Математическая статистика

7 iюня 2014 г.

Глава 1

Основы

1.1 Методы оценок характеристик наблюдаемых случайных величин

$x_1, \ldots, x_n$ — независимые одинаково распределённые случайные величины с неизвестной функцией распределения F. Логично, что вероятность выпадения каждого x_k (вероятность того, что наугад взятый из выборки x будет равен x_k) одинакова

$$P(x = x_k) = \frac{1}{n}$$

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 $x_1, \ldots, x_n$, называется функция

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

Теорема 1.1.2

Неизвестная функция распределения $F(x)$ может быть сколь угодно точно восстановлена по выборке достаточно большого объёма [1, стр. 25].

$$\mathbb{P}\left(F_n(x) \underset{n\to\infty}{\Longrightarrow} F(x)\right) = 1$$

Идея доказательства. Вспомним, чему равна эмпирическая функция распределения

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

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

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

Так как эмпирическая функция распределения является средним арифметическим индикаторов, то по усиленному закону больших чисел она сходится к неизвестной функции распределения почти наверное при устремлении длины выборки к бесконечности

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

Теорема доказана

$$F_n(x) \overset{a.s.}{\underset{n\to\infty}{\rightrightarrows}} F(x)$$

□

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

$$F(b) - F(a) = \int\limits_{a}^{b} p(x)\,dx$$

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

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

Делим обе части на Δ_x.

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

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

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

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

$$p\left(x\right) \approx \frac{F\left(x+\Delta\right) - F\left(x\right)}{\Delta}$$

Возьмём m полуинтервалов на числовой прямой $I_j = \left(a_{j-1}, a_j\right], i = \overline{1, m}$ таких, что все значения выборки попадают в один из них. 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

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

Введём функцию $q\left(y\right)$

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

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

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

Отметим, что q_n сходится к q почти наверное (согласно закону больших чисел), а q в свою очередь сходится к p (согласно центральной предельной теореме)

$$q_n\left(y\right) \xrightarrow[n\to\infty]{a.s.} q\left(y\right) \xrightarrow[m\to\infty]{} p\left(y\right)$$

Функция q_n называется **гистограммой**.

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

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

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

$$F_n\left(a_j\right) - F_n\left(a_{j-1}\right) =$$

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

Сгруппируем слагаемые и получим чуть более компактную запись разности

$$F_n\left(a_j\right) - F_n\left(a_{j-1}\right) = \\ = \frac{1}{n} \cdot \sum_{k=1}^{n} \left[\mathbb{1}(x_k \le a_j) - \mathbb{1}(x_k \le a_{j-1})\right] \tag{1.2}$$

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 $a_{j-1} \le a_j$, то можно обойтись тем, что $x \le a_{j-1}$

$$\begin{cases} \mathbb{1}(x_k \le a_j) = 1 \\ \mathbb{1}(x_k \le a_{j-1}) = 1 \\ a_{j-1} \le a_j \end{cases} \Rightarrow \begin{cases} x_k \le a_j \\ x_k \le a_{j-1} \\ a_{j-1} \le a_j \end{cases} \\ \Rightarrow x_k \le a_{j-1} \le a_j \Rightarrow x_k \le a_{j-1}$$

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 $(a_j < x_k \le a_{j-1})$ означало бы то, что $a_j < a_{j-1}$

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

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

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

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

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

Вспомним формулу (1.2)

$$F_n\left(a_j\right) - F_n\left(a_{j-1}\right) = \\ = \frac{1}{n} \cdot \sum_{k=1}^{n} \left[\mathbb{1}(x_k \le a_j) - \mathbb{1}(x_k \le a_{j-1})\right]$$

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

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

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

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

Упростим, введя функцию $\nu_j(X)$ [1, стр. 68], которая считает количество элементов выборки $X = x_1, \ldots, x_n$, 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

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

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

$$q_n^k = \frac{\nu_k(X)}{n \cdot |I_k|} \tag{1.4}$$

Что мы тут видим? Теперь k — номер "столбика" гистограммы (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). Кроме того, происходит деление на общее количество элементов. Деление нужно, чтобы $q(y)$ сходилось к $p(y)$.

Делителю же $|I_k|$ отведена особая роль — он предотвращает искажение гистограммы при различных длинах отрезков. 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 [1, стр. 24]), является площадью прямоугольника

$$S_k = \frac{\nu_k(X)}{n} = \mathbb{P}_n(x \in I_k)$$

Введём замену в формуле (1.4) и умножим обе части на длину отрезка

$$\mathbb{P}_n(x \in I_k) = q_n^k \cdot |I_k|$$

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

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

Очень напоминает ситуацию с плотностью распределения непрерывной случайной величины ξ

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

Нужно отметить, что количество элементов выборки должно стремиться к бесконечности ($n \to \infty$), 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 $x_1, \ldots, x_n$ — выборка из распределения F_θ, где θ — неизвестный параметр из множества Θ

Пример 1.1.3

Имеем нормальное распределение с известным СКО $\sigma = 1$ и неизвестным математическим ожиданием a — $N(a, 1)$. Тогда θ — математическое ожидание a

Пример 1.1.4

Есть нормальное распределение, в котором неизвестны оба параметра. Тогда θ будет парой (a, σ)

Главный вопрос — определение основных параметров распределения выборки.

Определение 1.1.5: Статистика

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

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

Определение 1.1.6: Оценка

Статистику, значение которой заменяет неизвестный параметр, называют оценкой

Пример 1.1.7

Предположим, что выборка сделана 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 неизвестный параметр — величина p (вероятность удачного эксперимента)

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

Введём разные оценки $\hat{p}$

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

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

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

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 $\mathrm{M}(\hat{p} - p)^2$ было самое маленькое из возможных.

2. Также логично желать того, чтобы оценка $\hat{p}$ сходилась к истинному значению параметра p по вероятности ($\hat{p} \xrightarrow[n\to\infty]{\mathbb{P}} p$) или почти всюду ($\hat{p} \xrightarrow[n\to\infty]{a.s.} p$)

3. При многократном повторении эксперимента даже самая (на первый взгляд) плохая оценка может оказаться полезной

$$\mathrm{M}\hat{p}_1 = p$$
$$\mathrm{M}\hat{p}_2 = p$$
$$\mathrm{M}\hat{p}_3 = p$$

   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 1.1.8: Состоятельная оценка

Оценка $\hat{\theta}$ называется состоятельной, если стремится к истинному значению θ по вероятности
$$\hat{\theta} \xrightarrow[n\to\infty]{\mathbb{P}} \theta$$

Определение 1.1.9: Сильно состоятельная оценка

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

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

Пример 1.1.10

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

Определение 1.1.11: Несмещённая оценка

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

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

Замечание 1.1.12

Несмещённая оценка существует не всегда

Определение 1.1.13

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

$$\mathrm{D}_\theta \hat{\theta} \leq \mathrm{D}_\theta \tilde{\theta}, \qquad \forall \theta \in \Theta$$

или же

$$\mathrm{M}_\theta \left(\hat{\theta} - \theta\right)^2 \leq \mathrm{M}_\theta \left(\tilde{\theta} - \theta\right)^2, \qquad \forall \theta \in \Theta$$

Замечание 1.1.14

В учебнике Боровкова А. А. “Математическая статистика” оценка, удовлетворяющая этим условиям, носит название **эффективная оценка** [1, стр. 130], но у нас этот термин будет использоваться далее в другом смысле

