Теоретические материалы для проекта

Лоптев Сергей

1 декабря 2020 г.

Содержание

Часть I

Теория вероятностей

1 Дискретные распределения

1.1 Вероятность

Пусть задано некоторое множество возможных исходов (эксперимента) $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. Это множество Ω называют **множеством элементарных исходов**. Всякое подмножество $A \subset \Omega$ называют **событием**. Функцию $\mathsf{P} : 2^\Omega \to [0, 1]$, удовлетворяющую следующим свойствам:

(i) $\mathsf{P}(\Omega) = 1$,

(ii) $A \cup B = \varnothing \implies \mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$ (правило суммы или аддитивность),

называют **вероятностной мерой**, а значение $\mathsf{P}(A)$ **вероятностью** события A.

1.2 Распределения

Распределение Бернулли Говорят, что случайная величина I_A обладает биномиальным распределением, если для некоторого события A и элементарного исхода ω

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

Биномиальное распределение Говорят, что случайная величина X обладает биномиальным распределением с параметром $p \in [0, 1]$, если для некоторого N

$$\forall k \in \{0, 1, \dots, N\} \;\; \mathsf{P}(X = n) = \mathrm{C}_N^k \, p^k \, (1 - p)^{N-k}$$

Геометрическое распределение Говорят, что случайная величина X обладает геометрическим распределением с параметром $p \in [0, 1]$, если

$$\forall n \in \mathbb{N} \;\; \mathsf{P}(X = n) = p \, (1 - p)^{n-1}$$

Распределение Пуассона Говорят, что случайная величина X обладает распределением Пуассона с параметром λ ($X \sim \mathrm{Pois}(\lambda)$), если

$$\forall n \in \mathbb{N} \cup \{0\} \;\; \mathsf{P}(X = n) = \frac{\lambda}{n!} \cdot e^{-\lambda}$$

1.3 Совместные распределения случайных величин

Пусть X, Y — две случайные величины на дискретном вероятностном пространстве с множествами (различных) значений $\{x_1, \dots, x_k, \dots\}$ и $\{y_1, \dots, y_k, \dots\}$ соответственно. Их **совместным распределением** называется вероятностная мера $\mu_{(X,Y)}$ на вероятностном пространстве всех пар (x_j, y_k), для которой

$$\mu_{(X,Y)}\left(\{(x_j, y_k)\}\right) = \mathsf{P}(\omega : \; X(\omega) = x_j, \; Y(\omega) = y_k) = \mathsf{P}(\{\omega : \; X(\omega) = x_j\} \cap \{\omega : \; Y(\omega) = y_k\})$$

1.4 Таблица совместного распределения и частное распределение

Совместное распределение двух случайных величин X, Y с множествами значений $\{x_1, \ldots, x_k, \ldots\}$ и $\{y_1, \ldots, y_k, \ldots\}$ однозначно задаётся **таблицей совместного распределения**, у которой в ячейке i, j стоит $\mathsf{P}(\omega : X(\omega) = x_i,\ Y(\omega) = y_k)$. Из такой таблицы мы можем получить **частное распределение** случайной величины X:

$$P(\omega : X(\omega) = x_j) = \sum_{k=1}^{\infty} P(\omega : X(\omega) = x_j, Y(\omega) = y_k)$$

Аналогично можно получить частное распределение случайной величины Y.

1.5 Независимость случайных величин и некореллированность случайных величин

Случайные величины X, Y с множествами значений $\{x_1, \ldots, x_k, \ldots\}$ и $\{y_1, \ldots, y_k, \ldots\}$ соответственно называются **независимыми**, если

$$\mu_{(X,Y)}\left(\{(x_j, y_k)\}\right) = \mu_X\left(\{x_j\}\right) \cdot \mu_Y\left(\{y_k\}\right) \quad \forall k, j$$

или другими словами

$$\mathsf{P}(\omega :\ X(\omega) = x_j,\ Y(\omega) = y_k) = \mathsf{P}(\omega :\ X(\omega) = x_j) \cdot \mathsf{P}(\omega :\ Y(\omega) = y_k) \quad \forall k, j$$

Аналогично определяется независимость трёх и более случайных величин.

