

Математическая статистика

24 декабря 2014 г.

Вступление

В курсе теории вероятностей мы рассматривали вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайные величины $(\xi, \eta, \varkappa, \dots)$ с известными распределениями. Нас интересовало поведение этих случайных величин. Математическая статистика исследует характер распределений случайных величин.

Пусть x_1, \dots, x_n — независимые одинаково распределённые случайные величины с неизвестной функцией распределения F . Такой набор случайных величин называют выборкой из распределения F . В каждом конкретном вероятностном эксперименте мы наблюдаем некоторую реализацию выборки, то есть набор значений $x_1(\omega), \dots, x_n(\omega)$ при некотором $\omega \in \Omega$. Математическая статистика занимается восстановлением характера распределений из реализаций выборок. Там, где это не вызовет путаницы, мы будем опускать символ ω в записи $x_1(\omega), \dots, x_n(\omega)$ и употреблять термин “выборка” в значении “реализация выборки”.

В настоящем курсе будут рассматриваться такие задачи:

1. x_1, \dots, x_n — значения, полученные в результате измерения случайной величины. Какое распределение имеет эта величина?
2. Наблюдается последовательность x_1, \dots, x_n . Являются ли величины x_1, \dots, x_n реализацией независимых одинаково распределённых случайных величин?
3. Есть две гипотезы относительно типа распределения x_1, \dots, x_n . Какая из гипотез верна?
4. Существует ли зависимость наблюдаемых значений от каких-либо других величин? Каков вид этой зависимости?

Глава 1

ОСНОВЫ

1.1 Методы оценок параметров распределения

Определение 1.1.1

Набор x_1, \dots, x_n независимых одинаково распределённых случайных величин с неизвестной функцией распределения F называется выборкой из распределения F объёма n .

Цель — найти F или сказать что-то о её свойствах.

1.1.1 Эмпирическая функция распределения

Определение 1.1.2: Эмпирическая функция распределения

Эмпирической (выборочной) функцией распределения, построенной по выборке x_1, \dots, x_n , называется функция

$$F_n(x) = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \leq x)$$

Пример 1.1.3: Эмпирическая функция распределения

Пускай числа 0.5, 0.2, 0.1, 0.2 — выборка из неизвестного распределения.

Для удобства выстроим числа в порядке их возрастания: 0.1, 0.2, 0.2, 0.5.

Видим, что слева от 0.1 эмпирическая функция распределения будет равна нулю.

В точке 0.1 произойдёт скачок на $\frac{1}{4}$, так как только один элемент выборки не превышает это значение.

В точке 0.2 будет скачок на $\frac{2}{4}$, а значение самой функции будет равно $\frac{3}{4}$, так как есть три элемента, не превышающих значение 0.2 — один равный 0.1 и два равных 0.2.

Последний скачок произойдёт в точке 0.5 и тут функция достигнет значения 1, так как нет элемента выборки, превышающего 0.5.

Эмпирическая функция будет выглядеть так, как показано на рисунке 1.1.

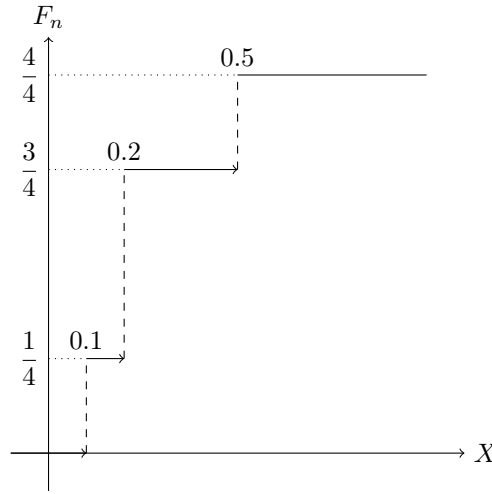


Рис. 1.1: Эмпирическая функция распределения для примера 1.1.3

Неизвестная функция распределения $F(x)$ может быть сколь угодно точно восстановлена по выборке достаточно большого объёма.

Теорема 1.1.4

$$\forall x \in \mathbb{R} : \quad F_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} F(x)$$

Доказательство. Вспомним, чему равна эмпирическая функция распределения

$$F_n(x) = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \leq x).$$

Заметим, что индикаторы $\mathbb{1}(x_k \leq x)$ являются независимыми одинаково распределёнными случайными величинами, а функцию распределения $F(x)$ можно записать следующим образом

$$F(x) = \mathbb{P}\{x_1 \leq x\} = \mathbb{M} \mathbb{1}(x_1 \leq x)$$

По усиленному закону больших чисел

$$F_n(x) = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \leq x) \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{M} \mathbb{1}(x_1 \leq x) = F(x)$$

Теорема доказана. \square

В приведённом утверждении речь шла о сходимости F_n к F в фиксированной точке. Сходятся ли при этом интегралы $\int_{\mathbb{R}} \varphi(x) dF_n(x)$ к $\int_{\mathbb{R}} \varphi(x) dF(x)$ для “хороших” функций φ ? Мы помним, что из поточечной сходимости не следует сходимости интегралов, но в нашем случае функции распределения F_n имеют одинаковый специальный вид, поэтому справедлива следующая теорема

Теорема 1.1.5

Пусть $\varphi \in C(\mathbb{R})$ — интегрируемая по F функция. Тогда

$$\int_{\mathbb{R}} \varphi(x) dF_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} \int_{\mathbb{R}} \varphi(x) dF(x)$$

Доказательство. Вспомним, что интеграл Лебега в нашем случае — сумма вида

$$\int_{\mathbb{R}} \varphi(x) dF_n(x) = \frac{1}{n} \cdot \sum_{k=1}^n \varphi(x_k)$$

Тогда, применив усиленный закон больших чисел, получим

$$\frac{1}{n} \cdot \sum_{k=1}^n \varphi(x_k) \xrightarrow[n \rightarrow \infty]{a.s.} M \varphi(x_1) = \int_{\mathbb{R}} \varphi(x) dF(x)$$

Что и требовалось доказать. \square

Пример 1.1.6

Возьмём $\varphi(x) = x^m$. Воспользовавшись теоремой 1.1.5, получаем

$$\frac{1}{n} \cdot \sum_{k=1}^n \varphi(x_k) \xrightarrow[n \rightarrow \infty]{a.s.} \int_{\mathbb{R}} x^m dF(x)$$

Поскольку в правой части стоит m -й момент распределения F , сумме в левой части называют m -м выборочным моментом.

1.1.2 Гистограмма

Как можно попытаться отследить плотность распределения? Допустим, F имеет хорошую (непрерывную) плотность. Как тогда из F получить p ? Мы знаем, что $F' = p$, но производная функции F_n почти везде будет равна нулю как производная ступенчатой функции. С другой стороны, по теореме Лагранжа для $a < b$ имеем

$$\exists z \in [a; b] : \quad \frac{F(b) - F(a)}{b - a} = \frac{dF(z)}{dz} = p(z)$$

Поскольку функция p непрерывна, то при сужении отрезка $[a; b]$ получим

$$p(x) \approx \frac{F(b) - F(a)}{b - a}, \quad x \in [a; b] \quad (1.1)$$

Это соображение легло в основу построения гистограммы. Разобьём числовую прямую на $m + 1$ полуинтервалов (два из них бесконечные) точками $a_1 < \dots < a_m$. Заменим на каждом полуинтервале

$$I_j = (a_j, a_{j+1}], \quad j = \overline{1, m-1}$$

функцию p выражением 1.1, а на $(-\infty; a_1]$ и $(a_m; +\infty)$ положим равной нулю. Определим последовательность функций $q_n(y)$, заменив $F(x)$ на $F_n(x)$ в предыдущем определении

$$q_n(y) = \sum_{j=1}^{m-1} \frac{F_n(a_{j+1}) - F_n(a_j)}{a_{j+1} - a_j} \cdot \mathbb{1}(y \in I_j) \quad (1.2)$$

Отметим, что q_n сходится к q почти наверное, так как сходятся почти наверное функции распределения согласно теореме 1.1.4

$$q_n(y) \xrightarrow[n \rightarrow \infty]{a.s.} q(y)$$

Уменьшим размеры полуинтервалов I_k

$$\max_{k=\overline{1, m-1}} |I_k| \rightarrow 0, \quad a_1 \rightarrow -\infty, a_m \rightarrow +\infty$$

Приходим к следующему выводу

$$\forall y \in \mathbb{R} : \quad q(y) \xrightarrow{m \rightarrow \infty} p(y)$$

Подставив в 1.2 выражение для F_n , введём определение

Определение 1.1.7: Гистограмма

Гистограммой выборки x_1, \dots, x_n называется функция

$$q_n(y) = \sum_{j=1}^m \frac{\nu_j}{n \cdot |I_j|} \cdot \mathbb{1}(y \in I_j)$$

ν_j — количество элементов выборки, попавших в полуинтервал I_j

$$\nu_j = \sum_{k=1}^n \mathbb{1}(x_k \in I_j).$$

1.1.3 Оценка неизвестных параметров

Пусть x_1, \dots, x_n — выборка из распределения F_θ , где θ — неизвестный параметр из множества Θ .

Пример 1.1.8

Пусть имеем нормальное распределение с известной дисперсией $\sigma^2 = 1$ и неизвестным математическим ожиданием $a \sim N(a, 1)$. Тогда θ — математическое ожидание a , а множество неизвестных параметров будет множеством действительных чисел $\Theta = \mathbb{R}$.

Пример 1.1.9

Есть нормальное распределение, в котором неизвестны оба параметра. Тогда θ будет парой (a, σ^2) , а Θ будет декартовым произведением $\mathbb{R} \times \mathbb{R}_+$.

Цель данного пункта — предложить процедуры построения функций от выборки, значения которых будут заменять неизвестные параметры.

Определение 1.1.10: Статистика

Статистикой называют функцию S от выборки $X = (x_1, x_2, \dots, x_n)$

$$S(X) = S(x_1, x_2, \dots, x_n)$$

Определение 1.1.11: Оценка

Статистику, значение которой заменяет неизвестный параметр, называют оценкой этого параметра

Пример 1.1.12

Пусть $\{x_i\}$ — выборка из распределения Бернулли, то есть набор независимых одинаково распределённых случайных величин, причём

$$\mathbb{P}(x_i = 1) = p,$$

$$\mathbb{P}(x_i = 0) = 1 - p$$

Неизвестным параметром будет величина p

$$\theta = p \in [0; 1] = \Theta$$

Рассмотрим разные оценки \hat{p}

$$\begin{aligned}\hat{p}_1 &= \frac{1}{n} \sum_{k=1}^n x_k \\ \hat{p}_2 &= x_1 \\ \hat{p}_3 &= \frac{2}{n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} x_k\end{aligned}$$

Замечание 1.1.13

Поскольку \hat{p} — случайная величина, то может оказаться, что она не равна настоящему параметру p

$$\mathbb{P} \{ \hat{p} = p \} = 0$$

Определение 1.1.14: Состоятельная оценка

Оценка $\hat{\theta}$ называется состоятельной, если стремится к истинному значению θ по вероятности

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta$$

Определение 1.1.15: Сильно состоятельная оценка

Оценка $\hat{\theta}$ называется сильно состоятельной, если она стремится к истинному значению θ почти наверное

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{a.s.} \theta$$

Пример 1.1.16

Оценка \hat{p}_1 из примера 1.1.12 является сильно состоятельной согласно усиленному закону больших чисел. Оценка \hat{p}_2 в свою очередь не является состоятельной.

Определение 1.1.17: Несмещённая оценка

Оценка $\hat{\theta}$ несмещённая, если

$$\forall \theta \in \Theta : \quad \mathbb{M}_{\theta} \hat{\theta} = \theta$$

Замечание 1.1.18

Несмещённая оценка существует не всегда

Пример 1.1.19

Пусть x_1, \dots, x_n — выборка из распределения Бернулли

$$x_k = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$$

Обозначим сумму элементов выборки S_n

$$S_n = x_1 + \dots + x_n$$

Тогда S_n — количество единиц, а $n - S_n$ — количество нулей, встретившихся в выборке. Не существует несмещённой оценки для e^p при $p \in (0; 1)$. Для проверки предположим противное. Пусть $\hat{\theta}$ — несмещённая оценка для e^p

$$\hat{\theta} = f(x_1, \dots, x_n)$$

Тогда справедливо следующее равенство

$$\forall p \in [0; 1] : \quad M_p f(x_1, \dots, x_n) = e^p$$

Или

$$\sum_{x_i \in \{0,1\}} f(x_1, \dots, x_n) \cdot p^{S_n} \cdot (1-p)^{n-S_n} = e^p$$

Слева стоит полином степени не выше n , который не может совпадать на интервале $(0; 1)$ с показательной функцией. То есть не существует несмещённой оценки для e^p .

Определение 1.1.20: Оптимальная оценка

Несмещённая оценка $\hat{\theta} \in K$ называется оптимальной в классе квадратично интегрируемых оценок K , если для всякой другой несмещённой оценки $\tilde{\theta} \in K$

$$\forall \theta \in \Theta : \quad D_{\theta} \hat{\theta} \leq D_{\theta} \tilde{\theta}$$

или же

$$\forall \theta \in \Theta : \quad M_{\theta} (\hat{\theta} - \theta)^2 \leq M_{\theta} (\tilde{\theta} - \theta)^2.$$

1.1.4 Выборочные оценки. Метод моментов

Как правило, неизвестный параметр может быть выражен через моменты распределения. Например:

1. Нормальное распределение $N(a, \sigma^2)$. В нём параметр a является средним, а параметр σ^2 — дисперсией
2. Пуассоновское распределение $Poi(\lambda)$. Тут параметр λ является и средним, и дисперсией
3. Экспоненциальное распределение $Exp(\lambda)$. Мы помним, что $M x_1 = \frac{1}{\lambda}$, $D x_1 = \frac{1}{\lambda^2}$, из чего следует $\lambda = \frac{1}{M x_1} = \frac{1}{\sqrt{D x_1}}$

Заменяя моменты распределения выборочными, можно получать оценки соответствующих параметров. В общем случае рецепт построения оценки параметра θ выглядит так. Пусть φ — непрерывная функция на \mathbb{R} , рассмотрим интеграл

$$\int_{\mathbb{R}} \varphi(x) dF_{\theta}(x) = g(\theta)$$

как функцию θ .

Заменяя F_{θ} эмпирической функцией распределения, получим уравнение на оценку $\hat{\theta}$

$$g(\hat{\theta}) = \int_{\mathbb{R}} \varphi(x) dF_n(x), \quad (1.3)$$

Пример 1.1.21

Если θ — среднее, то $\varphi(x) = x$

$$\int_{-\infty}^{+\infty} x dF_{\theta}(x) = \theta = g(\theta)$$

и после замены F_{θ} на F_n получим $\hat{\theta} = \bar{x}$.

Теорема 1.1.22: Оценка метода моментов

Пусть φ такова, что g — монотонная и непрерывная функция. Тогда оценка

$$\hat{\theta} = g^{-1} \left(\int_{\mathbb{R}} \varphi(x) dF_n(x) \right)$$

существует и является сильно состоятельной.

Доказательство. Поскольку функция g непрерывна и монотонна, то она имеет обратную g^{-1} . Применим обратную функцию к обеим частям уравнения (1.3)

$$\hat{\theta} = g^{-1} \left(\int_{\mathbb{R}} \varphi(x) dF_n(x) \right)$$

Воспользуемся теоремой 1.1.5 о сходимости интегралов по эмпирическим функциям распределения

$$\int_{\mathbb{R}} \varphi(x) dF_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} \int_{\mathbb{R}} \varphi(x) dF_{\theta}(x)$$

Функция g^{-1} непрерывна, следовательно,

$$\hat{\theta} = g^{-1} \left(\int_{\mathbb{R}} \varphi(x) dF_n(x) \right) \xrightarrow[n \rightarrow \infty]{a.s.} g^{-1} \left(\int_{\mathbb{R}} \varphi(x) dF_{\theta}(x) \right) = \theta$$

Теорема доказана. \square

Примерами интегралов по эмпирической функции распределения, часто встречающимся на практике, являются выборочные моменты.

Определение 1.1.23: Выборочное среднее

Выборочное среднее обозначается через \bar{x} и считается по следующей формуле

$$\bar{x} = \int_{\mathbb{R}} x dF_n(x) = \frac{1}{n} \cdot \sum_{k=1}^n x_k$$

Определение 1.1.24: Выборочная дисперсия

Выборочная дисперсия $\overline{\sigma^2}$ считается по формуле

$$\overline{\sigma^2} = \frac{1}{n-1} \cdot \sum_{k=1}^n (x_k - \bar{x})^2$$

Замечание 1.1.25

Очень полезным является равенство

$$\sum_{k=1}^n (x_k - \bar{x})^2 = \sum_{k=1}^n (x_k^2 - \bar{x}^2)$$

Вывод равенства. Действительно,

$$\begin{aligned} \sum_{k=1}^n (x_k - \bar{x})^2 &= \sum_{k=1}^n (x_k^2 - 2 \cdot x_k \cdot \bar{x} + \bar{x}^2) = \sum_{k=1}^n x_k^2 - 2 \cdot \sum_{k=1}^n x_k \cdot \bar{x} + \sum_{k=1}^n \bar{x}^2 \\ &= \sum_{k=1}^n x_k^2 - 2 \cdot n \cdot \bar{x}^2 + \sum_{k=1}^n \bar{x}^2 = \sum_{k=1}^n (x_k^2 - \bar{x}^2) \end{aligned}$$

Что и требовалось показать. \square

Замечание 1.1.26

Оценка метода моментов является сильно состоятельной, но необязательно несмещённой.

Если воспользоваться лишь методом моментов, то оценка дисперсии будет выглядеть так

$$\widehat{\sigma^2} = \int_{\mathbb{R}} x^2 dF_n(x) - \left(\int_{\mathbb{R}} x dF_n(x) \right)^2 = \frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})^2$$

Посчитаем математическое ожидание этой оценки

$$\mathbb{M} \widehat{\sigma^2} = \frac{1}{n} \cdot \mathbb{M} \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{M} x_k^2 - \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{M} \bar{x}^2 = \mathbb{M} x_1^2 - \mathbb{M} \bar{x}^2$$

Имеем

$$\begin{aligned} \mathbb{M} x_1^2 - \mathbb{M} \bar{x}^2 &= \mathbb{M} x_1^2 - \mathbb{D} \bar{x} - (\mathbb{M} \bar{x})^2 = \mathbb{M} x_1^2 - \mathbb{D} \bar{x} - (\mathbb{M} x_1)^2 \\ &= \mathbb{D} x_1 - \frac{1}{n^2} \cdot \mathbb{D} \sum_{k=1}^n x_k = \frac{n-1}{n} \cdot \mathbb{D} x_1 \end{aligned}$$

Мы помним, что x_1, \dots, x_n — независимые одинаково распределённые случайные величины, — этот факт и позволяет нам разбить дисперсию суммы на сумму дисперсий.

То есть чтобы оценка была несмещённой, нужно умножить её на $\frac{n}{n-1}$, — тогда математическое ожидание оценки дисперсии будет равно самой дисперсии. Именно так и поступим

$$\overline{\sigma^2} = \frac{n}{n-1} \cdot \widehat{\sigma^2} = \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{1}{n-1} \cdot \sum_{k=1}^n (x_k - \bar{x})^2$$

Оценка дисперсии, полученная методом моментов, называется смещённой выборочной дисперсией, а оптимизированная оценка $\overline{\sigma^2}$ называется несмещённой выборочной дисперсией.

1.2 Свойства оценок

1.2.1 Неравенство Рао-Крамера

Теорема 1.2.1: Колмогорова (теорема единственности)

Оптимальная оценка единственная или её нет вообще

Доказательство. Допустим, есть две разные оптимальные оценки θ_1 и θ_2 параметра θ . Напомним, что оптимальные оценки параметра θ являются

несмещёнными, а для любой несмещённой оценки $\hat{\theta}$

$$\forall \theta \in \Theta : \quad \begin{cases} D_{\theta} \theta_1 \leq D_{\theta} \hat{\theta} \\ D_{\theta} \theta_2 \leq D_{\theta} \hat{\theta} \end{cases}$$

Поскольку неравенство выполняется для каждой несмещённой оценки $\hat{\theta}$, а оценки θ_1 и θ_2 являются несмещёнными, то можем их и поставить в неравенство в роли $\hat{\theta}$

$$\forall \theta \in \Theta : \quad \begin{cases} D_{\theta} \theta_1 \leq D_{\theta} \theta_2 \\ D_{\theta} \theta_2 \leq D_{\theta} \theta_1 \end{cases}$$

А это возможно только если дисперсии этих оценок равны. Обозначим эту дисперсию через $\sigma^2(\theta)$

$$D_{\theta} \theta_1 = D_{\theta} \theta_2 = \sigma^2(\theta)$$

Если $\sigma^2(\theta) = 0$, то $\theta_1 = \theta_2 = \theta$ почти всюду. Рассмотрим случай $\sigma^2(\theta) > 0$.

Возьмём несмещённую оценку $\tilde{\theta}$, равную среднеарифметическому оценок θ_1 и θ_2

$$\tilde{\theta} = \frac{1}{2} \cdot \theta_1 + \frac{1}{2} \cdot \theta_2$$

Тогда по определению θ_1 и θ_2 получаем, что дисперсия новой оценки не меньше, чем у оптимальных

$$D_{\theta} \tilde{\theta} \geq \sigma^2(\theta) \quad (1.4)$$

Посчитаем дисперсию оценки $\tilde{\theta}$

$$\begin{aligned} D_{\theta} \tilde{\theta} &= M_{\theta} \left(\tilde{\theta} - \theta \right)^2 = M_{\theta} \left[\frac{1}{2} \cdot (\theta_1 - \theta) + \frac{1}{2} \cdot (\theta_2 - \theta) \right]^2 = \\ &= \frac{1}{4} \cdot D_{\theta} \theta_1 + \frac{1}{4} \cdot D_{\theta} \theta_2 + \frac{1}{2} \cdot M_{\theta} [(\theta_1 - \theta) \cdot (\theta_2 - \theta)] \end{aligned}$$

Воспользуемся неравенством Коши

$$\begin{aligned} M_{\theta} [(\theta_1 - \theta) \cdot (\theta_2 - \theta)] &\leq \sqrt{M_{\theta} (\theta_1 - \theta)^2 \cdot M_{\theta} (\theta_2 - \theta)^2} = \\ &= \sqrt{D_{\theta} \theta_1 \cdot D_{\theta} \theta_2} = \sigma^2(\theta) \end{aligned} \quad (1.5)$$

Вернёмся к вычислению дисперсии оценки $\tilde{\theta}$

$$\begin{aligned} \frac{1}{4} \cdot D_{\theta} \theta_1 + \frac{1}{4} \cdot D_{\theta} \theta_2 + \frac{1}{2} \cdot M_{\theta} [(\theta_1 - \theta) \cdot (\theta_2 - \theta)] &\leq \\ &\leq \frac{1}{2} \cdot \sigma^2(\theta) + \frac{1}{2} \cdot \sqrt{\sigma^2(\theta) \cdot \sigma^2(\theta)} = \sigma^2(\theta) \end{aligned}$$

То есть дисперсия оценки $\tilde{\theta}$ не больше дисперсии введённой оптимальной оценки

$$D_{\theta} \tilde{\theta} \leq \sigma^2(\theta) \quad (1.6)$$

Воспользовавшись неравенствами (1.4) и (1.6), получаем равенство

$$D_{\theta} \tilde{\theta} = \sigma^2(\theta)$$

Это значит, что в неравенстве (1.5) в данном случае тоже выходит равенство

$$M_{\theta}[(\theta_1 - \theta) \cdot (\theta_2 - \theta)] = \sqrt{M_{\theta}(\theta_1 - \theta)^2} \cdot \sqrt{M_{\theta}(\theta_2 - \theta)^2}$$

Для дальнейших рассуждений полезной оказывается аналогия с векторами в пространстве. Скалярное произведение двух векторов равно произведению их модулей тогда и только тогда, когда они сонаправлены

$$(\vec{a}, \vec{b}) = \|\vec{a}\| \cdot \|\vec{b}\| \Leftrightarrow \vec{a} \uparrow \vec{b} \quad (1.7)$$

Для ненулевых векторов это означает

$$\frac{\vec{a}}{\|\vec{a}\|} = \frac{\vec{b}}{\|\vec{b}\|} \quad (1.8)$$

Мы знаем, что в пространстве интегрируемых с квадратом случайных величин также может быть введено скалярное произведение $\langle \xi, \eta \rangle = M[\xi \cdot \eta]$ и норма $\|\xi\| = \sqrt{M\xi^2}$. Для пространства с такой метрикой также справедливы утверждения. То есть, когда $\xi \neq 0$ и $\eta \neq 0$, справедливо следующее (1.7) и (1.8):

$$\langle \xi, \eta \rangle = \|\xi\| \Leftrightarrow \frac{\xi}{\|\xi\|} = \frac{\eta}{\|\eta\|} \quad (1.9)$$

Соответственно, равенство

$$M_{\theta}[(\theta_1 - \theta) \cdot (\theta_2 - \theta)] = \sqrt{M_{\theta}(\theta_1 - \theta)^2} \cdot \sqrt{M_{\theta}(\theta_2 - \theta)^2}$$

влечёт за собой

$$\frac{\theta_1 - \theta}{\sqrt{M_{\theta}(\theta_1 - \theta)^2}} = \frac{\theta_2 - \theta}{\sqrt{M_{\theta}(\theta_2 - \theta)^2}}$$

Поскольку $M_{\theta}(\theta_1 - \theta)^2 = M_{\theta}(\theta_2 - \theta)^2$, получим $\theta_1 - \theta = \theta_2 - \theta$, откуда получаем $\theta_1 = \theta_2$, что противоречит предположению о том, что оценки различны. \square

Замечание 1.2.2

Для дальнейших действий будем считать, что функция распределения $F_{\theta}(x)$ имеет плотность $p(x, \theta)$, которая дважды дифференцируема по θ и её можно дифференцировать под знаком интеграла.*

$$\frac{\partial}{\partial \theta} \int_{\Delta} p(x, \theta) dx = \int_{\Delta} \frac{\partial}{\partial \theta} p(x, \theta) dx$$

Отметим, что выборка (x_1, \dots, x_n) имеет плотность распределения, так как является случайным вектором в \mathbb{R}^n , все координаты которого — независимые одинаково распределённые случайные величины, имеющие плотность.

Определение 1.2.3: Функция правдоподобия

Плотность распределения выборки называется функцией правдоподобия

$$L(\vec{x}, \theta) = \prod_{k=1}^n p(x_k, \theta)$$

Прологарифмировав функцию правдоподобия, получим сумму

$$\ln L(\vec{x}, \theta) = \sum_{k=1}^n \ln p(x_k, \theta)$$

Это сумма независимых одинаково распределённых случайных величин. Воспользовавшись законом больших чисел, можем сказать, что

$$\ln L(\vec{x}, \theta) = n \cdot \frac{\ln p(x_1, \theta) + \dots + \ln p(x_n, \theta)}{n} \approx n \cdot M_\theta \ln p(x_1, \theta)$$

при больших n .

Определение 1.2.4: Вклад выборки

Вклад выборки — частная производная по параметру θ от логарифма функции правдоподобия

$$U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta)$$

Замечание 1.2.5

Рекомендуется запомнить ещё две записи вклада выборки, так как они нам дальше пригодятся.

Первая:

$$U(\vec{x}, \theta) = \sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(x_k, \theta)$$

Вторая:

$$U(\vec{x}, \theta) = \frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)}$$

*Подробнее с дифференцированием под знаком интеграла Римана можно почитать во втором томе курса дифференциального и интегрального исчисления Фихтенгольца [14, с. 712]. Для тех, кто интересуется интегралом Лебега, прямой путь в книгу Дороговцева по общей теории меры и интеграла [11, с. 102]

Первая формула следует непосредственно из определения функции правдоподобия, вторая — из правила дифференцирования сложной функции.

