Московский государственный университет
Факультет вычислительной математики и кибернетики

Теория вероятностей. Математическая статистика

Рожков И., Рыгин А.

Москва
2020

Оглавление

Глава 1

Теория вероятностей

Глава 2

Математическая статистика

2.1 Статистическая структура. Выборка. Статистика. Порядковые статистики. Вариационный ряд. Эмпирическая функция распределения

Определение 2.1.1. *Статистическая структура* — совокупность $(\Omega, \mathcal{A}, \mathcal{P})$, где Ω —множество элементарных исходов, $\mathcal{A})$ — σ-алгебра событий, $\mathcal{P})$ — семейство вероятностных мер, определённых на $\mathcal{A}$, параметризованное одно- или многомерным числовым параметром: $\mathcal{P} = (\mathbb{P}_\theta \mid \theta \in \Theta \subset R^m)$.

Определение 2.1.2. *Выборка* $\vec{X} = (X_1, \ldots, X_n)$ объёма n — набор из n независимых и одинаково распределённых случайных величин, имеющих такое же распределение, как и наблюдаемая случайная величина ξ.

До того, как эксперимент проведён, выборка — набор случайных величин, после — набор чисел из множества возможных значений случайной величины. Числовой набор $\vec{X}(\omega_0) = (X_1(\omega_0), \ldots, X_n(\omega_0)) = (x_1, \ldots, x_n)$ — *реализация выборки* на элементарном исходе ω_0.

Определение 2.1.3. *Статистика* — измеримая функция от выборки $T(X)$.

Определение 2.1.4. *Вариационный ряд* — выборка $X_1, \ldots, X_n$, упорядоченная по возрастанию на каждом элементарном исходе:

$$X_{(1)}(\omega) = \min(X_1(\omega), \ldots, X_n(\omega))$$
$$X_{(k)}(\omega) = \{\forall \omega \in \Omega \Rightarrow \exists m \le i_1, \ldots, i_{k-1}, i_k, i_{k+1}, \ldots, i_n \le n, i_j \ne i_m (j \ne m) :$$
$$X_{(k)}(\omega) = X_{i_k}(\omega)$$
$$X_{i_1}(\omega), \ldots, X_{i_{k-1}}(\omega) \le X_{i_k}(\omega); X_{i_{k+1}}(\omega), \ldots, X_{i_n}(\omega) > X_{i_k}(\omega)\}, 2 \le k \le n-1$$
$$X_{(n)}(\omega) = \max(X_1(\omega), \ldots, X_n(\omega))$$

Элемент $X_{(k)}$ — *k-я порядковая статистика.*

Определение 2.1.5. *Эмпирическая функция распределения*, построенная по выборке $X_1, \ldots, X_n$ объёма n — случайная функция $F_n^* : \mathbb{R} \times \Omega \to [0,1]$, при каждом $y \in \mathbb{R}$ равная: $F_n^*(y) = \frac{1}{n}\sum_{i=1}^n \mathrm{I}\left(X_i < y\right)$

Эмпирическая функция распределенния строится по вариационному ряду следующим образом:

$$F_n^*(y) = \begin{cases} 0, & \text{если } y \leqslant X_{(1)} \\ \frac{k}{n}, & \text{если } X_{(k)} < y \leqslant X_{(k+1)} \\ 1 & \text{при } y > X_{(n)} \end{cases}$$

Пример. Найдём эмпирические функции распределения для крайних порядковых статистик.

1. $F_{(1)}(x) = \mathbb{P}(X_{(1)} < x) = 1 - \mathbb{P}(\mathrm{X}_{(1)} \geq x) = 1 - \mathbb{P}(x_1 \geq x, \ldots, x_n \geq x) = 1 - \prod_{i=1}^n \mathbb{P}(x_i \geq x) = 1 - (\mathbb{P}(x_1 \geq x))^{\mathrm{n}} = 1 - (1 - F(x))^n$

2. $F_{(n)}(x) = \mathbb{P}(X_{(n)} < x) = \mathbb{P}(x_1 < x, \ldots, x_n < x) = \prod_{i=1}^n \mathbb{P}(x_i < x) = (\mathbb{P}(x_1 < x))^{\mathrm{n}} = F^n(x)$

Теорема 2.1.1. *Свойства эмпирической функции распределения:*

1. *Пусть $X_1, \ldots, X_n$ –— выборка из распределения $\mathcal{F}$ с функцией распределения F и пусть F_n^* — эмпирическая функция распределения, построенная по этой выборке. Тогда $F_n^*(y) \xrightarrow{\mathrm{p}} F(y)$ при $n \to \infty$ для любого $y \in \mathbb{R}$.*

2. *Для любого $y \in \mathbb{R}$:*

   1) *$\mathbb{E}F_n^*(y) = F(y)$, т.е. $F_n^*(y)$ — несмещённая оценка для $F(y)$.*
   2) *$\mathbb{D}F_n^*(y) = \dfrac{F(y)(1-F(y))}{n}$*
   3) *$\sqrt{n}(F_n^*(y) - F(y)) \Rightarrow \mathrm{N}_{0,F(y)(1-F(y))}$, т.е. $F_n^*(y)$ — асимптотически нормальная оценка для $F(y)$.*

Доказательство.