Ковариацией пары случайных величин называется число

$$\operatorname{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}\, X)\,(Y - \mathbb{E}\, Y)\right]$$

Заметим, что ковариация является неотрицательно определённой билинейной формой,

$$\operatorname{cov}(X, Y) = \mathbb{E}\left[X \cdot Y\right] - \left[\mathbb{E}\, X\right] \cdot \left[\mathbb{E}\, Y\right]$$

в частности $\mathbb{D}\, X = \operatorname{cov}(X, X) = \mathbb{E}\left[X^2\right] - \left[\mathbb{E}\, X\right]^2$.

Величину

$$r(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathbb{D}\, X}\sqrt{\mathbb{D}\, Y}}$$

называют **коэффициентом корреляции**.

Две случайные величины X, Y называются **некоррелированными**, если $r(X, Y) = 0$.

Заметим, что если две случайные величины X, Y независимы, то они некоррелированы. Обратное, вообще говоря, неверно.

1.6 Моменты распределений

Число $\mathbb{E}\, \xi^k$ называется **моментом порядка** k или k**-ым моментом случайной величины** ξ.

2 Непрерывные распределения

2.1 Плотность распределения

Предположим, что определена вероятностная мера μ. Функция

$$F(t) = \mu((-\infty, t])$$

называется **функцией распределения** меры μ.

Из определения F следует, что $\mu((a, b]) = F(b) - F(a)$.

Функция F удовлетворяет следующим свойствам:

(i) $F : \mathbb{R} \to [0, 1]$ не убывает;

(ii) F непрерывна справа;

(iii) $\lim\limits_{t \to -\infty} F(t) = 0$ и $\lim\limits_{t \to +\infty} F(t) = 1$.

Во многих случаях задавать вероятностную меру на числовой прямой удобно плотностью.

Пусть ϱ — неотрицательная и интегрируемая (по Риману) функция на $\mathbb{R}$, причём

$$\int\limits_{-\infty}^{+\infty} \varrho(x) dx = 1$$

Если функция распределения F меры μ задаётся неравенством

$$F(t) = \int\limits_{-\infty}^{t} \varrho(x) dx$$

то говорят, что вероятностная мера μ задана **плотностью** ϱ. В этом случае

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

На самом деле можно доказать, что

$$\mu(A) = \int\limits_{A} \varrho dx$$

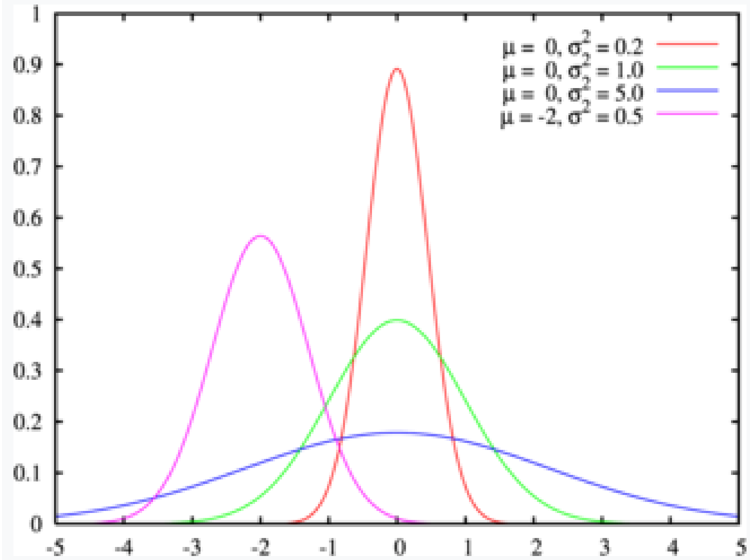
для всякого множества A, для которого имеет смысл интеграл в правой части.

2.2 Распределения

Нормальное распределение Нормальным распределением с параметрами μ и σ^2 называют вероятностную меру на числовой прямой, заданную плотностью

$$\varrho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x-\mu}{2\sigma^2}}$$

Случайную величину с нормальным распределением обозначают как $X \sim \mathcal{N}(\mu, \sigma^2)$.