Лемма 1.2.6

Математическое ожидание вклада выборки равно нулю

$$M_{\theta} U(\vec{x}, \theta) = 0$$

Доказательство. Посчитаем математическое ожидание вклада выборки

$$\begin{aligned} M_{\theta} U(\vec{x}, \theta) &= \int_{\mathbb{R}^n} U(\vec{u}, \theta) \cdot L(\vec{u}, \theta) d\vec{u} = \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\vec{x}, \theta) \cdot L(\vec{u}, \theta) d\vec{u} = \\ &= \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\vec{u}, \theta) d\vec{u} \end{aligned}$$

Воспользовавшись предположением о том, что функция распределения дважды дифференцируема, вынесем производную за знак интеграла

$$M_{\theta} U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\vec{u}, \theta) d\vec{u}$$

Так как L — плотность распределения, то

$$\int_{\mathbb{R}^n} L(\vec{u}, \theta) d\vec{u} = 1$$

и соответственно

$$M_{\theta} U(\vec{x}, \theta) = 0$$

□

Определение 1.2.7: Количество информации Фишера

Математическое ожидание квадрата вклада выборки называется количеством информации Фишера

$$I_n(\theta) = M_{\theta} U(\vec{x}, \theta)^2$$

Замечание 1.2.8

Между математическим ожиданием квадрата вклада выборки и второй

производной функции правдоподобия существует такое соотношение

$$M_{\theta} U(\vec{x}, \theta)^2 = -M_{\theta} \frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta)$$

Доказательство. Будем доказывать справа налево

$$\begin{aligned} -M_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta) \right] &= -M_{\theta} \left[\frac{\partial}{\partial \theta} \frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right] = \\ &= -M_{\theta} \left[\frac{\frac{\partial^2}{\partial \theta^2} L(\vec{x}, \theta) \cdot L(\vec{x}, \theta) - \left[\frac{\partial}{\partial \theta} L(\vec{x}, \theta) \right]^2}{L(\vec{x}, \theta)^2} \right] = \\ &= -M_{\theta} \left[\frac{\frac{\partial^2}{\partial \theta^2} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right] + M_{\theta} \left(\frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right)^2 \end{aligned}$$

Математическое ожидание вклада выборки равно нулю, а это значит, что

$$-M_{\theta} \frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta) = M_{\theta} \left[\frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right]^2$$

Воспользовавшись замечанием 1.2.5 и переписав равенство в другом виде, получаем

$$M_{\theta} U(\vec{x}, \theta)^2 = -M_{\theta} \frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta)$$

□

Пример 1.2.9

Пусть x_1, \dots, x_n — выборка из нормального распределения с параметрами θ и σ^2 . Неизвестный параметр — математическое ожидание. Тогда количество информации

$$I_n(\theta) = M_{\theta} U(\vec{x}, \theta)^2 = M_{\theta} \left(\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(x_k, \theta) \right)^2 = D_{\theta} \left[\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(x_k, \theta) \right]$$

Случайные величины x_k независимы и распределены одинаково, а это значит, что

$$D_{\theta} \left[\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(x_k, \theta) \right] = n \cdot I_1(\theta) = n \cdot M_{\theta} \left(\frac{\partial}{\partial \theta} \ln p(x_1, \theta) \right)^2$$

Производная логарифма плотности нормального распределения в данном случае считается просто

$$\frac{\partial}{\partial \theta} \ln p(x, \theta) = \frac{\partial}{\partial \theta} \left[\ln \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} - \frac{(x - \theta)^2}{2 \cdot \sigma^2} \right] = \frac{x - \theta}{\sigma^2}$$

Получаем результат

$$I_n(\theta) = n \cdot I_1(\theta) = n \cdot M_\theta \left[\frac{(x_1 - \theta)^2}{\sigma^4} \right] = n \cdot \frac{D_\theta x_1}{\sigma^4} = \frac{n}{\sigma^2}$$

То есть объём получаемой информации растёт линейно с увеличением длины выборки а в данном примере — ещё и обратно пропорционально дисперсии.

Количество информации позволяет оценить точность, с которой можем получить параметр θ .

Теорема 1.2.10: Неравенство Рао-Крамера

Пусть $\hat{\theta}$ — несмещённая оценка параметра θ . Тогда имеет место неравенство

$$\forall \theta \in \Theta : \quad D_\theta \hat{\theta} \geq \frac{1}{I_n(\theta)}$$

Доказательство. Выпишем, чему равно математическое ожидание оценки $\hat{\theta}$

$$\begin{cases} M_\theta \hat{\theta} = \theta \\ M_\theta \hat{\theta} = \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot L(\vec{u}, \theta) d\vec{u} \Rightarrow \theta = \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot L(\vec{u}, \theta) d\vec{u} \end{cases}$$

Продифференцируем с двух сторон полученное для θ равенство по параметру θ

$$\frac{\partial}{\partial \theta} \theta = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot L(\vec{u}, \theta) d\vec{u}$$

Левая часть равенства превращается в единицу, а справа происходит дифференцирование под знаком интеграла. Также помним, что оценка $\hat{\theta}(\vec{u})$ не зависит от параметра θ . Это значит, что производную нужно брать только от функции правдоподобия

$$1 = \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot \frac{\partial}{\partial \theta} L(\vec{u}, \theta) d\vec{u}$$

Далее нам нужно получить вклад выборки. Для этого умножим и поделим подынтегральное выражение на функцию правдоподобия

$$\int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot \frac{\partial}{\partial \theta} L(\vec{u}, \theta) d\vec{u} = \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot \frac{\frac{\partial}{\partial \theta} L(\vec{u}, \theta)}{L(\vec{u}, \theta)} \cdot L(\vec{u}, \theta) d\vec{u}$$

Видим, что дробь под интегралом — производная логарифма функции правдоподобия, которая является вкладом выборки

$$\int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot \frac{\frac{\partial}{\partial \theta} L(\vec{u}, \theta)}{L(\vec{u}, \theta)} \cdot L(\vec{u}, \theta) d\vec{u} = \int_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot U(\vec{x}, \theta) \cdot L(\vec{u}, \theta) d\vec{u}$$

У нас есть математическое ожидание произведения оценки и вклада выборки, которое равно единице

$$1 = M_{\theta} [\hat{\theta} \cdot U(\vec{x}, \theta)] \quad (1.10)$$

Помним, что математическое ожидание вклада выборки равно нулю. Значит, умножение его на константу ничего не меняет

$$M_{\theta} U(\vec{x}, \theta) = 0 \Rightarrow \theta \cdot M_{\theta} U(\vec{x}, \theta) = M_{\theta} [\theta \cdot U(\vec{x}, \theta)] = 0$$

Воспользовавшись полученным результатом, вернёмся к равенству (1.10). Отнимем от обеих частей ноль (то есть полученное только что выражение)

$$1 = M_{\theta} [\hat{\theta} \cdot U(\vec{x}, \theta)] - M_{\theta} [\theta \cdot U(\vec{x}, \theta)]$$

Получаем равенство

$$1 = M_{\theta} [(\hat{\theta} - \theta) \cdot U(\vec{x}, \theta)]$$

Воспользуемся неравенством Коши

$$\begin{aligned} 1 &= M_{\theta} [(\hat{\theta} - \theta) \cdot U(\vec{x}, \theta)] \leq \\ &\leq \sqrt{M_{\theta} (\hat{\theta} - \theta)^2} \cdot \sqrt{M_{\theta} U(\vec{x}, \theta)^2} = \sqrt{D_{\theta} \hat{\theta}} \cdot \sqrt{I_n(\theta)} \end{aligned} \quad (1.11)$$

Возводим обе части неравенства в квадрат и делим на количество информации

$$D_{\theta} \hat{\theta} \geq \frac{1}{I_n(\theta)}$$

Неравенство доказано. \square

Замечание 1.2.11

Иногда нужно оценивать не сам параметр, а функцию параметра. Тогда, если α — несмещённая оценка для $f(\theta)$, справедливо следующее неравенство

$$\forall \theta \in \Theta : \quad D_{\theta} \alpha \geq \frac{|f'(\theta)|}{I_n(\theta)}$$

1.2.2 Метод максимального правдоподобия

У нас есть нижняя оценка точности, с которой можно отыскать желаемую оценку, а это значит, что точнее определить просто не получится и нужно стремиться к равенству в неравенстве Рао-Крамера.

Определение 1.2.12: Эффективная оценка

Оценка $\hat{\theta}$, для которой в неравенстве Рао-Крамера стоит равенство,

называется эффективной

$$\forall \theta \in \Theta : \quad D_{\theta} \hat{\theta} = \frac{1}{I_n(\theta)}$$

Пример 1.2.13

Пусть x_1, \dots, x_n — выборка из распределения $N(0, 1)$. Докажем, что эффективная оценка параметра θ имеет вид

$$\hat{\theta} = \bar{x} = \frac{1}{n} \cdot \sum_{k=1}^n x_k$$

Для начала найдём её дисперсию

$$D_{\theta} \hat{\theta} = D_{\theta} \bar{x} = \frac{1}{n^2} \cdot \sum_{k=1}^n D_{\theta} x_k = \frac{1}{n} \cdot D_{\theta} x_1 = \frac{1}{n}$$

Теперь посмотрим на логарифм функции правдоподобия

$$\begin{aligned} \ln L(\vec{x}, \theta) &= \ln \left\{ \left(\frac{1}{\sqrt{2 \cdot \pi}} \right)^n \cdot \exp \left(-\frac{1}{2} \cdot \sum_{k=1}^n (x_k - \theta)^2 \right) \right\} \\ &= -n \cdot \ln \sqrt{2 \cdot \pi} - \frac{1}{2} \cdot \sum_{k=1}^n (x_k - \theta)^2 \end{aligned}$$

Не составит труда найти вклад выборки

$$U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) = \sum_{k=1}^n (x_k - \theta)$$

Посчитаем количество информации Фишера

$$I_n(\theta) = M_{\theta} \left[\sum_{k=1}^n (x_k - \theta) \right]^2 = D_{\theta} \left[\sum_{k=1}^n (x_k - \theta) \right] = n \cdot D_{\theta} x_1 = n$$

Видим, что условие $D_{\theta} \hat{\theta} = \frac{1}{I_n(\theta)}$ выполняется, а это значит, что оценка $\hat{\theta}$ является эффективной

Выясним, какими свойствами должна обладать плотность, чтобы можно было получить эффективную оценку. Для такой оценки

$$D_{\theta} \hat{\theta} = \frac{1}{I_n(\theta)}$$

Рассмотрим неравенство (1.11) и выясним, в каком случае в нём будет стоять знак равенства

$$1 = M_\theta \left[\left(\hat{\theta} - \theta \right) \cdot U(\vec{x}, \theta) \right] = \sqrt{M_\theta \left(\hat{\theta} - \theta \right)^2} \cdot \sqrt{M_\theta U(\vec{x}, \theta)^2}$$

В силу (1.9) последнее равенство влечёт за собой

$$\frac{U(\vec{x}, \theta)}{\sqrt{M_\theta U(\vec{x}, \theta)^2}} = \frac{\hat{\theta} - \theta}{M_\theta \left[\hat{\theta} - \theta \right]^2}$$

Распишем как

$$U(\vec{x}, \theta) = \left(\hat{\theta} - \theta \right) \cdot k(\theta)$$

Имеем

$$\ln L(\vec{x}, \theta) = \hat{\theta}(\vec{x}) \cdot \int k(\theta) \partial\theta - \int \theta \cdot k(\theta) \partial\theta + c(\vec{x})$$

Введём обозначения

$$a(\theta) = \int k(\theta) \partial\theta, \quad b(\theta) = - \int \theta \cdot k(\theta) \partial\theta$$

Получаем

$$\ln L(\vec{x}, \theta) = \hat{\theta}(\vec{x}) \cdot a(\theta) + b(\theta) + c(\vec{x})$$

Или же

$$L(\vec{x}, \theta) = \exp \left\{ \hat{\theta}(\vec{x}) \cdot a(\theta) + b(\theta) + c(\vec{x}) \right\}$$

Предположим, что существует эффективная оценка θ для $n = 1$. Тогда в силу доказанного

$$p(x, \theta) = \exp \left\{ \hat{\theta}(x) \cdot a(\theta) + b(\theta) + c(x) \right\}$$

При этом для $n > 1$ получим

$$L(\vec{x}, \theta) = \prod_{k=1}^n p(x_k, \theta) = \exp \left\{ \sum_{k=1}^n \hat{\theta}(x_k) \cdot a(\theta) + n \cdot b(\theta) + \sum_{k=1}^n c(x_k) \right\}$$

Тогда для несмещённой оценки $\hat{\theta}_n(\vec{x}) = \frac{1}{n} \cdot \sum_{k=1}^n \hat{\theta}(x_k)$ с помощью дифференцирования получим равенство

$$U(\vec{x}, \theta) = \left(\hat{\theta}_n - \theta \right) \cdot n \cdot k(\theta)$$

Значит, оценка $\hat{\theta}_n$ действительно эффективна.

Определение 1.2.14: Экспоненциальное распределение

Распределения следующего вида называются экспоненциальными

$$p(x, \theta) = \exp \left\{ \hat{\theta}(x) \cdot a(\theta) + b(\theta) + c(x) \right\}$$

Мы доказали следующий факт:

Утверждение 1.2.15

Если существует эффективная оценка $\hat{\theta}$ параметра θ при $n = 1$, то распределение экспоненциально, и существует эффективная оценка θ для любого n

$$\hat{\theta}_n(\vec{x}) = \frac{1}{n} \cdot \sum_{k=1}^n \hat{\theta}(x_k)$$

Попробуем предложить рецепт нахождения эффективной оценки. Начнём с примера

Пример 1.2.16

Есть выборка x_1, \dots, x_n из нормального распределения с неизвестным математическим ожиданием $N(\theta, 1)$. Тогда плотность распределения k -ой случайной величины будет следующей

$$p(x_k) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp \left\{ -\frac{(x_k - \theta)^2}{2} \right\}$$

Её логарифм, очевидно, имеет такой вид

$$\ln p(x_k) = \ln \frac{1}{\sqrt{2 \cdot \pi}} - \frac{(x_k - \theta)^2}{2}$$

Теперь выпишем логарифм функции правдоподобия

$$\begin{aligned} \ln L(\vec{x}, \theta) &= \sum_{k=1}^n \ln p(x_k) = \sum_{k=1}^n \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{(x_k - \theta)^2}{2} = \\ &= n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{(x_k - \theta)^2}{2} \end{aligned}$$

Преобразуем

$$\begin{aligned}\ln L(\vec{x}, \theta) &= n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{x_k^2}{2} + \sum_{k=1}^n x_k \cdot \theta - \frac{n \cdot \theta^2}{2}, \\ \sum_{k=1}^n x_k \cdot \theta &= \left(\frac{1}{n} \cdot \sum_{k=1}^n x_k \right) \cdot \theta \cdot n = \bar{x} \cdot \theta \cdot n, \\ \ln L(\vec{x}, \theta) &= n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{x_k^2}{2} + \bar{x} \cdot \theta \cdot n - \frac{n \cdot \theta^2}{2}\end{aligned}$$

Сгруппировав множители при n , получаем

$$\ln L(\vec{x}, \theta) = n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{x_k^2}{2} - n \cdot \frac{\theta^2 - 2 \cdot \bar{x} \cdot \theta}{2}$$

Добавим и вычтем в числителе дроби выборочное среднее

$$\ln L(\vec{x}, \theta) = n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{x_k^2}{2} - n \cdot \frac{\theta^2 - 2 \cdot \bar{x} \cdot \theta + (\bar{x}^2 - \bar{x}^2)}{2}$$

В результате получим

$$\ln L(\vec{x}, \theta) = n \cdot \ln \frac{1}{\sqrt{2 \cdot \pi}} - \sum_{k=1}^n \frac{x_k^2 - \bar{x}^2}{2} - n \cdot \frac{(\theta - \bar{x})^2}{2}$$

Видим, что последнее слагаемое не может быть положительным, так как это квадрат со знаком “минус”. Когда значение параметра θ равно эффективной оценке \bar{x} , функция правдоподобия принимает максимальное значение. Делаем предположение о том, как находить наилучшую оценку

$$\theta_* = \arg \max_{\theta} \ln L(\vec{x}, \theta)$$

Соответствующий рецепт поиска оценок во многих случаях оказывается эффективным.

Определение 1.2.17: Оценка максимального правдоподобия

Оценка максимального правдоподобия θ_* — такое значение параметра θ , при котором функция правдоподобия достигает своего максимального значения

$$\theta_* = \arg \max_{\theta} \ln L(\vec{x}, \theta)$$

Замечание 1.2.18

Оценок максимального правдоподобия может быть несколько, а может

не существовать ни одной.

Определение 1.2.19: Уравнение правдоподобия

Уравнением правдоподобия называется равенство вида

$$U(\vec{x}, \theta) = 0$$

Или же

$$\frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) = 0$$

Замечание 1.2.20

В гладком случае оценку θ_* можно искать с помощью уравнения правдоподобия. Тем не менее, нужно помнить, что равенство первой производной нулю является лишь необходимым условием максимума. Полученные критические точки нужно проверять на характер и наличие экстремума.

Пример 1.2.21

Пусть x_1, \dots, x_n — выборка из показательного распределения с параметром $\lambda > 0$. Тогда

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) &= \frac{\partial}{\partial \lambda} \ln \left[\prod_{k=1}^n \lambda \cdot e^{-\lambda \cdot x_k} \cdot \mathbb{1}\{x_k \in [0; +\infty)\} \right] = \\ &= \frac{\partial}{\partial \lambda} \sum_{k=1}^n (\ln \lambda - \lambda \cdot x_k) \cdot \mathbb{1}\{x_k \in [0; +\infty)\} = \\ &= \sum_{k=1}^n \left(\frac{1}{\lambda} - x_k \right) \cdot \mathbb{1}\{x_k \in [0; +\infty)\} \end{aligned}$$

Уравнение правдоподобия примет вид

$$\sum_{k=1}^n \left(\frac{1}{\lambda} - x_k \right) \cdot \mathbb{1}\{x_k \in [0; +\infty)\} = 0$$

Его решение будет следующим

$$\lambda_* = \frac{n}{\sum_{k=1}^n x_k}$$

Проверим полученный результат с помощью второй производной следующим образом

$$\frac{\partial^2}{\partial \lambda^2} \ln L(\vec{x}, \lambda_*) < 0$$

Действительно, получаем

$$-\frac{1}{\lambda_*^2} < 0$$

То есть обратное к среднему и есть оценка максимального правдоподобия для параметра λ

$$\lambda_* = \frac{1}{\bar{x}}$$

Замечание 1.2.22

Оценка максимального правдоподобия не обязана быть даже несмещённой — не только эффективной. В рассмотренном примере для $n = 1$ вообще не существует $M \lambda_* = M \frac{1}{x_1}$

Определение 1.2.23: Вариационный ряд

Вариационный ряд выборки x_1, \dots, x_n — значения выборки, упорядоченные в порядке неубывания

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}, \quad x_{(1)} = \min_k x_k, \quad x_{(n)} = \max_k x_k$$

Не всегда удобно и возможно использовать производную для решения уравнения правдоподобия. Следующий пример продемонстрирует это.

Пример 1.2.24

Пусть x_1, \dots, x_n — выборка из равномерного распределения на $[0; \theta]$, $\theta > 0$. Плотность принимает вид

$$p(x, \theta) = \frac{1}{\theta} \cdot \mathbb{1}\{x \in [0; \theta]\}$$

Нам нужно, чтобы все x_k оставались в пределах отрезка $[0; \theta]$, а это значит, что минимальный элемент не меньше, чем 0, а максимальный не больше θ . Сейчас

$$L(\vec{x}, \theta) = \frac{1}{\theta^n} \cdot \mathbb{1}\{0 \leq x_{(1)} \leq x_{(n)} \leq \theta\}$$

График функции правдоподобия (без индикаторов) изображён на рисунке 1.2. Чем меньше θ , тем больше значение функции правдоподобия, но есть одно важное ограничение — индикатор, который не позволяет быть параметру θ меньше максимального элемента выборки $x_{(n)}$. Значит, у нас есть оценка максимального правдоподобия

$$\theta_* = x_{(n)}$$

Возникает вопрос состоятельности этой оценки. При увеличении n мы располагаем всё большим количеством значений выборки, а θ_* при этом не убывает. Значит,

$$\exists \lim_{n \rightarrow \infty} \theta_* = \theta'$$

Также знаем, что $\theta' \leq \theta$, так как всегда $\theta_* \leq \theta$. Предположим, что

$$\mathbb{P}\{\theta - \theta' > 0\} > 0 \Rightarrow \exists \delta > 0 : \mathbb{P}\{\theta' < \theta - \delta\} > 0$$

Обозначим

$$\alpha = \mathbb{P}(\theta' < \theta - \delta)$$

Поскольку $\theta_* \leq \theta'$, имеем

$$\mathbb{P}(\theta_* < \theta - \delta) \geq \mathbb{P}(\theta' < \theta - \delta) = \alpha$$

С другой стороны

$$\mathbb{P}(\theta_* < \theta - \delta) = \mathbb{P}(x_1 < \theta - \delta, \dots, x_n < \theta - \delta) = \frac{(\theta - \delta)^n}{\theta^n} \xrightarrow{n \rightarrow \infty} 0$$

Противоречие. Значит, оценка $\theta_* = x_{(n)}$ сильно состоятельная

$$\theta_* \xrightarrow[n \rightarrow \infty]{a.s.} \theta$$

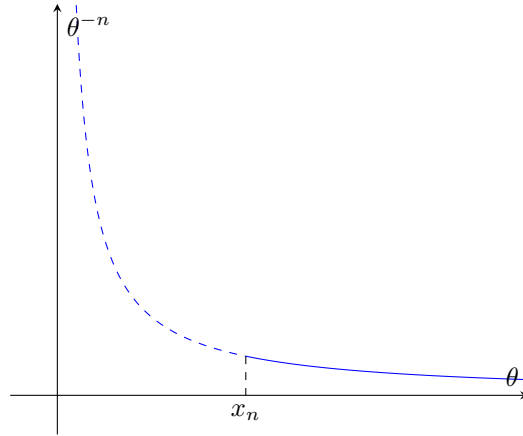


Рис. 1.2: Параметр θ не может быть меньше максимального значения выборки

Следующий результат мы приводим без доказательства.

Теорема 1.2.25

Если плотность $p(x, \theta)$ непрерывна и дифференцируема по параметру

θ , а производная не равна нулю $\frac{\partial}{\partial \theta} p(x, \theta) \neq 0$, то оценка максимального правдоподобия состоятельна

Глава 2

Достаточные статистики

2.1 Оптимальная оценка

Пример 2.1.1

Пусть $x_1, \dots, x_{2 \cdot n}$ — выборка из распределения Бернулли

$$\theta \in [0; 1] : \quad x_i = \begin{cases} 1, & \theta \\ 0, & 1 - \theta \end{cases}$$

Введём оценку

$$\hat{\theta} = \frac{2 \cdot x_1 + x_2 + 2 \cdot x_3 + x_4 + \dots + 2 \cdot x_{2 \cdot n-1} + x_{2 \cdot n}}{3 \cdot n}$$

Оценка $\hat{\theta}$ состоятельная

$$\hat{\theta} = \frac{2 \cdot \bar{x}}{3} + \frac{x_1 + x_3 + \dots + x_{2 \cdot n-1}}{3 \cdot n}$$

По закону больших чисел видим, что

$$\begin{cases} \frac{2 \cdot \bar{x}}{3} \xrightarrow{n \rightarrow \infty} \frac{2 \cdot \theta}{3} \\ \frac{x_1 + x_3 + \dots + x_{2 \cdot n-1}}{3 \cdot n} \xrightarrow{n \rightarrow \infty} \frac{\theta}{3} \end{cases}$$

Значит,

$$\hat{\theta} \xrightarrow{n \rightarrow \infty} \frac{2}{3} \cdot \theta + \frac{1}{3} \cdot \theta = \theta$$

Также оценка $\hat{\theta}$ является несмещённой

$$M \hat{\theta} = \frac{2 \cdot n \cdot \theta + n \cdot \theta}{3 \cdot n} = \theta$$

Хоть оценка $\hat{\theta}$ и обладает такими хорошими свойствами как состоятельность и несмещённость, она не оптимальная — не лучшая.

Упражнение 2.1.2

Проверить, что дисперсия $\hat{\theta}$ больше, чем дисперсия $\theta_* = \bar{x}$

Определение 2.1.3: Симметризация

Симметризация Λ оценки $\hat{\theta}$ — среднее оценок $\hat{\theta}$ для всевозможных перестановок $\sigma \in S_n$ элементов выборки x_1, \dots, x_n

$$\Lambda \hat{\theta} = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \hat{\theta}(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$$

Лемма 2.1.4

Для произвольной несмещённой оценки $\hat{\theta}$ её симметризация $\Lambda \hat{\theta}$ не хуже её самой в среднем квадратическом

$$M_{\theta} \hat{\theta} = \theta \Rightarrow \begin{cases} M_{\theta} \Lambda \hat{\theta} = M_{\theta} \hat{\theta} = \theta \\ D_{\theta} \Lambda \hat{\theta} \leq D_{\theta} \hat{\theta} \end{cases}$$

Доказательство. Как и раньше, обозначим $\vec{x} = [x_1, x_2, \dots, x_n]$. Для перестановки $\sigma \in S_n$ используем $\vec{x}_{\sigma} = [x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}]$ Тогда

$$\Lambda \hat{\theta} = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \hat{\theta}(\vec{x}_{\sigma})$$

1. Сначала докажем несмещённость $\Lambda \hat{\theta}$.

Нетрудно показать, что вектора \vec{x} и \vec{x}_{σ} имеют одинаковое распределение для любой перестановки σ , а это значит, что и оценки $\hat{\theta}(\vec{x})$ и $\hat{\theta}(\vec{x}_{\sigma})$ распределены одинаково как результаты применения одной и той же функции к одинаково распределённым случайным векторам. Следовательно, их математические ожидания равны между собой при любой перестановке σ

$$M_{\theta} \hat{\theta}(\vec{x}) = M_{\theta} \hat{\theta}(\vec{x}_{\sigma}) = \theta$$

Посчитаем математическое ожидание симметризации оценки $\hat{\theta}$

$$M_{\theta} \Lambda \hat{\theta} = M_{\theta} \left\{ \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \hat{\theta}(\vec{x}_{\sigma}) \right\} = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} M_{\theta} \hat{\theta}(\vec{x}_{\sigma}) = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \theta = \theta$$

2. Теперь посмотрим, чему равна дисперсия симметризации оценки $\hat{\theta}$. Воспользуемся определением

$$D_{\theta} \Lambda \hat{\theta} = M_{\theta} \left(\Lambda \hat{\theta} - \theta \right)^2 = M_{\theta} \left\{ \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \hat{\theta}(\vec{x}_{\sigma}) - \theta \right\}^2$$

Внесём параметр θ в сумму

$$\begin{aligned} M_{\theta} \left\{ \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \hat{\theta}(\vec{x}_{\sigma}) - \theta \right\}^2 &= M_{\theta} \left\{ \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right) \right\}^2 = \\ &= M_{\theta} \left\{ \sum_{\sigma \in S_n} \frac{1}{n!} \cdot \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right) \right\}^2 \end{aligned}$$

Воспользуемся неравенством Коши-Буняковского. Помним, что в S_n находится $n!$ перестановок

$$\begin{aligned} \left\{ \sum_{\sigma \in S_n} \frac{1}{n!} \cdot \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right) \right\}^2 &\leq \sum_{\sigma \in S_n} \left(\frac{1}{n!} \right)^2 \cdot \sum_{\sigma \in S_n} \left\{ \hat{\theta}(\vec{x}_{\sigma}) - \theta \right\}^2 \\ &= \frac{1}{n!} \cdot \sum_{\sigma \in S_n} \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right)^2 \end{aligned}$$

Тогда

$$M_{\theta} \left\{ \sum_{\sigma \in S_n} \frac{1}{n!} \cdot \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right) \right\}^2 \leq \frac{1}{n!} \cdot \sum_{\sigma \in S_n} M_{\theta} \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right)^2 \quad (2.1)$$

Видим сумму дисперсий. Дисперсии одинаковы, так как оценки имеют одинаковые распределения

$$\begin{aligned} \frac{1}{n!} \cdot \sum_{\sigma \in S_n} M_{\theta} \left(\hat{\theta}(\vec{x}_{\sigma}) - \theta \right)^2 &= \frac{1}{n!} \cdot \sum_{\sigma \in S_n} D_{\theta} \hat{\theta}(\vec{x}_{\sigma}) = \\ &= \frac{1}{n!} \cdot \sum_{\sigma \in S_n} D_{\theta} \hat{\theta}(\vec{x}) = \frac{1}{n!} \cdot n! \cdot D_{\theta} \hat{\theta}(\vec{x}) = D_{\theta} \hat{\theta}(\vec{x}) \end{aligned}$$

Из неравенства (2.1) видим, что дисперсия симметризации не хуже дисперсии самой оценки

$$D_{\theta} \Lambda \hat{\theta} \leq D_{\theta} \hat{\theta}(\vec{x})$$

То есть симметризация не ухудшает оценку, а в общем случае (когда неравенство строгое) даже делает её лучше. \square

Замечание 2.1.5

В предыдущем примере 2.1.1 выполняется равенство

$$\Lambda \hat{\theta} = \bar{x}$$

Замечание 2.1.6

Любая оптимальная оценка является функцией вариационного ряда

2.2 σ -алгебра, порождённая случайной величиной

Есть вероятностное пространство $(\Omega, \mathfrak{F}, \mathbb{P})$ и случайная величина ξ .

Определение 2.2.1: Сигма-алгебра, порождённая случайной величиной

$\mathfrak{F}_\xi = \sigma(\xi)$ — σ -алгебра, порождённая случайной величиной ξ

$$\mathfrak{F}_\xi = \{\xi^{-1}(\Delta) \mid \Delta \in \mathfrak{B}\}$$

\mathfrak{B} — борелевская σ -алгебра в \mathbb{R} .

Из курса теории вероятностей помним лемму, которая утверждает, что ξ — случайная величина тогда и только тогда, когда

$$\forall \Delta \in \mathfrak{B} : \quad \{\omega \mid \xi(\omega) \in \Delta\} = \{\xi \in \Delta\} = \xi^{-1}(\Delta) \in \mathfrak{F}$$

А это значит, что все элементы \mathfrak{F}_ξ входят в σ -алгебру \mathfrak{F} , а сама \mathfrak{F}_ξ является подмножеством \mathfrak{F}

$$\begin{cases} \mathfrak{F}_\xi = \{\xi^{-1}(\Delta) \mid \Delta \in \mathfrak{B}\} \\ \forall \Delta \in \mathfrak{B} : \xi^{-1}(\Delta) \in \mathfrak{F} \end{cases} \Rightarrow \mathfrak{F}_\xi \subset \mathfrak{F}$$

Проверим, что \mathfrak{F}_ξ действительно является σ -алгеброй.

