведем без доказательства несколько теорем о совместном распределении характеристических корней [1, гл. 13]. Будем употреблять обозначение $W_p(m,Q)$ как для распределения Уишарта (в данном случае: p-мерное распределение с m степенями свободы и матрицей ковариации Q, см. § 5), так и для случайных матриц с такими законами распределений.

Теорема 6.2. *Пусть $W_p(m,Q)$ и $W_p(n,Q)$ — независимые случайные матрицы, распределённые по Уишарту, причём $m,n \geqslant p \geqslant 1$. Тогда совместную плотность распределения характеристических корней уравнения*

$$\det[W_p(m,Q) - \lambda W_p(n,Q)] = 0$$

дает формула

$$\begin{cases} \mathrm{const} \prod\limits_{i=1}^{p} \lambda_i^{\frac{1}{2}(m-p-1)} \prod\limits_{i=1}^{p} (1+\lambda_i)^{-\frac{1}{2}(m+n)} \prod\limits_{i<j} (\lambda_i - \lambda_j) \\ \qquad\qquad\qquad\qquad \text{\textit{в области} } \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0; \\ 0 \quad \text{\textit{в остальной части пространства} } \mathbb{R}^p. \end{cases}$$

Постоянный (нормирующий) множитель известен, он выражается через Г-функции [1, теорема 13.2.2].

Теорема 6.3. *Пусть $W_p(m,Q)$ и $W_p(n,Q)$ — независимые случайные матрицы, распределённые по Уишарту, причём $n \geqslant p \geqslant m > 0$. Тогда плотность совместного распределения положительных характеристических корней уравнения*

$$\det[W_p(m,Q) - \beta\{W_p(m,Q) + W_p(n,Q)\}] = 0$$

дает формула

$$\begin{cases} \mathrm{const} \prod\limits_{i=1}^{m} \beta_i^{\frac{1}{2}(p-m-1)} \prod\limits_{i=1}^{m} (1-\beta_i)^{\frac{1}{2}(n-p-1)} \prod\limits_{1\leqslant i<j\leqslant m} (\beta_i - \beta_j) \\ \qquad\qquad\qquad\qquad \text{\textit{в области} } \beta_1 \geqslant \beta_2 \geqslant \ldots \geqslant \beta_m \geqslant 0; \\ 0 \quad \text{\textit{в остальной части пространства} } \mathbb{R}^m. \end{cases}$$

Постоянный (нормирующий) множитель известен, он выражается через Г-функции [1, теорема 12.2.3].

Для полноты картины приведем теорему о распределении характеристических корней одной матрицы Уишарта.

Теорема 6.4. *Пусть случайная матрица $W_p(m, Q)$ распределена по Уишарту, причём $n \geqslant p \geqslant 1$. Тогда плотность совместного распределения корней характеристического уравнения*

$$\det[W_p(m, Q) - \lambda I] = 0$$

дает формула

$$\begin{cases} \mathrm{const} \prod\limits_{i=1}^{p} \lambda_i^{\frac{1}{2}(n-p-1)} \prod\limits_{1 \leqslant i < j \leqslant p} (\lambda_i - \lambda_j) \exp\Big\{-\frac{1}{2}\sum\limits_{i=1}^{p} \lambda_i\Big\} \\ \qquad\qquad\qquad\qquad \text{в области } \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p \geqslant 0; \\ 0 \quad \text{в остальной части пространства } \mathbb{R}^p. \end{cases}$$

Постоянный (нормирующий) множитель известен, он выражается через Г-функции [1, теорема 12.3.2].

Т. Андерсон сообщает, что теоремы о распределении характеристических корней были практически одновременно (1939 г.) опубликованы Гиршиком, Роем, Фишером, Хеу и немного позже (1941 г.) Мудом.

Выражения для плотностей совместных распределений характеристических корней при альтернативах были получены значительно позже. К сожалению, эти выражения оказались сложны, и их трудно использовать для вычисления вероятностей. Упомянутые формулы плотностей можно найти, например, в [[17], §10.4]. Формулы для плотностей распределения ранее упомянутых критических статистик к настоящему времени получены. Их также можно найти в [[17], §10.6]. К сожалению, эти сложные формулы трудно использовать для вычисления функций распределения или функций квантилей. Теория предлагает для этих целей различные приближенные формулы, которые я не стану упоминать. Их можно найти, например, в работе [2].

Исключение из сказанного составляет лишь статистика отношения правдоподобия Λ. Приведём точный результат о её распределении при гипотезе.

Теорема 6.5. *Пусть $W_p(m, Q)$ и $W_p(n, Q)$ — независимые случайные матрицы, распределённые по Уишарту, причём $m \geqslant p$. Тогда случайная величина*

$$\Lambda = \frac{\det W_p(m, Q)}{\det\{W_p(m, Q) + W_p(n, Q)\}}$$

распределена как произведение p независимых случайных величин, следующих бета-распределениям с параметрами соответственно

$$\Big(\frac{m-p+1}{2}, \frac{n}{2}\Big), \Big(\frac{m-p+2}{2}, \frac{n}{2}\Big), \ldots, \Big(\frac{m}{2}, \frac{n}{2}\Big).$$

Напоминание. Говорят, что случайная величина следует бета-распределению с параметрами (a, b), причем $a > 0, b > 0$, если плотность ее распределения (сосредоточенного на отрезке $[0, 1]$) задается формулой

$$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \quad \text{для } x \in [0, 1].$$

Здесь $B(a, b)$ — бета-функция Эйлера,

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx.$$

В частности, можно показать, что случайная величина, следующая бета-распределению с параметрами $(\frac{m}{2}, \frac{n}{2})$, где m, n — натуральные числа, равна по распределению

$$\frac{\chi^2(m)}{\chi^2(m) + \chi^2(n)},$$

где $\chi^2(m)$ и $\chi^2(n)$ — независимые хи-квадрат величины со степенями свободы m и n соответственно.

Билодо и Бреннер (Bilodeau, Brenner) в своей работе дают ссылку на программы (на языке S), которые вычисляют функцию распределения и функцию квантилей случайной величины Λ (Appendix C, p. 261) [см. www.dms.umontreal.ca/∼bilodeau]. Авторы рассматривают лишь статистику отношения правдоподобия Λ. Другие функции корней, возможных в качестве критических статистик, не упоминаются.

§ 6. Линейная регрессия

В этом разделе мы рассмотрим важную для приложений линейную модель — многомерную множественную линейную регрессию. Мы уже упоминали эту задачу раньше, в § 1, как пример линейной модели. Там же кратко было дано и определение линейной регрессии. Сейчас мы повторим это определение с большими подробностями, а далее рассмотрим для этой модели основные задачи — оценивание неизвестных параметров и проверку статистических гипотез (линейных) о параметрах.

Многомерная линейная регрессия определяется по образцу одномерной, которую мы сейчас напомним.

В одномерной регрессии наблюдают *отклик* $y, y \in \mathbb{R}^1$, как функцию нескольких *факторов* $x_1, \ldots, x_m \in \mathbb{R}^1$: ожидаемое значение отклика есть $y = f(x_1, \ldots, x_m)$, называемое *регрессией* y *на* $x_1, \ldots, x_m$. Чтобы подчеркнуть, что отклик зависит от нескольких факторов, такую регрессию называют множественной. Регрессию y на $x_1, \ldots, x_m$ называют линейной, если

$$y = \sum_{i=1}^{m} a_j x_j, \qquad \text{где } a_1, \ldots, a_m \text{ — некоторые числа, параметры задачи.}$$

Название «линейная» указывает на то, что функция регрессии $f(x_1, \ldots, x_m)$ — это линейная функция.

Отклик и факторы наблюдаются в нескольких опытах. Число опытов обозначим n, отдельные опыты нумеруем числами от 1 до n. В опыте i наблюдаемое значение отклика есть y_i, значения факторов — $x_{1i}, x_{2i}, \ldots, x_{mi}$. Наблюдаемое значение отклика y_i отличается от его ожидаемого значения $y = \sum_{i=1}^{m} a_j x_j$. Разницу между ними называют ошибкой. Если обозначить ошибку в опыте i через ε_i, то

$$y_i = a_1 x_{1i} + a_2 x_{2i} + \ldots + a_m x_{mi} + \varepsilon_i, \qquad i = \overline{1, n}.$$

Основное предположение гауссовского статистического анализа: ошибки $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ – независимые случайные величины, одинаково распределённые по нормальному закону $N(0, \sigma^2)$.

Задачи статистического анализа: по наблюдениям $y_1, \ldots, y_n$ оценить параметры модели $a_1, \ldots, a_m$ и σ^2 и/или проверить гипотезы об этих параметрах.

Модель линейной множественной регрессии запишем в векторно-матричной форме:

$$y_i = (a_1, a_2, \ldots, a_m) \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{pmatrix} + \varepsilon_i, \qquad i = \overline{1, n},$$

или коротко:

$$y_i = aX_i + \varepsilon_i, \qquad i = \overline{1, n}.$$

Здесь $a = (a_1, a_2, \ldots, a_m)$ — вектор-строка, $X_i = (x_{1i}, x_{2i}, \ldots, x_{mi})^T$ — вектор-столбец; $i = \overline{1, n}$. Векторы $X_1, X_2, \ldots, X_n$ образуют *план* эксперимента. План эксперимента считается заданным. Значения откликов $y_1, y_2, \ldots, y_n$ становятся известны после проведения опытов.