1. $F_n^*(y) = \frac{1}{n}\sum_{i=1}^n \mathrm{I}(X_i < y)$, при этом случайные величины $\mathrm{I}(X_1 < y)$, $\mathrm{I}(X_2 < y), \ldots$ независимы и одинаково распределены, их математическое ожидание конечно:

$$\mathbb{E}\mathrm{I}(X_1 < y) = 1 \cdot \mathrm{P}(X_1 < y) + 0 \cdot \mathrm{P}(X_1 \geqslant y) = \mathrm{P}(X_1 < y) = F(y) < \infty$$

Следовательно, применим ЗБЧ в форме Хинчина:

$$F_n^*(y) = \frac{\sum_{i=1}^n \mathrm{I}(X_i < y)}{n} \xrightarrow{\text{p}} \mathbb{E}\mathrm{I}(X_1 < y) = F(y)$$

2. Заметим, что $\mathrm{I}(X_1 < y) \sim \mathrm{B}_{F(y)} \Rightarrow \mathbb{E}\mathrm{I}(X_1 < y) = F(y); \mathbb{D}\mathrm{I}(X_1 < y) = F(y)(1 - F(y))$.

   1) Случайные величины $\mathrm{I}(X_i < y)$ одинаково распределены, поэтому:

   $$\mathbb{E}F_n^*(y) = \mathbb{E}\frac{\sum_{i=1}^n \mathrm{I}(X_i < y)}{n} = \frac{\sum_{i=1}^n \mathbb{E}\mathrm{I}(X_i < y)}{n} = \frac{n\mathbb{E}\mathrm{I}(X_1 < y)}{n} = F(y)$$

   2) Случайные величины $\mathrm{I}(X_i < y)$ независимы и одинаково распределены, поэтому:

   $$\mathbb{D}\mathrm{I}_n^*(y) = \mathbb{D}\frac{\sum_{i=1}^n \mathrm{I}(X_i < y)}{n} = \frac{\sum_{i=1}^n \mathbb{D}\mathrm{I}(X_i < y)}{n^2} = \\ = \frac{n\mathbb{D}\mathrm{I}(X_1 < y)}{n^2} = \frac{F(y)(1 - F(y))}{n}$$

   3) Применим ЦПТ:

   $$\sqrt{n}\left(F_n^*(y) - F(y)\right) = \sqrt{n}\left(\frac{\sum I(X_i < y)}{n} - F(y)\right) = \\ = \frac{\sum_{i=1}^n I(X_i < y) - nF(y)}{\sqrt{n}} = \frac{\sum_{i=1}^n I(X_i < y) - n\mathbb{E}\mathrm{I}(X_1 < y)}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\mathbb{D}\mathrm{I}(X_1 < y)}$$

■

2.2 Выборочные моменты. Их свойства

Рассмотрим случайную величину ξ^* с эмпирическим распределением, введём для последнего числовые характеристики.

Определение 2.2.1. *Выборочное матожидание* — $\tilde{\mathbb{E}}\xi^* = \sum_{i=1}^n \frac{1}{n}X_i = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$.

Выборочное матожидание функции от ξ^* — $\tilde{\mathbb{E}}g\left(\xi^*\right) = \frac{1}{n}\sum_{i=1}^n g\left(X_i\right) = \overline{g(X)}$.

Определение 2.2.2. *Выборочная дисперсия* — $\tilde{\mathbb{D}}\xi^* = \sum_{i=1}^n \frac{1}{n}(X_i - \tilde{\mathbb{E}}\xi^*)^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = S^2$.

Определение 2.2.3. *Несмещённая выборочная дисперсия* — $S_0^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$.

Определение 2.2.4. *Выборочный момент k-го порядка* — $\tilde{\mathbb{E}}(\xi^*)^k = \sum_{i=1}^n \frac{1}{n} X_i^k = \frac{1}{n}\sum_{i=1}^n X_i^k = \overline{X^k}$.

Все вышеперечисленные характеристики являются случайными величинами как функции от выборки $X_1, \ldots, X_n$ и оценками для истинных моментов искомого распределения.

Теорема 2.2.1. *Выборочное среднее $\overline{X}$ является несмещённой, состоятельной и асимптотически нормальной оценкой для теоретического среднего (математического ожидания):*

1. *Если $\mathbb{E}|X_1| < \infty$, то $\mathbb{E}\overline{X} = \mathbb{E}X_1 = a$*

2. *Если $\mathbb{E}|X_1| < \infty$, то $\overline{X} \xrightarrow{\text{p}} \mathbb{E}X_1 = a$ при $n \to \infty$.*

3. *Если $\mathbb{D}X_1 < \infty, \quad \mathbb{D}X_1 \neq 0$, то $\sqrt{n}(\overline{X} - \mathbb{E}X_1) \Rightarrow \mathrm{N}_{0,\mathbb{D}X_1}$.*

Доказательство.

1. $\mathbb{E}\overline{X} = \frac{1}{n}(\mathbb{E}X_1 + \ldots + \mathbb{E}X_n) = \frac{1}{n} \cdot n\mathbb{E}X_1 = \mathbb{E}X_1 = a$

2. Из ЗБЧ в форме Хинчина:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{p}} \mathbb{E}X_1 = a$$

3. Из ЦПТ:

$$\sqrt{n}\left(\overline{X} - \mathbb{E}X_1\right) = \frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\mathbb{D}X_1}$$

Аналогичными свойствами обладает выборочный k-й момент $\overline{X^k}$. ■

Теорема 2.2.2. *Пусть $\mathbb{D}X_1 < \infty$.*

1. *Выборочные дисперсии S^2 и S_0^2 являются состоятельными оценками для истинной дисперсии:*

$$S^2 \xrightarrow{\text{p}} \mathbb{D}X_1 = \sigma^2, \quad S_0^2 \xrightarrow{\text{p}} \mathbb{D}X_1 = \sigma^2$$

2. *Величина* S^2 *— смещённая оценка дисперсии, а* S_0^2 *—– несмещённая:*

$$\mathbb{E}S^2 = \frac{n-1}{n}\mathbb{D}X_1 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad \mathbb{E}S_0^2 = \mathbb{D}X_1 = \sigma^2$$

3. *Если* $0 \neq \mathbb{D}(X_1 - \mathbb{E}X_1)^2 < \infty$, *то* S^2 *и* S_0^2 *являются асимптотически нормальными оценками истинной дисперсии:*

$$\sqrt{n}\left(S^2 - \mathbb{D}X_1\right) \Rightarrow \mathrm{N}_{0,\mathbb{D}(X_1 - \mathbb{E}X_1)^2}$$

Доказательство.

