Глава 1. Нормальные (гауссовские) вектора - простейшие свойства

1. Определение и простейшие характеристики нормальных векторов.

Определение 1.1. Пусть $\vec{\eta} = (\eta_1, \ldots, \eta_p)^T$ - случайный вектор, где $\eta_i \backsim \mathcal{N}(0,1), i = \overline{1,p}$ независимы и одинаково распределены. Будем говорить, что этот вектор - **p-мерный стандартный нормальный (или гауссовский) вектор**, т.е. он имеет **стандартное p-мерное нормальное (или гауссовское) распределение.** Случайный вектор

$$\vec{\xi} = \vec{a} + B\vec{\eta} \tag{1.1}$$

где B - постоянная (неслучайная) матрица размера $n \times p$ и $\vec{a} \in \mathbb{R}^n$ - постоянный (неслучайный) вектор, называется **n - мерным нормальным (или гауссовским) случайным вектором** , а его распределение - **n-мерным нормальным (или гауссовским) распределением.** Заметим, что представление этого вектора в форме $\vec{\xi} = \vec{a} + B\vec{\eta}$ не единственно (см. упражнение 1.1 в конце главы).

Замечание 1.2. В случае $n = 1$ мы имеем обычную (одномерную) нормальную случайную величину, т. к. если $a \in \mathbb{R}^1$, $B = (b_1, \ldots, b_p)$ - матрица $1 \times p$, т. е. p - мерный вектор, то $\xi = a + B\vec{\eta} = a + \sum\limits_{i=1}^{p} b_i\eta_i$ - линейная комбинация нормальных случайных величин и постоянной величины, т.е. сама является случайной величиной. (Ее распределение - $\mathcal{N}(a, \sum_{i=1}^p b_i^2)$.)

Замечание 1.3. Линейное (или, если говорить точнее, аффинное) преобразование $\vec{\xi}$, т.е. $\vec{a}_1 + B_1\vec{\xi}$ (где $\vec{a}_1 \in \mathbb{R}^q$ - постоянный вектор, B_1 - постоянная матрица размера $q \times n$) также является нормальным вектором, т.к. $\vec{a}_1 + B_1\vec{\xi} = \vec{a}_1 + B_1\vec{a} + B_1B\vec{\eta}$, т.е. $\vec{a}_1 + B_1\vec{\xi} = \tilde{B}\vec{\eta} + \tilde{a}$ для $q \times p$ мерной матрицы $\tilde{B} := B_1B$ и q-мерного вектора $\tilde{a} := \vec{a}_1 + B_1\vec{a}$; таким образом, для случайного вектора $\vec{a}_1 + B_1\vec{\xi}$ имеется выражение через $\vec{\eta}$ типа (1.1), и этот вектор - нормальный q-мерный.

Теперь найдем вектор математических ожиданий и ковариационную (дисперсионную) матрицу $R = \mathbf{D}\vec{\xi} = \operatorname{Var}\vec{\xi}$ для $\vec{\xi} = \vec{a} + B\vec{\eta}$.

Напомним некоторые определения из теории вероятностей.

Определение 1.4. Вектор математических ожиданий (средних значений) для случайного n-мерного вектора $\vec{\zeta} = (\zeta_1, \ldots, \zeta_n)^T$ (не обязательно нормального) определен, если и только если $\mathbf{E}\zeta_i$ при всех $i = \overline{1,n}$ существуют и конечны, и равен, по определению,

$$(\mathbf{E}\zeta_1, \ldots, \mathbf{E}\zeta_n)^T.$$

Обозначение: $\mathbf{E}\vec{\zeta}$. Для краткости этот вектор именуется просто *математическим ожиданием* или *средним значением* исходного вектора $\vec{\zeta}$. Оговорка про существование и конечность $\mathbf{E}\zeta_i$ при всех $i = \overline{1,n}$ необходима, т.к. это не всегда выполнено: например, если первая компонента $\vec{\mu}_1$ случайного вектора $\vec{\mu}$ имеет распределение Коши, то $\mathbf{E}\mu_1$, а, значит, и $\mathbf{E}\vec{\mu}$ не существует. Для таких векторов выполнены простые свойства (указанные в упражнении 1.2).

Определение 1.5. Матрицей ковариаций (или ковариационной матрицей, дисперсионной матрицей) данного n-мерного случайного вектора $\vec{\zeta}$ называется, по определению, квадратная матрица порядка n (обозначаемая как $\mathbf{D}\vec{\zeta}$, $\operatorname{Var}\vec{\zeta}$, $R_{\vec{\zeta}}$), ij-й элемент которой равен $\operatorname{cov}(\zeta_i, \zeta_j)$. И эта матрица определена, если и только если $\mathbf{E}\zeta_i^2$ конечны при всех $i = \overline{1,n}$ - это условие равносильно тому, что $\operatorname{cov}(\zeta_i, \zeta_j) = \mathbf{E}(\zeta_i - \mathbf{E}\zeta_i)(\zeta_j - \mathbf{E}\zeta_j)$ конечны при всех $i, j = \overline{1,n}$. Эта оговорка важна - по тем же причинам, по которым необходима аналогичная оговорка в определении вектора математических ожиданий: $\mathbf{E}\zeta_i^2$ может и не быть конечно (если, например, ζ_i имеет распределение Коши). Матрица ковариаций случайного вектора также обладает некоторыми важными свойствами - см. упражнение 1.3.

В дальнейшем нам понадобится понятие **математического ожидания матрицы**.

Определение 1.6. Пусть X - **случайная матрица**, т.е. матрица, элементы которой - это случайные величины; можно также рассматривать такую матрицу (размера $k \times l$) как случайный вектор размерности kl, записанный в виде матрицы. **Матрица математических ожиданий (средних значений)**, или просто **математическое ожидание (среднее значение)** данной матрицы X определена тогда и только тогда, когда у всех случайных величин - элементов X - существует и конечно математическое ожидание, и определяется как числовая матрица того же размера, ij-й элемент которой математическому ожиданию ij-го элемента матрицы X.

Для этой операции также выполнены некоторые простейшие свойства - они перечислены в упражнении 1.2.

Из свойств операции взятия математического ожидания для векторов (см. упражнение 1.2) $\mathbf{E}\vec{\xi} = \mathbf{E}\vec{a} + B\mathbf{E}\vec{\eta} = \vec{a}$. Мы использовали то, что $\eta_i \backsim \mathcal{N}(0,1)$, откуда $\mathbf{E}\eta_i = 0$ для $i = \overline{1,p}$ и $\mathbf{E}\vec{\eta} = (\mathbf{E}\eta_1, \ldots, \mathbf{E}\eta_p)^T = (0, \ldots, 0)^T$. Т. е. $\vec{a}$ - вектор математических ожиданий для $\vec{\xi}$.

Полагая (в упражнении 1.3.3) $\vec{X} = \vec{\eta}, A = B, \vec{b} = \vec{a}$, имеем: $\operatorname{Var}\vec{\xi} = B \operatorname{Var}\vec{\eta}\ B^T = BB^T$ (ведь $\eta_i \backsim \mathcal{N}(0,1), i = \overline{1,p}$ независимы и $\operatorname{cov}(\eta_i, \eta_j) = 0$ для $i, j = \overline{1,n}, i \neq j$, а $\operatorname{cov}(\mu_i, \mu_i) = \operatorname{Var}\eta_i = 1$ для $i = \overline{1,n}$, и, следовательно, $\operatorname{Var}\vec{\eta} = I_p$, где I_p - единичная матрица порядка p; это обозначение I_p будет использоваться и дальше).

Итак, верно

Утверждение 1.7. Вектор средних значений для $\vec{\xi}$ из формулы (1.1) равен $\vec{a}$, а матрица ковариаций - BB^T.

2. Общие свойства многомерных характеристических функций. Характеристические функции нормальных векторов.

Определение 1.8. Если $\vec{\xi} = (\xi_1, \ldots, \xi_n)^T$ - n-мерный случайный вектор, то его **многомерной (или n-мерной) характеристической функцией** называется, по определению, функция $\varphi : \mathbb{R}^n \to \mathbb{C}$, задаваемая так: для $t \in \mathbb{R}^n$

$$\varphi_{\vec{\xi}}(t) = \mathbf{E}\exp(it^T\vec{\xi}) = \mathbf{E}\exp(i(t, \vec{\xi})) = \int\limits_{\mathbb{R}^n} e^{it^T x} P(dx),$$

где скобки $(\cdot, \cdot)$ обозначают скалярное произведение, а P - распределение вектора $\vec{\xi}$ в $\mathbb{R}^n$.

(Это - естественное обобщение понятия характеристической функции случайной величины: функция $\varphi_\xi : \mathbb{R} \to \mathbb{R}$, задаваемая так: $\varphi_\xi(t) =$
$= \mathbf{E}\exp(it\xi)$, как мы знаем, называется *характеристической функцией случайной величины* ξ. В дальнейшем мы будем именно так - символами $\varphi_\xi, \varphi_{\vec{\xi}}$ - обозначать характеристические функции.)

Ее свойства:

1. $\varphi_{\vec{\xi}}(0) = 1, |\varphi_{\vec{\xi}}(t)| \leq 1$ при всех $t \in \mathbb{R}^n$.
2. $\varphi_{\vec{\xi}}$ равномерно непрерывна на $\mathbb{R}^n$.
3. Если $\vec{Y} = a + B\vec{X}$, где $\vec{X}$ - случайный n-мерный вектор, a - постоянный (неслучайный) вектор размерности m, B - неслучайная (постоянная) матрица размера $m \times n$, то
$$\varphi_{\vec{Y}}(t) = e^{i(t,a)}\varphi_{\vec{X}}(B^T t),$$
где скобки $(\cdot, \cdot)$ обозначают скалярное произведение.

4. Если $\vec{X}, \vec{Y}$ - случайные вектора размерности n, то $\varphi_{\vec{X}} \equiv \varphi_{\vec{Y}}$ на $\mathbb{R}^n$ тогда и только тогда, когда распределения векторов $\vec{X}$ и $\vec{Y}$ совпадают. (Поэтому также

$$\varphi(t) = \int\limits_{\mathbb{R}^n} e^{it^T x} P(dx)$$

называется **функцией распределения для распределения** P в $\mathbb{R}^n$. Т.е. функция распределения случайного вектора есть функция распределения для его распределения.)

5. Если $\vec{X}_j, j = \overline{1,m}$ - n-мерные случайные вектора, то они независимы, если и только если для всех $t \in \mathbb{R}^n$

$$\varphi_{\vec{X}}(t) = \prod_{j=1}^{m} \varphi_{\vec{X}_j}(t),$$

где $\vec{X} = \sum_{j=1}^m \vec{X}_j$.

6. Если $\vec{X}_j$ - k_j-мерный случайный вектор, $j = \overline{1,m}$, а у случайного $k = \sum_{j=1}^m k_j$-мерного вектора $\vec{X}$ первые k_1 компонент составляют вектор $\vec{X}_1$, следующие k_2 компонент - вектор $\vec{X}_2$, ..., последние k_m компонент - вектор $\vec{X}_m$, то $\vec{X}_j, j = \overline{1,m}$ независимы тогда и только тогда, когда при всех $t \in \mathbb{R}^k$

$$\varphi_{\vec{X}}(t) = \prod \varphi_{\vec{X}_j}(t_j),$$

где $t_j \in \mathbb{R}^{k_j}, j = \overline{1,m}$, а вектор t составлен из векторов $t_j, j = \overline{1,m}$: первые k_1 компонент вектора t составляют вектор t_1, следующие k_2 компонент - вектор t_2 и т.д.

В частности, компоненты случайного вектора $\vec{\xi} = (\xi_1, \dots, \xi_n)$ независимы, если и только если

$$\varphi_{\vec{\xi}}(t) = \prod_{j=1}^{n} \varphi_{\xi_j}(t_j)$$

при всех $t = (t_1, \dots, t_n) \in \mathbb{R}^n$.

7. **Теорема непрерывности.** Пусть $\{F_m\}_{m=1}^\infty$ - последовательность
функций распределения в $\mathbb{R}^n$, φ_m - n-мерная характеристическая функция, соответствующая распределению в $\mathbb{R}^n$ с функцией распределения F_m. Тогда:

- Если последовательность $\{F_m\}_{m=1}^\infty$ **слабо сходится** к некоторой n-мерной функции распределения F (это означает, что $F_n(x) \to F(x), n \to \infty$ в каждой точке $x \in \mathbb{R}^n$, в которой F непрерывна), то $\forall\ t \in \mathbb{R}^n$

$$\varphi_n(t) \to \varphi(t),$$

при $n \to \infty$, т.е. характеристические функции поточечно сходятся к характеристической функции предельного распределения. Более того, на компактах в $\mathbb{R}^n$ эта сходимость - равномерная.

- Если при всех $t \in \mathbb{R}^n$ предел $\lim_{m\to\infty} \varphi_m(t)$ существует, и функция

$$\lim_{m\to\infty} \varphi_m(t) = \varphi(t)$$

непрерывна в точке $t = 0$, то она является характеристической функцией некоторого распределения вероятностей в $\mathbb{R}^n$ с функцией распределения F, и последовательность $\{F_m\}_{m=1}^\infty$ слабо сходится к функции F на $\mathbb{R}^n$.

(Все вышеперечисленные свойства верны для произвольных случайных векторов, а не только для нормальных.)

Доказательства данных свойств можно найти в книге [2] (глава 2, \$ 12; глава 3, \$ 1). См. также упражнение 1.7.

Утверждение 1.9. Характеристическая функция для нормального вектора - для случайного вектора $\vec{\xi}$ вида (1.1) - равна

$$\varphi_{\vec{\xi}}(t) = e^{i(t,\vec{a}) - \frac{1}{2}(BB^T t,t)} = e^{i(t,\vec{a}) - \frac{1}{2}(Rt,t)},$$

где $R = BB^T$ - матрица ковариаций нормального вектора $\vec{\xi}$ - см. п.1 данной главы.

Доказательство. По свойству 3 характеристических функций, отмеченному выше,

$$\varphi_{\vec{\xi}}(t) = e^{i(t,\vec{a})}\varphi_\eta(B^T t).$$

Но если обозначить $v := B^T t = (v_1, \ldots, v_p) \in \mathbb{R}^p$, то

$$\varphi_\eta(B^T t) = \varphi_\eta(v) = \prod_{k=1}^{p} \varphi_{\eta_k}(v_k),$$

т.к. компоненты $\eta_k, k = \overline{1,p}$ случайного вектора η независимы (мы воспользовались свойством 6 характеристических функций). Известно, что $\varphi_{\eta_k}(v_k) = \exp(-v_k^2/2)$ - ведь $\eta_k \backsim \mathcal{N}(0,1)$ для всех $k = \overline{1,p}$. Значит,

$$\prod_{k=1}^{p} \varphi_{\eta_k}(v_k) = \prod_{k=1}^{p} e^{-v_k^2/2} = e^{-\frac{1}{2}\sum_{k=1}^{p} v_k^2} = e^{-\frac{1}{2}(v,v)} =$$

$$= e^{-\frac{1}{2}(B^T t, B^T t)} = e^{-\frac{1}{2}(BB^T t,t)} = e^{-\frac{1}{2}(Rt,t)}$$

(вспомним, что $R = BB^T$ - матрица ковариаций вектора $\vec{\xi}$). Отсюда

$$\varphi_{\vec{\xi}}(t) = \exp(i(\vec{a},t) - \frac{1}{2}(Rt,t)). \tag{1.2}$$

Доказательство завершено. ■

Замечание 1.10. Отметим как частный случай полученной формулы, что для p-мерного стандартного нормального вектора $\vec{\eta}$ $\varphi_{\vec{\eta}}(t) = \exp(-(t,t)/2)$. Это было вычислено в процессе только что проведенных выкладок; впрочем, это можно получить и непосредственно из формулы (1.2), если подставить $n = p$, $\vec{a} = \vec{0}$, $B = I_p$ - единичная матрица порядка p.

Замечание 1.11. Формула (1.2) может быть принята за эквивалентное определение многомерного нормального вектора $\vec{\xi}$: если случайный вектор имеет такую характеристическую функцию, как в (1.2), то он называется нормальным. (Ведь распределение случайного вектора однозначно восстанавливается по его характеристической функции - см. свойство 4 характеристических функций, отмеченное выше.)

3. Дополнительные замечания. Вычисление плотности распределения гауссовских векторов.

Заметим, что теория многомерных гауссовских векторов во многом аналогична теории (одномерных) случайных нормальных величин. В самом деле, если $\eta \backsim \mathcal{N}(0,1)$, то случайная величина $\xi = a + \sigma\eta$ имеет распределение $\mathcal{N}(a,\sigma)$, а это - аналог формулы (1.1). Многомерный аналог математического ожидания a - это вектор средних значений $\vec{a}$, а многомерный аналог среднеквадратического отклонения σ (а точнее, дисперсии σ^2) - это матрица ковариаций Σ.

Поэтому многомерное нормальное распределение с вектором математических ожиданий $\vec{a}$ и матрицей ковариаций Σ обозначается так: $\mathcal{N}(\vec{a},\Sigma)$, или так: $\mathcal{N}_n(\vec{a},\Sigma)$, где n - размерность данного распределения. Например, $\mathcal{N}_p(0, I_p)$, где I_p - единичная матрица порядка p - это стандартное p-мерное нормальное распределение. Характеристическая функция у распределения $\mathcal{N}(a,\sigma)$ равна $\varphi_\xi(t) = \exp(iat - t^2\sigma^2/2)$, и эта формула также аналогична формуле (1.2).

Теорема 1.13. Нормальный n - мерный вектор с вектором средних $\vec{a}$ и матрицей ковариаций Σ существует тогда и только тогда, когда матрица Σ **неотрицательно определена** (это записывается так: $\Sigma \geq 0$) , т.е. она симметрична и $\forall z \in \mathbb{R}^n\ z^t \Sigma z \geq 0$.

Доказательство. В одну сторону утверждение очевидно: любая матрица ковариаций любого случайного вектора симметрична и неотрицательно определена - см. упражнение 1.3.4. Обратно: если Σ неотрицательно определена, то, как известно из линейной алгебры, существует единственная неотрицательно определенная матрица B, для которой $B^2 = \Sigma$. (См. упражнение 1.5.)

Т.к. $B = B^T$, то $BB^T = \Sigma$. Пусть случайный вектор $\vec{\eta}$ имеет стандартное n - мерное нормальное распределение. Тогда положим $\vec{\xi} = \vec{a} + B\vec{\eta}$, как в формуле (1.1); как было показано в п.1 данной главы, $\vec{\xi}$ имеет нормальное n - мерное распределение с вектором математических ожиданий $\vec{a}$ и ковариационной матрицей $BB^T = \Sigma$.■

Замечание 1.14. В дальнейшем будем обозначать σ-алгебру борелевских множеств, лежащих в борелевском множестве $U \in \mathbb{R}^n$, через $\mathfrak{B}(U)$, а σ-алгебру всех борелевских множеств из $\mathbb{R}^n$, т.е. $\mathfrak{B}(\mathbb{R}^n)$, как $\mathfrak{B}^n$. Классическую меру Лебега в $\mathbb{R}^n$ обозначим через μ. Напомним, что функция $f : U \to \mathbb{R}^n$, $U \in \mathfrak{B}^m$ называется **борелевской**, если $\forall B \in \mathfrak{B}^n\ f^{-1}(B) \in \mathfrak{B}(U)$.

Как вычислить плотность n - мерного нормального распределения?

Напомним определение плотности случайного вектора.

Определение 1.15. Случайный вектор $\vec{\xi}$ (не обязательно гауссовский) **имеет плотность распределения** $p_{\vec{\xi}}$ (по определению, это - борелевская функция $\mathbb{R}^n \to \mathbb{R}$), если n - мерная функция распределения $F_\xi(x_1, \ldots, x_n)$ случайного вектора $\vec{\xi}$ представима в виде

$$F_\xi(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} p_\xi(t_1, \ldots, t_n) dt_1 \ldots dt_n,$$

где интеграл в правой части понимается в смысле Лебега.

Определение 1.16. Случайный вектор $\vec{\xi}$ (не обязательно гауссовский) **имеет плотность распределения** $p_{\vec{\xi}}$ (по определению, это - борелевская функция $\mathbb{R}^n \to \mathbb{R}$), если $\forall A \in \mathfrak{B}^n$

$$\mathbf{P}\{\ \vec{\xi} \in A\} = \int_A p_\xi(t) d\mu,$$

где интеграл в правой части понимается в смысле Лебега.

Эти определения эквивалентны. (См. упражнение 1.8.)

Докажем предварительно лемму.

Лемма 1.17. Пусть $D, G \subset \mathbb{R}^n$ - открытые множества, $\alpha : D \to G$ - биекция, причем $\alpha \in C^1(D), \alpha^{-1} \in C^1(G)$, и при этом $J(t)$ - якобиан функции α в точке $t \in D$ - отличен от нуля во всех точках $t \in D$. Пусть, далее, $f : G \to \mathbb{R}$ - борелевская функция, $A \in \mathfrak{B}(D)$ - некоторое подмножество, $\alpha(A)$ - его образ при отображении α.

Тогда $f \in L_1(\alpha(A))$, если и только если $f(\alpha(t))|J(t)| \in L_1(A)$. Если имеет место интегрируемость, то

$$\int_A f(\alpha(t))|J(t)| dt = \int_{\alpha(A)} f(x) dx. \tag{1.3}$$

(Доказательство - см. приложение.)

Лемма 1.18. Если $D, G \subset \mathbb{R}^n$ - открытые множества, $\alpha : D \to G$ - биективное отображение, причем $\alpha \in C^1(D), \alpha^{-1} \in C^1(G)$, а случайный n-мерный вектор $\vec{\xi}$ с вероятностью 1 принадлежит D и имеет в D плотность распределения $p(t)$, то

$\alpha(\vec{\xi})$ - случайный n-мерный вектор, лежащий в G с вероятностью 1 и имеющий в G плотность распределения

$$q(x) = \frac{1}{J(\alpha^{-1}(x))} p(\alpha^{-1}(x)),$$

где $t \in G$, $J(t)$ - якобиан отображения α в точке t. (См. упражнение 1.11.)

Утверждение 1.19. Рассмотрим распределение $\mathcal{N}(\vec{a}, \Sigma)$. Если матрица Σ ковариаций вектора, имеющего это распределение, невырождена (или, что равносильно, **положительно определена** - это означает, что она симметрична и $\forall z \in \mathrm{R}^n, z \neq 0$ $z^t \Sigma z > 0$; этот факт записывается как $\Sigma > 0$), то это распределение абсолютно непрерывно и имеет плотность

$$p(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{a})^T \Sigma^{-1} (\vec{x} - \vec{a})\right), \tag{1.4}$$

где $\vec{x} \in \mathbb{R}^n$.

Если же Σ вырождена и $\operatorname{rk} \Sigma = k < n$, найдется линейное многообразие (т.е. плоскость) размерности k в $\mathbb{R}^n$, которому с вероятностью 1 принадлежит случайный вектор с данным распределением. А в этом многообразии данный вектор имеет k - мерное нормальное распределение. (См. упражнение 1.12.)

4. Связь некоррелированности и независимости для гауссовских векторов.

