

Математическая статистика

31 июля 2014 г.

Вступление

В курсе теории вероятностей у нас было 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 занимается выяснением распределения случайных величин.

x_1, \dots, x_n — выборка из распределения F . То есть, это независимые одинаково распределённые случайные величины с неизвестной функцией распределения F .

Предмет математической статистики — выяснение характера вероятностного эксперимента, отвечающего данной модели. 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 случайные величины x_1, \dots, x_n реализацией независимых одинаково распределённых случайных величин?
3. Есть две гипотезы относительно типа распределения. Какая из гипотез верна?
4. Существует ли связь между параметрами распределения x и y ? Каков вид этой зависимости?

Глава 1

ОСНОВЫ

1.1 Методы оценок характеристик наблюдаемых случайных величин

x_1, \dots, x_n — независимые одинаково распределённые случайные величины с неизвестной функцией распределения F .

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

1.1.1 Эмпирическая функция распределения

Определение 1.1.1: Эмпирическая функция распределения

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

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

Неизвестная функция распределения $F(x)$ может быть сколь угодно точно восстановлена по выборке достаточно большого объёма [7, стр. 25].

Теорема 1.1.2

$$F_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} F(x)$$

Идея доказательства. 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

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

Так как эмпирическая функция распределения является средним арифметическим индикаторов, то по усиленному закону больших чисел

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

Теорема доказана

$$F_n(x) \xrightarrow[n \rightarrow \infty]{a.s.} F(x)$$

□

Также можно доказать более сильную теорему

Теорема 1.1.3

$$\mathbb{P}\left(F_n \xrightarrow[n \rightarrow \infty]{} F\right) = 1$$

1.1.2 Гистограмма

Как можно попытаться отследить плотность распределения? Постараемся найти функцию распределения, а потом и плотность.

Допустим, F имеет хорошую (непрерывную) плотность. Как тогда из F получить p ?

Мы знаем, что $F' = p$, но это никому не нужно, так как F'_n — производная ступенчатой функции, которая почти везде будет равна нулю.

Но также мы помним, что

$$F(b) - F(a) = \int_a^b p(x) dx$$

Положим $a = x$ и введём $\Delta_x = b - x$

$$F(x + \Delta_x) - F(x) = \int_x^{x+\Delta_x} p(y) dy$$

Делим обе части на Δ_x .

$$\frac{1}{\Delta_x} \cdot \int_x^{x+\Delta_x} p(y) dy = \frac{F(x + \Delta_x) - F(x)}{\Delta_x}$$

Несложно заметить, что при достаточно малых значениях Δ_x получаем плотность распределения $p(x)$

$$\frac{\Delta F(x)}{\Delta_x} \xrightarrow{\Delta_x \rightarrow 0} \frac{dF(x)}{dx} = p(x)$$

Значит, можем заменить $p(x)$ не производной, а такой разностью.

$$p(x) \approx \frac{F(x + \Delta_x) - F(x)}{\Delta_x}$$

Возьмём m полуинтервалов на числовой прямой $I_j = (a_{j-1}, a_j], i = \overline{1, m}$ таких, каждое значение выборки попадает в свой интервал. Для этого определим пару свойств точек, ограничивающих эти интервалы:

1. Каждая следующая точка строго правее (больше) предыдущей (так как зачем нам одинаковые точки?)

$$a_0 < a_1 < \dots < a_m$$

2. Каждое значение выборки должно попадать ровно в один полуинтервал. Очевидно, что данные полуинтервалы I_j не пересекаются между собой. Значит, осталось потребовать, чтобы крайнее левое значение было меньше минимального значения из выборки, а крайнее правое — не меньше максимального

$$a_0 < \min(X) \leq \max(X) \leq a_m$$

Введём функцию $q(y)$

$$q(y) = \sum_{j=1}^m \frac{F(a_j) - F(a_{j-1})}{a_j - a_{j-1}} \cdot \mathbb{1}(y \in I_j)$$

Определим последовательность функций $q_n(y)$, заменив $F(x)$ на $F_n(x)$ в предыдущем определении

$$q_n(y) = \sum_{j=1}^m \frac{F_n(a_j) - F_n(a_{j-1})}{a_j - a_{j-1}} \cdot \mathbb{1}(y \in I_j) \quad (1.1)$$

Отметим, что q_n сходится к q почти наверное, так как сходятся почти наверное функции распределения согласно теореме 1.1.2

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

Функция q в свою очередь сходится к p , так как приблизительно равна производной, а при бесконечно большом количестве интервалов их длины становятся бесконечно малыми

$$q(y) \xrightarrow{m \rightarrow \infty} p(y)$$

Функция q_n называется гистограммой.

Избавимся от a_j в формуле, а для этого вспомним, чему равно $F_n(x)$

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

Теперь посмотрим, чему равна разность $F_n(a_j) - F_n(a_{j-1})$, которая, как мы видим, является вероятностью того, что x попало в отрезок I_j

$$\begin{aligned} & F_n(a_j) - F_n(a_{j-1}) = \\ &= \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \leq a_j) - \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \leq a_{j-1}) \end{aligned}$$

Сгруппируем слагаемые и получим чуть более компактную запись разности

$$\begin{aligned} & F_n(a_j) - F_n(a_{j-1}) = \\ &= \frac{1}{n} \cdot \sum_{k=1}^n [\mathbb{1}(x_k \leq a_j) - \mathbb{1}(x_k \leq a_{j-1})] \end{aligned} \quad (1.2)$$

Рассмотрим возможные значения индикаторов

Если оба индикатора равны единице, это значит, что x_k не больше a_j и не больше a_{j-1} . Поскольку $a_{j-1} \leq a_j$, то можно обойтись тем, что $x \leq a_{j-1}$

$$\begin{aligned} & \begin{cases} \mathbb{1}(x_k \leq a_j) = 1 \\ \mathbb{1}(x_k \leq a_{j-1}) = 1 \\ a_{j-1} < a_j \end{cases} \Rightarrow \begin{cases} x_k \leq a_j \\ x_k \leq a_{j-1} \\ a_{j-1} < a_j \end{cases} \\ & \Rightarrow x_k \leq a_{j-1} < a_j \Rightarrow x_k \leq a_{j-1} \end{aligned}$$

Такая ситуация, что x больше, чем a_j , но не больше, чем a_{j-1} , невозможна, так как a_{j-1} не больше, чем a_j , а признать возможной такое положение дел ($a_j < x_k \leq a_{j-1}$) означало бы то, что $a_j < a_{j-1}$

$$\begin{aligned} & \begin{cases} \mathbb{1}(x_k \leq a_j) = 0 \\ \mathbb{1}(x_k \leq a_{j-1}) = 1 \\ a_{j-1} < a_j \end{cases} \Rightarrow \begin{cases} x_k > a_j \\ x_k \leq a_{j-1} \\ a_{j-1} < a_j \end{cases} \\ & \Rightarrow \begin{cases} a_j < x_k \leq a_{j-1} \\ a_{j-1} < a_j \end{cases} \end{aligned}$$

Если оба индикатора равны нулю, то это значит, что x строго больше как a_j , так и a_{j-1} . Опять же, поскольку $a_{j-1} \leq a_j$, то достаточно сказать, что $x > a_j$.

$$\begin{aligned} & \begin{cases} \mathbb{1}(x_k \leq a_j) = 0 \\ \mathbb{1}(x_k \leq a_{j-1}) = 0 \\ a_j > a_{j-1} \end{cases} \Rightarrow \begin{cases} x_k > a_j \\ x_k > a_{j-1} \\ a_j > a_{j-1} \end{cases} \\ & \Rightarrow x_k > a_j > a_{j-1} \Rightarrow x_k > a_j \end{aligned}$$

Если же x больше, чем a_{j-1} , но не больше, чем a_j , то x попадает в полуинтервал $(a_{j-1}, a_j]$

$$\begin{aligned} & \begin{cases} \mathbb{1}(x_k \leq a_j) = 1 \\ \mathbb{1}(x_k \leq a_{j-1}) = 0 \\ a_j > a_{j-1} \end{cases} \Rightarrow \begin{cases} x_k \leq a_j \\ x_k > a_{j-1} \\ a_j > a_{j-1} \end{cases} \\ & \Rightarrow a_{j-1} < x_k \leq a_j \end{aligned}$$

Вспомним формулу (1.2)

$$\begin{aligned} F_n(a_j) - F_n(a_{j-1}) &= \\ &= \frac{1}{n} \cdot \sum_{k=1}^n [\mathbb{1}(x_k \leq a_j) - \mathbb{1}(x_k \leq a_{j-1})] \end{aligned}$$

Очевидно, что нас интересуют те пары, разность которых не равна нулю. Это значит, что те случаи, когда $x > a_j$ или $x \leq a_{j-1}$, нас не интересуют. Поскольку такой случай, что $a_j < x \leq a_{j-1}$ невозможен, то его тоже отбросим. Значит, остался только тот вариант, когда x попадает в полуинтервал $(a_{j-1}, a_j]$

$$\frac{1}{n} \cdot \sum_{k=1}^n [\mathbb{1}(x_k \leq a_j) - \mathbb{1}(x_k \leq a_{j-1})] = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \in (a_{j-1}, a_j])$$

Видим знакомые полуинтервалы $(a_{j-1}, a_j] = I_j$. Воспользуемся этим

$$\frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \in (a_{j-1}, a_j]) = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \in I_j)$$

Получаем компактную запись для разности функций распределения

$$F_n(a_j) - F_n(a_{j-1}) = \frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}(x_k \in I_j) \quad (1.3)$$

Вернёмся к уравнению (1.1)

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

Воспользовавшись тем, что $(a_j - a_{j-1})$ — длина полуинтервала I_j , а разность $F_n(a_j) - F_n(a_{j-1})$ была только что переписана через индикаторы, получаем такую формулу

$$q_n(y) = \sum_{j=1}^m \frac{1}{n} \sum_{k=1}^n \mathbb{1}(x_k \in I_j) \cdot \frac{1}{|I_j|} \cdot \mathbb{1}(y \in I_j) \quad (1.4)$$

Определение 1.1.4: Гистограмма

Гистограммой выборки x_1, \dots, x_n называется функция

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

Где I_j — полуинтервалы ненулевой длины, покрывающие весь промежуток числовой оси, на который попадают числа из выборки

$$a_0 < \min(X) \leq \max(X) \leq a_m, \quad I_j = (a_{j-1}, a_j], \quad j = \overline{1, m}$$

Упростим формулу (1.4), введя функцию $\nu_j(X)$ [7, стр. 68], которая считает количество элементов выборки $X = x_1, \dots, x_n$, попавших в интервал I_j . Это будет сумма индикаторов того, что элемент x_k попал в I_j

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

Поскольку $\mathbb{1}(y \in I_j)$ зависит от j и не зависит от k , то его можно перенести во внешнюю сумму. Получаем следующую формулу

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

У этой суммы только один ненулевой элемент, так как y может попасть только в один полуинтервал. Тогда обозначим номер отрезка, в который попал y , через k ($y \in I_k$), а функцию $q_n(y)$ запишем как q_n^k

$$q_n^k = \frac{\nu_k(X)}{n \cdot |I_k|} \quad (1.5)$$

Что мы тут видим?

Теперь k — номер столбика гистограммы. В математическом смысле это номер интересующего нас полуинтервала — того, в который попал y .

Высота столбика — значение функции на определённом полуинтервале — пропорциональна количеству элементов, попавших в этот отрезок, что логично.

Делителю $|I_k|$ отведена особая роль — он предотвращает искажение гистограммы, когда длины отрезков разные; когда они одинаковые, можно вынести длину как нормирующий множитель. То есть, чем длиннее отрезок, тем ниже столбик, так как элементы более размазаны по отрезку, что тоже логично.

Представим, что значение функции — это высота прямоугольника, а длина отрезка — его ширина (графически это изображается именно так). Тогда отношение количества элементов, попавших в полуинтервал, к количеству всех элементов выборки (вероятность того, что случайно взятый элемент из выборки попадёт в k -ый отрезок [7, стр. 24]), является площадью прямоугольника. Воспользовавшись формулой (1.5), получаем равенство

$$S_k = q_n^k \cdot |I_k| = \frac{\nu_k(X)}{n} = \mathbb{P}_n(x \in I_k)$$

Если устремить количество полуинтервалов к бесконечности ($m \rightarrow \infty$), то каждый полуинтервал будет сжиматься в точку. При этом вероятность попадания x в отрезок будет стремиться к вероятности попадания x в точку y . Введём обозначения $|I_j| = \delta$, $I_j = \Delta_y$

$$\mathbb{P}_n(x = y) \approx \mathbb{P}_n(x \in \Delta_y) = q_n(y) \cdot \delta, \quad m \rightarrow \infty$$

Очень напоминает ситуацию с плотностью распределения непрерывной случайной величины ξ

$$\mathbb{P}(\xi = x) \approx p(x) \cdot \delta, \quad \delta \rightarrow 0$$

Нужно отметить, что количество элементов выборки должно стремиться к бесконечности ($n \rightarrow \infty$), так как плотность может быть лишь у непрерывных случайных величин. Чем больше будет элементов, тем плотнее они будут стоять на числовой прямой.

1.1.3 Оценка неизвестных параметров

Снова у нас есть x_1, \dots, x_n — выборка из распределения F_θ , где θ — неизвестный параметр из множества Θ .

Пример 1.1.5

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 1.1.6

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

Главный вопрос — определение основных параметров распределения выборки.

Определение 1.1.7: 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 1.1.8: Оценка

Статистику, значение которой заменяет неизвестный параметр, называют оценкой

Пример 1.1.9

Возьмём выборку из распределения из распределения Бернулли, то есть $\{x_i\}$ — набор одинаково распределённых случайных величин, причём

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

Тогда неизвестным параметром будет величина p (вероятность удачного эксперимента)

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

Введём разные оценки \hat{p}

$$\hat{p}_1 = \frac{1}{n} \sum_{k=1}^n x_k$$

$$\hat{p}_2 = x_1$$

$$\hat{p}_3 = \frac{2}{n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} x_k$$

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

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

1. Возникает мысль о том, что разность $\hat{p} - p$ должна быть “маленькой”. Например, чтобы $M(\hat{p} - p)^2$ было самое маленькое из возможных.
2. Также логично желать того, чтобы оценка \hat{p} сходилась к истинному значению параметра p по вероятности ($\hat{p} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} p$) или почти всюду ($\hat{p} \xrightarrow[n \rightarrow \infty]{a.s.} p$)
3. При многократном повторении эксперимента даже самая (на первый взгляд) плохая оценка может оказаться полезной

$$M \hat{p}_1 = p$$

$$M \hat{p}_2 = p$$

$$M \hat{p}_3 = p$$

Например, если целый год каждый день дают набор чисел, а статистик считает значение параметра p с помощью оценки \hat{p}_2 , то в среднем за год у него получится величина, близкая к истинному p .

Определение 1.1.10: 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 1.1.11: Сильно состоятельная оценка

Оценка $\hat{\theta}$ называется сильно состоятельной, если стремится к истин-

ному значению θ почти наверное

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

Пример 1.1.12

Оценка \hat{p}_1 из примера 1.1.9 является сильно состоятельной.

Определение 1.1.13: Несмещённая оценка

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

$$\forall \theta \in \Theta : M_{\theta} \hat{\theta} = \theta$$

Замечание 1.1.14

Несмещённая оценка существует не всегда

Определение 1.1.15

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

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

или же

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

^aВ учебнике Боровкова А. А. “Математическая статистика” оценка, удовлетворяющая этим условиям, носит название **эффективная оценка** [7, стр. 130], но у нас этот термин будет использоваться далее в другом смысле

Пример 1.1.16

Сравним \hat{p}_1 и \hat{p}_3

$$D_p \hat{p}_1 = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p) = \frac{p \cdot (1-p)}{n}$$

$$D_p \hat{p}_3 = \frac{2 \cdot p \cdot (1-p)}{n}$$

1.1.4 Выборочные оценки. Метод моментов

Как восстановить неизвестный параметр $\theta \in \Theta$ из функции распределения $F_\theta(x)$?