1. $S^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \overline{X^2} - (\overline{X})^2$

   Используя состоятельность первого и второго выборочных моментов и свойства сходимости по вероятности, получаем:

$$S^2 = \overline{X^2} - (\overline{X})^2 \xrightarrow{\mathrm{p}} \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \sigma^2$$

$$\frac{n}{n-1} \to 1 \Rightarrow S_0^2 = \frac{n}{n-1}S^2 \xrightarrow{\mathrm{p}} \sigma^2$$

2. Используя несмещённость первого и второго выборочных моментов:

   $\mathbb{E}S^2 = \mathbb{E}\left(\overline{X^2} - (\overline{X})^2\right) = \mathbb{E}\overline{X^2} - \mathbb{E}(\overline{X})^2 = \mathbb{E}X_1^2 - \mathbb{E}(\overline{X})^2 =$

$$= \mathbb{E}X_1^2 - \left((\mathbb{E}\overline{X})^2 + \mathbb{D}\overline{X}\right) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 - \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) =$$

$$= \sigma^2 - \frac{1}{n^2}n\mathbb{D}X_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

   Откуда следует:

   $\mathbb{E}S_0^2 = \frac{n}{n-1}\mathbb{E}S^2 = \sigma^2$

3. Введём случайные величины $Y_i = X_i - a$; $\mathbb{E}Y_i = 0, \mathbb{D}Y_1 = \mathbb{D}X_1 = \sigma^2$.

   $S^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - a - (\overline{X} - a)\right)^2 = \overline{Y^2} - (\overline{Y})^2$

   $\sqrt{n}\left(S^2 - \sigma^2\right) = \sqrt{n}\left(\overline{Y^2} - (\overline{Y})^2 - \sigma^2\right) = \sqrt{n}\left(\overline{Y^2} - \mathbb{E}Y_1^2\right) - \sqrt{n}(\overline{Y})^2 =$

$$= \frac{\sum_{i=1}^n Y_i^2 - n\mathbb{E}Y_1^2}{\sqrt{n}} - \overline{Y} \cdot \sqrt{n}\overline{Y} \Rightarrow \mathrm{N}_{0,\mathbb{D}(X_1 - a)^2}$$

   ...поскольку $\dfrac{\sum_{i=1}^n Y_i^2 - n\mathbb{E}Y_1^2}{\sqrt{n}} \Rightarrow \mathrm{N}_0, \mathbb{D}Y_1^2$ по ЦПТ, а $\overline{Y} \cdot \sqrt{n}\overline{Y} \Rightarrow 0$ как произведение последовательностей $\overline{Y} \xrightarrow[n\to\infty]{p} 0$ и $\sqrt{n}\overline{Y} \Rightarrow \mathrm{N}_{0,\mathbb{D}x_1}$.

■

2.3 Точечная оценка. Несмещенность, состоятельность, оптимальность. Теорема о единственности оптимальной оценки

Определение 2.3.1. *Статистика* или *оценка* — измеримая функция от выборки $T(X)$.

Определение 2.3.2. *Несмещённая оценка* функции $\tau(\theta)$ — статистика $T(X)$, т.ч. $\forall \theta \in \Theta : \mathbb{E}T(\theta) = \theta$.

Определение 2.3.3. *Асимптотически несмещённая оценка* функции $\tau(\theta)$ — статистика $T(X)$, т.ч. $\forall \theta \in \Theta : \mathbb{E}T(\theta) \xrightarrow[n\to\infty]{} \theta$.

Определение 2.3.4. *Состоятельная оценка* функции $\tau(\theta)$ — статистика $T(X)$, т.ч. $\forall \theta \in \Theta : T(\theta) \xrightarrow[n\to\infty]{p} \theta$.

Замечание. *Несмещённость означает отсутствие ошибки «в среднем», т. е. при систематическом использовании данной оценки. Несмещённость является желательным, но не обязательным свойством оценок. Достаточно, чтобы смещение оценки (разница между её средним значением и истинным параметром) уменьшалось с ростом объёма выборки. Поэтому асимптотическая несмещённость является весьма желательным свойством оценок. Свойство состоятельности означает, что последовательность оценок приближается к неизвестному параметру при увеличении количества наблюдений. В отсутствие этого свойства оценка совершенно «несостоятельна» как оценка.*

Пример. Дано распределение $\mathrm{Pois}(\theta), X = X_1$. Найти несмещённую оценку для функции $\tau(\theta) = \dfrac{1}{\theta}$.

$$\mathbb{E}T(\theta) = \sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta} \Rightarrow \sum_{x=0}^{\infty} T(x) \frac{\theta^{x+1}}{x!} = \sum_{x=0}^{\infty} \frac{\theta^x}{x!} \Rightarrow T(x) \equiv \frac{1}{\theta}$$

Т.к. полученная статистика зависит от θ, искомой несмещённой оценки для $\tau(\theta)$ не существует.

Пример. Дано распределение $\mathrm{Bi}(1, \theta), X = X_1$. Найти несмещённую оценку для параметра θ.

$$\mathbb{E}T(\theta) = \textstyle\sum_{x=0}^{\infty} T(x)\theta^x = \dfrac{\theta}{1-\theta} = \textstyle\sum_{r=1}^{\infty} \theta^r$$

$$T(x) = \begin{cases} 0, & \text{если } x = 0 \\ 1, & \text{если } x \geq 1 \end{cases}$$

Полученная оценка, очевидно, является бессмысленной.

