

Математическая Статистика

3 мая 2014 г.

Глава 1

ОСНОВЫ

1.1 Методы оценок характеристик распределения наблюдаемых случайных величин

x_1, \dots, 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, \dots, x_n называется функция

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

Теорема 1.1.1. Неизвестная функция распределения $F(x)$ может быть сколь угодно точно восстановлена по выборке достаточно большого объёма [1, стр. 25].

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

Идея доказательства. Вспомним, чему равна эмпирическая функция распределения

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

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

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

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

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

Теорема доказана

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

□

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

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

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

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

Делим обе части на Δ_x .

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

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

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

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

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

Возьмём m полуинтервалов на числовой прямой $I_j = (a_{j-1}, a_j], i = \overline{1, m}$ таких, что все значения выборки попадают в один из них. Для этого определим пару свойств точек, ограничивающих эти интервалы:

1.1. Методы оценок характеристик распределения наблюдаемых случайных величин

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 больше максимального

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

Введём функцию $q(y)$

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

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

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

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

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

Функция q_n называется **гистограммой**.

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

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

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

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

Сгруппируем слагаемые и получим чуть более компактную запись разности

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

Рассмотрим возможные значения индикаторов

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

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

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

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

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

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

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

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

Вспомним формулу (1.2)

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

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

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

1.1. Методы оценок характеристик распределения наблюдаемых случайных величин 7

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

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

Получаем компактную запись для разности функций распределения

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

Вернёмся к уравнению (1.1)

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

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

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

Упростим, введя функцию $\nu_j(X)$ [1, стр. 68], которая считает количество элементов выборки $X = x_1, \dots, x_n$, попавших в интервал I_j . Это будет сумма индикаторов того, что элемент x_k попал в I_j

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

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

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

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

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

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

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

Очень напоминает ситуацию с плотностью распределения непрерывной случайной величины ξ

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

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

1.1.3 Оценка неизвестных параметров

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

Пример 1.1.1. Имеем нормальное распределение с известным СКО $\sigma = 1$ и неизвестным математическим ожиданием a — $N(a, 1)$. Тогда θ — математическое ожидание a

Пример 1.1.2. Есть нормальное распределение, в котором неизвестны оба параметра. Тогда θ будет парой (a, σ)

Главный вопрос — определение основных параметров распределения выборки.

Определение 1.1.2 (Статистика). Статистикой называют функцию S от выборки $X = (x_1, x_2, \dots, x_n)$

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

Определение 1.1.3 (Оценка). Статистику, значение которой заменяет неизвестный параметр, называют оценкой

1.1. Методы оценок характеристик распределения наблюдаемых случайных величин⁹

Пример 1.1.3. 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}

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

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

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

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

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

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

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

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

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

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

Определение 1.1.4 (Состоятельная оценка). Оценка $\hat{\theta}$ называется состоятельной, если стремится к истинному значению θ по вероятности

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

Определение 1.1.5 (Сильно состоятельная оценка). Оценка $\hat{\theta}$ называется *сильно состоятельной*, если стремится к истинному значению θ почти наверное

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

Пример 1.1.4. Оценка \hat{p}_1 из прошлого примера является сильно состоятельной.

Определение 1.1.6 (Несмещённая оценка). Оценка $\hat{\theta}$ несмещённая, если

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

Замечание 1. Несмещённая оценка существует не всегда

Определение 1.1.7. Несмещённая оценка $\hat{\theta} \in K$ называется *оптимальной* в классе квадратично интегрируемых оценок K , если для всякой другой несмещённой оценки $\tilde{\theta} \in K$

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

или же

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

Замечание 2. 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 1.1.5. Сравним \hat{p}_1 и \hat{p}_3

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

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

1.1.4 Выборочные оценки. Метод моментов

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

Вспомним распределения и их параметры

1. Нормальное распределение $N(a, \sigma^2)$. В нём параметр a является средним, а параметр σ^2 — дисперсией
2. Пуассоновское распределение $Poi(\lambda)$. Тут параметр λ является и средним, и дисперсией
3. Экспоненциальное распределение $Exp(\lambda)$. $\frac{1}{\lambda}$ — среднее, $\frac{1}{\lambda^2}$ — дисперсия

И так далее...