Определение 1.20. Пусть $\vec{X}_1, \vec{X}_2$ - пара случайных векторов (не обязательно гауссовских), имеющих некоторое совместное распределение (т.е. заданных на одном и том же вероятностном пространстве; можно считать, что это наблюдения, регистрируемые в одном и том же эксперименте). **Матрица ковариаций** этих двух векторов, по определению, имеет $\operatorname{cov}(X_{1i}, X_{2j})$ в качестве ij - го элемента ($i = \overline{1,n}, j = \overline{1,m}$), где $n = \dim \vec{X}_1, m = \dim \vec{X}_2$ - размерности данных случайных векторов, $\vec{X}_1 = (X_{11}, \dots, X_{1n})$, $\vec{X}_2 = (X_{21}, \dots, X_{2m})$. Она обозначается через $\operatorname{cov}(\vec{X}_1, \vec{X}_2)$.

Операция взятия матрицы ковариаций двух случайных векторов обладает важными свойствами (см. упражнение 1.13).

Утверждение 1.21. Если $\vec{X}_1, \vec{X}_2$ независимы, то они **некоррелированы**, т.е. их матрица ковариаций - нулевая. (См. упражнение 1.14.)

Утверждение 1.22. Обратное, вообще говоря, неверно для произвольных случайных векторов. (См. упражнение 1.15.)

Утверждение 1.23. Если случайный вектор $\vec{X}$ размерности $n+m$ имеет гауссовское распределение, вектор$\vec{X}_1$ состоит из первых его n компонент, а $\vec{X}_2$ - из последних m, $\vec{X}_1 \backsim \mathcal{N}(\vec{a}_1, \Sigma_1)$, и при этом $\vec{X}_2 \backsim \mathcal{N}(\vec{a}_2, \Sigma_2)$ гауссовские (это следует из задачи 3), то обратное утверждение верно, т.е. из некоррелированности векторов $\vec{X}_1, \vec{X}_2$ следует их независимость.

Доказательство. Матрица ковариаций Σ вектора $\vec{X}$ (размерности $n+m$) равна, как нетрудно вычислить, используя ее определение,

$$\begin{pmatrix} \operatorname{Var} \vec{X}_1 & \operatorname{cov}(\vec{X}_1, \vec{X}_2) \\ \operatorname{cov}(\vec{X}_2, \vec{X}_1) & \operatorname{Var} \vec{X}_2 \end{pmatrix}$$

т.е. (в предположении, что векторы $\vec{X}_1, \vec{X}_2$ некоррелированы)

$$\begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$$

.

Т.е. если $\vec{a} = \left(\begin{smallmatrix} \vec{a}_1 \\ \vec{a}_2 \end{smallmatrix}\right)$ - вектор средних значений для случайного вектора $\vec{X}$, $\vec{t} \in \mathbb{R}^{n+m}$ - произвольный вектор, вектор $\vec{t}_1$ состоит из первых n его компонент, а вектор $\vec{t}_2$ - из последних m, то

$$\varphi_{\vec{X}}(\vec{t}) = \exp(i(\vec{a}, \vec{t}) - \frac{1}{2}\vec{t}^T\Sigma\vec{t}) = \exp(i(\vec{a}_1, \vec{t}_1) + i(\vec{a}_2, \vec{t}_2) -$$

$$-\frac{1}{2}\vec{t}_1^T\Sigma_1\vec{t}_1 - \frac{1}{2}\vec{t}_2^T\Sigma_2\vec{t}_2) = \exp(i(\vec{a}_1, \vec{t}_1) - \frac{1}{2}\vec{t}_1^T\Sigma_1\vec{t}_1) \cdot \exp(i(\vec{a}_2, \vec{t}_2) - \frac{1}{2}\vec{t}_2^T\Sigma_2\vec{t}_2) =$$

$$= \varphi_{\vec{X}_1}(\vec{t}_1)\varphi_{\vec{X}_2}(\vec{t}_2),$$

откуда (как известно из общих свойств характеристических функций - см. свойство 6 в п.2) следует, что векторы $\vec{X}_1, \vec{X}_2$ независимы. ■

Замечание 1.24. В этом доказательстве мы использовали правила блочного перемножения матриц и векторов.

Замечание 1.25. Можно было бы (в случае невырожденных матриц Σ_1, Σ_2) доказать этот факт через формулу (1.5) для плотностей $\vec{X}_1, \vec{X}_2$, но приведенное доказательство проще и, кроме того, оно не использует невырожденность матриц Σ_{11}, Σ_{22} ковариаций данных векторов. (См. упражнение 1.16.)

Утверждение 1.26. (Обобщение данного пункта.) Если первые m_1 компонент нормального n-мерного вектора $\vec{X}$ образуют вектор $\vec{X}_1$, следующие m_2 компонент - вектор $\vec{X}_2$,..., последние m_p компонент - вектор $\vec{X}_p$, $\sum_{k=1}^{p} m_k = n$, то из условия некоррелированности этих векторов, т.е. условия $\operatorname{cov}(\vec{X}_k, \vec{X}_l) = 0$ при $k, l = \overline{1,p}, k \neq l$ следует независимость этих векторов $X_k, k = \overline{1,p}$. Выше это было доказано для $p = 2$.

В частности, если $\vec{X} = (X_1, \ldots, X_n) - n$-мерный нормальный вектор, $\operatorname{cov}(X_i, X_j) = 0$ при $i, j = \overline{1,n}, i \neq j$, то случайные величины - компоненты $X_i, i = \overline{1,n}$ независимы. (См. упражнение 1.17.)

5. Условное распределение для нормальных векторов.

В дальнейшем нам понадобится понятие **условного математического ожидания** для случайных векторов.

Определение 1.27. Пусть $\vec{X} = (X_1, \ldots, X_n)$ - случайный вектор, $\mathfrak{A}$ - σ - подалгебра в исходной σ - алгебре $\mathfrak{F}$ (где $(\Omega, \mathfrak{F}, \mathbf{P})$ - исходное вероятностное пространство). Напомним, что **условным математическим ожиданием** случайной величины ξ относительно $\mathfrak{A}$ (если $\mathbf{E}\xi$ существует и конечно) называется $\mathfrak{A}$ - измеримая случайная величина $\eta : \Omega \to \mathbb{R}$ (т.е. такая, для которой прообраз любого борелевского множества из $\mathbb{R}$ лежит в $\mathfrak{A}$), для которой $\forall A \in \mathfrak{A}$

$$\int\limits_A \xi d\mathbf{P} = \int\limits_A \eta d\mathbf{P}.$$

Обозначение: $\eta = \mathbf{E}(\xi|\mathfrak{A})$. Такая величина η существует по теореме Радона - Никодима, т.к. $\int\limits_A \xi d\mathbf{P}$ - это счетно - аддитивная функция множества на $\mathfrak{A}$. Если же μ - (другая) случайная величина (или вектор) на $(\Omega, \mathfrak{F}, \mathbf{P})$, то **условным математическим ожиданием** $\mathbf{E}(\xi|\mu)$ случайной величины ξ относительно μ называется условное математическое ожидание ξ относительно σ - подалгебры $\mathfrak{A}_\mu$ в $\mathfrak{F}$, порожденной μ ($\mathfrak{A}_\mu$ - это совокупность прообразов всех борелевских множеств в $\mathbb{R}$, если μ - случайная величина, и в $\mathbb{R}^m$, если μ - m - мерный случайный вектор.) Можно доказать, что $\mathbf{E}(\xi|\mu) = f(\mu)$ для некоторой борелевской функции f.

Определение 1.28. Вектором условных математических ожиданий (или просто **условным математическим ожиданием**) случайного вектора $\vec{X}$ относительно μ (предполагается, что вектор $\vec{X}$ имеет математическое ожидание) называется

$$(\mathbf{E}(X_1|\mu), \ldots, \mathbf{E}(X_n|\mu)).$$

Этот вектор обозначается так: $\mathbf{E}(\vec{X}|\mu)$. Т.к. его компоненты $\mathfrak{A}$ - измеримы, то и он сам $\mathfrak{A}$ - измерим. Из того, что для некоторых борелевских функций $f_i, i = \overline{1,n}$ $\mathbf{E}(X_i|\mu) = f_i(\mu)$, следует, что для борелевской вектор - функции $\vec{f} = (f_1, \ldots, f_n)$ выполнено $\mathbf{E}(\vec{X}|\mu) = \vec{f}(\mu)$.

Выполнены следующие важные свойства:

1. $\mathbf{E}(A\vec{X}_1 + B\vec{X}_2|\vec{Y}) = A\mathbf{E}(\vec{X}_1|\vec{Y}) + B\mathbf{E}(\vec{X}_2|\vec{Y})$;
2. $\mathbf{E}(\vec{a}^T\vec{X}|\vec{Y}) = \vec{a}^T\mathbf{E}(\vec{X}|\vec{Y})$,

где $\vec{X}, \vec{X}_1, \vec{X}_2, \vec{Y}$ - случайные вектора, причем первые три из них имеют математические ожидания, A, B - постоянные (неслучайные) матрицы, и их размеры таковы, что операции сложения и умножения в п.1 выполнимы, а $\vec{a}$ - постоянный (неслучайный) вектор той же размерности, что и $\vec{X}$.

3. Если ξ - случайная величина с конечным математическим ожиданием, $\mathfrak{A}$ - σ-подалгебра в исходной σ - алгебре $\mathfrak{F}$, и ξ не зависит от этой σ-подалгебры (т.е. σ-подалгебры $\mathfrak{D}_\xi, \mathfrak{A}$ независимы, где мы обозначаем $\mathfrak{D}_{\vec{\zeta}} = \{\vec{\zeta}^{-1}(B) : B \in \mathfrak{B}(\mathbb{R}^n)\}, n$ - размерность $\vec{\zeta}$, для случайных векторов или, как частный случай, случайных величин $\vec{\zeta}$; независимость σ-подалгебр $\mathfrak{A}_1, \mathfrak{A}_2$ означает, по определению, независимость любых событий $A_1 \in \mathfrak{A}_1, A_2 \in \mathfrak{A}_2$), то выполнено следующее равенство:

$$\mathbf{E}(\xi|\mathfrak{D}) = \mathbf{E}\xi.$$

В частности, если ξ, $\vec{\eta}$ независимы, где $\vec{\eta}$ - любой случайный вектор или, как частный случай, случайная величина (это означает по определению, что ξ не зависит от $\mathfrak{D}_{\vec{\eta}}$, т.е. $\mathfrak{D}_\xi, \mathfrak{D}_{\vec{\eta}}$ независимы), то

$$\mathbf{E}(\xi|\vec{\eta}) = \mathbf{E}\xi.$$

4. Если $\xi, \vec{\eta}$ - случайная величина и случайный n-мерный вектор, $g : \mathbb{R}^n \to \mathbb{R}$ - борелевская функция (т.е. прообраз любого борелевского множества из $\mathbb{R}^n$ - борелевское множество из $\mathbb{R}$), $\mathbf{E}\xi, \mathbf{E}(\xi g(\vec{\eta}))$ существуют и конечны, то

$$\mathbf{E}(\xi g(\vec{\eta})|\vec{\eta}) = g(\vec{\eta})\mathbf{E}(\xi|\vec{\eta}).$$

Подробное изложение теории условных математических ожиданий можно найти в книге [2], гл.2, \$ 7. См. также упражнение 1.18.

Определение 1.29. Условное распределение нормального n - мерного вектора $\vec{X}_1$ относительно нормального m - мерного вектора $\vec{X}_2$ - это функция $f : \mathfrak{B}^n \times \mathbb{R}^m \to [0, 1]$, задаваемая так:

$$f(B, \vec{t}) := \mathbf{P}(\vec{X}_1 \in B|\vec{X}_2 = \vec{t}) = \mathbf{E}(I_{\{\vec{X}_1 \in B\}}|\vec{X}_2 = \vec{t}),$$

где $\vec{t} \in \mathbb{R}^m, B \in \mathfrak{B}^n, \mathfrak{B}^n$ - борелевская σ - алгебра в $\mathbb{R}^n$.

(Обозначение: $\mathrm{Law}(\vec{X}_1|\vec{X}_2)$. Заметим, что, в отличие от условного математического ожидания, для условного распределения нет общепринятых обозначений.)

Это распределение можно изучить, рассматривая *условную характеристическую функцию*, т.е. $\mathbf{E}(\exp(i\vec{u}^T\vec{X}_1|\vec{X}_2) = \mathbf{E}(\exp(i(\vec{u}, \vec{X}_1)|\vec{X}_2), \vec{u} \in \mathbb{R}^n$, где скобки $(\cdot, \cdot)$, как обычно, обозначают скалярное произведение. Далее в этом пункте будем считать, что матрица $\operatorname{cov} \vec{X}_2$ невырождена.

Утверждение 1.30. Такая характеристическая функция равна

$$\exp(i\vec{u}^T(\vec{a}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2)) - \vec{u}^T(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\vec{u}/2).$$

Доказательство. Подберем вектор $\vec{Y}$ размерности m, для которого $\vec{Y}, \vec{X}_2$ независимы и $\vec{Y} = \vec{X}_1 + T\vec{X}_2$ для некоторой постоянной матрицы T (размера $n \times m$). Найдем эту матрицу T. Обозначим $\Sigma_{kj} := \operatorname{cov}(\vec{X}_k, \vec{X}_j), k, j = 1, 2$, т.е. $\Sigma_{jj} = \operatorname{Var} \vec{X}_j, j = 1, 2, \Sigma_{12} = \Sigma_{21}^T$ по утверждению из упражнения 1.14. Независимость $\vec{Y}, \vec{X}$ по предыдущему пункту равносильна некоррелированности, т.е. равенству $\operatorname{cov}(\vec{X}, \vec{Y}) = 0 \Leftrightarrow \operatorname{cov}(\vec{X}_1 + T\vec{X}_2, \vec{X}_2) = 0 \Leftrightarrow \operatorname{cov}(\vec{X}_1, \vec{X}_2) + T\operatorname{cov}(\vec{X}_2, \vec{X}_2) = 0 \Leftrightarrow T = -\Sigma_{12}\Sigma_{22}^{-1}$. (Напомним, что, по предположению, $\operatorname{cov} \vec{X}_2 = \Sigma_{22}$ невырождена.)

Итак, $\vec{Y} = \vec{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2$, $\vec{X}_1 = \Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2 + \vec{Y}$.

Значит,

$$\mathbf{E}(\exp(i\vec{u}^T\vec{X}_1)|\vec{X}_2) = \mathbf{E}(\exp(i\vec{u}^T\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2 + i\vec{u}^T\vec{Y})|\vec{X}_2) =$$
$$= \exp(i\vec{u}^T\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2)\mathbf{E}(\exp(i\vec{u}^T\vec{Y}|\vec{X}_2) = \exp(i\vec{u}^T\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2)\mathbf{E}\exp(i\vec{u}^T\vec{Y}),$$

т.к. $\vec{Y}, \vec{X}_2$ независимы.

(В двух последних равенствах мы использовали вышеуказанные свойства 3, 4 условных математических ожиданий.)

Заметим далее, что нормальный вектор $\vec{Y} = \vec{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2$ имеет вектор средних, равный $\vec{a}_1 - \Sigma_{12}\Sigma_{22}^{-1}\vec{a}_2$ (где положим $\vec{a}_j := \vec{X}_j, i = 1, 2$), и матрицу ковариаций

$$\operatorname{Var}\vec{Y} = \operatorname{cov}(\vec{Y},\vec{Y}) = \operatorname{cov}(\vec{X}_1,\vec{X}_1) - \operatorname{cov}(\vec{X}_1,\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2) - \operatorname{cov}(\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_1,\vec{X}_2) +$$

$$+\operatorname{cov}(\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2,\Sigma_{12}\Sigma_{22}^{-1}\vec{X}_2) = \Sigma_{11} - \operatorname{cov}(\vec{X}_1,\vec{X}_2)(\Sigma_{12}\Sigma_{22}^{-1})^T -$$

$$-\Sigma_{12}\Sigma_{22}^{-1}\operatorname{cov}(\vec{X}_1,\vec{X}_2) + \Sigma_{12}\Sigma_{22}^{-1}\operatorname{cov}(\vec{X}_2,\vec{X}_2)(\Sigma_{12}\Sigma_{22}^{-1})^T =$$

$$= \Sigma_{11} - \Sigma_{12}(\Sigma_{12}\Sigma_{22}^{-1})^T - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}(\Sigma_{12}\Sigma_{22}^{-1})^T =$$

$$= \Sigma_{11} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Отсюда следует, что искомая условная характеристическая функция равна $\exp(i\vec{u}^T(\vec{a}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2)) - \vec{u}^T(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\vec{u}/2)$.■

При фиксированном значении $\vec{X}_2$ это - характеристическая функция m - мерного нормального закона распределения с вектором математических ожиданий $\vec{a}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2)$ и матрицей ковариаций $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Вектор условных математических ожиданий $\vec{a}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2)$(соответствующий данному условному распределению $\operatorname{Law}(\vec{X}_1|\vec{X}_2)$ и условной характеристической функции $\mathbf{E}(\exp(i\vec{u}^T\vec{X}_1|\vec{X}_2))$, есть линейная функция от $\vec{X}_2$, а условная матрица ковариаций $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ не зависит от $\vec{X}_2$.

Утверждение 1.31. Если положить $\vec{\xi} := \vec{X}_1 - \vec{a}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2)$, т.е. $\vec{\xi} + \mathbf{E}(\vec{X}_1|\vec{X}_2) = \vec{X}_1$, то векторы $\vec{X}_2, \vec{\xi}$ независимы.

Доказательство. Вектор $\left(\begin{smallmatrix}\vec{\xi}\\ \vec{X}_2\end{smallmatrix}\right)$ имеет нормальное распределение. Действительно, любая линейная комбинация его компонент есть линейная комбинация компонент вектора $\left(\begin{smallmatrix}\vec{X}_1\\ \vec{X}_2\end{smallmatrix}\right)$, а он имеет нормальное распределение. (Мы дважды используем утверждение упражнение 1.3 - см. выше) Но

$$\operatorname{cov}(\vec{\xi},\vec{X}_2) = \operatorname{cov}(\vec{X}_1 - \vec{a}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2),\vec{X}_2) =$$

$$= \operatorname{cov}(\vec{X}_1 - \vec{a}_1,\vec{X}_2) - \operatorname{cov}(\Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2),\vec{X}_2) =$$

$$= \operatorname{cov}(\vec{X}_1,\vec{X}_2) - \Sigma_{12}\Sigma_{22}^{-1}\operatorname{cov}(\vec{X}_2 - \vec{a}_2,\vec{X}_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$$

(в ходе выкладок мы несколько раз воспользовались указанными в этом пункте свойствами условного математического ожидания - см. выше.) Значит, по п.4 векторы $\vec{X}_2, \vec{\xi}$ независимы.■

6. Регрессия и линейная регрессия. Наилучшая оценка

Определение 1.32 Регрессия случайной величины (или случайного вектора) X_1 относительно случайной величины (или случайного вектора) X_2 - это $\mathbf{E}(X_1|X_2) = f(X_2)$, где f - некоторая (борелевская) функция. Тогда $X_1 = f(X_2) + \varepsilon$, где ε - случайная величина (или случайный вектор) с $\mathbf{E}(\varepsilon|X_2) = 0$, называемая **случайной ошибкой**. Если f линейна, то регрессия называется **линейной**. В разобранном выше случае мы имеем именно линейную регрессию, а $\vec{\xi} := \vec{X}_1 - \vec{a}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\vec{X}_2 - \vec{a}_2) = \vec{X}_1 - \mathbf{E}(\vec{X}_1|\vec{X}_2)$ - это случайная ошибка.

Определение 1.33 Несмещенная оценка случайного n-мерного вектора (или, как частный случай, случайной величины, при $n = 1$) $\vec{X}_1$ по случайному m-мерному вектору (или, как частный случай, случайной величине, при $m = 1$) X_2 называют борелевскую вектор - функцию $\vec{g} : \mathbb{R}^m \to \mathbb{R}^n$, т.е. такую функцию, для которой $\forall B \in \mathfrak{B}^n \vec{g}^{-1}(B) \in \mathfrak{B}^m$, где $\mathfrak{B}^k$ - борелевская σ - алгебра в $\mathbb{R}^k$ (или, в случае, когда $\vec{X}_1, \vec{X}_2$ - случайные величины, борелевскую функцию $g : \mathbb{R} \to \mathbb{R}$), для которой $\mathbf{E}\vec{g}(\vec{X}_2) = \mathbf{E}\vec{X}_1$.

Определение 1.34 Наилучшей несмещенной оценкой для случайного вектора (или, как частный случай, случайной величины) $\vec{X}_1$ по случайному вектору (или, как частный случай, случайной величине) $\vec{X}_2$ называется несмещенная оценка $\vec{g}$, для которой выполнено неравенство

$$\mathbf{E}(X_1 - g(X_2))^2 \leq \mathbf{E}(X_1 - h(X_2))^2, \tag{1.5}$$

если X_1 - случайная величина, и

$$\mathbf{E}(X_1 - g(X_2))(X_1 - g(X_2))^T \leq \mathbf{E}(X_1 - h(X_2))(X_1 - h(X_2))^T, \tag{1.6}$$

если X_1 - случайный вектор, для любой другой несмещенной оценки $\vec{h}$ вектора $\vec{X}_1$ по вектору $\vec{X}_2$.

Замечание 1.35. При этом необходимо пояснить, как сравнивать матрицы $\mathbf{E}(X_1 - g(X_2))(X_1 - g(X_2))^T$ и $\mathbf{E}(X_1 - h(X_2))(X_1 - h(X_2))^T$, если X_1 - случайный вектор. Такие матрицы симметричны, а для симметричных матриц A, B одного и того же порядка можно определить отношение порядка так: $A \geq B$, если и только если $A - B \geq 0$.

Утверждение 1.36. На множестве симметричных квадратных матриц данного порядка это отношение задает частичный порядок. (См. упражнение 1.19.)

Утверждение 1.37. Такое отношение порядка удовлетворяет многим обычным свойствам:

1) Если $A_1 \geq B_1, A_2 \geq B_2$, то $A_1 + A_2 \geq B_1 + B_2$;

2) Если $\lambda \in \mathbb{R}, A \geq B$, то $\lambda A \geq \lambda B$ при $\lambda \geq 0$, $\lambda A \leq \lambda B$ при $\lambda \leq 0$.

3) Если $A \geq B \geq 0$ и A, B невырождены, то $A^{-1} \leq B^{-1}$. (См. упражнение 1.20.)

Утверждение 1.38. Такая оценка имеет вид $\mathbf{E}(X_1|X_2)$, т.е. это как раз регрессия X_1 по X_2.

Доказательство. Рассмотрим сначала случай, когда X_1 - (одномерная) случайная величина.