Вспомним распределения и их параметры

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 $Exp(\lambda)$. $\frac{1}{\lambda}$ — среднее, $\frac{1}{\lambda^2}$ — дисперсия

Как правило, неизвестный параметр θ можно искать следующим образом

$$\exists \varphi \in C(\mathbb{R}) : \int_{\mathbb{R}} \varphi(x) dF_\theta(x) = g(\theta)$$

Значит, у нас есть уравнение для поиска оценки $\hat{\theta}$ при непрерывной и монотонной $g(\hat{\theta})$

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

Пример 1.1.17

Если θ — среднее, то $\varphi(x) = x$

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

Теорема 1.1.18: Оценка метода моментов

Есть такая непрерывная $\varphi(x)$, что функция $g(\hat{\theta})$ монотонная, непрерывная и ограниченная [7, с. 87]

$$g(\hat{\theta}) = \int_{\mathbb{R}} \varphi(x) dF_\theta(x) \tag{1.6}$$

Тогда оценка $\hat{\theta}$ существует и является сильно состоятельной.

Доказательство. Подставим эмпирическую функцию выборки вместо неизвестной в (1.6)

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

Поскольку функция $g(\hat{\theta})$ непрерывна и монотонна, то она имеет обратную функцию $g^{-1} : g^{-1}(g(\hat{\theta})) = \hat{\theta}$.

Применим обратную функцию к обеим частям уравнения

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

Поскольку выборочная функция распределения при достаточно большом объёме выборки равна неизвестной функции распределения почти всюду, то

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

Функция $g^{-1}(x)$ непрерывна

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

Теорема доказана

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

□

Определение 1.1.19: Выборочное среднее

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

$$\bar{x} = \int_{\mathbb{R}} x dF_n(x)$$

Поскольку все элементы выборки равновероятны, получаем математическое ожидание дискретной равномерно распределённой случайной величины, принимающей n значений

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

Определение 1.1.20: Выборочная дисперсия

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

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

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 и несмещённые оценки θ_1 и θ_2 . Тогда по определению для любой несмещённой оценки $\hat{\theta}$ будет

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

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 $\hat{\theta}$

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

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

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

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

$$D_{\theta} \tilde{\theta} \geq \sigma^2(\theta) \quad (1.7)$$

Честно 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 (частный случай неравенства Гёльдера)

$$\begin{aligned} M_{\theta} [(\theta_1 - \theta) \cdot (\theta_2 - \theta)] &\leq \sqrt{M_{\theta} (\theta_1 - \theta)^2 \cdot M_{\theta} (\theta_2 - \theta)^2} = \\ &= \sqrt{D_{\theta} \theta_1 \cdot D_{\theta} \theta_2} = \sqrt{\sigma_1^2 \cdot \sigma_2^2} \end{aligned} \quad (1.8)$$

И 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

$$D_{\theta} \tilde{\theta} \leq \sigma^2(\theta) \quad (1.9)$$

Воспользовавшись неравенствами (1.7) и (1.9), получаем равенство

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

Это значит, что в неравенстве (1.8) 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 размышлений вспомним аналогию с векторами, а именно смысл равенства в неравенстве Коши для скалярного произведения векторов

$$\vec{a} \cdot \vec{b} = |\vec{a}| \cdot |\vec{b}| \cdot \cos(\widehat{\vec{a}, \vec{b}}) = \sqrt{\vec{a}^2} \cdot \sqrt{\vec{b}^2} \cdot \cos(\widehat{\vec{a}, \vec{b}})$$

Скалярное произведение двух векторов равно произведению их модулей только тогда, когда они сонаправлены

$$(\widehat{\vec{a}, \vec{b}}) = 0 \Rightarrow \vec{a} \cdot \vec{b} = \sqrt{\vec{a}^2} \cdot \sqrt{\vec{b}^2}$$

Положим математическое ожидание нормой, а $\theta_1 - \theta$ и $\theta_2 - \theta$ векторами пространства случайных событий. Получаем, что нормы и направления этих векторов совпадают

$$\begin{aligned} M_{\theta} [(\theta_1 - \theta) \cdot (\theta_2 - \theta)] &= \sqrt{M_{\theta} (\theta_1 - \theta)^2} \cdot \sqrt{M_{\theta} (\theta_2 - \theta)^2} \\ &\Rightarrow (\widehat{\theta_1 - \theta, \theta_2 - \theta}) \end{aligned}$$

Это значит, что они равны, что противоречит предположению о том, что они разные

$$\begin{aligned} \begin{cases} (\widehat{\theta_1 - \theta, \theta_2 - \theta}) = 0 \\ M_{\theta} (\theta_1 - \theta)^2 = M_{\theta} (\theta_2 - \theta)^2 \end{cases} &\Rightarrow \theta_1 - \theta = \theta_2 - \theta \\ &\Rightarrow \theta_1 = \theta_2 \end{aligned}$$

Теорема доказана

□

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 . То есть, её можно дифференцировать под знаком интеграла.^a

Грубо говоря, проходит такой трюк

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

^aПодробнее с дифференцированием под знаком интеграла Римана можно почитать во втором томе курса дифференциального и интегрального исчисления Фихтенгольца [15, с. 712].

Для тех, кто интересуется интегралом Лебега, прямой путь в книгу Дороговцева по общей теории меры и интеграла [12, с. 102]

Также отметим, что выборка (x_1, \dots, x_n) имеет плотность распределения, так как является случайным вектором в \mathbb{R}^n , все компоненты которого — случайные величины.

Определение 1.2.3: Функция правдоподобия

Плотность распределения вектора независимых случайных величин, равная произведению плотностей распределения его компонент, называется функцией правдоподобия

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

Прологарифмировав функцию правдоподобия, получим симпатичную сумму

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

А симпатична она тем, что это сумма независимых одинаково распределённых случайных величин. Воспользовавшись законом больших чисел, можем сказать, что она стремится к сумме n одинаковых математических ожиданий при достаточно большом размере выборки

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

Проблема в том, что мы не знаем среднего. Для разрешения этого вопроса введём ещё одно определение

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:

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

Откуда это взялось. Первая формула следует непосредственно из определения функции правдоподобия

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

Чтобы получить вторую запись, нужно взять производную. Вспоминаем, как правильно дифференцировать сложные функции [14, с. 226], [13, с. 133]

$$\frac{\partial}{\partial x} \ln f(x) = \frac{1}{f(x)} \cdot \frac{\partial}{\partial x} f(x)$$

И считаем

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

□

Замечание 1.2.6

Математическое ожидание вклада выборки равно нулю

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

Доказательство. Посчитаем математическое ожидание вклада выборки

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

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

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

Поскольку интегрируем плотность распределения случайного вектора по всему пространству, то он равен единице. Производная же от единицы равна нулю. Это значит, что математическое ожидание вклада выборки равно нулю

$$M_\theta U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\vec{u}, \theta) d\vec{u} = \frac{\partial}{\partial \theta} 1 = 0$$

□

Замечание 1.2.7
Частная производная по оценке θ от функции правдоподобия $L(\vec{u}, \theta)$ равна нулю.

Доказательство. Выше у нас было равенство

$$\frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\vec{u}, \theta) d\vec{u} = 0$$

Так как производную можем заносить под знак интеграла (согласно замечанию 1.2.2), то получаем такое равенство

$$\int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\vec{u}, \theta) d\vec{u} = 0$$

Поскольку интеграл не зависит от θ , то такое возможно лишь в том случае, когда производная равна нулю

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

□

Определение 1.2.8: Количество информации Фишера

Математическое ожидание квадрата вклада выборки называется количеством информации Фишера

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

Замечание 1.2.9

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

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

Получили такое равенство

$$-M_\theta \frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta) = -M_\theta \frac{\frac{\partial^2}{\partial \theta^2} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} + M_\theta \left[\frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right]^2 \quad (1.10)$$

Помним, что производная от функции правдоподобия по θ равна нулю (замечание 1.2.7). Значит, вторая производная тоже равна нулю

$$\frac{\partial}{\partial \theta} L(\vec{x}, \theta) = 0 \Rightarrow -M_\theta \frac{\frac{\partial^2}{\partial \theta^2} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} = 0$$

От равенства (1.10) осталось лишь это

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

После недолгих преобразований, пользуясь лишь замечанием 1.2.5, видим, что получается нужный нам результат

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

Утверждение доказано

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

□

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

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

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 самому 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

$$1 = M_{\theta} [\hat{\theta} \cdot U(\vec{x}, \theta)] \quad (1.11)$$

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 (1.11). 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 компактное равенство

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

Воспользовавшись неравенством Коши, узнаём произведение корней дисперсии и количества информации больше, чем единица

$$\begin{aligned} 1 &= M_{\theta} [(\hat{\theta} - \theta) \cdot U(\vec{x}, \theta)] \leq \\ &\leq \sqrt{M_{\theta} (\hat{\theta} - \theta)^2} \cdot \sqrt{M_{\theta} U(\vec{x}, \theta)^2} = \sqrt{D_{\theta} \hat{\theta}} \cdot \sqrt{I_n(\theta)} \end{aligned} \quad (1.12)$$

Возводим обе части равенства 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

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

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

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

Выясним, какими свойствами должна обладать плотность, чтобы можно было получить эффективную оценку. Для этого в неравенстве Рао-Крамера нужно рассмотреть случай равенства (так как в этом случае оценка будет самой точной)

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

Рассмотрим неравенство (1.12) и выясним, в каком случае в нём будет стоять знак равенства

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

Снова проводим аналогию с векторами и видим, что скалярное произведение (математическое ожидание произведения) векторов (функций от параметра θ : $f_1(\theta) = \hat{\theta} - \theta$ и $f_2(\theta) = U(\vec{x}, \theta)$) равно произведению их норм (корней математических ожиданий квадратов).

Это в свою очередь означает, что угол между этими векторами (функциями) равен нулю, а сами функции являются линейными комбинациями друг друга. Значит, есть такая функция $k(\theta)$, что $f_2(\theta)$ равняется произведению $f_1(\theta)$ и $k(\theta)$.

$$\begin{aligned} U(\vec{x}, \theta) &= \left(\hat{\theta} - \theta \right) \cdot k(\theta) \\ \frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) &= \hat{\theta} \cdot k(\theta) - \theta \cdot k(\theta) \\ \partial \ln L(\vec{x}, \theta) &= \hat{\theta}(\vec{x}) \cdot k(\theta) \cdot \partial \theta - \theta \cdot k(\theta) \cdot \partial \theta \end{aligned}$$

Проинтегрируем обе части равенства

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

Получим следующее равенство

$$\ln L(\vec{x}, \theta) + c_1(\vec{x}) = \hat{\theta}(\vec{x}) \cdot [a(\theta) + c_2] - [b^*(\theta) + c_3]$$

Сгруппируем константы и введём замену $b(\theta) = -b^*(\theta)$

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

Избавимся от логарифма слева, а для этого возьмём экспоненту от обеих частей равенства

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

При конечном n положим такую плотность распределения

$$p(x_1, \theta) = \exp \left\{ \hat{\theta}(x_1) \cdot a_1(\theta) + b_1(\theta) + c_1(x_1) \right\}$$

В таком случае получим следующую функцию правдоподобия

$$L(\vec{x}, \theta) = \prod_{k=1}^n p(x_k, \theta) = \\ = \exp \left\{ \sum_{k=1}^n \hat{\theta}(x_k) \cdot a_1(\theta) + n \cdot b_1(\theta) + \sum_{k=1}^n c_1(x_k) \right\}$$

Отметим, что в этом случае оценка $\hat{\theta}(\vec{x})$ является суммой оценок по каждой координате (случайной величине)

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

Определение 1.2.13: Экспоненциальное распределение

Распределения следующего вида называются экспоненциальными

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

Также подведём итог предыдущих размышлений.

Утверждение 1.2.14

Для экспоненциальных распределений существует эффективная оценка.

Попробуем найти рецепт выяснения эффективной оценки. Начнём с примера

Пример 1.2.15

Есть выборка x_1, x_2, \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

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

Раскроем скобки

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

Воспользуемся формулой для несмещённой и эффективной оценки среднего

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

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

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

Теперь в числителе очевиден квадрат разности

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

Сворачиваем квадрат разности, а выборочное среднее заносим под знак суммы

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

Видим, что последнее слагаемое не может быть положительным, так как это квадрат со знаком “минус”. Когда оценка θ равна выборочному среднему (идеальный случай), то последнее слагаемое обращается в нуль, а сама функция правдоподобия в таком случае принимает максимальное значение.

Делаем предположение о том, как находить наилучшую оценку

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

Оказывается, именно так она и находится.

Определение 1.2.16: 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 1.2.17

Оценок максимального правдоподобия может быть несколько, а может не существовать ни одной.

Определение 1.2.18: 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 1.2.19

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, поэтому полученные результаты необходимо проверять.

Определение 1.2.20: Вариационный ряд

Вариационный ряд выборки x_1, x_2, \dots, x_n — значения выборки, упорядоченные по возрастанию.

доченные в порядке неубывания

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

Теорема 1.2.21

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

Определение 2.1.1: Симметризация

Симметризация Λ оценки $\hat{\theta}$ — среднее оценок $\hat{\theta}$ для всевозможных перестановок $\sigma \in S_n$ элементов выборки x_1, x_2, \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 2.1.2

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. Берём x_1, x_2, \dots, x_n — независимые одинаково распределённые случайные величины.

Введём обозначения для более короткой записи используемых в доказательстве случайных векторов.

Вектор, состоящий из элементов выборки в их изначальном порядке, обозначим привычным \vec{x}

$$(x_1, x_2, \dots, x_n) = \vec{x}$$

Вектор, состоящий из элементов, изменивших своё местоположение под влиянием перестановки σ (значение которой будет ясно из контекста), будем обозначать через \vec{x}_{σ}

$$(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = \vec{x}_{\sigma}$$

Тогда и оценки примут более красивый вид

$$\begin{aligned}\hat{\theta}(x_1, x_2, \dots, x_n) &= \hat{\theta}(\vec{x}) \\ \hat{\theta}(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) &= \hat{\theta}(\vec{x}_\sigma)\end{aligned}$$

Теперь приступим непосредственно к доказательству.

1. Начнём с первого пункта — докажем несмещённость симметризации оценки $\hat{\theta}$.