Пример 1.1.15

Сравним $\hat{p}_1$ и $\hat{p}_3$

$$\mathrm{D}_p \hat{p}_1 = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p) = \frac{p \cdot (1-p)}{n}$$
$$\mathrm{D}_p \hat{p}_3 = \frac{2 \cdot p \cdot (1-p)}{n}$$

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 $N\left(a, \sigma^2\right)$. 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 $Exp(\lambda)$. $\dfrac{1}{\lambda}$ — среднее, $\dfrac{1}{\lambda^2}$ — дисперсия

И так далее…

Как правило, неизвестный параметр θ можно искать следующим образом

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

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

$$g\left(\hat{\theta}\right) = \int\limits_{\mathbb{R}} \varphi\left(x\right) dF_n\left(x\right) \tag{1.5}$$

Пример 1.1.16

Если θ — среднее, то $\varphi\left(x\right) = x$

$$\int\limits_{-\infty}^{+\infty} x dF_\theta\left(x\right) = \theta = g\left(\theta\right)$$

Теорема 1.1.17

Пусть функция $\varphi\left(x\right)$ в (1.5) непрерывна, ограничена и строго монотонная. Тогда оценка $\hat{\theta}$ существует и является сильно состоятельной.

Доказательство. Имеем формулу (1.5)

$$g\left(\hat{\theta}\right) = \int\limits_{\mathbb{R}} \varphi\left(x\right) dF_n\left(x\right)$$

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

Применим обратную функцию к обеим частям уравнения

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

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

$$\int\limits_{\mathbb{R}} \varphi\left(x\right) dF_n\left(x\right) \xrightarrow[n\to\infty]{a.s.} \int\limits_{\mathbb{R}} \varphi\left(x\right) dF_n\left(x\right)$$

Функция $g^{-1}\left(x\right)$ непрерывна

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

Теорема доказана

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

□

Определение 1.1.18: Выборочное среднее

Выборочное средние обозначается через $\overline{x}$ и считается по следующей формуле

$$\overline{x} = \int\limits_{\mathbb{R}} x dF_n(x)$$

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

$$\overline{x} = \int\limits_{\mathbb{R}} x dF_n(x) = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k$$

Определение 1.1.19: Выборочная дисперсия

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

$$\overline{\sigma^2} = \int\limits_{\mathbb{R}} (x - \overline{x})^2 dF_n(x) = \frac{1}{n} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2$$

1.2 Свойства оценок

1.2.1 Неравенство Рао-Крамера

Теорема 1.2.1: Колмогорова

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

$$\begin{cases} \mathrm{D}_\theta \theta_1 \le \mathrm{D}_\theta \hat{\theta} \\ \mathrm{D}_\theta \theta_2 \le \mathrm{D}_\theta \hat{\theta} \end{cases}, \forall \theta \in \Theta$$

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 $\hat{\theta}$

$$\begin{cases} \mathrm{D}_\theta \theta_1 \le \mathrm{D}_\theta \theta_2 \\ \mathrm{D}_\theta \theta_2 \le \mathrm{D}_\theta \theta_1 \end{cases}, \forall \theta \in \Theta$$

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

$$\mathrm{D}_\theta \theta_1 = \mathrm{D}_\theta \theta_2 = \sigma^2(\theta)$$

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

$$\mathrm{D}_\theta \tilde{\theta} \geq \sigma^2(\theta) \tag{1.6}$$

Попробуем честно вычислить дисперсию оценки $\tilde{\theta}$

$$\mathrm{D}_\theta \tilde{\theta} = \mathrm{M}_\theta \left(\tilde{\theta} - \theta\right) = \mathrm{M}_\theta \left[\frac{1}{2} \cdot (\theta_1 - \theta) + \frac{1}{2} \cdot (\theta_2 - \theta)\right]^2 =$$
$$= \frac{1}{4} \cdot \mathrm{D}_\theta \theta_1 + \frac{1}{4} \cdot \mathrm{D}_\theta \theta_1 + \frac{1}{2} \cdot \mathrm{M}_\theta \left[(\theta_1 - \theta) \cdot (\theta_2 - \theta)\right]$$

Воспользуемся неравенством Коши (частный случай неравенства Гёльдера)

$$\begin{aligned} \mathrm{M}_\theta \left[(\theta_1 - \theta) \cdot (\theta_2 - \theta)\right] \leq \sqrt{\mathrm{M}_\theta (\theta_1 - \theta)^2 \cdot \mathrm{M}_\theta (\theta_2 - \theta)^2} = \\ = \sqrt{\mathrm{D}_\theta \theta_1 \cdot \mathrm{D}_\theta \theta_2} = \sqrt{\sigma_1^2 \cdot \sigma_2^2} \end{aligned} \tag{1.7}$$

И вернёмся к вычислению дисперсии оценки $\tilde{\theta}$

$$\frac{1}{4} \cdot \mathrm{D}_\theta \theta_1 + \frac{1}{4} \cdot \mathrm{D}_\theta \theta_1 + \frac{1}{2} \cdot \mathrm{M}_\theta \left[(\theta_1 - \theta) \cdot (\theta_2 - \theta)\right] \leq$$
$$\leq \frac{1}{2} \cdot \sigma^2(\theta) + \frac{1}{2} \cdot \sqrt{\sigma^2(\theta) \cdot \sigma^2(\theta)} = \sigma^2(\theta)$$

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

$$\mathrm{D}_\theta \tilde{\theta} \leq \sigma^2(\theta) \tag{1.8}$$

Воспользовавшись неравенствами (1.6) и (1.8), получаем равенство

$$\mathrm{D}_\theta \tilde{\theta} = \sigma^2(\theta)$$

Это значит, что в неравенстве (1.7) в данном случае тоже выходит равенство

$$\mathrm{M}_\theta \left[(\theta_1 - \theta) \cdot (\theta_2 - \theta)\right] = \sqrt{\mathrm{M}_\theta (\theta_1 - \theta)^2} \cdot \sqrt{\mathrm{M}_\theta (\theta_2 - \theta)^2}$$

Для дальнейших размышлений вспомним аналогию с векторами, а именно смысл равенства в неравенстве Коши для скалярного произведения векторов

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

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

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

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

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

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

$$\begin{cases} \left(\theta_1 - \widehat{\theta, \theta_2} - \theta\right) = 0 \\ \mathrm{M}_\theta \left(\theta_1 - \theta\right)^2 = \mathrm{M}_\theta \left(\theta_2 - \theta\right)^2 \end{cases} \Rightarrow \theta_1 - \theta = \theta_2 - \theta$$

$$\Rightarrow \theta_1 = \theta_2$$

Теорема доказана □

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.