Заметим, что $\mathbf{E}(X_1|X_2)$ - борелевская функция от X_2, причем $\mathbf{E}\mathbf{E}(X_1|X_2) = \mathbf{E}X_1$. Этим доказано, что $\mathbf{E}(X_1|X_2)$ - несмещенная оценка X_1 по X_2. Теперь докажем неравенство (1.5) для любой другой несмещенной оценки h X_1 по X_2. Заметим, что

$$\mathbf{E}(X_1 - h(X_2))^2 = \mathbf{E}((X_1 - \mathbf{E}(X_1|X_2))^2 + 2(X_1 - \mathbf{E}(X_1|X_2))(\mathbf{E}(X_1|X_2) - h(X_2)) +$$

$$+ (\mathbf{E}(X_1|X_2) - h(X_2))^2) = \mathbf{E}(X_1 - \mathbf{E}(X_1|X_2))^2 +$$

$$+ 2\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2))(\mathbf{E}(X_1|X_2) - h(X_2)) + \mathbf{E}(\mathbf{E}(X_1|X_2) - h(X_2))^2 \geq$$

$$\geq \mathbf{E}(X_1 - \mathbf{E}(X_1|X_2))^2,$$

т.к. $\mathbf{E}(\mathbf{E}(X_1|X_2) - h(X_2))^2 \geq 0$, а

$$\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2))(\mathbf{E}(X_1|X_2) - h(X_2)) = \mathbf{E}\mathbf{E}((X_1 - \mathbf{E}(X_1|X_2)) =$$

$$\mathbf{E}\mathbf{E}((X_1 - \mathbf{E}(X_1|X_2))(\mathbf{E}(X_1|X_2) - h(X_2))|X_2) =$$

$$= \mathbf{E}(\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2)|X_2) \cdot (\mathbf{E}(X_1|X_2) - h(X_2))),$$

ведь мы вынесли случайную величину $\mathbf{E}(X_1|X_2) - h(X_2)$ из-под знака условного математического ожидания относительно X_2, т.к. эта случайная величина есть борелевская функция от X_2, однако

$$\mathbf{E}(\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2)|X_2) \cdot (\mathbf{E}(X_1|X_2) - h(X_2))) = 0,$$

т.к. $\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2)|X_2) = \mathbf{E}(X_1|X_2) - \mathbf{E}(\mathbf{E}(X_1|X_2)|X_2) = \mathbf{E}(X_1|X_2) - \mathbf{E}(X_1|X_2) = 0$ (ведь $\mathbf{E}(X_1|X_2)$ - борелевская функция относительно X_2 и $\mathbf{E}(\mathbf{E}(X_1|X_2)|X_2) = \mathbf{E}(X_1|X_2)$.

Итак, неравенство (1.5) доказано, т.е. утверждение 1.38 доказано для случая, когда X_1 - случайная величина. Теперь рассмотрим общий случай: X_1 - n-мерный случайный вектор. Аналогично, $\mathbf{E}(X_1|X_2)$ - борелевская функция от X_2, причем $\mathbf{E}\mathbf{E}(X_1|X_2) = \mathbf{E}X_1$. Этим доказано, что $\mathbf{E}(X_1|X_2)$ - несмещенная оценка X_1 по X_2. Далее, докажем неравенство (1.6) для любой другой несмещенной оценки h X_1 по X_2.

Пусть $z \in \mathbb{R}^n$ - произвольный постоянный (неслучайный) вектор. Тогда по свойствам условного математического ожидания (см. упражнение 1.18) $\mathbf{E}(z^T X_1|X_2) = z^T\mathbf{E}(X_1|X_2)$. Но, по доказанному выше, $\mathbf{E}(z^T X_1|X_2)$ - наилучшая несмещенная оценка случайной величины $z^T X_1$ по X_2. Кроме того, $\mathbf{E}(z^T h(X_2)) = z^T\mathbf{E}h(X_2) = z^T\mathbf{E}X_1$ по задаче 1.2.5, т.е. $z^T h(X_2)$ - несмещенная оценка $z^T X_1$ по X_2. Т.е.

$$\mathbf{E}(z^T X_1 - z^T\mathbf{E}(X_1|X_2))^2 \leq \mathbf{E}(z^T X_1 - z^T h(X_2))^2,$$

или, эквивалентно,

$$z^T\mathbf{E}(X_1 - \mathbf{E}(X_1|X_2))(X_1 - \mathbf{E}(X_1|X_2))^T z \leq z^T\mathbf{E}(X_1 - h(X_2))\mathbf{E}(X_1 - h(X_2))^T z,$$

(мы использовали утверждение упражнения 1.2.11), но в силу произвольности вектора $\vec{z} \in \mathbb{R}^n$ это и означает, что выполнено неравенство (1.6). ■

Заметим, что это доказательство очень похоже на доказательство теоремы Блэкуелла - Рао (Blackwell - Rao) об улучшении несмещенной оценки с помощью достаточной статистики. (Она будет сформулирована в следующей главе.)

7. Датчик нормальных случайных величин.

Как написать программу, которая выдавала бы выборку требуемого объема из стандартного нормального (одномерного) распределения? Можно, используя датчик случайных чисел, дающий выборку $(\xi_1, \ldots, \xi_n)$ из равномерного на $[0,1]$ распределения, построить выборку $(\Phi^{-1}(\xi_1), \ldots, \Phi^{-1}(\xi_n))$ из $\mathcal{N}(0,1)$, где $\Phi(x) = \frac{1}{2\pi}\int_{-\infty}^{x}\exp(-t^2/2)dt$. (Здесь мы пользуемся тем, что если случайная величина ξ имеет равномерное на $[0,1]$ распределение, то $\Phi^{-1}(\xi) \backsim \mathcal{N}(0,1)$ - ведь $\mathbf{P}\{\Phi^{-1}(\xi) \leq x\} = \mathbf{P}\{\xi \leq \Phi(x)\} =$ (т.к. ξ равномерно распределена на $[0,1]$) $= \Phi(x)$)

Но функцию Φ^{-1} на практике трудно вычислять. Поэтому поступают иначе: если вектор(X_1, X_2) имеет двумерное стандартное нормальное распределение, то $\arg(X_1, X_2)$ имеет равномерное распределение на $[0, 2\pi]$. (Где $\arg(X_1, X_2)$ -это такой угол $\varphi \in [0, 2\pi)$, что $\cos\varphi = X_1/\sqrt{X_1^2 + X_2^2}, \sin\varphi = X_2/\sqrt{X_1^2 + X_2^2}$. Он определен тогда и только тогда, когда для X_1, X_2 неверно, что $X_1 = X_2 = 0$, т.е. почти наверное. При этом $\arg(X_1, X_2)$ распределен равномерно на $[0, 2\pi]$.

(Интуитивно это очевидно, т.к. совместное распределение (X_1, X_2) "изотропно т.е. не зависит от направления вектора (X_1, X_2), а $\arg(X_1, X_2)$ зависит только от этого направления. См. также упражнение 1.21.) Т.е. $\arg(X_1, X_2)/2\pi$ имеет равномерное распределение на $[0,1]$. Далее, $X_1^2 + X_2^2 \backsim \chi^2(2) = \Gamma(2/2, 1/2)$ - а это показательное распределение с параметром $1/2$. Здесь использовалось то, что $\chi^2(n)$ и $\Gamma(n/2, 1/2)$ - это одно и то же распределение - см. упражнение 1.22.

(Напомним, что $\Gamma(\alpha, \lambda)$ - это одномерное абсолютно непрерывное распределение, сосредоточенное на положительной полуоси, с плотностью

$$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},$$

где $\alpha, \lambda > 0$ - параметры; показательное распределение с параметром $\lambda > 0$ - это одномерное абсолютно непрерывное распределение, сосредоточенное на положительной полуоси, с плотностью $p(x) = \lambda e^{-\lambda x}$. Как видим, последнее распределение есть частный случай первого при $\alpha = 1$.)

Т.е. $X_1^2 + X_2^2$ имеет функцию распределения $F(x) = 1 - \exp(-x/2), x \geq 0, F(x) = 0, x < 0$. Отсюда следует, что $\exp(-(X_1^2 + X_2^2)/2)$ имеет равномерное на $[0,1]$ распределение.

(Доказательство: при $x > 0$ $\mathbf{P}\{\exp(-(X_1^2+X_2^2)/2) \le x\} = \mathbf{P}\{X_1^2+X_2^2 \ge -2\ln x\} = 1 - F(-2\ln x)$; при $x \ge 1 - 2\ln x \le 0, 1 - F(-2\ln x) = 1$; при $x < 1 - 2\ln x \ge 0, 1 - F(-2\ln x) = x$; если же $x \le 0$, то, очевидно, $\mathbf{P}\{\exp(-(X_1^2+X_2^2)/2) \le x\} = 0$.)

Если нам известны $a = \exp(-(X_1^2+X_2^2)/2), b = \arg(X_1/X_2)/2\pi$, то мы можем восстановить значения $\Delta = X_1^2 + X_2^2 = -2\ln a$, $\varphi = \arg(X_1/X_2) = 2b\pi$, и тогда $X_1 = \sqrt{\Delta}\cos\varphi, X_2 = \sqrt{\Delta}\sin\varphi$.

На практике поступают так: с помощью датчика случайных чисел получают выборки $(a_1, \ldots, a_n)$,
$(b_1, \ldots, b_n)$ из равномерного на $[0, 1]$ распределения, а по ним восстанавливают $(X_{11}, \ldots, X_{1n})$, $(X_{21}, \ldots, X_{2n})$ - выборки из распределения $\mathcal{N}(0, 1)$. (Т.е. указанным выше способом восстанавливают X_{1i}, X_{2i} по a_i, b_i, $i = \overline{1, n}$.)

Задачи и упражнения к главе 1

Упражнение 1.1. Представление нормального вектора $\vec{\xi}$ (см. формулу (1.1)) в форме $\vec{\xi} = \vec{a} + B\vec{\eta}$, где B - постоянная (неслучайная) матрица размера $n \times p$ и $\vec{a} \in \mathbb{R}^n$ - постоянный (неслучайный) вектор, $\vec{\eta}$ - p-мерный стандартный нормальный вектор, не единственно.

Упражнение 1.2. Обосновать следующие свойства операции взятия вектора средних значений (всюду здесь и в упражнении 1.3 мы рассматриваем произвольные случайные вектора, а не обязательно гауссовские):

1. $\mathbf{E}(\lambda_1\vec{X}_1 + \lambda_2\vec{X}_2) = \lambda_1\mathbf{E}\vec{X}_1 + \lambda_2\mathbf{E}\vec{X}_2$, где $\lambda_1, \lambda_2 \in \mathbb{R}$ - скаляры,
$\vec{X}_1, \vec{X}_2$ - случайные n-мерные векторы, предполагается, что у $\vec{X}_1, \vec{X}_2$ есть математические ожидания;

2 $\mathbf{E}\vec{b} = \vec{b}$;

3. $\mathbf{E}(A\vec{X} + \vec{b}) = A\mathbf{E}\vec{X} + \vec{b}$, если A - постоянная (неслучайная) матрица размера $p \times n$;

4. $\mathbf{E}(\vec{X}^T A + \vec{b}) = (\mathbf{E}\vec{X})^T A + \vec{b}$, если A - постоянная (неслучайная) матрица размера $n \times p$;

5. $\mathbf{E}(\vec{c}^T\vec{X}) = \vec{c}^T\mathbf{E}\vec{X}$,

где в пп.2-5 $\vec{b} \in \mathbb{R}^p, \vec{c} \in \mathbb{R}^n$ - постоянные (неслучайные) вектора, $\vec{X}$ - случайный вектор размерности n; в пп. 3-5 предполагается, что $\mathbf{E}\vec{X}$ существует.

Обосновать также следующие свойства операции взятия математического ожидания случайной матрицы:

6. $\mathbf{E}(AX + BY) = A\mathbf{E}X + B\mathbf{E}Y$;

7. $\mathbf{E}(XA + YB) = \mathbf{E}XA + \mathbf{E}YB$;

8. $\mathbf{E}A = A$;

9. $\mathbf{E}X^T = (\mathbf{E}X)^T$;

10. $\mathbf{E}(\alpha X + \beta Y) = \alpha\mathbf{E}X + \beta\mathbf{E}Y$,

где X, Y - случайные, а A, B - постоянные (неслучайные) матрицы, и в пп. 6-7 предполагается, что их размеры согласованы (так, чтобы были выполнимы операции сложения и умножения), а $\alpha, \beta \in \mathbb{R}$ - постоянные числа (скаляры). В пп. 6, 7, 10 предполагается, что у матриц X, Y существуют математические ожидания, в п.9 это предполагается только для матрицы X.

Также доказать следующее свойство:

11. $\mathbf{E}(\vec{z}^T\vec{\mu})^2 = \vec{z}^T\mathbf{E}(\vec{\mu}\vec{\mu}^T)\vec{z}$,

где $\vec{z} \in \mathbb{R}^n$ - постоянный (неслучайный) вектор, $\vec{\mu} = (\mu_1, \ldots, \mu_n)$ - случайный n-мерный вектор, причем предполагается, что $\mathbf{E}\mu_i^2$ существует и конечно при всех $i = \overline{1, n}$ или, эквивалентно, у матрицы $\vec{\mu}\vec{\mu}^T$ существует математическое ожидание. (Почему эти два условия эквивалентны?)

Упражнение 1.3. Доказать следующие свойства матриц ковариаций случайных векторов:

1. $\operatorname{Var}\vec{X} = \mathbf{E}(\vec{X}\vec{X}^T) - \mathbf{E}\vec{X}\mathbf{E}\vec{X}^T$;

2. $\operatorname{Var}(\lambda\vec{X} + \vec{b}) = \lambda^2 \operatorname{Var}\vec{X}$;

3. $\operatorname{Var}(A\vec{X} + \vec{b}) = A \operatorname{Var}\vec{X} A^T$, в частности, $\operatorname{Var}(\vec{c}^T\vec{X}) = \vec{c}^T \operatorname{Var}\vec{X}\vec{c}$;

4. $\operatorname{Var}\vec{X}$ - симметричная неотрицательно определенная матрица (свойство неотрицательной определенности для симметричной матрицы M размера $n \times n$ означает, как говорилось выше, следующее: $\forall \vec{z} \in \mathbb{R}^n \vec{z}^T M\vec{z} \geq 0$),

где в этих четырех пунктах $\vec{X}$ - случайный вектор (не обязательно гауссовский) размерности n, $\lambda \in \mathbb{R}$ - число, A - постоянная матрица размера $m \times n$, $\vec{b} \in \mathbb{R}^m, \vec{c} \in \mathbb{R}^n$ - постоянные векторы, и всюду предполагается, что матрица ковариаций у случайного вектора $\vec{X}$ существует. (А в п.1 $\mathbf{E}\vec{X}\vec{X}^T$ - математическое ожидание случайной матрицы XX^T.)

5. Если $\vec{X}_k, k = \overline{1,n}$ - независимые случайные векторы одной и той же размерности, имеющие ковариационные матрицы $\operatorname{Var}\vec{X}_k, k = \overline{1,n}$, то $\sum_{k=1}^{n}\vec{X}_k$ также имеет матрицу ковариаций, и она равна $\sum_{k=1}^{n}\operatorname{Var}\vec{X}_k$.

Упражнение 1.4. Доказать, что случайный вектор нормально распределен, если и только если любая линейная комбинация его компонент нормально распределена.

Упражнение 1.5. 1. Доказать, что для неотрицательно определенной симметричной квадратной матрицы A порядка n существует и единственен **квадратный корень**, т.е. такая неотрицательно определенная симметричная квадратная матрица B того же порядка n, что $B^2 = A$.

2. Доказать, что если A - неотрицательно определенная симметричная квадратная матрица порядка n, то ее невырожденность эквивалентна ее положительной определенности.

3. Доказать, что квадратный корень из неотрицательно определенной симметричной матрицы невырожден, если и только если сама эта матрица невырождена.

Упражнение 1.6. Если случайный вектор $\vec{X} = (X_1, \ldots, X_n)$ - нормальный, то любой случайный вектор, составленный из его компонент, т.е. вектор вида $(X_{i_1}, \ldots, X_{i_k})$, где $i_1, \ldots, i_k \in \{1, \ldots, n\}$ также имеет нормальное распределение.

В частности, компоненты этого нормального вектора - нормальные случайные величины: если $\vec{X} \backsim \mathcal{N}_n(\vec{a}, Q)$, где $\vec{a} = (a_1, \ldots, a_n)^T, q_{ij} - ij-$й элемент матрицы ковариаций Q, $i, j = \overline{1,n}$, то $X_i \backsim \mathcal{N}_n(a_i, q_{ii})$ при $i = \overline{1,n}$.

Привести пример случайного вектора $\vec{X} = (X_1, \ldots, X_n)$, не являющегося нормальным, для которого при всех $k < n$, $i_1, \ldots, i_k \in$
$\in \{1, \ldots, n\}$ вектор $(X_{i_1}, \ldots, X_{i_k})$ имеет нормальное распределение.

Упражнение 1.7. Доказать самостоятельно свойства 1 - 3, 5, 6 характеристических функций.

(*Подсказка:* при доказательстве свойств 5, 6 воспользоваться свойством 4.)

Упражнение 1.8. Доказать эквивалентность определений 1.15 и 1.16.

(**Указание:** при доказательстве того, что из определения 1.16 следует определение 1.15, воспользоваться **теоремой Каратеодори:**

если $\Omega \neq \varnothing$ - некоторое множество, $\mathcal{A}$ - алгебра его подмножеств, $\mu_0 - \sigma$ - конечная мера на $\mathcal{A}$, то ее можно единственным образом продолжить до меры на $\sigma(\mathcal{A})$ - на $\sigma-$ алгебре, порожденной $\mathcal{A}$.

Положить в рассматриваемом конкретном случае $\Omega = \mathbb{R}^n, \mathcal{A} = \mathcal{A}(\mathbb{R}^n), \sigma(\mathcal{A}) = \mathfrak{B}^n$.)

Упражнение 1.9. Если $X_m, Y_m, m = \overline{1, N}$ - множества (произвольные), то

$$(\bigcup_{m=1}^{N} X_i) \triangle (\bigcup_{m=1}^{N} Y_i) \subset \bigcup_{m=1}^{N} (X_i \triangle Y_i).$$

Упражнение 1.10. Подробно провести рассуждения в лемме 1.17.

Упражнение 1.11. Вывести лемму 1.18. из леммы 1.17.

Упражнение 1.11. Доказать утверждение 1.21.

Упражнение 1.12. Доказать утверждение 1.22.

Упражнение 1.12. Доказать утверждение 1.19. с помощью леммы 1.18. (**Указание:** в случае невырожденной Σ положить $D = G = \mathbb{R}^n, \alpha(t) = \Sigma^{1/2}t + a$. А если Σ вырождена, то положить $C\Sigma C^T = \Lambda$ - диагональная матрица с собственными значениями матрицы Σ на главной диагонали, $\vec{\zeta} := C\vec{\xi}$)

Упражнение 1.13. Проверить следующие свойства матриц ковариации:

1) $\operatorname{cov}(\vec{X}, \vec{Y}) = \operatorname{cov}(\vec{Y}, \vec{X})^T$;

2) $\operatorname{cov}(A\vec{X}, \vec{Y}) = A \operatorname{cov}(\vec{X}, \vec{Y})$;

3) $\operatorname{cov}(\vec{X}, B\vec{Y}) = \operatorname{cov}(\vec{X}, \vec{Y})B^T$;

4) $\operatorname{cov}(\vec{X} + \vec{a}, \vec{Y} + \vec{b}) = \operatorname{cov}(\vec{X}, \vec{Y})$;

5) $\operatorname{cov}(\alpha\vec{X}_1 + \beta\vec{X}_2, \vec{Y}) = \alpha \operatorname{cov}(\vec{X}_1, \vec{Y}) + \beta \operatorname{cov}(\vec{X}_2, \vec{Y})$;

6) $\operatorname{cov}(\vec{X}, \alpha\vec{Y}_1 + \beta\vec{Y}_2) = \alpha \operatorname{cov}(\vec{X}, \vec{Y}_1) + \beta \operatorname{cov}(\vec{X}, \vec{Y}_2)$;

7) $\operatorname{cov}(\vec{X}, \vec{X}) = \operatorname{Var} \vec{X}$,

где $\vec{X}, \vec{X}_1, \vec{X}_2, \vec{Y}, \vec{Y}_1, \vec{Y}_2$ - случайные вектора (не обязательно нормальные), A, B - постоянные (неслучайные) матрицы, $\vec{a}, \vec{b}$ - постоянные (неслучайные) вектора, $\alpha, \beta \in \mathbb{R}$ - постоянные числа (скаляры).

Упражнение 1.14. Доказать утверждение 1.21.

Упражнение 1.15. Доказать утверждение 1.22.

Упражнение 1.16. Доказать утверждение 1.23. с помощью формулы (1.4) для плотности нормальных векторов. (См. замечание 1.25.)

Упражнение 1.17. Доказать утверждение 1.26.

Упражнение 1.18. Доказать свойства 1 - 4 условного математического ожидания, указанные в п.5.

Упражнение 1.21. Доказать, что $\arg(X_1, X_2)$ - случайная величина, равномерно распределенная на отрезке $[0, 2\pi]$.

Упражнение 1.22. Доказать, что $\chi^2(n)$ и $\Gamma(n/2, 1/2)$ - одно и то же распределение.

(Подсказка: предварительно обосновать, что если случайные величины $X_i \backsim \Gamma(\alpha_i, \lambda), i = \overline{1, p}$ независимы, $\alpha_i > 0, i = \overline{1, p}, \lambda > 0$, то $\sum_{i=1}^{p} X_i \backsim \Gamma(\sum_{i=1}^{p} \alpha_i, \lambda)$.

Упражнение 1.23. Пусть A - симметричная неотрицательно определенная матрица порядка p, $\lambda_i, i = \overline{1, p}$ - ее собственные значения (не обязательно попарно различные), u_i - собственный вектор, соответствующий собственному значению $\lambda_i, i = \overline{1, p}$. Тогда матрицу A можно представить в виде $A = \sum_{i=1}^{p} \lambda_i u_i u_i^T$, и тогда $A^{1/2} = \sum_{i=1}^{p} \lambda_i^{1/2} u_i u_i^T$. Это - еще один, эквивалентный способ определения квадратного корня из неотрицательно определенной матрицы A.

Упражнение 1.24. 1) Пусть $\vec{X}_i \backsim \mathcal{N}(\vec{a}_i, \Sigma_i), i = \overline{1, n}$ имеют одинаковую размерность и независимы. Тогда $\sum_{i=1}^{n} \vec{X}_i \backsim \mathcal{N}(\sum_{i=1}^{n} \vec{a}_i, \sum_{i=1}^{n} \Sigma_i)$. (Подсказка: при доказательстве воспользоваться методом характеристических функций.)

2) Если $\vec{X} \backsim \mathcal{N}(\vec{a}, \Sigma)$, $\lambda \in \mathbb{R}$, то $\lambda\vec{X} \backsim \mathcal{N}(\lambda\vec{a}, \lambda^2\Sigma)$.

Задача 1.1. Рассмотрим двумерный гауссовский вектор $X = (X_1, X_2)^T$.

Положим $\mu_1 = \mathbf{E}X_1, \mu_2 = \mathbf{E}X_2, \sigma_1^2 = \operatorname{Var} X_1, \sigma_2^2 = \operatorname{Var} X_2, \rho = \operatorname{corr}(X_1, X_2)$. Пусть $\sigma_1 > 0, \sigma_2 > 0$.

1. Указать формулу для плотности случайного вектора

$$\left(\frac{X_1-\mu_1}{\sigma_1}, \frac{X_2-\mu_2}{\sigma_2}\right)^T.$$

2. Как выглядят линии уровня этой плотности при различных ρ, в частности, при $\rho=0$, $\rho\to 1$ $\rho\to -1$?