Зеленая линия соответствует стандартному нормальному распределению

Определение. *Введём понятие **Гамма-функции**.*

$$\Gamma(z) = \lim_{n\to+\infty} \frac{(n-1)!n^z}{z(z+1)(z+2)\cdots(z+n-1)}, \quad z \in \mathbb{C} \setminus \{0,\ -1,\ -2, \ldots\}$$

Распределение Стьюдента Пусть $Y_0,\ Y_1,\ \ldots,\ Y_n$ — независимые стандартные нормальные случайные величины, такие что $Y_i \sim \mathcal{N}(0,1),\ i = 0, \ldots, n$. Тогда распределение случайной величины t, где

$$t = \frac{Y_0}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} Y_i^2}},$$

называется **распределением Стьюдента** с n степенями свободы $t \sim t(n)$.

Это распределение абсолютно непрерывно с плотностью:

$$\varrho_t(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}},$$

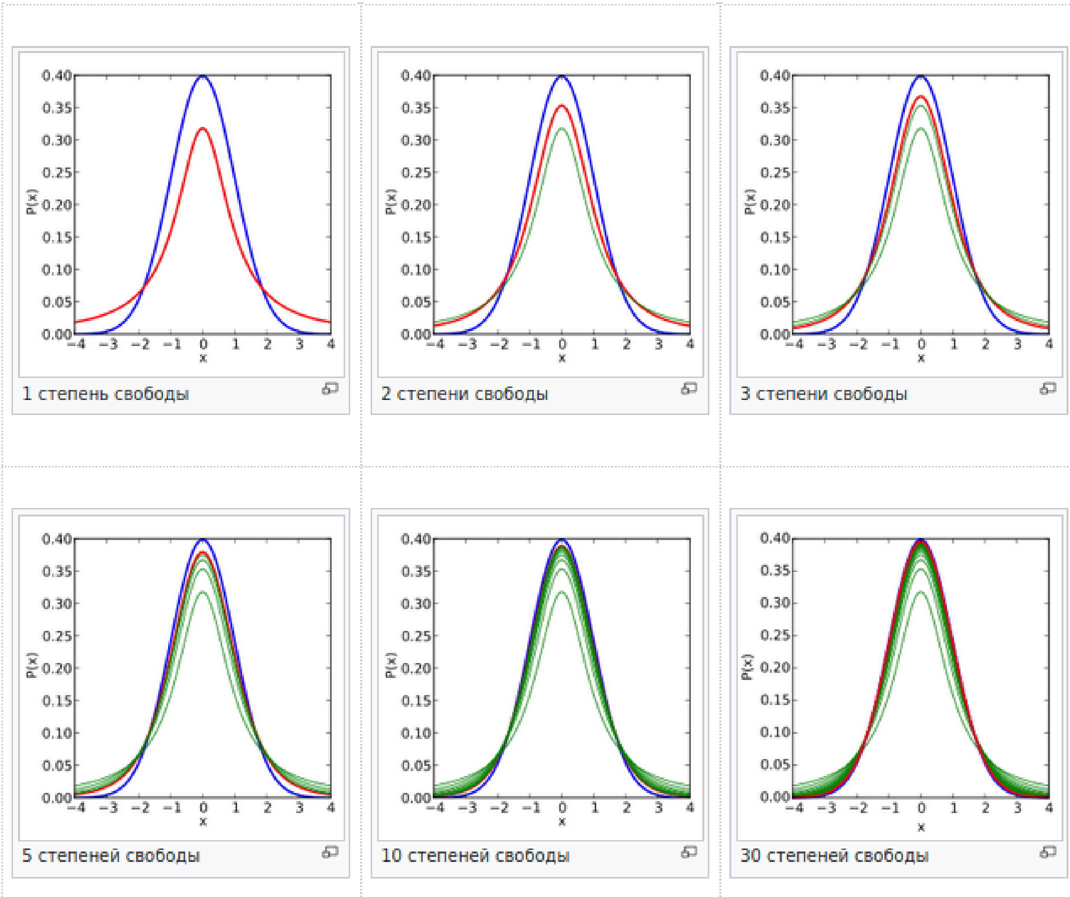
где Γ — гамма-функция. Таким образом:

$$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} = \frac{(n-1)(n-3)\cdots 5\cdot 3}{2\sqrt{n}(n-2)(n-4)\cdots 4\cdot 2}, \text{ для чётных } n$$

и соответственно

$$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} = \frac{(n-1)(n-3)\cdots 4\cdot 2}{\pi\sqrt{n}(n-2)(n-4)\cdots 5\cdot 3}, \text{ для нечётных } n.$$

Плотность t-распределения (красная линия) для 1, 2, 3, 5, 10 и 30 степеней свободы
в сравнении со стандартным нормальным распределением (синяя линия). Предыдущие графики показаны зеленым.