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

Определение 1.2.2: 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 незасимых одинаково распределённых случайных величин. Воспользовавшись законом больших чисел, можем сказать, что она стремится к сумме n одинаковых математических ожиданий при достаточно большом размере выборки

$$\ln L(\vec{x}, \theta) = n \cdot \frac{\ln p(x_1, \theta) + \cdots + \ln p(x_n, \theta)}{n} \approx$$
$$\approx n \cdot \mathrm{M}_\theta \ln p(x_1, \theta)$$

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

Определение 1.2.3: Вклад выборки

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

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

Замечание 1.2.4

Математическое ожидание вклада выборки равно нулю

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

Доказательство. Посчитаем математическое ожидание вклада выборки

$$\mathrm{M}_\theta U(\vec{x}, \theta) = \int\limits_{\mathbb{R}^n} U(\vec{u}, \theta) \cdot L(\vec{u}, \theta)\, d\vec{u} =$$
$$= \int\limits_{\mathbb{R}^n} \frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \cdot L(\vec{u}, \theta)\, d\vec{u} =$$
$$= \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\vec{u}, \theta)\, d\vec{u}$$

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

$$\mathrm{M}_\theta U(\vec{x}, \theta) = \frac{\partial}{\partial \theta} \int\limits_{\mathbb{R}^n} L(\vec{u}, \theta)\, d\vec{u}$$

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

$$\mathrm{M}_\theta U\left(\vec{x},\theta\right) = \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} L\left(\vec{u},\theta\right)\,d\vec{u} = \frac{\partial}{\partial\theta}1 = 0$$

□

Замечание 1.2.5

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

Доказательство. Выше у нас было равенство

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

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

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

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

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

□

Определение 1.2.6: Количество информации Фишера

Математическое ожидание квадрата вклада выборки называется количеством информации Фишера

$$I_n\left(\theta\right) = \mathrm{M}_\theta U\left(\vec{x},\theta\right)^2$$

Замечание 1.2.7

$$\mathrm{M}_\theta U\left(\vec{x},\theta\right)^2 = -\mathrm{M}_\theta\frac{\partial^2}{\partial\theta^2}\ln L\left(\vec{x},\theta\right)$$

Доказательство. Будем доказывать справа налево

$$-\mathrm{M}_\theta \frac{\partial^2}{\partial\theta^2} \ln L(\vec{x},\theta) = -\mathrm{M}_\theta \frac{\partial}{\partial\theta} \frac{\frac{\partial}{\partial\theta} L(\vec{x},\theta)}{L(\vec{x},\theta)} =$$

$$= -\mathrm{M}_\theta \left(\frac{\frac{\partial^2}{\partial\theta^2} L(\vec{x},\theta) \cdot L(\vec{x},\theta) - \left[\frac{\partial}{\partial\theta} L(\vec{x},\theta)\right]^2}{L(\vec{x},\theta)^2} \right) =$$

$$= -\mathrm{M}_\theta \frac{\frac{\partial^2}{\partial\theta^2} L(\vec{x},\theta)}{L(\vec{x},\theta)} + \mathrm{M}_\theta \left[\frac{\frac{\partial}{\partial\theta} L(\vec{x},\theta)}{L(\vec{x},\theta)} \right]^2$$

Помним, что производная от функции правдоподобия по θ равна нулю. Значит вторая производная тоже равна нулю и остаётся лишь математическое ожидание квадрата, который равен квадрату производной логарифма функции правдоподобия, что в свою очередь и есть вклад выборки

$$\frac{\partial}{\partial\theta} L(\vec{u},\theta) = 0 \Rightarrow -\mathrm{M}_\theta \frac{\frac{\partial^2}{\partial\theta^2} L(\vec{x},\theta)}{L(\vec{x},\theta)} = 0$$

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

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

Утверждение доказано

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

□

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

Теорема 1.2.8: Неравенство Рао-Крамера

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

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

Доказательство. Выпишем, чему равно математическое ожидание оценки θ

$$\begin{cases} \mathrm{M}_\theta \hat{\theta} & = \theta \\ \mathrm{M}_\theta \hat{\theta} & = \int\limits_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot L(\vec{u},\theta)\, d\vec{u} \end{cases}$$

$$\Rightarrow \theta = \int\limits_{\mathbb{R}^n} \hat{\theta}(\vec{u}) \cdot L(\vec{u},\theta)\, d\vec{u}$$

Продифференцируем с двух сторон полученное для θ равенство по самому параметру θ

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

Левая часть равенства превращается в единицу, а справа заносим взятие производной под знак интеграла. Также помним, что оценка $\theta\left(\vec{u}\right)$ не зависит от параметра θ. Это значит, что производную нужно брать только от функции правдоподобия

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

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

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

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

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

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

$$1 = \mathrm{M}_\theta\left(\hat{\theta}\cdot U\left(\vec{x},\theta\right)\right) \tag{1.9}$$

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

$$\mathrm{M}_\theta U\left(\vec{x},\theta\right) = 0$$
$$\Rightarrow \theta\cdot\mathrm{M}_\theta U\left(\vec{x},\theta\right) = \mathrm{M}_\theta\left(\theta\cdot U\left(\vec{x},\theta\right)\right) = 0$$

Воспользовавшись полученным результатом, вернёмся к равенству (1.9). Отнимем от обеих частей ноль (то есть, полученное только что выражение)

$$1 = \mathrm{M}_\theta\left(\hat{\theta}\cdot U\left(\vec{x},\theta\right)\right) - \mathrm{M}_\theta\left(\theta\cdot U\left(\vec{x},\theta\right)\right)$$

Получаем компактное равенство

$$1 = \mathrm{M}_\theta\left[\left(\hat{\theta}-\theta\right)\cdot U\left(\vec{x},\theta\right)\right]$$

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

$$
\begin{aligned}
1 &= \mathrm{M}_\theta \left[\left(\hat{\theta} - \theta\right) \cdot U\left(\vec{x}, \theta\right)\right] \leq \\
&\leq \sqrt{\mathrm{M}_\theta \left(\hat{\theta} - \theta\right)} \cdot \sqrt{\mathrm{M}_\theta U\left(\vec{x}, \theta\right)} = \\
&= \sqrt{\mathrm{D}_\theta \hat{\theta}} \cdot \sqrt{I_n\left(\theta\right)}
\end{aligned} \tag{1.10}
$$

Возводим обе части равенства в квадрат и делим на количество информации

$$\mathrm{D}_\theta \hat{\theta} \geq \frac{1}{I_n\left(\theta\right)}$$

Неравенство доказано □

Замечание 1.2.9

Иногда нужно оценивать не сам параметр, а функцию параметра

Если α — несмещённая оценка для $f\left(\theta\right)$, то справедливо следующее неравенство

$$\forall \theta \in \Theta : \mathrm{D}_\theta \alpha \geq \frac{\left|f'\left(\theta\right)\right|}{I_n\left(\theta\right)}$$

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 1.2.10: Эффективная оценка

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

$$\forall \theta \in \Theta : \mathrm{D}_\theta \hat{\theta} = \frac{1}{I_n\left(\theta\right)}$$

Попытаемся выяснить, 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)

$$\mathrm{D}_\theta \hat{\theta} = \frac{1}{I_n\left(\theta\right)}$$

Рассмотрим неравенство (1.10) и попытаемся понять, в каком случае в нём будет стоять знак равенства

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

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 и эти функции являются линейными комбинациями друг друга. Значит, есть такая функция $k(\theta)$, что $f_2(\theta)$ равняется произведению $f_1(\theta)$ и $k(\theta)$.

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

Проинтегрируем обе части равенства

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

Получим следующее равенство

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

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

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

Избавимся от логарифма слева, а для этого проэкспонируем обе части равенства

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

При конечном n положим такую плотность распределения

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

В таком случае получим следующую функцию правдоподобия

$$L\left(\vec{x}, \theta\right) = \prod_{k=1}^{n} p\left(x_1, \theta\right) =$$
$$= \exp\left\{ \sum_{k=1}^{n} \hat{\theta}\left(x_k\right) \cdot a_1\left(\theta\right) + n \cdot b_1\left(\theta\right) + \sum_{k=1}^{n} c_1\left(x_k\right) \right\}$$

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 1.2.11: Экспоненциальное распределение

Распределения следующего вида называются экспоненциальными

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

Попробуем найти рецепт выяснения эффективной оценки. Начнём с примера

Пример 1.2.12

Есть выборка $x_1, x_2, \ldots, x_n$ 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

$$\sum_{k=1}^{n} x_k \cdot \theta = \left(\frac{1}{n} \cdot \sum_{k=1}^{n} x_k\right) \cdot \theta \cdot n = \overline{x} \cdot \theta \cdot n$$

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

Сгруппировав множители при n, получаем

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

Добавим и вычтем в числителе дроби выборочное среднее

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

Теперь в числителе очевиден квадрат разности

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

Сворачиваем квадрат разности, а выборочное среднее заносим под знак суммы

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

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

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

Оказывается, именно так она и находится.