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 функции случайных одинаково распределённых векторов. 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 $\hat{\theta}$

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

Помним, что математическое ожидание линейно и константы можно выносить за знак математического ожидания, а математическое ожидание суммы равно сумме математических ожиданий

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

Не забываем, что математическое ожидание оценки любого вектора \vec{x}_σ одинаково и равно параметру θ

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

Сумма имеет $n!$ слагаемых (количество перестановок $\sigma \in S_n$)

$$\frac{1}{n!} \cdot \sum_{\sigma \in S_n} \theta = \frac{1}{n!} \cdot n! \cdot \theta = \theta$$

А это значит, что первый пункт доказан и симметризация несмещённой оценки $\hat{\theta}$ действительно несмещённая

$$M_\theta \Lambda \hat{\theta} = \theta$$

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. Для этого нужно умножить и поделить его на $n!$ (так как сумма имеет $n!$ слагаемых)

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

Вспомним неравенство Йенсена для выпуклой функции f [13, с. 167]

$$f \left(\sum_{i=1}^n q_i \cdot x_i \right) \leq \sum_{i=1}^n q_i \cdot f(x_i), \quad \sum_{i=1}^n q_i = 1$$

В нашем случае $x_i = \left(\hat{\theta}(\vec{x}_{\sigma_i}) - \theta \right)$, функция $f(x) = x^2$, сумма проходит по всевозможным перестановкам σ , а роль q_i выполняет $\frac{1}{n!}$, так как

$$\sum_{\sigma \in S_n} q_i = \sum_{\sigma \in S_n} \frac{1}{n!} = n! \cdot \frac{1}{n!} = 1$$

Перепишем неравенство Йенсена для нашего случая

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

Воспользуемся линейностью математического ожидания, внося его под знак суммы

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

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 Йенсена (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 2.1.3

Равенство в неравенстве Йенсена (в формуле (2.1) из доказательства выше) возможно только в случае симметричной функции. Значит, в качестве оценки достаточно брать только симметричные функции выборки

Определение 2.1.4: Функция вариационного ряда

Если оценка $\hat{\theta}$ симметрична относительно перестановок аргументов, то она является функцией вариационного ряда

Замечание 2.1.5

Все оценки, которые претендуют быть оптимальными, должны быть функциями вариационного ряда

2.2 σ -алгебра, порождённая случайной величиной

Имеем вероятностное пространство $(\Omega, \mathfrak{F}, \mathbb{P})$, также есть функция $\xi : \Omega \rightarrow \mathbb{R}$ такая, что связанные с ней множества измеримы по Лебегу

$$\{\omega \mid \xi(\omega) < c\} \in \mathfrak{F}, c \in \mathbb{R}$$

Но это будет неудобно при использовании, поэтому возьмём борелевские подмножества \mathfrak{B} множества \mathbb{R}

$$\mathbb{R} \supset \mathfrak{B} \ni \Delta : \xi^{-1}(\Delta) \in \mathfrak{F}$$

Рассмотрим более подробно, что же означает запись $\xi^{-1}(\Delta)$

$$\xi^{-1}(\Delta) = \{\omega \mid \xi(\omega) \in \Delta\}, \Delta \in \mathfrak{B}, \omega \in \Omega$$

Определение 2.2.1: Сигма-алгебра, порождённая случайной величиной

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

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

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

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

А это значит, что все элементы σ -алгебры \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 . А поскольку \mathbb{R} принадлежит борелевской σ -алгебре, то его прообраз по определению принадлежит σ -алгебре \mathfrak{F}_ξ

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

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

$$\begin{aligned} A &= \xi^{-1}(\Delta) = \{\omega \mid \xi(\omega) \in \Delta\} \\ \Rightarrow \bar{A} &= \{\omega \mid \xi(\omega) \notin \Delta\} = \{\omega \mid \xi(\omega) \in \bar{\Delta}\} \\ \bar{A} &= \xi^{-1}(\bar{\Delta}) \end{aligned}$$

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

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

3. Замкнутость относительно счётных пересечений.

Начнём с замкнутости относительно пересечения двух множеств

$$A = \xi^{-1}(\Delta_1), B = \xi^{-1}(\Delta_2)$$

Начинаем считать

$$\begin{aligned} A \cap B &= \xi^{-1}(\Delta_1) \cap \xi^{-1}(\Delta_2) = \\ &= \{\omega \mid \xi(\omega) \in \Delta_1\} \cap \{\omega \mid \xi(\omega) \in \Delta_2\} = \\ &= \{\omega \mid \xi(\omega) \in \Delta_1 \wedge \xi(\omega) \in \Delta_2\} = \\ &= \{\omega \mid \xi(\omega) \in \Delta_1 \cap \Delta_2\} = \xi^{-1}(\Delta_1 \cap \Delta_2) \end{aligned}$$

Значит, имеем равенство

$$\xi^{-1}(\Delta_1) \cap \xi^{-1}(\Delta_2) = \xi^{-1}(\Delta_1 \cap \Delta_2)$$

Пользуясь методом математической индукции нетрудно показать, что для любого n выполняется

$$\xi^{-1}\left(\bigcap_{i=1}^n \Delta_i\right) = \bigcap_{i=1}^n \xi^{-1}(\Delta_i), \Delta_i \in \mathfrak{B}$$

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

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

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 \mathfrak{F}_ξ на уровни — непересекающиеся подмножества

Определение 2.2.2: Множество уровня

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

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

Замечание 2.2.3

Уровни H_i составляют разбиение множества элементарных исходов Ω .

1. Множества H_i не пересекаются

$$H_{t_1} \neq H_{t_2} \Leftrightarrow t_1 \neq t_2$$

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 в свою очередь 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

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

Прежде, чем продолжить, зафиксируем явление и дадим ему название.

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 на-

бором гипотез и будет выглядеть следующим образом

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

Нам нет смысла пользоваться лишь одной случайной величиной ξ . Нас интересует, как устроены случайные величины, которые измеримы относительно σ -алгебры $\sigma(\xi)$.

Определение 2.3.2: Случайная величина, измеримая относительно сигма-алгебры

Тот факт, что случайная величина \varkappa измерима относительно \mathfrak{F}_1 , значит, что все прообразы случайной величины \varkappa лежат в σ -алгебре \mathfrak{F}_1

$$\forall c \in \mathbb{R} : \{\omega \mid \varkappa(\omega) \leq c\} \in \mathfrak{F}_1$$

Возьмём случайную величину κ , измеримую относительно σ -алгебры $\sigma(\xi)$

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

То есть, прообразы \varkappa выражаются через объединения уровней H_k

$$\{\omega \mid \varkappa(\omega) \leq c\} = \bigcup_{k=1}^n H_k^{\eta_k}$$

Введём обозначение

$$A(c) = \{\omega \mid \varkappa(\omega) \leq c\}$$

Очевидно, что при $c \rightarrow -\infty$ прообразом является пустое множество, а когда $c \rightarrow +\infty$, то прообразом является всё множество элементарных исходов

$$\begin{aligned} \{\omega \mid \varkappa(\omega) \leq -\infty\} &= \{\omega \mid \varkappa(\omega) \in \emptyset\} = \varkappa^{-1}(\emptyset) = \emptyset \\ \{\omega \mid \varkappa(\omega) \leq +\infty\} &= \{\omega \mid \varkappa(\omega) \in \mathbb{R}\} = \varkappa^{-1}(\mathbb{R}) = \Omega \end{aligned}$$

Также ясно, что, если имеются два элемента борелевского множества и один включён в другой, то полный прообраз первого элемента тоже будет включён в прообраз второго

$$\begin{aligned} \Delta_1, \Delta_2 \in \mathfrak{B}, \Delta_1 \subseteq \Delta_2 \\ \Rightarrow \varkappa^{-1}(\Delta_1) \subseteq \varkappa^{-1}(\Delta_1) \cup \varkappa^{-1}(\Delta_2) = \\ = \varkappa^{-1}(\Delta_1 \cup \Delta_2) = \varkappa^{-1}(\Delta_2) \end{aligned}$$

Ни у кого не возникает сомнений, что справедливо и такое утверждение

$$c_1, c_2 \in \mathbb{R}, c_1 \leq c_2 \Rightarrow A(c_1) \subseteq A(c_2)$$

Объединим и проанализируем вышеописанное:

1. Количество элементов в множестве $A(c)$ не уменьшается с ростом c

$$c_1 \leq c_2 \Rightarrow A(c_1) \subseteq A(c_2)$$

2. Множество $A(c)$ “разрастается” от пустого множества \emptyset до множества элементарных событий Ω с ростом c от $-\infty$ до $+\infty$

$$A(-\infty) = \emptyset, A(+\infty) = \Omega$$

3. Множество $A(c)$ растёт дискретными шагами. Это связано с тем, что уровни H_k в нашей σ -алгебре неделимые, а каждый её элемент должен состоять из объединений этих уровней и ничего другого.

Из этого всего делаем более конкретные выводы о том, как изменяется значение функции $A(c)$ с ростом параметра c . Должны быть опорные точки, на которых происходит “скачок” — точки, на которых к объединению добавляется ещё один или более уровней.

Поскольку имеется n уровней, то может быть не более n скачков: ведь самый “медленный” рост будет происходить, если добавлять по одному уровню на определённых константах, а нужно пройти всё от пустого множества \emptyset до множества элементарных исходов Ω .

Выделим m точек ($m \leq n$) $c_1 < c_2 < \dots < c_m$ на числовой прямой \mathbb{R} как значения случайной величины \varkappa

$$\varkappa : \Omega \rightarrow \{c_1, c_2, \dots, c_m\}$$

Посмотрим, как соотносятся между собой $A(c_i)$ и $A(c_{i-1})$, чтобы лучше понять природу скачков.

Сначала покажем, что $A(c_1)$ является прообразом c_1

$$\varkappa^{-1}(c_1) = \{\omega \mid \varkappa(\omega) = c_1\}$$

Поскольку случайная величина не принимает значений до c_1 , то множество $A(c_1 - 0) = \{\omega \mid \varkappa(\omega) < c_1\}$ пустое. Получаем то, что хотели

$$\begin{aligned} \varkappa^{-1}(c_1) &= \{\omega \mid \varkappa(\omega) = c_1\} \cup \emptyset = \\ &= \{\omega \mid \varkappa(\omega) = c_1\} \cup \{\omega \mid \varkappa(\omega) < c_1\} = \\ &= \{\omega \mid \varkappa(\omega) \leq c_1\} = A(c_1) \end{aligned}$$

Идём дальше. Обозначим $c_0 = -\infty$. Тогда в каждой точке $A(c_i)$, $i = \overline{1, m}$ происходит скачок на множество $\varkappa^{-1}(c_i)$, то есть

$$A(c_i) = A(c_{i-1}) \cup \varkappa^{-1}(c_i)$$

Так происходит, потому что имеет место равенство, которое выполняется из-за того, что функция имеет скачки лишь на параметрах c_i , а между ними не меняет значения

$$A(c_i) = A(c_{i+1} - 0)$$

В таком случае тождество очевидно

$$\begin{aligned} A(c_i) &= \{\omega \mid \varkappa(\omega) \leq c_i\} = \\ &= \{\omega \mid \varkappa(\omega) < c_i\} \cup \{\omega \mid \varkappa(\omega) = c_i\} = \\ &= A(c_{i-1} - 0) \cup \varkappa^{-1}(c_i) = A(c_{i-1}) \cup \varkappa^{-1}(c_i) \end{aligned}$$

Поскольку \varkappa — случайная величина, принимающая m значений, то её прообразы составляют разбиение пространства элементарных исходов Ω . А поскольку $A(c_{i-1})$ состоит из объединений этих прообразов, то оно не пересекается с $\varkappa^{-1}(c_i)$. То есть, мы знаем, как вычислять прообраз \varkappa

$$\begin{cases} A(c_{i-1}) \cap \varkappa^{-1}(c_i) = \emptyset \\ A(c_i) = A(c_{i-1}) \cup \varkappa^{-1}(c_i) \end{cases} \Rightarrow \varkappa^{-1}(c_i) = A(c_i) \setminus A(c_{i-1})$$

Значит, случайная величина \varkappa принимает значение c_i при выпадении любого элементарного исхода ω из множества $A(c_i) \setminus A(c_{i-1})$

$$\varkappa(\omega) = c_i, \omega \in A(c_i) \setminus A(c_{i-1}) \quad (2.2)$$

Запишем это в более удобном виде

$$\varkappa(\omega) = \sum_{i=1}^m c_i \cdot \mathbb{1}\{\omega \in A(c_i) \setminus A(c_{i-1})\}$$

Но эта сумма кажется уродливой из-за длинного индикатора и непонятного m . Попытаемся разобраться, в чём же дело и как прийти к изначальной n и милым H_k .

Помним, что $A(c_i) \setminus A(c_{i-1})$ — объединение нескольких множеств уровня H_k .

Для любого t разность множеств $A(c_t) \setminus A(c_{t-1}) \neq \emptyset$ (когда это множество пустое, то индикатор просто не сработает и нечего считать) можно представить как объединение двух непересекающихся множеств, которые обозначим $H_1^t \in \mathfrak{F}$ и $H_2^t \in \mathfrak{F}$, причём H_1^t — множество уровня, а H_2^t — произвольное множество из \mathfrak{F} (в том числе и пустое, если разность и есть множество уровня). Тогда t -ое слагаемое примет следующий вид

$$c_t \cdot \mathbb{1}\{\omega \in A(c_t) \setminus A(c_{t-1})\} = c_t \cdot \mathbb{1}\{\omega \in H_1^t \cup H_2^t\}$$

Поскольку множества H_1^t и H_2^t по условию не пересекаются, то можно разбить индикатор на сумму

$$\begin{aligned} c_t \cdot \mathbb{1}\{\omega \in H_1^t \cup H_2^t\} &= c_t \cdot (\mathbb{1}\{\omega \in H_1^t\} + \mathbb{1}\{\omega \in H_2^t\}) \\ &= c_t \cdot \mathbb{1}\{\omega \in H_1^t\} + c_t \cdot \mathbb{1}\{\omega \in H_2^t\} \end{aligned}$$

Если ввести две константы c_1^t и c_2^t , которые будут равны старой c_t , то равенство примет более симпатичный вид

$$c_t \cdot \mathbb{1}\{\omega \in H_1^t\} + c_t \cdot \mathbb{1}\{\omega \in H_2^t\} = c_1^t \cdot \mathbb{1}\{\omega \in H_1^t\} + c_2^t \cdot \mathbb{1}\{\omega \in H_2^t\}$$

Если же H_2^t не является пустым множеством \emptyset или множеством уровня H_k , то нужно повторить процедуру, разбив H_2^t на объединение двух непересекающихся множеств — на множество уровня и множество из \mathfrak{F} . В итоге (вследствие конечности множества \mathfrak{F}) индикатор разности $A(c_t) \setminus A(c_{t-1})$ будет разбита на сумму индикаторов множеств уровней.

Таким же образом можно поступить со всеми остальными индикаторами. В итоге получим n констант d_1, d_2, \dots, d_n вместо m чисел c_1, c_2, \dots, c_m .

Теперь сумма примет более приятный для глаз и понятный из контекста начала раздела вид

$$\varkappa(\omega) = \sum_{i=1}^m c_i \cdot \mathbb{1}\{\omega \in A(c_i) \setminus A(c_{i-1})\} = \sum_{i=1}^n d_i \cdot \mathbb{1}\{\omega \in H_i\}$$

Видим, что теперь можно определить отображение из множества значений, принимаемых случайной величиной ξ , в множество значений, принимаемых случайной величиной \varkappa

$$f : \{a_1, a_2, \dots, a_n\} \rightarrow \{d_1, d_2, \dots, d_n\}$$

Попробуем показать, что \varkappa является функцией от ξ . Очевидно, что случайная величина ξ имеет такой же вид, что и \varkappa — сумма констант, умноженных на индикаторы, так как мы только что показали, что все функции, измеримые относительно σ -алгебры, порождённой случайной величиной ξ , выглядят именно так

$$f(\xi(\omega)) = f\left(\sum_{i=1}^n a_i \cdot \mathbb{1}\{\omega \in H_i\}\right)$$

Поскольку уровни H_i не пересекаются, то лишь одно слагаемое не будет равно нулю: ω может принадлежать лишь одному уровню. В таком случае запись принимает свой изначальный вид без суммы (2.2)

$$f(\xi(\omega)) = f(a_i), \quad \omega \in H_i$$

Замечаем, что $f(a_i) = d_i$, а это и есть то значение, которое принимает случайная величина \varkappa на уровне H_i

$$f(\xi(\omega)) = f(a_i) = d_i = \varkappa(\omega), \quad \omega \in H_i$$

Поскольку мы не привязывались к конкретным i и конкретным ω , то получаем желаемое равенство

$$\varkappa = f(\xi)$$

Отсюда делаем следующий вывод

Утверждение 2.3.3

Случайной величине \varkappa необходимо и достаточно быть функцией случайной величины ξ , чтобы быть измеримой относительно σ -алгебры, порождённой случайной величиной ξ .

2.4 Условное математическое ожидание

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

2.4.1 Проекция вектора

Для наглядности начнём с геометрической интерпретации задачи. Если представить η как вектор в некоем пространстве \mathfrak{L} , а $\sigma(\xi)$ как подпространство пространства \mathfrak{L} , то $\tilde{\eta}$ будет ни что иное, как проекция случайной величины η на пространство $\sigma(\xi)$.

Отдохнём от случайных величин и вспомним геометрию.

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 .