Задача 1.2. (Обобщение задачи 1.19. и результатов п.7) Пусть $X \backsim \mathcal{N}_n(0,I)$, т.е. случайный вектор X имеет n-мерное стандартное нормальное распределение. С вероятностью 1 он не равен нулю, а все ненулевые вектора $\vec{a}\in\mathbb{R}^n$ однозначно задаются точкой ψ на единичной сфере (через которую проходит луч, выходящий из начала координат и содержащий вектор $\vec{a}$) и своей длиной r. Т.е. случайный вектор X порождает еще один случайный вектор ψ, с вероятностью 1 принадлежащий единичной сфере в $\mathbb{R}^n$, и положительную случайную величину r. Доказать:

1. $r^2 \backsim \chi^2(n-1)$;

2. ψ равномерно распределен на единичной сфере в $\mathbb{R}^n$;

3. ψ, r независимы.

Задача 1.3. Пусть $X \backsim \mathcal{N}_p(a,\Sigma)$. Найти условное распределение $\mathrm{Law}(X|AX)$, где A - заданная матрица. В частности, найти условное распределение X при условии $\sum\limits_{k=1}^{p} X_k = 0$.

Задача 1.4. Пусть $\{W(t)\}_{t\in[0,1]}$ - винеровский процесс. (Напомним, что это понятие означает, что это - гауссовский процесс с $\mathbf{E}W(t)=0$, $\mathrm{cov}(W(s),W(t)) = \min(s,t)$.) Тогда при наложении условия $W(1)=0$ мы получаем *броуновский мост*, т.е. такой гауссовский процесс на $[0,1]$ (обозначаемый так: $\{B^0_t\}_{t\in[0,1]}$), что $\mathbf{E}B^0_t = 0, \mathrm{cov}(B^0_s, B^0_t) = \min(s,t) - st$. Это означает, что при всех $t_1,\dots,t_n\in[0,1]$ $\mathrm{Law}(W(t_1),\dots,W(t_n)|W(1 0) = \mathrm{Law}(B^{t_1}_0,\dots,B^{t_n}_0)$.

Задача 1.5. (M. Bilodeau, D. Bremer - 1999). Пусть X - n-мерный случайный вектор, $\mathbf{E}X=a, \mathrm{Var}\,X=\Sigma$. Показать, что :

$$\min_{a\in\mathbb{R}^n}\mathbf{E}|X-a|^2 = \mathrm{tr}\,\Sigma,\ \min_{a\in\mathbb{R}^n}\mathbf{E}(X-a)(X-a)^T=\Sigma$$

и оба эти минимума достигаются при $a=\mu$.

Задача 1.6. (M. Bilodeau, D. Bremer - 1999). Пусть X - нормальный $n+m$-мерный вектор, случайный вектор X_1 составлен из его первых n компонент, а случайный вектор X_2 - из последних m компонент. Пусть, далее, $\mathbf{E}X_1=\mu_1, \mathbf{E}X_2=\mu_2, \mathrm{cov}(X_i,X_j) = \Sigma_{ij}$, т.е. $\mathrm{Var}\,X_1=\Sigma_1, \mathrm{Var}\,X_2=\Sigma_2, \Sigma^T_{21}=\Sigma_{12}$ по задаче 1.14. Показать, что

$$\min_{C,d}\mathbf{E}|X_1-(CX_2+d)|^2 = \mathrm{tr}\,\Sigma_{12},\ \min_{C,d}\mathbf{E}(X_1-(CX_2+d))(X_1-(CX_2+d))^T=\Sigma_{12},$$

где минимум берется по всевозможным матрицам C размера $n\times m$ и по всевозможным векторам $d\in\mathbb{R}^n$. При этом оба минимума достигаются при $C=\Sigma_{12}\Sigma^{-1}_{22}, d = \mu_1-\Sigma_{12}\Sigma^{-1}_{22}\mu_2$.

Литература. [1] Т.Андерсон. Введение в многомерный статистический анализ /Пер. с англ. - М., Физматгиз, 1963.

[2] А.Н. Ширяев. Вероятность , М., Наука, 1989.

[3] В.А. Ильин, Э.Г. Позняк. Основы математического анализа, ч.2, М., Наука, 1973.

[4] П.Л. Ульянов, А.Н. Бахвалов, М.И. Дьяченко, К.С. Казарян, П. Сифуэнтес. Действительный анализ в задачах, М., Физматлит, 2005.

Глава 2. Статистические модели для гауссовских векторов

1. Напоминание основных понятий статистики.

Определение 2.1. Базовой моделью математической статистики является **статистическое пространство** - тройка $(\mathfrak{X}, \mathfrak{F}, \{P_\theta\}_{\theta\in\Theta})$, где $\mathfrak{X} \neq \varnothing$ называется **выборочным пространством** или **пространством наблюдений** (а его элементы - **наблюдениями**), $\mathfrak{F}$ - σ-алгебра над $\mathfrak{X}$, а $\{P_\theta\}_{\theta\in\Theta}$ - совокупность вероятностных мер на $\mathfrak{F}$. Мы считаем, что какая - либо из этих мер - истинная, т.е. реально заданная на $\mathfrak{F}$, но нам неизвестно, какая именно эта мера. (Но заведомо известно, что она содержится в семействе $\{P_\theta\}_{\theta\in\Theta}$).

Задача статистики - по данному наблюдению $X \in \mathfrak{X}$ сделать заключение о том, какова же истинная мера, заданная на $\mathfrak{F}$. Для этого достаточно сделать выводы об истинном значении параметра θ, т.е. том значении, которое соответствует истинной мере. В частности, можно попытаться построить оценки этого параметра. Пусть далее $\Theta \subset \mathbb{R}^m$.

Определение 2.2. Оценкой параметра θ называется любая случайная величина δ (если $m = 1$) или случайный m-мерный вектор δ (в случае $m > 1$), при всех $X \in \mathfrak{X}$ приближенно равный истинному значению параметра θ.

Такое определение, разумеется, не может считаться математически строгим: слова "приближенно равна" необходимо как-то формализовать. Вот один из способов такой формализации:

Определение 2.3. Несмещенной оценкой параметра θ называется случайная величина δ (если $m = 1$) или случайный m-мерный вектор (в случае $m > 1$) $\delta : \mathfrak{X} \to \mathbb{R}^m$, для которой при всех $\theta \in \Theta$

$$\mathbf{E}_\theta \delta(X) = \theta,$$

где индекс θ у знака математического ожидания означает, что оно берется по мере P_θ, т.е. равно $\int_{\mathfrak{X}} \delta(X) P_\theta(dX)$ (интеграл понимается в смысле Лебега).

Определение 2.4. Наилучшей несмещенной оценкой параметра θ называется такая несмещенная оценка δ_0 этого параметра, что для любой другой несмещенной оценки δ этого параметра при всех $\theta \in \Theta$ выполнено неравенство

$$\mathbf{E}_\theta(\delta_0(X) - \theta)^2 \leq \mathbf{E}_\theta(\delta(X) - \theta)^2,$$

если $m = 1$ (т.е. δ, δ_0 - случайные величины), и

$$\mathbf{E}_\theta(\delta_0(X) - \theta)(\delta_0(X) - \theta)^T \leq \mathbf{E}_\theta(\delta(X) - \theta)(\delta(X) - \theta)^T, \tag{1}$$

если $m > 1$ (т.е. δ, δ_0 - случайные вектора). Можно записать эти два неравенства и так: $\operatorname{Var}\delta_0 \leq \operatorname{Var}\delta$, т.к. в силу несмещенности δ, δ_0 имеем: $\mathbf{E}_\theta\delta_0 = \mathbf{E}_\theta\delta = \theta$. (Матрицы в левой и правой частях неравенства (1) симметричны, ведь это - матрицы ковариаций - см. упражнение 1.3.4, т.е. их можно сравнивать; как именно - см. гл.1, п.6)

Определение 2.5. Статистика - это любая случайная величина (или случайный вектор), определенный на $\mathfrak{X}$. Статистика T называется **достаточной**, если для всех $A \in \mathfrak{F}$ $P_\theta(A|T)$ не зависит от θ. При этом, напомним, **условная вероятность** $\mathbf{P}(A|\xi)$ **измеримого множества** A относительно случайной величины или случайного вектора ξ - это, по определению, $\mathbf{E}(I_A|\xi)$. Часто говорят и так: T - достаточная статистика, если условное распределение $X \in \mathfrak{X}$ при заданном $T(X)$ не зависит от θ.

Теорема 2.6. (Blackwell, Rao, 1947 - 1949.) Если δ - несмещенная оценка θ, а T - достаточная статистика, то $\delta_0 := \mathbf{E}_\theta(\delta|T)$ - также несмещенная оценка θ, и при этом она *не хуже* δ - это означает, что для всех $\theta \in \Theta$ $\operatorname{Var}_\theta \delta_0 \leq \operatorname{Var}_\theta \delta$.

Кроме того, для любых $\theta \in \Theta$ $\operatorname{Var}_\theta \delta_0 = \operatorname{Var}_\theta \delta$ тогда и только тогда, когда $\forall\theta \in \Theta$ $\mathbf{P}_\theta\{\delta_0 = \delta\} = 1$.

Индексы θ при знаках $\mathbf{E}$, Var означают, что соответствующие математические ожидания и дисперсии - или матрицы ковариаций - вычисляются по вероятностной мере P_θ. (См. упражнение 2.1.)

Таким образом, наличие достаточной статистики позволяет улучшать несмещенные оценки.

Заметим также, что $\delta_0 = \mathbf{E}_\theta(\delta|T)$ представляет собой некоторую борелевскую функцию от T. (См. свойства условного математического ожидания, гл.1, п.5.) Значит, при поиске наилучшей несмещенной оценки - т.е. несмещенной оценки с мимимальной дисперсией или матрицей ковариаций - можно ограничиться лишь борелевскими функциями от достаточной статистики (если изначально мы ей располагаем).

Определение 2.7. Достаточная статистика $T : \mathfrak{X} \to \mathbb{R}^k$ называется **полной**, если для любой борелевской функции $f : \mathbb{R}^k \to \mathbb{R}$ такой, что у случайной величины $f(T)$ существует конечное математическое ожидание, из $\mathbf{E}_\theta f(T) = 0$ следует $f(T) = 0$ почти наверное.

Теорема 2.8. (Леман, Шефаре, 1955.) Если $T : \mathfrak{X} \to \mathbb{R}^k$ - полная достаточная статистика и $f(T)$ - несмещенная оценка θ, где $f : \mathbb{R}^k \to \mathbb{R}^m$ - борелевская функция, эта несмещенная оценка является наилучшей. (См. упражнение 2.2.)

Мы будем рассматривать частный случай вышеописанной общей базовой модели: $\mathfrak{X} = \mathbb{R}^{np}$, $\mathfrak{F} = \mathfrak{B}^{np}$, $\Theta = \{(a, Q) : a \in \mathbb{R}^p, Q$ - положительно определенная матрица порядка $p\}$, а для $\theta = (a, Q) \in \Theta$

$$P_\theta = \underbrace{\mathcal{N}_p(a, Q) \times \ldots \times \mathcal{N}_p(a, Q)}_{n}.$$

Это означает, что P_θ - прямое произведение n экземпляров вероятностной меры Q.

В нашем случае такое прямое произведение n экземпляров меры Q - это распределение случайного $np-$мерного вектора, являющегося выборкой размера n из распределения $N_p(a, Q)$. (Эта выборка есть n $p-$мерных случайных векторов, и ее можно записать как $np-$мерный вектор.)

Наша задача - искать, какое именно это распределение, т.е. какие параметры a, Q оно имеет.

2. Достаточная статистика $(\overline{X}, S)$.

Утверждение 2.9. Если $(\vec{X}_1, \ldots, \vec{X}_n)$ - выборка из $\mathcal{N}_p(\vec{a}, Q)$, $\vec{a} \in \mathbb{R}^p$ - постоянный вектор, Q - невырожденная матрица порядка p, то $(\bar{X}, S)$, где

$$\overline{X} = \frac{\vec{X}_1 + \ldots + \vec{X}_n}{n}, \; S = \frac{1}{n-1}\sum_{k=1}^{n}(\vec{X}_k - \overline{X})(\vec{X}_k - \overline{X})^T -$$

достаточная статистика.

Доказательство. Мы будем использовать **теорему о факторизации**:

$T(\xi)$ - достаточная статистика для семейства вероятностных мер, заданных плотностями $p(x, \theta)$, если и только если найдутся функции g, h, для которых $p(x, \theta) = g(T(x), \theta)h(x)$ при всех x, θ.

В данном случае совместная плотность распределения векторов
$\vec{X}_1, \ldots, \vec{X}_n$ равна

$$p(\vec{x}_1, \ldots, \vec{x}_n) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det Q}} \exp(-\frac{1}{2}(\vec{x}_i - \vec{a})^T Q^{-1}(\vec{x}_i - \vec{a})) =$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{pn} \left(\frac{1}{\sqrt{\det Q}}\right)^{n} \exp(-\frac{1}{2}\sum_{i=1}^{n}(\vec{x}_i-\vec{a})^T Q^{-1}(\vec{x}_i-\vec{a}))$$

Но в показателе экспоненты стоит $\operatorname{tr} C$, где $C = X^T Q^{-1} X, X$ - матрица размера $p \times n$, k - й столбец которой равен $\vec{X}_k - \vec{a}$, т.е. jk-й элемент которой равен $X_{kj} - a_j, j,k = \overline{1,n}$, где мы полагаем $\vec{X}_k = (X_{k1},\ldots,X_{kn}), k = \overline{1,n}$. Кроме того, пусть $\overline{X} = (\overline{X}_1,\ldots,\overline{X}_n)$. Заметим, что $\operatorname{tr} C = \operatorname{tr}(Q^{-1}XX^T)$ (мы применили свойство следа матрицы: $\operatorname{tr} AB = \operatorname{tr} BA$.)

Но по правилу умножения матриц kl-й элемент матрицы XX^T равен

$$\sum_{j=1}^{n}(X_{jk}-a_k)(X_{jl}-a_l) = \sum_{j=1}^{n}((X_{jk}-\overline{X}_k)+(\overline{X}_k-a_k))((X_{jl}-\overline{X}_l)+(\overline{X}_l-a_l)) =$$

$$= \sum_{j=1}^{n}(X_{jk}-\overline{X}_k)(X_{jl}-\overline{X}_l) + \sum_{j=1}^{n}(\overline{X}_k-a_k)(X_{jl}-\overline{X}_l) + \sum_{j=1}^{n}(X_{jk}-\overline{X}_k)(\overline{X}_l-a_l) +$$

$$+ \sum_{j=1}^{n}(\overline{X}_k-a_k)(\overline{X}_l-a_l) = (n-1)s_{kl} + (\overline{X}_k-a_k)\sum_{j=1}^{n}(X_{jl}-\overline{X}_l) +$$

$$+(\overline{X}_l-a_l)\sum_{j=1}^{n}(X_{jk}-\overline{X}_k) + n(\overline{X}_k-a_k)(\overline{X}_l-a_l) = (n-1)s_{kl} + n(\overline{X}_k-a_k)(\overline{X}_l-a_l),$$

где s_{kl} - это kl-й элемент матрицы S; мы использовали то, что $\sum_{j=1}^{n}(X_{jk}-\overline{X}_k) = \sum_{j=1}^{n} X_{jk} - n\overline{X}_k = n\overline{X}_k - n\overline{X}_k = 0$, аналогично $\sum_{j=1}^{n}(X_{jl}-\overline{X}_l) = 0$ при всех $k,l = \overline{1,n}$. Т.е. $XX^T = (n-1)S + n(\overline{X}-\vec{a})(\overline{X}-\vec{a})^T$. Значит, в показателе экспоненты в формуле для плотности стоит число

$$-\frac{1}{2}\operatorname{tr}((n-1)S + n(\overline{X}-\vec{a})(\overline{X}-\vec{a})^T),$$

и плотность $p(\vec{x}_1,\ldots,\vec{x}_n)$ имеет вид функции от $\vec{a}, Q, \overline{X}, S$, т.е по теореме факторизации $(\bar{X}, S)$ - достаточная статистика. Утверждение доказано.

Замечание 2.10.Матрица S называется **выборочной матрицей ковариаций**, а $\overline{X}$ - **эмпирическим средним.** Нетрудно видеть, что это - многомерные аналоги обычных характеристик выборки $(X_1,\ldots,X_n)$ из одномерного распределения:

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k,\ \hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2.$$

Утверждение 2.11. $\mathbf{E}\overline{X} = a, \mathbf{E}S = Q$.

Доказательство. Докажем первое равенство: $\mathbf{E}\overline{X} = \mathbf{E}(\frac{1}{n}\sum_{k=1}^{n}\vec{X}_k) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\vec{X}_k = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\vec{a} = n\vec{a}/n = \vec{a}$. Теперь докажем второе. Необходимо доказать, что $\mathbf{E}s_{ij} = q_{ij}$ (где $s_{ij},\ q_{ij}$ - ij-е элементы матриц S, Q соответственно), при всех $i,j = \overline{1,n}$. Но $ij-$й элемент матрицы S равен

$$\frac{1}{n-1}\sum_{k=1}^{n}(X_{ki}-\overline{X}_i)(X_{kj}-\overline{X}_j) = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki}X_{kj} - X_{ki}\overline{X}_j - X_{kj}\overline{X}_i + \overline{X}_i\overline{X}_j) =$$

$$= \frac{1}{n-1}(\sum_{k=1}^{n} X_{ki}X_{kj} - n\overline{X}_i\overline{X}_j - \overline{X}_i n\overline{X}_j + n\overline{X}_i\overline{X}_j) = \frac{1}{n-1}(\sum_{k=1}^{n} X_{ki}X_{kj} - n\overline{X}_i\overline{X}_j).$$

Вычислим его математическое ожидание.

Заметим, что

$$\mathbf{E}\sum_{k=1}^{n} X_{ki}X_{kj} = \sum_{k=1}^{n} \mathbf{E}X_{ki}X_{kj} = (\mathrm{cov}(X_{1i}, X_{1j}) + \mathbf{E}X_{1i}\mathbf{E}X_{1j}) = n(q_{ij} + a_i a_j).$$

Далее, $\mathbf{E}n\overline{X}_i\overline{X}_j = n(\mathrm{cov}(\overline{X}_i, \overline{X}_j) + \mathbf{E}\overline{X}_i\mathbf{E}\overline{X}_j)$. Но $\overline{X} \backsim \mathcal{N}(\vec{a}, Q/n)$, т.к.

$$\sum_{k=1}^{n} \vec{X}_k \backsim \mathcal{N}(n\vec{a}, nQ)$$

(мы использовали утверждения обоих пунктов упражнения 1.24). Значит,

$$\mathrm{cov}(\overline{X}_i, \overline{X}_j) = q_{ij}/n, \mathbf{E}\overline{X}_i = a_i, \mathbf{E}\overline{X}_j = a_j.$$

Отсюда заключаем, что $\mathbf{E}n\overline{X}_i\overline{X}_j = n(q_{ij}/n + a_i a_j) = q_{ij} + n a_i a_j$. Итак, $\mathbf{E}s_{ij} = \frac{1}{n-1}(n(q_{ij} + a_i a_j) - (q_{ij} + n a_i a_j)) = q_{ij}$. Утверждение доказано.

Итак, $\overline{X}, S$ - несмещенная оценка для $\vec{a}, Q$ и достаточная статистика. Можно показать, что это - полная статистика, тогда по теореме Лемана - Шефаре получим, что $\overline{X}, S$ - наилучшая несмещенная оценка $\vec{a}, Q$.

3. Независимость $\overline{X}, S$.

Утверждение 2.12. Если имеем $p = 1$, т.е. $X_i, i = \overline{1, n}$ - случайные величины, то $\overline{X} = (X_1 + \ldots + X_n)/n, \hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_i - \bar{X})^2$ независимы.

Доказательство. Пусть $X_1, \ldots, X_n \backsim \mathcal{N}(a, \sigma^2)$ независимы, $\vec{X} = (X_1, \ldots, X_n)^T \backsim \mathcal{N}(\vec{\mu}, \sigma^2 I)$, I - единичная матрица порядка n, $\vec{\mu} = (a, \ldots, a)^T$. Пусть D - ортогональная матрица порядка n с первой строкой $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$. (См. задачу 2.3.)

Т.к. $\vec{X} \backsim \mathcal{N}(\vec{\mu}, \sigma^2 I)$, то

$$\vec{Y} \backsim \mathcal{N}(D\vec{\mu}, D\sigma^2 D^T) = \mathcal{N}(D\vec{\mu}, \sigma^2 I)$$

по упражнению 1.3.3, откуда получаем, что компоненты вектора $\vec{Y}$ независимы. (Ведь $\mathrm{cov}(Y_k, Y_l)$ - это kl-й элемент в матрице $\mathrm{Var}\,\vec{Y} = \sigma^2 I$, т.е. 0 - применяем утверждение 1.13.) Значит, $Y_1, \sum_{k=2}^{n} Y_k^2$ независимы. Но $Y_1 = X_1/\sqrt{n} + \ldots + X_n/\sqrt{n} = \sqrt{n}\overline{X}$, т.к. первая строка матрицы D равна $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$. Отсюда $Y_1^2 = n\overline{X}^2$. Значит,

$$\sum_{k=2}^{n} Y_k^2 = |\vec{Y}|^2 - Y_1^2 = |D\vec{X}|^2 - n\overline{X}^2.$$

Т.к. D - ортогональная матрица, то $|D\vec{X}|^2 = |\vec{X}|^2$. Значит, $|D\vec{X}|^2 - n\overline{X}^2 = |\vec{X}|^2 - n\overline{X}^2$. Но

$$\hat{\sigma}^2 = \sum_{k=1}^{n}(X_k - \overline{X})^2 = \sum_{k=1}^{n}(X_k^2 - 2\overline{X}X_k + \overline{X}^2) =$$

$$= \sum_{k=1}^{n} X_k^2 - 2\overline{X}\sum_{k=1}^{n} X_k + \sum_{k=1}^{n}\overline{X}^2 = |\vec{X}|^2 - n\overline{X}^2.$$

Тем самым доказано, что $\hat{\sigma}^2 = \sum_{k=1}^{n} Y_k^2$ и $\overline{X} = Y_1$ независимы. Утверждение доказано.

Итак, независимость $\overline{X}, \hat{\sigma^2}$ в одномерном случае доказана. Как традиционно доказывается эта независимость в многомерном случае (т.е в случае, когда выборка производится из многомерного, а не из одномерного распределения)?

Лемма 2.13. Пусть $X_i, i = \overline{1,n}$ независимые гауссовские вектора, $X_i \backsim \mathcal{N}_p(\mu_i, Q), i = \overline{1,n}$. Пусть, далее, C - ортогональная матрица порядка n, $c_{\alpha\beta} - \alpha\beta$-й элемент матрицы C,

$$Y_\alpha = \sum_{\beta=1}^{n} c_{\alpha\beta} X_\beta$$

при $\alpha = \overline{1,n}$. Тогда:

(a) $(Y_1, \ldots, Y_n)$ - гауссовский вектор;

(b) $\mathbf{E}Y_\alpha = \sum\limits_{\beta=1}^{n} c_{\alpha\beta}\mathbf{E}X_\beta$;

(c) $\mathrm{cov}(Y_\alpha, Y_\beta) = \delta_{\alpha\beta}Q$ (следствие: при $\alpha \neq \beta Y_\alpha, Y_\beta$ некоррелированы, а, значит, в силу утверждения 1.10 и п.(a) данного утверждения, независимы);

(d) $\sum\limits_{i=1}^{n} X_i X_i^T = \sum\limits_{\alpha=1}^{n} X_\alpha X_\alpha^T$.