1. Множество элементарных исходов Ω входит в \mathfrak{F}_ξ . Поскольку случайная величина ξ принимает действительные значения, то прообраз множества действительных чисел \mathbb{R} и будет множеством элементарных исходов Ω .

$$\begin{cases} \xi^{-1}(\Delta \in \mathfrak{B}) \in \mathfrak{F} \\ \mathbb{R} \in \mathfrak{B} \\ \xi^{-1}(\mathbb{R}) = \Omega \end{cases} \Rightarrow \Omega \in \mathfrak{F}_\xi$$

2. Если событие A принадлежит \mathfrak{F}_ξ , то его дополнение \bar{A} тоже принадлежит \mathfrak{F}_ξ

$$A = \xi^{-1}(\Delta) = \{\omega \mid \xi(\omega) \in \Delta\}$$

Значит,

$$\overline{A} = \{\omega \mid \xi(\omega) \notin \Delta\} = \{\omega \mid \xi(\omega) \in \overline{\Delta}\} = \xi^{-1}(\overline{\Delta})$$

Поскольку \mathfrak{B} является σ -алгеброй, а Δ — её элемент, то дополнение $\overline{\Delta}$ тоже принадлежит σ -алгебре \mathfrak{B} . Из этого следует, что свойство выполняется

$$\begin{cases} \xi^{-1}(\Delta) \in \mathfrak{F} \\ \Delta \in \mathfrak{B} \Rightarrow \overline{\Delta} \in \mathfrak{B} \end{cases} \Rightarrow \overline{\xi^{-1}(\Delta)} = \xi^{-1}(\overline{\Delta}) \in \mathfrak{F}$$

3. Замкнутость относительно счётных пересечений. Пусть $A_n = \xi^{-1}(\Delta_n)$, $n \geq 1$, $\Delta_n \in \mathfrak{B}$. Тогда

$$\bigcap_{n=1}^{\infty} A_n = \xi^{-1}\left(\bigcap_{n=1}^{\infty} \Delta_n\right).$$

Так как \mathfrak{B} — σ -алгебра, то

$$\bigcap_{n=1}^{\infty} \Delta_n \in \mathfrak{B}.$$

Следовательно,

$$\xi^{-1}\left(\bigcap_{n=1}^{\infty} \Delta_n\right) \in \mathfrak{F}_{\xi}$$

Как устроена эта σ -алгебра? Каждому элементарному исходу отвечает одно и только одно значение случайной величины, а каждому значению случайной величины отвечает один и больше элементарных исходов. Допустим, есть некое $a \in \mathbb{R}$, которое является образом по крайней мере двух элементарных исходов ω_1 и ω_2 (рисунок 2.1)

$$\xi(\omega_1) = \xi(\omega_2) = a$$

Теперь рассмотрим элемент Δ борелевской σ -алгебры \mathfrak{B} . Из вышесказанного следует, что, если число a принадлежит множеству Δ , то прообраз этого множества содержит элементы ω_1 и ω_2 , в противном случае оба элементарных исхода не входят в прообраз

$$\begin{aligned} a \in \Delta &\Rightarrow \xi^{-1}(\Delta) \ni \omega_1, \omega_2 \\ a \notin \Delta &\Rightarrow \xi^{-1}(\Delta) \not\ni \omega_1, \omega_2 \end{aligned}$$

То есть множество \mathfrak{F}_{ξ} не будет различать элементы ω_1 и ω_2 . Это в свою очередь означает, что можно разбить Ω на уровни — непересекающиеся подмножества, не различимые с помощью \mathfrak{F}_{ξ} .

Определение 2.2.2: Множество уровня

Множество уровня H_t — полный прообраз значения $t \in \mathbb{R}$ случайной величины ξ

$$H_t = \{\omega \mid \xi(\omega) = t\} = \xi^{-1}(t)$$

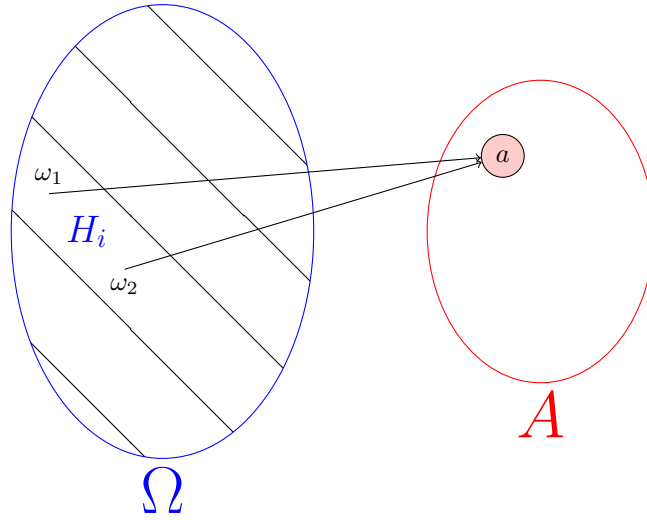


Рис. 2.1: Множества уровней

Замечание 2.2.3

Уровни H_i составляют разбиение множества элементарных исходов Ω .

1. Множества H_i не пересекаются

$$H_{t_1} \cap H_{t_2} = \emptyset \Leftrightarrow t_1 \neq t_2$$

2. Объединение всех H_i даёт множество элементарных исходов

$$\bigcup_{t \in \mathbb{R}} H_t = \bigcup_{t \in \mathbb{R}} \xi^{-1}(t) = \xi^{-1}(\mathbb{R}) = \Omega$$

Очень похоже на гипотезы из курса теории вероятностей с той лишь разницей, что уровней может быть бесконечное и даже континуальное количество, из чего также следует, что вероятность некоторых из них может быть нулевой.

2.3 Случайная величина, измеримая относительно σ -алгебры

В общем случае вероятностное пространство может быть разбито на континуальное количество множеств уровней (для σ -алгебры, порождённой непрерывной случайной величиной). Начнём же с рассмотрения того случая, когда случайная величина ξ принимает n значений a_1, a_2, \dots, a_n

$$\xi : \Omega \rightarrow \{a_1, a_2, \dots, a_n\}$$

Это означает, что у нас есть n уровней

$$H_k = \{\omega \mid \xi(\omega) = a_k\}, k = \overline{1, n}$$

Нетрудно понять, что $\sigma(\xi)$ содержит 2^n элементов

$$\sigma(\xi) = \left\{ \bigcup_{k=1}^n H_k^{\eta_k} \mid \eta_k = \overline{0, 1}, H_k^0 = \emptyset, H_k^1 = H_k \right\}$$

Определение 2.3.1: Сигма-алгебра, порождённая полным набором гипотез

Возьмём набор множеств H_1, \dots, H_n который является полным набором гипотез для пространства элементарных исходов Ω

$$\bigcap_{k=1}^n H_k = \emptyset, \bigcup_{k=1}^n H_k = \Omega, \mathbb{P}\{H_k\} \neq 0$$

В таком случае σ -алгебра, содержащая всевозможные объединения этих множеств, будет называться σ -алгеброй, порождённой полным набором гипотез и будет выглядеть следующим образом

$$\mathfrak{F}_1 = \left\{ \bigcup_{k=1}^n H_k^{\eta_k} \mid \eta_k = \overline{0, 1}, H_k^0 = \emptyset, H_k^1 = H_k \right\}$$

Определение 2.3.2: Случайная величина, измеримая относительно сигма-алгебры

Случайная величина \varkappa называется измеримой относительно σ -алгебры \mathfrak{F}_1 , если

$$\forall \Delta \in \mathfrak{B} : \quad \{\omega \mid \varkappa(\omega) \in \Delta\} \in \mathfrak{F}_1$$

Нас интересует, как устроены случайные величины, которые измеримы относительно σ -алгебры $\sigma(\xi)$.

Утверждение 2.3.3

Пусть случайная величина принимает n значений a_1, \dots, a_n

$$\xi : \Omega \rightarrow \{a_1, \dots, a_n\}$$

Случайная величина \varkappa измерима относительно $\sigma(\xi)$ тогда и только тогда, когда $\varkappa = f(\xi)$ для некоторой функции

$$f : \{a_1, \dots, a_n\} \rightarrow \mathbb{R}$$

Доказательство. Докажем сначала необходимость. Из вида σ -алгебры $\sigma(\xi)$ следует, что множества из $\sigma(\xi)$ либо целиком содержат определённые множества уровней H_k , либо не пересекаются с ними

$$\forall A \in \sigma(\xi), \forall k \in \{1, \dots, n\} : \quad A \cap H_k \neq \emptyset \Rightarrow H_k \subset A$$

Для произвольной измеримой относительно $\sigma(\xi)$ случайной величины \varkappa положим

$$\forall k \in \{1, \dots, n\}, \forall \omega_k \in H_k : \quad c_k = \varkappa(\omega_k)$$

Введём обозначение

$$\forall k \in \{1, \dots, n\} : \quad A_k = \{\omega \mid \varkappa(\omega) = c_k\}$$

Получаем

$$\forall k \in \{1, \dots, n\} : \quad \omega_k \in A_k \cap H_k \Rightarrow A_k \cap H_k \neq \emptyset \Rightarrow H_k \subset A_k$$

Значит,

$$\forall \omega \in H_k : \varkappa(\omega) = c_k$$

Введём функцию

$$\forall k \in \{1, \dots, n\} : \forall f(a_k) = c_k$$

Имеем

$$\varkappa(\omega) = \sum_{k=1}^n c_k \cdot \mathbb{1}\{\xi(\omega) = a_k\} = f(\xi(\omega))$$

Докажем достаточность. Пусть $\varkappa = f(\xi(\omega))$. Обозначим

$$\forall k \in \{1, \dots, n\} : \quad c_i = f(a_k)$$

Тогда для произвольного $\Delta \in \mathfrak{B}(\mathbb{R})$

$$\begin{aligned} \{\omega \mid \varkappa(\omega) \in \Delta\} &= \bigcup_{k=1}^n H_k \cap \{\omega \mid \varkappa(\omega) \in \Delta\} = \\ &= \bigcup_{k=1}^n \{\omega \mid \xi(\omega) = a_k, f(\xi(\omega)) \in \Delta\} = \\ &= \bigcup_{k=1}^n \{\omega \mid \xi(\omega) = a_k\}, f(a_k) \in \Delta = \\ &= \bigcup_{k=1}^n \{\omega \mid \xi(\omega) = a_k, c_i \in \Delta\} = \bigcup_{k=1}^n H_i^{\eta_i}, \end{aligned}$$

где

$$\eta_k = \begin{cases} 0, & c_k \notin \Delta \\ 1, & c_k \in \Delta \end{cases}$$

Видим, что

$$\{\omega \mid \varkappa(\omega) \in \Delta\} \in \sigma(\xi)$$

Теорема доказана. □

Легко увидеть, что все наши рассуждения остаются верными, если рассматривать не конечные наборы гипотез, а счётные (соответственно, случайные величины, принимающие счётное количество значений). В таком случае σ -алгебра \mathfrak{F}_1 содержит континуальное число элементов, а утверждение 2.3.3 остаётся в силе

$$\mathfrak{F}_1 = \left\{ \bigcup_{k=1}^{\infty} H_k^{\eta_k} \mid \eta_k = \overline{0, 1}, H_k^0 = \emptyset, H_k^1 = H_k \right\}$$

Справедливо гораздо более общее утверждение, касающееся случайных величин с произвольным распределением.

Теорема 2.3.4

Случайная величина η измерима относительно σ -алгебры, порождённой случайной величиной ξ , тогда и только тогда, когда существует борелевская функция

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

такая, что

$$f(\xi) = \eta$$

[15, с. 219].

2.4 Условное математическое ожидание

Пусть η — произвольная случайная величина, интегрируемая с квадратом. Нужно найти случайную величину $\tilde{\eta}$ которая измерима в $\sigma(\xi)$ и является наиболее близкой в среднем квадратическом к η .

2.4.1 Проекция вектора

Сейчас для нас снова будет полезна аналогия с векторами в пространстве. будет полезна аналогия с векторами в пространстве. Имеется точка x в пространстве L' . Мы ищем такую точку y в подпространстве $L \subset L'$, что расстояние между x и y минимальное. Значит, надо опустить перпендикуляр от y на L (рисунок 2.2). Пусть e_1, e_2, \dots, e_n — ортонормированный базис в L , тогда y можно найти по формуле

$$y = \sum_{k=1}^n (x, e_k) \cdot e_k \quad (2.2)$$

2.4.2 Проекция случайной величины

Воспользуемся приведённым геометрическим фактом применительно к случайным величинам. Возьмём в качестве L множество всех случайных величин, которые измеримы относительно $\sigma(\xi)$.

$$L \ni \sum_{k=1}^n c_k \cdot \mathbb{1}_{H_k}, c_k \in \mathbb{R}$$

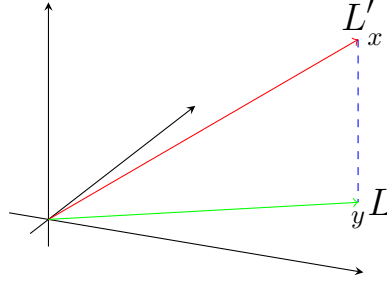


Рис. 2.2: Проекция

В качестве скалярного произведения случайных величин используем математическое ожидание произведения.

У нас есть ортонормированный базис

$$e_k = \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \quad (2.3)$$

Теперь проекция случайной величины η на L имеет вид

$$\begin{aligned} \tilde{\eta} &= \sum_{k=1}^n (\eta, e_k) \cdot e_k = \sum_{k=1}^n \left(\eta, \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \right) \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} = \\ &= \sum_{k=1}^n \mathbb{M} \left(\eta \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \right) \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} = \\ &= \sum_{k=1}^n \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}]}{\sqrt{\mathbb{P}(H_k)}} \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} = \\ &= \sum_{k=1}^n \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k} \end{aligned} \quad (2.4)$$

На что стоит обратить внимание в этой формуле:

1. $\tilde{\eta}$ — **случайная величина**, так как индикатор вне математического ожидания — случайная величина, и результат суммы будет зависеть от произошедшего ω , а точнее от множества уровня H_k , которому оно принадлежит

2. Когда ω принадлежит H_k , то результатом суммы будет среднее значение случайной величины η на событии H_k

Определение 2.4.1: Условное математическое ожидание случайной величины относительно случайного события

Условное математическое ожидание случайной величины ξ относительно **события** A [5, стр. 68] обозначается $M[\xi | A]$ и считается по формуле

$$M[\xi | A] = \frac{M[\xi \cdot \mathbb{1}_A]}{\mathbb{P}(A)} = \int_A \xi(\omega) \mathbb{P}(d\omega | A),$$

где

$$d\omega \in A, \quad \mathbb{P}(d\omega | A) = \frac{\mathbb{P}(d\omega)}{\mathbb{P}(A)}$$

Определение 2.4.2: Условное математическое ожидание случайной величины относительно конечной сигма-алгебры

Есть σ -алгебра \mathfrak{F}_1 , порождённая множествами H_1, H_2, \dots, H_n . Тогда условное математическое ожидание случайной величины η относительно этой σ -алгебры — **случайная величина**, которая обозначается $M[\eta | \mathfrak{F}_1]$ и вычисляется по формуле

$$M[\eta | \mathfrak{F}_1] = \sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k}$$

Замечание 2.4.3

У нас есть определения условного математического ожидания относительно σ -**алгебры** \mathfrak{F} (определение 2.4.2) и относительно **случайного события** A (определение 2.4.1). Из контекста будет ясно, какое именно определение используется, поэтому путаницы возникнуть не должно.

Перепишем последнее определение в терминах математического ожидания относительно случайных событий

$$M[\eta | \mathfrak{F}_1] = \sum_{k=1}^n M[\eta | H_k] \cdot \mathbb{1}_{H_k}$$

При детальном рассмотрении из записи очевиден её смысл: условное математическое ожидание относительно σ -алгебры — вектор в пространстве с базисом $\{H_1, \dots, H_n\}$, элементами которого являются проекции вектора η на соответствующие оси.

Ведь $M[\eta | H_k]$ — ни что иное, как проекция вектора (случайной величины) η на ось (уровень) H_k . Эта величина является скаляром, как и проекция вектора на ось.

Лемма 2.4.4: Равенство скалярных произведений для конечной сигма-алгебры

Для случайной величины η и её проекции $\tilde{\eta}$ на σ -алгебру \mathfrak{F}_ξ , порожд-

дённую случайной величиной ξ , принимающей конечное количество значений, выполняется равенство скалярных произведений

$$\forall A \in \mathfrak{F}_\xi : M[\tilde{\eta} \cdot \mathbb{1}_A] = M[\eta \cdot \mathbb{1}_A] \quad (2.5)$$

Доказательство.

$$M[\tilde{\eta} \cdot \mathbb{1}_A] = M\left(\sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k} \cdot \mathbb{1}_A\right) = \sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot M[\mathbb{1}_{H_k \cap A}]$$

Помним, что математическое ожидание индикатора — вероятность

$$\sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot M[\mathbb{1}_{H_k \cap A}] = \sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{P}(H_k \cap A)$$

Замечаем условную вероятность

$$\begin{aligned} \sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{P}(H_k \cap A) &= \sum_{k=1}^n M[\eta \cdot \mathbb{1}_{H_k}] \cdot \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(H_k)} = \\ &= \sum_{k=1}^n M[\eta \cdot \mathbb{1}_{H_k}] \cdot \mathbb{P}(A | H_k) \end{aligned}$$

Событие A принадлежит σ -алгебре случайных событий \mathfrak{F}_ξ . Значит, условная вероятность $\mathbb{P}(A | H_k)$ равна либо нулю, либо единице, поскольку A либо включает в себя уровень H_k , либо не пересекается с ним. То есть получился индикатор $\mathbb{1}(H_k \subseteq A)$. А этот индикатор говорит о том, что теперь надо суммировать лишь по тем H_k , которые являются частью события A

$$\begin{aligned} \sum_{k=1}^n M[\eta \cdot \mathbb{1}_{H_k}] \cdot \mathbb{P}(A | H_k) &= \sum_{k=1}^n M[\eta \cdot \mathbb{1}_{H_k}] \cdot \mathbb{1}(H_k \subseteq A) = \\ &= \sum_{H_k \subseteq A} M[\eta \cdot \mathbb{1}_{H_k}] = M\left[\sum_{H_k \subseteq A} \eta \cdot \mathbb{1}_{H_k}\right] \end{aligned}$$

Значит, равенство (2.5) выполняется. \square

Замечание 2.4.5

В связи с выполнением равенства скалярных произведений можем сделать вывод, что математическое ожидание случайной величины и её проекции тоже равны. Это нетрудно показать, установив A равным всему множеству элементарных исходов (индикатор в таком случае станет просто тождественной единицей)

$$M\eta = M[\eta \cdot \mathbb{1}_\Omega] = M[\tilde{\eta} \cdot \mathbb{1}_\Omega] = M\tilde{\eta}$$

2.4.3 Условное математическое ожидание

Введём общее определение для условного математического ожидания случайной величины относительно σ -алгебры

Определение 2.4.6: Условное математическое ожидание случайной величины относительно сигма-алгебры

Условным математическим ожиданием случайной величины η относительно σ -алгебры $\mathfrak{F}_1 \subset \mathfrak{F}$ называется такая случайная величина $\tilde{\eta}$, что

1. Случайная величина $\tilde{\eta}$ измерима относительно σ -алгебры \mathfrak{F}_1
2. Выполняется равенство

$$\forall A \in \mathfrak{F}_1 : M[\tilde{\eta} \cdot \mathbb{1}_A] = M[\eta \cdot \mathbb{1}_A]$$

Обозначение $\tilde{\eta} = M[\eta \mid \mathfrak{F}_1]$

Замечание 2.4.7

В силу леммы 2.4.4 определения 2.4.2 и 2.4.6 для σ -алгебры, порождённой дискретной случайной величиной, совпадают.

Замечание 2.4.8

Условное математическое ожидание случайной величины η относительно σ -алгебры, порождённой случайной величиной ξ , будем обозначать $M[\eta \mid \sigma(\xi)]$, а более кратко $M[\eta \mid \xi]$.

То есть имеем три эквивалентных записи

$$M[\eta \mid \mathfrak{F}_\xi] = M[\eta \mid \sigma(\xi)] = M[\eta \mid \xi]$$

Немного остановимся на примере, чтобы понять, что у нас есть на данный момент

Пример 2.4.9

У нас есть две дискретные случайные величины ξ и η с совместным дискретным распределением

$$\mathbb{P}\{\eta = a_i, \xi = b_j\} = p_{ij}$$

Числа p_{ij} обладают свойствами

$$p_{ij} \geq 0, \quad \sum_{i,j=1}^{\infty} p_{ij} = 1, \quad \mathbb{P}\{\xi = b_j\} = \sum_{i=1}^{\infty} p_{ij}$$

Посчитаем условное математическое ожидание согласно формуле

из определения (2.4.2)

$$M[\eta \mid \sigma(\xi)] = \sum_{k=1}^n \frac{M[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k}$$

Для этого выясним, чему равно математическое ожидание ξ при определённом значении η по формуле из определения (2.4.1)

$$M[\eta \mid \xi = b_j] = \frac{M[\eta \cdot \mathbb{1}_{\eta=b_j}]}{\mathbb{P}\{\xi = b_j\}} = \frac{\sum_{i=1}^{\infty} a_i \cdot p_{ij}}{\sum_{i=1}^{\infty} p_{ij}}$$

Попробуем обобщить определение условного математического ожидания, чтобы обладать универсальной формулой, из которой можно делать какие-то выводы. Начнём с формулы для случая конечного числа значений

$$M[\eta \mid \mathfrak{F}_{\xi}] = \sum_{k=1}^n M[\eta \mid H_k] \cdot \mathbb{1}_{H_k}$$

Вспомним, что

$$H_k = \{\omega \mid \xi(\omega) = a_k\}$$

Для краткости будем использовать запись $H_k = \{\xi = a_k\}$. Тогда

$$M[\eta \mid \sigma(\xi)] = \sum_{k=1}^n M[\eta \mid \xi = a_k] \cdot \mathbb{1}\{\xi = a_k\}$$

Введём функцию

$$\forall x \in \{a_1, \dots, a_n\} : \quad \varphi^{\eta}(x) = M[\eta \mid \xi = x]$$

Получаем

$$M[\eta \mid \sigma(\xi)] = \varphi^{\eta}(\xi)$$

Пусть случайные величины ξ и η имеют совместную плотность распределения

$$\mathbb{P}\{(\xi, \eta) \in \Delta\} = \iint_{\Delta} p(x, y) \, dx \, dy$$

В таком случае компонента ξ имеет плотность r

$$r(x) = \int_{\mathbb{R}} p(x, y) \, dy$$

Компонента η имеет плотность q

$$q(y) = \int_{\mathbb{R}} p(x, y) \, dx$$

Для $f \in C(\mathbb{R}^2)$ найдём $M[f(\xi, \eta) \mid \sigma(\xi)]$. Помним, что для некоторой борелевской функции φ справедливо равенство

$$M[f(\xi, \eta) \mid \sigma(\xi)] = \varphi(\xi)$$

В определении условного математического ожидания возьмём $A = \{\xi \in (-\infty; x] \in \sigma(\xi)\}$

$$\mathbb{M}[\varphi(\xi) \cdot \mathbb{1}\{\xi \leq x\}] = \mathbb{M}[f(\xi, \eta) \cdot \mathbb{1}\{\xi \leq x\}]$$

Перепишем равенство через интегралы

$$\int_{-\infty}^x \varphi(t) \cdot r(t) dt = \int_{-\infty}^x \int_{\mathbb{R}} f(t, y) \cdot p(t, y) dy dt$$

Продифференцируем его по dt с обеих сторон

$$\varphi(x) \cdot r(x) = \int_{\mathbb{R}} f(x, y) \cdot p(x, y) dy$$

Получаем

$$\varphi(x) = \frac{\int_{\mathbb{R}} f(x, y) \cdot p(x, y) dy}{r(x)} = \frac{\int_{\mathbb{R}} f(x, y) \cdot p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy}$$

Следовательно,

$$\mathbb{M}[f(\xi, \eta) \mid \sigma(\xi)] = \frac{\int_{\mathbb{R}} f(\xi, y) \cdot p(\xi, y) dy}{\int_{\mathbb{R}} p(\xi, y) dy} \quad (2.6)$$

Для случая $f(\xi, \eta) = \eta$ имеем

$$\mathbb{M}[\eta \mid \sigma(\xi)] = \frac{\int_{\mathbb{R}} y \cdot p(\xi, y) dy}{\int_{\mathbb{R}} p(\xi, y) dy} \quad (2.7)$$

Замечание 2.4.10

Предыдущее доказательство предполагает существование условного математического ожидания, что гарантирует следующая теорема.

Теорема 2.4.11: Существование условного математического ожидания

Для случайной величины ξ такой, что $\mathbb{M}|\xi| < +\infty$ и произвольной σ -алгебры $\mathfrak{F}' \subset \mathfrak{F}$ условное математическое ожидание $\mathbb{M}[\xi \mid \mathfrak{F}']$ существует и единственно с точностью до равенства почти наверное. [6, стр. 142]

2.4.4 Свойства условного математического ожидания

*

I Формула полной вероятности [6, стр. 144]

$$M M[\eta | \mathfrak{F}_1] = M \eta$$

II Условное математическое ожидание неотрицательной случайной величины неотрицательно почти наверное

$$\eta \geq 0 \Rightarrow M[\eta | \mathfrak{F}_1] \geq 0$$

III Неравенство Йенсена. Если функция φ выпуклая вниз, то

$$\varphi(M[\eta | \mathfrak{F}_1]) \leq M[\varphi(\eta) | \mathfrak{F}_1]$$

IV Телескопическое свойство

$$\mathfrak{F}_2 \subset \mathfrak{F}_1 \Rightarrow M[M[\eta | \mathfrak{F}_1] | \mathfrak{F}_2] = M[\eta | \mathfrak{F}_2]$$

V Если случайная величина η измерима относительно σ -алгебры \mathfrak{F}_1 , то её условное математическое ожидание равно ей самой

$$M[\eta | \mathfrak{F}_1] = \eta$$

VI Если случайная величина η измерима относительно \mathfrak{F}_1 , то для любой случайной величины ξ

$$M[\eta \cdot \xi | \mathfrak{F}_1] = \eta \cdot M[\xi | \mathfrak{F}_1]$$

VII Если η не зависит от \mathfrak{F}_1 , то её условное математическое ожидание равно простому математическому ожиданию

$$M[\eta | \mathfrak{F}_1] = M \eta$$

VIII Условное математическое ожидание линейно

$$\forall a, b \in \mathbb{R} : M[a \cdot \xi + b \cdot \eta | \mathfrak{F}_1] = M[a \cdot \xi | \mathfrak{F}_1] + M[b \cdot \eta | \mathfrak{F}_1]$$

IX Сохраняется теорема Лебега о возможности предельного перехода под знаком условного математического ожидания [3, стр. 302] [15, стр. 272]

$$|\xi_n| \leq \eta, M \eta < \infty, \xi_n \xrightarrow[n \rightarrow \infty]{a.s.} \xi \Rightarrow M[\xi_n | \mathfrak{F}_1] \xrightarrow[n \rightarrow \infty]{a.s.} M[\xi | \mathfrak{F}_1]$$

Пара полезных частных случая неравенства Йенсена (III свойство)

$$\begin{aligned} \varphi(x) = |x| : \quad & M[\eta | \mathfrak{F}_1] \leq M[|\eta| | \mathfrak{F}_1] \\ \varphi(x) = x^2 : \quad & (M[\eta | \mathfrak{F}_1])^2 \leq M[\eta^2 | \mathfrak{F}_1] \end{aligned}$$

*Также со свойствами и их доказательствами можно ознакомиться в книгах Ширяева [15, стр. 270] и Боровкова [6, стр. 143]

2.4.5 Условное математическое ожидание по функции от случайного вектора

Чем вызвала интерес эта тема? Допустим, у нас есть x_1, \dots, x_n — выборка с функцией правдоподобия L

$$L(\vec{x}, \theta) = \prod_{k=1}^n p(x_k, \theta)$$

Также есть $\hat{\theta}$ — несмещённая оценка параметра θ . Как улучшить $\hat{\theta}$? Возьмём статистику $T = T(\vec{x})$, обладающую определёнными свойствами. Тогда улучшенной оценкой θ будет условное математическое ожидание $M[\hat{\theta} | T]$.

О свойствах, которыми должна обладать статистика T , поговорим позже. Одно ясно уже сейчас: T является функцией от выборки \vec{x} , как и оценка $\hat{\theta}$. Это значит, что нам не нужно погружаться в слишком абстрактные размышления, а достаточно выяснить, как считать математическое ожидание одной функции выборки (случайного вектора) $f(\vec{x})$ при условии другой функции $g(\vec{x})$ **той же** выборки \vec{x} .

$$f, g : \mathbb{R}^n \rightarrow \mathbb{R}$$

Вспомним, что для поиска условного математического ожидания мы находили функцию $\varphi^\eta(x) = M[\eta | \xi = x]$. Сейчас изменилось совсем немного — лишь обозначения: вместо η у нас $f(\vec{x})$, а вместо ξ тут $g(\vec{x})$. Значит, нужно найти вид такого условного математического ожидания

$$M[f(\vec{x}) | g(\vec{x}) = t] = ?$$

Предположим, что случайная величина $g(\vec{x})$ имеет плотность $q(t)$. Тогда из равенства

$$M[f(\vec{x}) | g(\vec{x}) = t] = \varphi(g(\vec{x}))$$

аналогично доказательству формулы (2.6) получаем

$$M[f(\vec{x}) | g(\vec{x}) = t] = \frac{1}{q(t)} \cdot \frac{d}{dt} \int_{\{g(\vec{y}) \leq t\}} f(\vec{y}) \cdot p(\vec{y}) d\vec{y} \quad (2.8)$$

Пример 2.4.12

Найдём $M[x_1 | x_{(n)}]$ для равномерного на $[0; 1]$ распределения.