Всю совокупность наблюдений в модели множественной линейной регрессии можно кратко записать в матричной форме. Для этого введем n-строки откликов и случайных ошибок

$$y = (y_1, \dots, y_n),$$
$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$$

и матрицу плана эксперимента

$$\mathcal{X} = \| X_1, \dots, X_n \| = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}.$$

В этих обозначениях всю совокупность соотношений $y_i = aX_i + \varepsilon_i$, $i = \overline{1,n}$, передает единственная формула

$$y = a\mathcal{X} + \varepsilon.$$

В многомерной регрессии наблюдаемый отклик многомерный: $Y_i = (y_{1i}, y_{2i}, \dots, y_{pi})^T$, $i = \overline{1,n}$. Для каждой координаты многомерного отклика действует *своя* линейная регрессия. Значения факторов для этих одномерных линейных регрессий одинаковы, коэффициенты, однако, различны. В векторно-матричной форме наблюдаемое значение отклика Y_i есть

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{pi} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{pi} \end{pmatrix}, \qquad i = \overline{1,n},$$

или коротко:

$$Y_i = AX_i + E_i, \qquad i = \overline{1,n}.$$

Модель многомерной множественной регрессии можно кратко записать в чисто матричной форме. Для этого введем $(p \times n)$-матрицы откликов и случайных ошибок

$$\mathcal{Y} = \| Y_1, Y_2, \dots, Y_n \|,$$
$$\mathcal{E} = \| E_1, E_2, \dots, E_n \|.$$

Теперь всю совокупность наблюдений в многомерной модели множественной регрессии запишем единой формулой

$$\mathcal{Y} = A\mathcal{X} + \mathcal{E}.$$

По предположению, ошибки $E_1, \dots, E_n$ — независимые случайные p-векторы, распределенные одинаково по p-мерному нормальному закону $N_p(0, Q)$. Когда это будет необходимо, неизвестные истинные значения параметров в этой задаче мы будем обозначать (A_0, Q_0).

Задачи статистического анализа — по наблюдениям $\mathcal{Y}$ оценить параметры модели A_0 и Q_0 и/или проверить гипотезы об этих параметрах.

В § 1 было показано, что многомерная модель множественной регрессии — это линейная модель. Ожидаемое значение отклика $\mathcal{Y}$ принадлежит подмодулю

$$\mathcal{L} = \{\mathcal{Z} : \mathcal{Z} = A\mathcal{X} \mid (p \times m)\text{-матрица } A \text{ пробегает множество } \mathbb{R}^p_m\}.$$

Согласно результатам § 3, наилучшей несмещённой оценкой для $\mathrm{E}\,\mathcal{Y} = A_0\mathcal{X}$ является $\mathrm{proj_L Y}$ — проекция $\mathcal{Y}$ на подмодуль $\mathcal{L}$. Это матрица вида $\widehat{A}\mathcal{X}$, ближайшая к $\mathcal{Y}$. Матрица $\widehat{A}$ (оценка A_0) — решение экстремальной задачи

$$|\mathcal{Y} - A\mathcal{X}|^2 \to \min_{A \in \mathbb{R}^p_m}.$$

Наилучшей несмещённой оценкой матрицы Q_0 при этом служит матрица

$$\frac{1}{n-m}|\mathcal{Y} - \widehat{A}\mathcal{X}|^2 = \frac{1}{n-m}|\mathrm{proj_{L^\perp} Y}|^2.$$

Оценивание. Найдём явный вид оценки $\widehat{A}$, решив для параметра A указанную экстремальную задачу. Введём вспомогательную матрицу B того же размера, что и матрица A. Значение B определим позже. Положим $A = B + (A - B)$ и преобразуем $|\mathcal{Y} - A\mathcal{X}|^2$:

$$\begin{aligned}|\mathcal{Y} - A\mathcal{X}|^2 &= |(\mathcal{Y} - B\mathcal{X}) + (B\mathcal{X} - A\mathcal{X})|^2 \\ &= |\mathcal{Y} - B\mathcal{X}|^2 + (B - A)(\mathcal{X}\mathcal{X}^T)(B - A)^T \\ &\qquad + (\mathcal{Y} - B\mathcal{X})\mathcal{X}^T(B - A)^T + (B - A)\mathcal{X}(\mathcal{Y} - B\mathcal{X}).\end{aligned}$$

Выберем B так, чтобы $(\mathcal{Y} - B\mathcal{X})\mathcal{X}^T = 0$:

$$B = (\mathcal{Y}\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1}.$$

Заметим, что при таком выборе B в нуль обращаются оба последних слагаемых в разложении $|\mathcal{Y} - A\mathcal{X}|^2$. Следовательно,

$$|\mathcal{Y} - A\mathcal{X}|^2 = |\mathcal{Y} - B\mathcal{X}|^2 + (B - A)(\mathcal{X}\mathcal{X}^T)(B - A)^T.$$

Мы представили $|\mathcal{Y} - A\mathcal{X}|^2$ в виде суммы двух неотрицательно определённых матриц, причём первое слагаемое не зависит от выбора A. Ясно, что минимального значения — в смысле сравнения симметричных матриц — эта сумма достигает при выборе $A = B$: в этом случае второе слагаемое минимально (равно нулю). Итак, ответ: матричный метод наименьших квадратов даёт для матрицы A оценку

$$\widehat{A} = (\mathcal{Y}\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1}.$$

При этом

$$|\mathrm{proj}_{\mathrm{L}^\perp}\mathrm{Y}|^2 = |\mathcal{Y} - (\mathcal{Y}\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1}\mathcal{X}|^2.$$

Ту же оценку для A мы получим, приложив к задаче многомерной множественной регрессии метод Роя. Выбрав произвольно $u \in \mathbb{R}^p$, умножим слева на u^T соотношение $Y_i = AX_i + E_i$ для каждого $i = \overline{1,n}$, получим равенство

$$u^T Y_i = u^T A X_i + u^T E_i, \qquad i = \overline{1,n}.$$

Положим

$$y_i = u^T Y_i, \quad y = (y_1, \ldots, y_n), \quad a = u^T A, \quad \varepsilon_i = u^T E_i, \quad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n).$$

Отметим, что $\varepsilon_1, \ldots, \varepsilon_n$ — независимые $N(0, \sigma^2)$, где $\sigma^2 = u^T Q u$. В силу полученного равенства величины $y_1, \ldots, y_n$ связаны уравнением одномерной множественной регрессии y на $\mathcal{X}$:

$$y_i = aX_i + \varepsilon_i, \quad i = \overline{1,n}.$$

В матричной форме эти уравнения имеют вид

$$y = a\mathcal{X} + \varepsilon.$$

Оценки наименьших квадратов для a и σ^2 в этой задаче хорошо известны:

$$\widehat{a} = (y\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1} = \Big(\sum_{i=1}^n y_i X_i^T\Big)\Big(\sum_{i=1}^n X_i X_i^T\Big)^{-1}$$

$$(n-m)\widehat{\sigma}^2 = \sum_{i=1}^n (y_i - \widehat{a}X_i)^2 = |u^T\mathcal{Y} - \widehat{a}\mathcal{X}|^2.$$

Здесь m обозначает количество оцениваемых по наблюдениям параметров $a_1, a_2, \ldots, a_m$, т. е. размерность параметра a.

Возвращаясь к прежним обозначениям $a = u^T A$ и $y_i = u^T Y_i$, формулу для $\widehat{a}$ можно записать как

$$u^T \widehat{A} = u^T (\mathcal{Y}\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1}.$$

Это дает для $\widehat{A}$ уже знакомое выражение, полученное по матричному методу наименьших квадратов.

Проверка линейных гипотез. В одномерной модели множественной регрессии, коротко записываемой как $y = a\mathcal{X} + \varepsilon$, линейные гипотезы относятся к параметру $a = (a_1, \ldots, a_m)$. Линейная гипотеза в этом случае означает, что параметры $a_1, \ldots, a_m$ удовлетворяют одному или нескольким заданным линейным соотношениям (линейным уравнениям). Простая по форме (хотя и достаточно общая) линейная гипотеза для рассматриваемой модели состоит в том, что влияние на отклик оказывают не все участвующие в модели факторы, а, скажем для определенности, только первые $r < m$ факторов. Это означает гипотезу

$$H : a_{r+1} = \ldots = a_m = 0.$$

В многомерной модели множественной регрессии $\mathcal{Y} = A\mathcal{X} + \mathcal{E}$ линейная гипотеза о том, что влияние на отклик оказывают только первые $r < m$ факторов, означает, что последние $(m - r)$ столбцов матрицы A нулевые. Чтобы точно сформулировать это предположение, разобьем матрицу A на две подматрицы: $A^{(1)}$ состоит из r первых столбцов матрицы A, $A^{(2)}$ — из $(m - r)$ последних. Запишем это так:

$$A = (A^{(1)} \vdots A^{(2)}).$$

Гипотеза о том, что $(m - r)$ последних факторов не оказывают влияния на отклик, теперь выглядит так:

$$H : A^{(2)} = 0.$$

Эту гипотезу надо проверить по наблюдениям.

Мы знаем, что проверку линейной гипотезы H можно основать на сравнении $|\mathrm{proj}_{\mathcal{L}_2}\mathcal{Y}|^2$ и $|\mathrm{proj}_{\mathrm{L}^\perp}\mathrm{Y}|^2$ (см. § 4). Последняя матрица нам известна из процедуры оценивания:

$$|\mathrm{proj}_{\mathrm{L}^\perp}\mathrm{Y}|^2 = |\mathcal{Y} - (\mathcal{Y}\mathcal{X}^T)(\mathcal{X}\mathcal{X}^T)^{-1}\mathcal{X}|^2.$$

Остается вычислить $|\mathrm{proj}_{\mathcal{L}_2}\mathcal{Y}|^2$.