Определение 1.2.13: Оценка максимального правдоподобия

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

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

Замечание 1.2.14

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

Определение 1.2.15: Уравнение правдоподобия

Уравнением правдоподобия называется равенство вида

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

Или же

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

Замечание 1.2.16

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 1.2.17: Вариационный ряд

Вариационный ряд выборки $x_1, x_2, \ldots, x_n$ — значения выборки, упорядоченные в порядке неубывания

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}, \qquad x_{(1)} = \min_k x_k$$

Теорема 1.2.18

Если плотность $p\left(x, \theta\right)$ непрерывна и дифференцируема по параметру θ, а производная не равна нулю $\dfrac{\partial}{\partial \theta} p\left(x, \theta\right) \neq 0$, 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 $x_1, x_2, \ldots, x_n$

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

Лемма 2.1.2

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

$$\mathrm{M}_\theta\hat{\theta} = \theta \Rightarrow \begin{cases} \mathrm{M}_\theta\Lambda\hat{\theta} = \mathrm{M}_\theta\hat{\theta} = \theta \\ \mathrm{D}_\theta\Lambda\hat{\theta} \le \mathrm{D}_\theta\hat{\theta} \end{cases}$$

Доказательство. Берём $x_1, x_2, \ldots, x_n$ — 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 $\vec{x}$

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

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

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

Тогда и оценки примут более красивый вид

$$\hat{\theta}\left(x_1, x_2, \ldots, x_n\right) = \hat{\theta}\left(\vec{x}\right)$$
$$\hat{\theta}\left(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\right) = \hat{\theta}\left(\vec{x}_\sigma\right)$$

Теперь приступим непосредственно к доказательству.

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

   Нетрудно показать, что вектора $\vec{x}$ и $\vec{x}_\sigma$ имеют одинаковое распределение для любой перестановки σ, а это значит, что и оценки $\hat{\theta}\left(\vec{x}\right)$ и $\hat{\theta}\left(\vec{x}_\sigma\right)$ распределены одинаково как функции случайных одинаково распределённых векторов. Следовательно, их математические ожидания равны между собой при любой перестановке σ

   $$\mathrm{M}_\theta \hat{\theta}\left(\vec{x}\right) = \mathrm{M}_\theta \hat{\theta}\left(\vec{x}_\sigma\right) = \theta$$

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

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

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

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

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

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

   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

   $$\mathrm{M}_\theta \Lambda \hat{\theta} = \theta$$

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

Воспользуемся определением

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

Внесём параметр θ в сумму. Для этого нужно умножить и поделить его на $n!$ (так как сумма имеет $n!$ слагаемых)

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

Вспомним неравенство Йенсена для выпуклой функции f

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

В нашем случае $x_i = \left(\hat{\theta}\left(\vec{x}_{\sigma_i}\right) - \theta\right)$, функция $f\left(x\right) = x^2$, 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

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

Воспользуемся линейностью математического ожидания, внеся его под знак суммы

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

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

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

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

$$\mathrm{D}_\theta\Lambda\hat{\theta}\leq\mathrm{D}_\theta\hat{\theta}\left(\vec{x}\right)$$

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

Замечание 2.1.3

Равенство в неравенстве Йенсена (в доказательстве 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 $\xi:\Omega\to\mathbb{R}$ такая, что связанные с ней множества измеримы по Лебегу

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

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

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

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

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

Определение 2.2.1: Сигма-алгебра, порождённая случайной величиной

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

$$\mathfrak{F}_\xi = \left\{\xi^{-1}(\Delta) \,\middle|\, \Delta \in \mathfrak{B}\right\}$$

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 $\mathfrak{F}_\xi$ входят в σ-алгебру $\mathfrak{F}$, а сама $\mathfrak{F}_\xi$ является подмножеством $\mathfrak{F}$

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

Проверим, что $\mathfrak{F}_\xi$ действительно является σ-алгеброй

1. Множество элементарных исходов Ω входит в $\mathfrak{F}_\xi$. 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 $\mathfrak{F}_\xi$

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

2. Если событие A принадлежит $\mathfrak{F}_\xi$, то его дополнение $\overline{A}$ тоже принадлежит $\mathfrak{F}_\xi$

$$\begin{aligned} A &= \xi^{-1}(\Delta) = \{\omega \mid \xi(\omega) \in \Delta\} \\ \Rightarrow \overline{A} &= \{\omega \mid \xi(\omega) \notin \Delta\} = \left\{\omega \,\middle|\, \xi(\omega) \in \overline{\Delta}\right\} \\ \overline{A} &= \xi^{-1}\left(\overline{\Delta}\right) \end{aligned}$$

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

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

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 $\mathfrak{F}_\xi$ не будет различать элементы ω_1 и ω_2. Это в свою очередь означает, что можно разбить $\mathfrak{F}_\xi$ 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 $a_1, a_2, \ldots, a_n$

$$\xi : \Omega \to \{a_1, a_2, \ldots, a_n\}$$

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 σ-алгебра $\sigma(\xi)$ содержит 2^n элементов

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

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

Возьмём $\varkappa$ — случайная величина, измеримая относительно $\sigma(\xi)$. Это значит, что все прообразы случайной величины $\varkappa$ должны лежать в σ-алгебре $\sigma(\xi)$

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

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 $c \to -\infty$ прообразом является пустое множество, а когда $c \to +\infty$, то прообразом является всё множество элементарных исходов

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

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 неделимы, 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 $(m \leq n)$ $c_1 < c_2 < \cdots < c_m$ на числовой прямой $\mathbb{R}$ как значения случайной величины $\varkappa$

$$\varkappa : \Omega \to \{c_1, c_2, \ldots, c_m\}$$

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 $A\left(c_1-0\right)=\{\omega \mid \varkappa(\omega)<c_1\}$ пустое. Получаем то, что хотели

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

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

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

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

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

В таком случае тождество очевидно

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

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

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

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

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

Запишем это в более удобном виде

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

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

Для любого t разность множеств $A\left(c_t\right) \setminus A\left(c_{t-1}\right) \neq \emptyset$ (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

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

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

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

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

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

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 $d_1, d_2, \ldots, d_n$ вместо m чисел $c_1, c_2, \ldots, c_m$.