Доказательство: см. упражнение 2.4.

Теперь докажем основное

Утверждение 2.14. И в многомерном случае $\overline{X}$ и S тоже независимы.

Доказательство Пусть $X_i, i = \overline{1,n} \backsim N_p(a, Q)$ - компоненты данной выборки; они независимы. Тогда, если

$$Y_\alpha := \sum_{\beta=1}^{n} c_{\alpha\beta}(X_\beta - a), \alpha = \overline{1,n},$$

где C - указанная в лемме 2.13 ортогональная матрица с первой строкой $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ (см.задачу 2.1), то

$$Y_1 = \sum_{\beta=1}^{n} \frac{1}{\sqrt{n}}(X_\beta - a) = \frac{1}{\sqrt{n}}(n\overline{X} - na) = \sqrt{n}(\overline{X} - a).$$

Кроме того, $Y_1 Y_1^T + \sum_{\alpha=2}^{n} Y_\alpha Y_\alpha^T = \sum_{\alpha=1}^{n}(X_\alpha - a)(X_\alpha - a)^T$ и $Y_\alpha, \alpha = \overline{1,n}$, независимы. (Мы применили лемму 2.5, подставив $X_\alpha - a$ вместо X_α, $\alpha = \overline{1,n}$ в условие этой леммы.) Но

$$\sum_{\alpha=1}^{n}(X_\alpha - a)(X_\alpha - a)^T = \sum_{\alpha=1}^{n}((X_\alpha - \overline{X}) + (\overline{X} - a))((X_\alpha - \overline{X}) + (\overline{X} - a))^T =$$

$$= \sum_{\alpha=1}^{n}\big((X_\alpha - \overline{X})(X_\alpha - \overline{X})^T + (\overline{X} - a)(X_\alpha - \overline{X})^T + (X_\alpha - \overline{X})(\overline{X} - a)^T +$$

$$+(\overline{X} - a)(\overline{X} - a)^T\big) = \sum_{\alpha=1}^{n}(X_\alpha - \overline{X})(X_\alpha - \overline{X})^T + (\overline{X} - a)(\sum_{\alpha=1}^{n}(X_\alpha - \overline{X}))^T +$$

$$+\sum_{\alpha=1}^{n}(X_\alpha - \overline{X})(\overline{X} - a)^T + n(\overline{X} - a)(\overline{X} - a)^T =$$

$$= \sum_{\alpha=1}^{n}(X_\alpha - \overline{X})(X_\alpha - \overline{X})^T + n(\overline{X} - a)(\overline{X} - a)^T = (n-1)S + n(\overline{X} - a)(\overline{X} - a)^T.$$

Наконец, $Y_1 Y_1^T = \sqrt{n}(\overline{X} - a)(\sqrt{n}(\overline{X} - a))^T = n(\overline{X} - a)(\overline{X} - a)^T$. Значит, $\sum_{\alpha=2}^{n} Y_\alpha Y_\alpha^T = (n-1)S$,

$$S = \frac{1}{n-1}\sum_{\alpha=2}^{n} Y_\alpha Y_\alpha^T.$$

Т.к. $Y_\alpha, \alpha = \overline{1,n}$, независимы (см. выше), $\sqrt{n}(\overline{X} - a) = Y_1$, откуда $\overline{X} = Y_1/\sqrt{n} + a$, а S зависит только от $Y_\alpha, \alpha = \overline{2,n}$, то S и $\overline{X}$ независимы. Утверждение доказано.

Заметим, что $(n-1)S = \sum\limits_{\alpha=2}^{n} Y_\alpha Y_\alpha^T \backsim \mathcal{W}_p(n-1, Q)$, где

$$\mathcal{W}_p(m, Q) = \sum_{i=1}^{m} \xi_i \xi_i^T -$$

статистика Уишарта (Wishart) с m степенями свободы. (Предполагается, что $\xi_i \backsim \mathcal{N}(0, Q), i = \overline{1,m}$ независимы.)

4. Оценка максимального правдоподобия для параметров $\vec{a}, Q$.

Напомним понятие "оценка максимального правдоподобия".

Пусть $(\mathfrak{X}, \mathfrak{F}, \{P_\theta\}_{\theta \in \Theta})$ - рассмотренная в начале данной главы общая статистическая модель, и при этом $\mathfrak{X} = \mathbb{R}^m$. Пусть, далее, при всех $\theta \in \Theta$ P_θ - абсолютно непрерывная мера на $\mathbb{R}^m$, т.е. при любом $\theta \in \Theta$ P_θ имеет плотность; обозначим значение этой плотности в точке $x \in \mathbb{R}^m$ через $p(x, \theta)$. Положим $X \in \mathfrak{X}$, т.е. будем считать, что мы провели наблюдение и X - его результат.

Определение 2.15. Правдоподобием для данного наблюдения X называется функция из Θ в $\mathbb{R}$, сопоставляющая элементу $\theta \in \Theta$ значение $p(X, \theta)$.

Здравый смысл подсказывает, что "хорошей"оценкой для истинного значения θ является, по-видимому, такое θ, при котором достигается максимум $p(X, \theta)$.

Определение 2.16. Значение $\theta = \hat{\theta} \in \Theta$, при котором правдоподобие $p(X, \theta)$ достигает максимума, называется **оценкой максимального правдоподобия**.

Найдем эту оценку для рассматриваемой конкретной модели - для выборки из нормального многомерного распределения.

Утверждение 2.17.

$$\hat{Q} = S_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T, \hat{\vec{a}} = \overline{X} -$$

оценка максимального правдоподобия для неизвестных параметров $\vec{a}, Q$ распределения $\mathcal{N}_p(\vec{a}, Q)$ (по умолчанию считаем, что $Q > 0$) по известной выборке $(X_1, \ldots, X_n)$ из этого распределения.

Доказательство. Правдоподобие при данных значениях параметров $\vec{a}, Q$ - это плотность, если подставить $\vec{x}_i = \vec{X}_i, i = \overline{1,n}$. Необходимо показать, что при указанных в условии значениях $\vec{a}, Q$ это правдоподобие принимает наибольшее значение.

Но компоненты выборки независимы, т.е. если

$$p_0(\vec{x}; \vec{a}, Q) = \frac{1}{\sqrt{(2\pi)^p \det Q}} \exp(-\frac{1}{2}(\vec{x} - \vec{a})^T Q^{-1}(\vec{x} - \vec{a})) -$$

плотность $\mathcal{N}_p(\vec{a}, Q)$ (напомним, что по условию $Q > 0$, т.е. Q невырождена), то $p(\vec{x}_1, \ldots, \vec{x}_n; \vec{a}, Q) = p_0(\vec{x}_1; \vec{a}, Q) \cdot \ldots \cdot p_0(\vec{x}_n; \vec{a}, Q)$, т.е.

$$p(\vec{X}_1, \ldots, \vec{X}_n; \vec{a}, Q) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^p \det Q}} \exp(-\frac{1}{2}(\vec{X}_i - \vec{a})^T Q^{-1}(\vec{X}_i - \vec{a})) =$$

$$= \frac{1}{(\sqrt{2\pi})^{np}} \frac{1}{(\sqrt{\det Q})^n} \exp(-\frac{1}{2}\sum_{i=1}^{n}(\vec{X}_i - \vec{a})^T Q^{-1}(\vec{X}_i - \vec{a})).$$

Показатель полученной экспоненты равен

$$-\frac{1}{2} n(\overline{X} - \vec{a})^T Q^{-1}(\overline{X} - \vec{a}) + n\,\mathrm{tr}(Q^{-1} S_n).$$

(См. доказательство теоремы 2.9, формулу для показателя экспоненты; мы использовали также, что $(n-1)S = nS_n$.) Если бы $\hat{\vec{a}} \neq \overline{X}$, то

$$-\frac{1}{2}n(\overline{X}-\hat{\vec{a}})^T\hat{Q}^{-1}(\overline{X}-\hat{\vec{a}}) + n\operatorname{tr}(\hat{Q}^{-1}S_n) < n\operatorname{tr}(\hat{Q}^{-1}S_n)$$

и

$$p(\vec{X}_1,\ldots,\vec{X}_n;\hat{\vec{a}},\vec{Q}) < p(\vec{X}_1,\ldots,\vec{X}_n;\overline{X},\hat{Q}),$$

но это противоречит тому, что $\hat{\vec{a}},\hat{Q}$ - оценка наибольшего правдоподобия.

Т.е. $\hat{\vec{a}} = \overline{X}$. Теперь докажем, что $\hat{Q} = S_n$.

Заметим, что

$$\ln p(\vec{X}_1,\ldots,\vec{X}_n;\hat{\vec{a}},Q) = -\frac{n}{2}\ln\det Q - \frac{n}{2}\operatorname{tr}(Q^{-1}S_n) + \mathrm{const},$$

Пусть

$$q(Q) = -\frac{n}{2}\ln\det Q - \frac{n}{2}\operatorname{tr}(Q^{-1}S_n).$$

Это - функция, определенная на E - множестве всех положительно определенных симметричных матриц порядка p. Нам необходимо доказать, что при $Q = S_n$ у этого выражения достигается строгий максимум, т.е. для любой $Q \in E, Q \neq S_n$ $q(Q) < q(S_n)$. Для этого докажем, что при $T = I_p$ у выражения $\tilde{r}(T) := q(S_nT^{-1})$ достигается строгий максимум, т.е. если $T \in E, T \neq I_p$, то $r(T) < r(I)$. (Ведь $T \mapsto S_nT^{-1}$ - биекция $E \to E$ - см. упражнение 2.7.) Достаточно доказать, что у функции

$$r(T) := \frac{2}{n}\tilde{r}(T) + \ln\det S_n$$

в точке $T = I_p$ достигается строгий максимум. Преобразуем эту функцию:

$$r(T) = \frac{2}{n}(-\frac{n}{2}\ln\det(S_nT^{-1}) - \frac{n}{2}\operatorname{tr}((S_nT^{-1})^{-1}S_n)) + \ln\det S_n =$$

$$= -\ln\det(S_nT^{-1}) - \operatorname{tr}(TS_n^{-1}S_n) + \ln\det S_n = -\ln(\frac{\det S_n}{\det T}) - \operatorname{tr} T + \ln\det S_n =$$

$$= -(\ln\det S_n - \ln\det T) - \operatorname{tr} T + \ln\det S_n = \ln\det T - \operatorname{tr} T.$$

Пусть $T \in E, T \neq I_p$. Тогда найдется ортогональная матрица C порядка p, для которой $C^{-1}TC = C^TTC = \Lambda$, где Λ - диагональная матрица с собственными значениями $\lambda_i, i = \overline{1,p}$ матрицы T на главной диагонали. Но эти собственные значения положительны, т.к. матрица T положительно определена. (См. упражнение 2.8.) Далее, $\det T = \det(C^{-1}TC) = \det\Lambda = \lambda_1\lambda_2\ldots\lambda_p, \operatorname{tr} T = \operatorname{tr}(C^{-1}TC) = \operatorname{tr}\Lambda = \sum\limits_{i=1}^{p}\lambda_i$, т.к. матрицы T, Λ подобны, а у подобных матриц след и определитель равны.

Отсюда получаем: $r(T) = \ln(\lambda_1\lambda_2\ldots\lambda_p) - \sum\limits_{i=1}^{p}\lambda_i$.

Мы можем ограничиться рассмотрением случая, когда все числа $\lambda_i, i = \overline{1,p}$ равны. Действительно, если хотя бы два из этих чисел не равны, то по неравенству Коши для

$$\mu_1 = \mu_2 = \ldots = \mu_p := \frac{1}{n}\sum_{i=1}^{n}\lambda_i$$

выполнено следующее:

$$\sqrt[p]{\lambda_1\lambda_2\ldots\lambda_p} < \mu_1 = \sqrt[p]{\mu_1\mu_2\ldots\mu_p},$$

т.е.

$$\lambda_1\lambda_2\ldots\lambda_p < \mu_1\mu_2\ldots\mu_p,$$

а, кроме того,

$$\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \mu_i.$$

Отсюда имеем:

$$\ln(\lambda_1 \lambda_2 \dots \lambda_p) - \sum_{i=1}^{p} \lambda_i < \ln(\mu_1 \mu_2 \dots \mu_p) - \sum_{i=1}^{p} \mu_i.$$

Но все числа $\mu_i, i = \overline{1,p}$ равны.

Итак, пусть далее $\lambda_1 = \ldots = \lambda_p$. Тогда $\lambda_1 = \ldots = \lambda_p \neq 1$, иначе бы $\Lambda = I_p$, т.е. $C^{-1}TC = I_p, T = CI_pC^{-1} = I_p$, что противоречит предположению $T \neq I_p$. Значит, $r(T) = \ln(\lambda_1^p) - p\lambda_1 = p(\ln\lambda_1 - \lambda_1) < -p$, т.к. $\lambda_1 \neq 1$. (См. упражнение 2.9.) Однако $r(I_p)$ как раз равно $-p$: $r(I_p) = \ln\det I_p - \operatorname{tr} I_p = \ln 1 - p = -p$. Тем самым неравенство $r(T) < r(I_p)$ при $T \in E, T \neq I_p$ доказано, а вместе с ним - и исходное утверждение. ■

5. Скобка Тюрина для выборок из многомерных распределений.

Запишем выборку $(X_1, \ldots, X_n)$ из p-мерного распределения в виде матрицы $X = \|X_1|\ldots|X_n\|$ размера $p \times n$, столбцы которой - X_i, т.е. компоненты выборки.

Определение 2.18. Положим для двух таких матриц $X = \|X_1|\ldots|X_n\|, Y = \|Y_1|\ldots|Y_n\|$ $< X, Y > := \sum\limits_{\alpha=1}^{n} X_\alpha Y_\alpha^T = XY^T$. Т.е. $< X, Y >$ - квадратная матрица порядка p; она называется **скобкой Тюрина**. Если $p = 1$, то эта операция - обычное скалярное произведение n-мерных векторов
$(X_1, \ldots, X_n), (Y_1, \ldots, Y_n)$.

Утверждение 2.19. Скобка Тюрина обладает следующими свойствами (аналогичными свойствам обычного скалярного произведения):

1. $< X, Y > = < Y, X >^T$;
2. $< \alpha_1 X_1 + \alpha_2 X_2, Y > = \alpha_1 < X_1, Y > + \alpha_2 < X_2, Y >$;
3. $< X, \alpha_1 Y_1 + \alpha_2 Y_2 > = \alpha_1 < X, Y_1 > + \alpha_2 < X, Y_2 >$;
4. $< AX, Y > = A < X, Y >$;
5. $< X, AY > = < X, Y > A^T$;
6. $< X, X >$ неотрицательно определена для всех матриц X размера $p \times n$;
7. Если A ортогональна, то $< X, Y > = < XA, YA >$,

где всюду здесь X, Y - матрицы размера $p \times n$, A - квадратная матрица порядка p, $\alpha_1, \alpha_2 \in \mathbb{R}$. (См. упражнение 2.5.)

Пусть $(X_1, \ldots, X_n)$ - выборка из распределения $\mathcal{N}_p(a, Q)$, $X = \|X_1|\ldots|X_n\|$ - матрица размера $p \times n$. Тогда $\|X_1 - a|\ldots|X_n - a\| = X - ae$, $e = (1, \ldots, 1) - n$-мерная строка.

Утверждение 2.20. $X - ae \backsim \mathcal{W}_p(n, Q)$.

Доказательство. Пусть $Y = (X - ae)C$, C - ортогональная матрица порядка n с первым столбцом $(1/\sqrt{n}, \ldots, 1/\sqrt{n})^T = (e/\sqrt{n})^T$. (См. задачу 2.1.) Тогда $Y_1 = (X - ae)(e/\sqrt{n})^T =$

$5. Доверительные эллипсоиды (метод Шеффе).

Задачи и упражнения к главе 2

Упражнение 2.1. Доказать теорему 2.1.

(Подсказка: рассмотреть вначале случай одномерного θ и потом свести к нему случай многомерного θ. Доказательство этой теоремы очень похоже на доказательство теоремы 1.18 из гл.1.)

Упражнение 2.2. Доказать теорему 2.2.

Упражнение 2.3. Почему существует ортогональная матрица порядка n с первой строкой $(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ или с первым столбцом $(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})^T$

Упражнение 2.4. Доказать лемму 2.13.

Упражнение 2.5. Доказать утверждение 2.19.

Упражнение 2.6. Доказать, что для квадратной матрицы C порядка p

$$W_p(m, CQC^T) = CW_p(m, Q)C^T.$$

В частности, $W_p(m, Q) = Q^{1/2} W_p(m, I_p) Q^{1/2}$.

Глава 3.
Проверки статистических гипотез.

Рассмотрим проверку по выборке $X_1, \ldots, X_n$ из $N_p(a, Q)$ простой гипотезы

$$H : a = a_0, \tag{$*$}$$

где a_0 - заданный p-мерный вектор.

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. Второй метод связан с применением леммы Неймана-Пирсона и широко используется для проверки простых гипотез. Для одной выборки, при проверке $(*)$, оба метода приводят к одному и тому же статистическому критерию (критерию Хотеллинга), но для более сложных многомерных линейных моделей статистические правила получаются разными. Об этом будет рассказано в последующих главах. Сопоставляя эти конкретные критерии, мы придем к пониманию, как в многомерном линейном анализе устроены разумные критерии для проверки линейных гипотез (и как распределены их статистики при гипотезах и альтернативах).

Замечание. Для выборки $X_1, \ldots, X_n$ из $N_p(a, Q)$ параметр a называется параметром положения, а параметр Q — параметром масштаба. Обычно при изменении условий проведения эксперимента, в результате которого мы получаем выборку $X_1, \ldots, X_n$, вначале меняется параметр положения, затем — параметр масштаба и в последнюю очередь — распределение выборки.

1. Метод Роя для проверки простой гипотезы.

Пусть u обозначает переменный p-мерный вектор, $u \in \mathbb{R}^p$. От выборки $X_1, \ldots, X_n$ из $N_p(a, Q)$, где матрица Q невырождена, 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 $N(u^T a, u^T Q u)$. Таким образом, гипотеза $(*)$ для данного u переходит в гипотезу

$$H^{(u)} : u^T a = u^T a_0. \tag{3.1}$$

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

В зависимости от того, задана наблюдателю матрица Q или нет, надо пользоваться либо гауссовским, либо стьюдентовским распределениями для проверки гипотезы $H^{(u)}$. Положим (как обычно)

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} u^T X_i = u^T \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = u^T \overline{X}.$$

Рассмотрим сначала более трудный случай, когда матрица Q неизвестна. Тогда неизвестна и дисперсия наблюдений $x_1, \ldots, x_n$. В этом случае для проверки гипотезы $H^{(u)}$ (3.1) следует использовать "отношение Стьюдента"или, что то же самое, "Стьюдентову дробь":

$$\sqrt{n}\frac{u^T(\overline{X} - a_0)}{s}, \tag{3.2}$$

где

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x^i - \overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(u^T X_i - u^T \overline{X})^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[u^T(X_i-\overline{X})][u^T(X_i-\overline{X})]^T =$$

$$= \frac{1}{n-1}u^T\left(\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})\right)u = \frac{1}{n-1}u^TSu$$

Утверждение 3.1. Статистика (3.2) имеет распределение Стьюдента с $(n-1)$ степенями свободы и параметром нецентральности

$$\delta = \frac{\sqrt{n}u^T(a-a_0)}{\sqrt{u^TQu}}. \tag{3.3}$$

(См. упражнение 3.1.)

Гипотезу $H^{(u)} : \mu = \mu_0$, или эквивалентно $H^{(u)} : \delta = 0$, следует отвергнуть (на уровне $2\alpha,\ 0 < \alpha < 0.5$), если

$$\left|\sqrt{n}\frac{u^T(\overline{X}-a_0)}{s}\right| > t_{1-\alpha}, \tag{3.4}$$

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

Поскольку правая часть (3.4) одна и та же для всех $u \in \mathbb{R}^p$, представляется разумным отвергать гипотезу $(*)$, когда статистика

$$\max_{u\in\mathbb{R}^p}\sqrt{n}\frac{u^T(\overline{X}-a_0)}{s}, \tag{3.5}$$

где $s = \sqrt{\frac{u^TSu}{n-1}}$, принимает "слишком большие"значения. При этом ясно, что критические значения для (3.5) надо выбирать, руководствуясь ее распределением при гипотезе $(*)$.

Если же матрица Q задана, то нужно использовать статистику

$$\max_{u\in\mathbb{R}^p}\sqrt{n}\frac{u^T(\overline{X}-a_0)}{\sqrt{u^TQu}}, \tag{3.6}$$

и отвергать гипотезу, когда эта статистика принимает достаточно большие значения. (См. упражнение 3.2.)

2. Вычисление статистик (3.5) **и** (3.6)**.**

Лемма 3.2. Пусть R - произвольная положительно определенная симметричная матрица порядка p, Y - произвольный p-мерный вектор. Тогда

$$\max_{u\in\mathbb{R}^p}\frac{u^TY}{\sqrt{u^TRu}} = \sqrt{Y^TR^{-1}Y}.$$

Доказательство. Заметим, что функция

$$f(u) = \max_{u\in\mathbb{R}^p}\frac{u^TY}{\sqrt{u^TRu}} \tag{3.10}$$

не меняет своего значения при замене $v = \alpha u$, где $\alpha > 0$. (См. упражнение 3.3.) Таким образом, задачу о нахождении данного максимума можно заменить задачей о нахождении

$$\max_{\substack{u\in\mathbb{R}^p\\ u^TRu=1}} u^TY \tag{3.11}$$

Для нахождения полученного максимума воспользуемся методом Лагранжа отыскания локального условного экстремума.

Напомним его. (См. также [3], гл. 15, $ 5.)

Пусть требуется найти экстремум дифференцируемой функции $f : U \to \mathbb{R}$, где $U \subset \mathbb{R}^{k+m}$ - открытое множество, при наличии m условий связи:

$$\begin{gathered} F_1(x_1, \ldots, x_k, y_1, \ldots, y_m) = 0 \\ F_2(x_1, \ldots, x_k, y_1, \ldots, y_m) = 0 \\ \ldots \\ F_m(x_1, \ldots, x_k, y_1, \ldots, y_m) = 0 \end{gathered}$$

При этом будем предполагать, что функции $F_i, i = \overline{1, m}$, определены и непрерывно дифференцируемы на множестве U, и в любой точке этого множества якобиан

$$\begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{vmatrix}$$

отличен от нуля.

Если в точке $P_0 \in U$ функция f имеет локальный условный экстремум, то найдутся $\lambda_i \in \mathbb{R}, i = \overline{1, m}$, для которых функция $\Psi := f + \sum\limits_{i=1}^{m} \lambda_i F_i$ удовлетворяет в точке P_0 условиям:

$$\begin{gathered} \frac{\partial \Psi}{\partial x_i} = 0, i = \overline{1, k} \\ \frac{\partial \Psi}{\partial y_i} = 0, i = \overline{1, m}, \end{gathered}$$

а, кроме того, $F_i(P_0) = 0, i = \overline{1, m}$.