Для начала найдём плотность $r(t)$ величины $x_{(n)}$

$$\mathbb{P}\{x_{(n)} \leq t\} = \mathbb{P}\{x_1 \leq t, x_2 \leq t, \dots, x_n \leq t\} = t^n$$

откуда

$$q(t) = \frac{d}{dt}(t^n) = n \cdot t^{n-1}$$

Интеграл в (2.8)

$$\int_{\{\max(y_1, y_2, \dots, y_n) \leq t\}} y_1 d\vec{y} = \int_{y_1=0 \dots y_n=0}^{y_1=t \dots y_n=t} y_1 d\vec{y} = \frac{t^2}{2} \cdot t^{n-1} = \frac{t^{n+1}}{2}$$

Окончательно

$$M[x_1 | x_{(n)}] = \frac{\frac{d}{dt} \left(\frac{t^{n+1}}{2} \right)}{n \cdot t^{n-1}} = \frac{(n+1) \cdot t^n}{2 \cdot n \cdot t^{n-1}} = \frac{(n+1) \cdot t}{2 \cdot n}$$

Разберёмся теперь, что из себя представляет множество точек $S_t = \{\vec{u} \mid g(\vec{u}) = t\}$. Функция g описывает скалярное поле в n -мерном пространстве. Для скалярного поля множество S_t имеет своё название — поверхность уровня (изоповерхность) — то есть поверхность, на которой функция принимает одно и то же значение.

Для понимания ситуации рассмотрим несколько примеров.

Пример 2.4.13

Пусть

$$n = 2, \quad g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

Функция g имеет вид

$$g(x, y) = x$$

Очевидно, что поверхности уровней — вертикальные линии (параллельные оси ординат) вида

$$S_t = \{(x, y) \mid x = t\}$$

Пример 2.4.14

$$n = 2, g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

В этот раз функция $g(x, y)$ будет квадратом расстояния от начала координат $(0, 0)$ до точки (x, y)

$$g(x, y) = x^2 + y^2$$

Поверхностями уровня будут окружности радиуса \sqrt{t}

$$S_t = \{(x, y) \mid x^2 + y^2 = t\} = \{(x, y) \mid x^2 + y^2 = t\}$$

Пример 2.4.15

Рассмотрим n -мерное пространство ($n \geq 1$). Возьмём единичный вектор $\vec{e} : \|\vec{e}\| = 1$. Функция g будет скалярным произведением аргумента

\vec{u} с вектором \vec{e}

$$g(\vec{u}) = (\vec{u}, \vec{e})$$

В таком случае поверхностью уровня S_t будет гиперплоскость, проходящая через точку $t \cdot \vec{e}$, с нормалью \vec{e} , которую описывает следующее уравнение

$$S_t = \{\vec{u} \mid (\vec{u}, \vec{e}) = t\}$$

Попробуем получить отличное от (2.8) выражение для $M[x_1 \mid x_{(n)}]$, воспользовавшись введённым понятием поверхностей уровня S_t . Кажется логичным, чтобы в формуле для условного математического ожидания при $g(\vec{x}) = t$ учитывались только точки из множества S_t . Интегралы, в которых учитываются только точки некоторой поверхности, называются поверхностными интегралами. Мы получим формулу для условного математического ожидания в виде частного двух поверхностных интегралов по S_t .

Попробуем найти плотность распределения $g(\vec{x})$.

Вероятность того, что значение $g(\vec{x})$ отдалено от t не больше, чем на $\tilde{\varepsilon}$, будет приблизительно равна плотности распределения $g(\vec{x})$ в этой точке (если таковая имеется), умноженной на это расстояние. Обозначим плотность случайной величины $g(\vec{x})$ в точке t через $q(t)$. Тогда

$$\mathbb{P}\{g(\vec{x}) \in [t - \tilde{\varepsilon}, t + \tilde{\varepsilon}]\} \approx q(t) \cdot 2 \cdot \tilde{\varepsilon}$$

Множество таких точек $\vec{u} \in \mathbb{R}^n$, что $|g(\vec{u}) - t| \leq \tilde{\varepsilon}$ можно считать приблизительно равным множеству точек, удовлетворяющих соотношению

$$\rho(\vec{u}, S'_t) \leq \varepsilon = \frac{\tilde{\varepsilon}}{\|\vec{\nabla} \cdot g(\vec{u}_0)\|}, \quad (2.9)$$

где \vec{u}_0 — ближайшая к \vec{u} точка поверхности S_t . Обозначим это множество точек через $G_{\tilde{\varepsilon}}$. Окончательно

$$q(t) \cdot 2 \cdot \tilde{\varepsilon} \approx \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u}$$

Следовательно, у нас почти готова формула для плотности $q(t)$ случайной величины $g(\vec{x})$

$$q(t) \cdot 2 \cdot \tilde{\varepsilon} \approx \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u} \Rightarrow q(t) \approx \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u}$$

Чтобы убрать неточность и было обычное равенство, устремим ширину коридорчика к нулю

$$q(t) = \lim_{\tilde{\varepsilon} \rightarrow 0} \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u} \quad (2.10)$$

$$q(t) = \lim_{\tilde{\varepsilon} \rightarrow 0} \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u} = \lim_{\varepsilon \rightarrow 0} \frac{1}{2 \cdot \varepsilon \cdot \|\vec{\nabla} \cdot g(\vec{u})\|} \cdot \int_{G_{\varepsilon}} p(\vec{u}) d\vec{u}$$

Поскольку ε играет роль половины толщины (отсюда и возникает множитель двойка), а $d\vec{u}$ — маленький элемент объёма, то при делении объёма на толщину получим площадь. Поскольку толщина стремится к нулю, то она становится соразмерна с объёмом и мы получаем ненулевое значение площади, а коридорчик G_ε вырождается в поверхность уровня S_t . Обозначив меру площади на поверхности S_t как $\sigma_t(d\vec{u})$, получаем поверхностный интеграл первого рода

$$q(t) = \int_{S_t} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u}) \quad (2.11)$$

А теперь вспомним, как обстояло дело со случайными величинами, имеющими совместную плотность (2.6)

$$\varphi^n(x) = \frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{r(x)} = \frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy}$$

В знаменателе у нас стоял вес поверхности уровня, раскрыв который, мы получали интеграл от совместной плотности. Что же есть у нас? Вес поверхности уровня — функция одного аргумента $q(t)$, которая равна интегралу от плотности по всей той части пространства, где случайная величина $g(\vec{x})$ принимает одно и то же значение t — по поверхности уровня S_t .

То есть в нашем случае роль \mathbb{R} играет поверхность S_t , роль совместной плотности $p(x, y)$ играет плотность случайного вектора $p(\vec{u})$, а вместо дифференциала dy у нас мера площади, делённая на норму градиента $\frac{\sigma_t(d\vec{u})}{\|\vec{\nabla} \cdot g(\vec{u})\|}$. В числителе дроби в формуле (2.6) стоит y , в нашем же случае это функция $f(\vec{u})$, поскольку там случайная величина присутствовала в плотности сама по себе, тут же у нас есть плотность случайного вектора $p(\vec{x})$, а найти нужно среднее функции случайной величины $f(\vec{x})$. Итого, получается переход

$$\frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{r(x)} \rightarrow \frac{\int_{S_t} f(\vec{u}) \cdot p(\vec{u}) \frac{\sigma_t(d\vec{u})}{\|\vec{\nabla} \cdot g(\vec{u})\|}}{q(t)}$$

И конечная формула

$$M[f(\vec{x}) | g(\vec{x}) = t] = \frac{\int_{S_t} f(\vec{u}) \cdot p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}$$

Теорема 2.4.16: Условное математическое ожидание гладких функций

Если есть случайный вектор \vec{x} с известной плотностью распределения $p(\vec{x})$, а также гладкая функция $g(\vec{x})$ с невырожденным градиентом,

то математическое ожидание случайной величины $f(\vec{x})$ при условии $g(\vec{x}) = t$ считается по формуле

$$M[f(\vec{x}) | g(\vec{x}) = t] = \frac{\int_{S_t} f(\vec{u}) \cdot p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}$$

Замечание 2.4.17

Формула остаётся справедливой, если поверхности уровня функции g состоят из нескольких гладких кусков.

2.4.6 Пример

Поскольку этот пример оказался достаточно громоздким, было решено посвятить ему целый подраздел.

Есть выборка x_1, \dots, x_n из равномерного распределения с на отрезке $[0, \theta]$. Нужно посчитать условное математическое ожидание первого элемента выборки $f(\vec{x}) = x_1$ при условии максимального $g(\vec{x}) = x_{(n)}$

$$M[x_1 | x_{(n)}] = ?$$

Поверхности уровня

Для начала сообразим, что из себя представляет поверхность уровня S_t

$$S_t = \left\{ \vec{u} \mid u_i \geq 0, \max_{i=1, n} u_i = t \right\}$$

На двумерной плоскости это будет два отрезка, перпендикулярных друг другу и осям. Они будут выходить из точки (t, t) и заканчиваться на осях.

Рассмотрим, почему всё именно так

1. Нас не устраивают точки, которые находятся за пределами квадрата, ограниченного прямыми, проходящими через точки $(0, t)$ и $(t, 0)$, так как в таком случае максимум будет больше t , что по условию быть не может

2. Максимальное значение зафиксировано, а это значит, что хотя бы одна координата должна всегда равняться максимуму

Таким образом, закрасив квадрат, нижняя левая грань которого находится в начале координат, а верхняя правая в точке (t, t) , удаляем все те точки, где нет такой координаты, которая равна t . Все точки внутри контура пропадут, так как лишь на контуре квадрата могут находиться точки, имеющие хоть одну координату равной t . Теперь осталось отбросить те отрезки, что лежат на координатных осях, потому что на них у точек меняется лишь одна координата (например, на оси ординат меняется лишь x , а $y = 0$ на всей оси), а желаемое значение t принимается лишь на самом конце отрезка, но это лишь одна точка, и в пространстве размерностью $n \geq 1$ имеет меру Лебега 0, поэтому её тоже можно отбросить.

Так уж получилось, что только что был выведен практически универсальный способ построения поверхности уровня $g(\vec{u}) = \max_{i=1,n} u_i = t$. Осталось лишь ввести небольшие правки и распространить его на пространство размерности n .

Например, в трёхмерном пространстве поверхностью уровня будет совокупность граней куба, где исключены те грани, что соприкасаются с одной из осей — те грани, одна точка которых находится в начале координат $(0, 0)$.

Так же будет и в многомерном пространстве — чертим гиперкуб и отбрасываем те его грани, один из углов которых находится в начале координат.

Также отметим, что каждая грань перпендикулярна $n - 1$ осям, а пересекается с ними лишь в точках со значением t . Так как нельзя провести две разные гиперплоскости, имеющих одну нормаль и проходящих через одну точку (гиперплоскость по определению определяется этой парой), то делаем вывод: поверхность S_t у нас состоит из n граней.

Норма градиента

Найдём норму градиента функции $g(\vec{x}) = \max_{i=1,n} x_i$.

Возможны два случая:

1. Максимальный элемент в векторе один
2. Максимальных элементов в векторе несколько (от двух до n) и они равны друг другу (если $n \geq 2$)

Рассмотрим первый случай. Без потери общности предположим, что максимальный элемент — первый

$$\max_{i=1,n} x_i = x_1$$

Очевидно, что очень малое изменение x_1 приведёт к очень малому (причём такому же) изменению максимального элемента выборки. Грубо говоря, если у нас есть числа $x_1 = 3, x_2 = 2, x_3 = 1$, то максимальное из них — x_1 . Если оно изменится на 1 в какую-либо сторону, то максимальное значение выборки изменится так же на 1.

Мы это всё рассматриваем для того, чтобы показать, что производная по максимальной координате будет равна единице, так как производная — отношение очень малого прироста функции к очень малому изменению аргумента (который привёл к такому изменению функции). Даже если у нас были числа $x_1 = 1$ и $x_2 = 1.001$, мы всё равно сможем найти такой маленький прирост $\delta < 0.001$ (между двумя разными действительными числами найдётся ещё континуальное число действительных чисел, поэтому такое δ найдётся всегда), при котором x_2 останется максимальным элементом и прирост градиента по этой координате будет равен приросту самой координаты.

Не забываем, что мы сейчас рассматриваем тот случай, когда максимальное значение одно, а это значит, что остальные значения выборки строго меньше максимального.

Поскольку очень малые изменения других элементов выборки не приведут к тому, что они станут максимальными (а если приведут, то возьмём

ещё более маленький прирост), то их изменение не повлияет на значение функции g , а это значит, что частные производные по ним обращаются в нули.

Итого, к чему мы пришли. Когда x_1 является максимальным элементом выборки, то градиент функции g в точке \vec{x} равен

$$\vec{\nabla} \cdot g(\vec{x}) = \frac{\partial g}{\partial x_1} \cdot \vec{e}_1 + \frac{\partial g}{\partial x_2} \cdot \vec{e}_2 + \dots + \frac{\partial g}{\partial x_n} \cdot \vec{e}_n = \vec{e}_1 + 0 + \dots + 0 = \vec{e}_1$$

Когда у нас не x_1 является максимальным элементом, а x_k , где $1 \leq k \leq n$, то очевидно, что результат будет следующим:

$$\vec{\nabla} \cdot g(\vec{x}) = \vec{e}_k$$

Нас интересует лишь норма градиента, которая в данном случае равна единице

$$\left\| \vec{\nabla} \cdot g(\vec{x}) \right\| = \|\vec{e}_k\| = 1$$

Второй случай (когда несколько элементов максимальны и равны между собой) нас не интересует, так как такое возможно лишь на рёбрах ($n \geq 3$) и в точках пересечения рёбер ($n \geq 2$). Что те, что другие (прямые и точки в пространстве) имеют нулевой объём. В одномерном пространстве (с одной осью) у нас есть лишь одна случайная величина, и она же является максимальной.

Условное математическое ожидание относительно события

Посчитаем условное математическое ожидание по формуле

$$\mathbb{M}[x_1 \mid x_{(n)} = t] = \frac{\int_{S_t} u_1 \cdot \theta^{-n} \cdot \frac{1}{t} \sigma_t(d\vec{u})}{\int_{S_t} \theta^{-n} \cdot \frac{1}{t} \sigma_t(d\vec{u})} = \frac{\int_{S_t} u_1 \sigma_t(d\vec{u})}{\int_{S_t} \sigma_t(d\vec{u})}$$

Поверхность уровня S_t в верхнем интеграле можно (и нужно!) разбить на две части: ту, где u_1 меняется, и ту, где u_1 принимает постоянное значение.

Вспоминаем, что поверхность S_t — часть гиперкуба, содержащая лишь грани, на которых та случайная величина, что не меняется, имеет значение t (максимум). Частью поверхности, на которой u_1 принимает постоянное значение t , будет одна грань — та грань гиперкуба, что проходит через точку $u_1 = t$ и перпендикулярна оси (вектору \vec{e}_1).

Обозначим грань, где $u_1 = t$, как U_t , а оставшуюся часть поверхности как $Y_t = S_t \setminus U_t$.

Имеем более конкретную формулу для подсчёта условного математического ожидания

$$\mathbb{M}[x_1 \mid x_{(n)} = t] = \frac{\int_{S_t} u_1 \sigma_t(d\vec{u})}{\int_{S_t} \sigma_t(d\vec{u})} = \frac{\int_{U_t} u_1 \Big|_{u_1=t} \sigma_t(d\vec{u}) + \int_{Y_t} u_1 \sigma_t(d\vec{u})}{\int_{S_t} \sigma_t(d\vec{u})}$$

Вспоминаем, что имеется n граней, а это значит, что в знаменателе у нас n объёмов $(n-1)$ -мерных гиперкубов (n площадей n -мерных квадратов) со стороной t , которые в свою очередь равняются числу t^{n-1} .

В числителе у нас два интеграла. Первый интеграл — интеграл по одной грани, что опять же является $(n-1)$ -мерным объёмом (n -мерной площадью). Не забываем про константу t , что там находится: она умножается на результат интегрирования t^{n-1} и в результате получаем t^n .

Теперь подошли к самому сложному кусочку этой дроби — интеграл по оставшимся $n-1$ граням. Поскольку интеграл не зависит ни от чего кроме u_1 , то мы преспокойнейше можем вынести $(n-2)$ -мерный объём, а по оставшемуся измерению придётся интегрировать от 0 до t

$$\frac{\int_{U_t} u_1 \Big|_{u_1=t} \sigma_t(d\vec{u}) + \int_{Y_t} u_1 \sigma_t(d\vec{u})}{\int_{S_t} \sigma_t(d\vec{u})} = \frac{t^n + (n-1) \cdot t^{n-2} \cdot \int_0^t u_1 du_1}{n \cdot t^{n-1}}$$

Дальше идут нехитрые математические преобразования, которые называются интегрированием

$$\begin{aligned} M[x_1 \mid x_{(n)} = t] &= \frac{t^n + (n-1) \cdot t^{n-2} \cdot \int_0^t u_1 du_1}{n \cdot t^{n-1}} = \\ &= \frac{t^n + (n-1) \cdot t^{n-2} \cdot \frac{t^2}{2}}{n \cdot t^{n-1}} = \frac{t + (n-1) \cdot t \cdot \frac{1}{2}}{n} = t \cdot \frac{n+1}{2 \cdot n} \end{aligned}$$

Условное математическое ожидание

Мы получили условное математическое ожидание относительно события $\{g(x) = t\}$ — конкретное значение $t \cdot \frac{n+1}{2 \cdot n}$. Чтобы получить условное математическое ожидание одной случайной величины относительно другой, нужно лишь подставить вместо t наше значение $g(\vec{x}) = x_{(n)}$

$$M[x_1 \mid x_{(n)}] = M[x_1 \mid x_{(n)} = t] \Big|_{t=x_{(n)}} = t \cdot \frac{n+1}{2 \cdot n} \Big|_{t=x_{(n)}} = x_{(n)} \cdot \frac{n+1}{2 \cdot n}$$

Видим, что получили случайную величину, которая не зависит от параметра θ , а это значит, что мы на правильном пути

$$M[x_1 \mid x_{(n)}] = \frac{n+1}{2 \cdot n} \cdot x_{(n)}$$

Проверка

Проведём небольшую очевидную проверку: положим $n = 1$ (одна случайная величина, одномерное пространство). Тогда формула примет следующий вид

$$M[x_1 \mid x_{(1)}] = \frac{1+1}{2 \cdot 1} \cdot x_{(1)} = \frac{2}{2} \cdot x_{(1)} = x_{(1)} = x_1$$

Всё сходится с нашими интуитивными предположениями: у нас имеется лишь один элемент в выборке, а мы знаем значение максимального. Значит, мы знаем значение этого единственного элемента (иначе кому ещё быть максимальным в этой выборке?).

2.5 Условные распределения

Определение 2.5.1: Условное распределение

Условное распределение случайной величины ξ при известной σ -алгебре \mathfrak{F}_1 — это функция π

$$\pi : \Omega \times \mathfrak{B} \rightarrow [0, 1]$$

Функция π должна обладать следующими свойствами

1. На любом элементе Δ борелевского множества \mathfrak{B} функция $\pi(\cdot, \Delta)$ является измеримой относительно σ -алгебры \mathfrak{F}_1
2. На любом элементарном исходе из множества Ω функция $\pi(\omega, \cdot)$ является вероятностной мерой
3. Выполняется равенство

$$\forall \Delta \in \mathfrak{B} : \pi(\cdot, \Delta) = M[\mathbb{1}\{\xi \in \Delta\} \mid \mathfrak{F}_1]$$

Это равенство нам уже знакомо, поэтому ничего принципиально нового не добавилось

$$\mathbb{P}(\xi \in \Delta) = M[\mathbb{1}\{\xi \in \Delta\}]$$

Обозначение

$$\pi(\cdot, \Delta) = \mathbb{P}(\xi \in \Delta \mid \mathfrak{F}_1)$$

Если σ -алгебра \mathfrak{F}_1 порождена случайной величиной η ($\mathfrak{F}_1 = \sigma(\eta)$), будем использовать следующее обозначение

$$\mathbb{P}(\xi \in \Delta \mid \sigma(\eta)) = p(\eta, \Delta)$$

Когда нас интересует событие $\eta = t$, обозначаем это так

$$\mathbb{P}(\xi \in \Delta \mid \eta = t) = p(t, \Delta)$$

Связь с условным математическим ожиданием

$$M[f(\xi) \mid \eta = t] = \int_{\mathbb{R}} f(u) p(t, du)$$

Замечание 2.5.2

В обозначениях выше точка вместо аргумента означает, что на выходе мы получаем не определённое значение, а функцию от того аргумента, который заменён точкой.

Например, запись $\pi(\cdot, \Delta)$ означает некую функцию ρ

$$\rho : \Omega \rightarrow [0, 1]$$

Значение этой функции будет считаться по формуле

$$\rho(\omega) = \pi(\omega, \Delta)$$

Рассмотрим примеры вычисления условных распределений

Пример 2.5.3: См. пример 2.4.9

Случайные величины ξ и η имеют совместное дискретное распределение

$$\mathbb{P}\{\xi = a_i, \eta = b_j\} = p_{ij}$$

В таком случае условное распределение считается по формуле

$$\mathbb{P}\{\xi = a_i \mid \eta = b_j\} = \frac{p_{ij}}{\sum_i p_{ij}}$$

Пример 2.5.4: См. формулу 2.6

Случайные величины ξ и η имеют совместную плотность распределения $p(x, y)$

$$\mathbb{P}\{\xi \in \Delta \mid \eta = x\} = \frac{\int_{\Delta} p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy}$$

Пример 2.5.5: См. теорему 2.4.16

У случайного вектора \vec{x} есть плотность распределения $p(\vec{u})$. Тогда условное распределение $f(\vec{x})$ относительно гладкой функции $g(\vec{x})$ считается по формуле

$$\mathbb{P}(f(\vec{x}) \in \Delta \mid g(\vec{x}) = t) = \frac{\int_{S_t \cap \Delta} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u})}$$

2.6 Достаточные статистики

Как говорилось в подразделе 2.4.5, условное математическое ожидание нам понадобилось из-за наличия статистик T , обладающих особыми свойствами. Пришло время о них поговорить

Определение 2.6.1: Достаточная статистика

Статистика $T(x_1, x_2, \dots, x_n)$ — достаточная статистика для параметра θ , если условное распределение выборки при известном T не зависит

от θ

Осмыслим написанное

1. Речь идёт об условном распределении всей выборки

$$\pi(T, \Delta) = \mathbb{P}(\vec{x} \in \Delta \mid T)$$

2. Почему возникает определение достаточных статистик? Пусть T — достаточная статистика. Как с её помощью получить распределение всей выборки?

$$\mathbb{P}\{\vec{x} \in \Delta\} = \mathbb{M} \mathbb{1}\{\vec{x} \in \Delta\}$$

Далее воспользуемся I свойством условного математического ожидания (формула полной вероятности)

$$\mathbb{M} \mathbb{1}\{\vec{x} \in \Delta\} = \mathbb{M} \mathbb{M}[\mathbb{1}\{\vec{x} \in \Delta\} \mid T]$$

Из определения условного математического ожидания и утверждения 2.3.3 следует, что для некоторой измеримой функции f справедливо следующее равенство

$$\mathbb{M}[\mathbb{1}\{\vec{x} \in \Delta \mid T\}] = f(T)$$

Тогда получаем

$$\mathbb{P}\{\vec{x} \in \Delta\} = \mathbb{M} f(T)$$

Теорема 2.6.2: Об улучшении оценки с помощью достаточной статистики

Пусть x_1, \dots, x_n — выборка из распределения $F_\theta(x)$, $\theta \in \Theta$. Есть T — достаточная статистика для параметра θ , а также несмещённая оценка $\hat{\theta}$ параметра θ . Введём оценку θ_*

$$\theta_* = \mathbb{M}[\hat{\theta} \mid T]$$

Оценка θ_* не хуже, чем оценка $\hat{\theta}$, в среднем квадратическом

$$\begin{cases} \mathbb{M}_\theta \theta_* = \theta \\ D_\theta \theta_* \leq D_\theta \hat{\theta} \end{cases}$$

Замечание 2.6.3

Оценка $\theta_* = \mathbb{M}[\hat{\theta} \mid T]$ не зависит от θ , так как T — достаточная статистика

Замечание 2.6.4

“Как правило”, одномерная достаточная статистика для одномерного параметра даёт не улучшаемую статистику

Доказательство. 1. С первым пунктом всё просто: пользуемся формулой полной вероятности (I свойство)

$$M_{\theta} \theta_* = M_{\theta} M \left[\hat{\theta} \mid T \right] = M_{\theta} \hat{\theta} = \theta$$

2. Для доказательства второго утверждения распишем дисперсию и оценки θ_* по определению

$$D_{\theta} \theta_* = M_{\theta} (\theta_* - \theta)^2 = M_{\theta} \left(M \left[\hat{\theta} \mid T \right] - \theta \right)^2$$

Поскольку T — достаточная статистика и не зависит от θ , то θ измерима относительно $\sigma(T)$ и является константой. Значит, можно переписать статистику θ как условное математическое ожидание (V свойство), а затем воспользоваться линейностью условного математического ожидания (VIII свойство)

$$\begin{aligned} M_{\theta} \left(M \left[\hat{\theta} \mid T \right] - \theta \right)^2 &= M_{\theta} \left(M \left[\hat{\theta} \mid T \right] - M[\theta \mid T] \right)^2 = \\ &= M_{\theta} \left(M \left[(\hat{\theta} - \theta) \mid T \right] \right)^2 \end{aligned}$$

Далее воспользуемся неравенством Йенсена (III свойство) и формулой полной вероятности (I свойство)

$$\begin{aligned} M_{\theta} \left(M \left[(\hat{\theta} - \theta) \mid T \right] \right)^2 &\leq M_{\theta} M \left[(\hat{\theta} - \theta)^2 \mid T \right] = \\ &= M_{\theta} (\hat{\theta} - \theta)^2 = D_{\theta} \hat{\theta} \end{aligned}$$

□

Замечание 2.6.5

Равенство в неравенстве Йенсена выше возможно, когда условное распределение вырождается в одну точку. Когда условное распределение не вырождено, то неравенство оказывается строгим.

Замечание 2.6.6

Оценка $\theta_* = M \left[\hat{\theta} \mid T \right]$ измерима относительно T по определению условного математического ожидания, а это значит, что она является функцией от T .

Пусть $\tilde{\theta}$ — оптимальная несмещённая оценка. Тогда $\theta_* = M \left[\hat{\theta} \mid T \right]$ — оптимальная, а значит, эти оценки равны $\theta_* = \tilde{\theta}$ по теореме един-

ственности (теорема Колмогорова 1.2.1), поскольку оптимальная оценка либо одна, либо не существует вовсе. Значит, оптимальная оценка — функция достаточной статистики.

Теорема 2.6.7: Факторизационная теорема (о характеристизации достаточной статистики)

Пусть x_1, \dots, x_n — выборка из распределения с плотностью $p(x, \theta)$, $\theta \in \Theta$.

Статистика T является достаточной тогда и только тогда, когда функция правдоподобия $L(\vec{x}, \theta)$ допускает факторизацию, то есть может быть представлена произведением двух функций следующего вида

$$L(\vec{x}, \theta) = h(T, \theta) \cdot g(\vec{x})$$

Замечание 2.6.8

Также с теоремой и её доказательствами можно ознакомиться в источниках [9, стр. 78], [6, стр. 158].

Наброски доказательства для гладкой функции правдоподобия. Условное распределение выборки при известной статистике $T = f(\vec{x})$ определяется формулой

$$\mathbb{P}(\vec{x} \in \Delta \mid T = t) = \frac{\int_{S_t \cap \Delta} L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})} \quad (2.12)$$

Достаточность Пусть функция правдоподобия допускает факторизацию, то есть существуют такие функции $h(T, \theta)$ и $g(\vec{x})$, что

$$L(\vec{x}, \theta) = h(T, \theta) \cdot g(\vec{x})$$

Тогда интеграл (2.12) примет следующий вид

$$\mathbb{P}(\vec{x} \in \Delta \mid T = t) = \frac{\int_{S_t \cap \Delta} h(T, \theta) \cdot g(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} h(T, \theta) \cdot g(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}$$

Хоть T и является функцией выборки, мы зафиксировали его значение, а θ является константой. Это значит, что функция $h(T, \theta)$ тоже не зависит от \vec{u} , поэтому сверху и снизу её можно сократить, избавив зависимость распределения от параметра θ

$$\mathbb{P}(\vec{x} \in \Delta \mid T = t) = \frac{\int_{S_t \cap \Delta} g(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} g(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}$$

Условное распределение \vec{x} не зависит от θ при известном T , что и требовалось доказать для того, чтобы показать достаточность факторизации.

Необходимость Пусть T — достаточная статистика. Выпишем плотность распределения T в точке t согласно формуле (2.11), но с небольшими поправками: плотность распределения выборки \vec{x} заменяется функцией правдоподобия, а плотность T будет зависеть не только от t , но и от параметра θ .

$$q(t, \theta) = \int_{S_t} L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})$$

Поверхностная мера $\sigma_t(\vec{u})$ не зависит от параметра θ , а это значит, что можно поделить обе части равенства на $q(t, \theta)$ и внести плотность под знак интеграла.

$$\int_{S_t} \frac{1}{q(t, \theta)} \cdot L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u}) = 1$$

Область интегрирования и мера не зависят от θ , а это значит, что и подынтегральное выражение тоже не зависит от θ , а зависит лишь от вектора \vec{u}

$$\frac{1}{q(t, \theta)} \cdot L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} = c(\vec{u})$$

Значит, функцию правдоподобия можно представить в следующем виде

$$L(\vec{u}, \theta) = q(t, \theta) \cdot \|\vec{\nabla} \cdot f(\vec{u})\| \cdot c(\vec{u})$$

И тут мы видим, что это и есть факторизация!