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

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

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

$$f : \{a_1, a_2, \ldots, a_n\} \to \{d_1, d_2, \ldots, d_n\}$$

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)

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

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

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

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

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

Отсюда делаем следующий вывод

Утверждение 2.3.1

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 $e_1, e_2, \ldots, e_n$ — ортонормированный базис в L, тогда y можно найти по формуле

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

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, $e_2, \ldots, e_n$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Начнём с определения математического ожидания

$$\begin{aligned}\mathrm{M}\left[\eta \cdot \mathbb{1}_{H_k}\right] &= \int\limits_{\Omega} \eta(\omega) \cdot \mathbb{1}_{H_k} \mathbb{P}(d\omega) = \\ &= \int\limits_{H_k} \eta(\omega)\, \mathbb{P}(d\omega) + \int\limits_{\Omega \setminus H_k} 0\, \mathbb{P}(d\omega)\end{aligned} \tag{2.6}$$

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 $\mathbb{P}_k(A) = \dfrac{\mathbb{P}(A)}{\mathbb{P}(H_k)}$, 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

$$\mathrm{M}\left[\eta \cdot \mathbb{1}_{H_k}\right] = \mathbb{P}(H_k) \cdot \int\limits_{H_k} \eta(\omega)\, \mathbb{P}(d\omega \mid H_k)$$

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 $\mathbb{P}(H_k)$

$$\frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{H_k}\right]}{\mathbb{P}(H_k)} = \int\limits_{H_k} \eta(\omega)\, \mathbb{P}(d\omega \mid H_k)$$

Определение 2.4.1: Условное математическое ожидание случайной величины относительно случайного события

Условное математическое ожидание случайной величины ξ относительно **события** A [2, стр. 68] обозначается $\mathrm{M}[\xi \mid A]$ и считается по формуле

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

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

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

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 сигма-алгебры, порождённой случайной величиной, принимающей конечное количество значений

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

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

Замечание 2.4.3

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

Например, последнее определение может выглядеть немного странно

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

Зато при более детальном рассмотрении из самой записи очевиден её смысл: условное математическое ожидание относительно σ-алгебры — вектор, для получения которого нужно умножить проекции на базисные векторы. Ведь $\mathrm{M}\left[\eta \mid H_k\right]$ — ни что иное, как проекция вектора (случайной величины) η на ось (уровень) H_k, также 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 $\mathfrak{F}_\xi$, порождённую случайной величиной ξ, принимающей конечное количество значений, выполняется равенство скалярных произведений

$$\forall A \in \mathfrak{F}_\xi : \mathrm{M}\left[\tilde{\eta} \cdot \mathbb{1}_A\right] = \mathrm{M}\left[\eta \cdot \mathbb{1}_A\right] \tag{2.7}$$

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

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

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

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

Помним, что математическое ожидание индикатора — вероятность

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

Замечаем условную вероятность

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

Поскольку A принадлежит множеству случайных событий $\mathfrak{F}_\xi$, то условная вероятность $\mathbb{P}\left(A \mid H_k\right)$ 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

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

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

Далее мы имеем полное математическое и моральное право вынести η за знак суммы. Если с математикой всё очевидно (работает закон дистрибутивности), то напомню о морально-этической стороне дела: нам нужно, пройтись по всем возможным индикаторам $\mathbb{1}_{H_k}$, из которых лишь один сработает (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)

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

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)

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

Значит, равенство (2.7) выполняется. □

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)

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

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

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

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

Замечание 2.4.7

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

То есть, имеем три эквивалентных записи

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

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

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

Очевидно, что числа p_{ij} обладают некоторым свойствам

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

Посчитаем условное математическое ожидание согласно формуле из определения (2.4.2)

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

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

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

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

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

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

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

Вспомним альтернативные записи прообраза

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

И перепишем формулу (2.8)

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

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

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

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

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

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

$$\mathrm{M}\left[\eta \mid \xi\right] = \varphi^\eta\left(\xi\right)$$

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

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

Тогда будет корректна следующая запись

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

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

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

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

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

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.

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

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

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

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

$$\int\limits_{\Omega} \varphi^{\eta}(\xi) \cdot \mathbb{1}_A \, d\mathbb{P} = \int\limits_{A} \varphi^{\eta}(\xi(\omega)) \, \mathbb{P}(d\omega) \tag{2.9}$$

*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

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[†]. 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 $\varphi^\eta(x)$

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

А теперь распишем условное математическое ожидание

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

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)

$$\int\limits_A \varphi^\eta(\xi(\omega))\,\mathbb{P}(d\omega) = \int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{\{\xi = x\}}\right]}{\mathbb{P}\{\xi = x\}}\,\mathbb{P}(d\omega) \tag{2.10}$$

Дальше происходит магия, которую можно трактовать по-разному

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

$$\int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{\{\xi = x\}}\right]}{\mathbb{P}\{\xi = x\}}\,\mathbb{P}(d\omega) = \int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{d\omega}\right]}{\mathbb{P}(d\omega)}\,\mathbb{P}(d\omega)$$

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

$$\int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{d\omega}\right]}{\mathbb{P}(d\omega)}\,\mathbb{P}(d\omega) = \int\limits_A \int\limits_\Omega \eta \cdot \mathbb{1}_{d\omega} \cdot \mathbb{P}(d\tilde{\omega})$$

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

$$\int\limits_A \int\limits_\Omega \eta \cdot \mathbb{1}_{d\omega} \cdot \mathbb{P}(d\tilde{\omega}) = \int\limits_A \int\limits_{d\omega} \eta(\tilde{\omega})\,\mathbb{P}(d\tilde{\omega})$$

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

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

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 $d\omega$

$$\int\limits_{d\omega} \eta\left(\tilde{\omega}\right) \mathbb{P}\left(d\tilde{\omega}\right) = \eta\left(\omega\right) \cdot \mathbb{P}\left(d\omega\right)$$

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

$$\int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{\{\xi = x\}}\right]}{\mathbb{P}\left\{\xi = x\right\}} \mathbb{P}\left(d\omega\right) = \int\limits_A \mathrm{M}\left[\eta \mid d\omega\right] \mathbb{P}\left(d\omega\right)$$

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 $d\omega$[‡] (если значение на промежутке $d\omega$ — константа, то очевидно, что среднее значение будет равно ей же).

$$\int\limits_A \mathrm{M}\left[\eta \mid d\omega\right] \mathbb{P}\left(d\omega\right) = \int\limits_A \eta\, \mathbb{P}\left(d\omega\right)$$

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

$$\int\limits_A \frac{\mathrm{M}\left[\eta \cdot \mathbb{1}_{\{\xi = x\}}\right]}{\mathbb{P}\left\{\xi = x\right\}} \mathbb{P}\left(d\omega\right) = \int\limits_A \eta\, \mathbb{P}\left(d\omega\right)$$

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

$$\int\limits_A \eta\, \mathbb{P}\left(d\omega\right) = \int\limits_\Omega \eta \cdot \mathbb{1}_A \cdot \mathbb{P}\left(d\omega\right) = \mathrm{M}\left[\eta \cdot \mathbb{1}_A\right]$$

$\square$

Теперь вернёмся к менее абстрактным вещам и посмотрим, как выглядит условное математическое ожидание, когда случайные величины ξ и η имеют совместную плотность распределения

$$\mathbb{P}\left\{(\xi, \eta) \in \Delta\right\} = \iint\limits_\Delta p\left(x, y\right)\, dx\, dy$$

[‡]Нам достаточно постоянства значения $\xi\left(\omega\right)$ **почти всюду** на событии $d\omega$, так как интеграл Лебега простой функции (функции, что принимает конечное число значений [3, стр. 53]) — сумма значений функции, умноженных на меры соответствующих им прообразов [3, 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 $\xi\left(\omega\right)$ равна одному и тому же значению почти всюду на ω)