Обратно, если в точке $P_0 \in U$ выполнены данные условия и при этом $d^2\Psi$ в этой точке - положительно (отрицательно) определенная квадратичная форма, то P_0 есть точка локального условного минимума (соответственно, максимума).

Замечание. Напомним, что квадратичная форма $b : \mathbb{R}^m \to R$ называется **отрицательно определенной**, если при всех $x \in \mathbb{R}^m, x \neq 0$ $b(x) < 0$. Симметричная матрица B порядка m называется **отрицательно определенной**, если при всех $x \in \mathbb{R}^m, x \neq 0$ $x^T B x < 0$. Будем обозначать это так: $B < 0$. Очевидно, квадратичная форма отрицательно определена, если и только если ее матрица отрицательно определена. (См. упражнение 3.2.) Для данного понятия выполнены некоторые простые свойства - см. упражнение 3.3.

Заметим, что в изложенном утверждении все $m + k$ переменных $x_1, \ldots, x_k$, $y_1, \ldots, y_m$ равноправны, т.е. можно не требовать, чтобы именно якобиан функций $F_i, i = \overline{1, m}$ по переменным $y_i, i = \overline{1, m}$ не обращался в ноль. Переформулируем метод Лагранжа следующим образом.

Пусть $f : U \in \mathbb{R}$ - дифференцируемая функция на открытом множестве $U \subset \mathbb{R}^l$, даны $q < l$ условий связи $F_i = 0$, $F_i, i = \overline{1, q}$, определены и непрерывно дифференнцируемы на U, причем ранг матрицы Якоби

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_l} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_q}{\partial x_1} & \cdots & \frac{\partial F_q}{\partial x_l} \end{pmatrix}$$

максимален, т.е. равен q. Тогда необходимые условия условного локального экстремума в точке $P_0 \in U$ таковы: найдутся числа $\lambda_i \in \mathbb{R}, i = \overline{1, q}$, для которых

$$\frac{\partial \Psi}{\partial x_i} = 0, i = \overline{1, l}$$

в точке P_0, где

$$\Psi := f + \sum_{i=1}^{q} \lambda_i F_i,$$

а также $F_i(P_0) = 0, i = \overline{1, q}$.

Если к этим условиям добавить еще подожительную (отрицательную) определенность квадратичной формы $d^2\Psi$ в точке P_0, то получим достаточные условия того, что в точке P_0 функция f имеет локальный условный минимум (соответственно, максимум).

Применим этот метод к данной задаче. В нашем случае $q = 1, l = p, f(u_1, \ldots, u_p) = u^T Y = \sum_{i=1}^{p} u_i Y_i$, где $u_i, Y_i - i -$е компоненты векторов $u, Y \in \mathbb{R}^p$. Условие связи - лишь одно:

$$u^T R u = 1 \Leftrightarrow \sum_{i,j=1}^{p} u_i u_j r_{ij} = 1 \Leftrightarrow F_1(u_1, \ldots, u_p) = 0,$$

где $F_1(u_1, \ldots, u_p) := \sum\limits_{i,j=1}^{p} u_i u_j r_{ij} - 1$. Функция F_1 определена и непрерывно дифференцируема на открытом множестве $U := \mathbb{R}^p \setminus \{0\}$, и матрица Якоби в любой точке этого множества (в данном случае это вектор-строка - градиент функции F_1) отлична от нуля, т.е., эквивалентно, имеет ранг 1, иными словами, максимальный ранг. Действительно,

$$\frac{\partial F_1}{\partial u_k} = \frac{\partial}{\partial u_k}(\sum_{\substack{i,j=1 \\ i,j \neq k}}^{p} u_i u_j r_{ij} + 2 \sum_{\substack{i=1^p \\ i \neq k}}{}_{u_i u_k r_{ik} + p_k^2 r_{kk}}$$

(мы использовали то, что матрица

■

Следствие 3.3. Статистика (3.6) равна

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

Следствие 3.4. Статистика (3.5) равна

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

Поскольку мы отвергаем гипотезу $(*)$ при больших значениях статистик (3.5) или (3.6), в качестве критериальных статистик для $(*)$ можно взять статистики

$$n(\overline{X} - a_0)^T Q^{-1} (\overline{X} - a_0) \tag{3.12}$$

в случае,когда матрица Q известна, или

$$T^2 := n(\overline{X} - a_0)^T S^{-1} (\overline{X} - a_0) \tag{3.13}$$

в случае, когда неизвестную матрицу Q приходится замещать ее выборочной оценкой S.

Статистику T^2 называют статистикой Хотеллинга (Hotelling,Harolds;1931). Она является многомерным обобщением статистики Стьюдента $t = \sqrt{n}\frac{\bar{x} - \mu}{s}$. Хотеллинг нашел распределение T^2 при гипотезе $(*)$.

3. Метод отношения правдоподобий.

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

Гипотеза, подлежащая проверке по наблюдению $H : \theta \in \mathcal{K}$, 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:

$$Q_z := \left\{ x : \frac{p(x, \widehat{\theta})}{p(x, \widetilde{\theta})} > z \right\},$$

где $\widehat{\theta}$ - оценка наибольшего правдоподобия для θ, при условии $\theta \in \Theta$; $\widetilde{\theta}$ - оценка наибольшего правдоподобия для θ при условии $\theta \in \mathcal{K}$ (то есть при гипотезе).

Статистику этого критерия можно записать как отношение двух правдоподобий:

$$\lambda := \frac{p(\xi, \widehat{\theta})}{p(\xi, \widetilde{\theta})}, \text{ или } \lambda := \frac{\max\limits_{\theta \in \Theta} p(\xi, \theta)}{\max\limits_{\theta \in \mathcal{K}} p(\xi, \theta)}. \tag{3.14}$$

Гипотезу $H : \theta \in \mathcal{K}$ предлагается отвергать, если наблюденное λ чрезмерно велико, то есть если λ превосходит критическое значение. Это критическое значение следует вычислять, имея распределение статистики λ при гипотезе $\theta \in \mathcal{K}$.

Не исключается, вообще говоря, что распределение статистики λ (3.14) окажется зависящим от истинного значения параметра θ, даже если $\theta \in \mathcal{K}$, то есть даже если проверяемая гипотеза верна. К счастью, в этой и последующих гауссовских задачах распределение отношения правдоподобий λ оказывается свободным от влияния θ, если $\theta \in \mathcal{K}$ (одним и тем же для всех $\theta \in \mathcal{K}$). Как говорят, при гипотезе статистика λ оказывается свободной (distribution free). Именно это обстоятельство, уже отмеченное ранее для одномерных гауссовских линейных моделей, оправдывает применение здесь метода отношения правдоподобий.

В нашей задаче параметром θ служит пара (a, Q), гипотеза $(*)$ относится к вектору a. Множества $\mathcal{K}$ и Θ определяются в зависимости от того, известна наблюдателю матрица Q или нет.

В этом случае Θ - это множество возможных значений параметра $a \in \mathbb{R}^p$, то есть $\Theta = \mathbb{R}^p$; множество $\mathcal{K}$ состоит из одной точки $a_0 \in \mathbb{R}^p$. Поэтому $\widetilde{\theta} = (a_0, Q), \widehat{\theta} = (\overline{X}, Q)$.

Напомним, что правдоподобие для выборки $X_1, \ldots, X_n$ из $N_p(a, Q)$ имеет вид:

$$p(\vec{X}_1, \ldots, \vec{X}_n; \vec{a}, Q) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^p \det Q}} \exp(-\frac{1}{2}(\vec{X}_i - \vec{a})^T Q^{-1} (\vec{X}_i - \vec{a})) =$$

$$= \frac{1}{(\sqrt{2\pi})^{np}} \frac{1}{(\sqrt{\det Q})^n} \exp(-\frac{1}{2} \sum_{i=1}^{n} (\vec{X}_i - \vec{a})^T Q^{-1} (\vec{X}_i - \vec{a})).$$

Замещая в выражении для правдоподобий параметр указанными значениями, находим статистику:

$$\lambda = \exp\left\{ -\frac{1}{2} \left[\sum_{i=1}^{n} (X_i - \overline{X})^T Q^{-1} (X_i - \overline{X}) - \sum_{i=1}^{n} (X_i - a_0)^T Q^{-1} (X_i - a_0) \right] \right\}. \tag{3.15}$$

Воспользуемся формулой, которая уже встречалась нам в предыдущих главах:

$$\sum_{i=1}^{n} (X_i - a)^T Q^{-1} (X_i - a) = \sum_{i=1}^{n} (X_i - \overline{X})^T Q^{-1} (X_i - \overline{X}) + n(\overline{X} - a)^T Q^{-1} (\overline{X} - a). \tag{3.16}$$

Таким образом, статистика (3.15) преобразуется к виду:

$$\lambda = \exp\left\{\frac{n}{2}(\overline{X}-a_0)^T Q^{-1}(\overline{X}-a_0)\right\}.$$

Следовательно, критической статистикой для проверки $(*)$ может служить следующая статистика:

$$n(\overline{X}-a_0)^T Q^{-1}(\overline{X}-a_0). \tag{3.17}$$

Отвергать гипотезу $(*)$ следует при больших значениях статистики (3.17).

Нетрудно видеть, что мы получили ту же статистику (3.12) и то же правило для проверки $(*)$, что мы нашли, применив метод Роя.

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 в главе 2,

$$\widehat{\theta} = (\overline{X}, S_n), \widetilde{\theta} = (a_0, \widetilde{Q}),$$

где

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

Следовательно, отношение правдоподобий в данном случае примет вид:

$$\lambda = \left(\frac{\sqrt{\det \widetilde{Q}}}{\sqrt{\det S_n}}\right)^n \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}(X_i-\overline{X})^T S_n^{-1}(X_i-\overline{X}) - \sum_{i=1}^{n}(X_i-a_0)^T\widetilde{Q}^{-1}(X_i-a_0)\right]\right\}. \tag{3.18}$$

Воспользуемся уже знакомым нам приемом, который использовался в главе 2:

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

Аналогично,

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

Поэтому статистика λ (3.18) равна:

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

Таким образом, получаем следующее промежуточное правило для проверки гипотезы $(*)$:

отвергать гипотезу $H : a = a_0$ при больших значениях статистики

$$\frac{\det\left[\sum\limits_{i=1}^{n}(X^i-\overline{X})(X^i-\overline{X})^T + n(\overline{X}-a_0)(\overline{X}-a_0)^T\right]}{\det\left[\sum\limits_{i=1}^{n}(X^i-\overline{X})(X^i-\overline{X})^T\right]} \tag{3.19}$$

Статистику (2.5) можно упростить и довести до вида T^2 (1.13), если воспользоваться следующей леммой об определителях блочных матриц.

Упростим статистику (3.19), воспользовавшись следующей леммой.

Лемма 3.2: Пусть квадратная матрица Σ разбита на блоки, причем Σ_{11}, Σ_{22} - квадратные невырожденные матрицы:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Тогда

$$\det(\Sigma) = \det(\Sigma_{11})\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) = \det(\Sigma_{22})\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \qquad (3.20)$$

Доказательство. $\Box$ При доказательстве воспользуемся правилом перемножения блочных матриц. Кроме того, будет использовано следующее тождество:

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

где A, B, C - матрицы размера $n \times n, m \times n, m \times m$ соответственно, а 0 - нулевая матрица размера $n \times m$. (Такая лемма известна из линейной алгебры.)

Таким образом, воспользовавшись сделанными замечаниями, получим:

$$\det \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \det \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \det \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{pmatrix} =$$

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

Вторая формула доказывается аналогично:

$$\det \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \det \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \det \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} =$$

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

$\blacksquare$

Теперь вернемся к отношению (3.19). Определитель, который стоит в числителе этой дроби, представим в следующем виде:

$$det \begin{pmatrix} \sum\limits_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T & -\sqrt{n}(\overline{X} - a_0) \\ \sqrt{n}(\overline{X} - a_0)^T & 1 \end{pmatrix} \qquad (3.21)$$

В том, что это верно, можно убедиться, воспользовавшись второй из формул (3.20) только что доказанной леммы. Теперь для вычисления (3.21) применим первую из формул (3.20). Попутно заменим на $(n-1)S$ выражение $\sum\limits_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T$. Таким образом, получаем, что числитель (3.19) равен:

$$\det[(n-1)S]\det[1 + n(\overline{X} - a_0)^T[(n-1)S]^{-1}(\overline{X} - a_0)] = \det[(n-1)S]\det[1 + \frac{1}{n-1}T^2],$$

где определение статистики Хотеллинга T^2 было введено ранее, см. (3.13). После таких преобразований числителя дробь (3.19) упрощается, и правило для проверки гипотезы (*) можно теперь сформулировать следующим образом:

отвергать гипотезу (*) при больших значениях статистики

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

Заметим, что символ det можно опустить, т.к. $1 + \frac{1}{n-1}T^2$ - число, т.е. матрица размера 1×1, а определитель матрицы такого размера, т.е. числа - это само это число. В итоге получаем, что (*) следует отвергать, если

$$T^2 \geq Const,$$

что совпадает с (3.13).

Глава 4. Проверки статистических гипотез - продолжение; доверительные эллипсоиды.

1. Распределение статистики $\mathbf{T^2}$.

Для нахождения распределения статистики T^2 докажем следующую теорему:

Лемма 4.1. Пусть $Y, Y_1, \ldots, Y_k$ - независимые одинаково распределенные случайные величины с распределением $N_p(0, \Sigma)$, причем матрица Σ невырождена (т.е., эквивалентно, она положительно определена - см. упражнение 5.2 в главе 1) . Пусть

$$W = \sum_{i=1}^{k} Y_i Y_i^T, \quad \mathcal{T}^2 = (\nu + Y)^T W^{-1} (\nu + Y), \tag{3.21),}$$

где ν - произвольный вектор, $\nu \in \mathbb{R}^p$. Тогда для $k \geq p$

$$\mathcal{T}^2 \stackrel{d}{=} \frac{\chi^2(p, \Delta)}{\chi^2(k - p + 1)}, \tag{3.22}$$

причем случайные величины в числителе и знаменателе независимы, и параметр нецентральности Δ определяется равенством

$$\Delta = \nu^T \Sigma^{-1} \nu. \tag{3.23}$$

Замечание 4.2. Из (3.22) следует, что

$$\frac{k - p + 1}{p} \mathcal{T}^2 \stackrel{d}{=} F(p, k - p + 1, \Delta). \tag{3.24}$$

Замечание 4.3. При $k \geq p$ матрица W невырождена с вероятностью 1, а при $k < p$ вырождена почти наверное. (См. упражнения 3.1, 3.2.)

Введем обозначения для элементов матриц $\Sigma, \Sigma^{-1}, W, W^{-1}$. Пусть $\sigma_{ij}, \sigma^{ij}, w_{ij}, w^{ij}$ - ij-е элементы этих матриц: $\Sigma, \Sigma^{-1}, W, W^{-1}$ соответственно.

Доказательство теоремы существенно опирается на следующие леммы.

Лемма 4.4. $\dfrac{\sigma^{11}}{w^{11}} \stackrel{d}{=} \chi^2(k - p + 1)$.

Доказательство. □ Рассмотрим случайный вектор $X \sim N_p(0, \Sigma), X = (x_1, \ldots, x_p)^T$. Разобъем на блоки матрицы Σ и W:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}.$$

Здесь $\Sigma_{11} = \sigma_{11}, W_{11} = w_{11}$ - матрицы размера 1×1, т.е являющиеся числами, матрицы Σ_{22} ? W_{22} — квадратные матрицы размера $(p - 1) \times (p - 1)$.

Ранее было отмечено, что

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

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

$$(\sigma^{11})^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{3.25}$$

Действительно, по определению

$$\sigma^{11} = \frac{\det(\Sigma_{22})}{\det(\Sigma)},$$

а в силу леммы 3.2

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

Аналогично,

$$(w^{11})^{-1} = W_{11} - W_{12}W_{22}^{-1}W_{21}. \tag{3.26}$$

Как уже отмечалось,

$$X^{(1)} = \Sigma_{12}\Sigma_{22}^{-1}X^{(2)} + \xi, \tag{3.27}$$

где ξ– независимая от $X^{(2)}$ случайная величина, распределенная по гауссовскому закону
$N(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, то есть в силу (3.25) - по закону $N(0, (\sigma^{11})^{-1})$.

Для случайных векторов, фигурирующих в условии теоремы 3.1, равенство (3.27) означает, что

$$Y_i^{(1)} = \Sigma_{12}\Sigma_{22}^{-1}Y_i^{(2)} + \xi_i \ ??? \ i = \overline{1,k}, \tag{3.28}$$

где $\xi_1, \ldots, \xi_k$ суть независимые и не зависящие от $Y_1^{(2)}, \ldots, Y_k^{(2)}$ гауссовские случайные величины, распределенные по закону $N(0, (\sigma^{11})^{-1})$.

Положим

$$R^2 = \min_{\beta \in \mathbb{R}^{p-1}} \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^2. \tag{3.29}$$

При фиксированных значениях $Y_1^{(2)}, \ldots, Y_k^{(2)}$ случайная величина R^2 распределена как
$(\sigma^{11})^{-1}\chi^2(k-p+1)$, что следует из теории гауссовской линейной регрессии, примененной к модели (3.29). Поскольку $Y_1^{(2)}, \ldots, Y_k^{(2)}$ — случайные величины, сказанное выше относится к условному расчпределению R^2 при фиксированных $Y_1^{(2)}, \ldots, Y_k^{(2)}$. Поскольку при любых значениях $Y_1^{(2)}, \ldots, Y_k^{(2)}$ условное распределение R^2 одно и то же, то оно не зависит от значений $Y_1^{(2)}, \ldots, Y_k^{(2)}$ и совпадает с безусловным (т.е. обычным) распределением R^2. Итак, установлено, что случайная величина R^2 распределена по закону

$$R^2 \sim (\sigma^{11})^{-1}\chi^2(k-p+1). \tag{3.30}$$

Остается лишь указать для R^2 явное выражение. Для этого можно воспользоваться готовыми формулами из главы 1, п.5 (условное математическое ожидание для гауссовских векторов, т.е. линейная регрессия - см. также п.6 в той же главе), но проще проделать необходимые выкладки непосредственно. Найдем сначала значение $\hat{\beta}$, при котором правая часть (3.29) достигает минимума. Введем функцию

$$Q(\beta) := \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^2 = \sum_{i=1}^{k} \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right) \left(Y_i^{(1)} - \beta^T Y_i^{(2)}\right)^T \tag{3.31}$$

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 $\nabla Q(\beta)$ обозначает 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 подученное для ∇Q выражение, получим, что градиент обращается в ноль (а функция $Q(\beta)$, соответственно, достигает минимального значения) при

$$\hat{\beta} = \left[\sum_{i=1}^{k} \left(Y_i^{(2)}\right)\left(Y_i^{(2)}\right)^T\right]^{-1}\left[\sum_{i=1}^{k} \left(Y_i^{(1)}\right)\left(Y_i^{(2)}\right)^T\right]. \tag{3.32}$$

Подставив полученное выражение (3.32) в (3.31), найдем, что R^2 (3.29) равно

$$R_0^2 = \sum_{i=1}^{k} Y_i^{(1)}Y_i^{(1)} - \left(\sum_{i=1}^{k} Y_i^{(1)}Y_i^{(2)}\right)^T \left(\sum_{i=1}^{k} Y_i^{(2)}Y_i^{(2)}\right)^{-1} \left(\sum_{i=1}^{k} Y_i^{(1)}Y_i^{(2)}\right)^T. \tag{3.33}$$

Теперь заметим, что составляющие выражения (3.33) — это блоки матрицы $W = \sum\limits_{i=1}^{k} Y_iY_i^T$, которые мы указали ранее. Поэтому

$$R^2 = W_{11} - W_{12}W_{22}^{-1}W_{21},$$

и, согласно (3.26), равно $(w^{11})^{-1}$. Поэтому

$$R^2 = (w^{11})^{-1} \tag{3.34}$$

Сопоставив (3.34) с (3.30), найдем, что

$$(w^{11})^{-1} \stackrel{d}{=} (\sigma^{11})^{-1}\chi^2(k-p+1).$$

Таким образом, утверждение леммы доказано. ■

Лемма 4.5. $\dfrac{l^T\Sigma^{-1}l}{l^TW^{-1}l} \stackrel{d}{=} \chi^2(k-p+1)$ для любого $l \in \mathbb{R}^p,\ l \neq 0$.

□ Выберем l таким, что $|l| = 1$. Рассмотрим ортогональную матрицу B, первая строка которой есть l^T. Введем случайные векторы $\widetilde{Y}_1, \ldots, \widetilde{Y}_k$, положив $\widetilde{Y}_i = BY_i, i = \overline{1,k}$. Ясно, что $\widetilde{Y}_1, \ldots, \widetilde{Y}_k$ суть независимые $N_p(0, \widetilde{\Sigma})$, где $\widetilde{\Sigma} = B\Sigma B^T$. Рассмотрим матрицу $\widetilde{W} = \sum\limits_{i=1}^{k} \widetilde{Y}_i\widetilde{Y}_i^T = BWB^T$. Введем обозначения для элементов матрицы $\tilde{\Sigma}, \tilde{W}, \tilde{\Sigma}^{-1}, \tilde{W}^{-1}$. Положим

$$\tilde{\Sigma} = \|\tilde{\sigma}_{ij}\|, \tilde{\Sigma}^{-1} = \|\tilde{\sigma}^{ij}\|, \tilde{W} = \|\tilde{w}_{ij}\|, \tilde{W}^{-1} = \|\tilde{w}^{ij}\|.$$

Заметим, что в силу выбора матрицы B

$$l^T\Sigma^{-1}l = (Bl)^T \left[B\Sigma B^T\right]^{-1}(Bl) = (1,0,\ldots,0)^T\tilde{\Sigma}^{-1}(1,0,\ldots,0) = \tilde{\sigma}^{11}.$$

Аналогично:

$$l^TW^{-1}l = \tilde{w}^{11}.$$

Поэтому

$$\frac{l^T\Sigma^{-1}l}{l^TW^{-1}l} = \frac{\tilde{\sigma}^{11}}{\tilde{w}^{11}} \stackrel{d}{=} \chi^2(k-p+1)$$

по доказанной лемме 3. ■

Обратимся теперь к **доказательству теоремы 1**.

□ Представим (3.21) в виде:

$$\mathcal{T}^2 = \frac{(Y+\nu)^TW^{-1}(Y+\nu)}{(Y+\nu)^T\Sigma^{-1}(Y+\nu)}(Y+\nu)^T\Sigma^{-1}(Y+\nu). \tag{3.35}$$

В силу леммы 4.4 условное распределение при фиксированном Y случайной величины

$$\frac{(Y+\nu)^T W^{-1}(Y+\nu)}{(Y+\nu)^T \Sigma^{-1}(Y+\nu)} \tag{3.36}$$

— это распределение $\chi^2(k-p+1)$. Заметим, что это распределение - одно и то же при всех значениях Y. Следовательно, "безусловное"распределение, т.е. просто распределение (3.36) есть $\chi^2(k-p+1)$. Кроме того, как случайная величина, (3.36) не зависит от Y. Следовательно, (3.36) не зависит от

$$(Y+\nu)^T \Sigma^{-1}(Y+\nu). \tag{3.37}$$

Как известно, случайная величина (3.37) распределена по закону $\chi^2(k-p+1,\Delta)$, где значение параметра нецентральности Δ определяется формулой (3.23).