Замечание. *Для несмещённой оценки $T(X)$ функции $\tau(\theta)$: $\mathbb{E}_\theta(T(X) - T(\theta))^2 = \mathbb{D}_\theta T(X)$. Т.к. для двух разных оценок $T_1(X)$, $T_2(X)$ соответствующие дисперсии могут быть несравнимыми, введём понятие оптимальной оценки.*

Определение 2.3.5. *Оптимальная оценка* параметра θ — статистика $\theta^* = \theta^*(X)$, т.ч.:

1. $T(X)$ — несмещённая.

2. $T(X)$ имеет равномерно минимальную дисперсию, т.е. для любой другой несмещённой оценки $T^*(X)$ функции $\tau(\theta)$: $\mathbb{D}_\theta T(X) \leq \mathbb{D}_\theta T_1(X)\ \forall X$.

Теорема 2.3.1. *Если существует оптимальная оценка функции $\tau(\theta)$, то она единственна.*

Доказательство. Предположим обратное: пусть существуют две оптимальныеоценки $T_1(X)$ и $T_2(X)$ функции $\tau(\theta)$. Тогда в силу их несмещённости: $\mathbb{E}_\theta T_1(X) = \mathbb{E}_\theta T_2(X) = T(\theta)$, а а в силу того, что они имеют равномерно минимальную дисперсию: $\mathbb{D}_\theta T_1(X) = \mathbb{D}_\theta T_2(X)\ \forall\theta$.

Введём новую статистику: $T_3(X) = \dfrac{T_1(X) + T_2(X)}{2}$.

Так как $\mathbb{E}_\theta T_3(X) = \dfrac{\mathbb{E}_\theta T_1(X) + \mathbb{E}_\theta T_2(X)}{2} = \tau(\theta)$, то $T_3(X)$ — несмещённая оценка функции $\tau(\theta)$.

Имеем также:
$$\mathbb{D}_\theta T_3(X) = \frac{\mathbb{D}_\theta\left(T_1(X) + T_2(X)\right)}{4} = \frac{\mathbb{D}_\theta T_1(X) + \mathbb{D}_\theta T_2(X) + 2\operatorname{cov}\left(T_1(X) T_2(X)\right)}{4}$$

В силу свойства $\mathbb{E}_\theta\xi^2 < \infty, \mathbb{E}_\theta\eta^2 < \infty \Rightarrow |\operatorname{cov}(\xi,\eta)| = |\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)| \leq \sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta}$, где равенство достигается тогда и только тогда, когда $\xi = a\eta + b$, получаем:
$$\mathbb{D}_\theta T_3(X) \leq \frac{\mathbb{D}_\theta T_1(X) + \mathbb{D}_\theta T_2(X) + 2\sqrt{\mathbb{D}_\theta T_1(X)}\sqrt{\mathbb{D}_\theta T_2(X))}}{4} = \mathbb{D}_\theta T_1(X)$$

В силу того, что $T_1(X)$ и $T_2(X)$ — оптимальные, дисперсия $T_3(X)$ не может быть меньше дисперсии $T_1(X)$, следовательно, справедливо равенство, достигаемое при следующих условиях:

$T_1(X) = aT_2(X) + b \Rightarrow \mathbb{E}T_1(X) = a\mathbb{E}T_2(X) + b \Leftrightarrow T(\theta) = aT(\theta) + b \forall\theta \Rightarrow a = 1, b = 0$ ■

2.4 Функция правдоподобия. Достаточные статистики, полные статистики. Теорема факторизации

В зависимости от типа распределения $\mathcal{F}_\theta$ обозначим через $f_\theta(y)$ одну из следующих функций:

$$f_\theta(y) = \begin{cases} \text{плотность } f_\theta(y), & \text{если } \mathcal{F}_\theta \text{ абсолютно непрерывно,} \\ P_\theta\left(X_1 = y\right), & \text{если } \mathcal{F}_\theta \text{ дискретно.} \end{cases}$$

Определение 2.4.1. *Функция правдоподобия* выборки $\overline{X}$ — $L(\vec{X};\theta) = f_\theta\left(X_1\right) \cdot f_\theta\left(X_2\right) \cdot \ldots \cdot f_\theta\left(X_n\right) = \prod_{i=1}^n f_\theta\left(X_i\right)$.

Замечание. *В дискретном случае функция правдоподобия принимает вид:*
$L(\vec{x};\theta) = \prod_{i=1}^n f_\theta(x_i) = \mathrm{P}_\theta(X_1 = x_1) \cdot \ldots \cdot \mathrm{P}_\theta(X_n = x_n) = \mathrm{P}_\theta(X_1 = x_1, \ldots, X_n = x_n)$

Таким образом, смысл функции правдоподобия — вероятность попасть в заданную точку при соответствующем параметре θ в дискретном случае; для абсолютно непрерывного аналогично — вероятность попасть в куб с центром в $x_1, \ldots, x_n$ и сторонами $dx_1, \ldots, dx_n$.

Определение 2.4.2. *Достаточная статистика* для параметра θ — статистика $T(X)$, т.ч. $\forall t$, $\forall B \in \mathfrak{B}(\mathbb{R}^n)$ условное распределение $\mathbb{P}(X_1, \ldots, X_n \in B \mid T = t)$ не зависит от параметра θ.

Иными словами, если значение статистики T известно и фиксировано, то даже знание её распределения больше не даёт никакой информации о параметре; достаточно лишь вычислить T по выборке.