В таком случае компонента ξ имеет плотность r

$$r\,(x) = \int\limits_{\mathbb{R}} p\,(x, y)\; dy$$

Компонента η имеет плотность q

$$q\,(y) = \int\limits_{\mathbb{R}} p\,(x, y)\; dx$$

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

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

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 $\xi = x$

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

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

$$\varphi^{\eta}\,(x) = \frac{\int\limits_{\mathbb{R}} y \cdot p\,(x, y)\; dy}{r\,(x)} = \frac{\int\limits_{\mathbb{R}} y \cdot p\,(x, y)\; dy}{\int\limits_{\mathbb{R}} p\,(x, y)\; dy} \tag{2.11}$$

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

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

$$\varphi^{\eta}\,(\xi) = \left.\frac{\int\limits_{\mathbb{R}} y \cdot p\,(x, y)\; dy}{\int\limits_{\mathbb{R}} p\,(x, y)\; dy}\right|_{x=\xi}$$

Является условным математическим ожиданием $\mathrm{M}\left[\eta \mid \xi\right]$

Доказательство. Первое свойство снова очевидно, поэтому надо доказать

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

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 $\sigma\left(\xi\right)$, то оно представимо в виде $\xi^{-1}\left(\Delta\right), \Delta \in \mathfrak{B}$. 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

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

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 $p\left(x, y\right)$ появляется $r\left(x\right)$. Также учтём индикатор и сузим область интегрирования с $\mathbb{R}$ до Δ

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

Дальше распишем функцию φ^{η}, пользуясь формулой (2.11)

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

Сократим одинаковые плотности и получим интересный двойной интеграл

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

Вернём индикатор обратно в интеграл

$$\int\limits_{\Delta} \int\limits_{\mathbb{R}} y \cdot p\left(x, y\right) \cdot\, dy\, dx = \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} y \cdot \mathbb{1}(x \in \Delta) \cdot p\left(x, y\right)\, dy\, dx$$

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

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

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

$$\varphi^{\eta}\left(x\right) = \frac{\int\limits_{\mathbb{R}} y \cdot p\left(x, y\right)\, dy}{\int\limits_{\mathbb{R}} p\left(x, y\right)\, dy}$$

По формуле
$$\mathrm{M}\left[\eta \mid \xi\right] = \varphi^{\eta}\left(\xi\right) = \left.\varphi^{\eta}\left(x\right)\right|_{x=\xi}$$
□

Теорема 2.4.11: Существование условного математического ожидания

Условное математическое ожидание существует всегда и единственное почти наверное

Доказательство. [1, стр. 142] □

2.4.4 Свойства условного математического ожидания

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

I Формула полной вероятности [1, стр. 144]
$$\mathrm{M}\mathrm{M}\left[\eta \mid \mathfrak{F}_1\right] = \mathrm{M}\eta$$

II Условное математическое ожидание неотрицательной случайной величины неотрицательно почти наверное
$$\eta \geq 0 \Rightarrow \mathrm{M}\left[\eta \mid \mathfrak{F}_1\right] \geq 0$$

III Неравенство Йенсена. Если функция φ выпуклая вниз, то
$$\varphi\left(\mathrm{M}\left[\eta \mid \mathfrak{F}_1\right]\right) \leq \mathrm{M}\left[\varphi\left(\eta\right) \mid \mathfrak{F}_1\right]$$

IV Теорема о трёх перпендикулярах
$$\mathfrak{F}_2 \subset \mathfrak{F}_1 \Rightarrow \mathrm{M}\left[\mathrm{M}\left(\eta \mid \mathfrak{F}_1 \mid \mathfrak{F}_2\right)\right] = \mathrm{M}\left[\eta \mid \mathfrak{F}_2\right]$$

V Если случайная величина η измерима относительно σ-алгебры $\mathfrak{F}_1$, то её условное математическое ожидание равно ей самой
$$\mathrm{M}\left[\eta \mid \mathfrak{F}_1\right] = \eta$$

VI Если случайная величина η измерима относительно $\mathfrak{F}_1$, то для любой случайной величины ξ
$$\mathrm{M}\left[\eta \cdot \xi \mid \mathfrak{F}_1\right] = \eta \cdot \mathrm{M}\left[\xi \mid \mathfrak{F}_1\right]$$

VII Если η не зависит от $\mathfrak{F}_1$, то её условное математическое ожидание равно простому математическому ожиданию
$$\forall \Delta \in \mathfrak{B}, A \in \mathfrak{F}_1 : \mathbb{P}\left(\{\eta \in \Delta\} \mid A\right) = \{\eta \in \Delta\} \Rightarrow \mathrm{M}\left[\eta \mid \mathfrak{F}_1\right] = \mathrm{M}\eta$$

[§] Также со свойствами и их доказательствами можно ознакомиться в книгах Ширяева [4, стр. 270] и Боровкова [1, стр. 143]

VIII Условное математическое ожидание линейно

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

IX Сохраняется теорема Лебега о возможности предельного перехода под знаком условного математического ожидания [5, стр. 302]. В книге Ширяева это называется теоремой о сходимости под знаком условных ожиданий [4, стр. 272]

$$|\xi_n| \le \eta,\ \mathrm{M}\eta < \infty,\ \xi_n \xrightarrow[n\to\infty]{a.s.} \xi \Rightarrow \mathrm{M}\left[\xi_n \mid \mathfrak{F}_1\right] \xrightarrow[n\to\infty]{a.s.} \mathrm{M}\left[\xi \mid \mathfrak{F}_1\right]$$

Пара полезных частных случаев неравенства Йенсена (III свойство)

$$\begin{array}{lcl} \varphi(x) = |x| : & \left|\mathrm{M}\left[\eta \mid \mathfrak{F}_1\right]\right| & \le & \mathrm{M}\left[|\eta| \mid \mathfrak{F}_1\right] \\ \varphi(x) = x^2 : & \left(\mathrm{M}\left[\eta \mid \mathfrak{F}_1\right]\right)^2 & \le & \mathrm{M}\left[\eta^2 \,\middle|\, \mathfrak{F}_1\right] \end{array}$$

2.4.5 Условное математическое ожидание функции произвольной случайной величины

Чем вызвала интерес эта тема? Допустим, у нас есть $x_1, \ldots, x_n$ — 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 $\mathrm{M}\left[\hat{\theta} \,\middle|\, T\right]$.

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 $\vec{x}$.

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

Вспомним, что для поиска условного математического ожидания мы находили функцию $\varphi^{\eta}(x) = \mathrm{M}\left[\eta \mid \xi = x\right]$. 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

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

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

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

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

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

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 одинаковые)

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

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 $\left\|\vec{\nabla} \cdot g(\vec{u})\right\|$. Напрашивается формула

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

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

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

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

$$q\left(t\right) \cdot 2 \cdot \tilde{\varepsilon} \approx \int\limits_{G_{\tilde{\varepsilon}}} p\left(\vec{u}\right)\, d\vec{u} \Rightarrow q\left(t\right) \approx \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int\limits_{G_{\tilde{\varepsilon}}} p\left(\vec{u}\right)\, d\vec{u}$$

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

$$q\left(t\right) = \lim_{\tilde{\varepsilon} \to 0} \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int\limits_{G_{\tilde{\varepsilon}}} p\left(\vec{u}\right)\, d\vec{u} \tag{2.14}$$