Для начала вспомним, что у нас распределение при условии $T = t$, а это значит, что плотность $q(t, \theta)$ может быть расписана следующим образом

$$q(t, \theta) = q(T, \theta)$$

Теперь, выделив функции g и h , получаем необходимый результат

$$\begin{cases} L(\vec{u}, \theta) = q(t, \theta) \cdot \|\vec{\nabla} \cdot f(\vec{u})\| \cdot c(\vec{u}) \\ h(T, \theta) = q(T, \theta) = q(t, \theta) \\ g(\vec{u}) = \|\vec{\nabla} \cdot f(\vec{u})\| \cdot c(\vec{u}) \end{cases} \Rightarrow L(\vec{u}, \theta) = g(\vec{u}) \cdot h(T, \theta)$$

То есть чтобы T было достаточной статистикой, необходимо, чтобы функция правдоподобия допускала факторизацию.

□

Замечание 2.6.9

Теорема остаётся справедливой для дискретных распределений, где

функция правдоподобия выглядит следующим образом

$$L(\vec{u}, \theta) = \prod_{k=1}^n \mathbb{P}\{x_k = u_k\}$$

Пример 2.6.10

x_1, \dots, x_n — выборка из нормального распределения с неизвестным математическим ожиданием $N(\theta, 1)$, $\theta \in \mathbb{R}$.

Рассмотрим функцию правдоподобия

$$L(\vec{x}, \theta) = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k - \theta)^2 \right\}$$

Рассмотрим сумму под экспонентой и выделим из неё выборочное среднее

$$\begin{aligned} \sum_{k=1}^n (x_k - \theta)^2 &= \sum_{k=1}^n x_k^2 - 2 \cdot \theta \cdot \sum_{k=1}^n x_k + n \cdot \theta^2 = \\ &= \sum_{k=1}^n x_k^2 - 2 \cdot \theta \cdot n \cdot \bar{x} + n \cdot \theta^2 \end{aligned}$$

Теперь выделим квадрат разности выборочного среднего и параметра θ

$$\begin{aligned} &\sum_{k=1}^n x_k^2 - 2 \cdot \theta \cdot n \cdot \bar{x} + n \cdot \theta^2 = \\ &= \sum_{k=1}^n x_k^2 - 2 \cdot \theta \cdot n \cdot \bar{x} + n \cdot \theta^2 + n \cdot \bar{x}^2 - n \cdot \bar{x}^2 = \\ &= \sum_{k=1}^n (x_k^2 - \bar{x}^2) + n \cdot (\bar{x} - \theta)^2 \end{aligned}$$

Вернёмся к исходному выражению и видим экспоненту суммы. Заменим её на произведение экспонент

$$\begin{aligned} &\frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k - \theta)^2 \right\} = \\ &= \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k^2 - \bar{x}^2) + n \cdot (\bar{x} - \theta)^2 \right\} = \\ &= \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k^2 - \bar{x}^2) \right\} \cdot \exp \left\{ n \cdot (\bar{x} - \theta)^2 \right\} \end{aligned}$$

Мы получили произведение двух функций: одна зависит лишь от статистики \bar{x} и параметра θ , а другая, зависит лишь от выборки (так как выборочное среднее является функцией выборки)

$$\begin{cases} L(\vec{x}, \theta) = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k^2 - \bar{x}^2) \right\} \cdot \exp \left\{ n \cdot (\bar{x} - \theta)^2 \right\} \\ h(\bar{x}, \theta) = \exp \left\{ n \cdot (\bar{x} - \theta)^2 \right\} \\ g(\vec{x}) = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (x_k^2 - \bar{x}^2) \right\} \end{cases}$$

$$\Rightarrow L(\vec{x}, \theta) = g(\vec{x}) \cdot h(\bar{x}, \theta)$$

Функция выборки допускает факторизацию со статистикой \bar{x} , а это значит, что выборочное среднее является достаточной статистикой.

Глава 3

Случайные вектора

Мы уже знаем, что нам не нужна вся выборка для построения хороших оценок — нам хватит достаточных статистик. Введя метод наименьших квадратов, мы избавимся от неприятной процедуры вычисления интегралов.

Тем не менее, чтобы перейти непосредственно к изучению метода, необходимо владеть инструментарием, коим являются случайные вектора.

3.1 Основные характеристики случайного вектора

Вспомним, что случайный вектор $\vec{\xi}$ — это набор случайных величин $\vec{\xi} = [\xi_1, \xi_2, \dots, \xi_n]$, которые определены на одном вероятностном пространстве. Распределением случайного вектора называют совместное распределение его компонент.

Если величины $\xi_1, \xi_2, \dots, \xi_n$ имеют совместную плотность распределения, говорят о плотности распределения случайного вектора. Обозначим эту плотность p . Тогда

1. Вероятность того, что вектор $\vec{\xi}$ окажется в множестве Δ , равна интегралу от плотности по этой области

$$\mathbb{P} \left\{ \vec{\xi} \in \Delta \right\} = \int_{\Delta} p(\vec{u}) d\vec{u}$$

2. Во всех точках плотность неотрицательна

$$\forall \vec{x} \in \mathbb{R}^n : p(\vec{x}) \geq 0$$

3. Выполняется условие нормировки

$$\int_{\mathbb{R}^n} p(\vec{u}) d\vec{u} = 1$$

Естественным образом вводится определение характеристической функции.

Определение 3.1.1: Характеристическая функция случайного вектора

Значение характеристической функции случайного вектора $\vec{\xi}$ в точке $\vec{\lambda}$ считается по формуле

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = M e^{i(\vec{\lambda}, \vec{\xi})} = M \exp \left\{ i \cdot \sum_{k=1}^n \lambda_k \cdot \xi_k \right\}$$

Когда существует плотность, имеем преобразование Фурье

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = \int_{\mathbb{R}^n} p(\vec{u}) \cdot e^{i(\vec{\lambda}, \vec{u})} d\vec{u}$$

Определение 3.1.2: Математическое ожидание случайного вектора

Математическое ожидание случайного вектора $\vec{\xi} = (\xi_1, \dots, \xi_n)$ — вектор, элементы которого — математические ожидания компонент случайного вектора $\vec{\xi}$

$$M \vec{\xi} = (M \xi_1, \dots, M \xi_n)$$

Важным свойством является линейность математического ожидания.

Утверждение 3.1.3

Пусть $\vec{\xi} = [\xi_1, \xi_2, \dots, \xi_n]$ — случайный вектор, A — матрица $m \times n$, вектор $\vec{b} \in \mathbb{R}^m$. Тогда случайный вектор $\vec{\eta} = A \vec{\xi} + \vec{b}$ имеет математическое ожидание

$$M \vec{\eta} = A M \vec{\xi} + \vec{b}$$

Доказательство. Действительно, для каждой компоненты $k = \overline{1, n}$ справедливо

$$\begin{aligned} (M \vec{\eta})_k &= M \left[(A \vec{\xi})_k + b_k \right] = M \sum_{i=1}^n (a_{k,i} \cdot \xi_i) + b_k = \\ &= \sum_{i=1}^n (a_{k,i} \cdot M \xi_i) + b_k = (A M \vec{\xi} + \vec{b})_k \end{aligned}$$

□

Но что же является дисперсией случайного вектора? Набор дисперсий компонент $[D \xi_1, D \xi_2, \dots, D \xi_n]$ даёт о случайном векторе слишком мало информации. Например, в теории оценивания нас интересовали выражения

вида $D f(\vec{\xi})$. По набору $[D \xi_1, D \xi_2, \dots, D \xi_n]$ невозможно в общем случае вычислить $D f(\vec{\xi})$ даже для линейной функции $f(\vec{x}) = \alpha_1 \cdot x_1 + \dots + \alpha_n \cdot x_n$. Нужна какая-либо информация о совместном распределении компонент вектора.

3.1.1 Ковариационная матрица случайного вектора

Начнём с определения ковариации двух случайных величин.

Определение 3.1.4: Ковариация

Ковариация двух случайных величин ξ и η , принимающих действительные значения, обозначается $\text{cov}(\xi, \eta)$ и считается по формуле

$$\text{cov}(\xi, \eta) = M[(\xi - M \xi) \cdot (\eta - M \eta)] \quad (3.1)$$

Замечание 3.1.5

Ковариация случайной величины ξ с ней же — её дисперсия

$$\text{cov}(\xi, \xi) = M[(\xi - M \xi) \cdot (\xi - M \xi)] = M[(\xi - M \xi)^2] = D \xi$$

Замечание 3.1.6

Ковариация симметрична

$$\text{cov}(\xi, \eta) = M[(\xi - M \xi) \cdot (\eta - M \eta)] = M[(\eta - M \eta) \cdot (\xi - M \xi)] = \text{cov}(\eta, \xi)$$

Замечание 3.1.7

Ковариация двух независимых случайных величин равна нулю [13, с. 244]

Раскрыв скобки в (3.1), получим ещё одно выражение для ковариации

$$\text{cov}(\xi, \eta) = M[\xi \cdot \eta] - M \xi \cdot M \eta$$

Определение 3.1.8: Ковариационная матрица случайного вектора

Ковариационная матрица случайного вектора $\vec{\xi} = (\xi_1, \dots, \xi_n)$ — матрица, на пересечении i строки и j столбца которой находятся ковариации i и j элементов вектора ξ

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = \|\text{cov}(\xi_i, \xi_j)\|_{i,j=1}^n = \|M\{(\xi_i - M \xi_i) \cdot (\xi_j - M \xi_j)\}\|_{i,j=1}^n$$

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} \text{cov}(\xi_1, \xi_1) & \cdots & \text{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(\xi_n, \xi_1) & \cdots & \text{cov}(\xi_n, \xi_n) \end{bmatrix}$$

Замечание 3.1.9

На диагонали ковариационной матрицы $\text{Cov}_{\vec{\xi}, \vec{\xi}}$ случайного вектора ξ стоят дисперсии компонент вектора.

Случайный вектор находится во многомерном пространстве, а это значит, что имеется много направлений его размазывания, поэтому в качестве дисперсии нам нужна матрица.

Пример 3.1.10

Возьмём двумерный вектор с одним и тем же элементом в каждой координате — случайной величиной из стандартного нормального распределения

$$\vec{\xi} = (\xi, \xi), \quad \xi \sim N(0, 1)$$

Нетрудно посчитать, что ковариационная матрица будет заполнена единицами, так как во всех ячейках будет ковариация $\text{cov}(\xi, \xi)$, равная дисперсии случайной величины ξ , то есть единице

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Пример 3.1.11

Возьмём опять же двумерный вектор, но с двумя независимыми случайными величинами из стандартного нормального распределения

$$\vec{\xi} = (\xi_1, \xi_2), \quad \xi_1, \xi_2 \sim N(0, 1)$$

На диагонали будут стоять единицы — дисперсии случайных величин. Если две случайные величины независимы, то их ковариация равна нулю (замечание 3.1.7). Это в свою очередь означает, что вне диагонали будут нули

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Определение 3.1.12: Сопряжённая матрица

Есть матрица A размером $n \times m$ с комплексными элементами. Тогда сопряжённая к ней матрица A^* получается путём транспонирования матрицы A и замены всех элементов на комплексно-сопряжённые [10,

с. 243], то есть

$$(a_{i,j}^* = \overline{a_{j,i}}), \quad A \in \mathbb{C}^{n \times m}, A^* \in \mathbb{C}^{m \times n}$$

Или же в таком виде

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix} \Rightarrow A^* = \begin{bmatrix} \overline{a_{1,1}} & \cdots & \overline{a_{n,1}} \\ \vdots & \ddots & \vdots \\ \overline{a_{1,m}} & \cdots & \overline{a_{n,m}} \end{bmatrix}$$

Замечание 3.1.13

Отметим, что к матрице с действительными коэффициентами сопряжённой будет транспонированная матрица

$$A \in \mathbb{R}^{n \times m} \Rightarrow A^* = A^T$$

Перейдём к свойствам ковариационной матрицы

1. Симметричность. Ковариационная матрица случайного вектора $\vec{\xi}$ равна своей сопряжённой

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} = \text{Cov}_{\vec{\xi}, \vec{\xi}}^*$$

2. Неотрицательная определённость*

$$\text{Cov}_{\vec{\xi}, \vec{\xi}} \geq 0$$

Это значит следующее

$$\forall \vec{u} \in \mathbb{R}^n : (\text{Cov}_{\vec{\xi}, \vec{\xi}} \cdot \vec{u}, \vec{u}) = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j) \cdot u_j \cdot u_i \geq 0$$

Доказательство. Распишем ковариацию по определению

$$\begin{aligned} \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j) \cdot u_j \cdot u_i &= \sum_{i,j=1}^n \text{M}[(\xi_i - \text{M} \xi_i) \cdot (\xi_j - \text{M} \xi_j)] \cdot u_j \cdot u_i = \\ &= \text{M} \left[\sum_{i=1}^n (\xi_i - \text{M} \xi_i) \cdot u_i \cdot \sum_{j=1}^n (\xi_j - \text{M} \xi_j) \cdot u_j \right] = \text{M} \left(\sum_{t=1}^n u_t \cdot (\xi_t - \text{M} \xi_t) \right)^2 \geq 0 \end{aligned}$$

□

3. Для неслучайного вектора $\vec{a} \in \mathbb{R}^n$

$$\text{Cov}_{\vec{\xi} + \vec{a}, \vec{\xi} + \vec{a}} = \text{Cov}_{\vec{\xi}, \vec{\xi}}$$

*Больше о неотрицательно определённых операторах можно почитать в книге Ильина и Позняка “Линейная алгебра” [7, с. 139]. В ней такой оператор называется положительным.

Замечание 3.1.14

Вспомним линейную алгебру.

Самосопряжённая неотрицательно определённая матрица $\text{Cov}_{\vec{\xi}, \vec{\xi}}$ имеет собственный ортонормированный базис, в котором она превращается в диагональную матрицу с неотрицательными элементами

$$\begin{bmatrix} \lambda_1 & & \emptyset \\ & \ddots & \\ \emptyset & & \lambda_n \end{bmatrix}, \lambda_k \geq 0$$

Далее будем упускать нули, подразумевая диагональные матрицы. Как эта матрица преобразует пространство?

Единичная матрица не меняет ничего

$$\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$$

Если первый элемент единичной матрицы сделать нулём, то такой оператор убивает первую координату вектора, на который действует

$$\begin{bmatrix} 0 & & \\ & 1 & \\ & & \ddots \\ & & & 1 \end{bmatrix}$$

А такая матрица усиливает первую составляющую в десять раз и ослабляет остальные в десять раз

$$\begin{bmatrix} 10 & & \\ & 0.1 & \\ & & \ddots \\ & & & 0.1 \end{bmatrix}$$

Оказывается, через ковариационную матрицу вычисляются все характеристики линейных преобразований.

3.1.2 Ковариационная матрица

Логичным обобщением ковариационной матрицы случайного вектора является ковариационная матрица двух случайных векторов. Сейчас станет ясно, зачем мы дважды писали вектор $\vec{\xi}$ в индексе оператора $\text{Cov}_{\vec{\xi}, \vec{\xi}}$.

Определение 3.1.15: Ковариационная матрица

Ковариационная матрица двух случайных векторов $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$ и $\vec{\beta} = (\beta_1, \dots, \beta_m)$ — матрица, в которой на пересечении i строки и j столбца стоит ковариация случайных величин α_i и β_j

$$\text{Cov}_{\vec{\alpha}, \vec{\beta}} = \|\text{cov}(\alpha_i, \beta_j)\|_{\substack{i=\overline{1,n}, \\ j=\overline{1,m}}} = \|\text{M}\{(\alpha_i - \text{M}\alpha_i) \cdot (\beta_j - \text{M}\beta_j)\}\|_{\substack{i=\overline{1,n}, \\ j=\overline{1,m}}}$$

$$\text{Cov}_{\vec{\alpha}, \vec{\beta}} = \begin{bmatrix} \text{cov}(\alpha_1, \beta_1) & \cdots & \text{cov}(\alpha_1, \beta_m) \\ \vdots & \ddots & \vdots \\ \text{cov}(\alpha_n, \beta_1) & \cdots & \text{cov}(\alpha_n, \beta_m) \end{bmatrix}$$

Теорема 3.1.16: Свойства ковариационной матрицы

Ковариационная матрица обладает следующими свойствами

1. Перестановка векторов ведёт к транспонированию матрицы (симметричность)

$$\text{Cov}_{\beta, \alpha} = \text{Cov}_{\alpha, \beta}^T$$

2. Линейность относительно первого аргумента

$$\text{Cov}_{B\alpha_1 + C\alpha_2, \beta} = B \text{Cov}_{\alpha_1, \beta} + C \text{Cov}_{\alpha_2, \beta}$$

3. Линейность относительно второго аргумента

$$\text{Cov}_{\alpha, D\beta_1 + E\beta_2} = \text{Cov}_{\alpha, \beta_1} D^T + \text{Cov}_{\alpha, \beta_2} E^T$$

4. Ковариация вектора при линейном преобразовании

$$\text{Cov}_{B\vec{\xi}, B\vec{\xi}} = B \text{Cov}_{\vec{\xi}, \vec{\xi}} B^T$$

Доказательство. 1. Следует из равенства

$$\|\text{cov}(\vec{\beta}, \vec{\alpha})\|_{i,j} = \|\text{M}[(\beta_i - \text{M}\beta_i) \cdot (\alpha_j - \text{M}\alpha_j)]\|_{i,j} = \|\text{cov}(\vec{\alpha}, \vec{\beta})\|_{j,i}$$

2. Имеем

$$\begin{aligned} & \|\text{cov}(B\vec{\alpha}_1 + C\vec{\alpha}_2, \vec{\beta})\|_{i,j} = \\ & = \|\text{M}[(B\vec{\alpha}_1 + C\vec{\alpha}_2 - \text{M}[B\vec{\alpha}_1 + C\vec{\alpha}_2]) \cdot (\beta_j - \text{M}\beta_j)]\|_{i,j} = \\ & = \left\| \text{M} \left[\sum_{k=1}^n (b_{ik} \cdot \alpha_1^k + c_{ik} \cdot \alpha_2^k - b_{ik} \text{M}\alpha_1^k - c_{ik} \text{M}\alpha_2^k) \cdot (\beta_j - \text{M}\beta_j) \right] \right\|_{i,j} = \\ & = \left\| \sum_{k=1}^n b_{ik} \cdot \text{cov}(\alpha_1^k, \beta_j) + \sum_{k=1}^n c_{ik} \cdot \text{cov}(\alpha_2^k, \beta_j) \right\|_{i,j} = B \text{Cov}_{\vec{\alpha}_1, \vec{\beta}} + C \text{Cov}_{\vec{\alpha}_2, \vec{\beta}} \end{aligned}$$

3. Аналогично 2

4. Получается последовательным применением 2 и 3

□

3.2 Линейные преобразования случайных векторов

Рассмотрим всё тот же случайный вектор $\vec{\xi} = (\xi_1, \dots, \xi_n)$ и произвольный постоянный вектор $\vec{\lambda} \in \mathbb{R}^n$.

3.2.1 Скалярное произведение

Определим случайную величину η как скалярное произведение векторов $\vec{\xi}$ и $\vec{\lambda}$

$$\eta = (\vec{\xi}, \vec{\lambda})$$

Посчитаем математическое ожидание случайной величины η .

$$M \eta = M \sum_{k=1}^n \lambda_k \cdot \xi_k = \sum_{k=1}^n \lambda_k \cdot M \xi_k = (\vec{\lambda}, M \vec{\xi}) \quad (3.2)$$

Теперь посчитаем дисперсию

$$D \eta = M (\eta - M \eta)^2 = M \left(\sum_{k=1}^n \lambda_k \cdot \xi_k - \lambda_k \cdot M \xi_k \right)^2$$

Полученное выражение сворачивается в математическое ожидание квадрата суммы, которая превращается в двойную сумму произведений

$$M \left\{ \sum_{k=1}^n \lambda_k \cdot (\xi_k - M \xi_k) \right\}^2 = \sum_{i,j=1}^n M [(\xi_i - M \xi_i) \cdot (\xi_j - M \xi_j)] \cdot \lambda_i \cdot \lambda_j$$

А это, как мы уже знаем, произведение ковариационной матрицы вектора $\vec{\xi}$ на вектор $\vec{\lambda}$, умноженное на тот же вектор $\vec{\lambda}$. То есть дисперсия η выражается следующим образом

$$D \eta = D (\vec{\xi}, \vec{\lambda}) = (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda}) \quad (3.3)$$

Оформим вывод в виде утверждения.

Утверждение 3.2.1

Скалярное произведение случайного вектора $\vec{\xi}$ и произвольного вектора $\vec{\lambda}$ является случайной величиной с математическим ожиданием a и дисперсией σ^2

$$a = (\vec{\lambda}, M \vec{\xi}), \quad \sigma^2 = (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda})$$

Обобщением утверждения 3.2.1 на случай линейного преобразования вектора является следующее утверждение

Утверждение 3.2.2

Если на случайный вектор $\vec{\xi}$ подействовать линейным оператором T , то получим случайный вектор с математическим ожиданием \vec{a} и ковариационной матрицей A

$$\begin{aligned} M T \vec{\xi} &= T M \vec{\xi} \\ \text{Cov}_{T \vec{\xi}, T \vec{\xi}} &= T \text{Cov}_{\vec{\xi}, \vec{\xi}} T^* \end{aligned}$$

Доказательство. Утверждение 3.1.3 и теорема 3.1.16 (пункт 4). \square

Утверждение 3.2.3

Есть два линейных оператора $C \in \mathbb{R}^{k \times n}$ и $D \in \mathbb{R}^{k \times m}$, два случайных вектора: вектор $\vec{\xi}$ из \mathbb{R}^n с параметрами \vec{a} и A , вектор $\vec{\eta}$ из \mathbb{R}^m с параметрами \vec{b} и B .

В таком случае сумма случайных векторов, полученных с помощью операторов C и D , будет случайным вектором в \mathbb{R}^k с параметрами \vec{m} и M

$$\begin{aligned} \vec{m} &= C M \vec{\xi} + D M \vec{\eta} \\ M &= C \text{Cov}_{\vec{\xi}, \vec{\xi}} C^T + C \text{Cov}_{\vec{\xi}, \vec{\eta}} D^T + D \text{Cov}_{\vec{\eta}, \vec{\xi}} C^T + D \text{Cov}_{\vec{\eta}, \vec{\eta}} D^T \end{aligned}$$

Доказательство. Утверждение 3.1.3 и теорема 3.1.16 (пункты 1, 2 и 3). \square

3.3 Гауссовские случайные вектора

3.3.1 Определения

Определение 3.3.1: Гауссовский случайный вектор (определение Максвелла)

Случайный вектор $\vec{\xi}$ в \mathbb{R}^n называется гауссовским, если его проекция на произвольный вектор из пространства \mathbb{R}^n является гауссовской случайной величиной

$$\forall \vec{\lambda} \in \mathbb{R}^n : (\vec{\lambda}, \vec{\xi}) \sim N(a_{\vec{\lambda}}, \sigma_{\vec{\lambda}}^2)$$

Пример 3.3.2

Возьмём случайный вектор $\vec{\xi}$, координаты которого между собой рав-

ны и являются гауссовской случайной величиной

$$\vec{\xi} = (\xi, \dots, \xi), \quad \xi \sim N(a, \sigma^2)$$

Очевидно, что математическое ожидание — вектор (a, \dots, a) , а ковариационная матрица состоит из σ^2 , так как на каждом месте стоит дисперсия случайной величины ξ

$$M \vec{\xi} = (a, \dots, a), \quad \text{Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} \sigma^2 & \dots & \sigma^2 \\ \vdots & \ddots & \vdots \\ \sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

Проверим, является ли вектор $\vec{\xi}$ гауссовским. Возьмём произвольный вектор $\vec{\lambda} \in \mathbb{R}^n$ и посмотрим, чему равно скалярное произведение

$$(\vec{\lambda}, \vec{\xi}) = \sum_{k=1}^n (\lambda_k \cdot \xi) = \xi \cdot \sum_{k=1}^n \lambda_k$$

Получилось произведение случайной гауссовской величины и константы, а распределение такой величины мы знаем. Для компактности записи заменим сумму большой буквой “лямбда” Λ

$$\xi \cdot \sum_{k=1}^n \lambda_k = \xi \cdot \Lambda \sim N(a \cdot \Lambda, \sigma^2 \cdot \Lambda^2)$$

Вывод: данный вектор $\vec{\xi} = (\xi, \dots, \xi)$ является гауссовским для любой нормально распределённой случайной величины ξ .

Пример 3.3.3

Теперь возьмём случайный вектор $\vec{\xi}$, состоящий из n независимых случайных гауссовских величин со своими математическими ожиданиями и дисперсиями

$$\vec{\xi} = (\xi_1, \dots, \xi_n), \quad \xi_k \sim N(a_k, \sigma_k^2)$$

С математическим ожиданием всё очевидно: это вектор из a_k . Ковариация — диагональная матрица дисперсий, так как вне диагонали должны стоять ковариации случайных величин между собой, но они независимы, а это значит, что их ковариации равны нулю (замечание 3.1.7)

$$M \vec{\xi} = (a_1, \dots, a_n), \quad \text{Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$

Снова рассматриваем скалярное произведением и видим, что ре-

зультат сложнее, но схож с полученным в предыдущем примере 3.3.2

$$(\vec{\lambda}, \vec{\xi}) = \sum_{k=1}^n \lambda_k \cdot \xi_k \sim N \left(\sum_{k=1}^n \lambda_k \cdot a_k, \sum_{k=1}^n \lambda_k^2 \cdot \sigma_k^2 \right)$$

Вывод: вектор, состоящий из независимых гауссовских случайных величин, является гауссовским.

Определение 3.3.4: Стандартный гауссовский вектор

Гауссовский случайный вектор $\vec{\xi}$ называется стандартным гауссовским вектором, если его математическое ожидание — нулевой вектор, а ковариационная матрица — единичная матрица

$$M \vec{\xi} = \vec{0}, \text{ Cov}_{\vec{\xi}, \vec{\xi}} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$$

Тут возникает вопрос: однозначно ли математическое ожидание и ковариационная матрица определяют распределение гауссовского случайного вектора?

Следующая лемма показывает, что такого определения нам вполне достаточно.

Лемма 3.3.5

Распределение гауссовского вектора однозначно определяется его средним и ковариационной матрицей.

Доказательство. Для краткости введём новые обозначения

$$M \vec{\xi} = \vec{a}, \text{ Cov}_{\vec{\xi}, \vec{\xi}} = A$$

Поскольку характеристическая функция однозначно определяет распределение, то достаточно показать, что она является функцией математического ожидания и ковариационной матрицы

$$\varphi_{\vec{\xi}} = f(\vec{a}, A)$$

Введём старое обозначение скалярного произведения случайного вектора $\vec{\xi}$ с произвольным вектором $\vec{\lambda}$

$$\eta = (\vec{\lambda}, \vec{\xi})$$

И начнём писать, чему равна характеристическая функция. Ясно, что она будет функцией не числа t , а вектора $\vec{\lambda}$

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = M e^{i \cdot (\vec{\lambda}, \vec{\xi})} = M e^{i \cdot \eta} = \varphi_{\eta}(1)$$

Случайная величина η является гауссовской по определению гауссовского вектора 3.3.1, а это значит, что её характеристическая функция имеет вид

$$\varphi_{\eta}(t) = \exp \left\{ i \cdot t \cdot M \eta - \frac{t^2 \cdot D \eta}{2} \right\}$$

Очевидно, что в точке $t = 1$ она принимает значение

$$\varphi_{\eta}(1) = \exp \left\{ i \cdot M \eta - \frac{D \eta}{2} \right\}$$

Из начала подраздела 3.2 о линейных преобразованиях помним формулы для математического ожидания (3.2) и дисперсии (3.3) случайной величины η , которая является скалярным произведением случайного вектора $\vec{\xi}$ с произвольным вектором $\vec{\lambda}$

$$M \eta = (\vec{\lambda}, M \vec{\xi}), \quad D \eta = (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda}) = (A \vec{\lambda}, \vec{\lambda})$$

Перепишем характеристическую функцию, воспользовавшись тем, что имеем

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = \exp \left\{ i \cdot (\vec{\lambda}, \vec{a}) - \frac{1}{2} \cdot (A \vec{\lambda}, \vec{\lambda}) \right\}$$

Видим, что характеристическая функция полностью восстанавливающая распределение, определяется исключительно математическим ожиданием и ковариационной матрицей, что и требовалось доказать. \square

Определение 3.3.6: Гауссовское распределение

Тот факт, что случайный вектор $\vec{\xi}$ имеет гауссовское распределение со средним \vec{a} и ковариационной матрицей A , будем обозначать привычным образом

$$\vec{\xi} \sim N(\vec{a}, A)$$

Замечание 3.3.7: Характеристическая функция гауссовского распределения

Как было показано в предыдущей лемме 3.3.5, характеристическая функция вектора $\vec{\xi}$, имеющего гауссовское распределение с параметрами \vec{a} и A , имеет следующий вид

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = \exp \left\{ i \cdot (\vec{\lambda}, \vec{a}) - \frac{1}{2} \cdot (A \vec{\lambda}, \vec{\lambda}) \right\}, \quad \vec{\xi} \sim N(\vec{a}, A)$$

3.3.2 Линейные преобразования

Лемма 3.3.8

Пусть $\vec{\xi}$ — случайный n -элементный вектор, имеющий гауссовское

распределение с параметрами \vec{a} и A

$$\vec{\xi} \sim N(\vec{a}, A), \quad \vec{b} \in \mathbb{R}^m, \quad T \in \mathbb{R}^{m \times n}$$

Тогда

$$T\vec{\xi} + \vec{b} \sim N(T\vec{a} + \vec{b}, T A T^*)$$

Доказательство. Из подраздела 3.2 мы знаем, что оператор меняет среднее значение и ковариационную матрицу именно таким образом, как это указано в лемме. Значит, нам нужно проверить то, что вектор остался гауссовским. Воспользовавшись определением 3.3.1, видим, что всё прекрасно

$$\forall \vec{e} \in \mathbb{R}^n : \quad (\vec{e}, T\vec{\xi} + \vec{b}) = (T^* \vec{e}, \vec{\xi}) + (\vec{e}, \vec{b})$$

Первое слагаемое — гауссовская случайная величина, поскольку $\vec{\xi}$ — гауссовский случайный вектор, а второе — число. В сумме получаем гауссовскую случайную величину, что и требовалось доказать. \square

Теорема 3.3.9

Для произвольных $\vec{a} \in \mathbb{R}^n$ и симметричной неотрицательной матрицы A существует гауссовский вектор с распределением $N(\vec{a}, A)$

Доказательство. Пускай $\vec{\xi}$ — стандартный гауссовский вектор в \mathbb{R}^n

$$\vec{\xi} \sim N(0, I)$$

Тогда возьмём неотрицательную матрицу V , вектор \vec{a} и рассмотрим случайный вектор $\vec{\eta}$

$$\vec{\eta} = V\vec{\xi} + \vec{a}$$

Из леммы 3.3.8 знаем, что новый вектор будет иметь гауссовское распределение с параметрами \vec{a} и $VIV^* = VV^*$

$$\eta \sim N(\vec{a}, VV^*)$$

Теперь задача состоит в том, чтобы подобрать такую матрицу V , чтобы её произведение с сопряжённой равнялось нужной нам A

$$VV^* = A$$

В замечании 3.1.14 мы вспомнили, что самосопряжённая неотрицательная определённая матрица A имеет собственный ортонормированный базис, в котором она превращается в диагональную матрицу с неотрицательными элементами

$$\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \quad \lambda_k \geq 0$$

Создадим в этом базисе самосопряжённую матрицу V , в ячейках которой которой будут корни соответствующих элементов исходной матрицы (в этом базисе)

$$\begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}, \lambda_k \geq 0$$

Диагональная матрица с действительными элементами очевидно является самосопряжённой. То есть произведение матрицы V на сопряжённую к ней матрицу V^* в этом базисе будет давать исходную матрицу A .