У нас есть e_1, e_2, \dots, e_n — ортонормированный базис в L , тогда y можно найти по формуле

$$y = \sum_{k=1}^n (x, e_k) \cdot e_k \quad (2.3)$$

Потому что $y \in L$ должен лежать в пространстве L по условию, а это значит, что он должен быть линейной комбинацией базисных векторов e_1, e_2, \dots, e_n и это очевидно выполняется

Также разностью $x - y$ должен быть вектор, перпендикулярный пространству L . То есть, скалярное произведение этой разности с любым вектором z из пространства L должно равняться нулю

$$(x - y) \perp L \Leftrightarrow \forall z \in L : (x - y, z) = 0$$

Вследствие линейности скалярного произведения можно переписать это условие иначе

$$\begin{cases} \forall z \in L : (x - y, z) = 0 \\ (a + b, c) = (a, c) + (b, c) \end{cases} \Rightarrow \forall z \in L : (x, z) = (y, z)$$

Покажем, что и это выполняется. z является линейной комбинацией базисных векторов. Запишем это

$$z = \sum_{k=1}^n \beta_k \cdot e_k$$

В таком случае скалярное произведение (x, z) будет таким

$$(x, z) = \sum_{k=1}^n \beta_k \cdot (x, e_k)$$

С произведением (y, x) придётся чуть-чуть повозиться

$$(y, x) = \left(\sum_{k=1}^n (x, e_k) \cdot e_k, \sum_{k=1}^n \beta_k \cdot e_k \right) = \sum_{k=1}^n (x, e_k) \cdot \beta_k$$

Как видим, суммы равны, а значит, проекция x на L найдена верно.

2.4.2 Проекция случайной величины

Возьмём L — множество всех случайных величин, которые измеримы относительно $\sigma(\xi)$.

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

Но что же взять в качестве ортонормированного базиса? По внешнему виду элементов пространства L кажется, что это $\mathbb{1}_{H_k}$. В качестве скалярного произведения случайных величин возьмём математическое ожидание произведения.

Оказывается, H_k действительно ортогональны

$$k_1 \neq k_2 \Rightarrow H_{k_1} \cap H_{k_2} = \emptyset \Rightarrow M[\mathbb{1}_{H_{k_1}} \cdot \mathbb{1}_{H_{k_2}}] = 0$$

Теперь нужно нормировать эти базисные вектора, а для этого их надо поделить на их нормы. В нашем пространстве норма порождена скалярным произведением, то есть

$$\|x\| = \sqrt{(x, x)} = \sqrt{M[x \cdot x]} = \sqrt{M[x^2]}, x \in L$$

Теперь у нас есть всё необходимое для того, чтобы представить ортонормированный базис. Начнём преобразования H_k

$$e_k = \frac{\mathbb{1}_{H_k}}{\sqrt{M(\mathbb{1}_{H_k})^2}}$$

Поскольку индикатор может принимать лишь одно из двух значений 0 или 1, а их квадраты равны им самим, то в формуле квадрат тоже можно убрать

$$e_k = \frac{\mathbb{1}_{H_k}}{\sqrt{M \mathbb{1}_{H_k}}}$$

Также помним, что математическое ожидание в знаменателе есть ни что иное, как вероятность события H_k , и теперь у нас есть красивый ортонормированный базис

$$e_k = \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \quad (2.4)$$

Идём дальше, ищем проекцию. Вспомним снова пример с векторами (2.3)

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

Если заменить y на $\tilde{\eta}$, а x на η , то получаем следующую картину, имеющую непосредственное отношение к задаче

$$\tilde{\eta} = \sum_{k=1}^n (\eta, e_k) \cdot e_k$$

Осталось заменить e_k на то, что получили выше (2.4)

$$\tilde{\eta} = \sum_{k=1}^n \left(\eta, \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \right) \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}}$$

Заменяем скалярное произведение на математическое ожидание произведения и получаем то, с чем можно дальше работать, не отвлекаясь на геометрию

$$\tilde{\eta} = \sum_{k=1}^n \mathbb{M} \left(\eta \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}} \right) \cdot \frac{\mathbb{1}_{H_k}}{\sqrt{\mathbb{P}(H_k)}}$$

Поскольку вероятность $\mathbb{P}(H_k)$ — константа, то её можно вынести за математическое ожидание

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

При умножении знаменателей получаем вероятность события H_k . Теперь у нас есть красивая формула для проекции случайной величины

$$\tilde{\eta} = \sum_{k=1}^n \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k} \quad (2.5)$$

На что стоит обратить внимание в этой формуле:

1. $\tilde{\eta}$ — **случайная величина**, так как индикатор вне математического ожидания никуда не девается и результат суммы будет зависеть от произошедшего ω , а точнее от того, какому уровню H_k оно принадлежит
2. Когда ω принадлежит H_k , то результатом суммы будет среднее значение случайной величины η на событии H_k

Если с первым пунктом всё очевидно, то небольшое пояснение ко второму не помешает.

Нужно показать, что k -я “координата” случайной величины $\tilde{\eta}$ действительно даёт среднее значение случайной величины η на событии H_k

$$\frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)}$$

Начнём с определения математического ожидания

$$\begin{aligned} \mathbb{M}[\eta \cdot \mathbb{1}_{H_k}] &= \int_{\Omega} \eta(\omega) \cdot \mathbb{1}_{H_k} \mathbb{P}(d\omega) = \\ &= \int_{H_k} \eta(\omega) \mathbb{P}(d\omega) + \int_{\Omega \setminus H_k} 0 \mathbb{P}(d\omega) \end{aligned} \quad (2.6)$$

Видим математическое ожидание случайной величины, которая гарантированно принимает нулевое значение на множестве $\Omega \setminus H_k$, что в свою очередь искажает желаемую картину и притягивает результат к нулю с силой, которая пропорциональна $\mathbb{P}(\Omega \setminus H_k)$. То есть, “вес” каждого ненулевого значения случайной величины уменьшился.

Почему так происходит? Потому что вероятность события H_k в общем случае не равна единице. Если ввести новую меру $\mathbb{P}_k(A) = \frac{\mathbb{P}(A)}{\mathbb{P}(H_k)}$, то наступит гармония, а вероятность $\mathbb{P}_k(H_k)$ будет равна единице.

Из контекста понятно, что эта мера будет использоваться лишь в интеграле по событию H_k , поэтому её значение будет колебаться в пределах $[0; 1]$, но строгости ради введём небольшую поправку (и увидим, что не напрасно)

$$\mathbb{P}_k(A) = \frac{\mathbb{P}(A \cap H_k)}{\mathbb{P}(H_k)}$$

Видим условную вероятность, а это значит, что мы на правильном пути! Логично, что в поисках условного математического ожидания должна была встретиться условная вероятность

$$\mathbb{P}_k(A) = \frac{\mathbb{P}(A \cap H_k)}{\mathbb{P}(H_k)} = \mathbb{P}(A \mid H_k)$$

Теперь математическое ожидание (2.6) принимает несколько иной вид

$$\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}] = \mathbb{P}(H_k) \cdot \int_{H_k} \eta(\omega) \mathbb{P}(d\omega \mid H_k)$$

Тут уже уровень H_k играет роль целого множества элементарных исходов, его мера $\mathbb{P}(H_k \mid H_k)$ равна единице, а мы получаем действительно среднее значение случайной величины η на множестве H_k , умноженное на вероятность $\mathbb{P}(H_k)$. Значит, осталось лишь поделить обе части на $\mathbb{P}(H_k)$

$$\frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_k}]}{\mathbb{P}(H_k)} = \int_{H_k} \eta(\omega) \mathbb{P}(d\omega \mid H_k)$$

Определение 2.4.1: Условное математическое ожидание случайной величины относительно случайного события

Условное математическое ожидание случайной величины ξ относительно события A [6, стр. 68] обозначается $\mathbb{M}[\xi \mid A]$ и считается по формуле

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

Пользуясь только что введённым обозначением, можно более красиво переписать формулу (2.5) для получения проекции случайной величины η на σ -алгебру, порождённую уровнями H_1, H_2, \dots, H_n

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

Забегая наперёд, введём определение частного случая условного математического ожидания случайной величины относительно σ -алгебры, чтобы обратить внимание на этот важный момент.

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

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

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

Замечание 2.4.3

У нас есть определения условного математического ожидания относительно σ -алгебры \mathfrak{F} (определение 2.4.15) и относительно **случайного события** A (определение 2.4.1). Из контекста будет ясно, какое именно определение используется, поэтому путаницы возникнуть не должно.

Например, последнее определение может выглядеть немного странно

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

Зато при более 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 $M[\eta \mid H_k]$ — 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 , порождённую случайной величиной ξ , принимающей конечное количество значений, выполняется равенство скалярных произведений

$$\forall A \in \mathfrak{F}_\xi : M[\tilde{\eta} \cdot \mathbb{1}_A] = M[\eta \cdot \mathbb{1}_A] \quad (2.7)$$

Доказательство. Для начала распишем $\tilde{\eta}$ по определению

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

Произведение индикаторов $\mathbb{1}_{H_k}$ и $\mathbb{1}_A$ — индикатор пересечения $\mathbb{1}_{H_k \cap A}$. Воспользуемся линейностью математического ожидания, не забывая, что

дробь в каждом слагаемом — константа и выносится за знак математического ожидания

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

Помним, что математическое ожидание индикатора — вероятность

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

Замечаем условную вероятность

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

Поскольку A принадлежит множеству случайных событий \mathfrak{F}_ξ , то 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 уровням, которые являются частью события A , а дальше можно смело воспользоваться линейностью математического ожидания

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

Далее мы имеем полное математическое и моральное право вынести η за знак суммы. Если с математикой всё очевидно (работает закон дистрибутивности), то напомним о морально-этической стороне дела: нам нужно, пройтись по всем возможным индикаторам $\mathbb{1}_{H_k}$, из которых лишь один работает (будет равен единице, а не нулю), поэтому сумма нужна лишь для того, чтобы не писать в конце каждой строчки “для тех ω , что входят в H_k ” (помним, что случайная величина и индикатор — функции от элементарного события ω)

$$\mathbb{M} \left[\sum_{H_k \subseteq A} \eta \cdot \mathbb{1}_{H_k} \right] = \mathbb{M} \left[\eta(\omega) \cdot \sum_{H_k \subseteq A} \mathbb{1}_{H_k}(\omega) \right]$$

Сумма индикаторов непересекающихся событий — индикатор их объединения, которое является множеством A . Не забываем, что оно может состоять из объединений уровней и только из них (или же быть пустым)

$$\mathbb{M} \left[\eta \cdot \sum_{H_k \subseteq A} \mathbb{1}_{H_k} \right] = \mathbb{M}[\eta \cdot \mathbb{1}_A]$$

Значит, равенство (2.7) 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 же 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 называется такая случайная величина $\tilde{\eta}$, что

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

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

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

Замечание 2.4.7

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 2.4.8

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

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

Очевидно, что числа p_{ij} обладают некоторым свойством

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

Посчитаем условное математическое ожидание согласно формуле из определения (2.4.2)

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

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

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

Попробуем обобщить определение условного математического ожидания, чтобы обладать универсальной формулой, из которой можно делать какие-то выводы. Начнём с того, что у нас уже есть

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

Множество уровня H_k — прообраз одного из значений случайной величины ξ , которой порождена σ -алгебра \mathfrak{F}_{ξ} . Если назвать эти значения a_1, a_2, \dots, a_n , то запись примет следующий вид

$$\mathbb{M}[\eta \mid \sigma(\xi)] = \sum_{k=1}^n \mathbb{M}[\eta \mid \xi^{-1}(a_k)] \cdot \mathbb{1}_{\{\xi^{-1}(a_k)\}} \quad (2.8)$$

Вспомним альтернативные записи прообраза

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

И перепишем формулу (2.8)

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

Теперь введём функцию $\varphi^{\eta}(x) = \mathbb{M}[\eta \mid \xi = x]$ и условное математическое ожидание примет следующий вид

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

Вновь вспоминаем роль суммы и индикаторов и видим, что условное математическое ожидание в нашей формуле принимает значение $\varphi^{\eta}(x)$ в зависимости от того, какое значение приняла случайная величина $\xi(\omega)$. То есть, можно переписать равенство следующим образом

$$\mathbb{M}[\eta \mid \xi] = \varphi^{\eta}(a_k) : \xi(\omega) = a_k$$

То есть, можно просто подставить значение случайной величины ξ в качестве аргумента функции φ^η и получим условное математическое ожидание

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

Остановимся ещё немного на функции φ^η . Она является случайной величиной, поэтому перепишем равенство следующим образом

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

Тогда будет корректна следующая запись

$$\varphi^\eta(\xi)(\omega) = \mathbb{M}[\eta \mid \xi = t] \Big|_{t=\xi(\omega)}$$

Не путаем случайную величину $\xi(\omega)$ саму по себе со случайной величиной в случайном событии

$$H_t = \{\xi = t\} = \{\tilde{\omega} \mid \xi(\tilde{\omega}) = t\}$$

Для удобства вернёмся к обозначению H_t

$$\varphi^\eta(\xi)(\omega) = \mathbb{M}[\eta \mid H_t] \Big|_{t=\xi(\omega)} = \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{H_t}]}{\mathbb{P}(H_t)} \Big|_{t=\xi(\omega)}$$

Покажем, что такая формула вычисления условного математического ожидания подходит для общего случая.*

Лемма 2.4.9: Равенство скалярных произведений в общем случае

В общем случае случайная величина $\varphi^\eta(\xi)$ является условным математическим ожиданием случайной величины η относительно σ -алгебры, порождённой случайной величиной ξ

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

Доказательство. Нужно доказать то, что выполняются оба свойства условного математического ожидания.

То, что $\varphi^\eta(\xi)$ измерима относительно $\sigma(\xi)$, очевидно из определения: $\varphi^\eta(\xi)$ является функцией случайной величины ξ , а это и есть измеримость.

Дальше придётся немного повозиться.

$$\forall A \in \sigma(\xi) : \mathbb{M}[\varphi^\eta(\xi) \cdot \mathbb{1}_A] = \mathbb{M}[\eta \cdot \mathbb{1}_A]$$

Следуем определению. Пока что ничего очевидного нет кроме надежды на то, что была выведена достаточно общая формула, которая должна работать

$$\mathbb{M}[\varphi^\eta(\xi) \cdot \mathbb{1}_A] = \int_{\Omega} \varphi^\eta(\xi) \cdot \mathbb{1}_A d\mathbb{P}$$

*Так как формула была выведена из условного математического ожидания относительно σ -алгебры, порождённой случайной величиной, принимающей конечное количество значений, то справедливость формулы для этого случая доказывать уже нет нужды

Применим индикатор и будем интегрировать не по всему множеству элементарных исходов, а лишь по событию A , а также в явном виде покажем элементарный исход ω , так как сейчас с ним надо будет поработать основательно

$$\int_{\Omega} \varphi^{\eta}(\xi) \cdot \mathbb{1}_A d\mathbb{P} = \int_A \varphi^{\eta}(\xi(\omega)) \mathbb{P}(d\omega) \quad (2.9)$$

Теперь нужно немного остановиться и подумать, что же делать дальше. Немного выше оказалось, что сама по себе запись $\varphi^{\eta}(\xi)$ не даёт ничего полезного. Копнём немного глубже и посмотрим на то, что есть у нас. Значение случайной величины использовалось лишь для восстановления случайного события, которому принадлежит произошедший элементарный исход ω^{\dagger} . То есть, мы знали, чему равна случайная величина, но не знали, какое именно событие произошло, зато могли определить, какому уровню принадлежит произошедшее событие. Тут же у нас есть интеграл и мы проходим по каждому мельчайшему событию $d\omega$. Вспомним, чему равна $\varphi^{\eta}(x)$

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

А теперь распишем условное математическое ожидание

$$\varphi^{\eta}(x) = \mathbb{M}[\eta \mid \xi = x] = \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{\{\xi=x\}}]}{\mathbb{P}\{\xi = x\}}$$

В общем случае для непрерывных случайных величин такая запись не имеет смысла, но мы как раз рассматриваем очень маленькие значения, а усложнять нет желания. Поэтому просто подставляем получившееся выражение в интеграл (2.9)

$$\int_A \varphi^{\eta}(\xi(\omega)) \mathbb{P}(d\omega) = \int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{\{\xi=x\}}]}{\mathbb{P}\{\xi = x\}} \mathbb{P}(d\omega) \quad (2.10)$$