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

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

Вернёмся к плотности в формуле (2.14). Заменив $\tilde{\varepsilon}$ на $\varepsilon \cdot \left\|\vec{\nabla} \cdot g\left(\vec{u}\right)\right\|$, 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)

$$q\left(t\right) = \lim_{\tilde{\varepsilon} \to 0} \frac{1}{2 \cdot \tilde{\varepsilon}} \cdot \int\limits_{G_{\tilde{\varepsilon}}} p\left(\vec{u}\right)\, d\vec{u} = \lim_{\varepsilon \to 0} \frac{1}{2 \cdot \varepsilon \cdot \left\|\vec{\nabla} \cdot g\left(\vec{u}\right)\right\|} \cdot \int\limits_{G_{\tilde{\varepsilon}}} p\left(\vec{u}\right)\, d\vec{u}$$

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 $G_{\tilde{\varepsilon}}$ вырождается в поверхность уровня S_t. Обозначив меру площади на поверхности S_t как $\sigma_t\left(d\vec{u}\right)$, получаем поверхностный интеграл первого рода

$$q\left(t\right) = \int\limits_{S_t} p\left(\vec{u}\right) \cdot \frac{1}{\left\|\vec{\nabla} \cdot g\left(\vec{u}\right)\right\|}\, \sigma_t\left(d\vec{u}\right) \tag{2.15}$$

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

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

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 $\dfrac{\sigma_t(d\vec{u})}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}$. 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

$$\frac{\int\limits_{\mathbb{R}} y\cdot p(x,y)\,dy}{r(x)} \to \frac{\int\limits_{S_t} f(\vec{u})\cdot p(\vec{u})\,\frac{\sigma_t(d\vec{u})}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}}{q(t)}$$

И конечная формула

$$\mathrm{M}\left[f(\vec{x})\mid g(\vec{x})=t\right]=\frac{\int\limits_{S_t} f(\vec{u})\cdot p(\vec{u})\cdot\frac{1}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}\,\sigma_t(d\vec{u})}{\int\limits_{S_t} p(\vec{u})\cdot\frac{1}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}\,\sigma_t(d\vec{u})}$$

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

$$\mathrm{M}\left[f(\vec{x})\mid g(\vec{x})=t\right]=\frac{\int\limits_{S_t} f(\vec{u})\cdot p(\vec{u})\cdot\frac{1}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}\,\sigma_t(d\vec{u})}{\int\limits_{S_t} p(\vec{u})\cdot\frac{1}{\left\|\vec{\nabla}\cdot g(\vec{u})\right\|}\,\sigma_t(d\vec{u})}$$

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 $g(\vec{x}) = x_{(n)}$

$$\mathrm{M}\left[x_1 \mid x_{(n)}\right] = ?$$

Поверхности уровня

Для начала сообразим, что из себя представляет поверхность уровня S_t

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

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

Рассмотрим, почему всё именно так

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

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

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

Так уж получилось, что только что был выведен практически универсальный способ построения поверхности уровня $g(\vec{u}) = \max\limits_{i=\overline{1,n}} u_i = t$. 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 $g(\vec{x}) = \max\limits_{i=\overline{1,n}} x_i$.

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

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

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

$$\vec{\nabla} \cdot g(\vec{x}) = \frac{\partial g}{\partial x_1} \cdot \vec{e_1} + \frac{\partial g}{\partial x_2} \cdot \vec{e_2} + \cdots + \frac{\partial g}{\partial x_n} \cdot \vec{e_n} = \vec{e_1} + 0 + \cdots + 0 = \vec{e_1}$$

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

$$\mathrm{M}\left[x_1 \mid x_{(n)} = t\right] = \frac{\int\limits_{S_t} u_1 \cdot \theta^{-n} \cdot \frac{1}{1} \sigma_t\left(d\vec{u}\right)}{\int\limits_{S_t} \theta^{-n} \cdot \frac{1}{1} \sigma_t\left(d\vec{u}\right)} = \frac{\int\limits_{S_t} u_1 \, \sigma_t\left(d\vec{u}\right)}{\int\limits_{S_t} \sigma_t\left(d\vec{u}\right)}$$

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

$$\mathrm{M}\left[x_1 \mid x_{(n)} = t\right] = \frac{\int\limits_{S_t} u_1 \, \sigma_t\left(d\vec{u}\right)}{\int\limits_{S_t} \sigma_t\left(d\vec{u}\right)} = \frac{\int\limits_{U_t} u_1\Big|_{u_1 = t} \sigma_t\left(d\vec{u}\right) + \int\limits_{Y_t} u_1 \, \sigma_t\left(d\vec{u}\right)}{\int\limits_{S_t} \sigma_t\left(d\vec{u}\right)}$$

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

$$\frac{\int\limits_{U_t} u_1\Big|_{u_1 = t} \sigma_t\left(d\vec{u}\right) + \int\limits_{Y_t} u_1 \, \sigma_t\left(d\vec{u}\right)}{\int\limits_{S_t} \sigma_t\left(d\vec{u}\right)} = \frac{t^n + (n-1) \cdot t^{n-2} \cdot \int\limits_0^t u_1 \, du_1}{n \cdot t^{n-1}}$$

Дальше идут нехитрые математические преобразования, которые называются интегрированием

$$\mathrm{M}\left[x_1 \mid x_{(n)} = t\right] = \frac{t^n + (n-1) \cdot t^{n-2} \cdot \int\limits_0^t u_1\, du_1}{n \cdot t^{n-1}} =$$
$$= \frac{t^n + (n-1) \cdot t^{n-2} \cdot \frac{t^2}{2}}{n \cdot t^{n-1}} = \frac{t + (n-1) \cdot t \cdot \frac{1}{2}}{n} = t \cdot \frac{n+1}{2 \cdot n}$$

Условное математическое ожидание

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

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

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

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

Проверка

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

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

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

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

Функция π должна обладать следующими свойствами

1. На любом элементе Δ борелевского множества $\mathfrak{F}_1$ функция $\pi(\cdot, \Delta)$ является измеримой относительно $\mathfrak{F}_1$

2. На любом элементарном исходе из множества Ω функция $\pi(\omega, \cdot)$ является вероятностной мерой

3. Выполняется равенство

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

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

$$\mathbb{P}(\xi \in \Delta) = \mathrm{M}\,\mathbb{1}\{\xi \in \Delta\}$$

Обозначение

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

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

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

Когда нас интересует событие $\eta = t$, работает следующее обозначение

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

Связь с условным математическим ожиданием

$$\mathrm{M}\left[f(\xi) \mid \eta = t\right] = \int\limits_{\mathbb{R}} f(u)\; p(t, du)$$

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

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

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

$$\mathbb{P}\left\{\xi = a_i, \eta = b_j\right\} = \frac{p_{ij}}{\sum_j p_{ij}}$$

Пример 2.5.4: См. формулу 2.11

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

$$\frac{\int\limits_{\Delta} y \cdot p\left(x, y\right)\, dy}{\int\limits_{\mathbb{R}} p\left(x, y\right)\, dy}$$

Пример 2.5.5: См. теорему 2.4.15

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

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

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 $T\left(x_1, x_2, \ldots, x_n\right)$ — 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:

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

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

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

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

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

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 2.3.1), а это значит, что существует функция f такая, что

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

Тогда получаем такое красивое равенство

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

Теорема 2.6.2: Об улучшении оценки с помощью достаточной статистики

Пусть $x_1, \ldots, x_n$ — 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_*

$$\theta_* = \mathrm{M}\left[\hat{\theta} \,\middle|\, T\right]$$

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

$$\begin{cases} \mathrm{M}_\theta \theta_* = \theta \\ \mathrm{D}_\theta \theta_* \leq \mathrm{D}_\theta \hat{\theta} \end{cases}$$

Замечание 2.6.3

Оценка $\theta_* = \mathrm{M}\left[\hat{\theta} \,\middle|\, T\right]$ 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)

$$\mathrm{M}_\theta \theta_* = \mathrm{M}_\theta \mathrm{M}\left[\hat{\theta} \,\middle|\, T\right] = \mathrm{M}_\theta \hat{\theta} = \theta$$

2. Тут же доказательство пройдёт в несколько этапов.