Распределение хи-квадрат Пусть $z_1, \ldots, z_k$ — совместно независимые стандартные нормальный случайные величины, то есть $z_i \sim \mathcal{N}(0,1)$. Тогда случайная величина

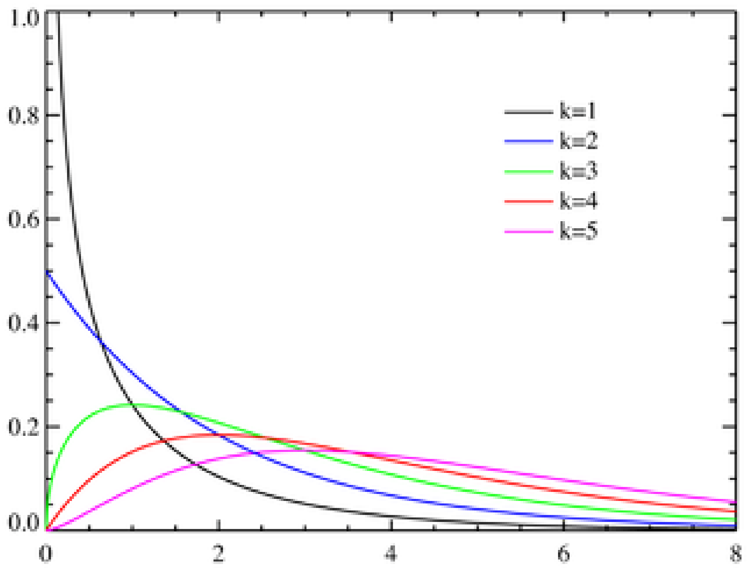
$$x = z_1^2 + \ldots + z_k^2$$

имеет **распределение хи-квадрат** с k степенями свободы, то есть $x \sim \varrho_{\chi^2(k)}(x)$ или, если записать по-другому:

$$x = \sum_{i=1}^{k} z_i^2 \sim \chi^2(k)$$

Плотность распределения хи-квадрат имеет вид:

$$\varrho_{\chi^2(k)}(x) = \frac{(1/2)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\, x^{\frac{k}{2}-1}\, e^{-\frac{x}{2}}$$



Плотность вероятности

Определение. *Введём понятие* ***Бета-функции****.*

Формула через Гамма-функции:

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

где $\Gamma(x)$ *— Гамма-функция.*

Формула через нисходящий факториал:

$$\mathrm{B}(x,y) = \frac{1}{y}\sum_{n=0}^{\infty}(-1)^n \frac{(y)_{n+1}}{n!\,(x+n)},$$

где $(x)_n$ *— нисходящий факториал, равный* $x \cdot (x-1) \cdot (x-2) \cdot \ldots \cdot (x-n+1)$.

Распределение Фишера Пусть $Y_1,\ Y_2$ — две независимые случайные величины, имеющие распределение хи-квадрат: $Y_i \sim \chi^2(d_i)$, где $d_i \in \mathbb{N},\ i = 1, 2$. Тогда распределение случайной величины