Как правило, неизвестный параметр θ можно искать следующим образом

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

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

$$g(\hat{\theta}) = \int_{\mathbb{R}} \varphi(x) dF_n(x) \quad (1.5)$$

Пример 1.1.6. Если θ — среднее, то $\varphi(x) = x$

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

Теорема 1.1.2. Пусть функция $\varphi(x)$ в (1.5) непрерывна, ограничена и строго монотонная. Тогда оценка $\hat{\theta}$ существует и является сильно состоятельной.

Доказательство. Имеем формулу (1.5)

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

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

Применим обратную функцию к обеим частям уравнения

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

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

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

Функция $g^{-1}(x)$ непрерывна

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

Теорема доказана

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

□

Определение 1.1.8 (Выборочное среднее). *Выборочное среднее обозначается через \bar{x} и считается по следующей формуле*

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

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

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

Определение 1.1.9 (Выборочная дисперсия). *Выборочная дисперсия $\overline{\sigma^2}$ считается формуле*

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

1.2 Свойства оценок

1.2.1 Неравенство Рао-Крамера

Теорема 1.2.1 (Колмогорова). *Оптимальная оценка единственная или её нет вообще*

Доказательство. Допустим, есть две разные оптимальные и несмещённые оценки θ_1 и θ_2 . Тогда по определению для любой несмещённой оценки $\hat{\theta}$ будет

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

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

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

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

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

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

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

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

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

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

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

Воспользуемся неравенством Коши (частный случай неравенства Гёльдера)

$$\begin{aligned} M_{\theta}[(\theta_1 - \theta) \cdot (\theta_2 - \theta)] &\leq \sqrt{M_{\theta}(\theta_1 - \theta)^2 \cdot M_{\theta}(\theta_2 - \theta)^2} = \\ &= \sqrt{D_{\theta}\theta_1 \cdot D_{\theta}\theta_2} = \sqrt{\sigma_1^2 \cdot \sigma_2^2} \end{aligned} \quad (1.7)$$

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

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

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

$$D_{\theta}\tilde{\theta} \leq \sigma^2(\theta) \quad (1.8)$$

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

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

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

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

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

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

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

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

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

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

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

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

Теорема доказана □

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

Определение 1.2.1 (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

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

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

Определение 1.2.2 (Вклад выборки). *Вклад выборки — частная производная по параметру θ от логарифма функции правдоподобия*

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

Замечание 3. *Математическое ожидание вклада выборки равно нулю*

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

Доказательство. Посчитаем математическое ожидание вклада выборки

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

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

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

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

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

□

Замечание 4. *Частная производная по оценке θ от функции правдоподобия $L(\vec{u}, \theta)$ равна нулю.*

Доказательство. Выше у нас было равенство

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

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

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

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

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

□

Определение 1.2.3 (Количество информации Фишера). Математическое ожидание квадрата вклада выборки называется количеством информации Фишера

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

Замечание 5.

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

Доказательство. Будем доказывать справа налево

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

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

$$\begin{aligned} \frac{\partial}{\partial \theta} L(\vec{u}, \theta) = 0 &\Rightarrow -M_\theta \frac{\frac{\partial^2}{\partial \theta^2} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} = 0 \\ \Rightarrow -M_\theta \frac{\partial^2}{\partial \theta^2} \ln L(\vec{x}, \theta) &= M_\theta \left[\frac{\frac{\partial}{\partial \theta} L(\vec{x}, \theta)}{L(\vec{x}, \theta)} \right]^2 = \\ &= M_\theta \left[\frac{\partial}{\partial \theta} \ln L(\vec{x}, \theta) \right]^2 = M_\theta U(\vec{x}, \theta)^2 \end{aligned}$$

Утверждение доказано

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

□

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

Теорема 1.2.2 (Неравенство Рао-Крамера). Пусть $\hat{\theta}$ — несмещённая оценка параметра θ . Тогда имеет место неравенство

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

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

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

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

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

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

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

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

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

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

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

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

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

$$1 = M_{\theta} \left(\hat{\theta} \cdot U(\vec{x}, \theta) \right) \quad (1.9)$$

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

$$\begin{aligned} M_{\theta} U(\vec{x}, \theta) &= 0 \\ \Rightarrow \theta \cdot M_{\theta} U(\vec{x}, \theta) &= M_{\theta} (\theta \cdot U(\vec{x}, \theta)) = 0 \end{aligned}$$

