

# Математическая Статистика

9 марта 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].

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

*Доказательство.* Вспомним, чему равна эмпирическая функция распределения

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

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

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

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  $\Delta_x = b - x$

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

Делим обе части на  $\Delta_x$ .

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

Несложно заметить, что при достаточно малых значениях  $\Delta_x$  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 выборку из  $m$  случайных величин в порядке возрастания  $a_1, \dots, a_m$ , обозначим отрезки  $I_j = [a_{j-1}, a_j]$  и введём функцию  $q(y)$

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

### 1.1. Методы оценок характеристик распределения наблюдаемых случайных величин

Теперь введём последовательность функций  $q_n(y)$  и видим, что она сходится к  $q(y)$  почти наверное согласно закону больших чисел, а та в свою очередь имеет сходимость порядка  $\frac{1}{n}$  к плотности распределения  $p(y)$  гистограмма

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

Отметим, что  $q_n$  сходится к  $q$  почти наверное, а  $q$  в свою очередь сходится к  $p$

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

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  $I_j$

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

Сгруппируем слагаемые и получим чуть более компактную запись разности

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

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  $x \leq a_{j-1}$

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

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  $a_j < a_{j-1}$

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

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  $x > a_j$ .

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

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

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

Вспомним формулу (1.2)

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

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  $(a_{j-1}, a_j]$

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

Пренебрегаем тем, что у нас полуинтервал, и будем считать, что вероятность попадания  $x$  чётко на границу интервала пренебрежимо мала и заменим индикатор на более удобный.

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

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 отрезка  $I_j$ , а разность  $F_n(a_j) - F_n(a_{j-1})$  была только что компактизирована, 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 формулу. Введём 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 интервал  $I_j$

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

### 1.1. Методы оценок характеристик распределения наблюдаемых случайных величин<sup>7</sup>

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 отрезок (пренебрегаем возможностью его попадания на границу между двумя отрезками). Тогда обозначим номер отрезка, в который попал  $y$ , как  $k$ , а функцию  $q_n(y)$  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 отрезка — номер отрезка, в который попал  $y$ ).

“Высота” столбика (значение функции на определённом отрезке) пропорциональна количеству элементов, попавших в этот отрезок (что логично). Кроме того, происходит деление на общее количество элементов, которое возникло, чтобы  $q(y)$  сходилось к  $p(y)$ .

Делителю же  $|I_k|$  отведена особая роль — он предотвращает искажение гистограммы при различных длинах отрезков. Получается, что, чем длиннее отрезок, тем ниже столбик, так как элементы более “размазаны” по отрезку — тоже логично.

Если рассматривать значение функции как высоту прямоугольника, а длину отрезка как его ширину (графически это изображается именно так), то оказывается, что отношение количества элементов, попавших в отрезок, к количеству всех элементов выборки (вероятность того, что случайно взятый элемент из выборки попадёт в  $k$ -ый отрезок [1, стр. 24]) есть 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 отрезков к бесконечности ( $m \rightarrow \infty$ ), то каждый отрезок 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  $\xi$

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

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.

неизвестный параметр  
оценка

### 1.1.3 Оценка неизвестных параметров

Снова у нас есть  $x_1, \dots, x_n$  — выборка из распределения  $F_\theta$ , где  $\theta$  — неизвестный параметр из множества  $\Theta$

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

**Пример 1.1.2.** Есть нормальное распределение, в котором неизвестны оба параметра. Тогда  $\theta$  будет парой  $(a, \sigma)$

Главный вопрос — определение основных параметров распределения выборки.

**Определение 1.1.2** (Оценка). Функцию от выборки, значение которой заменяет неизвестный параметр, называют оценкой

**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 ( $\hat{p} \xrightarrow[n \rightarrow \infty]{a.s.} p$ )



### 1.1. Методы оценок характеристик распределения наблюдаемых случайных величин

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

оценка!состоятельная  
оценка!сильно  
состоятельная  
оценка!несмещённая

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

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

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

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

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

**Определение 1.1.4** (Сильно состоятельная оценка). Оценка  $\hat{\theta}$  называется *сильно состоятельной*, если стремится к истинному значению  $\theta$  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 1.1.5** (Несмещённая оценка). Оценка  $\hat{\theta}$  *несмещённая*, если

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

**Замечание 1.** Несмещённая оценка существует не всегда

**Определение 1.1.6.** Несмещённая оценка  $\hat{\theta} \in K$  называется *оптимальной (эффективной [1, стр. 130])* 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

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

**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  $\theta \in \Theta$ , имея и так далее функцию распределения  $F_\theta(x)$ ?

Вспомним распределения и их параметры

1. Нормальное распределение  $N(a, \sigma^2)$ . В нём параметр  $a$  является средним, а параметр  $\sigma^2$  — дисперсией
2. Пуассоновское распределение  $Poi(\lambda)$ . Тут параметр  $\lambda$  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  $\theta$  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  $\hat{\theta}$

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

**Пример 1.1.6.** Если  $\theta$  — 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) 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

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

### 1.1. Методы оценок характеристик распределения наблюдаемых случайных величин 11

Функция  $g^{-1}(x)$  непрерывна

выборочное среднее  
выборочная дисперсия

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

Теорема доказана

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

□

**Определение 1.1.7** (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 1.1.8** (Выборочная дисперсия). *Выборочная дисперсия  $\overline{\sigma^2}$  считается формуле*

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



# Литература

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

# Предметный указатель

- функция распределения
  - эмпирическая, 3
  - неизвестная, 3
  - выборочная, 3
- гистограмма, 5
- неизвестный параметр, 8
- оценка, 8
  - несмещённая, 9
  - сильно состоятельная, 9
  - состоятельная, 9
- выборочная дисперсия, 11
- выборочное среднее, 11

# Оглавление

<b>1</b>	<b>Основы</b>	<b>3</b>
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