Выпишем явно подмодуль $\mathcal{L}_1$. Это совокупность $(p \times n)$-матриц вида $(A^{(1)} \vdots\, 0)\mathcal{X}$. В этом произведении матриц существенны только первые r строк матрицы $\mathcal{X}$ (r — это число столбцов матрицы $(A^{(1)} \vdots\, 0)$, не равных тождественно нулю). Обозначим $\mathcal{X}^{(1)}$ матрицу размера $r \times n$, состоящую из r первых строк матрицы $\mathcal{X}$. В этих обозначениях

$$\mathcal{L}_1 = \{\mathcal{Y} \mid \mathcal{Y} = A^{(1)}\mathcal{X}^{(1)},\ \text{где } A^{(1)} \text{ пробегает множество } (p \times r)\text{-матриц}\}.$$

Оценку матрицы $A^{(1)}$ по наблюдениям и остаточное расстояние $|\mathcal{Y} - \operatorname{proj}_{\mathcal{L}_1}\mathcal{Y}|^2$ теперь можно получить, используя результат процедуры оценивания:

$$\widehat{A}^{(1)} = (\mathcal{Y}\mathcal{X}^{(1)T})(\mathcal{X}^{(1)}\mathcal{X}^{(1)T})^{-1}.$$

Остаточная сумма квадратов $|\mathcal{Y} - \widehat{A}^{(1)}\mathcal{X}^{(1)}|^2$ – это квадрат проекции $\mathcal{Y}$ на $\mathcal{L}_1^{\perp}$, ортогональное дополнение подмодуля $\mathcal{L}_1$ до всего модуля $\mathbb{R}_n^p$. Так как

$$\mathbb{R}_n^p = \mathcal{L}^{\perp} \oplus \mathcal{L}_1 \oplus \mathcal{L}_2,$$

то

$$|\operatorname{proj}_{\mathcal{L}_1^{\perp}}\mathcal{Y}|^2 = |\operatorname{proj}_{\mathcal{L}^{\perp}\oplus\mathcal{L}_2}\mathcal{Y}|^2 = |\operatorname{proj}_{\mathcal{L}^{\perp}}\mathcal{Y}|^2 + |\operatorname{proj}_{\mathcal{L}_2}\mathcal{Y}|^2.$$

Поэтому

$$\begin{aligned}|\operatorname{proj}_{\mathcal{L}_2}\mathcal{Y}|^2 &= |\operatorname{proj}_{\mathcal{L}_1^{\perp}}\mathcal{Y}|^2 - |\operatorname{proj}_{\mathcal{L}^{\perp}}\mathcal{Y}|^2 \\ &= |\mathcal{Y} - \operatorname{proj}_{\mathcal{L}_1}\mathcal{Y}|^2 - |\mathcal{Y} - \operatorname{proj}_{\mathcal{L}}\mathcal{Y}|^2 = |\mathcal{Y} - \widehat{A}^{(1)}\mathcal{X}^{(1)}|^2 - |\mathcal{Y} - \widehat{A}\mathcal{X}|^2.\end{aligned}$$

Обозначим $\widehat{A}_H$ оценку матрицы A при гипотезе H:

$$\widehat{A}_H = (\widehat{A}^{(1)} \vdots\, 0).$$

С этим обозначением можно явно указать, что

$$|\operatorname{proj}_{\mathcal{L}_2}\mathcal{Y}|^2 = (\widehat{A} - \widehat{A}_H)(\mathcal{X}\mathcal{X}^T)(\widehat{A} - \widehat{A}_H).$$

В § 4, используя традиционные обозначения, мы назвали эти остаточные квадраты

$$S_1 = |\operatorname{proj}_{\mathcal{L}^{\perp}}\mathcal{Y}|^2, \qquad S_2 = |\operatorname{proj}_{\mathcal{L}_2}\mathcal{Y}|^2.$$

Согласно общей теории, критические статистики для проверки гипотезы $H : A^{(2)} = 0$ должны быть функциями характеристических корней уравнения

$$\det(S_2 - \lambda S_1) = 0.$$

О некоторых критических статистиках такого рода и их распределениях мы уже говорили.

§ 7. Упражнения

1. Рассмотрим модель линейной регрессии с выделенным свободным членом

$$\mathcal{Y} = A\left(\mathcal{X} - \overline{\mathcal{X}}\right) + b + \mathcal{E},$$

где $\mathcal{X} - \overline{\mathcal{X}} = \left\{X_i - \overline{X} | i = \overline{1, n}\right\}$, A – неизвестная матрица, b – неизвестный p-столбец. Вычислите для A и b оценки наименьших квадратов $\widehat{A}$, $\widehat{b}$.

Покажите, что $\widehat{A}$ и $\widehat{b}$ статистически независимы.

2. Рассмотрим задачу линейной регрессии (см. § 6)

$$\mathcal{Y} = A\mathcal{X} + \mathcal{E}.$$

Пусть B — такая неслучайная заданная матрица, что произведение AB существует.

(a) Покажите, что гипотеза $H_0 : AB = 0$ (о матрице A) линейна.

(b) В предположении, что $AB = 0$, дайте оценку A.

(c) Предложите статистики для проверки гипотезы $H_0 : AB = 0$.

3. Рассмотрим двухфакторную аддитивную модель (одно наблюдение в клетке):

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},\ i = \overline{1, r},\ j = \overline{1, s},$$

все элементы — p-столбцы. Постоянные (не случайные) p-столбцы μ,$\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s$ неизвестны, причем $\sum\limits_{i=1}^{r} \alpha_i = 0$, $\sum\limits_{j=1}^{s} \beta_j = 0$. Случайные p-столбцы ε_{ij}, $i = \overline{1, r}$, $j = \overline{1, s}$, — независимые $N_p\,(0, Q)$, матрица Q неизвестна и не вырождена.

(a) Найдите оценки для неизвестных $\mu, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s, Q$.

(b) Покажите, что гипотеза $H_0 : \beta_1 = \ldots = \beta_s$ линейна.

(c) Предложите статистики для проверки гипотезы H_0.

Глава 7

МАТРИЧНАЯ КОРРЕЛЯЦИЯ

§ 1. Определения

По определению *коэффициентом корреляции* между двумя *одномерными* случайными величинами X_1 и X_2, заданными на одном вероятностном пространстве (измеряемыми в одном случайном эксперименте, т.е. имеющими совместное распределение вероятностей), называют

$$\rho = \rho(X_1, X_2) = \frac{\mathrm{E}(X_1 - \mathrm{E}\,X_1)(X_2 - \mathrm{E}\,X_2)}{\sqrt{\mathrm{D}\,X_1}\,\sqrt{\mathrm{D}\,X_2}}.$$

Заметим, что для существования коэффициента корреляции необходимо и достаточно, чтобы существовали невырожденные дисперсии $\mathrm{D}\,X_1$ и $\mathrm{D}\,X_2$.

Перечислим основные свойства коэффициента корреляции.

1. Коэффициент корреляции симметричен: $\rho = \rho(X_1, X_2) = \rho(X_2, X_1)$.
2. Для независимых случайных величин $\rho = 0$.
3. Коэффициент корреляции (в силу неравенства Коши–Буняковского) удовлетворяет соотношению
$$\rho^2 \leqslant 1.$$
4. Равенство $\rho^2 = 1$ достигается, только если случайные величины X_1, X_2 с вероятностью 1 удовлетворяют некоторому линейному уравнению (с неслучайными коэффициентами a_1, a_2, b)
$$a_1 X_1 + a_2 X_2 + b = 0.$$
5. Модуль коэффициента корреляции $|\rho|$ инвариантен относительно аффинных преобразований случайных величин. Пусть $a_1 \neq 0$, $a_2 \neq 0$, b_1 и b_2 – произвольные числа. Тогда
$$|\rho(a_1 X_1 + b_1,\, a_2 X_2 + b_2)| = |\rho(X_1, X_2)|.$$

Коэффициентам корреляций как индикаторам взаимозависимости случайных признаков (величин) традиционно уделяют большое внимание статистики-практики. Для многомерных векторов идея о корреляции случайных переменных реализуется в нескольких вариантах.

Примечание. Сейчас, помимо коэффициента корреляции ρ, в статистике рассматривают и используют и многие другие коэффициенты корреляции. Чтобы отличить от ρ, их обычно сопровождают эпитетами (например, коэффициенты ранговой корреляции, коэффициенты автокорреляции и т.п.). Коэффициент корреляции ρ часто называют коэффициентом корреляции Пирсона. Удачное его английское название – product moment correlation coefficient.

Матричная корреляция, которую мы сейчас определим, многими своими свойствами сходна с коэффициентом корреляции Пирсона ρ. Матричная корреляция – это корреляция между случайными векторами, скажем, $X^{(1)}$ и $X^{(2)}$. Должно существовать их совместное распределение. В этой главе, за исключением последнего § 6, мы не предполагаем нормальности совместного распределения пары случайных векторов $X^{(1)}$ и $X^{(2)}$. Мы лишь предполагаем для простоты, что их матрицы ковариаций Q_{11} и Q_{22} невырождены. Здесь и далее мы пользуемся обозначениями § 5 главы 1:

$$Q_{ij} = \mathrm{E}(X^{(i)} - \mathrm{E}\, X^{(i)})(X^{(j)} - \mathrm{E}\, X^{(j)})^T, \qquad i,j = 1,2.$$

Матричной корреляцией между случайными векторами $X^{(1)}$ и $X^{(2)}$ мы назовём (неслучайную) матрицу

$$\mathcal{P} = \mathcal{P}(X^{(1)}, X^{(2)}) = Q_{11}^{-\frac{1}{2}} Q_{12} Q_{22}^{-\frac{1}{2}}.$$

Матричная корреляция $\mathcal{P}$ – это матрица той же размерности, что и Q_{12}.

Отметим для $\mathcal{P}$ аналоги свойств 1–2 коэффициента корреляции Пирсона. При перестановке переменных

$$\mathcal{P}(X^{(2)}, X^{(1)}) = Q_{22}^{-\frac{1}{2}} Q_{21} Q_{11}^{-\frac{1}{2}} = [\mathcal{P}(X^{(1)}, X^{(2)})]^T.$$

Равенство $\mathcal{P}(X^{(1)}, X^{(2)}) = 0$ возможно, только если $Q_{12} = 0$ (тогда и $\mathcal{P}(X^{(2)}, X^{(1)}) = Q_{21} = 0$). Мы ранее назвали $X^{(1)}$ и $X^{(2)}$ некоррелированными в этом случае. Теперь мы можем сказать с полным правом, что некоррелированность в векторном случае означает обращение в нуль матричной корреляции (с той оговоркой, что в нашем изложении требуется невырожденность матриц $\mathrm{D}\, X^{(1)}, \mathrm{D}\, X^{(2)}$). Тем самым для независимых случайных векторов $X^{(1)}, X^{(2)}$ матричная корреляция $\mathcal{P} = 0$.

Мы далее покажем, что матричная корреляция $\mathcal{P}$ обладает также свойствами, сходными свойствам 3–5 пирсоновской корреляции ρ, тем самым являясь её многомерным аналогом.