Теорема 2.4.1. *Критерий факторизации: $T(X)$ — достаточная статистика $\Leftrightarrow$ её функция правдоподобия представима в виде $L(X_1, \ldots, X_n;\theta) \overset{п.н.}{=} h(\vec{X}) \cdot \Psi(S, \theta)$*

Доказательство. Рассмотрим только дискретный случай. Пусть $T(X)$ — достаточная статистика. Тогда:
$f_\theta(x) = p_\theta(X = x) = \mathrm{P}_\theta(X = x, T(X) = T(x)) = \underbrace{\mathrm{P}_\theta(X = x | T(X) = T(x)}_{h(x)} \underbrace{\mathrm{P}_\theta(T(X) = T(x))}_{g(\theta, T(X))}$

Пусть, наоборот, $\mathrm{P}_\theta(X = x) = h(x) g_\theta(T(x))$. Тогда:

$$\mathrm{P}_\theta(X = x | T(X) = 1) = \begin{cases} 0, & T(X) \neq 1 \\ \mathrm{P}_\theta(X = x | T(X) = T(x)), & T(X) = 1 \end{cases}$$

...откуда следует:

$$\mathrm{P}_\theta(X=x|T(X)=T(x)) = \frac{\mathrm{P}_\theta(X=x, T(X)=T(x))}{\mathrm{P}(T(X)=T(x))} =$$
$$= \frac{\mathrm{P}_\theta(X=x, T(X)=T(x))}{\sum_{y:T(y)=T(x)} \mathrm{P}_\theta(T(X)=T(x), X=y)} = \frac{\mathrm{P}_\theta(X=x)}{\sum_{y:T(y)=T(x)} \mathrm{P}_\theta(x=y)} = \quad \blacksquare$$
$$\frac{h(x)g_\theta(T(x))}{\sum_{y:T(y)=T(x)} h(y)g_\theta(T(x))} = \frac{h(x)}{\sum_{y:T(y)=T(x)} h(y)}$$

Определение 2.4.3. *Полная статистика* для параметра θ — статистика $T(X)$, т.ч. $\mathbb{E}g(T) = 0 \ \forall\theta \in \Theta \Rightarrow g(T) \overset{\text{п.н.}}{=} 0$

2.5 Неравенство Рао—Крамера. Эффективные оценки

Пусть $X_1, \ldots, X_n$ — некоторая выборка с функцией правдоподобия $L(X, \theta)$ относительно некоторой меры μ. Введём функцию $\varphi(\theta) = \int_{\mathbf{R}^n} T(x)L(x,\theta)\mu(dx) < \infty$, в дальнейшем считая, что она дифференцируема необходимое число раз.

Определение 2.5.1. Функция правдоподобия $L(X, \theta)$ *удовлетворяет условиям регулярности для m-й производной*, если существует $\dfrac{d^m\phi(\theta)}{d\theta^m} = \int_{\mathbb{R}^n} T(x)\dfrac{\partial^m L(x,\theta)}{\partial\theta^m}\mu(dx)$, причём множество $\{x \mid L(x,\theta) > 0\}$ не зависит от параметра θ.

Теорема 2.5.1. *Неравенство Рао-Крамера: Пусть $X_1, \ldots, X_n$ — выборка, $L(X, \theta)$ удовлетворяет условиям регулярности для первой производной и $\tau(\theta)$ — дифференцируемая функция θ. Тогда:*

1. *$\forall$ $T(X)$, — несмещённой оценки функции $\tau(\theta)$, справедливо неравенство:*
   $$\mathbf{D}_\theta T(X) \geq \frac{(\tau'(\theta))^2}{\mathbf{E}_\theta U^2(X,\theta)} \ \forall\theta \in \Theta,$$
   где $U(X,\theta) = \dfrac{\partial \ln L(X,\theta)}{\partial\theta}$ (функция вклада)
2. *Равенство достигается $\Leftrightarrow \exists\ a_n(\theta):\ T(X) - \tau(\theta) = a_n(\theta) \cdot U(X,\theta)$*

Доказательство. $\int L(x,\theta)\mu(dx) = 1 \Rightarrow \int \dfrac{\partial L(x,\theta)}{\partial\theta}\mu(dx) = 0$
Из условий регулярности $L(X, \theta)$ для следует:

$\int T(x)L(x,\theta)\mu(dx) = \mathbf{E}_\theta T(X) = T(\theta) \Rightarrow \int T(x)\dfrac{\partial L(x,\theta)}{\partial \theta}\mu(dx) = \tau'(\theta)$
Заметим, что

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

Откуда следует:

$$\int U(x,\theta)L(x,\theta)\mu(dx) = 0 \Leftrightarrow \mathbf{E}_\theta U(X,\theta) = 0$$

$$\int T(x)U(x,\theta)L(x,\theta)\mu(dx) = \tau'(\theta) \Leftrightarrow \mathbf{E}_\theta T(X)U(X,\theta) = \tau'(\theta)$$

Вычитая из первого равенства, помноженного на $\tau(\theta)$, второе, получаем:

$$\mathbf{E}_\theta(T(X) - T(\theta))U(X,\theta) = \tau'(\theta)$$

В левой части полученного равенства стоит ковариация случайных величин $T(X)$ и $U(X,\theta)$: $\mathrm{cov}_\theta(T(X), U(X,\theta)) = T'(\theta)$

Из неравенства Коши-Буняковского:

$$(\tau'(\theta))^2 = \mathrm{cov}^2_\theta(T(X)U(X,\theta)) \le \mathbf{D}_\theta T(X)\mathbf{D}_\theta U(X,\theta) = \mathbf{D}_\theta T(X)\mathbf{E}_\theta U^2(X,\theta)$$

...что равносильно п.1 теоремы:

$$\mathbf{D}_\theta T(X) \ge \frac{[T'(\theta)]^2}{\mathbf{E}_\theta U^2(X,\theta)}$$

Неравенство достигается, если линейно связаны:

$$T(X) = \varphi(\theta)U(X,\theta) + \psi(\theta) \Rightarrow T(\theta) = \psi(\theta) \Rightarrow a_n(\theta) = \varphi(\theta)$$

■

Определение 2.5.2. *Эффективная оценка* — оценка $T(X)$, для которой в неравенстве Рао-Крамера достигается равенство.