В результате получаем, что распределение $\mathcal{T}^2$ (3.35) - такое же, как у отношения двух независимых случайных величин, распределенных по закону χ^2, как это и утверждается в (3.32). Остается лишь напомнить, что по определению

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

так что

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

где $F(p,k-p+1,\Delta)$ — это распределение Фишера со степенями свободы $p, k-p+1$ и параметром нецентральности Δ. Таким образом, теорема полностью доказана. ■

Теперь мы можем доказать следующую теорему, касающуюся статистики T^2 Хотеллинга:

Теорема 4.6.Если $n>p$, то

$$T^2=n(\overline{X}-a_0)^T S^{-1}(\overline{X}-a_0) \overset{d}{=} \frac{(n-1)p}{n-p}F(p,n-p,\Delta), \tag{3.38}$$

где параметр нецентральности Δ определяется формулой:

$$\Delta=n(a-a_0)^T Q^{-1}(a-a_0), \tag{3.39}$$

□ Для доказательства надо учесть, что

(a)

$$(n-1)S=W \overset{d}{=} \sum_{i=1}^{n-1} Y_i Y_i^T,$$

где $Y_1,\dots,Y_{n-1}$ суть независимые $N_p(0,Q)$;

(b)

$$\sqrt{n}(\overline{X}-a_0) \overset{d}{=} Y \sim N_p(\nu,Q),$$

где $\nu=\sqrt{n}(a-a_0)$, и случайный вектор Y не зависит от $Y_1,\dots,Y_{n-1}$.

Следовательно, к статистике

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

применима теорема 1 с заменой k на $n-1$. Отсюда

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

где

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

$2. Доверительные эллипсоиды.

Определение 4.7. Пусть $(\mathfrak{X}, \mathfrak{F}, \{\mathbf{P}_\theta | \theta \in \Theta\})$ - исходное статистическое пространство (напомним, это определение было введено в $1 главы 2), $\tau : \Theta \to \mathbb{R}^n$ - заданная функция, $\alpha \in [0,1]$ - некоторое число. Тогда **доверительное множество для функции τ с доверительной вероятностью** α - это любое подмножество $A_X \subset \mathbb{R}^n$, зависящее от $X \in \mathfrak{X}$ (или, более формально, функция $\mathfrak{X} \to 2^{\mathbb{R}^n}$, где $2^{\mathbb{R}^n}$ - множество всех подмножеств $\mathbb{R}^n$), для которого при всех $\theta \in \Theta$

$$\mathbf{P}_\theta\{X \in \mathfrak{X} | \tau(\theta) \in A_X\} \geq \alpha.$$

Возникает естественный частный случай этого очень общего определения. Пусть $\Theta := \Theta_1 \times \Theta_2 \times \ldots \times \Theta_m$, где $\Theta_k \subset \mathbb{R}^{n_k}$. Т.е. неизвестный параметр θ есть набор из m неизвестных параметров $(\theta_1, \ldots, \theta_m)$, которые представляют из себя конечномерные векторы. Тогда доверительное множество для θ_k, где $k = \overline{1,m}$ - некоторый фиксированный индекс, есть, по определению, доверительное множество для функции $\tau : \Theta \to \mathbb{R}^{n_k}$, $\tau(\theta) := \theta_k$, т.е., говоря по-другому, соответствие, которое каждому $X \in \mathfrak{X}$ сопоставляет подмножество $A_X \subset \mathbb{R}^{n_k}$ так, что при всех $\theta \in \Theta$

$$\mathbf{P}_\theta\{X \in \mathfrak{X} | \theta_k \in A_X\} \geq \alpha.$$

Здесь $\alpha \in [0;1]$ - заданная доверительная вероятность.

Если α достаточно близко к единице (например, $\alpha = 0,99$), то, построив доверительное множество, мы можем сказать, что функция τ от неизвестного параметра θ, по-видимому, лежит именно в этом множестве - хотя неизвестно точно, чему же эта функция равна. Это - другой подход к оцениванию неизвестного параметра θ и функций τ от него; в известном смысле он противоположен построению оценок.

Замечание 4.8. Частный случай доверительных множеств, доверительные интервалы - хорошо известное понятие в математической статистике. Напомним, что для одномерной нормальной выборки $X = (X_1, \ldots, X_n)^T$, где $X_k \backsim \mathcal{N}(a, \sigma^2), k = \overline{1,n}$ - независимые и одинаково распределенные случайные величины, а дисперсия σ^2 известна,

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) -$$

доверительный интервал для a с доверительной вероятностью $1-2\alpha$, где $\alpha \in (0;1/2)$, а через z_β обозначен квантиль стандартного нормального распределения $\mathcal{N}(0,1)$, соответствующий уровню $\beta \in (0;1)$.

Если же дисперсия σ^2 неизвестна, то

$$\left(\overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}, \overline{X} + \frac{s}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) -$$

доверительный интервал для a, соответствующий уровню значимости $1-2\alpha$, где опять $\alpha \in (0;1/2)$, но t_β - квантиль распределения Стьюдента с $n-1$ степенями свободы, соответствующий уровню $\beta \in (0;1)$. Это - классические результаты статистики.

Теперь мы стоим перед аналогичной, совершенно естественно возникающей многомерной задачей: как же построить доверительный интервал для неизвестного параметра a (вектора математических ожиданий для многомерного гауссовского распределения) по выборке из этого распределения? При этом не только вектор средних a, но и матрица ковариаций Q этого распределения нам не известна; однако мы считаем, что она заведомо невырождена, т.е., эквивалентно, положительно определена.

Более формально: если $\Theta := \{(a,Q)|a \in \mathbb{R}^p, Q > 0\}$, $\mathbf{P}_\theta = \mathcal{N}_p(a,Q) \times \mathcal{N}_p(a,Q) \times \ldots \times \mathcal{N}_p(a,Q)$ для $\theta = (a,Q) \in \Theta$,

$$(\mathfrak{X}, \mathfrak{F}, \{\mathbf{P}_\theta | \theta \in \Theta\}) = (\mathbb{R}^{pn}, \mathfrak{B}(\mathbb{R}^{pn}), \{\mathcal{N}_p(a,Q) \times \mathcal{N}_p(a,Q) \times \ldots \times \mathcal{N}_p(a,Q) | a \in \mathbb{R}^p, Q > 0\})$$

(произведение состоит из n сомножителей) - исходное статистическое пространство, являющееся, как отмечалось в \$1 главы 2, базой для наших многомерных моделей, то каково доверительное множество $A_X \subset \mathbb{R}^p$ для функции $\tau : \Theta \to \mathbb{R}^p$, определяемой так: $\tau(\theta) := a$ для $\theta = (a,Q) \in \Theta$?

Но мы уже располагаем инструментами для решения этой задачи. А именно: мы знаем, что если $X = (X_1, \ldots, X_n)^T$ - выборка из распределения $\mathcal{N}_p(a,Q)$ (мы будем считать ее np-мерным вектором - столбцом), то

$$n(\overline{X} - a)^T S^{-1} (\overline{X} - a) \backsim \frac{(n-1)p}{n-p} F(p, n-p),$$

где, напомним, $F(p, n-p)$ - распределение Фишера (Снедекора) с p, $n-p$ степенями свободы, $\overline{X} = \frac{1}{n} \sum\limits_{k=1}^{n} X_k$ - эмпирическое среднее, $S := \frac{1}{n-1} \sum\limits_{k=1}^{n} (X_k - \overline{X})(X_k - \overline{X})^T$ - эмпирическая матрица ковариаций.

Отсюда немедленно получаем следующий результат.

Утверждение 4.9. Если для $X \in \mathbb{R}^{pn}$ мы определим множество

$$E_X := \left\{ a \in \mathbb{R}^p | n(\overline{X} - a)^T S^{-1} (\overline{X} - a) < \frac{(n-1)p}{n-p} F_{1-\varepsilon}(p, n-p) \right\} \subset \mathbb{R}^p,$$

то оно будет доверительным множеством для параметра a, соответствующим доверительной вероятности $1 - \varepsilon$ ($\varepsilon \in (0;1)$).

В случае, когда матрица Q ковариаций известна, т.е., формально говоря, $\Theta := \mathbb{R}^p, \mathbf{P}_\theta = \mathcal{N}_p(a,Q) \times \mathcal{N}_p(a,Q) \times \ldots \times \mathcal{N}_p(a,Q)$ для $\theta = a$ (сомножителей, как всегда, n; кроме того, $Q > 0$), имеем:

$$n(\overline{X} - a)^T Q^{-1} (\overline{X} - a) \backsim \chi^2(p),$$

и очевидным образом строится доверительное множество.

Утверждение 4.10. Если $\varepsilon \in (0;1)$, то для $X \in \mathfrak{X} = \mathbb{R}^{pn}$

$$E'_X := \left\{ a \in \mathbb{R}^p | n(\overline{X} - a)^T Q^{-1} (\overline{X} - a) < \chi^2_{1-\varepsilon}(p) \right\} \subset \mathbb{R}^p -$$

доверительное множество с доверительной вероятностью $1 - \varepsilon$.

Эти доверительные множества являются n-мерными эллипсоидами (или, вернее, областями, ограниченными эллипсоидами - ведь, формально говоря, в аналитической геометрии эллипсоидом называется не область в $\mathbb{R}^p$, а поверхность.) Они так и называются - **доверительными эллипсоидами.**

После этих простых соображений изложим более содержательные - так называемый **метод Шеффе (H. Sheffe)**. Зададимся некоторым набором векторов $C \subset \mathbb{R}^p$, не содержащим нулевой вектор, и будем строить доверительные множества для a с помощью набора $\{c^T a | c \in C\}$.

Определение 4.11. Выражения $c^T a, c \in C$ являются линейными формами от аргумента a. Такие формы называются в данном методе **контрастами**. Одновременное построение доверительных интервалов для этих контрастов называется **множественным сравнением**. Метод будет основан именно на таком одновременном построении.

Теорема 4.12.(Шеффе, 1952) Введем следующее вспомогательное обозначение: если $\alpha \in (0;1)$ - заданная доверительная вероятность, то пусть $k^2 := \frac{p(n-1)}{(n-p)n} F_{1-\alpha}(p, n-p)$. Тогда множество

$$A_X := \{a \in \mathbb{R}^p | -k\sqrt{c^T S c} + c^T \overline{X} < c^T k\sqrt{c^T S c} + c^T \overline{X} \ \forall c \in C\}$$

есть доверительное множество для a с доверительной вероятностью $1-\alpha$.

$\square$ Необходимо доказать, что при любом $\theta \in \Theta$ (где $\Theta = \{(a, Q) | a \in \mathbb{R}^p, Q > 0\}$)

$$\mathbf{P}_\theta\{X \in \mathbb{R}^{pn} | a \in A_X\} > 1-\alpha.$$

Для этого достаточно показать, что

$$\{X \in \mathbb{R}^{pn} | a \in E_X\} \subset \{X \in \mathbb{R}^{pn} | a \in A_X\},$$

или что $E_X \subset A_X$ при любом $X \in \mathbb{R}^{pn}$. Но

$$A_X = \bigcap_{c \in C} \{a \in \mathbb{R}^p | -k\sqrt{c^T S c} + c^T \overline{X} < c^T a < k\sqrt{c^T S c} + c^T \overline{X}\}.$$

Значит, будет достаточно показать, что

$$E_X \subset \{a \in \mathbb{R}^p | -k\sqrt{c^T S c} + c^T \overline{X} < c^T a < k\sqrt{c^T S c} + c^T \overline{X}\}$$

при всех $X \in \mathbb{R}^{pn}$, $c \in C$. Этот последний факт следует из того, что (мы фиксировали $X \in \mathbb{R}^{pn}, c \in C$) для $E_X (a-\overline{X})^T S^{-1}(\overline{X}) \le k^2$, а

$$\max_{(a-\overline{X})^T S^{-1}(\overline{X}) \le k^2} c^T(a-\overline{X}) = \max_{(a-\overline{X})^T S^{-1}(\overline{X}) = k^2} c^T(a-\overline{X}) = k \max_{(a-\overline{X})^T S^{-1}(\overline{X}) = 1} c^T(a-\overline{X}) = k\sqrt{c^T S c}.$$

Последнее равенство следует из леммы 3.2 главы 3. (Обосновать другие два равенства в этой цепочке!) Этим все и доказано. ■

Мы можем рассмотреть частный случай конструкции Шеффе: $C = \{c_k | k = \overline{1, m}\}$ - конечное множество. Тогда доверительный эллипсоид E_X вкладывается в множество A_X, которое представляет из себя область в пространстве $\mathbb{R}^p$, ограниченную m парами параллельных гиперплоскостей, касательных к границе - поверхности эллипсоида E_X. Т.е. эта область - многогранник. (Проведите это рассуждение подробнее!)

3. Проверка приближенной гипотезы.

Статистические гипотезы часто высказывают как точные утверждения о параметре θ из базовой модели статистики - статистического пространства, или, иными словами, о неизвестном параметре (параметрах) распределения наблюдений. Очевидный пример - гипотеза $H : a = a_0$, которой была посвящена глава 3. Но в данном случае при неограниченном увеличении размера выборки возникает неприятная ситуация: рано или поздно мы отвергнем любую наперед заданную гипотезу $H : a = a_0$, кроме той, которая является действительно истинной - т.е. той, где a_0 - истинное значение параметра a распределения.

Однако как можно, предлагая вектора a_0 для гипотез $a = a_0$, угадать действительно истинное значение этого параметра? Совершенно невероятно, чтобы нам удалось так точно угадать - но только тогда мы не попадем в описанную выше ловушку. Как же быть?

Заметим, что мы отвергаем гипотезу о точном равенстве вовсе не тогда, когда истинное значение параметра a не равно точно объявленному. (Иначе, по-видимому, пришлось бы практически всегда отвергать такие гипотезы.) Мы отвергаем гипотезу, если видим: расхождение между объявленным значением параметра и тем, которое, по-видимому, истинно, настолько велико, что мы не можем пренебречь этой разницей. И это надо иметь в виду при формулировании и проверке таких гипотез. Однако содержащееся в них утверждение - это все-таки именно равенство. Сейчас же мы рассмотрим гипотезы, изначально в своей формулировке содержащие неравенства, касающиеся неизвестного параметра a. Эти неравенства будут как раз строго выражать мысль о том, что истинное значение a близко к a_0 - тому значению, которое назвали мы.

Сперва рассмотрим одномерный случай. Пусть $X = (X_1, \dots, X_n)$ - выборка из $\mathcal{N}(a, \sigma^2)$. Тогда из классической математической статистики нам известно следующее:

Утверждение 4.13. Вероятность того, что

$$\left| \sqrt{n} \frac{\overline{X} - a}{s} \right| < t_\varepsilon,$$

равна $1 - 2\varepsilon$ при любых значениях параметров распределения a, σ. Здесь $\varepsilon \in (0; 1/2)$, $s := \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$ - эмпирическая дисперсия (несмещенная оценка для σ^2), а $t_{1-2\varepsilon}$ - квантиль распределения Стьюдента с $n - 1$ степенями свободы, соответствующий уровню $1 - 2\varepsilon$.

Перефразируем этот факт так: если a_0 - объявленное значение параметра a распределения $\mathcal{N}(a, \sigma^2)$, то мы отвергаем на уровне 2ε эту гипотезу тогда и только тогда, когда Так мы выразили мысль

В многомерном случае мысль о малости $a - a_0$, где a_0 - объявленное, а a - истинное значение параметра, можно выразить с помощью метода Роя сведения многомерных задач к одномерным. Другими словами, мы считаем, что $a - a_0$ мало тогда и только тогда, когда $u^T a - u^T a_0$ мало при всех $u \in \mathbb{R}^p, u \neq 0$. Но мы же располагаем способом формально объяснить, что значит, что $u^T a - u^T a_0$ мало для фиксированного $u \in \mathbb{R}^p, u \neq 0$. А именно, пусть $(X_1, \dots, X_n)$ - наша исходная выборка из $\mathcal{N}_p(a, Q)$ ($a \in \mathbb{R}^p$, $Q > 0$ - матрица размера $p \times p$ - неизвестные параметры распределения. Тогда $(u^T X_1, \dots, u^T X_n)$ есть выборка из уже одномерного гауссовского распределения $\mathcal{N}(u^T a, u^T Q u)$. Выборчная дисперсия s для данной выборки равна (подсчитайте это!) $u^T S u$. Значит, $u^T a - u^T a_0$ мало, если

$$\left| \sqrt{n} \frac{u^T a - u^T a_0}{u^T S u} \right| < \delta. \tag{4.5}$$

(Здесь $\delta > 0$ - некоторое число; мы можем его выбирать, вообще говоря, произвольно.) Но нам необходимо, чтобы $u^T a - u^T a_0$ было мало не при каком-то отдельном u, а при всех $u \in \mathbb{R}^p, u \neq 0$. Это можно формализовать таким образом: неравенство (4.5) выполнено при всех $u \in \mathbb{R}^p, u \neq 0$. Иначе говоря,

$$\max_{u \in \mathbb{R}^p, u \neq 0} \left| \sqrt{n} \frac{u^T a - u^T a_0}{u^T S u} \right| < \delta. \tag{4.6}$$

По лемме 3.2 из главы 3 (проведите подробнее в этом месте рассуждение!) это выполнено тогда и только тогда, когда

$$(a - a_0)^T S^{-1} (a - a_0) < \delta^2.$$

Это можно сформулировать так: параметр нецентральности статистики Хотеллинга $T^2 := n(\overline{X} - a_0)^T S^{-1}(\overline{X} - a_0)$ не превосходит $n\delta^2$. Поэтому станем отвергать гипотезу (4.6) при больших значениях T^2. Критическое значение уровня α - обозначим его $t(\alpha)$ - определим как решение относительно t экстремальной задачи

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

Эта задача равносильна уравнению

$$\mathbf{P}\left\{\frac{(n-1)p}{n-p} F(p, n-p, n\delta^2) \leq t\right\} = 1 - \alpha,$$

т.к. распределения Фишера $F(\nu_1, \nu_2, \Delta)$ стохастически возрастают при возрастании Δ.

Глава 7. Проверка независимости.

Пусть $(X_k, Y_k)^T, k = \overline{1, n}$ - выборка из распределения $\mathcal{N}_2((a, b)^T, \Sigma)$, где матрица ковариаций Σ - квадратная матрица порядка 2 - имеет вид $\begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$, здесь $\sigma_x^2 = \operatorname{Var} X, \sigma_y^2 \operatorname{Var} Y, \rho = \operatorname{corr}(X, Y)$ для $(X, Y)^T \backsim \mathcal{N}_2((a, b)^T, \Sigma)$. Построим правило для проверки гипотезы $\rho = 0$, т.е. гипотезы о том, что X, Y некоррелированы (или, что то же самое, независимы - это равносильные предположения, т.к. X, Y имеют совместное гауссовское распределение, см. \$4 главы 1).

Приложение.

Доказательство теоремы о замене переменных в интеграле Лебега по множествам в евклидовом $n-$мерном пространстве.

Пусть $D, G \subset \mathbb{R}^n$ - открытые множества, $\alpha : D \to G$ - биекция, причем $\alpha \in C^1(D), \alpha^{-1} \in C^1(G)$, и при этом $J(t)$ - якобиан функции α в точке $t \in D$ - отличен от нуля во всех точках $t \in D$. Пусть, далее, $f : G \to \mathbb{R}$ - борелевская функция, $A \in \mathfrak{B}(D)$ - некоторое подмножество, $\alpha(A)$ - его образ при отображении α.

Тогда $f \in L_1(\alpha(A))$, если и только если $f(\alpha(t))|J(t)| \in L_1(A)$. Если имеет место интегрируемость, то

$$\int\limits_A f(\alpha(t))|J(t)|dt = \int\limits_{\alpha(A)} f(x)dx. \tag{1.3}$$

Доказательство. Шаг 1. Сначала заметим, что мы можем перейти от множеств $A, \alpha(A)$ к множествам D, G: $f \in L_1(\alpha(A))$ тогда и только тогда, когда $I_{\alpha(A)} \in L_1(G)$, и в случае интегрируемости

$$\int\limits_{\alpha(A)} f(x)dx = \int\limits_G f(x)I_{\alpha(A)}(x)dx.$$

Аналогично, $f(\alpha(t))|J(t)| \in L_1(A)$ тогда и только тогда, когда $f(\alpha(t))|J(t)|I_A(t) \in L_1(D)$, и в случае интегрируемости

$$\int\limits_A f(\alpha(t))|J(t)|dt = \int\limits_D f(\alpha(t))|J(t)|I_A(t)dt.$$

Т.е. достаточно доказать следующий **факт** для функции $g(x) = f(x)I_{\alpha(A)}(x)$: $g(x) \in L_1(G)$ тогда и только тогда, когда $g(\alpha t)|J(t)| \in L_1(D)$, и в случае интегрируемости

$$\int\limits_G g(x)dx = \int\limits_D g(\alpha t)|J(t)|dt.$$

Шаг 2. Пусть сначала $g = I_B$, $B = \prod_{k=1}^n [a_k, b_k)$, где при $k = \overline{1, n}$ $[a_k, b_k)$ - ограниченный полуинтервал на действительной прямой (будем называть такие множества B *n-мерными брусами*), и при этом $\overline{B} = \prod_{k=1}^n [a_k, b_k] \subset G$ (т.е., как говорят, B компактно содержится в G). Тогда доказываемое утверждение верно. Действительно, сформулируем **теорему о замене переменных в кратном интеграле Римана**:

Пусть $\alpha : V \to W$ - биекция ограниченных замкнутых кубируемых областей $V, W \subset \mathbb{R}^n$ (напомним, что замкнутая область - это замыкание области в обычном смысле этого слова, т.е. открытого связного множества), причем $\psi \in C^1(V), \psi^{-1} \in C^1(W)$, $J(t)$ - якобиан функции ψ в точке $t \in V$ - отличен от нуля во всех точках $t \in V$. Пусть, далее, функция $h : W \to \mathbb{R}$ интегрируема по Риману на W. Тогда функция $h(\psi(t))|J(t)|$ интегрируема по Риману на V и их интегралы равны.

(Доказательство - см. [3], том 2, гл. 2, \$5.)