§ 2. Неравенство Коши–Буняковского

Рассмотрим ковариационную матрицу разности $X^{(1)} - AX^{(2)}$ для произвольной неслучайной матрицы A (для которой возможны указанные операции):

$$\mathrm{E}(X^{(1)} - AX^{(2)})(X^{(1)} - AX^{(2)})^T \geqslant 0.$$

Не ограничивая общности, положим $\mathrm{E}\,X^{(1)} = 0$ и $\mathrm{E}\,X^{(2)} = 0$. Раскрыв скобки, получим отсюда, что

$$\mathrm{E}\,X^{(1)}X^{(1)T} \geqslant A(\mathrm{E}\,X^{(2)}X^{(1)T}) + (\mathrm{E}\,X^{(1)}X^{(2)T})A^T - A(\mathrm{E}\,X^{(2)}X^{(2)T})A^T.$$

Равенство достигается лишь в случае $X^{(1)} = AX^{(2)}$ с вероятностью 1, и как следствие при $A = (\mathrm{E}\,X^{(1)}X^{(2)T})(\mathrm{E}\,X^{(2)}X^{(2)T})^{-1}$. При этом значении A последнее матричное неравенство принимает вид

$$Q_{11} \geqslant Q_{12}Q_{22}^{-1}Q_{21}.$$

Умножим обе части последнего неравенства слева и справа на $Q_{11}^{-1/2}$ и одновременно заменим Q_{22}^{-1} произведением $Q_{22}^{-1/2}Q_{22}^{-1/2}$. (О том, почему можно производить с матричными неравенствами указанные операции, см. упражнение 1 к гл. 7). Получим, что

$$I \geqslant (Q_{11}^{-\frac{1}{2}}Q_{12}Q_{22}^{-\frac{1}{2}})(Q_{22}^{-\frac{1}{2}}Q_{21}Q_{11}^{-\frac{1}{2}}) = \mathcal{P}\mathcal{P}^T,$$

где I – единичная матрица.

Полученное неравенство

$$\mathcal{P}\mathcal{P}^T \leqslant I$$

можно рассматривать как *матричный аналог неравенства Коши–Буняковского*.

Сделаем еще один шаг. Заметим, что $\mathcal{P}\mathcal{P}^T$ – симметричная неотрицательно определенная матрица. Следовательно, существует матрица $(\mathcal{P}\mathcal{P}^T)^{1/2}$. Эту симметричную неотрицательно определенную матрицу называют (левым) матричным модулем матрицы $\mathcal{P}$ и обозначают $|\mathcal{P}| = (\mathcal{P}\mathcal{P}^T)^{1/2}$. Собственные значения матрицы $|\mathcal{P}|$ называют *сингулярными числами* матрицы $\mathcal{P}$.

Неравенство $\mathcal{P}\mathcal{P}^T \leqslant I$ равносильно тому, что все собственные значения $|\mathcal{P}|^2$ не превосходят единицы (докажите). Легко видеть, что собственные значения $|\mathcal{P}|^2$ – это возведенные в квадрат собственные значения матрицы $|\mathcal{P}|$ (докажите). Мы получили, что собственные значения неотрицательно определенной матрицы $|\mathcal{P}|$ (сингулярные числа матрицы $\mathcal{P}$) неотрицательны и не превосходят единицы. Это свойство матричной корреляции $\mathcal{P} = \mathcal{P}(X^{(1)}, X^{(2)})$ — обобщение свойства 3 из § 1 коэффициента корреляции Пирсона ρ.

Если $|\mathcal{P}| = I$ или $\mathcal{P}\mathcal{P}^T = I$, то, как уже нами было отмечено, ковариационная матрица разности $X^{(1)} - AX^{(2)}$ равна нулю. Это означает существование с вероятностью 1 аффинной связи между $X^{(1)}$ и $X^{(2)}$. Данное свойство можно считать обобщением свойства 4 из § 1 обычного коэффициента корреляции.

§ 3. Аффинная инвариантность

Вместе с матрицей $\mathcal{P}\mathcal{P}^T$, где $\mathcal{P} = \mathcal{P}(X^{(1)}, X^{(2)})$, можно рассматривать и матрицу $\mathcal{P}^T\mathcal{P}$, где $\mathcal{P}^T = \mathcal{P}(X^{(2)}, X^{(1)})$. Матрицу $(\mathcal{P}^T\mathcal{P})^{1/2}$ называют правым матричным модулем матрицы $\mathcal{P}$. Матрицы $\mathcal{P}\mathcal{P}^T$ и $\mathcal{P}^T\mathcal{P}$, вообще говоря, имеют разные размерности, но их ненулевые (положительные) собственные значения совпадают. Поэтому совпадают ненулевые собственные значения матриц $|\mathcal{P}|$ и $|\mathcal{P}^T|$.

Теорема 7.1. *Собственные значения матриц $|\mathcal{P}|$ и $|\mathcal{P}^T|$ инвариантны при невырожденных аффинных преобразованиях исходных случайных векторов.*

Доказательство. Достаточно доказать, например, что собственные значения матрицы $|\mathcal{P}|$ инвариантны при переходе от $X^{(1)}$ к $AX^{(1)}$, где A — невырожденная матрица. (Мы продолжаем для простоты предполагать, что $EX^{(1)} = 0$.) Рассмотрим

$$\mathcal{P}_1 \underset{df}{=} \mathcal{P}(AX^{(1)}, X^{(2)}) = (AQ_{11}A^T)^{-\frac{1}{2}} AQ_{12}Q_{22}^{-\frac{1}{2}}.$$

Поскольку ненулевые собственные значения $|\mathcal{P}_1|$ и $|\mathcal{P}_1^T|$ совпадают, вместо $\mathcal{P}_1\mathcal{P}_1^T$ рассмотрим $\mathcal{P}_1^T\mathcal{P}_1$. Покажем, что $\mathcal{P}_1^T\mathcal{P}_1 = \mathcal{P}^T\mathcal{P}$. Имеем

$$\begin{aligned}\mathcal{P}_1^T\mathcal{P}_1 &= [(AQ_{11}A^T)^{-\frac{1}{2}} AQ_{12}Q_{22}^{-\frac{1}{2}}]^T [(AQ_{11}A^T)^{-\frac{1}{2}} AQ_{12}Q_{22}^{-\frac{1}{2}}] \\ &= Q_{22}^{-\frac{1}{2}} Q_{21}A^T (AQ_{11}A^T)^{-\frac{1}{2}} (AQ_{11}A^T)^{-\frac{1}{2}} AQ_{12}Q_{22}^{-\frac{1}{2}} \\ &= Q_{22}^{-\frac{1}{2}} Q_{21}A^T (AQ_{11}A^T)^{-1} AQ_{21}^T Q_{22}^{-\frac{1}{2}} \\ &= (Q_{22}^{-\frac{1}{2}} Q_{21} Q_{11}^{-\frac{1}{2}})(Q_{11}^{-\frac{1}{2}} Q_{12} Q_{22}^{-\frac{1}{2}}) = \mathcal{P}^T\mathcal{P},\end{aligned}$$

что и требовалось доказать. □

§ 4. Канонические корреляции

В этом разделе мы покажем, что (ненулевые) собственные значения матричных корреляций $|\mathcal{P}|$ и $|\mathcal{P}^T|$ – это канонические корреляции пары случайных векторов $X^{(1)}$, $X^{(2)}$. Поскольку нас будут интересовать только корреляции и ковариации, мы (без потери общности) будем предполагать, что $EX^{(1)} = 0$, $EX^{(2)} = 0$. Пусть $p \underset{df}{=} \dim X^{(1)}$, $q \underset{df}{=} \dim X^{(2)}$ и для определенности $p \leqslant q$.

Определим канонические корреляции. Пусть $\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^q$ – переменные (неслучайные) векторы-столбцы. Рассмотрим коэффициенты корреляции между случайными величинами $\lambda^T X^{(1)}$ и $\mu^T X^{(2)}$:

$$\rho(\lambda^T X^{(1)}, \mu^T X^{(2)}) = \frac{\lambda^T Q_{12}\mu}{\sqrt{\lambda^T Q_{11}\lambda}\,\sqrt{\mu^T Q_{22}\mu}}.$$

Подберем переменные λ и μ так, чтобы этот коэффициент корреляции оказался наибольшим возможным. Ясно, что, не ограничивая общности, можно на λ и μ наложить ограничения: $\lambda^T Q_{11}\lambda = 1$, $\mu^T Q_{22}\mu = 1$. Итак, получили экстремальную задачу:

$$\max_{\substack{\lambda:\lambda^T Q_{11}\lambda=1,\\ \mu:\mu^T Q_{22}\mu=1}} \lambda^T Q_{12}\mu\ .$$

Каноническими корреляциями случайных векторов $X^{(1)}$ и $X^{(2)}$ назовем локальные экстремальные (ненулевые) значения коэффициентов корреляций их линейных форм $\lambda^T X^{(1)}$ и $\mu^T X^{(2)}$, т.е. $\lambda^T Q_{12}$. Легко видеть, что число канонических корреляций не превосходит $\min(p, q)$. В частном случае, когда $1 = p < q$, (единственная) каноническая корреляция имеет специальное название – *коэффициент множественной корреляции*.