Останется лишь перейти в исходный базис и матрица будет готова. То есть существует такая матрица V , что вектор $\vec{\eta}$ будет иметь нужное нам распределение

$$\exists \vec{a}, V : \vec{\eta} = V V^* \vec{\xi} + \vec{a} \sim N(\vec{a}, A)$$

□

Замечание 3.3.10

Поскольку самосопряжённая матрица остаётся таковой в любом ортонормированном базисе, видим, что построенная матрица V является самосопряжённой.

Замечание 3.3.11

Так как определитель инвариантен, имеем

$$\det V = \sqrt{\lambda_1} \cdot \sqrt{\lambda_2} \cdot \sqrt{\lambda_n} = \sqrt{\det A}$$

Теорема 3.3.12: Плотность распределения случайного гауссовского вектора

Если матрица A невырожденная $A > 0 \Leftrightarrow \det A \neq 0$, то у случайного вектора, имеющего гауссовское распределение с параметрами \vec{a} и A , есть плотность распределения

$$p(\vec{u}) = \frac{1}{\sqrt{2 \cdot \pi^n} \cdot \sqrt{\det A}} \cdot \exp \left\{ -\frac{1}{2} \cdot (A^{-1}(\vec{u} - \vec{a}), \vec{u} - \vec{a}) \right\}$$

Доказательство. Из доказательства прошлой теоремы 3.3.9 помним, что для невырожденной матрицы A найдётся такая матрица V , что при её умножении на сопряжённую получится матрица A

$$\exists V = A^{\frac{1}{2}} : V V^* = V^2 = A$$

Там же мы построили вектор, имеющий гауссовское распределение с параметрами \vec{a} и A

$$\vec{\xi} \sim N(\vec{0}, I), \vec{\eta} = V \vec{\xi} + \vec{a} \sim N(\vec{a}, A)$$

В примере 3.3.3 было показано, что случайный вектор, состоящий из случайных гауссовских величин, является гауссовским. Если взять ξ_1, \dots, ξ_n из стандартного нормального распределения, то получим стандартный гауссовский вектор. Плотность такого вектора — произведение плотностей его координат

$$q(\vec{v}) = q(v_1, \dots, v_n) = \prod_{k=1}^n \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot v_k^2} = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot e^{-\frac{1}{2} \cdot \sum_{k=1}^n v_k^2}$$

То есть плотность распределения стандартного гауссовского вектора выглядит следующим образом

$$q(\vec{v}) = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot e^{-\frac{1}{2} \cdot (\vec{v}, \vec{v})}$$

Распишем, чему равна вероятность того, что вектор $\vec{\eta}$ очутился в области $\Delta \in \mathfrak{B}$

$$\mathbb{P}\{\vec{\eta} \in \Delta\} = \mathbb{P}\{V\vec{\xi} + \vec{a} \in \Delta\} = \mathbb{P}\{\vec{\xi} \in \{\vec{v} : V\vec{v} + \vec{a} \in \Delta\}\} \quad (3.4)$$

Перепишем вероятность через интеграл

$$\mathbb{P}\{\vec{\xi} \in \{\vec{v} : V\vec{v} + \vec{a} \in \Delta\}\} = \int_{\{\vec{v} : V\vec{v} + \vec{a} \in \Delta\}} q(\vec{v}) d\vec{v}$$

Введём замену

$$\vec{u} = V\vec{v} + \vec{a} \Rightarrow \vec{v} = V^{-1}(\vec{u} - \vec{a})$$

Поскольку дифференциал $d\vec{u}$ — площадь n -мерного параллелограмма, а матрица V^{-1} ортогональна, то она будет увеличивать площадь параллелограмма в $\det V^{-1}$ раз

$$d\vec{v} = \det V^{-1} \cdot d\vec{u}$$

Говоря формальным языком, имеем Якобиан перехода

$$\det \left\| \frac{\partial \vec{v}}{\partial \vec{u}} \right\| = \det V^{-1} = \frac{1}{\det V} = \frac{1}{\sqrt{\det A}}$$

Можем переписать интеграл

$$\int_{\{\vec{v} : V\vec{v} + \vec{a} \in \Delta\}} q(\vec{v}) d\vec{v} = \int_{\Delta} q(V^{-1}(\vec{u} - \vec{a})) \cdot \frac{1}{\sqrt{\det A}} d\vec{u} \quad (3.5)$$

Рассмотрим скалярное произведение, которое получится в результате расписывания плотности и перенесём матрицу с правой части в левую

$$(V^{-1}(\vec{u} - \vec{a}), V^{-1}(\vec{u} - \vec{a})) = ((V^{-1})^* V^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))$$

Сопряжённая к обратной матрице — обратная к сопряжённой, а наша матрица самосопряжённая, поэтому просто получаем обратную

$$\begin{aligned} ((V^{-1})^* V^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a})) &= ((V^*)^{-1} V^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a})) = \\ &= (V^{-1} V^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a})) = (V^{-2}(\vec{u} - \vec{a}), (\vec{u} - \vec{a})) \end{aligned}$$

Помним, что квадрат матрицы V — это матрица A , и вводим эту замену

$$V^2 = A \Rightarrow (V^{-2}(\vec{u} - \vec{a}), (\vec{u} - \vec{a})) = (A^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))$$

Возвращаемся к интегралу (3.5) и вводим только что оговоренную замену

$$\int_{\Delta} q(V^{-1}(\vec{u} - \vec{a})) \cdot \det V^{-1} d\vec{u} = \int_{\Delta} \frac{1}{\sqrt{2 \cdot \pi^n}} \cdot \frac{1}{\sqrt{\det A}} \cdot e^{(A^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))} d\vec{u}$$

Таким образом

$$\mathbb{P}\{\vec{\eta} \in \Delta\} = \int_{\Delta} p(\vec{u}) d\vec{u} = \int_{\Delta} \frac{\exp\{(A^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))\}}{\sqrt{2 \cdot \pi^n} \cdot \sqrt{\det A}} d\vec{u}$$

Отсюда уже очевидно, что плотность случайного гауссовского вектора с параметрами \vec{a} и A считается по формуле

$$p(\vec{u}) = \frac{1}{\sqrt{2 \cdot \pi^n} \cdot \sqrt{\det A}} \cdot \exp\left\{-\frac{1}{2} \cdot (A^{-1}(\vec{u} - \vec{a}), \vec{u} - \vec{a})\right\}$$

□

Далее очевидно было бы рассмотреть тот случай, когда ковариационная матрица вырождена. Именно этим мы сейчас и займёмся.

Утверждение 3.3.13

Если матрица A вырождена $\det A = 0$, то не существует плотности распределения у гауссовского вектора с параметрами \vec{a} и A

Доказательство. По определению 3.1.8 ковариационная матрица считается по формуле

$$A = \text{Cov}_{\vec{\xi}, \vec{\xi}} = \|\text{cov}(\xi_i, \xi_j)\|_{i,j=1}^n$$

Если определитель равен нулю, то это значит, что строки матрицы линейно зависимы. То есть найдётся такой ненулевой вектор $\vec{\alpha}$, что его произведение с любым столбиком матрицы будет давать ноль

$$(\exists \vec{\alpha} \neq \vec{0}) (\forall j) : \sum_{i=1}^n \alpha_i \cdot \text{cov}(\xi_i, \xi_j) = 0$$

Имеем право внести сумму и коэффициенты под знак ковариации[†]

$$(\exists \vec{\alpha} \neq \vec{0}) (\forall j) : \text{cov}\left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \xi_j\right) = 0$$

[†]Третье свойство ковариации [4, с. 310]

Поскольку это равенство выполняется для любых j , то их сумма тоже будет равняться нулю, причём каждое слагаемое можно умножить на любую константу. Пусть это будут α_j в каждом слагаемом

$$\exists \vec{\alpha} \neq \vec{0} : \sum_{j=1}^n \alpha_j \cdot \text{cov} \left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \xi_j \right) = 0$$

Эту сумму мы внесём во второй аргумент ковариации

$$\exists \vec{\alpha} \neq \vec{0} : \text{cov} \left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \sum_{j=1}^n \alpha_j \cdot \xi_j \right) = 0$$

Заменяя j на i во второй сумме (ничего не поменяется кроме обозначения), явно видно, что перед нами ковариация случайной величины с самой собой, что равняется её дисперсии

$$\exists \vec{\alpha} \neq \vec{0} : \text{cov} \left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \sum_{i=1}^n \alpha_i \cdot \xi_i \right) = D \sum_{i=1}^n \alpha_i \cdot \xi_i = D \left(\vec{\alpha}, \vec{\xi} \right) = 0$$

То есть случайная величина, равная скалярному произведению некоего ненулевого вектора $\vec{\alpha}$ и случайного вектора $\vec{\xi}$, не рассеивается. Значит, она равна константе с единичной вероятностью

$$D \left(\vec{\alpha}, \vec{\xi} \right) = 0 \Rightarrow \exists c : \mathbb{P} \left\{ \left(\vec{\alpha}, \vec{\xi} \right) = c \right\} = 1$$

То есть мы получили уравнение гиперплоскости — $(n-1)$ -мерное подпространство

$$H = \{ \vec{x} \in \mathbb{R}^n \mid (\vec{\alpha}, \vec{x}) = c \}$$

Вероятность того, что случайный вектор $\vec{\xi}$ попадёт в это подпространство, равна единице

$$\mathbb{P} \left\{ \vec{\xi} \in H \right\} = 1$$

Объём подпространства H в n -мерном пространстве равен нулю □

Раздел о линейных преобразованиях гауссовских векторов было бы логично закончить утверждением о результате линейных преобразований гауссовских векторов.

Утверждение 3.3.14

Есть два линейных оператора $C \in \mathbb{R}^{k \times n}$ и $D \in \mathbb{R}^{k \times m}$, два случайных гауссовских вектора: вектор $\vec{\xi}$ из \mathbb{R}^n с параметрами \vec{a} и A , вектор $\vec{\eta}$ из \mathbb{R}^m с параметрами \vec{b} и B .

В таком случае сумма случайных векторов, полученных с помощью операторов C и D будет случайным гауссовским вектором в \mathbb{R}^k с параметрами \vec{m} и M

$$C \vec{\xi} + D \vec{\eta} \sim N(\vec{m}, M)$$

Параметры следующие

$$\vec{m} = C M \vec{\xi} + D M \vec{\eta}$$

$$M = C \text{Cov}_{\vec{\xi}, \vec{\xi}} C^T + C \text{Cov}_{\vec{\xi}, \vec{\eta}} D^T + D \text{Cov}_{\vec{\eta}, \vec{\xi}} C^T + D \text{Cov}_{\vec{\eta}, \vec{\eta}} D^T$$

Доказательство. Начнём с того, что результирующий вектор будет действительно гауссовский по аналогии с доказательством леммы 3.3.8

$$\forall \vec{e} \in \mathbb{R}^k : (\vec{e}, C \vec{\xi} + D \vec{\eta}) = (\vec{e}, C \vec{\xi}) + (\vec{e}, D \vec{\eta}) = (C^* \vec{e}, \vec{\xi}) + (D^* \vec{e}, \vec{\eta})$$

Поскольку и вектор $\vec{\xi}$, и вектор $\vec{\eta}$ являются гауссовскими случайными векторами, то любые скалярные произведения, где в левой части стоит произвольный вектор, а в правой — гауссовский вектор, будут по определению 3.3.1 неотрицательными. Это значит, что и сумма таких скалярных произведений будет неотрицательна

$$\forall \vec{e} \in \mathbb{R}^k : 0 \leq (C^* \vec{e}, \vec{\xi}) + (D^* \vec{e}, \vec{\eta}) = (\vec{e}, C \vec{\xi} + D \vec{\eta})$$

Дальнейшее уже было показано в утверждении 3.2.3. \square

3.3.3 Независимость компонент гауссовского вектора

Определение 3.3.15: Конкатенация векторов

Есть два вектора $\vec{\xi}$ из пространства \mathbb{R}^n и $\vec{\eta}$ из пространства \mathbb{R}^m

$$\vec{\xi} = (\xi_1, \dots, \xi_n), \vec{\eta} = (\eta_1, \dots, \eta_m)$$

Результатом конкатенации (сцепления) будем называть вектор $\vec{\xi} \circ \vec{\eta}$, первая половина которого состоит из элементов вектора $\vec{\xi}$, а вторая половина — из элементов вектора $\vec{\eta}$. Естественно, результат будет в пространстве \mathbb{R}^{n+m}

$$\vec{\xi} \circ \vec{\eta} = (\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m)$$

Пусть $\vec{\xi} \circ \vec{\eta}$ — гауссовский случайный вектор.

Тогда вектор математического ожидания будет выглядеть следующим образом

$$\vec{a} = M [\vec{\xi} \circ \vec{\eta}] = M \vec{\xi} \circ M \vec{\eta} = (M \xi_1, \dots, M \xi_n, M \eta_1, \dots, M \eta_m)$$

Матрицу ковариации запишем в виде такой мозаики

$$A = \text{Cov}_{\vec{\xi} \circ \vec{\eta}, \vec{\xi} \circ \vec{\eta}} = \begin{bmatrix} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \text{Cov}_{\vec{\xi}, \vec{\eta}} \\ \text{Cov}_{\vec{\eta}, \vec{\xi}} & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{bmatrix} \quad (3.6)$$

Теорема 3.3.16

Для гауссовских случайных векторов независимость эквивалентна некоррелированности. То есть два случайных гауссовских вектора $\vec{\xi}$ и $\vec{\eta}$ независимы тогда и только тогда, когда их ковариационная матрица нулевая

$$\text{Cov}_{\vec{\xi}, \vec{\eta}} = 0$$

Доказательство. Необходимость очевидна: ковариация двух независимых случайных величин равна нулю (смотрите замечание 3.1.7). Поскольку ковариационная матрица этих векторов по определению 3.1.15 состоит из ковариаций элементов векторов, то все ячейки матрицы будут заполнены нулями

$$\text{Cov}_{\vec{\xi}, \vec{\eta}} = \|\text{cov}(\xi_i, \eta_j)\|_{\substack{i=\overline{1,n}, \\ j=\overline{1,m}}} = \|0\|_{\substack{i=\overline{1,n}, \\ j=\overline{1,m}}}$$

Достаточность тоже доказать несложно. Помним, что характеристическая функция суммы двух случайных величин равна произведению характеристических функций тогда и только тогда, когда случайные величины независимы [15, с. 354].

Из определения характеристической функции случайного вектора 3.1.1 видим, что она раскладывается в характеристическую функцию суммы его компонент

$$\varphi_{\vec{\xi}}(\vec{\lambda}) = M e^{i \cdot (\vec{\lambda}, \vec{\xi})} = M \exp \left\{ i \cdot \sum_{k=1}^n \lambda_k \cdot \xi_k \right\}$$

Поэтому тот факт, что характеристическая функция конкатенации $\vec{\xi} \circ \vec{\eta}$ распадётся на произведение характеристических функций вектора $\vec{\xi}$ и вектора $\vec{\eta}$, эквивалентен тому, что составляющие вектора $\vec{\xi}$ не зависят от элементов вектора $\vec{\eta}$. То есть нужно показать следующее

$$\varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) = \varphi_{\vec{\xi}}(\vec{\lambda}) \cdot \varphi_{\vec{\eta}}(\vec{\mu})$$

Естественно, вектора $\vec{\lambda}$ и $\vec{\mu}$ имеют размерности n и m соответственно, как случайные вектора $\vec{\xi}$ и $\vec{\eta}$.

Для профилактики вспомним, как выглядит характеристическая функция такой конкатенации, используя определение математического ожидания, а также плотность случайного гауссовского вектора из теоремы 3.3.12

$$\begin{aligned} \varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) &= \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_{n+m} e^{i \cdot \lambda_1 \cdot x_1} \dots e^{i \cdot \lambda_n \cdot x_n} \cdot e^{i \cdot \mu_1 \cdot y_1} \dots e^{i \cdot \mu_m \cdot y_m} \times \\ &\quad \times \frac{e^{-\frac{1}{2} \cdot (A^{-1}(\vec{u} - \vec{a}), \vec{u} - \vec{a})}}{\sqrt{2 \cdot \pi^n} \cdot \sqrt{\det A}} dx_1 \dots dx_n dy_1 \dots dy_m \end{aligned}$$

К счастью, это уже пройденный этап, и мы просто возьмём готовую формулу из замечания 3.3.7

$$\varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) = \exp \left\{ i \cdot \left(\vec{\lambda} \circ \vec{\mu}, M [\vec{\xi} \circ \vec{\eta}] \right) - \frac{1}{2} \cdot \left(A [\vec{\lambda} \circ \vec{\mu}], \vec{\lambda} \circ \vec{\mu} \right) \right\} \quad (3.7)$$

Прделаем небольшой трюк со скалярным произведением, в котором присутствует математическое ожидание

$$\begin{aligned} \left(\vec{\lambda} \circ \vec{\mu}, M \left[\vec{\xi} \circ \vec{\eta} \right] \right) &= \lambda_1 \cdot M \xi_1 + \dots + \lambda_n \cdot M \xi_n + \mu_1 \cdot M \eta_1 + \dots + \mu_n \cdot M \eta_n = \\ &= \left(\vec{\lambda}, M \vec{\xi} \right) + \left(\vec{\mu}, M \vec{\eta} \right) \end{aligned}$$

Из формулы (3.6) помним, как выглядит ковариационная матрица конкатенации. Применим то, что совместная ковариация случайных векторов $\vec{\xi}$ и $\vec{\eta}$ нулевая

$$\begin{cases} A = \left[\begin{array}{c|c} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \text{Cov}_{\vec{\xi}, \vec{\eta}} \\ \hline \text{Cov}_{\vec{\eta}, \vec{\xi}} & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{array} \right] \\ \text{Cov}_{\vec{\xi}, \vec{\eta}} = \text{Cov}_{\vec{\eta}, \vec{\xi}} = 0 \end{cases} \Rightarrow A = \left[\begin{array}{c|c} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \emptyset \\ \hline \emptyset & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{array} \right]$$

Дальше будем действовать по аналогии с математическим ожиданием

$$\begin{aligned} \left(A \left[\vec{\lambda} \circ \vec{\mu} \right], \vec{\lambda} \circ \vec{\mu} \right) &= \left(\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda} \circ \text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\lambda} \circ \vec{\mu} \right) = \\ &= \left(\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda} \right) + \left(\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu} \right) \end{aligned}$$

С учётом вышесказанного характеристическая функция (3.7) принимает вид

$$\begin{aligned} \varphi_{\vec{\xi} \circ \vec{\eta}} \left(\vec{\lambda} \circ \vec{\mu} \right) &= \exp \left\{ i \cdot \left(\vec{\lambda} \circ \vec{\mu}, M \left[\vec{\xi} \circ \vec{\eta} \right] \right) - \frac{1}{2} \cdot \left(A \left[\vec{\lambda} \circ \vec{\mu} \right], \vec{\lambda} \circ \vec{\mu} \right) \right\} = \\ &= \exp \left\{ i \cdot \left(\vec{\lambda}, M \vec{\xi} \right) + i \cdot \left(\vec{\mu}, M \vec{\eta} \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda} \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu} \right) \right\} \end{aligned}$$

Сгруппировав переменные, видим, что характеристическая функция разбилась на произведение двух характеристических функций

$$\begin{aligned} \varphi_{\vec{\xi} \circ \vec{\eta}} \left(\vec{\lambda} \circ \vec{\mu} \right) &= e^{i \cdot \left(\vec{\lambda}, M \vec{\xi} \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda} \right)} \cdot e^{i \cdot \left(\vec{\mu}, M \vec{\eta} \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu} \right)} = \\ &= \varphi_{\vec{\xi}} \left(\vec{\lambda} \right) \cdot \varphi_{\vec{\eta}} \left(\vec{\mu} \right) \end{aligned}$$

Характеристическая функция конкатенации разбилась в произведение характеристических функций, а это значит, что теорема доказана. \square

Замечание 3.3.17

Важно, чтобы вектор $\vec{\xi} \circ \vec{\eta}$ был гауссовским, иначе $\vec{\xi}$ и $\vec{\eta}$ могут быть зависимыми даже при нулевой ковариации $\text{Cov}_{\vec{\xi}, \vec{\eta}}$. Если $\vec{\xi}$ гауссовский и $\vec{\eta}$ тоже, это ещё не гарантирует того, что их конкатенация будет гауссовским вектором.

3.3.4 Теорема о нормальной корреляции

У нас есть гауссовский вектор $\vec{\xi} \circ \vec{\eta}$ с ненулевой ковариацией

$$\text{Cov}_{\vec{\xi} \circ \vec{\eta}, \vec{\xi} \circ \vec{\eta}} \neq 0$$

Какую нужно подобрать матрицу B , чтобы случайные величины $\vec{\xi} - B \vec{\eta}$ и $\vec{\eta}$ были независимыми?

Воспользуемся только что доказанной теоремой 3.3.16 и выпишем, ковариацию, которая должна равняться нулю

$$\text{Cov}_{\vec{\xi} - B \vec{\eta}, \vec{\eta}} = 0$$

Далее воспользуемся свойством линейности ковариационной матрицы относительно первого аргумента (свойство 2)

$$0 = \text{Cov}_{\vec{\xi} - B \vec{\eta}, \vec{\eta}} = \text{Cov}_{\vec{\xi}, \vec{\eta}} - B \text{Cov}_{\vec{\eta}, \vec{\eta}}$$

Значит, можем выразить матрицу B через ковариационные матрицы

$$B = \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \quad (3.8)$$

Утверждение 3.3.18

Если $\vec{\xi} \circ \vec{\eta}$ — гауссовский вектор, то следующие вектора независимы

$$\vec{\alpha} = \vec{\xi} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \vec{\eta}, \quad \vec{\beta} = \vec{\eta}$$

Теорема 3.3.19: О нормальной корреляции

есть гауссовский вектор $\vec{\xi} \circ \vec{\eta}$ с ненулевой ковариацией

$$\text{Cov}_{\vec{\xi} \circ \vec{\eta}, \vec{\xi} \circ \vec{\eta}} \neq 0$$

Определитель ковариационной матрицы вектора $\vec{\eta}$ положителен

$$\det \text{Cov}_{\vec{\eta}, \vec{\eta}} \geq 0$$

Тогда вектор $\vec{\xi}$ при условии $\vec{\eta}$ — гауссовский случайный вектор

$$\vec{\xi} \Big| \vec{\eta} \sim N(\vec{m}, D)$$

Параметры \vec{m} и D имеют следующий вид

$$\vec{m} = M \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - M \eta)$$

$$D = \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}}$$

Доказательство. Чтобы доказать, что условное математическое ожидание имеет именно то распределение, что нам нужно, мы воспользуемся характеристической функцией. Определим математическое ожидание характеристической функции случайного вектора $\vec{\xi}$ при условии известности вектора $\vec{\eta}$

$$\mathbb{M} \left[\varphi_{\vec{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = \mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi})} \mid \vec{\eta} \right]$$

Прделаем небольшой трюк — добавим и отнимем в правой части скалярного произведения вектор $\mathbf{B} \vec{\eta}$, где матрица \mathbf{B} известна из равенства (3.8)

$$\mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi})} \mid \vec{\eta} \right] = \mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta} + \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right]$$

Воспользуемся аддитивностью (распределительное свойство [10, с. 82]) и симметричностью скалярного произведения, разделив его на сумму двух скалярных произведений

$$\mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta} + \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right] = \mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta}) + i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right]$$

Экспонента суммы разбивается на произведение экспонент. Вторая экспонента очевидно является случайной величиной, измеримой относительно случайного вектора $\vec{\eta}$, так как является её функцией, матрица \mathbf{B} — всего лишь линейный оператор, содержащий константы, а вектор $\vec{\lambda}$ — некий константный вектор

$$e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} = f(\vec{\eta})$$

Значит, эту экспоненту можно вынести за знак условного математического ожидания согласно VI свойству условного математического ожидания

$$\mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta}) + i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right]$$

Случайный вектор, находящийся во второй экспоненте, как было выяснено в начале раздела (утверждение 3.3.18), не зависит от вектора $\vec{\eta}$. Это значит, что и скалярное произведение со всей экспонентой тоже не зависят от него: если мы ничего не знаем о случайном векторе, то как увеличатся наши знания о нём после подсчёта скалярного произведения или даже взятия экспоненты?

Воспользовавшись VII свойством условного ожидания, меняем условное математическое ожидание на простое

$$e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \mathbb{M} \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta})} \mid \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \mathbb{M} e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta})}$$

Видим, что имеется характеристическая функция

$$\mathbb{M} e^{i \cdot (\vec{\lambda}, \vec{\xi} - \mathbf{B} \vec{\eta})} = \varphi_{\vec{\xi} - \mathbf{B} \vec{\eta}}(\lambda)$$

Зафиксируем полученный результат

$$\mathbb{M} \left[\varphi_{\vec{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \varphi_{\vec{\xi} - \mathbf{B} \vec{\eta}}(\lambda) \quad (3.9)$$

Настало время посчитать значение самой характеристической функции. Воспользовавшись формулой из замечания 3.3.7, получим

$$\varphi_{\vec{\xi}-B\vec{\eta}}(\vec{\lambda}) = \exp \left\{ i \cdot \left(\vec{\lambda}, M [\vec{\xi} - B\vec{\eta}] \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\xi}-B\vec{\eta}, \vec{\xi}-B\vec{\eta}} \vec{\lambda}, \vec{\lambda} \right) \right\}$$

Формулы для подсчёта математического ожидания и ковариации у нас есть в утверждении 3.3.14. В нашем случае

$$C\vec{\xi} + D\vec{\eta} = I\vec{\xi} - B\vec{\eta}$$

То есть на вектор $\vec{\xi}$ действует оператор эквивалентного преобразования и его можно будет дальше упустить.