Дальше происходит магия, которую можно трактовать по-разному

Формулировка 1: Воспользовавшись вышесказанным, заменим событие $\{\xi = x\}$ на $d\omega$ и продолжим колдовать

$$\int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{\{\xi=x\}}]}{\mathbb{P}\{\xi = x\}} \mathbb{P}(d\omega) = \int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{d\omega}]}{\mathbb{P}(d\omega)} \mathbb{P}(d\omega)$$

Вероятности сокращаются, хоть это и немного смущает, а $d\omega$ находится в индикаторе, что ещё больше нагнетает обстановку. Учтём внесённые изменения и перепишем математическое ожидание через интеграл

$$\int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{d\omega}]}{\mathbb{P}(d\omega)} \mathbb{P}(d\omega) = \int_A \int_{\Omega} \eta \cdot \mathbb{1}_{d\omega} \cdot \mathbb{P}(d\tilde{\omega})$$

[†]Ведь именно по значению случайной величины мы и находили уровни, элементарные исходы которых для нас неразличимы внутри одного множества уровня

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

Не путаемся: $d\omega$ принадлежит внешнему интегралу, а $d\tilde{\omega}$ внутреннему. Индикатор упрощает нашу задачу, сужая пределы интегрирования внутреннего интеграла до маленького события $d\omega$

$$\int_A \int_{\Omega} \eta \cdot \mathbb{1}_{d\omega} \cdot \mathbb{P}(d\tilde{\omega}) = \int_A \int_{d\omega} \eta(\tilde{\omega}) \mathbb{P}(d\tilde{\omega})$$

Поскольку событие $d\omega$ и без того маленькое, дробить его на более мизерные $d\tilde{\omega}$ смысла нет, а это значит, что внутренний интеграл просто уничтожается и остаётся произведение случайной величины η на вероятность события $d\omega$

$$\int_{d\omega} \eta(\tilde{\omega}) \mathbb{P}(d\tilde{\omega}) = \eta(\omega) \cdot \mathbb{P}(d\omega)$$

Формулировка 2: Если посмотреть на исходный двойной интеграл, то можно увидеть условное математическое ожидание η относительно события $\{\xi = x\} = d\omega$

$$\int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{\{\xi=x\}}]}{\mathbb{P}\{\xi=x\}} \mathbb{P}(d\omega) = \int_A \mathbb{M}[\eta \mid d\omega] \mathbb{P}(d\omega)$$

Если определить $d\omega$ как случайное событие, на котором случайная величина η принимает одно и то же значение почти всюду, то математическое ожидание равно значению η при появлении почти любого события из $d\omega^{\ddagger}$ (если значение на промежутке $d\omega$ — константа, то очевидно, что среднее значение будет равно ей же).

$$\int_A \mathbb{M}[\eta \mid d\omega] \mathbb{P}(d\omega) = \int_A \eta \mathbb{P}(d\omega)$$

С этим моментом разобрались, вернёмся же к нашему двойному интегралу (2.10). Получаем такой вот результат

$$\int_A \frac{\mathbb{M}[\eta \cdot \mathbb{1}_{\{\xi=x\}}]}{\mathbb{P}\{\xi=x\}} \mathbb{P}(d\omega) = \int_A \eta \mathbb{P}(d\omega)$$

Но ведь это и есть искомое математическое ожидание! Значит, свойство доказано, формула верна

$$\int_A \eta \mathbb{P}(d\omega) = \int_{\Omega} \eta \cdot \mathbb{1}_A \cdot \mathbb{P}(d\omega) = \mathbb{M}[\eta \cdot \mathbb{1}_A]$$

□

[‡]Нам достаточно постоянства значения $\xi(\omega)$ **почти всюду** на событии $d\omega$, так как интеграл Лебега простой функции (функции, что принимает конечное число значений [12, стр. 53]) — сумма значений функции, умноженных на меры соответствующих им прообразов [12, стр. 69]; в противном случае результатом будет наибольшее значение из интегралов Лебега всех простых функций, не превышающих данную в каждой точке. А это значит, что, если и будут отклонения от основного значения функции ξ на событии $d\omega$, то они будут уничтожаться мерой своих прообразов, равными нулю (в связи с тем, что функция $\xi(\omega)$ равна одному и тому же значению почти всюду на ω)

Теперь вернёмся к менее абстрактным вещам и посмотрим, как выглядит условное математическое ожидание, когда 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

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

Уточним определение функции $\varphi^\eta(x)$ для данного случая. Вот первоначальный вариант

$$\varphi^\eta(x) = \mathbb{M}[\eta \mid \xi = x] = \frac{\mathbb{M}[\eta \cdot \mathbb{1}(\xi = x)]}{\mathbb{P}\{\xi = x\}}$$

В данном (непрерывном) случае вероятность события $\mathbb{P}\{\xi = x\}$ является плотностью случайной величины ξ в точке x

$$\mathbb{P}\{\xi = x\} = r(x)$$

Математическое ожидание случайной величины η , умноженной на индикатор $\mathbb{1}(\xi = x)$, есть ни что иное как математическое ожидание η при фиксированном $\xi = x$

$$\mathbb{M}[\eta \cdot \mathbb{1}(\xi = x)] = \int_{\mathbb{R}} y \cdot p(x, y) \, dy$$

Теперь у нас есть конкретная формула для $\varphi^\eta(x)$ для случая непрерывных случайных величин с общей плотностью распределения

$$\varphi^\eta(x) = \frac{\int_{\mathbb{R}} y \cdot p(x, y) \, dy}{r(x)} = \frac{\int_{\mathbb{R}} y \cdot p(x, y) \, dy}{\int_{\mathbb{R}} p(x, y) \, dy} \quad (2.11)$$

Докажем снова, что $\varphi^\eta(\xi)$ является условным математическим ожиданием случайной величины η относительно σ -алгебры, порождённой случайной величиной ξ . Чтобы не было скучно, будем доказывать несколько иначе, чем ранее.

Лемма 2.4.10: Равенство скалярных произведений условного математического ожидания случайных величин с совместной плотностью

Пусть имеются две случайные величины (ξ, η) с совместной плотностью $p(x, y)$. Тогда функция

$$\varphi^\eta(\xi) = \frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{\int_{\mathbb{R}} p(x, y) dy} \Big|_{x=\xi}$$

Является условным математическим ожиданием $M[\eta | \xi]$

Доказательство. Первое свойство снова очевидно, поэтому надо доказать

$$\forall A \in \sigma(\xi) : M[\varphi^\eta(\xi) \cdot \mathbb{1}_A] = M[\eta \cdot \mathbb{1}_A] \quad (2.12)$$

У нас есть совместная плотность и мы хотим посчитать математическое ожидание, пользуясь именно ею. Для этого превратим индикатор $\mathbb{1}(\omega \in A)$ в функцию случайной величины ξ . Поскольку любое событие A принадлежит $\sigma(\xi)$, то его можно представить в виде $\xi^{-1}(\Delta)$, $\Delta \in \mathfrak{B}$. Перепишем индикатор следующим образом: $\mathbb{1}(\omega \in A) = \mathbb{1}(\xi \in \Delta)$. И вот теперь мы готовы к тому, чтобы записать определение математического ожидания

$$M[\varphi^\eta(\xi) \cdot \mathbb{1}_A] = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi^\eta(x) \cdot \mathbb{1}(x \in \Delta) \cdot p(x, y) dx dy$$

От y зависит лишь совместная плотность, а интеграл от неё по всей оси y является плотностью распределения ξ . То есть, интеграл по y уходит, а вместо $p(x, y)$ появляется $r(x)$. Также учтём индикатор и сузим область интегрирования с \mathbb{R} до Δ

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \varphi^\eta(x) \cdot \mathbb{1}(x \in \Delta) \cdot p(x, y) dx dy = \int_{\Delta} \varphi^\eta(x) \cdot r(x) dx$$

Далее распишем функцию φ^η , пользуясь формулой (2.11)

$$\int_{\Delta} \varphi^\eta(x) \cdot r(x) dx = \int_{\Delta} \left(\frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{r(x)} \cdot r(x) \right) dx$$

Сократим одинаковые плотности и получим интересный двойной интеграл

$$\int_{\Delta} \left(\frac{\int_{\mathbb{R}} y \cdot p(x, y) dy}{r(x)} \cdot r(x) \right) dx = \int_{\Delta} \int_{\mathbb{R}} y \cdot p(x, y) dy dx$$

Вернём индикатор обратно в интеграл

$$\int_{\Delta} \int_{\mathbb{R}} y \cdot p(x, y) \cdot dy dx = \int_{\mathbb{R}} \int_{\mathbb{R}} y \cdot \mathbb{1}(x \in \Delta) \cdot p(x, y) dy dx$$

Видим, что это и есть то математическое ожидание, которое нам нужно

$$\int_{\mathbb{R}} \int_{\mathbb{R}} y \cdot \mathbb{1}(x \in \Delta) \cdot p(x, y) dy dx = M[\eta \cdot \mathbb{1}_{\xi \in \Delta}] = M[\eta \cdot \mathbb{1}_A]$$

Это значит, что тождество доказано и условное математическое ожидание для случайных величин с совместной плотностью считается с помощью

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

По формуле

$$M[\eta | \xi] = \varphi^n(\xi) = \varphi^n(x)|_{x=\xi}$$

□

Теорема 2.4.11: Существование условного математического ожидания

Условное математическое ожидание существует всегда и единственное почти наверное

Доказательство. [7, стр. 142]

□

2.4.4 Свойства условного математического ожидания

Были даны определения условного математического ожидания для разных случаев, теперь настало время привести основные свойства, которые позволят облегчить процедуру вычисления.[§]

I Формула полной вероятности [7, 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

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

[§]Также со свойствами и их доказательствами можно ознакомиться в книгах Ширяева [17, стр. 270] и Боровкова [7, стр. 143]

IV Теорема о трёх перпендикулярах

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

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

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

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

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

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

$$\forall \Delta \in \mathfrak{B}, A \in \mathfrak{F}_1 : \mathbb{P}(\{\eta \in \Delta\} \mid A) = \{\eta \in \Delta\} \Rightarrow M[\eta \mid \mathfrak{F}_1] = M\eta$$

VIII Условное математическое ожидание линейно

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

IX Сохраняется теорема Лебега о возможности предельного перехода под знаком условного математического ожидания [4, стр. 302]. В книге Ширяева это называется теоремой о сходимости под знаком условных ожиданий [17, стр. 272]

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

Пара полезных частных случая неравенства Йенсена (III свойство)

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

2.4.5 Условное математическое ожидание функции произвольной случайной величины

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 $M[\hat{\theta} \mid T]$.

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 $\varphi^n(x) = M[\eta \mid \xi = x]$. Тут 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

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

Для начала нужно понять, что из себя представляет множество точек $S_t = \{\vec{u} \mid g(\vec{u}) = t\}$.

Очевидно, что функция $g(\vec{x})$ описывает скалярное поле в n -мерном пространстве. А 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 2.4.12

Имеем двумерное пространство

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

Функция $g(x, y)$ будет возвращать первую координату

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

Очевидно, что поверхности уровней — вертикальные линии (параллельные оси ординат), так как при изменении y значение функции не меняется

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

Пример 2.4.13

Опять возьмём двумерное пространство

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

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

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

Тут поверхностями уровня будут окружности радиуса \sqrt{t} , так как окружность по определению является геометрическим местом точек, равноудалённых от определённой точки (расстояния до которых оди-

наковые)

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

Пример 2.4.14

Рассмотрим n -мерное пространство ($n \geq 1$). Возьмём единичный вектор \vec{e} : $\|\vec{e}\| = 1$. Функция g будет скалярным произведением аргумента \vec{u} с только что определённым единичным 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^a

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

^a Уравнение гиперплоскости с нормалью \vec{e} , проходящую через точку с радиус-вектором \vec{x} , выглядит следующим образом

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

Вследствие линейности скалярного произведения получаем

$$(t \cdot \vec{e}, \vec{e}) = t \cdot (\vec{e}, \vec{e}) = t$$

Значит, указав $\vec{x} = t \cdot \vec{e}$, получим уравнение, данное в примере

$$(\vec{u}, \vec{e}) = (\vec{x}, \vec{e}) = (t \cdot \vec{e}, \vec{e}) = t \cdot (\vec{e}, \vec{e}) = t$$

В примерах увидели, что получаемые поверхности уровня не имеют объёма в n -мерном пространстве, но у них есть площадь — $(n - 1)$ -мерный объём.

Опираясь на предыдущий опыт (для величин с совместной плотностью), хотелось бы найти совместную плотность случайных величин $f(\vec{x})$ и $g(\vec{x})$. И оказывается, что это желание является верной догадкой.

Нетрудно догадаться, что для того, чтобы найти “вес” поверхности уровня, нужно будет взять поверхностный интеграл от плотности.

Чтобы мы не получали нулевой вес поверхности $S_t = \{g(\vec{x}) = t\}$, будем считать объём её раздутия. Поместим поверхность в своеобразный кокон, толщина которого в каждой точке будет тем меньше, чем больше скорость перехода в этой точке от текущего уровня к следующему.

Чтобы значение t было близким к $g(\vec{u})$, нужно, чтобы точка \vec{u} была близка к поверхности S_t . Обозначим расстояние между t и $g(\vec{u})$ как $\tilde{\varepsilon}$, а расстояние между точкой \vec{u} и поверхностью S_t как ε . Чему равны эти расстояния, будет выяснено ниже, а значение t и точки \vec{u} будет ясно из контекста.

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 — погрешность, которая нас устроит (потом мы, естественно, устремим её к нулю). Обозначим

плотность случайной величины $g(\vec{x})$ в точке t через $q(t)$

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

Вернёмся к раздутию. Помним, что ε — расстояние от точки \vec{u} до ближайшей к нему точки кокона, а также то, что это расстояние должно быть обратно пропорционально стремительности изменения уровней в этой окрестности. Понимаем, что нам необходима численная мера этой скорости. Под описание такой величины прекрасно подходит модуль градиента. Поскольку значение $g(\vec{x})$ не меняется вдоль поверхности S_t и равно t , то градиент будет направлен по нормали к данной точке поверхности.

Норма градиента — отношение прироста функции к приросту координат. Нас интересует прирост координат ε в окрестности точки \vec{u} , мы располагаем приростом функции $\tilde{\varepsilon}$ и нормой градиента $\|\vec{\nabla} \cdot g(\vec{u})\|$. Напрашивается формула

$$\varepsilon \approx \frac{\tilde{\varepsilon}}{\|\vec{\nabla} \cdot g(\vec{u})\|} \quad (2.13)$$

Обозначим раздутие поверхности S_t как $G_{\tilde{\varepsilon}}$, тогда вероятность попадания значения $g(\vec{x})$ в коридорчик ширины $\tilde{\varepsilon}$ будет приблизительно равно интегралу плотности распределения вектора \vec{x} по этому коридорчику

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

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

$$q(t) = \lim_{\tilde{\varepsilon} \rightarrow 0} \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int_{G_{\tilde{\varepsilon}}} p(\vec{u}) d\vec{u} \quad (2.14)$$

Распишем $\tilde{\varepsilon}$, воспользовавшись формулой (2.13)

$$\varepsilon \approx \frac{\tilde{\varepsilon}}{\|\vec{\nabla} \cdot g(\vec{u})\|} \Rightarrow \tilde{\varepsilon} \approx \varepsilon \cdot \|\vec{\nabla} \cdot g(\vec{u})\|$$

Вернёмся к плотности в формуле (2.14). Заменяя $\tilde{\varepsilon}$ на $\varepsilon \cdot \|\vec{\nabla} \cdot g(\vec{u})\|$, нужно разобраться, что теперь нужно устремлять к нулю. Поскольку модуль градиента — величина, зависящая от координат, и стремиться к нулю будет лишь при изменении поведения функции, то устремлять будем ε (толщину кокона)

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

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

$$q(t) = \int_{S_t} p(\vec{u}) \cdot \frac{1}{\|\vec{\nabla} \cdot g(\vec{u})\|} \sigma_t(d\vec{u}) \quad (2.15)$$

А теперь вспомним, как обстояло дело со случайными величинами, имеющими совместную плотность (2.11)

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

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 (2.11) 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

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

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 2.4.15: 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 2.4.16

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 хотя 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 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 2.4.8

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 2.11

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 2.4.15

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. Никаких новых инструментов, к счастью, не появилось:

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

2. Почему возникает определение достаточных статистик? Пускай 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)

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

Помним, что условное математическое ожидание относительно σ -алгебры, порождённой случайной величиной T (функция случайного вектора является случайной величиной), является случайной величиной, измеримой относительно $\sigma(T)$. Мы также помним, что “быть измеримой относительно σ -алгебры, порождённой случайной величиной T ”, это то же самое, что “быть функцией случайной величины T ” (утверждение [2.3.3](#)), а это значит, что существует функция f такая, что

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

Тогда получаем такое красивое равенство

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

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 $\hat{\theta}$

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

Замечание 2.6.3

Оценка $\theta_* = M \left[\hat{\theta} \mid T \right]$ 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. Тут же доказательство пройдёт в несколько этапов.