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

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

Получаем компактное равенство

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

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

$$\begin{aligned} 1 &= M_{\theta} \left[(\hat{\theta} - \theta) \cdot U(\vec{x}, \theta) \right] \leq \\ &\leq \sqrt{M_{\theta} (\hat{\theta} - \theta)} \cdot \sqrt{M_{\theta} U(\vec{x}, \theta)} = \\ &= \sqrt{D_{\theta} \hat{\theta}} \cdot \sqrt{I_n(\theta)} \end{aligned} \quad (1.10)$$

Возводим обе части равенства в квадрат и делим на количество информации

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

Неравенство доказано \square

Замечание 6. Иногда нужно оценивать не сам параметр, а функцию параметра

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

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

1.2.2 Метод максимального правдоподобия

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

Определение 1.2.4 (Эффективная оценка). Оценка $\hat{\theta}$, для которой в неравенстве Рао-Крамера стоит равенство, называется эффективной

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

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)

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

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

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

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 квадратов математических ожиданий).

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 $k(\theta)$.

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

Проинтегрируем обе части равенства

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

Получим следующее равенство

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

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

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

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

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

При конечном n положим такую плотность распределения

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

В таком случае получим следующую функцию правдоподобия

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

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 1.2.5 (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 1.2.1. *Есть выборка x_1, x_2, \dots, x_n из нормального распределения с неизвестным математическим ожиданием $N(\theta, 1)$. Тогда плотность распределения k -ой случайной величины будет следующей*

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

Её логарифм, очевидно, имеет такой вид

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

Теперь выпишем логарифм функции правдоподобия

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

Раскроем скобки

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

Воспользуемся формулой для несмещённой (ещё и эффективной) оценки среднего

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

Сгруппировав множители n , получаем

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

Добавим и отнимем в числителе дроби выборочное среднее

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

Теперь в числителе очевиден квадрат разности

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

Записываем квадрат разности короче, а выборочное среднее вносим под знак суммы

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

Видим, что последнее вычитаемое не может быть отрицательным, а когда оценка θ равна выборочному среднему, 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 1.2.6 (Оценка максимального правдоподобия). Оценка максимального правдоподобия θ_* — такое значение параметра θ , при котором функция правдоподобия достигает своего максимального значения

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

Замечание 7. Оценок максимального правдоподобия может быть несколько, а может не существовать ни одной.

Определение 1.2.7 (Уравнение правдоподобия). Уравнением правдоподобия называется равенство вида

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

Или же

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

Замечание 8. 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 1.2.8 (Вариационный ряд). *Вариационный ряд выборки x_1, x_2, \dots, x_n — значения выборки, упорядоченные в порядке неубывания*

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

Теорема 1.2.3. *Если плотность $p(x, \theta)$ непрерывна и дифференцируема по параметру θ , а производная не равна нулю $\frac{\partial}{\partial \theta} p(x, \theta) \neq 0$, то оценка максимального правдоподобия состоятельна*

Глава 2

Достаточные статистики

2.1 Оптимальная оценка

Определение 2.1.1 (Симметризация). Симметризация Λ оценки $\hat{\theta}$ — среднее оценок $\hat{\theta}$ для всевозможных перестановок $\sigma \in S_n$ элементов выборки x_1, x_2, \dots, x_n

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

Лемма 2.1.1. Для произвольной несмещённой оценки $\hat{\theta}$ её симметризация $\Lambda \hat{\theta}$ не хуже её самой в среднем квадратическом

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

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

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

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

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

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

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

Тогда и оценки примут более красивый вид

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

Теперь приступим непосредственно к доказательству.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Воспользуемся определением

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

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

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

Вспомним неравенство Йенсена для выпуклой функции f

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

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

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

Перепишем неравенство Йенсена для нашего случая

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

Воспользуемся линейностью математического ожидания, внеся его под знак суммы

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

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

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

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

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

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

Замечание 9. Равенство в неравенстве Йенсена (в доказательстве выше) возможно только в случае симметричной функции. Значит, в качестве оценки достаточно брать только симметричные функции выборки