Теорема 7.2. *Канонические корреляции пары случайных векторов $X^{(1)}$ и $X^{(2)}$ совпадают с ненулевыми собственными значениями их матричных корреляций $|\mathcal{P}|$ и $|\mathcal{P}^T|$.*

Доказательство. Решим описанную выше экстремальную задачу поиска канонических корреляций. Экстремальное значение в этой задаче обозначим ρ. Для нахождения связанного экстремума используем метод Лагранжа. Функция Лагранжа

$$\Phi(\lambda, \mu) = \lambda^T Q_{12}\mu - \frac{\alpha}{2}(\lambda^T Q_{11}\lambda) - 1 - \frac{\beta}{2}(\mu^T Q_{22}\mu) - 1.$$

Локальные экстремумы в рассматриваемой задаче — решения системы уравнений
$$\frac{\partial \Phi}{\partial \lambda} = 0, \qquad \frac{\partial \Phi}{\partial \mu} = 0$$
с присоединением уравнений связей
$$\lambda^T Q_{11} \lambda = 1, \qquad \mu^T Q_{22} \mu = 1.$$
После вычисления производных получаем систему
$$Q_{12}\mu - \alpha Q_{11}\lambda = 0, \qquad Q_{21}\lambda - \beta Q_{22}\mu = 0.$$

Умножив (слева) первое уравнение на λ^T, второе – на μ^T, с учетом связей получим, что
$$\lambda^T Q_{12}\mu = \alpha, \qquad \mu^T Q_{21}\lambda = \beta.$$
Отсюда заключаем, что $\alpha = \beta = \rho$ (ρ – экстремальное значение). С этими значениями вернемся к предыдущей системе уравнений. Первое уравнение $Q_{12}\mu - \alpha Q_{11}\lambda = 0$ умножим (слева) на $Q_{21}Q_{11}^{-1}$, второе уравнение $Q_{21}\lambda - \beta Q_{22}\mu = 0$ умножим на ρ и результаты сложим. Получим уравнение относительно ρ^2 и μ:
$$Q_{21}Q_{11}^{-1}Q_{12}\mu - \rho^2 Q_{22}\mu = 0.$$
Для существования нетривиального (относительно μ) решения этого уравнения необходимо, чтобы
$$\det(Q_{21}Q_{11}^{-1}Q_{12} - \rho^2 Q_{22}) = 0.$$

Каждый корень ρ^2 этого уравнения – это одно из локальных экстремальных значений в задаче поиска канонических корреляций. Полученное уравнение равносильно уравнению
$$\det(Q_{22}^{-\frac{1}{2}} Q_{21}Q_{11}^{-1}Q_{12}Q_{22}^{-\frac{1}{2}} - \rho^2 I) = 0.$$
Последнее уравнение – это уже известное нам уравнение
$$\det(\mathcal{P}^T\mathcal{P} - \rho^2 I) = 0,$$
определяющее собственные значения матрицы $\mathcal{P}^T\mathcal{P}$. Нам уже известно, что собственные значения (ненулевые) матриц $\mathcal{P}^T\mathcal{P}$ и $\mathcal{P}\mathcal{P}^T$ одинаковы и являются квадратами ненулевых собственных значений матриц $|\mathcal{P}| = (\mathcal{P}\mathcal{P}^T)^{1/2}$ и $|\mathcal{P}^T| = (\mathcal{P}^T\mathcal{P})^{1/2}$. □

Следствие 7.1. *Если $1 = p < q$, то $|\mathcal{P}|$ — коэффициент множественной корреляции.*

Вернемся к уравнению

$$Q_{21}Q_{11}^{-1}Q_{12}\mu - \rho^2 Q_{22}\mu = 0$$

и введем новую (векторную) переменную $m = Q_{22}^{1/2}\mu$, откуда $\mu = Q_{22}^{-1/2}m$. Рассматриваемое уравнение относительно m и ρ приобретет форму

$$Q_{21}Q_{11}^{-1}Q_{12}Q_{22}^{-\frac{1}{2}}m - \rho^2 Q_{22}^{\frac{1}{2}}m = 0,$$

или

$$Q_{22}^{-\frac{1}{2}}Q_{21}Q_{11}^{-1}Q_{12}Q_{22}^{-\frac{1}{2}}m = \rho^2 m.$$

Относительно ρ^2 мы уже установили, что ρ^2 – это собственное значение матрицы $\mathcal{P}^T\mathcal{P}$. Теперь выясняется, что векторы m, решения уравнения, — это соответствующие собственные векторы матрицы $\mathcal{P}^T\mathcal{P}$. Отметим, что эти собственные векторы m являются и собственными векторами матрицы $|\mathcal{P}^T| = (\mathcal{P}^T\mathcal{P})^{\frac{1}{2}}$. К аналогичным выводам, но касающимся матрицы $\mathcal{P}\mathcal{P}^T$, мы придем, если в исследовании системы

$$Q_{12}\mu - \alpha Q_{11}\lambda = 0, \qquad Q_{21}\lambda - \beta Q_{22}\mu = 0$$

отправимся от второго уравнения. Умножим второе уравнение слева на $Q_{12}Q_{22}^{-1}$, первое уравнение умножим на ρ и результаты сложим. Получим уравнение относительно ρ и λ:

$$Q_{12}Q_{22}^{-1}Q_{21}\lambda - \rho^2 Q_{11}\lambda = 0.$$

Введём новую (векторную) переменную $l = Q_{11}^{1/2}\lambda$, $\lambda = Q_{11}^{-1/2}l$. Получим уравнение относительно ρ^2 и l:

$$Q_{11}^{-\frac{1}{2}}Q_{12}Q_{22}^{-1}Q_{21}Q_{11}^{-\frac{1}{2}}l = \rho^2 l.$$

Относительно ρ^2 можно повторить, что значение ρ^2 – это собственное значение матрицы $\mathcal{P}\mathcal{P}^T = |\mathcal{P}|^2$. Значение ρ при этом – это собственное значение матрицы $|\mathcal{P}|$. Теперь выясняется, что векторы l – это собственные векторы матрицы $\mathcal{P}\mathcal{P}^T$. Поэтому эти l – собственные векторы также и матрицы $|\mathcal{P}|$; они соответствуют ее собственным значениям ρ.

Как известно, собственные векторы симметричных матриц всегда могут быть выбраны ортонормированными. Для собственных векторов $m_1, \ldots, m_s$ и $l_1, \ldots, l_r$ это означает, что

$$(m_i, m_j) = \mu_i^T Q_{22}\mu_j = \delta_{ij}, \quad 1 \leqslant i, j \leqslant s,$$
$$(l_i, l_j) = \lambda_i^T Q_{11}\lambda_j = \delta_{ij}, \quad 1 \leqslant i, j \leqslant r.$$

§ 5. Канонические величины

Пусть $\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_k > 0$ — ненулевые канонические корреляции пары случайных векторов $X^{(1)}, X^{(2)}$. Их число k не превосходит $\min(p,q)$ и равно рангу матрицы Q_{12}. Мы уже установили, что канонические корреляции – это собственные значения матриц $|\mathcal{P}|$ и $|\mathcal{P}^T|$. Этим собственным значениям, включая и нулевые собственные значения, соответствуют векторы $\lambda_1, \ldots, \lambda_p \in \in \mathbb{R}^p$, причём $p \geqslant k$, и векторы $\mu_1, \ldots, \mu_q \in \mathbb{R}^q$, $q \geqslant k$.

Рассмотрим случайные величины

$$\lambda_1^T X^{(1)}, \lambda_2^T X^{(1)}, \ldots \text{ и } \mu_1^T X^{(2)}, \mu_2^T X^{(2)}, \ldots .$$

Их называют *каноническими случайными величинами* случайных векторов $X_1^{(1)}$ и $X_2^{(2)}$ соответственно.

Из § 4 нам известно, что

$$\operatorname{Cov}(\lambda_i^T X^{(1)}, \lambda_j^T X^{(1)}) = \lambda_i^T Q_{11} \lambda_j = \delta_{ij}, \quad 1 \leqslant i,j \leqslant p;$$
$$\operatorname{Cov}(\mu_i^T X^{(2)}, \mu_j^T X^{(2)}) = \mu_i^T Q_{22} \mu_j = \delta_{ij}, \quad 1 \leqslant i,j \leqslant q.$$

Установим дополнительно, что

$$\operatorname{Cov}(\lambda_i^T X^{(1)}, \mu_j^T X^{(2)}) = \rho_i \delta_{ij}, \quad 1 \leqslant i,j \leqslant k.$$

Равенство $\operatorname{Cov}(\lambda_i^T X^{(1)}, \mu_j^T X^{(2)}) = \rho_i$ при $1 \leqslant i = j \leqslant k$ следует из связей векторов λ_i, μ_i со значениями коэффициента корреляции $\rho(\lambda X^{(1)}, \mu X^{(2)})$: ρ_i — локальные экстремумы этой билинейной функции, которых она достигает именно при $\lambda = \lambda_i, \mu = \mu_i$

Покажем, что $\operatorname{Cov}(\lambda_i^T X^{(1)}, \mu_j^T X^{(2)}) = 0$ при $1 \leqslant i \neq j \leqslant k$. Вернёмся для этого к системе

$$Q_{12}\mu - \alpha Q_{11}\lambda = 0, \qquad Q_{21}\lambda - \beta Q_{22}\mu = 0.$$

Для μ_i, λ_i, ρ_i уравнения системы превращаются в равенства:

$$Q_{12}\mu_i = \rho_i Q_{11}\lambda_i, \qquad Q_{12}\lambda_j = \rho_j Q_{22}\mu_j, \quad 1 \leqslant i,j \leqslant k.$$

Умножим первое из равенств на λ_j^T слева; получаем, что

$$\lambda_j^T Q_{12}\mu_i = \rho_i \lambda_j^T Q_{11}\lambda_i = 0, \quad \text{если } i \neq j.$$

Умножим второе из равенств на μ_i^T слева; получаем, что

$$\mu_i^T Q_{21}\lambda_j = \rho_j \mu_i^T Q_{22}\mu_j = 0, \quad \text{если } i \neq j.$$

Свойства канонических случайных величин и канонических корреляций, полученные выше, сформулируем в виде нескольких утверждений.

1. Канонические корреляции суть ненулевые собственные значения матриц $|\mathcal{P}|$ и $|\mathcal{P}^T|$.
2. Число ненулевых канонических корреляций $\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_k > 0$ равно рангу матрицы Q_{12}.
3. Канонические случайные величины $\lambda_i^T X^{(1)}, \mu_j^T X^{(2)}, i = \overline{1,r}, j = \overline{1,s}$, обладают следующими свойствами:
$$\begin{aligned} \operatorname{Cov}(\lambda_i^T X^{(1)}, \lambda_j^T X^{(1)}) &= \delta_{ij}, \quad 1 \leqslant i,j \leqslant r; \\ \operatorname{Cov}(\mu_i^T X^{(2)}, \mu_j^T X^{(2)}) &= \delta_{ij}, \quad 1 \leqslant i,j \leqslant s; \\ \operatorname{Cov}(\lambda_i^T X^{(1)}, \mu_j^T X^{(2)}) &= \rho_i \delta_{ij}, \quad 1 \leqslant i,j \leqslant k. \end{aligned}$$

§ 6. Проверка независимости многомерных признаков

Пусть $X^{(1)}$ и $X^{(2)}$ — p- и q-мерные случайные величины, имеющие совместное нормальное распределение, причём $p \leqslant q$. Гауссовские признаки $X^{(1)}$ и $X^{(2)}$ независимы как случайные величины в том и только в том случае, если $Q_{12} = 0$ (тогда и $Q_{21} = 0$). В этом разделе мы обсудим проверку по наблюдениям гипотезы
$$H_0 : Q_{12} = 0.$$