Замечание. *Если существует эффективная оценка для функции $\tau(\theta)$, то ни для какой другой функции от θ, кроме линейного преобразования $\tau(\theta)$, эффективной оценки существовать не будет.*

2.6 Теорема Рао—Блекуэлла—Колмогорова. Оптимальность оценок являющихся функцией полной достаточной статистики

Теорема 2.6.1. *Теорема Рао—Блекуэлла—Колмогорова: Если оптимальная оценка функции $\tau(\theta)$ существует, то она является функцией от достаточной статистики.*

Доказательство. В доказательстве используются следующие свойства условного матожидания:

$\mathbb{E}f(x,z) = \mathbb{E}(\mathbb{E}(f(x,z)|z)), \mathbb{E}(g(z)|z) = g(z)$

Пусть $T(X)$ — достаточная статистика, $T_1(X)$ — несмещённая оценка функции $\tau(\theta)$, т.е. $\mathbb{E}T_1(X) = \tau(\theta)$. Рассмотрим функцию $H(T) = \mathbb{E}\left(T_1|T\right)$. Тогда из первого свойства следует:

$\mathbb{E}H(T) = \mathbb{E}\left(\mathbb{E}\left(T_1|T\right)\right) = \mathbb{E}T_1 = \tau(\theta) \Rightarrow H(T)$ — несмещённая оценка $\tau(\theta)$.

Докажем равномерную минимальность её дисперсии:

$\mathbb{E}((T_1 - H(T))(H(T) - \tau(\theta)) = \mathbb{E}(\mathbb{E}((T_1 - H(T))(H(T) - \tau(\theta))|T)) = \mathbb{E}((H(T) - H(T))(H(T) - \tau(\theta))) = 0$

Таким образом, $H(T)$ — оптимальная оценка $\tau(\theta)$. ■

Теорема 2.6.2. *Теорема Колмогорова: Если $T(X)$ — полная достаточная статистика, то она является оптимальной оценкой своего математического ожидания.*

Доказательство. Докажем, что $T(X)$ является единственной несмещенной оценкой для $\mathbb{E}T(X)$. Тогда $T(X)$ будет оптимальной оценкой. Предположим, что $T_1(X)$ — оптимальная оценка для $\mathbb{E}T(X)$. Из теоремы Рао-Блекуэлла-Колмогорова получаем, что $T_1 = H(T)$ и $\mathrm{E}T_1 = \mathrm{E}T$. Тогда:

$\mathrm{E}\underbrace{(T(\mathbf{X}) - H(T(\mathbf{X})))}_{\varphi(T)} = 0$

Из условия полноты $T(X)$ следует, что $\varphi(T) = 0$ с вероятностью 1, т.е. $T = H(T)$ с вероятностью 1. ■

2.7 Метод моментов. Свойства оценок, полученных методом моментов

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$. Выберем функцию $g(y) : \mathbb{R} \to \mathbb{R}$ так, чтобы существовал момент $\mathbb{E}g\left(X_1\right) = h(\theta)$ и функция $h(\theta)$ была обратима на Θ. Разрешим полученное уравнение относительно θ, а затем вместо истинного момента возьмём выборочный:

$\theta = h^{-1}\left(\mathbb{E}g\left(X_1\right)\right), \quad \theta^* = h^{-1}(\overline{g(X)}) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n g\left(X_i\right)\right)$

Полученная оценка θ^* — *оценка метода моментов* для параметра θ. Чаще всего берут $g(y) = y^k$. В этом случае, при условии обратимости функции h на Ω:

$\mathbb{E}X_1^k = h(\theta), \quad \theta = h^{-1}\left(\mathbb{E}X_1^k\right), \quad \theta^* = h^{-1}(\overline{X^k}) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^k\right)$

Теорема 2.7.1. *Пусть $\theta^* = h^{-1}(\overline{g(X)})$ — оценка параметра θ, полученная методом моментов, причём функция h^{-1} непрерывна. Тогда оценка θ^* состоятельна.*

Доказательство. По ЗБЧ Хинчина имеем:
$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^n g\left(X_i\right) \xrightarrow{\text{p}} \mathbb{E}g\left(X_1\right) = h(\theta)$
Ввиду непрерывности функции h^{-1}:
$\theta^* = h^{-1}(\overline{g(X)}) \xrightarrow{\text{p}} h^{-1}\left(\mathbb{E}g\left(X_1\right)\right) = h^{-1}(h(\theta)) = \theta$ ■

Определение 2.7.1. *Асимптотически нормальная оценка* параметра θ с коэффициентом $\sigma^2(\theta)$ — оценка θ^*, т.ч. при $n \to \infty$ имеет место слабая сходимость к стандартному нормальному распределению: $\sqrt{n}(\theta^* - \theta) \Rightarrow \mathrm{N}_{0,\sigma^2(\theta)}$.