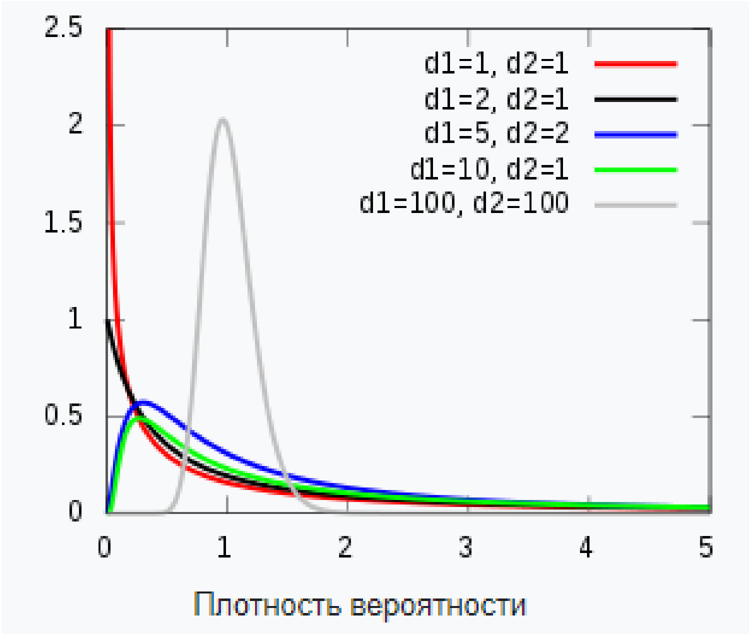
$$F = \frac{Y_1/d_1}{Y_2/d_2}$$

называется **распределением Фишера** со степенями свободы d_1 и d_2. Пишут $F \sim \mathrm{F}(d_1, d_2)$.

Опишем теперь функцию плотности случайной величины F:

$$\varrho_F(x) = \frac{\sqrt{\frac{(d_1\,x)^{d_1}\ d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$

Картинка:



2.3 Совместное распределение случайных величин

Пусть ξ и η — случайные величины. Рассмотрим отображение $\omega \to (\xi(\omega), \eta(\omega))$. Это отображение определяет на плоскости вероятностную меру μ следующим образом:

$$\mu(B) = \mathsf{P}(\{\omega : \ (\xi(\omega), \eta(\omega)) \in B\})$$

Меру μ называют **совместным распределением** случайных величин ξ и η. Функцию

$$F(x, y) = \mu((-\infty, x] \times (-\infty, y]) = \mathsf{P}(\{\omega : \ \xi(\omega) \leqslant x \text{ и } \eta(\omega) \leqslant y\})$$

называют **функцией совместного распределения** случайных величин ξ и η.

Аналогичным образом определяется совместное распределение любого конечного числа случайных величин.

2.4 Совместная и частная плотность

Если существует интегрируемая (по Риману) и неотрицательная функция $\varrho(x, y)$ такая, что

$$\mu(A) = \iint\limits_A \varrho(x, y) dx dy$$

для всякого допустимого (измеримого по Жордану) множества A, то ϱ называется **совместной плотностью распределения**.

Если известна плотность ϱ совместного распределения ξ и η, то можно найти плотности распределения каждой из случайных величин. Например, для случайной величины ξ:

$$\mu_\xi((a, b]) = \mu((a, b] \times \mathbb{R}) = \int\limits_a^b \left(\int\limits_{-\infty}^{+\infty} \varrho(x, y) dy \right) dx$$

и, следовательно,
$$\varrho_\xi(x) = \int\limits_{-\infty}^{+\infty} \varrho(x,y)dy$$
Если распределение каждой из случайных величин задаётся плотностью, то совместное распределение может не иметь плотность.

2.5 Независимые случайные величины и некореллированность случайных величин

Случайные величины ξ и η называются **независимыми**, если для всяких промежутков U и V выполняется равенство
$$\mathsf{P}(\{\omega:\ \xi(\omega) \in U \text{ и } \eta(\omega) \in V\}) = \mathsf{P}(\{\omega:\ \xi(\omega) \in U\}) \cdot \mathsf{P}(\{\omega:\ \eta(\omega) \in V\})$$
В терминах совместного распределения это равенство записывается так:
$$\mu(U \times V = \mu_\xi(U) \cdot \mu_\eta(V)$$
Говорят, что мера μ является произведением мер μ_ξ и μ_η и пишут $\mu = \mu_\xi \otimes \mu_\eta$.

Независимость ξ и ν равносильна тому, что
$$F(x,y) = F_\xi(x) \cdot F_\eta(y)$$
где F — совместная функция распределения ξ и η.

Для непрерывных случайных величин меняется определение математического ожидания. Математическое ожидание непрерывной случайной величины, распредление которой задаётся плотностью $\varrho(x)$, равно
$$\mathbb{E}[\xi] = \int\limits_{-\infty}^{+\infty} x\varrho(x)dx$$
Определения дисперсии, ковариации и коэффициента коррелляции требуют лишь дополнительного условия $\mathbb{E}\,\xi^2 < \infty$. Таким образом, определение некорреллированности случайных величин остаётся тем же: две случайные величины X, Y называются **некорреллированными**, если $r(X,Y) = 0$.