Применим эту теорему к $V = \alpha^{-1}(\overline{B}), W = \overline{B}, \psi = \alpha, h = 1$. Заметим, что $\overline{B}$ - замкнутая ограниченная кубируемая область в $\mathbb{R}^n$. Кроме того, $\alpha^{-1}(\overline{B})$ - также ограниченная замкнутая кубируемая область в $\mathbb{R}^n$. (См. [3], том 2, гл. 2, \$ 5.) Т.к. функция h, тождественно равная единице, интегрируема по Риману на W (а кроме того, и по Лебегу - ведь она измерима, ограничена, а множество W - борелевское и имеет конечную меру Лебега), то по этой теореме функция $h(\alpha(t))|J(t)|$ интегрируема по Риману на V(а кроме того, и по Лебегу - ведь она измерима, ограничена, а множество V - борелевское и имеет конечную меру Лебега) и верно равенство

$$\int\limits_{\overline{B}} h(x)dx = \int\limits_{\alpha^{-1}(\overline{B})} h(\alpha(t))|J(t)|dt,$$

или, вспоминая то, что $h = 1$ на W,

$$\int\limits_{\overline{B}} dx = \int\limits_{\alpha^{-1}(\overline{B})} |J(t)|dt.$$

Далее, т.к. $\overline{B} \setminus B \subset \partial B$ и $\mu(\overline{B} \setminus B) \leq \mu(\partial B)$, а n-мерный брус B кубируем, т.е. $\mu(\partial B) = 0$, то $\mu(\overline{B} \setminus B) = 0$. Кроме того, $\alpha^{-1}(\overline{B}) \setminus \alpha^{-1}(B) =$
$= \alpha^{-1}(\overline{B} \setminus B)$, а $\mu(\overline{B} \setminus B) = 0$. Значит, $\mu(\alpha^{-1}(\overline{B}) \setminus \alpha^{-1}(B)) = \mu(\alpha^{-1}(\overline{B} \setminus B)) = 0$. (Обоснование того факта, что непрерывно дифференцируемое отображение α^{-1} переводит множество меры нуль в множество меры нуль, см. в книге [3], том 2, гл. 2, \$ 5, лемма 5.) Отсюда следует, что $h \in L_1(B), h(\alpha(t))|J(t)| \in L_1(\alpha^{-1}(B))$ и при этом

$$\int\limits_{B} dx = \int\limits_{\alpha^{-1}(B)} |J(t)|dt.$$

Т.е. $g = I_B$ интегрируема по Лебегу на G, $g(\alpha(t))|J(t)| = I_{\alpha^{-1}(B)}|J(t)|$ интегрируема по Лебегу на D, и последнее равенство дает нам равенство (1.4) для функции $g = I_B$. Т.е. доказываемое утверждение верно для $g = I_B$.

Шаг 3. Если $W \in \mathfrak{B}(G)$, то пусть $\mathcal{A}(W)$ - множество, состоящее из всех конечных объединений n-мерных брусов, лежащих в W, а также пустого множества. Заметим, что множество n-мерных брусов, лежащих в W - полукольцо, а конечные объединения таких брусов образуют минимальное кольцо, содержащее это полукольцо; такое минимальное кольцо и есть $\mathcal{A}(W)$. Если же $W = U - n-$мерный брус, то это минимальное кольцо является алгеброй подмножеств U, ведь само U лежит в $\mathcal{A}(W)$. Можно определить $\mathcal{A}(W)$ как совокупность всевозможных конечных объединений непересекающихся n-мерных брусов, лежащих в W. (См. книгу [4], гл. 5, задачи 5.10, 5.21, 5.22, 5.31.)

В силу линейности интеграла Лебега доказываемое утверждение верно и для функций $g = I_C,\ C \in \mathcal{A}(U)$, т.е. для функций вида $g = \sum_{j=1}^{k} I_{B_j}$, где $B_j, j = \overline{1,k}$ - непересекающиеся n-мерные брусы, лежащие в U.

Заметим также, что $\mathfrak{B}(U) = \sigma(\mathcal{A}(U))$. (Это будет использовано в дальнейшем доказательстве.) Напомним **теорему об аппроксимации**:

если $\mathcal{A}$ - алгебра подмножеств $\Omega \neq \varnothing$, $\sigma(\mathcal{A})$ – σ-алгебра, порожденная ей, P-вероятностная мера на $\sigma(\mathcal{A})$, то

$$\forall \varepsilon > 0\ \forall A \in \sigma(\mathcal{A})\ \exists B \in \mathcal{A}:\ P(A \triangle B) < \varepsilon.$$

Применим эту теорему для доказательства промежуточного предложения:

Предложение. Если $E \in \mathfrak{B}(G), \mu E < \infty$, то при всех $\varepsilon > 0$ найдется $C \in \mathcal{A}(G)$, для которого $\mu(C \triangle E) < \varepsilon$.

Доказательство предложения.

Представляем G как $\bigcup\limits_{m=1}^{\infty} U_m$, где для всех $m \in \mathbb{N}$ U_m- n- мерный брус, компактно содержащийся в G. Подробнее: пусть для $x \in G$ $\varepsilon(x) > 0$ таково, что $B_{2\sqrt{n}\varepsilon(x)}(x) \subset G$ - мы используем то, что G открыто, т.е. точка $x \in G$ - внутренняя. $\{B_{\varepsilon(x)}(x)\}_{x \in G}$ - открытое покрытие G; т.к. G сепарабельно как метрическое пространство - ведь

G - подпространство метрического пространства $\mathbb{R}^n$ - то из этого покрытия можно выделить счетное подпокрытие $\{B_{\varepsilon_j}(x_j)\}_j$, где $\varepsilon_j = \varepsilon(x_j)$. Пусть, далее, $B_j = \prod\limits_{i=1}^{n}[x_{ji} - \varepsilon_j, x_{ji} + \varepsilon_j)$ – n-мерный брус, где $x_j := (x_{j1}, \ldots, x_{jn})$. Он компактно содержится в $B_{2\sqrt{n}\varepsilon_j}(x_j)$, а, значит, и в G. Но $B_{\varepsilon_j}(x_j) \subset B_j$, т.е. $G \subset \bigcup\limits_j B_j$ и, таким образом, $\{B_j\}_j$ - не более чем счетное покрытие G.

и рассматриваем $\mathcal{A} = \mathcal{A}(U_m), \sigma(\mathcal{A}) = \sigma(\mathcal{A}(U_m)) = \mathfrak{B}(U_m)$ и вероятностную меру на $\sigma(\mathcal{A})$, равную $P = P_m$, где $P_m(A) = \mu(A)/\mu(U_m)$. Пусть $\varepsilon > 0$. Разберем случаи:

1) Полученное покрытие счетно. Тогда $\bigcap\limits_{m=1}^{\infty}(E \setminus U_m) = E \setminus (\bigcup\limits_{m=1}^{\infty} U_m = E \setminus G = \varnothing$, откуда в силу непрерывности меры μ и того, что $\mu E < \infty$, следует, что $\exists M \in \mathbb{N}$: $\mu(\bigcap\limits_{m=M+1}^{\infty}(E \setminus U_m)) < \varepsilon/2$. Полагая $V := \bigcup\limits_{m=1}^{M} U_m$, имеем:

$$\mu E - \mu(E \setminus V) = \mu(E \setminus (E \setminus V)) = \mu(E \cap V) = \mu\left(\bigcap_{m=M+1}^{\infty}(E \setminus U_m)\right) < \frac{\varepsilon}{2}.$$

Но при всех $m \in \mathbb{N}$ $E \cap U_m \in \mathfrak{B}(U_m)$ и по теореме об аппроксимации $\exists\, C_m \in \mathcal{A}(U_m)$:

$$P_m((E \cap U_m) \triangle C_m) < \frac{\varepsilon}{2^{m+1}\mu(U_m)}.$$

Отсюда $\mu((E \cap U_m) \triangle C_m) < \varepsilon/2^{m+1}$. Отметим теоретико - множественный факт: если $X_m, Y_m, m = \overline{1, N}$ - множества (произвольные), то

$$(\bigcup_{m=1}^{N} X_i) \triangle (\bigcup_{m=1}^{N} Y_i) \subset \bigcup_{m=1}^{N}(X_i \triangle Y_i).$$

(См. упражнение 1.9.) Значит, имеем:

$$\mu\left(\bigcup_{m=1}^{M}(E \cap U_m) \triangle \left(\bigcup_{m=1}^{M} C_m\right)\right) \le \mu\left(\bigcup_{m=1}^{M}((E \cap U_m) \triangle C_m)\right) \le$$

$$\le \sum_{m=1}^{M} \mu((E \cap U_m) \triangle C_m) < \sum_{m=1}^{M} \frac{\varepsilon}{2^{m+1}} < \frac{\varepsilon}{2}.$$

Т.к.

$$\bigcup_{m=1}^{M}(E \cap U_m) = E \cap \left(\bigcup_{m=1}^{M} U_m\right) = E \cap V,$$

то для $C_0 := \bigcup\limits_{m=1}^{M} C_m \mu((E \cap V) \triangle C_0) < \varepsilon/2$. Заметим, что $E \triangle C_0 \subset ((E \cap V) \triangle C_0) \cup (E \setminus V)$, откуда мы получаем, что $\mu(E \triangle C_0) \subset \mu(E \cap V) \triangle C_0) + \mu(E \setminus V) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. Осталось лишь заметить, что $C_0 = \bigcup\limits_{m=1}^{M} C_m \in \mathcal{A}(G)$, $C_m \in \mathcal{A}(U_m) \subset \mathcal{A}(G)$.

2) Полученное покрытие конечно. Этот случай разбирается совершенно аналогично, но он проще первого.

Тем самым предложение доказано.■

Шаг 4. Если $E \in \mathfrak{B}(G), \mu E < \infty$, то, используя только что доказанное предложение, подберем $C^{(m)} \in \mathcal{A}(G)$, для которых при $m \in \mathbb{N}$ $\mu(E \triangle C^{(m)}) < 1/2^m$. Тогда

$$\mu\left(E \triangle \left(\bigcap_{m=1}^{\infty} C^{(m)}\right)\right) = \mu\left(\bigcap_{m=1}^{\infty}(E \triangle C^{(m)}\right) = 0,$$

а при $x \in E, x \in \bigcap\limits_{m=1}^{\infty} C^{(m)}$ $I_{C^{(m)}}(x) = 1 = I_E(x)$ для всех $m \in \mathbb{N}$; при $x \notin E, x \notin \bigcap\limits_{m=1}^{\infty} C^{(m)}$ $I_{C^{(m)}}(x) = 0 = I_E(x)$ для всех $m \in \mathbb{N}$. Мы заключаем, что $\lim\limits_{m\to\infty} I_{C^{(m)}} = I_E$ почти всюду. Кроме того,

$$\bigcup_{m=1}^{\infty} C^{(m)} = \left(\bigcup_{m=1}^{\infty}(C^{(m)} \setminus E)\right) \cup E \subset \left(\bigcup_{m=1}^{\infty}(C^{(m)} \triangle E)\right) \cup E,$$

т.е.

$$\mu(\bigcup_{m=1}^{\infty} C^{(m)}) \leq \sum_{m=1}^{\infty} \mu(C^{(m)} \triangle E) + \mu E < \sum_{m=1}^{\infty} \frac{1}{2^m} + \mu E = 1 + \mu E.$$

Значит, $I_H \in \mathrm{Ł}_1(G)$, где $H := \bigcup\limits_{m=1}^{\infty} C^{(m)}$. Но $0 \leq I_{C^{(m)}} \leq I_H$ для всех $m \in \mathbb{N}$.

Применяя теорему Лебега о мажорируемой сходимости, имеем: доказываемый факт верен и для функций $g = I_E, E \in \mathfrak{B}(G), \mu E < \infty$.

Отсюда получаем, в силу линейности интеграла Лебега, что доказываемое утверждение верно и для простой g, т.е. являющейся конечной линейной комбинацией индикаторов $I_E, E \in \mathfrak{B}(G), \mu E < \infty$. Но любую неотрицательную борелевскую функцию $g : G \to \mathbb{R}$ можно представить в виде предела п.в. монотонной последовательности простых функций. Применяя теорему Б.Леви о монотонной сходимости, доказываем данный факт и для неотрицательных борелевских функций $g : G \to \mathbb{R}$.

А любую борелевскую функцию $g : G \to \mathbb{R}$ произвольного знака можно представить в виде разности $g_+ - g_-, g_+ = \max(g, 0), g_- = \max(-g, 0)$ - неотрицательные борелевские функции. Таким образом, утверждение полностью доказано. См. также упражнение 1.10. ■

Приложение 2. Напоминание некоторых основных понятий математической статистики.

Мы предполагаем, что наш читатель знаком с классической математической статистикой - такое знакомство существенно необходимо для чтения главы 2 и последующих. Тем не менее, напомним хотя бы частично фундаментальные понятия этой науки. Надеемся, что нижеприведенные материалы облегчат понимание как математической статистики вообще, так и многомерного гауссовского анализа в частности.

Определение A2.1. Базовой моделью математической статистики является **статистическое пространство**. Это - тройка $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathfrak{X} \neq \varnothing$ - какое-то множество, называемое **пространством наблюдений** (а его элементы $X \in \mathfrak{X}$, разумеется, называются **наблюдениями**); $\mathfrak{F}$ - $\sigma-$алгебра на этом множестве $\mathfrak{X}$, т.е. $(\mathfrak{X}, \mathfrak{F})$ - измеримое пространство; а $\mathcal{P}$ - некоторое непустое множество (разумеется, не обязательно всевозможных) вероятностных мер, заданных на $\mathfrak{F}$. Мы считаем, что какая-то (только одна) из этих мер истинна, т.е. реально, в действительности задана на $\mathfrak{F}$, но нам неизвестно, какова же именно эта мера. (Однако заведомо известно, что она лежит в этом семействе $\mathcal{P}$.)

Основная задача математической статистики - по наблюдению $X \in \mathfrak{X}$ сделать выводы о том, какова же истинная вероятностная мера $\mathbf{P} \in \mathcal{P}$. Что это значит - "сделать выводы"? Точно мы, по-видимому, не сможем указать эту искомую меру $\mathbf{P}$. (Почти во всех статистических моделях это на самом деле невозможно.) Но можно попытаться разрешить эту задачу приближенно. Значит, нужно формализовать это понятие - "приближенно". И оказывается, что это можно сделать по-разному - мы увидим это ниже.

Можно сказать так: статистика занимается вопросами, в известном смысле противоположными вопросам теории вероятностей. Ведь последняя по заданной вероятностной мере $\mathbf{P}$ на измеримом пространстве $(\Omega, \mathfrak{F})$ делает выводы о вероятностях событий, связанных с какими-то случайными величинами, векторами и т.д. - иногда более простых, а иногда довольно сложных, как, например, в законе больших чисел или центральной предельной теореме. А статистика занимается тем, что по данной величине (наблюдению) $X \in \mathfrak{X}$ пытается сколько-нибудь точно восстановить вероятностную меру $\mathbf{P} \in \mathcal{P}$ на $(\mathfrak{X}, \mathfrak{F})$.

Часто множество $\mathcal{P}$ параметризовано, т.е. $\mathcal{P} = \{\mathbf{P}_\theta | \theta \in \Theta\}$, где $\Theta \neq \varnothing$ - некоторое множество индексов, причем $\mathbf{P}_\theta \neq \mathbf{P}_{\theta'}$ при любых $\theta, \theta' \in \Theta$, $\theta \neq \theta'$. (Проще говоря, каждой мере $\mathbf{P} \in \mathcal{P}$ взаимно-однозначно сопоставляется индекс $\theta \in \Theta$.) Разумеется, тогда существует какое-то неизвестное нам истинное значение индекса $\theta \in \Theta$ (только одно). И тогда сделать вывод об истинной вероятностной мере $\mathbf{P}$ - равносильно тому, чтобы сделать вывод об истинном значении индекса $\theta \in \Theta$. Вы уже знаете, что это можно сделать разными способами, а именно тремя:

- построить точечную оценку;
- построить доверительное множество;
- проверить гипотезу.

В этой главе мы остановимся на точечных оценках для конкретного случая этой общей модели - для параметров многомерного нормального распределения; а позже, в следующих главах, пойдет речь о двух других способах.

Проиллюстрируем абстрактное понятие статистического пространства на конкретном примере, который, кстати, для нас еще очень пригодится. Пусть мы делаем n последовательных независимых наблюдений $X_1, \ldots, X_n$, и нам заведомо известно, что для $k = \overline{1,n}$ $X_k \backsim \mathcal{N}(l_k, \sigma^2)$. При этом мы не знаем ни параметров $l_k, k = \overline{1,n}$ (будем также записывать их в виде вектора - столбца $l = (l_1, \ldots, l_n)^T$), ни среднеквадратического отклонения σ. Правда, считается известным, что $\sigma > 0$. Ставится естественная задача: сделать выводы об этих неизвестных параметрах l и σ. Но как это формализовать - как построить статистическое пространство для такой ситуации?

Можно сделать это специальным приемом, используемым в теории вероятностей, а именно: непосредственно задать случайный вектор $X = (X_1, \dots, X_n)^T$. Что это означает?

Определение A2.2. Пусть дана функция $\xi : \Omega \to S$, заданная на измеримом пространстве $(\Omega, \mathcal{F})$ со значениями в измеримом пространстве $(S, \mathcal{B})$, являющаяся $\mathcal{B}/\mathcal{F}$-измеримой, т.е. при $B \in \mathcal{B}$ $\xi^{-1}(B) \in \mathcal{F}$. Она называется **непосредственно заданной**, если $\Omega = S, \mathcal{F} = \mathcal{B}$ и ξ - тождественное отображение, т.е. при $\omega \in \Omega$ $\xi(\omega) = \omega$.

Такие функции иногда удобны в теоретических доказательствах существования. Например, если вы изучали доказательство теоремы Колмогорова о существовании случайного процесса, то знаете: там искомый процесс строится именно как непосредственно заданный. И здесь мы построим случайный вектор X как непосредственно заданный - правда, ситуация у нас несравненно проще, чем в теореме Колмогорова. А именно: пусть $\mathfrak{X} := \mathbb{R}^n, \mathfrak{F} := \mathfrak{B}(\mathbb{R}^n)$ - борелевская $\sigma-$алгебра на $\mathbb{R}^n$. Пусть, далее, $\Theta := \{(l, \sigma) | l \in \mathbb{R}^n, \sigma > 0\}$ и для $\theta = (l, \sigma) \in \Theta$ имеем:

$$\mathbf{P}_\theta = \mathcal{N}(l_1, \sigma^2) \bigotimes \mathcal{N}(l_2, \sigma^2) \bigotimes \dots \bigotimes \mathcal{N}(l_n, \sigma^2). \qquad (A2.1)$$

(Знак $\bigotimes$ обозначает прямое произведение мер; это понятие считается известным, см., впрочем, [2], гл. 2, \$6 или [4], гл. 12, или любое руководство по теории меры и интеграла.)

И наконец, пусть $X : \Omega \to \mathbb{R}^n$ - непосредственно заданный случайный вектор, а проще говоря - тождественное отображение $\mathbb{R}^n \to \mathbb{R}^n$, и пусть при $k = \overline{1, n}$ X_k - его k-я компонента, т.е. функция $\Omega \to \mathbb{R}$, заданная так: для $x = (x_1, \dots, x_n)^T \in \mathbb{R}$ $X_k(x) := x_k$. Тогда $X_k, k = \overline{1, n}$ - независимые случайные величины, причем если $\theta = (l, \sigma)$ - истинное значение индекса, $l = (l_1, \dots, l_k)^T$, то $X_k \backsim \mathcal{N}(l_k, \sigma^2)$. (Обосновать это, исходя из того, что $\mathbf{P}_\theta$ - это прямое произведение одномерных мер!) Так мы построили статистическое пространство, реализующее возникшую ситуацию независимых случайных величин $X_1, \dots, X_n$, распределенных следующим образом: $X_k \backsim \mathcal{N}(l_k, \sigma^2)$.

Случайный вектор X имеет распределение $\mathcal{N}_n(l, \sigma^2 I_n)$ (проверьте это сами!) Такой вектор есть основная составляющая линейной гауссовской модели - классической главы математической статистики.

Если в предыдущей ситуации нам заведомо известно, что $l \in L$, где L - некоторое фиксированное линейное подпространство пространства $\mathbb{R}^n$, а мы хотим сделать выводы о неизвестных параметрах l и σ, уже зная этот факт (т.е. что $l \in L$), то соответствующее статистическое пространство строится так: $\mathfrak{X} := \mathbb{R}^n$, $\mathfrak{F} := \mathfrak{B}(\mathbb{R}^n)$, $\Theta := \{(l, \sigma) | l \in L, \sigma > 0\}$ и для $\theta := (l, \sigma) \in \Theta$ мера $\mathbf{P}_\theta$ определяется так же, как и для предыдущего статистического пространства - т.е. по формуле (A2.1).

Теперь приступим к введению важнейшего понятия статистики - а именно, понятия оценки параметра распределения.

Пусть $\mathfrak{T}$ - некоторое непустое множество, $\tau : \Theta \to \mathfrak{T}$ - заданная функция. Мы можем говорить, что существует некоторое (единственное) истинное значение этой функции - разумеется, это $\tau(\theta_0)$, где θ_0 - истинное значение параметра θ. Будем делать выводы об этом истинном значении функции.

Определение A2.3. Оценка истинного значения τ - это любая функция $\delta : \mathfrak{X} \to \mathfrak{T}$, достаточно хорошо приближающая неизвестное истинное значение τ. Часто говорят просто: **оценка** τ. Частный случай этого понятия - оценка τ для $\mathfrak{T} := \Theta$

и для τ - тождественного отображения Θ. Такая функция $\delta : \mathfrak{X} \to \mathfrak{T}$ называется **оценкой истинного значения** θ. Это, как легко видеть

Для чего служит такое определение? Мы получили наблюдение $X \in \mathfrak{X}$ и хотели бы по нему восстановить истинное значение τ - пусть не точно, хотя бы приближенно. Это можно сделать, если мы располагаем оценкой τ.

Но последнее определение, естественно, не может считаться математически строгим. Что означают слова "достаточно хорошо приближающая"? Необходимо как-то математически формализовать это.

Определение А2.4. Пусть $\mathfrak{X} \subset \mathbb{R}^n$ - борелевское подмножество $\mathbb{R}^n$, $\mathfrak{F} = \mathfrak{B}(\mathfrak{X})$ - σ-алгебра всех борелевских подмножеств $\mathfrak{X}$ и при всех $\theta \in \Theta$ $\mathbf{P}_\theta$ - вероятностная мера, абсолютно непрерывная относительно классической меры Лебега на $\mathfrak{B}(\mathfrak{X})$, т.е. имеющая плотность (или, более точно, производную Радона - Никодима относительно этой меры Лебега) $f_\theta : \mathfrak{X} \to \mathbb{R}^n$. (Эти понятия - абсолютная непрерывность, производная Радона - Никодима - считаются известными из курса действительного анализа.) Вместо $f_\theta(x)$ будем писать так: $f(x, \theta)$. Тогда при фиксированном $X \in \mathfrak{X}$ функция $\Theta \to \mathbb{R}^n$, $\theta \mapsto f(X, \theta)$ называется **правдоподобием**.

Определение А2.5. Если при любом $X \in \mathfrak{X}$ правдоподобие, соответствующее этому X, имеет строгий глобальный максимум - или, более формально, найдется такое значение $\theta_X \in \Theta$, что при всех $\theta \in \Theta$ $f(X, \theta) \leq f(X, \theta_X)$, причем равенство достигается только при $\theta = \theta_X$, то функция $\delta : \mathfrak{X} \to \Theta$, сопоставляющая наблюдению $X \in \mathfrak{X}$ значение θ_X, называется **оценкой наибольшего правдоподобия** (или **оценкой максимального правдоподобия**).

Почему же так определенная функция $\delta : \mathfrak{X} \to \Theta$ действительно является оценкой, т.е. почему оправдано название "оценка наибольшего правдоподобия"? По нестрогому определению А2.3, оценка должна хорошо