Лемма 2.7.2. *Пусть функция $g(y)$ такова, что $0 \neq \mathrm{D}g\left(X_1\right) < \infty$. Тогда статистика $\overline{g(X)}$ является асимптотически нормальной оценкой для $\mathrm{E}g\left(X_1\right)$ с коэффициентом $\sigma^2(\theta) = \mathrm{D}g\left(X_1\right)$:*

$$\sqrt{n}\frac{\overline{g(X)} - \mathrm{E}g\left(X_1\right)}{\sqrt{\mathrm{D}g\left(X_1\right)}} \Rightarrow \mathrm{N}_{0,1}$$

Доказательство. Следует непосредственно из ЦПТ. ■

Замечание. *Следующая теорема утверждает асимптотическую нормальность оценок вида*

$$\theta^* = H(\overline{g(X)}) = H\left(\frac{g\left(X_1\right) + \ldots + g\left(X_n\right)}{n}\right),$$

которые обычно получаются при использовании метода моментов, при этом всегда $\theta = H\left(\mathrm{E}g\left(X_1\right)\right)$.

Теорема 2.7.3. *Пусть функция $g(y)$ такова, что $0 \neq \mathrm{D}g\left(X_1\right) < \infty$, функция $H(y)$ дифференцируема в точке $a = \mathrm{E}g\left(X_1\right)$ и её производная в этой точке $H'(a) = H'(y)\big|_{y=a}$ отлична от нуля. Тогда оценка $\theta^* = H(\overline{g(X)})$ является асимптотически нормальной оценкой для параметра $\theta = H\left(\mathrm{E}g\left(X_1\right)\right) = H(a)$ с коэффициентом асимптотической нормальности $\sigma^2(\theta) = \left(H'(a)\right)^2 \cdot Dg\left(X_1\right)$.*

Доказательство. Согласно ЗБЧ последовательность $\overline{g(X)}$ стремится к $a = \mathrm{E}g\left(X_1\right)$ по вероятности с ростом n: Функция

$$G(y) = \begin{cases} \dfrac{H(y) - H(a)}{y - a}, & y \neq a \\ H'(a), & y = a \end{cases}$$

по условию непрерывна в точке a: Поскольку сходимость по веро- ятности сохраняется под действием непрерывной функции, получим, что $G(\overline{g(X)}) \xrightarrow{\text{p}} G(a) = H'(a)$.

Заметим также, что по вышеприведённой лемме величина $\sqrt{n}(\overline{g(X)} - a)$ слабо сходится к нормальному распределению $\mathrm{N}_{0,\mathrm{D}g}(X_1)$: Пусть ξ — случайная величина из этого распределения. Тогда

$$\sqrt{n}(H(\overline{g(X)}) - H(a)) = \sqrt{n}(\overline{g(X)} - a) \cdot G(\overline{g(X)}) \Rightarrow \xi \cdot H'(a)$$

Мы использовали следующее свойство слабой сходимости: если $\xi_n \Rightarrow \xi$ и $\eta_n \xrightarrow{\text{p}} c = \text{const}$, то $\xi_n \eta_n \Rightarrow c\xi$. Но распределение случайной величины $\xi \cdot H'(a)$ есть $\mathrm{N}_{0,(H'(a))^2 \cdot \mathrm{D}g(X_1)}$, откуда следует

$$\sigma^2(\theta) = (H'(a))^2 \cdot \mathrm{D}g\,(X_1). \qquad \blacksquare$$

2.8 Метод максимального правдоподобия. Свойства оценок максимального правдоподобия

Определение 2.8.1. *Оценка максимального правдоподобия $\hat{\theta}$ параметра θ —* точка параметрического множества Θ, в которой функция правдоподобия $L(X, \theta)$ при заданном X достигает максимума, т.е.:

$$L(\boldsymbol{x}, \hat{\theta}) = \sup_{\theta \in \Theta} L(\boldsymbol{x}, \theta)$$

Замечание. *Поскольку функция $\ln y$ монотонна, то точки максимума функций $L(X, \theta)$ и $lnL(X, \theta)$ совпадают.*

Если для каждого X максимум функции правдоподобия достигается во внутренней точке Θ, и $L(X, \theta)$ дифференцируема по θ, то оценка максимального правдопо-добия $\hat{\theta}$ удовлетворяет уравнению:

$$\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta} = 0$$

Если θ — векторный параметр: $\theta = (\theta_1, \ldots, \theta_n)$, то это уравнение заменяется системой уравнений:

$$\frac{\partial \ln L_n(\mathbf{x}, \theta)}{\partial \theta_i} = 0, \quad i = 1, \ldots, n$$

Теорема 2.8.1. *Если существует эффективная оценка $T(X)$ скалярного параметра θ, то она совпадает с оценкой максимального правдоподобия.*

Доказательство. Если оценка $T(X)$ скалярного параметра θ эффективна, то в неравенстве Рао-Крамера достигается равенство:

$U(X,\theta) = \dfrac{\partial \ln L_n(\mathbf{x},\theta)}{\partial\theta} = \dfrac{T(X)-\theta}{a_n(\theta)}$ ■

Теорема 2.8.2. *Если $T(X)$ достаточная статистика, а оценка максимального правдоподобия $\hat{\theta}$ существует и единственна, то она является функцией от $T(X)$.*

Доказательство. Из критерия факторизации следует, что если T=T(X) достаточная статистика, то имеет место представление:

$L(X,\theta) = g(T(X),\theta)h(X)$

Таким образом, максимизации $L(X,\theta)$ сводится к максимизации $g(T(X),\theta)$ по θ, Следовательно $\hat{\theta}$ есть функция от $T(X)$. ■

Добавить асимптотическую нормальность и эффективность + Чернова стр 39 теорема для состоятельности.

2.9 Интервальное оценивание. Методы центральной статистики и использования точечной оценки

Определение 2.9.1. *Доверительный интервал* для параметра θ с коэффициентом доверия $0 \le \alpha \le 1$ — интервал $(T_1(X), T_2(X))$, т.ч. $\mathbb{P}_\theta(T_1(X) < \theta < T_2(X)) \ge \alpha$.