Подготовительные результаты. Пусть $X_1, X_2, \ldots, X_n$, где $X_i = (X_i^{(1)T}, X_i^{(2)T})^T$, $i = \overline{1,n}$, — независимые случайные векторы, $\mathrm{E}\, X_i = 0$, $\mathrm{D}\, X_i = Q_{ii}$. Предположим, что $n \geqslant p + q + 1$. Введём обозначения
$$S_{11} = \sum_{i=1}^n (X_i^{(1)} - \overline{X}^{(1)})(X_i^{(1)} - \overline{X}^{(1)})^T = \sum_{i=1}^n X_i^{(1)} X_i^{(1)T} - n\overline{X}^{(1)}\overline{X}^{(1)T},$$
$$S_{12} = \sum_{i=1}^n (X_i^{(1)} - \overline{X}^{(1)})(X_i^{(2)} - \overline{X}^{(2)})^T = \sum_{i=1}^n X_i^{(1)} X_i^{(2)T} - n\overline{X}^{(1)}\overline{X}^{(2)T},$$
$$S_{22} = \sum_{i=1}^n (X_i^{(2)} - \overline{X}^{(2)})(X_i^{(2)} - \overline{X}^{(2)})^T = \sum_{i=1}^n X_i^{(2)} X_i^{(2)T} - n\overline{X}^{(2)}\overline{X}^{(2)T}.$$

Так как $n > p$, $n > q$, то матрицы S_{11} и S_{22} не вырождены. Выборочным матричным коэффициентом корреляции (выборочной матричной корреляцией) признаков $X^{(1)}$ и $X^{(2)}$ назовем
$$R_n \underset{df}{=} S_{11}^{-1/2} S_{12} S_{22}^{-1/2}.$$

Лемма 7.

$$S_{11} \overset{d}{=} \sum_{i=1}^{n-1} X_i^{(1)} X_i^{(1)T}, \qquad S_{12} \overset{d}{=} \sum_{i=1}^{n-1} X_i^{(1)} X_i^{(2)T}, \qquad S_{22} \overset{d}{=} \sum_{i=1}^{n-1} X_i^{(2)} X_i^{(2)T}.$$

Доказательство. Введём ортогональную $(n \times n)$-матрицу C, последний столбец которой $(\frac{1}{\sqrt{n}}, \ldots \frac{1}{\sqrt{n}})^T$, в остальном произвольную. Таблицу

$$\{X_i \mid i = \overline{1,n}\}$$

умножим (справа) на матрицу C. Элементы новой таблицы обозначим $U_i = (U_i^{(1)T}, U_i^{(2)T})^T$, $i = \overline{1,n}$:

$$\{U_i \mid i = \overline{1,n}\} = \{X_i \mid i = \overline{1,n}\}C.$$

В силу свойств ортогональных преобразований (см. Лемму из п. 5 Главы 1)векторы $U_1, U_2, \ldots, U_n$ независимы и распределены по закону $N_{p+q}(0, Q)$ – так же, как векторы $X_1, \ldots, X_n$. При этом

$$\sum_{i=1}^{n} X_i X_i^T = \sum_{i=1}^{n} U_i U_i^T.$$

В силу выбора C

$$U_n^{(1)} = \sqrt{n}\,\overline{X}^{(1)}, \qquad U_n^{(2)} = \sqrt{n}\,\overline{X}^{(2)}.$$

Поэтому

$$S_{11} = \sum_{i=1}^{n} X_i^{(1)} X_i^{(1)T} - n\overline{X}^{(1)}\overline{X}^{(1)T} = \sum_{i=1}^{n-1} U_i^{(1)} U_i^{(1)T} \overset{d}{=} \sum_{i=1}^{n-1} X_i^{(1)} X_i^{(1)T}.$$

Аналогично доказываются и другие утверждения. □

Следствие 7.2.

$$R_n \overset{d}{=} \left(\sum_{i=1}^{n-1} X_i^{(1)} X_i^{(1)T}\right)^{-1/2} \left(\sum_{i=1}^{n-1} X_i^{(1)} X_i^{(2)T}\right) \left(\sum_{i=1}^{n-1} X_i^{(2)} X_i^{(2)T}\right)^{-1/2}.$$

Далее будет доказана теорема.

Теорема 7.3. *Пусть $X_1, X_2, \ldots, X_n$ — независимые $N_{p+q}(0, Q)$, $n \geqslant p + q + 1$. Тогда*

(a) $S_{11} - S_{12} S_{22}^{-1} S_{21} \overset{d}{=} W_p(n - 1 - q, Q_{11} - Q_{12} Q_{22}^{-1} Q_{21})$;

(b) *при фиксированных* $X_1^{(2)}, \ldots, X_n^{(2)}$

$$S_{12}S_{22}^{-1}S_{21} \stackrel{d}{=} W_p(q, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}, D),$$

где параметр нецентральности $D = Q_{12}Q_{22}^{-1}S_{22}Q_{22}^{-1}Q_{21}$;

(c) *случайные матрицы* $S_{11} - S_{12}S_{22}^{-1}S_{21}$ *и* $S_{12}S_{22}^{-1}S_{21}$ *независимы;*

(d) *если* $Q_{12} = 0$ *(признаки* $X^{(1)}$ *и* $X^{(2)}$ *независимы), то*

$$S_{12}S_{22}^{-1}S_{21} \stackrel{d}{=} W_p(q, Q_{11}), \qquad S_{11} - S_{12}S_{22}^{-1}S_{21} \stackrel{d}{=} W_p(n-1-q, Q_{11}).$$

Доказательство. Мы используем результаты, полученные ранее для линейной регрессии (глава 6, § 6). В модели линейной регрессии многомерные переменные Y и X связаны соотношением $\mathrm{E}\,Y = AX$, где A – неслучайная матрица. В текущем исследовании таким соотношением мы свяжем случайные переменные $X^{(1)}$ и $X^{(2)}$. Чтобы без изменений использовать формулы гл. 6, на время переменим обозначения. Положим $X^{(1)} = Y$, $X^{(2)} = X$. Соответственно выборку (объёма $n-1$) из $N_{p+q}(0, Q)$ обозначим

$$\left\{ \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \mid i = \overline{1, n-1} \right\}.$$

Рассмотрим для каждого $i = \overline{1, n-1}$ условное распределение Y_i при фиксированном X_i. Как известно (см. гл. 1, § 6)

$$Y_i = Q_{12}Q_{22}^{-1}X_i + Z_i,$$

где $Z_i \sim N_p(0, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})$ и X_i, Z_i независимы.

Введём таблицы

$$\mathcal{Y} = \{Y_i \mid i = \overline{1, n-1}\}, \quad \mathcal{X} = \{X_i \mid i = \overline{1, n-1}\}, \quad \mathcal{Z} = \{Z_i \mid i = \overline{1, n-1}\}.$$

Эти таблицы связаны моделью линейной регрессии

$$\mathcal{Y} = Q_{12}Q_{22}^{-1}\mathcal{X} + \mathcal{Z}.$$

Найдём условное распределение таблицы $\mathcal{Y}$ при данном $\mathcal{X}$.

Рассмотрим общую задачу линейной регрессии

$$\mathcal{Y} = A\mathcal{X} + \mathcal{Z}.$$

Укажем для $(p \times q)$-матрицы A оценку наименьших квадратов $\widehat{A}$ по наблюдениям $\mathcal{X}, \mathcal{Y}$, как это делается в гл. 6:

$$\widehat{A} = \langle \mathcal{Y},\ \mathcal{X} \rangle \langle \mathcal{X},\ \mathcal{X} \rangle^{-1},$$

или

$$\widehat{A} = \left(\sum_{i=1}^{n-1} Y_i X_i^T\right)\left(\sum_{i=1}^{n-1} X_i X_i^T\right)^{-1}.$$

Возвращаясь к переменным $X_i^{(1)}, X_i^{(2)}$, получаем

$$\widehat{A} = S_{12}S_{22}^{-1}, \quad \mathrm{proj_L Y} = \widehat{A}\mathcal{X} = S_{12}S_{22}^{-1}\mathcal{X},$$
$$|\mathrm{proj_L Y}|^2 = \langle \widehat{A}\mathcal{X},\ \widehat{A}\mathcal{X}\rangle = S_{12}S_{22}^{-1}S_{21}.$$

Остаточная сумма квадратов при этом

$$\begin{aligned}\langle \mathcal{Y} - \widehat{A}\mathcal{X},\ \mathcal{Y} - \widehat{A}\mathcal{X}\rangle &= \sum_{i=1}^{n-1}(Y_i - \widehat{A}X_i)(Y_i - \widehat{A}X_i)^T \\ &= \sum_{i=1}^{n-1} Y_i Y_i^T - \Big(\sum_{i=1}^{n-1} Y_i X_i\Big)^T \Big(\sum_{i=1}^{n-1} X_i X_i^T)^{-1}(\sum_{i=1}^{n-1} X_i Y_i^T\Big).\end{aligned}$$

После возвращения к переменным $X_i^{(1)}, X_i^{(2)}$ получаем

$$|\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2 = \langle \mathcal{Y} - \widehat{A}\mathcal{X},\ \mathcal{Y} - \widehat{A}\mathcal{X}\rangle = S_{11} - S_{12}S_{22}^{-1}S_{21}.$$

По свойствам ортогональных проекций (гл. 5, § 5) при фиксированном $\mathcal{X}$ случайные величины $|\mathrm{proj_L Y}|^2$ и $|\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2$ независимы и при этом

$$\langle \mathcal{Y} - \widehat{A}\mathcal{X},\ \mathcal{Y} - \widehat{A}\mathcal{X}\rangle = |\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2 \stackrel{d}{=} W_p(n-1-q, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}),$$
$$\langle \widehat{A}\mathcal{X},\ \widehat{A}\mathcal{X}\rangle = |\mathrm{proj_L Y}|^2 \stackrel{d}{=} W_p(q, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21}, D),$$

где параметр нецентральности $D = Q_{12}Q_{22}^{-1}S_{22}Q_{22}^{-1}Q_{21}$.