Заметим, что если две случайные величины X, Y независимы, то они некорреллированы. Обратное, вообще говоря, неверно.

2.6 Моменты распределений

Число $\mathbb{E}\,\xi^k$ называется **моментом порядка** k или k**-ым моментом случайной величины** ξ.

Часть II

Математическая статистика

3 Оценки параметров

3.1 Точечные оценки параметров

Основная задача математической статистики состоит в нахождении неизвестного распределения. Приведём пример: пусть в ящике N шаров, M из которых чёрные. Вынимаем n шаров. Вопрос теории вероятностей: с какой вероятностью среди них m чёрных шаров? Вопрос математической статистики: пусть среди n шаров оказалось m чёрных, сколько всего чёрных шаров в коробке (чему равно M)?

Предположим, нам изначально известно, что неизвестное распределение принадлежит некоторому семейству распределений с функциями распределения $F_\theta, \theta \in \Theta \subset \mathbb{R}^k$. Тогда задача сводится к нахождению неизвестного параметра $\theta_0 \in \Theta$, соответствующего нашему неизвестному распределению.

Определение (Выборка и статистики)**.** *Вектор $X = (X_1, \ldots, X_n)$ с независимыми компонентами, где каждая случайная величина имеет одно и то же распределение, называется **выборкой**.*

*Все случайные величны вида $T(X)$ называются **статистиками**.*

Проведя серию независимых экспериментов с неизвестным распределением, мы получаем выборку $X = (X_1, \ldots, X_n)$. По выборке нам бы хотелось определить значение $\widehat{\theta}_n$, которое в каком-то смысле близко к реальному неизвестному значению θ_0. Тем самым, мы строим функцию от выборки $\widehat{\theta}_n(X)$ со значением во множестве параметров. Такие случайные величины называются оценками неизвестного параметра.

3.2 Интервальные оценки. Доверительные интервалы

Ясно, что информация о состоятельности оценки $\widehat{\theta}_n(X)$ не достаточно для того, чтобы что-то говорить о возможном значении параметра θ. Предположим, что мы знаем “скорость сходимости”, то есть для фиксированного $\alpha \in (0, 1)$, для фиксированного $\varepsilon > 0$ мы смогли подобрать номер n, для которого $\mathsf{P}_\theta(\left|\widehat{\theta}_n(X) - \theta\right| < \varepsilon) > 1 - \alpha$. Тогда параметр $\theta \in (\widehat{\theta}_n(X) - \varepsilon, \widehat{\theta}_n(X) + \varepsilon)$ с большой вероятностью.

Обобщая данное наблюдение, приходим к определению доверительного интервала. Пусть заданы две статистики $\theta(X)$ и $\theta_2(X)$. Случайный интервал $(\theta_1(X), \theta_2(X))$ называется **доверительным интервалом** для параметра θ с уровнем доверия $1 - \alpha$, если при всех θ выполняется неравенство

$$\mathsf{P}_\theta(\theta_1^n(X) < \theta < \theta_2^n(X)) \geqslant 1 - \alpha.$$

Если же последовательность пар статистик $\theta_1^n(X)$ и $\theta_2^n(X)$, для которой

$$\liminf_{n\to\infty} \mathsf{P}_\theta(\theta_1^n(X) < \theta < \theta_2^n(X)) \geqslant 1 - \alpha,$$

то говорят, что задан асимптотический интервал уровня доверия $1 - \alpha$.

3.3 Свойства точечных оценок параметров

Сформулируем теперь, в каком смысле можно понимать близость оценки $\widehat{\theta}_n(X)$ к неизвестному параметру θ.

I. **Несмещённость.** $\mathbb{E}_\theta\, \widehat{\theta}_n(X) = \theta$. Это естественное свойство, которое означает, что в среднем оценка даёт правильное значение неизвестного параметра.

II. **Состоятельность.** $\lim\limits_{n\to\infty} \widehat{\theta}_n(X) = \theta$ по вероятности P_θ (соответствующей функции распределения F_θ). Обычно это свойство является следствием закона больших чисел и теорем непрерывности.