Пример. Пусть $X_1, \ldots, X_n$ — выборка из $N(\theta, 1)$. Тогда

$$\theta^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\theta, \frac{1}{n}\right) \Rightarrow (\overline{X} - \theta)\sqrt{n} \sim N(0,1)$$

Для величины, имеющей стандартное нормальное распределение, строим доверительный интервал, т.е. находим такое $t_{\alpha/2}$, что

$$\mathrm{P}_\theta\left(|(\bar{X}-\theta)\sqrt{n}| < t_{\alpha/2}\right) = \alpha$$

Решаем уравнение относительно θ и получаем

$$\mathrm{P}_\theta\left(\bar{X} - \frac{t_{\alpha/2}}{\sqrt{n}} < \theta < \bar{X} + \frac{t_{\alpha/2}}{\sqrt{n}}\right) = \alpha$$

Определение 2.9.2. *Центральная статистика* — функция $G(X,\theta)$, т.ч.:

1. $G(X, \theta)$ непрерывна и строго монотонна по θ при любом фиксированном X.

2. $\mathbf{P}_\theta(G(X, \theta) < t) = F(t)$ непрерывна и не зависит от θ.

Замечание. *Формально определённая выше величина не является статистикой, т.к. зависит от неизвестного параметра θ.*

Построение доверительного интервала с помощью центральной статистики:

1. Зафиксируем $\alpha_1, \alpha_2 \in \mathbf{R}$, т.ч.

$$\mathbf{P}_\theta(\alpha_1 \leq G_1(X, \theta) \leq \alpha_2) = \alpha \ \forall\theta \Leftrightarrow F(\alpha_2) - F(\alpha_1) = \alpha$$

2. Пусть $G(X, \theta)$ возрастает. Из условий

$$\begin{cases} G(X, \theta) \leq \alpha_2 \\ G(X, \theta) \geq \alpha_1 \end{cases}$$

находятся статистики

$$\begin{cases} T_2(X) : G\left(X, T_2(X)\right) = \alpha_2, \\ T_1(X) : G\left(X, T_1(X)\right) = \alpha_1, \end{cases} \quad T_1(X) \leq \theta \leq T_2(X)$$

откуда $\mathbf{P}_\theta\left(T_1(X) \leq \theta \leq T_2(X)\right) \geq \alpha \ \forall\theta$.

Определение 2.9.3. *Центральный доверительный предел* для параметра θ с коэффициентом доверия $0 \leq \alpha \leq 1$ — интервал $(T_1(X), T_2(X))$, т.ч.

$$\mathbf{P}_\theta\left(T_1(X) > \theta\right) = \frac{1-\alpha}{2}$$
$$\mathbf{P}_\theta\left(T_2(X) < \theta\right) = \frac{1-\alpha}{2}$$

Построение доверительного интервала с помощью точечной оценки:

2.10 Проверка гипотез. Лемма Неймана—Пирсона

Определение 2.10.1. *Гипотеза H* — любое предположение о распределении наблюдаемой случайной величины: $H = \{\mathcal{F} = \mathcal{F}_1\}$ или $H = \{\mathcal{F} \in \mathbb{F}\}$, где $\mathbb{F}$ — некоторое подмножество в множестве всех распределений. Гипотеза называется *простой* в первом случае, *сложной* во втором. Если гипотез всего две, то одну из них принято называть *основной*, а другую — *альтернативой*.

Замечание. *Типичные задачи проверки гипотез:*

1. *Гипотезы о виде распределения;*

2. *Гипотезы о проверке однородности выборки: дано несколько выборок; основная гипотезасостоит в том, что эти выборки извлечены из одного распределения;*

3. *Гипотеза независимости: по выборке* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *из* n *независимых наблюдений пары случайных величин проверяется гипотеза* $H_1 = \{X_i \text{ и } Y_i \text{ независимы } \}$ *при альтернативе* $H_2 = \{H_1 \text{ неверна } \}$. *Обе гипотезы являются сложными;*

4. *Гипотеза случайности: в эксперименте наблюдаются* n *случайныхвеличин* $X_1, \ldots, X_n$ *и проверяется сложная гипотеза* $H_1 = \{X_1, \ldots, X_n \text{ независимы и одинаково распределены}\}$

Пусть дана выборка $X_1, \ldots, X_n$, относительно распределения которой выдвинуты гипотезы $H_1, \ldots, H_k$.

Определение 2.10.2. *Критерий* $\varphi = \varphi\,(X_1, \ldots, X_n)$ — измеримое отображение $\varphi : \mathbb{R}^n \to \{H_1, \ldots, H_k\}$ из множества всех возможных значений выборки в множество гипотез.

Замечание. *В случае одной основной гипотезы и одной альтернативы критерий задаёт критическую область* $S \in \mathbb{R}^n$ *для гипотезы* H_0 — $\varphi(x) = I(x \in S)$; *тогда правило проверки гипотезы* H_0 *при альтернативе* H_1 *можно сформулировать следующий образом:*

- $\varphi(x) = 1 \Rightarrow$ *отвергаем* H_0, *принимаем* H_1;

- $\varphi(x) = 0 \Rightarrow$ *отвергаем* H_1, *принимаем* H_0;

Определение 2.10.3. Говорят, что произошла *ошибка i-го рода* критерия δ, если критерий отверг верную гипотезу H_i. *Вероятностью ошибки i-го рода* — число $\alpha_i(\varphi) = P_{H_i}\left(\delta(\vec{X}) \neq H_i\right)$.

Истинная гипотеза	Результат принятия решения	
	H_0 отклонена	H_0 принята
H_0	α	$1-\alpha$
H_1	$1-\beta$	β