Сначала распишем дисперсию и оценку 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)

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

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

Пусть $\tilde{\theta}$ — оптимальная несмещённая оценка. Тогда $\theta_* = \mathbb{M} [\hat{\theta} \mid T]$ — оптимальная, а значит, эти оценки равны $\theta_* = \tilde{\theta}$ по теореме единственности (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 [10, стр. 78], [7, стр. 158].

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

$$\mathbb{P}(\vec{x} \in \Delta \mid T = t) = \frac{\int_{S_t \cap \Delta} L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})}{\int_{S_t} L(\vec{u}, \theta) \cdot \frac{1}{\|\vec{\nabla} \cdot f(\vec{u})\|} \sigma_t(d\vec{u})} \quad (2.16)$$

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 (2.16) 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. Пускай T — достаточная статистика. Выпишем плотность распределения T в точке t согласно формуле (2.15), 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 смело 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 и вытянем из неё что-то полезное

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

Рассмотрим сумму 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. За-

меним 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

Есть $\vec{\xi} = (\xi_1, \dots, \xi_n)$ — случайный вектор. С функцией распределения $F(\vec{\xi})$ возникают проблемы (скучновато и громоздко), поэтому будем использовать плотность распределения.

Определение 3.1.1: Плотность распределения случайного вектора

p — плотность распределения случайного вектора $\vec{\xi} = (\xi_1, \dots, \xi_n)$, если

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 3.1.2: Характеристическая функция случайного вектора

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

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

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 3.1.3: Математическое ожидание случайного вектора

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

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

Но что же является дисперсией случайного 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

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

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 [16, с. 244]

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 рас-

пределения

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

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 [11, 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

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

Чтобы не разрывать целостность дальнейших повествований, введём наперёд небольшое утверждение. Точнее, просто вспомним комбинаторику.

Утверждение 3.1.14

Квадрат суммы раскладывается в двойную сумму следующим образом

$$\left(\sum_{k=1}^n x_k \right)^2 = \sum_{i,j=1}^n x_i \cdot x_j$$

Доказательство. Чтобы убедиться в правильности формулы, вспомним мультиномиальные коэффициенты — их значение и определение.

Мультиномиальные коэффициенты — множители при слагаемых $x_1^{m_1} \cdot x_n^{m_n}$ после разложения $(x_1 + \dots + x_n)^m$ в сумму и считаются по следующей формуле [2, с. 23]

$$\binom{m_1, \dots, m_n}{m} = \frac{m!}{m_1! \dots m_n!}$$

$$0 \leq m_1, \dots, m_n \leq m, \quad m_1 + \dots + m_n = m$$

То есть, вот общая формула раскрытия натуральной степени m произвольной суммы выглядит так

$$(x_1 + \dots + x_n)^m = \sum_{\substack{m_1 + \dots + m_n = m \\ m_1, \dots, m_n \geq 0}} \binom{m_1, \dots, m_n}{m} \cdot x_1^{m_1} \dots x_n^{m_n}$$

Теперь вернёмся к нашему частному случаю: $m = 2$. Тогда мультиномиальные коэффициенты будут иметь следующий вид

$$\binom{m_1, \dots, m_n}{2} = \frac{2}{m_1! \dots m_n!}$$

$$0 \leq m_1, \dots, m_n \leq 2, \quad m_1 + \dots + m_n = 2$$

Из накладываемых ограничений видно, что в знаменателе будет либо одна двойка, либо две единицы, так как сумма должна равняться двойке.

Таким образом, сумму можно разбить на две части — квадраты ($m_k = 2$) и попарные произведения ($m_i \cdot m_j = 1, i \neq j$). Запишем

$$\begin{aligned} (x_1 + \dots + x_n)^2 &= \sum_{k=1}^n \frac{2}{2} \cdot x_k^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{1} \cdot x_i \cdot x_j = \\ &= \sum_{k=1}^n x_k^2 + 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j \end{aligned} \quad (3.1)$$

В связи с коммутативностью умножения последнюю удвоенную двойную сумму можно раскрыть как сумму по всем недиагональным элементам

$$2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j = \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j + \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j =$$

$$\begin{aligned}
& \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j + \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j = \\
& \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j + \sum_{j=1}^{n-1} \sum_{i=j+1}^n x_j \cdot x_i = \\
& = \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j + \sum_{i=2}^n \sum_{j=1}^{i-1} x_j \cdot x_i = \sum_{i \neq j}^n x_i \cdot x_j
\end{aligned}$$

Вместе с суммой квадратов диагональных элементов получится сумма по всем произведением. Перепишем, во что превратится формула (3.1)

$$\begin{aligned}
(x_1 + \dots + x_n)^2 &= \sum_{k=1}^n x_k^2 + 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i \cdot x_j = \\
&= \sum_{k=1}^n x_k^2 + \sum_{i \neq j}^n x_i \cdot x_j = \sum_{i,j=1}^n x_i \cdot x_j
\end{aligned}$$

То есть, действительно квадрат суммы равен сумме попарных произведений всех элементов, что и требовалось доказать

$$\left(\sum_{k=1}^n x_k \right)^2 = \sum_{i,j=1}^n x_i \cdot x_j$$

□

Простое доказательство. Также можно доказать это утверждение, просто расписав квадрат как произведение

$$\begin{aligned}
\left(\sum_{k=1}^n x_k \right)^2 &= (x_1 + \dots + x_n) \cdot (x_1 + \dots + x_n) = \\
&= x_1 \cdot x_1 + x_1 \cdot x_2 + \dots + x_1 \cdot x_n + x_2 \cdot x_1 + x_2 \cdot x_2 + \dots + x_n \cdot x_n
\end{aligned}$$

Видим, что каждый элемент умножается с каждым, и всё это дело суммируется. Запишем в виде суммы (с красивым значком сигма)

$$\begin{aligned}
\left(\sum_{k=1}^n x_k \right)^2 &= (x_1 + \dots + x_n) \cdot (x_1 + \dots + x_n) = \\
&= \sum_{i=1}^n (x_i \cdot x_1 + x_i \cdot x_2 + \dots + x_i \cdot x_n) = \sum_{i=1}^n \sum_{j=1}^n x_i \cdot x_j
\end{aligned}$$

Что и требовалось доказать.

□

Теперь мы готовы перейти 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*

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

Это значит следующее

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

Доказательство. Распишем ковариацию по определению и воспользуемся утверждением 3.1.14

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

Поскольку все коэффициенты действительные, а математическое ожидание константы равно самой константе, то делаем вывод, что сумма неотрицательна

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

Вот мы и получили желаемый результат

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

□

Замечание 3.1.15

Вспомним линейную алгебру.

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

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

*Больше о неотрицательно определённых операторах можно почитать в книге Ильина и Позняка “Линейная алгебра” [8, с. 139]. В ней такой оператор называется положительным.

Далее будем упускать символы пустоты \emptyset , 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 3.1.16: 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

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

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

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. Следствие из свойств 2 и 3

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

3.2 Линейные преобразования случайных векторов

Рассмотрим всё тот же случайный вектор $\vec{\xi} = (\xi_1, \dots, \xi_n)$ и произвольный константный 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 из утверждения 3.1.14, 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

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

3.2.2 Поворот случайного вектора

Обобщим задачу и попробуем выяснить, каким образом зависит случайный вектор $\vec{\eta}$, полученный путём линейных преобразований вектора $\vec{\xi}$, имеющего известное математическое ожидание и ковариационную матрицу.

Для линейных преобразований вектора нужен линейный оператор. Назовём его T . Этот оператор будет действовать из пространства \mathbb{R}^n в пространство \mathbb{R}^m , где n — размерность вектора $\vec{\xi}$, а m — размерность вектора $\vec{\eta}$, который будет получен в результате преобразования

$$\vec{\eta} = T \vec{\xi}, \quad T \in \mathbb{R}^{m \times n}$$

Посчитаем математическое ожидание

$$M \eta = M [T \vec{\xi}]$$

Очевидно, что в связи с линейностью математического ожидания можно вынести оператор T наружу.

Мы всё-таки проделаем математические выкладки по-честному. Итак, у нас есть математическое ожидание случайного вектора

$$M [T \vec{\xi}] = M \left\| \sum_{j=1}^n (T_{i,j} \cdot \xi_j) \right\|_{i=1}^m$$

Математическое ожидание случайного вектора — вектор математических ожиданий соответствующих координат. Дальше воспользуемся линейностью математического ожидания

$$M \left\| \sum_{j=1}^n (T_{i,j} \cdot \xi_j) \right\|_{i=1}^m = M \left\| \sum_{j=1}^n (T_{i,j} \cdot \xi_j) \right\|_{i=1}^m = \left\| \sum_{j=1}^n (T_{i,j} \cdot M \xi_j) \right\|_{i=1}^m$$

Видим, что перед нами произведение матрицы T на вектор математических ожиданий координат случайного вектора $\vec{\xi}$

$$\left\| \sum_{j=1}^n (T_{i,j} \cdot M \xi_j) \right\|_{i=1}^m = T M \vec{\xi}$$

То есть, интуиция нам подсказывала правильно и конечная формула такова

$$M \eta = M [T \vec{\xi}] = T M \vec{\xi}$$

Теперь нужно посчитать ковариацию. Мы могли бы решать эту задачу, расписав произведение матрицы или воспользовавшись свойством 4, но в этот раз, пожалуй, освежим наши знания в линейной алгебре.

Возьмём произвольный вектор $\vec{e} \in \mathbb{R}^n$ и выпишем квадратичную форму ковариационной матрицы вектора η с аргументом \vec{e} . Из начала подраздела (3.3) помним, что такая квадратичная форма равна дисперсии скалярного произведения, а дальше воспользуемся свойством симметричности скалярного произведения (для удобства дальнейших вычислений)

$$(\text{Cov}_{\vec{\eta}, \vec{\eta}} \cdot \vec{e}, \vec{e}) = D(\vec{\eta}, \vec{e}) = D(\vec{e}, \vec{\eta})$$

Распишем наш случайный вектор $\vec{\eta}$ через случайный вектор $\vec{\xi}$ и матрицу T

$$D(\vec{e}, \vec{\eta}) = D(\vec{e}, T \vec{\xi})$$

Далее воспользуемся ещё одним определением сопряжённого оператора[†] и перенесём оператор T в левую часть скалярного произведения

$$D(\vec{e}, T \vec{\xi}) = D(T^* \vec{e}, \vec{\xi})$$

Перейдём от дисперсии к квадратичной форме и посмотрим, что происходит

$$D(T^* \vec{e}, \vec{\xi}) = (\text{Cov}_{\vec{\xi}, \vec{\xi}} T^* \vec{e}, T^* \vec{e})$$

Снова воспользуемся определением сопряжённого оператора и перенесём его из правой стороны скалярного произведения в левую. Не забываем, что сопряжённый оператор к сопряжённому оператору — исходный оператор $(T^*)^* = T$

$$(\text{Cov}_{\vec{\xi}, \vec{\xi}} T^* \vec{e}, T^* \vec{e}) = (T \text{Cov}_{\vec{\xi}, \vec{\xi}} T^* \vec{e}, \vec{e})$$

Видим, что квадратичные формы совпадают, а это значит, что и операторы равны

$$(T \text{Cov}_{\vec{\xi}, \vec{\xi}} T^* \vec{e}, \vec{e}) = (\text{Cov}_{\vec{\eta}, \vec{\eta}} \cdot \vec{e}, \vec{e}) \Rightarrow T \text{Cov}_{\vec{\xi}, \vec{\xi}} T^* = \text{Cov}_{\vec{\eta}, \vec{\eta}}$$

Подведём итоги: если на случайный вектор $\vec{\xi}$ с известным математическим ожиданием и ковариационной матрицей подействовать оператором T , то математическое ожидание полученного вектора будет считаться по формуле

$$M T \vec{\xi} = T M \vec{\xi}$$

Расчёт ковариационной матрицы происходит в базисе вектора $\vec{\xi}$ с матрицей перехода T и матрицей T^* для перехода обратно

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

Подведём итоги в виде утверждения

[†] На самом деле, это и есть изначальное определение сопряжённого оператора [11, с. 241], [8, с. 126]

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

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

3.2.3 Линейные преобразования случайных векторов

Логичным продолжением всего вышесказанного будет вычисление характеристик суммы двух случайных векторов $\vec{\xi}$ и $\vec{\eta}$, умноженных слева на матрицы C и D .

Условимся, что вектор $\vec{\xi}$ находится в пространстве \mathbb{R}^n , а вектор $\vec{\eta}$ находится в \mathbb{R}^m . Тогда очевидно, что оператор C должен принадлежать $\mathbb{R}^{k \times n}$, а оператор D в свою очередь должен быть из множества $\mathbb{R}^{k \times m}$.

Итого, задание: посчитать характеристики случайного вектора $\vec{\gamma}$

$$\vec{\gamma} = C \vec{\xi} + D \vec{\eta}$$

Математическое ожидание считается крайне просто — достаточно воспользоваться линейностью

$$M \vec{\gamma} = C M \vec{\xi} + D M \vec{\eta}$$

С ковариационной матрицей проблем возникнуть тоже не должно — будем использовать свойства из раздела 3.1.2.

Для начала распишем ковариационную матрицу в терминах исходной задачи

$$\text{Cov}_{\vec{\gamma}, \vec{\gamma}} = \text{Cov}_{C \vec{\xi} + D \vec{\eta}, C \vec{\xi} + D \vec{\eta}}$$

Сначала разобьём ковариацию на сумму двух ковариаций

$$\text{Cov}_{C \vec{\xi} + D \vec{\eta}, C \vec{\xi} + D \vec{\eta}} = \text{Cov}_{C \vec{\xi}, C \vec{\xi} + D \vec{\eta}} + \text{Cov}_{D \vec{\eta}, C \vec{\xi} + D \vec{\eta}}$$

Воспользуемся симметричностью

$$\text{Cov}_{C \vec{\xi}, C \vec{\xi} + D \vec{\eta}} + \text{Cov}_{D \vec{\eta}, C \vec{\xi} + D \vec{\eta}} = \text{Cov}_{C \vec{\xi} + D \vec{\eta}, C \vec{\xi}}^T + \text{Cov}_{C \vec{\xi} + D \vec{\eta}, D \vec{\eta}}^T$$

Разобьём на суммы и снова воспользуемся симметричностью, чтобы избавиться от транспонированных матриц

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

Вынесем операторы за знаки ковариаций

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

Мне больше нравится, чтобы по бокам стояли ковариации $\vec{\xi}$ и $\vec{\eta}$, а внутри уже ковариации обоих векторов. Запишем результат именно в этом виде

$$\text{Cov}_{C\vec{\xi}+D\vec{\eta}, C\vec{\xi}+D\vec{\eta}} = C \text{Cov}_{\vec{\xi}, \vec{\xi}} C^T + C \text{Cov}_{\vec{\xi}, \vec{\eta}} D^T + D \text{Cov}_{\vec{\eta}, \vec{\xi}} C^T + D \text{Cov}_{\vec{\eta}, \vec{\eta}} D^T$$

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

$$\vec{m} = C M \vec{\xi} + D M \vec{\eta}$$

$$M = C \text{Cov}_{\vec{\xi}, \vec{\xi}} C^T + C \text{Cov}_{\vec{\xi}, \vec{\eta}} D^T + D \text{Cov}_{\vec{\eta}, \vec{\xi}} C^T + D \text{Cov}_{\vec{\eta}, \vec{\eta}} D^T$$

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

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

Пример 3.3.2

Возьмём случайный вектор $\vec{\xi}$, координаты которого между собой равны 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 произведение и видим, что результат 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 ко-

вариационная 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 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 $\vec{\lambda}$

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

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

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

3.3.2 Линейные преобразования

Лемма 3.3.8

Пускай $\vec{\xi}$ — случайный n -элементный вектор, имеющий гауссовское распределение с параметрами \vec{a} и A

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

Тогда при воздействии оператора T на случайный вектор $\vec{\xi}$ получим случайный вектор, имеющий гауссовское распределение с параметрами $T\vec{a}$ и TAT^*

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

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

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

В силу того, что вектор $\vec{\xi}$ является гауссовским, то значение с правой стороны равенства является случайной гауссовской величиной. Так как вектор \vec{a} после воздействия на него оператором T^* всё равно остаётся одним из векторов пространства (разве что размерность вектора стала такой, какая нам нужна), а скалярное произведение в итоге будет гауссовской величиной по определению случайного гауссовского вектора.