Математическое ожидание выглядит вот так

$$M [\vec{\xi} - B\vec{\eta}] = IM\vec{\xi} - BM\vec{\eta} = M\vec{\xi} - BM\vec{\eta}$$

Ковариация несколько длиннее (не забываем про минус возле матрицы B)

$$\text{Cov}_{\vec{\xi}-B\vec{\eta}, \vec{\xi}-B\vec{\eta}} = I\text{Cov}_{\vec{\xi}, \vec{\xi}}I^T - I\text{Cov}_{\vec{\xi}, \vec{\eta}}B^T - B\text{Cov}_{\vec{\eta}, \vec{\xi}}I^T + B\text{Cov}_{\vec{\eta}, \vec{\eta}}B^T \quad (3.10)$$

C математическим ожиданием уже почти покончили, а вот в ковариации лучше расписать оператор B, поскольку тогда сократится несколько некрасивых вещей. Также учтём, что результат транспонирования произведения двух матриц — произведение транспонированных матриц в обратном порядке

$$(AB)^T = B^T A^T$$

Итак, приступим. У нас есть матрица

$$B = \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1}$$

Это значит, что при транспонировании мы получим следующее

$$B^T = \left(\text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \right)^T \text{Cov}_{\vec{\xi}, \vec{\eta}}^T$$

Помним, что транспонированная обратная матрица — обратная к транспонированной, а также вспоминаем то, что ковариация симметрична и транспонирование лишь меняет порядок аргументов в ней

$$B^T = \left(\text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \right)^T \text{Cov}_{\vec{\xi}, \vec{\eta}}^T = \left(\text{Cov}_{\vec{\eta}, \vec{\eta}}^T \right)^{-1} \text{Cov}_{\vec{\xi}, \vec{\eta}}^T = \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}}$$

Подставим полученный результат в равенство (не забываем про минус перед матрицей B) 3.10

$$\begin{aligned} & I\text{Cov}_{\vec{\xi}, \vec{\xi}}I^T - I\text{Cov}_{\vec{\xi}, \vec{\eta}}B^T - B\text{Cov}_{\vec{\eta}, \vec{\xi}}I^T + B\text{Cov}_{\vec{\eta}, \vec{\eta}}B^T = \\ & = \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}}B^T - B\text{Cov}_{\vec{\eta}, \vec{\xi}} + B\text{Cov}_{\vec{\eta}, \vec{\eta}}B^T = \\ & = \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} + \\ & \quad + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} \end{aligned}$$

Вторая и третья матрицы одинаковые, поэтому можно поставить множитель 2 и сэкономить место. В последнем произведении сократятся матрицы $\text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1}$ и $\text{Cov}_{\vec{\eta}, \vec{\eta}}$

$$\begin{aligned} & \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} + \\ & \quad + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\xi}} = \\ & = \text{Cov}_{\vec{\xi}, \vec{\xi}} - 2 \cdot \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} \end{aligned}$$

Последняя матрица уходит вместе с двойкой и остаётся только разность

$$\text{Cov}_{\vec{\xi} - \text{B } \vec{\eta}, \vec{\xi} - \text{B } \vec{\eta}} = \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}}$$

Итого, равенство 3.9 теперь примет вид

$$\begin{aligned} & \text{M} \left[\varphi_{\vec{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, \text{B } \vec{\eta})} \cdot \varphi_{\vec{\xi} - \text{B } \vec{\eta}}(\lambda) = \\ & = e^{i \cdot (\vec{\lambda}, \text{B } \vec{\eta})} \cdot \exp \left\{ i \cdot \left(\vec{\lambda}, \text{M} \left[\vec{\xi} - \text{B } \vec{\eta} \right] \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\xi} - \text{B } \vec{\eta}, \vec{\xi} - \text{B } \vec{\eta}} \vec{\lambda}, \vec{\lambda} \right) \right\} = \\ & = \exp \left\{ i \cdot \left(\vec{\lambda}, \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \vec{\eta} \right) \right\} \cdot \exp \left\{ i \cdot \left(\vec{\lambda}, \text{M } \vec{\xi} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{M } \vec{\eta} \right) \right\} \times \\ & \quad \times \exp \left\{ -\frac{1}{2} \cdot \left(\left(\text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} \right) \vec{\lambda}, \vec{\lambda} \right) \right\} \end{aligned}$$

Для красоты объединим две экспоненты

$$\begin{aligned} & \text{M} \left[\varphi_{\vec{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = \exp \left\{ i \cdot \left(\vec{\lambda}, \text{M } \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - \text{M } \vec{\eta}) \right) \right\} \times \\ & \quad \times \exp \left\{ -\frac{1}{2} \cdot \left(\left(\text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} \right) \vec{\lambda}, \vec{\lambda} \right) \right\} \end{aligned}$$

То есть при известном $\vec{\eta}$ случайный гауссовский вектор $\vec{\xi}$ имеет гауссовское распределение с такими характеристиками

$$\vec{\xi} \mid \vec{\eta} \sim N(\vec{m}, D)$$

$$\begin{aligned} \vec{m} &= \text{M} \left[\vec{\xi} \mid \vec{\eta} \right] = \text{M } \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - \text{M } \eta) \\ D &= \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}} \end{aligned}$$

□

Замечание 3.3.20

Мы получили формулу для подсчёта условного математического ожидания одного гауссовского вектора относительно другого

$$\text{M} \left[\vec{\xi} \mid \vec{\eta} \right] = \text{M } \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - \text{M } \eta)$$

3.4 Распределения, связанные с нормальным

3.4.1 Распределение Пирсона (хи-квадрат, χ^2)

Определение 3.4.1: Распределение Пирсона

Если ξ_1, \dots, ξ_n — независимые стандартные гауссовские величины, то случайная величина с распределением Пирсона с n степенями свободы является суммой квадратов случайных величин ξ_1, \dots, ξ_n и обозначается греческой буквой “хи” χ_n^2

$$\eta = \xi_1^2 + \dots + \xi_n^2 \sim \chi_n^2$$

Выясним, какой вид имеет функция распределения случайной величины χ_n^2 .

Начнём с определения

$$F_{\chi_n^2}(t) = \mathbb{P}(\xi_1^2 + \dots + \xi_n^2 \leq t)$$

Это, ясное дело, интеграл по всему объёму, где сумма квадратов случайных гауссовских величин не превышает число t . Интегрировать нужно плотность

$$\mathbb{P}(\xi_1^2 + \dots + \xi_n^2 \leq t) = \int \dots \int_{x_1^2 + \dots + x_n^2 \leq t} (2 \cdot \pi)^{-\frac{n}{2}} \cdot e^{-\frac{u_1^2 + \dots + u_n^2}{2}} du_1 \dots du_n$$

Перейдём в полярные координаты в этом многомерном пространстве, вынеся некую константу за знак интеграла. Потом мы эту константу найдём по условию нормировки. Якобиан перехода в полярные координаты из n -мерного декартового пространства равен ρ^{n-1}

$$\int \dots \int_{x_1^2 + \dots + x_n^2 \leq t} (2 \cdot \pi)^{-\frac{n}{2}} \cdot e^{-\frac{u_1^2 + \dots + u_n^2}{2}} du_1 \dots du_n = c' \cdot \int_0^{\sqrt{t}} \rho^{n-1} \cdot e^{-\frac{\rho^2}{2}} d\rho$$

Введём новое обозначение $S = \frac{\rho^2}{2}$

$$\begin{aligned} c' \cdot \int_0^{\sqrt{t}} \rho^{n-1} \cdot e^{-\frac{\rho^2}{2}} d\rho &= \left| \begin{array}{l} S = \frac{\rho^2}{2} \\ \rho = \sqrt{2 \cdot S} \\ d\rho = \frac{dS}{\sqrt{2 \cdot S}} \end{array} \right| = \\ &= c' \cdot \int_0^{\frac{t}{2}} S^{\frac{n-1}{2}} \cdot e^{-S} \cdot S^{-\frac{1}{2}} dS = c' \cdot \int_0^{\frac{t}{2}} S^{\frac{n-1}{2}-1} \cdot e^{-S} dS \end{aligned}$$

Вспомним определение Γ -функции [12, с. 416], устремим t к бесконечности, и найдём константу c' . Эта константа не зависит от t , поэтому имеем

право

$$c' \cdot \int_0^{\frac{t}{2}} S^{\frac{n-1}{2}-1} \cdot e^{-S} dS \xrightarrow[t \rightarrow \infty]{a} c' \cdot \int_0^{+\infty} S^{\frac{n-1}{2}-1} \cdot e^{-S} dS =$$

$$= \left| \begin{array}{l} \Gamma(\alpha) = \int_0^{+\infty} e^{-x} \cdot x^{\alpha-1} dx \\ \alpha = \frac{n-1}{2} \\ x = S \end{array} \right| = c' \cdot \Gamma\left(\frac{n-1}{2}\right) \Rightarrow c' = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)}$$

Итого, у нас есть функция распределения

$$F_{\chi_n^2}(t) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \cdot \int_0^{\frac{t}{2}} \rho^{\frac{n}{2}-1} \cdot e^{-S} dS$$

Чтобы найти плотность распределения, нужно взять производную от интеграла с переменным верхним пределом [8, с. 353]. Делаем, но не забываем, что результирующая величина будет неотрицательной

$$p_{\chi_n^2}(t) = \frac{t^{\frac{n}{2}-1} \cdot e^{-\frac{t}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \mathbb{1}(t > 0)$$

С двумя степенями свободы распределение Пирсона превращается в экспоненциальное

$$\chi_2^2 = \text{Exp}\left(\frac{1}{2}\right)$$

Математическое ожидание и дисперсию посчитать несложно

$$\begin{aligned} M \chi_n^2 &= M [\xi_1^2 + \dots + \xi_n^2] = n \\ D \chi_n^2 &= 2 \cdot n \end{aligned}$$

Лемма 3.4.2: Фишера

Есть стандартный гауссовский вектор $\vec{\xi} = (\xi_1, \dots, \xi_n)$, есть ортогональная матрица C ($C^T = C^{-1}$) и случайный вектор $\vec{\eta} = (\eta_1, \dots, \eta_n)$, полученная следующим образом

$$\vec{\eta} = C \vec{\xi}$$

Тогда, если отнять от суммы квадратов элементов вектора $\vec{\xi}$ сумму квадратов элемента вектора $\vec{\eta}$, получится случайная величина с рас-

пределением Пирсона с $n - r$ степенями свободы

$$\sum_{k=1}^n \xi_k^2 - \sum_{k=1}^r \eta_k^2 \sim \chi_{n-r}^2$$

Доказательство. Ортогональное преобразование сохраняет расстояния (вращение, поворот), а расстояние в Евклидовом пространстве — сумма квадратов координат. Поскольку случайный вектор $\vec{\eta}$ получен путём ортогонального преобразования случайного вектора $\vec{\xi}$, то их нормы совпадают

$$\sum_{k=1}^n \xi_k^2 = \sum_{k=1}^n \eta_k^2$$

Несложными математическими преобразованиями получаем следующий результат

$$\sum_{k=1}^n \xi_k^2 - \sum_{k=1}^r \eta_k^2 = \sum_{k=1}^n \eta_k^2 - \sum_{k=1}^r \eta_k^2 = \sum_{k=r+1}^n \eta_k^2$$

Видим сумму $n - r$ квадратов независимых гауссовских случайных величин, а это значит, что имеет место распределение χ^2 с $n - r$ степенями свободы

$$\sum_{k=r+1}^n \eta_k^2 \sim \chi_{n-r}^2$$

□

3.4.2 Распределение Фишера

Определение 3.4.3: Распределение Фишера

Отношение независимых случайных $\chi_{k_1}^2$ и $\chi_{k_2}^2$ называется распределением Фишера

$$F_{k_1, k_2} = \frac{\chi_{k_1}^2}{\chi_{k_2}^2}$$

3.4.3 Распределение Стюдента

Определение 3.4.4: Распределение Стюдента

Есть $n + 1$ независимых стандартных гауссовских случайных величин ξ_0, \dots, ξ_n . Отношение первой (нулевой) случайной величины к корню суммы квадратов остальных, делённой на n , имеет распределение

Стьюдента с n степенями свободы

$$\frac{\xi_0}{\sqrt{\frac{1}{n} \cdot \sum_{k=1}^n \xi_k^2}} \sim t_n$$

Отметим, что квадрат этой случайной величины, делённой на n , имеет распределение Фишера с параметрами 1 и n

$$\frac{t_n^2}{n} = \frac{\xi_0^2}{\sum_{k=1}^n \xi_k^2} \sim F_{1,n}$$

Пример 3.4.5

Есть выборка x_1, \dots, x_n из нормального распределения $N(a, \sigma^2)$.

Выборочное среднее \bar{x} и выборочная дисперсия $\frac{1}{n-1} \cdot \sum_{k=1}^n (x_k - \bar{x})^2$ не зависят друг от друга, а это значит, что случайные величины \bar{x} и $\sum_{k=1}^n (x_k - \bar{x})^2$ тоже независимы^a.

Распределения выглядят следующим образом

$$\frac{\sqrt{n} \cdot (\bar{x} - a)}{\sigma} \sim N(0, 1)$$

$$\frac{1}{\sigma^2} \cdot \sum_{k=1}^n (x_k - \bar{x})^2 \sim \chi_{n-1}^2$$

В таком случае получаем распределение Стьюдента следующим образом. Обратите внимание, что распределение не зависит от дисперсии σ^2 , потому что она сокращается

$$\frac{\sqrt{n} \cdot (\bar{x} - a)}{\sqrt{\frac{1}{n-1} \cdot \sum_{k=1}^n (x_k - \bar{x})^2}}$$

^aпервым доказал Фишер [1], сам же Стьюдент показал лишь некоррелированность этих случайных величин [2]

Глава 4

Методы оценивания неизвестных параметров

4.1 Вступление

Мы занимаемся гауссовскими векторами, поскольку они геометричны. Об этом говорят такие важные свойства

1. Некоррелированность гауссовских векторов эквивалентна их независимости (теорема 3.3.16)
2. После поворота гауссовский вектор остаётся гауссовским (лемма 3.3.8)
3. Чтобы вычислить совместное распределение гауссовских векторов, достаточно матриц (теорема 3.3.19)

Рассмотрим пример, который наведёт нас на необходимые размышления.

Пример 4.1.1

Есть выборка x_1, \dots, x_n из гауссовского распределения с неизвестным средним $x_i \sim N(\theta, 1)$.

Как проверить, что оценка θ_* оптимальная?

Поскольку гауссовское распределение является экспоненциальным (определение 1.2.14), то для него существует эффективная оценка (утверждение 1.2.15).

Выпишем функцию правдоподобия (определение 1.2.3)

$$L(\vec{x}, \theta) = \frac{1}{\sqrt{2 \cdot \pi}^n} \cdot e^{-\frac{1}{2} \cdot \sum_{k=1}^n (x_k - \theta)^2}$$

Эффективная оценка считается по формуле (определение 1.2.17)

$$\theta_* = \arg \max_{\theta} \ln L(\vec{x}, \theta)$$

Поскольку θ находится лишь в экспоненте, то нужно максимизиро-

вать экспоненту

$$\theta_* = \arg \max_{\theta} \left(e^{-\frac{1}{2} \cdot \sum_{k=1}^n (x_i - \theta)^2} \right)$$

Сумма неотрицательна, а это значит, что для того, чтобы экспонента приняла максимальное значение, нужно минимизировать эту сумму

$$\theta_* = \arg \min_{\theta} \sum_{k=1}^n (x_i - \theta)^2$$

Вспомнив геометрию с физикой,^a понимаем, что в качестве оценки θ_* нужно взять выборочное среднее, так как момент инерции минимален у центра масс

$$\theta_* = \frac{1}{n} \cdot \sum_{k=1}^n x_k$$

^a Есть вектор \vec{y} , есть вектора $\vec{a}_1, \dots, \vec{a}_n$.

В точке \vec{a} находится центр масс тела, состоящего из точек \vec{a}_k (массу каждой точки считаем равной единице)

$$\vec{a} = \frac{1}{n} \cdot \sum_{k=1}^n \vec{a}_k$$

Момент инерции относительно точки \vec{y} — сумма квадратов расстояний, умноженных на массы точек

$$\sum_{k=1}^n \|\vec{y} - \vec{a}_k\|^2 = n \cdot \|\vec{y} - \vec{a}\|^2 + \sum_{k=1}^n \|\vec{a} - \vec{a}_k\|^2$$

Момент инерции минимальный у центра масс.

Пример 4.1.2

Обобщим задачу 4.1.1. Мы ведь изучали гауссовские вектора для того, чтобы иметь дело с многомерным случаем.

Возьмём m -мерную оценку $\vec{\theta}$

$$\vec{\theta} \in \mathbb{R}^m, \quad m \leq n$$

Матрица A

$$A : \mathbb{R}^m \rightarrow \mathbb{R}^n$$

Вместо выборки x_1, \dots, x_n будем рассматривать вектор \vec{y}

$$\vec{y} = A\vec{\theta} + \vec{\xi}, \quad \vec{\xi} \sim N(\vec{0}, I)$$

Эта ситуация является модификацией исходной задачи. Из леммы 3.3.8 Помним, что гауссовский вектор в сумме с произвольным — сме-

щённый гауссовский вектор

$$(x_1, \dots, x_n) = \overbrace{(\theta, \dots, \theta)}^n + \vec{\xi}, \quad m = 1 \Rightarrow A\theta = \overbrace{(\theta, \dots, \theta)}^n$$

Мы чуть-чуть усложнили задачу и теперь нужно найти оптимальную оценку для векторной величины

$$\vec{\theta}_* = ?$$

Поскольку имеем дело с экспоненциальным распределением, то дальше делаем то же, что и в предыдущем примере. Для начала выпишем функцию правдоподобия

$$L(\vec{\xi}, \vec{\theta}) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot \|\vec{x} - A\vec{\theta}\|^2}$$

Чтобы максимизировать функцию правдоподобия, нужно минимизировать ту норму, что в экспоненте

$$\vec{\theta}_* = \arg \min_{\vec{\theta} \in \mathbb{R}^m} \|\vec{\xi} - A\vec{\theta}\|^2$$

Увидели, что поиск оптимальной оценки среднего для гауссовского вектора теперь сводится не к дифференциальным уравнениям, а к поиску минимума.

4.2 Оценка метода наименьших квадратов

4.2.1 Размышления

Есть случайные величины x и y , между которыми нужно определить зависимость.

Гипотеза: эта зависимость может быть линейной

$$y = \sum_{k=1}^m f_k(x) \cdot \theta_k$$

В формуле f_k — известная функция, θ_k — неизвестный коэффициент.

Пример 4.2.1

$$\begin{cases} f_1(x) = 1 \\ f_2(x) = x \end{cases} \Rightarrow f(x) = \theta_1 + \theta_2 \cdot x$$

Пример 4.2.2

Пусть функции будут степенями x

$$\begin{cases} f_1(x) = 1 \\ f_2(x) = x \\ \vdots \\ f_m(x) = x^{m-1} \end{cases} \Rightarrow y = \sum_{k=1}^m \theta_k \cdot x^{k-1}$$

Получили полиномиальную зависимость.

Когда есть наблюдения y_1, \dots, y_n , то возникают и помехи $\xi_1, \dots, \xi_n \sim N(0, 1)^*$

$$\begin{aligned} y_1 &= \sum_{k=1}^m f_k(x_1) \cdot \theta_k + \xi_1 \\ &\vdots \\ y_n &= \sum_{k=1}^m f_k(x_n) \cdot \theta_k + \xi_n \end{aligned}$$

Какая же оценка будет оптимальной для $\vec{\theta} = (\theta_1, \dots, \theta_m)$?

Рассмотрим в контексте задачи [4.1.2](#).

Пример 4.2.3

Помехи составляют стандартный гауссовский вектор как совокупность независимых стандартных гауссовских величин

$$\vec{\xi} \sim N(\vec{0}, I)$$

Также у нас есть параметр $\vec{\theta} \in \mathbb{R}^m$.

Матрица A принимает следующий вид

$$A = \|f_j(x_i)\|_{\substack{i=\overline{1,n} \\ j=\overline{1,m}}}$$

Тогда вектор \vec{y} примет понятный из контекста примера [4.1.2](#) вид

$$\vec{y} = A\vec{\theta} + \vec{\xi}$$

Тут мы уже знаем, как искать минимум

$$\arg \min_{\vec{\theta} \in \mathbb{R}^m} \|\vec{y} - A\vec{\theta}\|^2 = \arg \min_{\vec{\theta} \in \mathbb{R}^m} \sum_{k=1}^n \left(y_k - (A\vec{\theta})_k \right)^2$$

*Считаем, что смещение помех нулевое, так как иначе это будет означать неправильную постановку эксперимента. Также считаем, что эти случайные величины независимы между собой по той же причине.

У нас идеальные условия, и экспериментаторы — люди толковые.

Ковариация выбрана единичной для удобства. Потом мы это дело обобщим.

Если для гауссовских случайных величин этот метод работает хорошо, то обобщим его.

4.2.2 Общая постановка задачи метода наименьших квадратов

Вектор $\vec{\xi}$ — случайный вектор в \mathbb{R}^n с нулевым средним и единичной ковариацией.

Оператор A действует из \mathbb{R}^m в \mathbb{R}^n . Чтобы он не уничтожал параметры, скажем, что его ранг равен m

$$\text{rank } A = m$$

Нужно найти оптимальную линейную оценку для параметра $\vec{\theta}$ по наблюдениям \vec{y}

$$\vec{\theta}_* = B \vec{y}, \quad \vec{y} = A \vec{\theta} + \vec{\xi}$$

Теорема 4.2.4: Гаусса-Маркова

Пусть матрица $A \in \mathbb{R}^{n \times m}$ имеет ранг m , а случайный вектор $\vec{\xi}$ находится в \mathbb{R}^n и обладает нулевым средним и единичной ковариацией.

Тогда оценка $\vec{\theta}_*$ является единственным решением задачи минимизации

$$\vec{\theta}_* = \arg \min_{\vec{\theta} \in \mathbb{R}^m} \left\| \vec{y} - A \vec{\theta} \right\|^2$$

То есть ковариация оценки $\vec{\theta}_*$ не больше, чем ковариация любой несмещённой оценки $\vec{\tau}$ параметра $\vec{\theta}$ (разность ковариационных матриц — неотрицательно определённая матрица)

$$\text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} \leq \text{Cov}_{\vec{\tau}, \vec{\tau}} \Rightarrow \text{Cov}_{\vec{\tau}, \vec{\tau}} - \text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} \geq 0$$

Доказательство. Разобьём доказательство на три части

1. Нужно найти минимум — значит, возьмём производную по m аргументам. Это оператор набла по параметру $\vec{\theta}$

$$\begin{aligned} \nabla_{\vec{\theta}} \cdot \left\| \vec{y} - A \vec{\theta} \right\|^2 &= \nabla_{\vec{\theta}} \cdot (\vec{y} - A \vec{\theta}, \vec{y} - A \vec{\theta}) = \\ &= \nabla_{\vec{\theta}} \cdot (\vec{y}, \vec{y}) - 2 \cdot \nabla_{\vec{\theta}} \cdot (A \vec{\theta}, \vec{y}) + \nabla_{\vec{\theta}} \cdot (A \vec{\theta}, A \vec{\theta}) = \\ &= \vec{0} - 2 \cdot A^* \vec{y} + 2 \cdot A^* A \vec{\theta} + \vec{0} \end{aligned}$$

Приравниваем к нулю

$$A^* A \vec{\theta} = A^* \vec{y}$$

И тут важный момент

$$\text{rank } A = \text{rank } A^* = m$$

С помощью оператора A мы переходим из пространства \mathbb{R}^m в пространство \mathbb{R}^n , а с помощью оператора A^* мы переходим обратно.

Выполняется следующее тождество

$$A^* A \mathbb{R}^m = \mathbb{R}^m$$

Значит, матрица $A^* A$ невырожденная

$$\det A^* A > 0$$

Значит, можно решить уравнения единственным образом

$$\vec{\theta}_* = \left(A^* A \right)^{-1} A^* \vec{y} \quad (4.1)$$

Это и есть явный вид оценки метода наименьших квадратов.

2. Возьмём случайный вектор $\vec{\eta}$ со средним \vec{m} и ковариацией B .

Возьмём матрицу C . Тогда, согласно утверждению 3.2.2, среднее и ковариационная матрица вектора $C \vec{\eta}$ примут следующий вид

$$\begin{aligned} M C \vec{\eta} &= C \vec{m} \\ \text{Cov}_{C \vec{\eta}, C \vec{\eta}} &= C B C^* \end{aligned}$$

Вектор \vec{y} — случайный вектор со средним $A \vec{\theta}$ и единичной ковариацией.

Оценка $\vec{\theta}_*$ действительно несмещённая. Чтобы это показать, воспользуемся определением 1.1.17 и формулой 4.1

$$M \vec{\theta}_* = \left(A^* A \right)^{-1} A^* M \vec{y} = \left(A^* A \right)^{-1} A^* A \vec{\theta} = \vec{\theta}$$

Посмотрим, какая у оценки ковариационная матрица $\vec{\theta}_*$

$$\text{Cov}_{\vec{\theta}_*, \vec{\theta}_*} = \left(A^* A \right)^{-1} A^* I \left\{ \left(A^* A \right)^{-1} A^* \right\}^*$$

Вспомним, что сопряжение меняет местами множители

$$\begin{aligned} \left(A^* A \right)^{-1} A^* I \left\{ \left(A^* A \right)^{-1} A^* \right\}^* &= \\ = \left(A^* A \right)^{-1} A^* A \left\{ \left(A^* A \right)^{-1} \right\}^* &= \left(A^* A \right)^{-1} \end{aligned}$$

3. Почему же оценка $\vec{\theta}_*$ будет оптимальной в классе линейных несмещённых оценок?

Обозначим пространство линейных несмещённых оценок параметра $\vec{\theta}$ с вектором шума $\vec{\xi}$ как $\mathcal{L}_{\vec{\xi}}$.

Пространство всех несмещённых оценок параметра $\vec{\theta}$ с помехами $\vec{\xi}$ будет $\mathcal{M}_{\vec{\xi}}$.

Очевидно, что пространство линейных несмещённых оценок параметра $\vec{\theta}$ является подпространством всех несмещённых оценок параметра $\vec{\theta}$

$$\mathcal{L}_{\vec{\xi}} \subset \mathcal{M}_{\vec{\xi}}$$

Введём ещё одну группу обозначений.

Пусть $\vec{\eta}$ — случайный вектор в \mathbb{R}^n с нулевым средним и единичной ковариацией, как у вектора $\vec{\xi}$

$$\begin{aligned}\vec{\eta} &\sim N(\vec{0}, \mathbf{I}) \\ \mathbf{M} \vec{\eta} &= \mathbf{M} \vec{\xi} = \vec{0} \\ \text{Cov}_{\vec{\eta}, \vec{\eta}} &= \text{Cov}_{\vec{\xi}, \vec{\xi}} = \mathbf{I}\end{aligned}$$

Вектору $\vec{\eta}$ поставим в соответствие пространства $\mathcal{L}_{\vec{\eta}}$ и $\mathcal{M}_{\vec{\eta}}$.

Что такое линейная несмещённая оценка параметра $\vec{\theta}$? Это матрицы, применённые к выборке

$$\mathcal{L}_{\vec{\xi}} \ni \mathbf{T} \vec{y}: \mathbf{M} \mathbf{T} \vec{y} = \mathbf{T} \mathbf{A} \vec{\theta} = \vec{\theta}$$

Возьмём последнее равенство из формулы и видим, что тут ничего не сказано о помехах — ни о $\vec{\eta}$, ни о $\vec{\xi}$

$$\mathbf{T} \mathbf{A} \vec{\theta} = \vec{\theta}$$

Тогда делаем вывод, что пространство линейных несмещённых оценок едино для любых шумов

$$\mathcal{L}_{\vec{\eta}} = \mathcal{L}_{\vec{\xi}} \ni \mathbf{T} \vec{y}$$

Подсчитаем ковариацию вектора $\mathbf{T} \vec{y}$ и видим, что она от шумов тоже не зависит

$$\text{Cov}_{\mathbf{T} \vec{y}, \mathbf{T} \vec{y}} = \mathbf{T} \mathbf{T}^*$$

Следовательно, искать оптимальную линейную несмещённую оценку в $\mathcal{L}_{\vec{\xi}}$ — то же самое, что искать оптимальную линейную несмещённую оценку в $\mathcal{L}_{\vec{\eta}}$.

Значит, мы будем использовать пространство $\mathcal{L}_{\vec{\eta}}$, поскольку вектор $\vec{\eta}$ гауссовский, а для него мы уже умеем искать оптимальную оценку.

Оценка в формуле (4.1) является оптимальной в пространстве несмещённых оценок $\mathcal{M}_{\vec{\xi}}$. Ещё она линейна, а это значит, что входит в пространство $\mathcal{L}_{\vec{\xi}}$. То есть она лучшая в пространстве линейных несмещённых оценок.

□

4.2.3 Оценка метода наименьших квадратов при неединичной ковариации

Есть вектор помех $\vec{\xi}$ с нулевым математическим ожиданием и положительной ковариацией. Между “случайными” помехами есть связь

$$\begin{aligned} M \vec{\xi} &= \vec{0} \\ \text{Cov}_{\vec{\xi}, \vec{\xi}} &= V > 0 \end{aligned}$$

Умножим вектор $\vec{\xi}$ на матрицу B , квадрат которой равен обратной к V матрице, и назовём полученный вектор $\vec{\eta}$

$$\begin{aligned} B &= V^{-\frac{1}{2}} \\ \vec{\xi} &= \vec{\eta} \\ \text{Cov}_{\vec{\eta}, \vec{\eta}} &= B \text{Cov}_{\vec{\xi}, \vec{\xi}} B^* = V^{-\frac{1}{2}} V V^{-\frac{1}{2}} = I \end{aligned}$$

Тогда, подействовав оператором B на вектор измерений \vec{y} , получим следующее

$$B \vec{y} = B A \vec{\theta} + B \vec{\xi} = V^{-\frac{1}{2}} A \vec{\theta} + \vec{\eta}$$

Матрица $V^{-\frac{1}{2}}$ невырожденная и квадратная, а это значит, что при умножении её на матрицу A получим результат с рангом m

$$\text{rank } V^{-\frac{1}{2}} A = \text{rank } A = m$$

Тогда оценка метода минимальных квадратов будет считаться как минимум такого выражения

$$\vec{\theta}_* = \arg \min_{\vec{\theta}} \left\| V^{-\frac{1}{2}} A V^{-\frac{1}{2}} \vec{y} \right\|^2$$

Подставим наши данные в исходную формулу (4.1)

$$\vec{\theta}_* = \left(A^* V^{-\frac{1}{2}} V^{-\frac{1}{2}} A \right)^{-1} V^{-\frac{1}{2}} A \vec{y}$$

Получили конечную формулу метода оценки наименьших квадратов при неединичной ковариации

$$\vec{\theta}_* = \left(A^* V^{-1} A \right)^{-1} V^{-\frac{1}{2}} A \vec{y} \quad (4.2)$$

4.3 Интервальное оценивание

4.3.1 Определения

Есть выборка x_1, \dots, x_n из распределения $F_\theta(x)$, неизвестный параметр θ — число действительное $\theta \in \Theta \subset \mathbb{R}$.