III. **Сильная состоятельность.** $\lim\limits_{n\to\infty} \widehat{\theta}_n(X) = \theta\, \mathsf{P}_\theta$-п.н.

IV. **Асимптотическая нормальность.** Оценка $\widehat{\theta}_n(X)$ является асимптотически нормальной, если
$$\sqrt{n}\left(\widehat{\theta}_n(X) - \theta\right) \to \mathbb{Z} \sim \mathcal{N}\left(\imath, \sigma^{\in}(\theta)\right)$$
по распределению для некоторого числа $\sigma^2(\theta)$. Это условие влечёт состоятельность и позволяет оценивать вероятности событий $\alpha < \theta_n(X) < \beta$ с помощью нормального распределения.

V. **Эффективность.** Несмещённая оценка $\widehat{\theta}_n(X)$ называется эффективной (в классе всех несмещённых оценок), если она имеет наименьшую дисперсию среди всех несмещённых оценок, т.е. $\mathbb{E}_\theta\left(\widehat{\theta}_n(X) - \theta\right)^2 \leqslant \mathbb{E}_\theta(\theta_n^*(X) - \theta)^2$ для произвольной несмещённой оценки $\theta_n^*(X)$. Это очень естественное требование, особенно, если принять во внимание неравенство Чебышева:
$$\mathsf{P}(\left|\widehat{\theta}_n(X) - \theta\right| \geqslant \varepsilon) \leqslant \frac{\mathbb{D}\, \widehat{\theta}_n(X)}{\varepsilon^2}.$$

4 Метод максимального правдоподобия

4.1 Метод максимального правдоподобия для оценки параметров распределений

Определение (Обобщённая плотность)**.** *Пусть случайная величина ξ имеет дискретное или непрерывное распределение.* ***Обобщённой плотностью*** *ϱ этого распределения (этой случайной величины) назовём обычную плотность в случае, когда ξ имеет непрерывное распределение, а в дискретном случае положим $\varrho(t) = \mathsf{P}(\xi = t)$, где t принадлежит области значений ξ.*

Определение (Функция правдоподобия)**.** *Пусть $X_1, \ldots, X_n$ — выборка из распределения с обобщённой плотностью ϱ_θ. Обобщённая плотность вектора $X = (X_1, \ldots, X_{})$*
$$p(x,\theta) := \varrho_\theta(x_1) \cdot \ldots \cdot \varrho_\theta(x_n)$$
называется ***функцией правдоподобия****, а её логарифм*
$$L(x,\, \theta) = \ln p(x,\, \theta) = \sum_{i=1}^{n} \ln \varrho_\theta(x_i)$$
называют ***логарифмической функцией правдоподобия****.*

Определение (Оценка максимального правдоподобия)**.** *Пусть $X_1, \ldots, X_n$ — выборка из распределения с обобщённой плотностью ϱ_θ. Оценка $\widehat{\theta}_n(X)$ называется* ***оценкой максимального правдоподобия****, если при фиксированном $x = (x_1, \ldots, x_n)$ число $\widehat{\theta}_n(x)$ есть точка максимума функции правдоподобия $p(x, \theta)$. То есть, в качестве оценки выбирается такое значение θ, при котором наблюдаемые значения $X = (X_1, \ldots, X_N)$ имеют максимальную вероятность.*

Замечание. *Ясно, что точки максимума функций $p(x, \theta)$ и $L(x, \theta)$ совпадают. Обычно для поиска точки максимума просто приравнивают к нулю производную функции $L(x, \theta)$, так как считать производную функции для суммы проще, чем для произведения.*

Пример. *Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с вероятностью успеха θ. Найдём оценку по методу максимального правдоподобия. В нашем случае обобщённая плотность распределения имеет вид $\varrho_\theta(t) = \theta^t(1-\theta)^{1-t}$, где $t \in \{0, 1\}$. Тем самым, функция правдоподобия*

$$p(x,\, \theta) = \theta^{\sum\limits_{j=1}^{n} x_j}(1-\theta)^{n-\sum\limits_{j=1}^{n} x_j}$$

и логарифмическая функция правдоподобия

$$L(x,\, \theta) = \left(\sum_{j=1}^{n} x_j\right)\ln\theta + \left(n - \sum_{j=1}^{n} x_j\right)\ln(1-\theta).$$

Дифференцируя, получаем уравнения

$$\frac{\sum x_j}{\widehat{\theta}_n} - \frac{n - \sum x_j}{1 - \widehat{\theta}_n} = 0,$$

откуда $\widehat{\theta}_n = \frac{\sum x_j}{n} = \bar{X}_n$.