Поскольку справа имеем случайную величину, имеющую нормальный закон распределения, то слева от знака равенства стоит та же случайная величина. То есть, гауссовский вектор остаётся таковым после линейных преобразований. \square

Лемма 3.3.9

Есть гауссовский вектор $\vec{\xi}$ с параметрами \vec{a} и A и произвольный вектор \vec{b} из \mathbb{R}^n . Сумма $\vec{\xi} + \vec{b}$ будет иметь гауссовское распределение с параметрами $\vec{a} + \vec{b}$ и A

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

Доказательство. Начнём с того, что полученный вектор действительно гауссовский. Снова воспользуемся определением 3.3.1, а также аддитивностью (распределительное свойство [11, с. 82]) и симметричностью скалярного произведения

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

Очевидно, что последнее скалярное произведение (\vec{e}, \vec{b}) — константа.

Поскольку скалярное произведение $(\vec{e}, \vec{\xi})$ является гауссовской случайной величиной, то константа лишь сдвинет его математическое ожидание. То есть, $\vec{\xi} + \vec{b}$ действительно является гауссовским вектором.

Математическое ожидание распишем по формуле. Тут всё элементарно — лишь воспользуемся линейностью математического ожидания и векторов

$$M[\vec{\xi} + \vec{b}] = \begin{bmatrix} M[\xi_1 + b_1] \\ \vdots \\ M[\xi_n + b_n] \end{bmatrix} = \begin{bmatrix} M\xi_1 + b_1 \\ \vdots \\ M\xi_n + b_n \end{bmatrix} = \begin{bmatrix} M\xi_1 \\ \vdots \\ M\xi_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = M\vec{\xi} + \vec{b}$$

Осталась ковариация. Начнём с определения 3.1.8 и опять же используем линейность математического ожидания

$$\begin{aligned} \text{Cov}_{\vec{\xi} + \vec{b}, \vec{\xi} + \vec{b}} &= \|M\{(\xi_i + b_i - M[\xi_i + b_i]) \cdot (\xi_j + b_j - M[\xi_j + b_j])\}\|_{i,j=1}^n = \\ &= \|M\{(\xi_i + b_i - M[\xi_i] - b_i) \cdot (\xi_j + b_j - M[\xi_j] - b_j)\}\|_{i,j=1}^n = \\ &= \|M\{(\xi_i - M\xi_i) \cdot (\xi_j - M\xi_j)\}\|_{i,j=1}^n = \text{Cov}_{\vec{\xi}, \vec{\xi}} = A \end{aligned}$$

То есть, вектор $\vec{\xi} + \vec{b}$ — случайный гауссовский вектор с параметрами $\vec{a} + \vec{b}$ и A , что и требовалось доказать

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

□

Теорема 3.3.10

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{K}

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

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

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

Из предыдущих лемм 3.3.8 и 3.3.9 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 3.1.15 мы вспоминали линейную алгебру, а именно — тот момент, что самосопряжённая неотрицательная определённая матрица A имеет собственный ортонормированный базис, в котором она превращается в диагональную матрицу с неотрицательными элементами

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

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

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

□

Теорема 3.3.11: 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 3.3.10 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

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

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 $\Delta \in \mathfrak{B}$

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

Теперь у нас появилось более чёткое представление об области интегрирования. 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

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

Можем переписать интеграл

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

Теперь вспомним несколько моментов, а именно:

1. Определитель — мультипликативная функция матрицы. Из этого следует красивый вывод

$$V^2 = A \Rightarrow \det V = \sqrt{\det A} \Rightarrow \det V^{-1} = \frac{1}{\sqrt{\det A}}$$

2. 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 оговоренные замены

$$\int_{\Delta} q(V^{-1}(\vec{u} - \vec{a})) \cdot \det V^{-1} d\vec{u} = \int_{\Delta} \frac{1}{\sqrt{2 \cdot \pi^n}} \cdot \frac{1}{\sqrt{\det A}} \cdot e^{(A^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))} d\vec{u}$$

Остановимся и внимательно посмотрим, что же это такое. Из равенства (3.4) вспоминаем, что этот интеграл — вероятность того, что вектор $\vec{\eta}$ попадёт в область Δ . Напишем это, чтобы не забыть

$$\mathbb{P}\{\vec{\eta} \in \Delta\} = \int_{\Delta} p(\vec{u}) d\vec{u} = \int_{\Delta} \frac{\exp\{(A^{-1}(\vec{u} - \vec{a}), (\vec{u} - \vec{a}))\}}{\sqrt{2 \cdot \pi^n} \cdot \sqrt{\det A}} d\vec{u}$$

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 3.3.12

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

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

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[‡]

$$(\exists \vec{\alpha} \neq \vec{0}) (\forall j) : \text{cov} \left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \xi_j \right) = 0$$

Поскольку это равенство выполняется для любых j , то их сумма тоже будет равняться нулю, причём каждое слагаемое можно умножить на любую константу. Пускай 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

$$\exists \vec{\alpha} \neq \vec{0} : \text{cov} \left(\sum_{i=1}^n \alpha_i \cdot \xi_i, \sum_{j=1}^n \alpha_j \cdot \xi_j \right) = 0$$

Заменяв 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

$$D \left(\vec{\alpha}, \vec{\xi} \right) = 0 \Rightarrow \exists c : \mathbb{P} \left\{ \left(\vec{\alpha}, \vec{\xi} \right) = c \right\} = 1$$

[‡]Третье свойство ковариации [5, с. 310]

То есть, мы получили уравнение гиперплоскости — $(n - 1)$ -мерное подпространство

$$H = \{\vec{x} \in \mathbb{R}^n \mid (\vec{a}, \vec{x}) = c\}$$

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 3.3.13

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

$$\vec{m} = CM\vec{\xi} + DM\vec{\eta}$$

$$M = C \text{Cov}_{\vec{\xi}, \vec{\xi}} C^T + C \text{Cov}_{\vec{\xi}, \vec{\eta}} D^T + D \text{Cov}_{\vec{\eta}, \vec{\xi}} C^T + D \text{Cov}_{\vec{\eta}, \vec{\eta}} D^T$$

Доказательство. Начнём с того, что результирующий вектор будет действительно гауссовский по аналогии с доказательством леммы 3.3.9

$$\forall \vec{e} \in \mathbb{R}^k : (\vec{e}, C\vec{\xi} + D\vec{\eta}) = (\vec{e}, C\vec{\xi}) + (\vec{e}, D\vec{\eta}) = (C^* \vec{e}, \vec{\xi}) + (D^* \vec{e}, \vec{\eta})$$

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. □

3.3.3 Независимость компонент гауссовского вектора

Определение 3.3.14: 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 3.3.15

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 3.1.16 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 [17, с. 354].

Из определения характеристической функции случайного вектора 3.1.2 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 $\vec{\xi} \circ \vec{\eta}$ распадается 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 3.3.11

$$\begin{aligned} \varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) &= \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_{n+m} e^{i \cdot \lambda_1 \cdot x_1} \dots e^{i \cdot \lambda_n \cdot x_n} \cdot e^{i \cdot \mu_1 \cdot y_1} \dots e^{i \cdot \mu_m \cdot y_m} \times \\ &\quad \times \frac{e^{-\frac{1}{2} \cdot (\vec{A}^{-1}(\vec{u} - \vec{a}), \vec{u} - \vec{a})}}{\sqrt{2 \cdot \pi}^n \cdot \sqrt{\det \vec{A}}} dx_1 \dots dx_n dy_1 \dots dy_m \end{aligned}$$

К счастью, это уже пройденный этап, и мы просто возьмём готовую формулу из замечания 3.3.7

$$\varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) = \exp \left\{ i \cdot (\vec{\lambda} \circ \vec{\mu}, \mathbf{M} [\vec{\xi} \circ \vec{\eta}]) - \frac{1}{2} \cdot (\vec{A} [\vec{\lambda} \circ \vec{\mu}], \vec{\lambda} \circ \vec{\mu}) \right\} \quad (3.7)$$

Прделаем небольшой трюк со скалярным произведением, в котором присутствует математическое ожидание

$$\begin{aligned} (\vec{\lambda} \circ \vec{\mu}, \mathbf{M} [\vec{\xi} \circ \vec{\eta}]) &= \lambda_1 \cdot \mathbf{M} \xi_1 + \dots + \lambda_n \cdot \mathbf{M} \xi_n + \mu_1 \cdot \mathbf{M} \eta_1 + \dots + \mu_m \cdot \mathbf{M} \eta_m = \\ &= (\vec{\lambda}, \mathbf{M} \vec{\xi}) + (\vec{\mu}, \mathbf{M} \vec{\eta}) \end{aligned}$$

Из формулы (3.6) помним, как выглядит ковариационная матрица конкатенации. Применим то, что совместная ковариация случайных векторов $\vec{\xi}$ и $\vec{\eta}$ нулевая

$$\begin{cases} \vec{A} = \begin{bmatrix} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \text{Cov}_{\vec{\xi}, \vec{\eta}} \\ \text{Cov}_{\vec{\eta}, \vec{\xi}} & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{bmatrix} \\ \text{Cov}_{\vec{\xi}, \vec{\eta}} = \text{Cov}_{\vec{\eta}, \vec{\xi}} = 0 \end{cases} \Rightarrow \vec{A} = \begin{bmatrix} \text{Cov}_{\vec{\xi}, \vec{\xi}} & \emptyset \\ \emptyset & \text{Cov}_{\vec{\eta}, \vec{\eta}} \end{bmatrix}$$

Дальше будем действовать по аналогии с математическим ожиданием

$$\begin{aligned} (\vec{A} [\vec{\lambda} \circ \vec{\mu}], \vec{\lambda} \circ \vec{\mu}) &= (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda} \circ \text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\lambda} \circ \vec{\mu}) = \\ &= (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda}) + (\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu}) \end{aligned}$$

С учётом вышесказанного характеристическая функция (3.7) принимает вид

$$\begin{aligned} \varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) &= \exp \left\{ i \cdot (\vec{\lambda} \circ \vec{\mu}, \mathbf{M} [\vec{\xi} \circ \vec{\eta}]) - \frac{1}{2} \cdot (\vec{A} [\vec{\lambda} \circ \vec{\mu}], \vec{\lambda} \circ \vec{\mu}) \right\} = \\ &= \exp \left\{ i \cdot (\vec{\lambda}, \mathbf{M} \vec{\xi}) + i \cdot (\vec{\mu}, \mathbf{M} \vec{\eta}) - \frac{1}{2} \cdot (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda}) - \frac{1}{2} \cdot (\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu}) \right\} \end{aligned}$$

Сгруппировав переменные, видим, что характеристическая функция разбилась на произведение двух характеристических функций

$$\begin{aligned}\varphi_{\vec{\xi} \circ \vec{\eta}}(\vec{\lambda} \circ \vec{\mu}) &= e^{i \cdot (\vec{\lambda}, M \vec{\xi}) - \frac{1}{2} \cdot (\text{Cov}_{\vec{\xi}, \vec{\xi}} \vec{\lambda}, \vec{\lambda})} \cdot e^{i \cdot (\vec{\mu}, M \vec{\eta}) - \frac{1}{2} \cdot (\text{Cov}_{\vec{\eta}, \vec{\eta}} \vec{\mu}, \vec{\mu})} = \\ &= \varphi_{\vec{\xi}}(\vec{\lambda}) \cdot \varphi_{\vec{\eta}}(\vec{\mu})\end{aligned}$$

Характеристическая функция конкатенации разбилась в произведение характеристических функций, а это значит, что теорема доказана. \square

Замечание 3.3.16

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 3.3.15 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 3.3.17

Если $\vec{\xi} \circ \vec{\eta}$ — гауссовский вектор, то следующее 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 3.3.18: 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 $\vec{\eta}$

$$M \left[\varphi_{\vec{\xi}}(\vec{\lambda}) \Big| \vec{\eta} \right] = M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi})} \Big| \vec{\eta} \right]$$

Прделаем небольшой трюк — добавим и отнимем в правой части скалярного произведения вектор $B \vec{\eta}$, где матрица B известна из равенства (3.8)

$$M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi})} \Big| \vec{\eta} \right] = M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - B \vec{\eta} + B \vec{\eta})} \Big| \vec{\eta} \right]$$

Воспользуемся аддитивностью (распределительное свойство [11, с. 82]) и симметричностью скалярного произведения, разделив его на сумму двух скалярных произведений

$$M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - B \vec{\eta} + B \vec{\eta})} \Big| \vec{\eta} \right] = M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - B \vec{\eta}) + i \cdot (\vec{\lambda}, B \vec{\eta})} \Big| \vec{\eta} \right]$$

Экспонента суммы разбивается на произведение экспонент. Вторая экспонента очевидно является случайной величиной, измеримой относительно случайного вектора $\vec{\eta}$, так как является её функцией, матрица B — всего лишь линейный оператор, содержащий константы, а вектор $\vec{\lambda}$ — некий константный вектор

$$e^{i \cdot (\vec{\lambda}, B \vec{\eta})} = f(\vec{\eta})$$

Значит, эту экспоненту можно вынести за знак условного математического ожидания согласно VI свойству условного математического ожидания

$$M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - B \vec{\eta}) + i \cdot (\vec{\lambda}, B \vec{\eta})} \Big| \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, B \vec{\eta})} \cdot M \left[e^{i \cdot (\vec{\lambda}, \vec{\xi} - B \vec{\eta})} \Big| \vec{\eta} \right]$$

Случайный вектор, находящийся во второй экспоненте, как было выяснено в начале раздела (утверждение 3.3.17), 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

$$\varphi_{\vec{\xi} - \mathbf{B} \vec{\eta}}(\vec{\lambda}) = \exp \left\{ i \cdot \left(\vec{\lambda}, \mathbb{M} \left[\vec{\xi} - \mathbf{B} \vec{\eta} \right] \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\vec{\xi} - \mathbf{B} \vec{\eta}, \vec{\xi} - \mathbf{B} \vec{\eta}} \vec{\lambda}, \vec{\lambda} \right) \right\}$$

Формулы для подсчёта математического ожидания и ковариации у нас есть в утверждении 3.3.13. В нашем случае

$$\mathbf{C} \vec{\xi} + \mathbf{D} \vec{\eta} = \mathbf{I} \vec{\xi} - \mathbf{B} \vec{\eta}$$

То есть, на вектор $\vec{\xi}$ действует оператор эквивалентного преобразования и его можно будет дальше упустить.