Определение 2.1.2 (Функция вариационного ряда). Если оценка $\hat{\theta}$ симметрична относительно перестановок аргументов, то она является функцией вариационного ряда

Замечание 10. Все оценки, которые претендуют быть оптимальными, должны быть функциями вариационного ряда

2.2 Условное математическое ожидание и условные распределения

2.2.1 σ -алгебра, порождённая случайной величиной

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

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

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

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

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

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

Определение 2.2.1 (Сигма-алгебра, порождённая случайной величиной). $\mathfrak{F}_{\xi} = \sigma(\xi)$ — σ -алгебра, порождённая случайной величиной ξ

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

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

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

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

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

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

1. Множество элементарных исходов Ω входит в \mathfrak{F}_ξ . Поскольку случайная величина ξ принимает действительные значения, то прообраз множества действительных чисел \mathbb{R} и будет множеством элементарных исходов Ω . А поскольку \mathbb{R} принадлежит борелевской σ -алгебре, то его прообраз по определению принадлежит σ -алгебре \mathfrak{F}_ξ

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

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

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

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

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

3. Замкнутость относительно счётных пересечений.

Начнём с замкнутости относительно пересечения двух множеств

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

Начинаем считать

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

Значит, имеем равенство

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

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

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

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

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

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

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

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

Определение 2.2.2 (Множество уровня). *Множество уровня H_t — полный прообраз значения $t \in \mathbb{R}$ случайной величины ξ*

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

Замечание 11. 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.

2.2.2 Случайная величина, измеримая относительно σ -алгебры, порождённой случайной величиной

В общем случае вероятностное пространство может быть разбито на континуальное количество множеств уровней.

Начнём же с рассмотрения того случая, когда случайная величина ξ принимает n значений: a_1, a_2, \dots, a_n

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

Это в свою очередь означает, что у нас есть n уровней

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

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

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

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

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

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

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

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

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

Объединим и проанализируем вышеописанное:

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

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

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

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

3. Множество $A(c)$ растёт дискретными шагами. Это связано с тем, что уровни H_k в нашей σ -алгебре 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 ($m \leq n$) $c_1 < c_2 < \dots < c_m$ на числовой прямой \mathbb{R} как значения случайной величины \varkappa

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

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

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

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

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

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

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

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

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

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

В таком случае тождество очевидно

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

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

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

Это в свою очередь означает, что случайная величина \varkappa принимает значение c_i при выпадении любого элементарного исхода ω из множества $A(c_i) \setminus A(c_{i-1})$

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

Запишем это в более удобном виде

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

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

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

Предположим, есть некое t такое, что $A(c_t) \setminus A(c_{t-1})$ является объединением двух (нетрудно показать, что для любого количества, в том числе и одного) уровней, которые обозначим H_1^t и H_2^t . Тогда t -ое слагаемое примет следующий вид

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

Поскольку уровни не пересекаются, то можно разбить индикатор на сумму

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

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

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

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

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

Попробуем показать, что \varkappa является функцией от ξ . Очевидно, что случайная величина ξ имеет такой же вид, что и \varkappa — сумма с константами и индикаторами

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

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 первый вид

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

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 вывод, что случайной величине \varkappa необходимо и достаточно быть функцией случайной величины ξ , чтобы быть измеримой относительно σ -алгебры $\sigma(\xi)$, порождённой случайной величиной ξ

2.2.3 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 .

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 к L .

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

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

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

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

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

Вследствие линейности скалярного произведения можно переписать это условие иначе

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

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

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

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

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

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

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

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

Проекция случайной величины

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

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

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

Оказывается, H_k действительно ортогональны

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

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

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

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

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

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

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

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

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

Идём дальше, ищем проекцию. Вспомним снова пример с векторами (2.2)

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

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

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

Осталось заменить e_k на то, что получили выше (2.3)

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

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

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

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

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

Произведения корней вероятности даёт саму вероятность. Теперь у нас есть красивая формула для проекции случайной величины

$$\tilde{\eta} = \sum_{k=1}^n \frac{M(\eta \cdot \mathbb{1}_{H_k})}{\mathbb{P}(H_k)} \cdot \mathbb{1}_{H_k} \quad (2.4)$$