Заметим, что в точке максимума $\widehat{\theta}_n(x)$ выполнено

$$L(x, \theta) \approx L(x, \widehat{\theta}_n(x)) + \frac{1}{2}\delta^2_{\theta\theta}L(x, \widehat{\theta}_n(x))(\theta - \widehat{\theta}_n(x))^2,$$

то есть вторая производная $\delta^2_{\theta\theta}L(x, \theta)$ описывает сколь быстро логарифмическая функция правдоподобия меняется в окрестности точки максимума. Пусть теперь на самом деле выборка задана распределением с параметром θ_0. Вычисляем среднее значение данной второй производной в точке θ_0 со знаком минус (в дискретном случае все знаки интегралов надо заменить на знаки сумм):

$$-\mathbb{E}_{\theta_0}\,\delta^2_{\theta\theta}L(X, \theta_0) = -\int\left[\delta^2_{\theta\theta}\ln p(x,\, \theta_0)\right]p(x, \theta_0)dx = \int\left[\frac{\delta^2_{\theta\theta}p(x,\, \theta_0)}{p(x,\, \theta_0)} - \frac{(\delta_\theta p(x, \theta_0))^2}{p^2(x, \theta_0)}\right]p(x,\, \theta_0)dx.$$

Предположим, что для $p(x,\, \theta)$ выполнено следующее условие регулярности: $p(x, \theta)$ положительна, непрерывна и дифференцируема по θ, причём операции дифференцирования и интегрирования перестановочны.

Так как

$$\int p(x, \theta)dx = 1,$$

то

$$\int \delta_\theta p(x, \theta)dx = \int \delta^2_{\theta\theta}p(x, \theta_0)dx = 0,$$

поэтому

$$\mathbb{E}_{\theta_0}\, \delta^2_{\theta\theta} L(X,\,\theta_0) = \int \frac{(\delta_\theta p(x,\theta_0))^2}{p(x,\theta_0)} dx = \int \frac{(\delta_\theta p(x,\theta_0))^2}{p^2(x,\theta_0)} p(x,\theta_0) dx =$$
$$= \int \left[\delta_\theta \ln p(x,\theta_0)\right]^2 p(x,\,\theta_0) dx = \mathbb{E}_{\theta_0} \left[\delta_\theta L(X,\,\theta_0)\right]^2.$$

4.2 Свойства ММП-оценок

Теорема (Состоятельность)**.** *Пусть $\theta \in (\alpha,\beta)$, $\varrho(x)$ непрерывна по θ и при каждом x функция $L(x,\theta)$ имеет ровно одну точку локального максимума $\theta_n(x)$ в (α,β). Тогда $\theta_n(X)$ сходится по вероятности к θ_0 (то есть является состоятельной оценкой).*

Теорема. *Пусть в дополнение к предположениям предыдущей теоремы выполнены условия регулярности $\varrho_\theta(x)$ положительна, трижды непрерывно дифференцируема по θ, причём операции дифференцирования и интегрирования перестановочны и $\left|\frac{\delta^3 \ln \varrho_\theta(x)}{\delta\theta^3}\right| \leqslant H(x)$, $\mathbb{E}_\theta\, H(X_1) = \int H(x)\varrho\theta(x)dx \leqslant M$ — не зависит от θ.*

Тогда оценка, полученная по методу максимального правдоподобия, будет асимптотически нормальной с асимптотической дисперсией $\frac{1}{i(\theta)}$.

5 Центральная предельная теорема

Пусть $\{\xi\}$ — последовательность независимых одинаково распределённых случайных величин, причём $\mathbb{E}\,\xi_1 = \mu$ и $\mathbb{D}\,\xi_1 = \sigma^2$. Тогда для всех x

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{\xi_1 + \ldots + \xi_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-y^2/2} dy.$$

6 Закон больших чисел

Теорема (Закон больших чисел)**.** *Предположим, что $\{\xi_n\}$ — последовательность независимых одинаково распределённых случайных величин таких, что $\mathbb{E}\,\xi_n^4 < \infty$. Пусть $\mathbb{E}\,\xi_n = \mu$. Тогда с вероятностью единица*

$$\lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} = \mu.$$