Заметим, что условное (при фиксированном $\mathcal{X}$) распределение $\langle \mathcal{Y} - \widehat{A}\mathcal{X},\ \mathcal{Y} - \widehat{A}\mathcal{X}\rangle = |\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2 = S_{11} - S_{12}S_{22}^{-1}S_{21}$ одно и то же при любом фиксированном $\mathcal{X}$. Поэтому:

(a) $S_{11} - S_{12}S_{22}^{-1}S_{21} \stackrel{d}{=} W_p(n-1-q, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})$;

(b) $S_{11} - S_{12}S_{22}^{-1}S_{21}$ и $\mathcal{X}$ независимы (как случайные объекты);

(c) так как $|\mathrm{proj_L Y}|^2 = S_{12}S_{22}^{-1}S_{21}$ и $|\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2 = S_{11} - S_{12}S_{22}^{-1}S_{21}$ статистически независимы при фиксированном $\mathcal{X}$ и так как $|\mathrm{proj}_{\mathcal{L}^\perp}\mathcal{Y}|^2$ и $\mathcal{X}$ статистически независимы, то статистически независимы и статистики $S_{12}S_{22}^{-1}S_{21}$ и $S_{11} - S_{12}S_{22}^{-1}S_{21}$;

(d) для независимых признаков, когда $\mathcal{P} = 0$, из сказанного выше следует утверждение (d) теоремы.

□

Распределения канонических корреляций. Рассмотрим уравнение, определяющее выборочные канонические коэффициенты корреляций между признаками $X^{(1)}$ и $X^{(2)}$:

$$\det(R_n R_n^T - r^2 I) = 0.$$

Пусть $r_1^2 \geqslant r_2^2 \geqslant \ldots \geqslant r_p^2$ – решения этого уравнения, т.е. выборочные канонические корреляции. В развернутой форме это уравнение

$$\det[S_{11}^{-1/2} S_{12} S_{22}^{-1} S_{21} S_{11}^{-1/2} - r^2 I] = 0,$$

или

$$\det[S_{12} S_{22}^{-1} S_{21} - r^2 S_{11}] = 0,$$

или

$$\det\Big[S_{12} S_{22}^{-1} S_{21} - \frac{r^2}{1-r^2}(S_{11} - S_{12} S_{22}^{-1} S_{21})\Big] = 0.$$

В силу доказанной выше теоремы 7.3 корни последнего уравнения при фиксированных ${X_1}^{(2)}, \ldots, {X_n}^{(2)}$ распределены так же, как корни характеристического уравнения

$$\det\Big[W_p(q, Q_{11} - Q_{12} Q_{22}^{-1} Q_{21}, D) \\ - \frac{r^2}{1-r^2} \times W_p(n-1-q, Q_{11} - Q_{12} Q_{22}^{-1} Q_{21})\Big] = 0,$$

причём участвующие в этом уравнении уишартовские матрицы независимы.

Для независимых признаков параметр нецентральности в первой из уишартовских матриц обращается в нуль; всё уравнение превращается в уравнение

$$\det\Big[W_p(q, Q_{11}) - \frac{r^2}{1-r^2} W_p(n-1-q, Q_{11})\Big] = 0.$$

Совместное распределение характеристических корней, т.е. совместное распределение при гипотезе независимости случайных величин

$$\frac{r_1^2}{1-r_1^2} > \frac{r_2^2}{1-r_2^2} > \ldots > \frac{r_p^2}{1-r_p^2},$$

даёт теорема 6.2 § 5. (Заметим, что условие $q, n-1-q \geqslant p$ выполнено).

В случае если гипотеза независимости ложна, присутствие положительного (хоть и случайного) параметра нецентральности D приводит к смещению (возрастанию) характеристических корней.

Поэтому для проверки гипотезы независимости $H : Q_{12} = 0$ разумно рассматривать в качестве критических статистик функции от случайных величин

$r_1^2/(1-r_1^2), \ldots, r_p^2/(1-r_p^2)$, которые возрастают при возрастании аргументов. Некоторые такие функции уже упоминались при статистических проверках линейных гипотез в линейных моделях. Эти же функции можно использовать и как критические статистики при проверках гипотезы независимости.

Выпишем некоторые критические статистики явно. Но прежде преобразуем матрицы $S_{12}S_{22}^{-1}S_{21}$ и $S_{11}-S_{12}S_{22}^{-1}S_{21}$, используя выборочные матричные корреляции $R_n = S_{11}^{-1/2}S_{12}S_{22}^{-1/2}$, $R_n^T = S_{22}^{-1/2}S_{21}S_{11}^{-1/2}$:

$$S_{12}S_{22}^{-1}S_{21} = S_{11}^{1/2}R_nR_n^TS_{11}^{1/2},$$
$$S_{11}-S_{12}S_{22}^{-1}S_{21} = S_{11}^{1/2}(I-R_nR_n^T)S_{11}^{1/2}.$$

После такого преобразования характеристическое уравнение приобретает вид

$$\det\Big[R_nR_n^T - \frac{r^2}{1-r^2}(I-R_nR_n^T)\Big] = 0.$$

Возможные критические статистики:
статистика Роя

$$V = \frac{r_1^2}{1-r_1^2} = \max_{u\in\mathbb{R}^p}\frac{u^TR_nR_n^Tu}{u^T(I-R_nR_n^T)u};$$

статистика Лоули–Хотеллинга

$$T^2 = \prod_{i=1}^{p}\frac{r_i^2}{1-r_i^2} = \operatorname{tr}[R_nR_n^T(I-R_nR_n^T)^{-1}];$$

статистика Уилкса

$$\Lambda = \prod_{i=1}^{p}(1-r_i^2) = \det(I-R_nR_n^T);$$

статистика Пилая

$$P = \sum_{i=1}^{p}(1-r_i^2) = \operatorname{tr}(I-R_nR_n^T).$$

Распределение всех перечисленных статистик неизвестно даже при гипотезе, исключая статистику Уилкса (это статистика критерия отношения правдоподобий).

Теорема 7.4. *Пусть $n \geqslant p+q+1$. При гипотезе независимости $H: Q_{12} = 0$ статистика*

$$\Lambda = \det(I-R_nR_n^T)$$

распределена как произведение p независимых случайных величин, следующих бета-распределениям с параметрами соответственно

$$\Big(\frac{n-1-q-p+1}{2},\frac{q}{2}\Big),\Big(\frac{n-1-q-p+2}{2},\frac{q}{2}\Big),\dots,\Big(\frac{n-1-q}{2},\frac{q}{2}\Big).$$

Доказательство. Как мы уже выяснили, для независимых признаков квадраты выборочных канонических корреляций r_i^2, т.е. собственные числа $R_nR_n^T$, удовлетворяют характеристическому уравнению

$$\det\Big[W_p(q)-\frac{r^2}{1-r^2}W_p(n-1-q)\Big]=0$$

с независимыми уишартовскими матрицами $W_p(q)$ и $W_p(n-1-q)$. Данное уравнение эквивалентно

$$\det\Big[W_p(q)+W_p(n-1-q)-\frac{1}{1-r^2}W_p(n-1-q)\Big]=0,$$

или

$$\det[W_p(n-1-q)-(1-r^2)(W_p(q)+W_p(n-1-q))]=0.$$

Следовательно,

$$\Lambda=\det(I-R_nR_n^T)=\prod_i(1-r_i^2)=\frac{\det W_p(n-1-q)}{\det(W_p(q)+W_p(n-1-q))}.$$

Утверждение теоремы теперь следует из теоремы 6.5 главы 6. □

В заключение приведём один результат о распределении (при гипотезе независимости) матричной функции от R_n.

Теорема 7.5. *Пусть $n\geqslant p+q+1$. При гипотезе независимости $H:Q_{12}=$ $=0$ матричная статистика*

$$|R_n(I-R_n^TR_n)^{-1/2}|^2\overset{d}{=}|ZW^{-1/2}|^2,$$

где $Z\perp\!\!\!\perp W$, Z – случайная $(p\times q)$-матрица, все элементы которой суть независимые $N(0,1)$-величины, $W\sim W_q(n-1-p)$.

§ 7. Упражнения

1. S_1,S_2,A невырожденные $(p\times p)$-матрицы такие, что $S_1\geqslant 0, S_2>0$. Докажите эквивалентность следующих условий:

   (a) $S_1\leqslant S_2$;

   (b) $AS_1A^T\leqslant AS_2A^T$;

(c) все собственные числа матрицы $S_1 S_2^{-1}$ не превосходят 1.

2. Пусть $\mathcal{P} = \mathcal{P}(X^{(1)}, X^{(2)})$ – матричная корреляция случайных векторов $X^{(1)}$ и $X^{(2)}$.
Докажите, что:
(a) $|\mathcal{P}(X^{(1)}, X^{(2)})|^2 \leqslant (\mathrm{D}\,X^{(1)})^{-\frac{1}{2}}\,\mathrm{D}[\mathrm{E}(X^{(1)} \mid X^{(2)})](\mathrm{D}\,X^{(1)})^{-\frac{1}{2}}$;
(b) равенство достигается, если и только если $\mathrm{E}(X^{(1)} \mid X^{(2)})$ – регрессия $X^{(1)}$ по $X^{(2)}$ — линейная по $X^{(2)}$ функция.

Замечание. Для одномерных $X^{(1)}, X^{(2)}$ величину, стоящую в правой части неравенства, называют квадратом *корреляционного отношения*.