   Сначала распишем дисперсию и и оценку θ_* по определению

$$\mathrm{D}_\theta \theta_* = \mathrm{M}_\theta \left(\theta_* - \theta\right)^2 = \mathrm{M}_\theta \left(\mathrm{M}\left[\hat{\theta} \,\middle|\, T\right] - \theta\right)^2$$

   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)

$$\mathrm{M}_\theta \left(\mathrm{M}\left[\hat{\theta} \,\middle|\, T\right] - \theta\right)^2 = \mathrm{M}_\theta \left(\mathrm{M}\left[\hat{\theta} \,\middle|\, T\right] - \mathrm{M}\left[\theta \mid T\right]\right)^2 = \\ = \mathrm{M}_\theta \left(\mathrm{M}\left[\left(\hat{\theta} - \theta\right) \,\middle|\, T\right]\right)^2$$

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

$$\mathrm{M}_\theta \left(\mathrm{M}\left[\left(\hat{\theta} - \theta\right) \,\middle|\, T\right]\right)^2 \leq \mathrm{M}_\theta \mathrm{M}\left[\left(\hat{\theta} - \theta\right)^2 \,\middle|\, T\right] = \\ = \mathrm{M}_\theta \left(\hat{\theta} - \theta\right)^2 = \mathrm{D}_\theta \hat{\theta}$$

□

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

Пусть $\tilde{\theta}$ — оптимальная несмещённая оценка. Тогда $\theta_* = \mathrm{M}\left[\hat{\theta} \,\middle|\, T\right]$ — 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

Пусть $x_1, \ldots, x_n$ — 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 [6, стр. 78], [1, стр. 158].

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

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

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

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

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

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

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 функцей правдоподобия, а плотность T будет зависеть не только от t, но и от параметра

θ.

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

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.

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

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 $\vec{u}$

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

Значит, функция правдоподобия представима в следующем виде

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

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

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

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

$x_1, \ldots, x_n$ — 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

$$\sum_{k=1}^{n} (x_k - \theta)^2 = \sum_{k=1}^{n} x_k^2 - 2 \cdot \theta \cdot \sum_{k=1}^{n} x_k + n \cdot \theta^2 = \\ = \sum_{k=1}^{n} x_k^2 - 2 \cdot \theta \cdot n \cdot \overline{x} + n \cdot \theta^2$$

Теперь выделим квадрат разности выборочного среднего и параметра θ

$$\sum_{k=1}^{n} x_k^2 - 2 \cdot \theta \cdot n \cdot \overline{x} + n \cdot \theta^2 = \\ = \sum_{k=1}^{n} x_k^2 - 2 \cdot \theta \cdot n \cdot \overline{x} + n \cdot \theta^2 + n \cdot \overline{x}^2 - n \cdot \overline{x}^2 = \\ = \sum_{k=1}^{n} \left(x_k^2 - \overline{x}^2\right) + n \cdot (\overline{x} - \theta)^2$$

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

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

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

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

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

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

Глава 3

Метод наименьших квадратов

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.

3.1 Гауссовские случайные вектора

3.1.1 Основные характеристики случайного вектора

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

Определение 3.1.1: Плотность распределения случайного вектора

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

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

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

2. Во всех точках плотность неотрицательна

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

3. Выполняется условие нормировки

$$\int\limits_{\mathbb{R}^n} p\left(\vec{u}\right)\, d\vec{u} = 1$$

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

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

Когда существует плотность, имеем преобразование Фурье

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

Определение 3.1.3: Математическое ожидание случайного вектора

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

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

Но что же является дисперсией случайного вектора?

3.1.2 Ковариационная матрица

Начнём с определения ковариации двух случайных величин.

Определение 3.1.4: Ковариация

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

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

Замечание 3.1.5

Ковариация случайной величины ξ с ней же — её дисперсия

$$\operatorname{cov}(\xi,\xi) = \mathrm{M}\left[(\xi - \mathrm{M}\xi)\cdot(\xi - \mathrm{M}\xi)\right] = \mathrm{M}\left[(\xi - \mathrm{M}\xi)^2\right] = \mathrm{D}\xi$$

Замечание 3.1.6

Ковариация симметрична

$$\operatorname{cov}(\xi,\eta) = \mathrm{M}\left[(\xi - \mathrm{M}\xi)\cdot(\eta - \mathrm{M}\eta)\right] = \mathrm{M}\left[(\eta - \mathrm{M}\eta)\cdot(\xi - \mathrm{M}\xi)\right] = \operatorname{cov}(\eta,\xi)$$

Определение 3.1.7

Ковариационная матрица двух случайных векторов $\vec{\xi} = (\xi_1, \dots, \xi_n)$ и $\vec{\eta} = (\eta_1, \dots, \eta_n)$ — матрица, в которой на пересечении i строки и j столбца стоит ковариация случайных величин ξ_i и η_j

$$\mathrm{Cov}_{\vec{\xi},\vec{\eta}} = \|\operatorname{cov}(\xi_i,\eta_j)\|_{i,j=1}^n = \|\mathrm{M}\left\{(\xi_i - \mathrm{M}\xi_i)\cdot(\eta_j - \mathrm{M}\eta_j)\right\}\|_{i,j=1}^n$$

$$\mathrm{Cov}_{\vec{\xi},\vec{\eta}} = \begin{bmatrix} \operatorname{cov}(\xi_1,\eta_1) & \cdots & \operatorname{cov}(\xi_1,\eta_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(\xi_n,\eta_1) & \cdots & \operatorname{cov}(\xi_n,\eta_n) \end{bmatrix}$$

Ковариационная матрица случайного вектора $\vec{\xi} = (\xi_1, \dots, \xi_n)$ имеет следующий вид

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

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

Замечание 3.1.8

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

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

Пример 3.1.9

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

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

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

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

Пример 3.1.10

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

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

На диагонали будут стоять единицы — дисперсии случайных величин. Если две случайные величины независимы, то их ковариация равна нулю [7, с. 244]. Это в свою очередь означает, что вне диагонали будут нули

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

3.1.3 Свойства ковариационной матрицы

Литература

[1] Боровков А. А. Математическая статистика. Санкт-Петербург: Лань, 2010. 705 с.

[2] Боровков А. А. Теория Вероятностей. Москва: Эдиториал УРСС, 1999. 472 с.

[3] Дороговцев А. Я. Элементы общей теории меры и интеграла. Киев: Выща школа. Головное издательство, 1989. 152 с.

[4] Ширяев А. Н. Вероятность-1. Москва: МЦНМО, 2004. 520 с.

[5] А. Н. Колмогоров С. В. Фомин. Элементы теории функций и функционального анализа. Москва: Наука, 1976. 543 с.

[6] В.Б. Горяинов И.В. Павлов Г.М. Цветкова и др. Математическая статистика. Москва: МГТУ им. Н. Э. Баумана, 2001. 424 с.

[7] Феллер В. Введение в теорию вероятностей и её приложения. Том 1. Москва: Мир, 1984. 528 с.

Предметный указатель

Оглавление