Математическая Статистика

30 марта 2014 г.

функция распределения!неизвестная
функция распределения!эмпирическая
функция распределения!выборочная

Глава 1

Основы

1.1 Методы оценок характеристик распределения 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*

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

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

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

Идея доказательства. Вспомним, чему равна эмпирическая функция распределения

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

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

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

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

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

Теорема доказана

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

□

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 $p(x)$ 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. Каждая следующая точка строго правее (больше) предыдущей. (так как зачем нам одинаковые точки?) гистограмма

$$a_0 < a_1 < \cdots < a_m$$

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

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

Введём функцию $q\left(y\right)$

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

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

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

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 $F_n\left(x\right)$

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

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

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

Сгруппируем слагаемые и получим чуть более компактную запись разности

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

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 $x \le a_{j-1}$

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

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 $a_j < a_{j-1}$

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

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 $x > a_j$.

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

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

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

Вспомним формулу (1.2)

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

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

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

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

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

Получаем компактную запись для разности функций распределения

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

Вернёмся к уравнению (1.1)

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

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

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

Упростим, введя функцию $\nu_j\left(X\right)$ [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

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

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

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

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 $q_n\left(y\right)$ запишем как q_n^k

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

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 $q\left(y\right)$ сходилось к $p\left(y\right)$.

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.

неизвестный параметр
статистика
оценка

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 ($m \to \infty$), 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 1.1.1. *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 $X = (x_1, x_2, \ldots, x_n)$*

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

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

оценка!состоятельная
оценка!сильно состоятельная

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 ($\hat{p} \xrightarrow[n\to\infty]{\mathbb{P}} p$) 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

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

   Например, если целый год каждый день дают набор чисел, а статистик считает значение параметра p с помощью оценки $\hat{p}$, 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*

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

оценка!несмещённая

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

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

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 $\tilde{\theta} \in K$*

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

или же

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

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

И так далее... выборочное среднее

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 $\hat{\theta}$

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

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

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

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

Доказательство. Имеем формулу (1.5)

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

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

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

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

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

Функция $g^{-1}(x)$ непрерывна

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

Теорема доказана

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

$\square$

выборочная дисперсия
теорема!Колмогорова

Определение 1.1.8 (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 1.1.9 (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

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

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 $\hat{\theta}$

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

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

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

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

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

Воспользуемся неравенством Коши (частный случай неравенства Гёльдера)

$$M_\theta \left[(\theta_1 - \theta) \cdot (\theta_2 - \theta)\right] \leq \sqrt{M_\theta (\theta_1 - \theta)^2 \cdot M_\theta (\theta_2 - \theta)^2} = \tag{1.7}$$
$$= \sqrt{D_\theta \theta_1 \cdot D_\theta \theta_2} = \sqrt{\sigma_1^2 \cdot \sigma_2^2}$$

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

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

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

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

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

Для дальнейших размышлений вспомним аналогию с векторами.

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

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 направление этих векторов совпадает,

функция!правдобия
вклад выборки

$$M_\theta\left[(\theta_1-\theta)\cdot(\theta_2-\theta)\right]=\sqrt{M_\theta\left(\theta_1-\theta\right)^2}\cdot\sqrt{M_\theta\left(\theta_2-\theta\right)^2}$$
$$\Rightarrow\left(\theta_1-\widehat{\theta,\theta_2}-\theta\right)$$

Это значит, что они равны, что противоречит предположению о том, что они разные. Теорема доказана

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

□

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

В таком случае выбора $(x_1,\ldots,x_n)$ имеет плотность распределения, так как является случайным вектором в $\mathbb{R}^n$.

Плотность распределения вектора независимых случайных величин равна произведению плотностей распределения его компонент.

Определение 1.2.1 (Функция правдоподобия)**.**

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

Прологорифмировав 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

$$\ln L(\vec{x},\theta)=n\cdot\frac{\ln p(x_1,\theta)+\cdots+\ln p(x_n,\theta)}{n}\approx$$
$$\approx n\cdot M_\theta\left[\ln p(x_1,\theta)\right]$$

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

Определение 1.2.2 (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 3. *Математическое ожидание вклада выборки равно нулю*

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

Доказательство. Посчитаем математическое ожидание вклада выборки

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

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

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

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

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

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

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

□

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

Доказательство. Выше у нас было равенство

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

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

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

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

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

□

количество информации Фишера
неравенство!Рао-Крамера

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

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

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

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

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 θ

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

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

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

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

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

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

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

оценка!эффективная

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

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

Получаем компактное равенство

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

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

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

Возводим обе части равенства в квадрат и делим на количество информации

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

Неравенство доказано □

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

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

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 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 $k(\theta)$.

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

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

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

распределение!экспоненциальноеОтметим, 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 $x_1, x_2, \ldots, x_n$ 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 (ещё и эффективной) оценки среднего

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

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

Сгруппировав множители n, получаем

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

Добавим и отнимем в числителе дроби выборочное среднее

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

Теперь в числителе очевиден квадрат разности

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

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

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

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.

Таким образом, делаем предположение о том, как находить наилучшую оценку

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

Оказывается, именно так она и находится

Литература

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

Предметный указатель

функция
    правдобия, 14
функция распределения
    эмпирическая, 3
    неизвестная, 3
    выборочная, 3
гистограмма, 5
количество информации Фишера, 16
неизвестный параметр, 8
неравенство
    Рао-Крамера, 16
оценка, 8
    эффективная, 18
    несмещённая, 10
    сильно состоятельная, 9
    состоятельная, 9
распределение
    экспоненциальное, 20
статистика, 8
теорема
    Колмогорова, 12
выборочная дисперсия, 12
выборочное среднее, 11
вклад выборки, 14

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