3. Рассмотрим разбиение случайного вектора X на подвекторы $X^{(1)}, X^{(2)}$ с $\mathrm{D}\,X^{(1)} > 0$, $\mathrm{D}\,X^{(2)} > 0$ и $\mathcal{P} = \mathcal{P}(X^{(1)}, X^{(2)})$ – их матричная корреляция. Пусть $g(\cdot)$ – произвольная борелевская функция, $\mathrm{D}\,g(X^{(2)}) > 0$.
Покажите, что:
(a) $|\mathcal{P}(X^{(1)}, g(X^{(2)}))|^2 \leqslant (\mathrm{D}\,X^{(1)})^{-\frac{1}{2}}\,\mathrm{D}[\mathrm{E}(X^{(1)} \mid X^{(2)})](\mathrm{D}\,X^{(1)})^{-\frac{1}{2}}$;
(b) если $X \sim N(a, Q)$, то
$$\max_{g(\cdot)} |\mathcal{P}(X^{(1)}, g(X^{(2)}))|^2 = |\mathcal{P}(X^{(1)}, X^{(2)})|^2.$$

4. Пусть X – случайный вектор с распределением F_ϑ, где $\vartheta \in \Theta$ – заданное открытое множество. Рассмотрим $\mathcal{K}_\vartheta = \{T = T(X) \mid \mathrm{E}_\vartheta\,T(X) = \vartheta\}$ – класс несмещённых оценок параметра ϑ. Пусть существует $T_0 \in \mathcal{K}_\vartheta$:
$$\text{для любых } T \in \mathcal{K}_\vartheta,\ \text{для любых } \vartheta \in \Theta \quad [\mathrm{D}_\vartheta T \geqslant \mathrm{D}_\vartheta T_0 > 0].$$
Докажите, что для всякого $T \in \mathcal{K}_\vartheta$ справедливо тождество
$$\text{для любых } \vartheta \in \Theta \quad [\mathcal{P}(T, T_0) = (\mathrm{D}_\vartheta\,T)^{-\frac{1}{2}}(\mathrm{D}_\vartheta\,T_0)^{\frac{1}{2}}].$$

5. Пусть случайный вектор X разбит на подвекторы $X^{(1)}, X^{(2)}$. Рассмотрим группу преобразований X
$$\mathcal{G} = \{Y = (Y^{(1)T}, Y^{(2)T})^T \mid Y^{(i)} = A_i X^{(i)} + b_i,\ i = 1, 2\},$$
где A_1 и A_2 (b_1 и b_2) пробегают множество неслучайных матриц (векторов).
Покажите, что всякая инвариантная относительно группы $\mathcal{G}$ функция $f(\mathrm{E}\,X,\ dX)$, т.е. такая, что
$$\text{для любых } Y \in \mathcal{G} \quad [f(\mathrm{E}\,X, \mathrm{D}\,X) = f(\mathrm{E}\,Y, \mathrm{D}\,Y)],$$
есть функция канонических корреляций векторов $X^{(1)}, X^{(2)}$.

Список обозначений

$(p \times n)$-**матрица** – элемент множества R_n^p матриц с p строками и n столбцами.

p-**столбец** – элемент множества $R_p = R_1^p$ p-мерных столбцов.

n-**строка** – элемент множества $R_n = R_n^1$ n-мерных строк.

$\boldsymbol{T}_X = \{X_i \mid i = \overline{1,n}\}$ – n-строка p-столбцов.

$\mathcal{X} = \|X_1, X_2, \ldots, X_n\|$ – $(p \times n)$-матрица.

$X = (x_1, \ldots, x_p)^T$ – представление p-мерной случайной величины в виде столбца, X-случайный p-столбец.

$\mathsf{E}\, X = (\mathsf{E}\, x_1, \ldots, \mathsf{E}\, x_p)^T$ – математическое ожидание p-мерной случайной величины X, p-столбец.

$\mathsf{D}\, X = \mathsf{E}(X - \mathsf{E}\, X)(X - \mathsf{E}\, X)^T$ – матрица ковариаций p-мерной случайной величины X, т.е. $(p \times p)$-матрица. Другое обозначение $\operatorname{Var} \mathbf{X}$

$\operatorname{Cov}(X^{(1)}, X^{(2)}) = \mathsf{E}(X^{(1)} - \mathsf{E}\, X^{(1)})(X^{(2)} - \mathsf{E}\, X^{(2)})^T$. – ковариация пары многомерных случайных величин $X^{(1)}, X^{(2)}$.

$N(a, \sigma^2)$ – нормальное распределение на числовой прямой, a – математическое ожидание, σ^2 – дисперсия.

$N(0,1)$ – стандартное нормальное распределение.

$N_p(a, Q)$ – нормальное распределение в p-мерном пространстве, a – математическое ожидание, Q – матрица ковариаций.

$\overline{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ – выборочное среднее, где $X_1, X_2, \ldots, X_n$ – p-столбцы.

$S = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^T$ – выборочная матрица ковариаций, где $X_1, X_2, \ldots, X_n$ – p-столбцы.

$W_p(m, Q) = \sum_{i=1}^{m} X_i X_i^T$ – распределение Уишарта с m степенями свободы, где $X_1, X_2, \ldots, X_m$ независимые многомерные случайные величины с распределением $N_p(0, Q)$

$W_p(m, Q, \Delta) = \sum_{i=1}^{m} (X_i + a_i)(X_i + a_i)^T$ – нецентральное распределение Уишарта с m степенями свободы и параметром нецентральности

$\Delta = \sum_{i=1}^{m} a_i\, a_i^T$, где $X_1, X_2, \ldots, X_m$ независимые многомерные случайные величины с распределением $N_p(0, Q)$.

$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \sum_{i=1}^{n} X_i Y_i^T$ – скалярное произведение таблиц.

$\mathcal{L}$ – подмодуль в пространстве матриц, замкнутый относительно линейных операций.

$\mathcal{L}^{\perp}$ – ортогональное дополнение подмодуля $\mathcal{L}$.

$\mathsf{proj}_{\mathcal{L}}\, T = \arg\min\limits_{S \in \mathcal{L}} \langle T - S, T - S \rangle$ – проекция таблицы T на подмодуль $\mathcal{L}$.

$Corr(X^{(1)}, X^{(2)}) = (\mathsf{D}\, X^{(1)})^{-\frac{1}{2}} \mathrm{Cov}(X^{(1)}, X^{(2)}) (\mathsf{D}\, X^{(2)})^{-\frac{1}{2}}$ – матричная корреляция случайных величин $X^{(1)}, X^{(2)}$, где $\mathrm{D}\, X^{(1)}$ и $\mathrm{D}\, X^{(2)}$ – невырожденные матрицы ковариаций.

Предметный указатель

Список литературы

1. *Андерсон Т.* Введение в многомерный статистический анализ. М.: Физматлит, 1963. 500 с.
2. *Аренс Х.*, *Лёйтер Ю.* Многомерный дисперсионный анализ. М.: Финансы и статистика, 1985. 232 с.
3. *Боровков А.А.* Математическая статистика: учебник. 3-е изд. М.: Физматлит, 2007. 704 с.
4. *Гантмахер Ф.Р.* Теория матриц. М.: Физматлит, 2005. 560 с.
5. *Колмогоров А.Н.* К обоснованию метода наименьших квадратов. Успехи математических наук. 1946, №1.
6. *Колмогоров А.Н.* Теория вероятностей и математическая статистика: сб. статей. М.: Наука, 1986. С. 267–283.
7. *Леман Э.* Проверка статистических гипотез/ пер. с англ. М.: Наука, 1964. 500 с.
8. *Суханова Е.М.* Матричная корреляция. Теория вероятностей и её применение, №54. С. 383–391.
9. *Тюрин Ю.Н.* Многомерный статистический анализ: геометрическая теория. Теория вероятностей и её применение, №55. С. 36–58.
10. *Тюрин Ю.Н.* Многомерные статистические модели в геометрическом изложении. Сб. [11]. С. 107–118.
11. Современные проблемы математики и механики. Т. IV. Математика. Теория вероятностей и математическая статистика, вып. 3. М.: Изд-во Моск. ун-та, 2009. 180 с.
12. *Шеффе Г.* Дисперсионный анализ. М.: Физматлит, 1963. 628 с.
13. *Ширяев А.Н.* Вероятность. М.: Наука, 1989. 640 с.
14. *Bilodeau M.*, *Brenner D.* Theory of Multivariate Statistics. Springer-Verlag, 1999. 288 p.
15. *Durbin J.*, *Kendall M.G.* The Geometry of Estimation. Biometrika, 1951, Vol. 38. P. 150–158.
16. *Horn R.A.*, *Johnson C.R.* Topics in Matrics Analysis. Cambridge: Cambridge University Press, 1994. 615 p.
17. *Muirhead R.J.* Aspects of Multivariate Statistical Theory. Wiley, 1982, 2005. 673 p.
18. *Roy S.N.* Some Aspects of Multivariate Analysis. Wiley, 1957. 214 p.

Abstract

The book presents the most developed to this day theory of multivariate statistics: the statistics of Gaussian (normally distributed) random variables.

A general theory of multivariate linear models which is represented geometrically constitutes the core of the book. This theory uniformly considers all their partial forms (analysis of variances, linear regression) previously studied separately. The Concept of modules over the rings of square matrices with matrix-valued scalar multiplication serves as a mathematical tool of this new theory. For multivariate data this mathematical structure is analogous to linear algebra. Based on the above theory of linear models and our new notion of matrix correlation a general normal correlation theory is developed.

We expect the reader to be well acquainted with calculus, linear algebra, and basic probability and statistics. The book can be used for a one-semester course of lectures. This book is For all who are interested in multivariate statistics, especially students and graduate students in economics and mathematics.

Key words: Multivariate Normal distribution. Arrays of multivariate date. Modules of Arrays over Rings of Square Matrices. Submodules Matrix-valued scalar Product. Matrix-valued Least Square Method.

Multivariate Linear Models. Linear Hypotheses. Testing linear Hypotheses. Matrix-valued Correlation Coefficient. Testing for Independence of Multivariate Random Variables.

Учебное издание

ЮРИЙ НИКОЛАЕВИЧ ТЮРИН

**МНОГОМЕРНАЯ СТАТИСТИКА:
ГАУССОВСКИЕ ЛИНЕЙНЫЕ МОДЕЛИ**

Художественный редактор
Г.Д. Колоскова

Обложка художника
Н.Н. Аникушина

Корректор
Г.Л. Семенова

Технический редактор
Н.И. Матюшина

Верстка
В.И. Громыко

Подписано в печать 15.04.2011. Формат 60x88/16.
Гарнитура Таймс. Печать офсетная. Усл. печ. л. 8,5.
Тираж 500 экз. Изд. № 9325. Заказ № .

Ордена “Знак Почета” Издательство Московского университета.
125009, Москва, ул. Б.Никитская, 5/7.
Тел.: (495)629-50-91. Факс: (495)697-66-71.
Отдел реализации: (495)939-34-93.

E-mail: secretary-msu-press@yandex.ru.
Сайт: www.msu.ru/depts/MSUPubl2005.

Отпечатано В ФГУП “ПИК ВИНИТИ”,
140010, г. Люберцы Московской обл., Октябрьский пр-т, 403.
Тел. 554-21-86.