Математическое ожидание выглядит вот так

$$\mathbb{M} \left[\vec{\xi} - \mathbf{B} \vec{\eta} \right] = \mathbf{I} \mathbb{M} \vec{\xi} - \mathbf{B} \mathbb{M} \vec{\eta} = \mathbb{M} \vec{\xi} - \mathbf{B} \mathbb{M} \vec{\eta}$$

Ковариация несколько длиннее (не забываем про минус возле матрицы \mathbf{B})

$$\text{Cov}_{\vec{\xi} - \mathbf{B} \vec{\eta}, \vec{\xi} - \mathbf{B} \vec{\eta}} = \mathbf{I} \text{Cov}_{\vec{\xi}, \vec{\xi}} \mathbf{I}^T - \mathbf{I} \text{Cov}_{\vec{\xi}, \vec{\eta}} \mathbf{B}^T - \mathbf{B} \text{Cov}_{\vec{\eta}, \vec{\xi}} \mathbf{I}^T + \mathbf{B} \text{Cov}_{\vec{\eta}, \vec{\eta}} \mathbf{B}^T \quad (3.10)$$

С математическим ожиданием уже почти покончили, а вот в ковариации лучше расписать оператор \mathbf{B} , поскольку тогда сократится несколько некрасивых вещей. Также учтём, что результат транспонирования произведения двух матриц — произведение транспонированных матриц в обратном порядке

$$(\mathbf{A} \mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T$$

Итак, приступим. У нас есть матрица

$$\mathbf{B} = \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1}$$

Это значит, что при транспонировании мы получим следующее

$$\mathbf{B}^T = (\text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1})^T \text{Cov}_{\tilde{\xi}, \tilde{\eta}}^T$$

Помним, что транспонированная обратная матрица — обратная к транспонированной, а также вспоминаем то, что ковариация симметрична и транспонирование лишь меняет порядок аргументов в ней

$$\mathbf{B}^T = (\text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1})^T \text{Cov}_{\tilde{\xi}, \tilde{\eta}}^T = (\text{Cov}_{\tilde{\eta}, \tilde{\eta}}^T)^{-1} \text{Cov}_{\tilde{\xi}, \tilde{\eta}}^T = \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}}$$

Подставим полученный результат в равенство (не забываем про минус перед матрицей \mathbf{B}) 3.10

$$\begin{aligned} & \mathbf{I} \text{Cov}_{\tilde{\xi}, \tilde{\xi}} \mathbf{I}^T - \mathbf{I} \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \mathbf{B}^T - \mathbf{B} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} \mathbf{I}^T + \mathbf{B} \text{Cov}_{\tilde{\eta}, \tilde{\eta}} \mathbf{B}^T = \\ & = \text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \mathbf{B}^T - \mathbf{B} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} + \mathbf{B} \text{Cov}_{\tilde{\eta}, \tilde{\eta}} \mathbf{B}^T = \\ & = \text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} + \\ & \quad + \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} \end{aligned}$$

Вторая и третья матрицы одинаковые, поэтому можно поставить множитель 2 и сэкономить место. В последнем произведении сократятся матрицы $\text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1}$ и $\text{Cov}_{\tilde{\eta}, \tilde{\eta}}$

$$\begin{aligned} & \text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} + \\ & \quad + \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} = \\ & = \text{Cov}_{\tilde{\xi}, \tilde{\xi}} - 2 \cdot \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} + \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} \end{aligned}$$

Последняя матрица уходит вместе с двойкой и остаётся только разность

$$\text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}}$$

Итого, равенство 3.9 теперь примет вид

$$\begin{aligned} & \mathbf{M} \left[\varphi_{\tilde{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \varphi_{\tilde{\xi} - \mathbf{B} \vec{\eta}}(\lambda) = \\ & = e^{i \cdot (\vec{\lambda}, \mathbf{B} \vec{\eta})} \cdot \exp \left\{ i \cdot \left(\vec{\lambda}, \mathbf{M} \left[\vec{\xi} - \mathbf{B} \vec{\eta} \right] \right) - \frac{1}{2} \cdot \left(\text{Cov}_{\tilde{\xi} - \mathbf{B} \vec{\eta}, \tilde{\xi} - \mathbf{B} \vec{\eta}} \vec{\lambda}, \vec{\lambda} \right) \right\} = \\ & = \exp \left\{ i \cdot \left(\vec{\lambda}, \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \vec{\eta} \right) \right\} \cdot \exp \left\{ i \cdot \left(\vec{\lambda}, \mathbf{M} \vec{\xi} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \mathbf{M} \vec{\eta} \right) \right\} \times \\ & \quad \times \exp \left\{ -\frac{1}{2} \cdot \left(\left(\text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} \right) \vec{\lambda}, \vec{\lambda} \right) \right\} \end{aligned}$$

Для красоты объединим две экспоненты

$$\begin{aligned} & \mathbf{M} \left[\varphi_{\tilde{\xi}}(\vec{\lambda}) \mid \vec{\eta} \right] = \exp \left\{ i \cdot \left(\vec{\lambda}, \mathbf{M} \vec{\xi} + \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} (\vec{\eta} - \mathbf{M} \vec{\eta}) \right) \right\} \times \\ & \quad \times \exp \left\{ -\frac{1}{2} \cdot \left(\left(\text{Cov}_{\tilde{\xi}, \tilde{\xi}} - \text{Cov}_{\tilde{\xi}, \tilde{\eta}} \text{Cov}_{\tilde{\eta}, \tilde{\eta}}^{-1} \text{Cov}_{\tilde{\eta}, \tilde{\xi}} \right) \vec{\lambda}, \vec{\lambda} \right) \right\} \end{aligned}$$

То есть, при известном $\vec{\eta}$ случайный гауссовский вектор $\vec{\xi}$ имеет гауссовское распределение с такими характеристиками

$$\vec{\xi} \mid \vec{\eta} \sim N(\vec{m}, D)$$

$$\begin{aligned}\vec{m} &= M \left[\vec{\xi} \mid \vec{\eta} \right] = M \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - M \eta) \\ D &= \text{Cov}_{\vec{\xi}, \vec{\xi}} - \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} \text{Cov}_{\vec{\eta}, \vec{\xi}}\end{aligned}$$

□

Замечание 3.3.19

Мы получили формулу для подсчёта условного математического ожидания одного гауссовского вектора относительно другого

$$M \left[\vec{\xi} \mid \vec{\eta} \right] = M \vec{\xi} + \text{Cov}_{\vec{\xi}, \vec{\eta}} \text{Cov}_{\vec{\eta}, \vec{\eta}}^{-1} (\vec{\eta} - M \eta)$$

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}

$$\int_{x_1^2 + \dots + x_n^2 \leq t} (2 \cdot \pi)^{-\frac{n}{2}} \cdot e^{-\frac{u_1^2 + \dots + u_n^2}{2}} du_1 \dots du_n = c' \cdot \int_0^{\sqrt{t}} \rho^{n-1} \cdot e^{-\frac{\rho^2}{2}} d\rho$$

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 [13, с. 416], устремим t к бесконечности, и найдём константу c' . Эта константа не зависит от t , поэтому имеем право

$$\begin{aligned} c' \cdot \int_0^{\frac{t}{2}} S^{\frac{n-1}{2}-1} \cdot e^{-S} dS &\xrightarrow[t \rightarrow \infty]{a} c' \cdot \int_0^{+\infty} S^{\frac{n-1}{2}-1} \cdot e^{-S} dS = \\ &= \left| \begin{array}{l} \Gamma(\alpha) = \int_0^{+\infty} e^{-x} \cdot x^{\alpha-1} dx \\ \alpha = \frac{n-1}{2} \\ x = S \end{array} \right| = c' \cdot \Gamma\left(\frac{n-1}{2}\right) \Rightarrow c' = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} \end{aligned}$$

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 [9, 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 распределением 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 Стьюдента**Определение 3.4.4: Распределение Стьюдента**

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 [3]

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 3.3.15)
2. После поворота гауссовский вектор остаётся гауссовским (лемма 3.3.8)
3. Чтобы вычислить совместное распределение гауссовских векторов, достаточно матриц (теорема 3.3.18)

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 1.2.13), то для него существует эффективная оценка (утверждение 1.2.14).

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 1.2.16)

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

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 3.3.9 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 выписываем 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}

$$\vec{\theta}_* = B \vec{y}, \vec{y} = A \vec{\theta} + \vec{\xi}$$

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 1.1.13 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 $\mathfrak{L}_{\vec{\xi}}$.

Пространство всех несмещённых оценок параметра $\vec{\theta}$ с помехами $\vec{\xi}$ будет $\mathfrak{M}_{\vec{\xi}}$.

Очевидно, что пространство линейных несмещённых оценок параметра $\vec{\theta}$ является подпространством всех несмещённых оценок параметра $\vec{\theta}$.

$$\mathfrak{L}_{\vec{\xi}} \subset \mathfrak{M}_{\vec{\xi}}$$

Введём ещё одну группу обозначений.

Пусть $\vec{\eta}$ — случайный вектор в \mathbb{R}^n с нулевым средним и единичной ковариацией, как у вектора $\vec{\xi}$

$$\begin{aligned}\vec{\eta} &\sim N(\vec{0}, I) \\ M \vec{\eta} &= M \vec{\xi} = \vec{0} \\ \text{Cov}_{\vec{\eta}, \vec{\eta}} &= \text{Cov}_{\vec{\xi}, \vec{\xi}} = I\end{aligned}$$

Вектору $\vec{\eta}$ поставим в соответствие пространства $\mathfrak{L}_{\vec{\eta}}$ и $\mathfrak{M}_{\vec{\eta}}$.

Что такое линейная несмещённая оценка параметра $\vec{\theta}$? Это матрицы, применённые к выборке

$$\mathfrak{L}_{\vec{\xi}} \ni T \vec{y}: M T \vec{y} = T A \vec{\theta} = \vec{\theta}$$

Возьмём последнее равенство из формулы и видим, что тут ничего не сказано о помехах — ни о $\vec{\eta}$, ни о $\vec{\xi}$

$$T A \vec{\theta} = \vec{\theta}$$

Тогда делаем вывод, что пространство линейных несмещённых оценок едино для любых шумов

$$\mathfrak{L}_{\vec{\eta}} = \mathfrak{L}_{\vec{\xi}} \ni T \vec{y}$$

Подсчитаем ковариацию вектора $T \vec{y}$ и видим, что она от шумов тоже не зависит

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

Следовательно, искать оптимальную линейную несмещённую оценку в $\mathfrak{L}_{\vec{\xi}}$ — то же самое, что искать оптимальную линейную несмещённую оценку в $\mathfrak{L}_{\vec{\eta}}$.

Значит, мы будем использовать пространство $\mathfrak{L}_{\vec{\eta}}$, 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 $\mathfrak{M}_{\vec{\xi}}$. Ещё она линейна, а это значит, что входит в пространство $\mathfrak{L}_{\vec{\xi}}$. 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 1.1.7) $T_1 \leq T_2$ 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 1.2.8) 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

$$A = \|\delta_{ij} - \sqrt{p_i} \cdot \sqrt{p_j}\|_{i=1}^n$$

Матрица A должна быть симметричная и неотрицательно определённая. Симметричность очевидна, неотрицательную определённость нужно проверить. С этой целью введём вектор \vec{e}

$$\vec{e} = (\sqrt{p_1}, \dots, \sqrt{p_m}) \|e\| = 1$$

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] Ralf P. Grimaldi. *Discrete and combinatorial mathematics an applied introduction*. Addison-Wesley Publishing Company, 1999.
- [3] Student. The probable error of a mean. *Biometrika*, VI.
- [4] С. В. Фомин А. Н. Колмогоров. *Элементы теории функций и функционального анализа*. Наука, Москва, 1976.
- [5] Г.М. Цветкова и др. А.В. Печинкин, О.И. Тескин. *Теория вероятностей*. МГТУ им. Н. Э. Баумана, Москва, 2004.
- [6] А. А. Боровков. *Теория Вероятностей*. Эдиториал УРСС, Москва, 1999.
- [7] А. А. Боровков. *Математическая статистика*. Лань, Санкт-Петербург, 2010.
- [8] Э. Г. Позняк В. А. Ильин. *Линейная алгебра*. Наука. Физматлит, Москва, 1999.
- [9] Э. Г. Позняк В. А. Ильин. *Основы математического анализа. Часть I*. Физико-математическая литература, Москва, 2005.
- [10] Г.М. Цветкова и др. В.Б. Горяинов, И.В. Павлов. *Математическая статистика*. МГТУ им. Н. Э. Баумана, Москва, 2001.
- [11] В. В. Воеводин. *Линейная алгебра*. Наука. Главная редакция физико-математической литературы, Москва, 1980.
- [12] А. Я. Дороговцев. *Элементы общей теории меры и интеграла*. Выща школа. Головное издательство, Киев, 1989.
- [13] А. Я. Дороговцев. *Математический анализ. Краткий курс в современном изложении*. Факт, Киев, 2004.
- [14] Фихтенгольц Г. М. *Курс дифференциального и интегрального исчисления. Том 1*. Физматлит, Москва, 2003.
- [15] Фихтенгольц Г. М. *Курс дифференциального и интегрального исчисления. Том 2*. Физматлит, Москва, 2003.

- [16] В. Феллер. *Введение в теорию вероятностей и её приложения. Том 1.* Мир, Москва, 1984.
- [17] А. Н. Ширяев. *Вероятность-1.* МЦНМО, Москва, 2004.

Оглавление

Вступление	3
1 Основы	5
1.1 Методы оценок характеристик наблюдаемых случайных величин	5
1.1.1 Эмпирическая функция распределения	5
1.1.2 Гистограмма	6
1.1.3 Оценка неизвестных параметров	11
1.1.4 Выборочные оценки. Метод моментов	14
1.2 Свойства оценок	16
1.2.1 Неравенство Рао-Крамера	16
1.2.2 Метод максимального правдоподобия	23
2 Достаточные статистики	29
2.1 Оптимальная оценка	29
2.2 σ -алгебра, порождённая случайной величиной	32
2.3 Случайная величина, измеримая относительно σ -алгебры	35
2.4 Условное математическое ожидание	39
2.4.1 Проекция вектора	40
2.4.2 Проекция случайной величины	41
2.4.3 Условное математическое ожидание	46
2.4.4 Свойства условного математического ожидания	53
2.4.5 Условное математическое ожидание функции произвольной случайной величины	54
2.4.6 Пример	59
2.5 Условные распределения	63
2.6 Достаточные статистики	64
3 Случайные вектора	71
3.1 Основные характеристики случайного вектора	71
3.1.1 Ковариационная матрица случайного вектора	72
3.1.2 Ковариационная матрица	78
3.2 Линейные преобразования случайных векторов	79
3.2.1 Скалярное произведение	79
3.2.2 Поворот случайного вектора	80
3.2.3 Линейные преобразования случайных векторов	82
3.3 Гауссовские случайные вектора	83
3.3.1 Определения	83

3.3.2	Линейные преобразования	86
3.3.3	Независимость компонент гауссовского вектора	92
3.3.4	Теорема о нормальной корреляции	95
3.4	Распределения, связанные с нормальным	99
3.4.1	Распределение Пирсона (хи-квадрат, χ^2)	99
3.4.2	Распределение Фишера	101
3.4.3	Распределение Стьюдента	102
4	Методы оценивания неизвестных параметров	105
4.1	Вступление	105
4.2	Оценка метода наименьших квадратов	107
4.2.1	Размышления	107
4.2.2	Общая постановка задачи метода наименьших квадратов	109
4.2.3	Оценка метода наименьших квадратов при неединич- ной ковариации	112
4.3	Интервальное оценивание	112
4.3.1	Определения	112
4.3.2	Рецепт построения доверительного интервала с помо- щью центральной статистики	113
4.3.3	Асимптотические доверительные интервалы	113
4.4	Критерий согласия	114
4.4.1	Критерий Колмогорова-Смирнова	114
4.4.2	Рецепт применения критерия Колмогорова-Смирнова .	115
4.4.3	Критерий Пирсона χ^2	116
4.4.4	Рецепт применения критерия Пирсона	120
	Литература	121
	Оглавление	123