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

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

Начнём с определения математического ожидания

$$\begin{aligned} M(\eta \cdot \mathbb{1}_{H_k}) &= \int_{\Omega} \eta(\omega) \cdot \mathbb{1}_{H_k} \cdot \mathbb{P}(d\omega) = \\ &= \int_{H_k} \eta(\omega) \mathbb{P}(d\omega) + \int_{\Omega \setminus H_k} 0 \cdot \mathbb{P}(d\omega) \end{aligned} \quad (2.5)$$

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

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

Из контекста понятно, что нигде кроме как в интеграле по событию H_k эта мера использоваться не будет, поэтому она 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 значит, 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 (2.5) принимает несколько иной вид

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

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

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

Определение 2.2.3 (Условное математическое ожидание случайной величины относительно случайного события). *Условное математическое ожидание случайной величины ξ относительно события A [2, стр. 68] обозначается $M(\xi | A)$ и считается по формуле*

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

Пользуясь только что введённым обозначением, можно более красиво переписать формулу (2.4) для получения проекции случайной величины η на σ -алгебру, порождённую уровнями H_1, H_2, \dots, H_n

$$\tilde{\eta} = \sum_{k=1}^n M(\eta | H_k) \cdot \mathbb{1}_{H_k}$$

Забегая наперёд, введём определение частного случая условного математического ожидания, чтобы обратить внимание на этот важный момент.

Определение 2.2.4 (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, \dots, H_n . Тогда условное математическое ожидание дискретной случайной величины η относительно этой σ -алгебры — случайная величина, которая обозначается $M[\eta | \mathfrak{F}_1]$ и вычисляется по формуле*

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

Замечание 12. *Далее не будем использовать условное математическое ожидание случайной величины относительно случайного события, так как это может вызвать путаницу. Например, последнее определение может вызвать недоумение, если его переписать, пользуясь этим определением*

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

Поэтому далее запись $M[\eta | \mathfrak{F}_1]$ будет означать лишь условное математическое ожидание случайной величины η относительно σ -алгебры \mathfrak{F}

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.

Предметный указатель

- Сигма-алгебра, порождённая случай-
ной величиной, 26
- функция
 - правдоподобия, 14
 - вариационного ряда, 26
- функция распределения
 - эмпирическая, 3
 - неизвестная, 3
 - выборочная, 3
- гистограмма, 5
- количество информации Фишера, 16
- математическое ожидание
 - условное, 32, 36
- множество уровня, 28
- неизвестный параметр, 8
- неравенство
 - Рао-Крамера, 16
- оценка, 8
 - эффективная, 18
 - максимального правдоподобия, 21
 - несмещённая, 10
 - сильно состоятельная, 9
 - состоятельная, 9
- проекция
 - случайной величины, 34
- распределение
 - экспоненциальное, 20
- сигма-алгебра
 - порождённая случайной величи-
ной, 26
- симметризация, 23
- случайная величина
 - измеримая относительно сигма-
алгебры, 28
- статистика, 8
- теорема
 - Колмогорова, 12
- уравнение
 - правдоподобия, 21
- условное математическое ожидание, 32, 36
- вариационный ряд, 22
- выборочная дисперсия, 12
- выборочное среднее, 12
- вклад выборки, 15

Оглавление

| | | |
|----------|---|-----------|
| 1 | Основы | 3 |
| 1.1 | Методы оценок характеристик распределения наблюдаемых случайных величин | 3 |
| 1.1.1 | Эмпирическая функция распределения | 3 |
| 1.1.2 | Гистограмма | 4 |
| 1.1.3 | Оценка неизвестных параметров | 8 |
| 1.1.4 | Выборочные оценки. Метод моментов | 10 |
| 1.2 | Свойства оценок | 12 |
| 1.2.1 | Неравенство Рао-Крамера | 12 |
| 1.2.2 | Метод максимального правдоподобия | 18 |
| 2 | Достаточные статистики | 23 |
| 2.1 | Оптимальная оценка | 23 |
| 2.2 | Условное математическое ожидание и условные распределения | 26 |
| 2.2.1 | σ -алгебра, порождённая случайной величиной | 26 |
| 2.2.2 | Случайная величина, измеримая относительно σ -алгебры, порождённой случайной величиной | 28 |
| 2.2.3 | Условное математическое ожидание | 32 |