Есть две статистики (функции от выборки, определение 1.1.10) $T_1 \leq T_2$ и константа $\alpha \in (0; 1)$

Определение 4.3.1: Доверительный интервал

Промежуток $[T_1, T_2]$ ($T_1 \leq T_2$ — статистики) называется доверительным интервалом для действительного параметра $\theta \in \mathbb{R}$ с уровнем доверия α , если

$$\forall \theta \in \Theta : \mathbb{P}_\theta \{ \theta \in [T_1; T_2] \} \geq \alpha$$

Определение 4.3.2: Центральная статистика

Функция $G(\vec{x}, \theta)$, зависящая от выборки и неизвестного параметра θ , называется центральной статистикой, если выполняются два свойства

1. $G(\vec{x}, \theta)$ имеет известное распределение
2. $G(\vec{x}, \cdot)$ строго монотонная и непрерывная функция ($G(\vec{\xi}, \theta)$ строго монотонна и непрерывна по θ)

4.3.2 Рецепт построения доверительного интервала с помощью центральной статистики

Нам дано $\alpha \in [0; 1]$, центральная статистика имеет распределение с функцией распределения F .

Выберем числа $a, b \in \mathbb{R}$ такие, что $a < b$. Тогда разность функций распределения в этих точках будет вероятностью попадания достаточной статистики в этот интервал

$$F(b) - F(a) \geq \alpha \Rightarrow \mathbb{P} \left\{ G(\vec{\xi}, \theta) \in [a; b] \right\} \geq \alpha$$

Поскольку центральная статистика G монотонна относительно параметра θ , то событие, заключающееся в том, что она попала в промежуток $[a; b]$, эквивалентно тому событию, что параметр θ попал в отрезок, ограниченный некими функциями от a и b

$$a \leq G(\vec{x}, \theta) \leq b \Leftrightarrow T_1 \leq \theta \leq T_2$$

4.3.3 Асимптотические доверительные интервалы

Пусть x_1, \dots, x_n — выборка из гладкого распределения с дважды интегрируемой плотностью (смотрите замечание 1.2.2) $p(\vec{x}, \theta)$.

Вспомним, чему равен вклад выборки (определение 1.2.4 и замечание 1.2.5)

$$U(\vec{x}, \theta) = \sum_{k=1}^n \frac{\partial}{\partial \theta} \ln p(x_k, \theta)$$

Математическое ожидание каждого слагаемого равно нулю (посмотрите замечание 1.2.6)

$$M_{\theta} \frac{\partial}{\partial \theta} \ln p(x_k, \theta) = 0$$

Также вспомним, чему равно количество информации Фишера (определение 1.2.7) по одному элементу выборки

$$D_{\theta} \frac{\partial}{\partial \theta} \ln p(x_k, \theta) = I_1(\theta)$$

Выпишем следующее соотношение

$$\frac{U(\vec{x}, \theta)}{\sqrt{n \cdot I_1(\theta)}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\frac{\partial}{\partial \theta} \ln p(x_k, \theta)}{\sqrt{I_1(\theta)}}$$

Видим сумму независимых случайных величин с нулевым математическим ожиданием и единичной дисперсией, делённую на \sqrt{n} , как в центральной предельной теореме. Это значит, что при больших n такое отношение является центральной статистикой со стандартным нормальным распределением.

$$\frac{U(\vec{x}, \theta)}{\sqrt{n \cdot I_1(\theta)}} \approx N(0, 1)$$

4.4 Критерий согласия

Критерий согласия даёт ответ на вопрос, является ли данная выборка x_1, \dots, x_n выборкой из распределения F .

4.4.1 Критерий Колмогорова-Смирнова

Есть выборка x_1, \dots, x_n . Проверим согласие с распределением F .
Функция F — непрерывная, строго возрастающая.

Лемма 4.4.1

Пусть ξ — случайная величина с непрерывной и строго монотонной функцией распределения F .

Тогда $F(\xi)$ имеет равномерное распределение на $[0; 1]$

Доказательство.

$$F_{F(\xi)}(a) = \mathbb{P}\{F(\xi) \leq a\} = \mathbb{P}\{\xi \leq F^{-1}(a)\} = F(F^{-1}(a)) = a \in (0; 1)$$

□

Если x_1, \dots, x_n — выборка из F , то $F(x_1), \dots, F(x_n)$ — выборка из $U([0; 1])$.

$$\frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}\{y \leq F(x_k)\} \xrightarrow{n \rightarrow \infty} y$$

Если это выполняется для любого y из $[0; 1]$, то справедливо следующее тождество

$$\frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}\{\cdot \leq F(x_k)\} \xrightarrow[n \rightarrow \infty]{} \cdot$$

Отметим, что центральная предельная теорема работает и в функциональном пространстве и следующая случайная величина сходится к стандартному гауссовскому распределению в пространстве непрерывных функций на $[0; 1]$

$$\frac{1}{\sqrt{n}} \cdot \left(\sum_{k=1}^n \mathbb{1}\{\cdot \leq F(x_k)\} - \cdot \right) \xrightarrow[n \rightarrow \infty]{} N(0, 1)$$

Распределение супремума такой случайной величины имеет распределение Смирнова

$$\frac{1}{\sqrt{n}} \cdot \sup_{y \in [0; 1]} \left| \sum_{k=1}^n (\mathbb{1}\{y \leq F(x_k)\} - y) \right|$$

Для распределения Смирнова нет аналитической функции, но есть таблицы.

Поскольку следующее тождество очевидно

$$F(x_{(k)}) = F(x)_{(k)}$$

Вместо того, чтобы брать супремум по всем точкам, достаточно перебирать лишь точки, в которых происходит скачок

$$\frac{1}{\sqrt{n}} \cdot \left(\sum_{k=1}^n \mathbb{1}\{\cdot \leq F(x_k)\} - \cdot \right) = \max \frac{1}{\sqrt{n}} \cdot \left| \frac{k}{n} - F(x_{(k)}) \right| \sim Q$$

Буква Q означает распределение Колмогорова-Смирнова (в литературе встречается под названием распределение Смирнова).

Обозначим случайную величину, равную максимуму, через R

$$R = \max_{k=1, n} \frac{1}{\sqrt{n}} \cdot \left| \frac{k}{n} - F(x_{(k)}) \right| \sim Q$$

4.4.2 Рецепт применения критерия Колмогорова-Смирнова

Систематизируем полученные знания для фиксированного α .

1. Вычисляем по табличкам значение r_α

$$r_\alpha : Q(\xi \leq r_\alpha) = \alpha$$

2. Вычисляем R

$$R = \max_{k=1, n} \frac{1}{\sqrt{n}} \cdot \left| \frac{k}{n} - F(x_{(k)}) \right|$$

3. Если R не больше, чем r_α , то ответ — “да”. Если же R больше, чем r_α , то ответ — “нет”

Но как же быть с дискретными распределениями? Иногда видоизменяют центральную предельную теорему, но мы же пойдём иным путём.

4.4.3 Критерий Пирсона χ^2

Есть выборка x_1, \dots, x_n . Имеет ли она распределение F ?

Известно, что F принадлежит какому-то интервалу $[a; b]$. Разбиваем интервал возможных значений выборки на m полуинтервалов так же, как для построения гистограммы в подразделе 1.1.2.

Введём несколько обозначений

1. Δ_i — полуинтервал номер i
2. ν_i — количество элементов выборки, которые попали в Δ_i

$$\nu_i = \sum_{k=1}^n \mathbb{1}\{x_k \in \Delta_i\}$$

3. p_i — вероятность случайных точек попасть в Δ_i

$$p_i = \mathbb{P}_F \{x_1 \in \Delta_i\}$$

Идея следующая: после применения центральной предельной теоремы (с некоторой правкой) получим распределение Пирсона

$$\sum_{i=1}^m \frac{(\nu_i - n \cdot p_i)^2}{n \cdot p_i} \sim \chi_{m-1}^2$$

Почему так происходит?

Предположим, мы угадываем F . В таком случае $n \cdot p_i$ — математическое ожидание. Смещаем ν_i для применения центральной предельной теоремы. Без квадрата нужно было бы делить на \sqrt{n} , с квадратом же — на n .

Чем больше линейных связей, тем меньше степеней свободы, поэтому их у нас $m - 1$.

Теорема 4.4.2

Если x_1, \dots, x_n — выборка из F , ν_i и p_i определены следующим образом

$$\nu_i = \sum_{k=1}^n \mathbb{1}\{x_k \in \Delta_i\}$$

$$p_i = \mathbb{P}_F \{x_1 \in \Delta_i\}$$

Тогда

$$\sum_{i=1}^m \frac{(\nu_i - n \cdot p_i)^2}{n \cdot p_i} \xrightarrow[n \rightarrow \infty]{} \chi_{m-1}^2$$

Доказательство. Обозначим вектор $\vec{\eta}_n$ следующим образом

$$\vec{\eta}_n = \left(\frac{\nu_1 - n \cdot p_1}{\sqrt{n \cdot p_1}}, \dots, \frac{\nu_m - n \cdot p_m}{\sqrt{n \cdot p_m}} \right)$$

Координаты вектора $\vec{\eta}_n$ зависимы между собой, поэтому центральная предельная теорема тут не поможет.

Вспомним, что слабая сходимость — то же самое, что поточечная сходимость характеристических функций. Вычислим характеристическую функцию вектора $\vec{\eta}_n$.

Берём произвольный вектор $\lambda \in \mathbb{R}^m$. Нам нужно найти характеристическую функцию

$$\varphi_{\vec{\eta}_n} = \mathbb{M} e^{i \cdot \vec{\lambda} \cdot \vec{\eta}_n}$$

Для начала разберём скалярное произведение в терминах исходной выборки

$$(\vec{\lambda}, \vec{\eta}_n) = \sum_{k=1}^m \lambda_k \cdot \frac{\nu_k - n \cdot p_k}{\sqrt{n \cdot p_k}} = \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot \sum_{j=1}^n (\mathbb{1}\{x_j \in \Delta_k\} - p_k)$$

Поменяем суммы местами и получим сумму n независимых одинаково распределённых случайных величин

$$\sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot \sum_{j=1}^n (\mathbb{1}\{x_j \in \Delta_k\} - p_k) = \sum_{j=1}^n \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot (\mathbb{1}\{x_j \in \Delta_k\} - p_k)$$

Помним, что характеристическая функция суммы независимых случайных величин — произведение характеристических функций, а когда это сумма одной и той же случайной величины, то получается характеристическая функция в степени

$$\varphi_{(\sum_{i=1}^n \xi_i)} = \prod_{i=1}^n \varphi_{\xi_i} = \varphi_{\xi}^n$$

Получаем следующее

$$\mathbb{M} e^{i \cdot \vec{\lambda} \cdot \vec{\eta}_n} = \left\{ \mathbb{M} e^{i \cdot \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot (\mathbb{1}\{x_1 \in \Delta_k\} - p_k)} \right\}^n$$

Посчитаем, чему равняется математическое ожидание, которое нужно возвести в степень n

$$\mathbb{M} \exp \left\{ i \cdot \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot (\mathbb{1}\{x_1 \in \Delta_k\} - p_k) \right\} = ? \quad (4.3)$$

Для начала проанализируем индикатор в экспоненте. Разобьём сумму на две

$$\begin{aligned} \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot (\mathbb{1}\{x_1 \in \Delta_k\} - p_k) &= \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot \mathbb{1}\{x_1 \in \Delta_k\} - \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot p_k = \\ &= \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot \mathbb{1}\{x_1 \in \Delta_k\} - \sum_{k=1}^m \frac{\lambda_k \cdot \sqrt{p_k}}{\sqrt{n}} \end{aligned}$$

То есть в первой сумме остаётся лишь то слагаемое, которое соответствует промежутку, в который попал элемент выборки. Имеем право переписать всё это дело в следующем виде

$$\begin{aligned} & \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot \mathbb{1}\{x_1 \in \Delta_k\} - \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot p_k = \\ & = \sum_{k=1}^m \mathbb{1}\{x_1 \in \Delta_k\} \cdot \left(\frac{\lambda_k}{\sqrt{n \cdot p_k}} - \sum_{l=1}^m \frac{\lambda_l \cdot \sqrt{p_l}}{\sqrt{n}} \right) \end{aligned}$$

А это значит, что экспоненту (4.3) можем переписать в следующем виде

$$\begin{aligned} & \mathbb{M} \exp \left\{ i \cdot \sum_{k=1}^m \frac{\lambda_k}{\sqrt{n \cdot p_k}} \cdot (\mathbb{1}\{x_1 \in \Delta_k\} - p_k) \right\} = \\ & = \mathbb{M} \sum_{k=1}^m \mathbb{1}\{x_1 \in \Delta_k\} \cdot \exp \left\{ i \cdot \left(\frac{\lambda_k}{\sqrt{n \cdot p_k}} - \sum_{l=1}^m \frac{\lambda_l \cdot \sqrt{p_l}}{\sqrt{n}} \right) \right\} \end{aligned}$$

Воспользуемся линейностью математического ожидания, а так же тем фактом, что математическое ожидание индикатора $\mathbb{1}\{x_1 \in \Delta_k\}$ есть ни что иное, как вероятность p_k

$$\begin{aligned} & \mathbb{M} \sum_{k=1}^m \mathbb{1}\{x_1 \in \Delta_k\} \cdot \exp \left\{ i \cdot \left(\frac{\lambda_k}{\sqrt{n \cdot p_k}} - \sum_{l=1}^m \frac{\lambda_l \cdot \sqrt{p_l}}{\sqrt{n}} \right) \right\} = \\ & = \sum_{k=1}^m p_k \cdot \exp \left\{ i \cdot \frac{1}{\sqrt{n}} \cdot \left(\frac{\lambda_k}{\sqrt{p_k}} - \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} \right) \right\} \end{aligned}$$

Добавим и отнимем единицу, возьмём предел и видим замечательный предел!

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \sum_{k=1}^m p_k \cdot \left[\exp \left\{ i \cdot \frac{1}{\sqrt{n}} \cdot \left(\frac{\lambda_k}{\sqrt{p_k}} - \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} \right) \right\} - 1 \right] \cdot \frac{n}{n} \right)^n = \\ \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n = e^x \end{aligned}$$

Вытащим наше “ x ” из предела, но не забываем, что он содержит n , которое стремится к бесконечности

$$x = \lim_{n \rightarrow \infty} n \cdot \sum_{k=1}^m p_k \cdot \left[\exp \left\{ i \cdot \frac{1}{\sqrt{n}} \cdot \left(\frac{\lambda_k}{\sqrt{p_k}} - \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} \right) \right\} - 1 \right]$$

Тут видим ещё одну экспоненту. Она хороша тем, что в ней происходит деление на бесконечно большое значение \sqrt{n} , а это значит, что допускается следующее разложение

$$e^\alpha - 1 \approx \alpha + \frac{\alpha^2}{2}, \quad \alpha \ll 1$$

Рассмотрим сумму, которой принадлежит экспонента. Пока что подставим лишь α , а квадрат уже потом

$$\sum_{k=1}^m p_k \cdot i \cdot \left(\frac{\lambda_k}{\sqrt{p_k}} - \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} \right) = i \cdot \sum_{k=1}^m \lambda_k \cdot \sqrt{p_k} - i \cdot \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} = 0$$

Теперь рассмотрим квадрат

$$\lim_{n \rightarrow \infty} -\frac{n}{n} \sum_{k=1}^m p_k \cdot \left(\frac{\lambda_k}{\sqrt{p_k}} - \sum_{l=1}^m \lambda_l \cdot \sqrt{p_l} \right)^2 = -\sum_{k=1}^m \lambda_k^2 + \left(\sum_{k=1}^m \lambda_k \cdot \sqrt{p_k} \right)^2$$

Получаем следующий результат

$$\lim_{n \rightarrow \infty} M e^{i \cdot (\vec{\lambda}, \vec{\eta}_n)} = \exp \left\{ -\frac{1}{2} \cdot \left[\sum_{k=1}^n \lambda_k^2 - \left(\sum_{k=1}^m \lambda_k \cdot \sqrt{p_k} \right)^2 \right] \right\} = e^{-\frac{1}{2} \cdot (A \vec{\lambda}, \vec{\lambda})}$$

Матрица A имеет следующий вид

$$A = \left\| \delta_{ij} - \sqrt{p_i} \cdot \sqrt{p_j} \right\|_{i=1}^n$$

Матрица A должна быть симметричная и неотрицательно определённая. Симметричность очевидна, неотрицательную определённость нужно проверить. С этой целью введём вектор \vec{e}

$$\begin{aligned} \vec{e} &= (\sqrt{p_1}, \dots, \sqrt{p_m}) \\ \|\vec{e}\| &= 1 \end{aligned}$$

Очевидно следующее тождество

$$(A \vec{\lambda}, \vec{\lambda}) = \|\vec{\lambda}\|^2 - (\vec{\lambda}, \vec{e})^2$$

Далее используем неравенство Коши

$$|(\vec{\lambda}, \vec{e})| \leq \|\vec{\lambda}\| \cdot \|\vec{e}\| = \|\vec{\lambda}\|$$

Делаем вывод, что матрица A неотрицательно определённая

$$(A \vec{\lambda}, \vec{\lambda}) \geq 0$$

Это значит, что вектор $\vec{\eta}_n$ имеет гауссовское распределение с ковариацией A при достаточно большом значении n

$$\vec{\eta}_n \xrightarrow[n \rightarrow \infty]{} N(\vec{0}, A)$$

Но откуда взялось распределение χ_{m-1}^2 в теореме?

Введём стандартный гауссовский вектор $\vec{\kappa}$

$$\vec{\kappa} \sim N(\vec{0}, I)$$

Из линейной алгебры помним, что если у нас есть ортонормированный базис, то при воздействии на него ортогональным оператором U получим другой ортонормированный базис.

Возьмём $\vec{f}_1, \dots, \vec{f}_m$ — стандартный базис в \mathbb{R}^m

$$\vec{\kappa} = \sum_{k=1}^m \kappa_k \cdot \vec{f}_k, \quad \kappa_k \sim N(0, 1)$$

Введём новый базис $\vec{e} = \vec{e}_1, \vec{e}_2, \dots, \vec{e}_m$ (обратим внимание на то, что для первого элемента базиса мы дали два имени — с индексом и без него). Также есть U — ортогональная матрица перехода $\{\vec{f}\}$ к $\{\vec{e}\}$.

$$U \vec{\kappa} = \sum_{k=1}^m \hat{\kappa}_k \cdot \vec{e}_k \sim N(\vec{0}, U U^*) = N(\vec{0}, I)$$

Рассмотрим новый вектор $\vec{\xi}_*$, в который мы возьмём все элементы вектора $\{\hat{\kappa}_k\}$ кроме первого. Отметим, что элементы вектора $\vec{\xi}_*$ независимы между собой

$$\vec{\xi}_* = \sum_{k=2}^m \hat{\kappa}_k \cdot \vec{e}_k \sim N(\vec{0}, ?)$$

Посмотрим, чему равна квадратичная форма ковариационной матрицы вектора $\vec{\xi}_*$

$$\begin{aligned} M[(\vec{\xi}_*, \vec{\lambda}) \cdot (\vec{\xi}_*, \vec{\lambda})] &= \sum_{k=2}^m (\vec{\lambda}, \vec{e}_k)^2 = \sum_{k=1}^m (\vec{\lambda}, \vec{e}_k)^2 - (\vec{\lambda}, \vec{e})^2 = \\ &= \|\vec{\lambda}\|^2 - (\vec{\lambda}, \vec{e})^2 = (A \vec{\lambda}, \vec{\lambda}) \end{aligned}$$

То есть это и есть распределение вектора $\vec{\eta}_n$. Норма вектора $\vec{\xi}_*$ имеет распределение Пирсона с $m - 1$ степенями свободы, что и требовалось доказать

$$\|\vec{\xi}_*\| = \sum_{k=2}^m \hat{\kappa}_k^2 \sim \chi_{m-1}^2$$

То есть линейной связью убиваем одну степень свободы. \square

4.4.4 Рецепт применения критерия Пирсона

У нас есть $\alpha \in (0; 1)$, выборка x_1, \dots, x_n и предполагаемая функция распределения F .

1. Разбиваем область значений выборки на $\Delta_1, \dots, \Delta_m$
2. Вводим определение p_i

$$p_i = \mathbb{P}_F \{x_i \in \Delta_i\}$$

3. Считаем R

$$R = \sum_{k=1}^m \frac{(\nu_k - n \cdot p_k)^2}{n \cdot p_k}$$

4. Находим r_α по таблице

$$\mathbb{P} \{ \chi_{m-1}^2 \leq r_\alpha \} = \alpha$$

5. Если $R \leq r_\alpha$, то ответ “да”. Если же $R > r_\alpha$, то ответ “нет”.

Если мы не угадали распределение, то сумма будет вести себя как \sqrt{n} — будет очень большой.

Литература

- [1] Ronald Aylmer Fisher. Applications of “student’s, distribution. *Metron*, 5(3):90–104, 1925.
- [2] Student. The probable error of a mean. *Biometrika*, VI.
- [3] С. В. Фомин А. Н. Колмогоров. *Элементы теории функций и функционального анализа*. Наука, Москва, 1976.
- [4] Г.М. Цветкова и др. А.В. Печинкин, О.И. Тескин. *Теория вероятностей*. МГТУ им. Н. Э. Баумана, Москва, 2004.
- [5] А. А. Боровков. *Теория Вероятностей*. Эдиториал УРСС, Москва, 1999.
- [6] А. А. Боровков. *Математическая статистика*. Лань, Санкт-Петербург, 2010.
- [7] Э. Г. Позняк В. А. Ильин. *Линейная алгебра*. Наука. Физматлит, Москва, 1999.
- [8] Э. Г. Позняк В. А. Ильин. *Основы математического анализа. Часть I*. Физико-математическая литература, Москва, 2005.
- [9] Г.М. Цветкова и др. В.Б. Горяинов, И.В. Павлов. *Математическая статистика*. МГТУ им. Н. Э. Баумана, Москва, 2001.
- [10] В. В. Воеводин. *Линейная алгебра*. Наука. Главная редакция физико-математической литературы, Москва, 1980.
- [11] А. Я. Дороговцев. *Элементы общей теории меры и интеграла*. Выща школа. Головное издательство, Киев, 1989.
- [12] А. Я. Дороговцев. *Математический анализ. Краткий курс в современном изложении*. Факт, Киев, 2004.
- [13] В. Феллер. *Введение в теорию вероятностей и её приложения. Том 1*. Мир, Москва, 1984.
- [14] Г. М. Фихтенгольц. *Курс дифференциального и интегрального исчисления. Том 2*. Физматлит, Москва, 2003.
- [15] А. Н. Ширяев. *Вероятность-1*. МЦНМО, Москва, 2004.

Предметный указатель

В

вариационный ряд, 27
вектор
 конкатенация, 80
 проекция, 39
 случайный, 63
 стандартный гауссовский, 73
вклад выборки, 17
выборка, 5
выборочная дисперсия, 13
выборочное среднее, 13

Г

гамма-функция, 88
 Г-функция, 88
гауссовский вектор
 независимость, 81
 некоррелированность, 81
 плотность распределения, 76
 поворот, 74
 распределение, 74
 совместное распределение, 83
 стандартный, 73, 88
 условное математическое
 ожидание, 86
 характеристическая функ-
 ция, 74
гистограмма, 8

Д

доверительный интервал, 99
 асимптотический, 99
 построение, 99

К

ковариационная матрица, 68
 свойства, 67, 69
 случайного вектора, 65
ковариация, 65

количество информации Фишера,
 18

Колмогоров

 критерий согласия, 100

Колмогорова-Смирнова

 распределение, 101

 критерий согласия, 100

 Колмогорова-Смирнова, 100

 рецепт, 101

Пирсона

 рецепт, 106

Л

лемма

 критерий Колмогорова-
 Смирнова, 100

 о симметризации несмещён-
 ной оценки, 32

 распределение гауссовского
 вектора

 после поворота, 74

 распределения гауссовского
 вектора

 однозначность, 73

 условное математическое
 ожидание

 конечная сигма-алгебра, 41

 Фишера, 88

М

математическое ожидание

 случайного вектора, 64

 условное, 39

 в общем виде, 43

 гауссовского вектора, 86

 гладких функций, 50

 относительно конечной
 сигма-алгебры, 41

 свойства, 46

матрица
 сопряжённая, 66
 метод
 моментов, 11
 оценка, 12
 наименьших квадратов, 93
 неединичная ковариация, 98
 множество уровня, 35

Н

независимость
 выборочного среднего и выборочной дисперсии, 90
 неизвестный параметр, 8
 неравенство
 Рао-Крамера, 20

О

оценка, 9
 линейная несмещённая, 97
 максимального правдоподобия, 25
 теорема о состоятельности, 28
 метода наименьших квадратов, 93
 при неединичной ковариации, 98
 формула, 96, 98
 несмещённая, 10
 пространство, 97
 сильно состоятельная, 10
 состоятельная, 10
 улучшенная, 57
 эффективная, 21

П

Пирсон
 критерий согласия, 102
 распределение, 87
 плотность распределения
 случайного вектора, 63
 проекция
 вектора, 39
 пространство
 несмещённых оценок, 97
 линейных, 97

Р

распределение

гауссовское, 74
 Колмогорова-Смирнова, 101
 Пирсона, 87
 скалярного произведения, 70
 Смирнова, 101
 Стьюдента, 89
 Фишера, 89
 хи-квадрат (χ^2), 87
 экспоненциальное, 23
 эффективная оценка, 24

С

сигма-алгебра
 порождённая
 полным набором гипотез, 37
 случайной величиной, 34
 симметризация, 32
 случайная величина
 измеримая относительно
 сигма-алгебры, 37
 случайный вектор
 вращение, 71
 гауссовский
 независимость, 81
 некоррелированность, 81
 плотность распределения, 76
 распределение, 74
 стандартный, 73
 условное математическое
 ожидание, 86
 характеристическая функция, 74
 конкатенация, 80
 ковариационная матрица, 80
 математическое ожидание, 80
 линейные преобразования, 71
 математическое ожидание, 64
 плотность распределения, 63
 скалярное произведение, 70
 характеристическая функция, 64
 Смирнов
 критерий согласия, 100
 распределение, 101
 сопряжённая матрица, 66
 статистика, 9

достаточная, 56
центральная, 99
Стьюдента распределение, 89

Т

теорема

Гаусса-Маркова, 95
единственности, 14
Колмогорова, 14
критерий Пирсона, 102
о восстановлении
 неизвестной функции рас-
 пределения, 6
о нормальной корреляции, 83
о состоятельности оценки
 максимального правдо-
 подобия, 28
о существовании
 условного математического
 ожидания, 45
о характеристизации
 достаточной статистики, 59
об улучшении
 оценки с помощью доста-
 точной статистики, 57
оценка метода моментов, 12
условное математическое
 ожидание гладких функ-
 ций, 50
факторизационная, 59

У

уравнение
 правдоподобия, 26
условное

математическое ожидание, 39
 в общем виде, 43
 гауссовского вектора, 86
 гладких функций, 50
 относительно конечной
 сигма-алгебры, 41
 свойства, 46
распределение, 55

Ф

Фишера распределение, 89

функция

 правдоподобия, 17
 факторизация, 59

функция распределения

 выборочная, 5
 неизвестная, 5
 эмпирическая, 5

Х

характеристическая функция

 гауссовского вектора, 74
 случайного вектора, 64

хи-квадрат (χ^2) распределение, 87

Ц

центральная предельная теорема
 функциональное простран-
 ство, 101

центральная статистика, 99

Э

экспоненциальное

 распределение, 23
 эффективная оценка, 24

эффективная оценка, 24

Оглавление

Вступление	3
1 Основы	5
1.1 Методы оценок параметров распределения	5
1.1.1 Эмпирическая функция распределения	5
1.1.2 Гистограмма	7
1.1.3 Оценка неизвестных параметров	8
1.1.4 Выборочные оценки. Метод моментов	11
1.2 Свойства оценок	14
1.2.1 Неравенство Рао-Крамера	14
1.2.2 Метод максимального правдоподобия	21
2 Достаточные статистики	31
2.1 Оптимальная оценка	31
2.2 σ -алгебра, порождённая случайной величиной	34
2.3 Случайная величина, измеримая относительно σ -алгебры	36
2.4 Условное математическое ожидание	39
2.4.1 Проекция вектора	39
2.4.2 Проекция случайной величины	39
2.4.3 Условное математическое ожидание	43
2.4.4 Свойства условного математического ожидания	46
2.4.5 Условное математическое ожидание по функции от случайного вектора	47
2.4.6 Пример	51
2.5 Условные распределения	55
2.6 Достаточные статистики	56
3 Случайные вектора	63
3.1 Основные характеристики случайного вектора	63
3.1.1 Ковариационная матрица случайного вектора	65
3.1.2 Ковариационная матрица	68
3.2 Линейные преобразования случайных векторов	70
3.2.1 Скалярное произведение	70
3.3 Гауссовские случайные вектора	71
3.3.1 Определения	71
3.3.2 Линейные преобразования	74
3.3.3 Независимость компонент гауссовского вектора	80
3.3.4 Теорема о нормальной корреляции	83

3.4	Распределения, связанные с нормальным	87
3.4.1	Распределение Пирсона (хи-квадрат, χ^2)	87
3.4.2	Распределение Фишера	89
3.4.3	Распределение Стьюдента	89
4	Методы оценивания неизвестных параметров	91
4.1	Вступление	91
4.2	Оценка метода наименьших квадратов	93
4.2.1	Размышления	93
4.2.2	Общая постановка задачи метода наименьших квадратов	95
4.2.3	Оценка метода наименьших квадратов при неединичной ковариации	98
4.3	Интервальное оценивание	98
4.3.1	Определения	98
4.3.2	Рецепт построения доверительного интервала с помощью центральной статистики	99
4.3.3	Асимптотические доверительные интервалы	99
4.4	Критерий согласия	100
4.4.1	Критерий Колмогорова-Смирнова	100
4.4.2	Рецепт применения критерия Колмогорова-Смирнова	101
4.4.3	Критерий Пирсона χ^2	102
4.4.4	Рецепт применения критерия Пирсона	106
	Литература	107
	Предметный указатель	109
	Оглавление	113