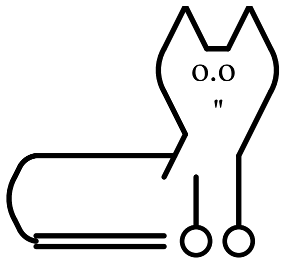


Конспект находится в разработке, могут присутствовать опечатки и неточности

Математическая статистика, DS-поток

I Статистики, оценки и их свойства

I - 1 Метод Монте-Карло

Рассмотрим задачу минимизации функции с использованием градиентного спуска.

<u>Задача:</u>

$\frac{1}{n}\sum\limits_{i=1}^{n} f_i(x) \to \min_x$

Задача очень важная, потому что именно это мы и делаем в ML, когда минимизируем ошибку:

- MSE: $\frac{1}{n}\sum\limits_{i=1}^{n} (y_i - f(x_i))^2$
- LogLoss: $-\frac{1}{n}\sum\limits_{i=1}^{n} y_i \ln(f(x_i)) + (1 - y_i)\ln(1 - f(x_i))$

Решается с помощью градиентного спуска:

$\text{GD} : x_{t+1} = x_t - \lambda \cdot \frac{1}{n}\sum\limits_{i=1}^{n} f_i'(x)$

$\text{SGD} : x_{t+1} = x_t - \frac{\lambda}{k} \cdot \frac{1}{n}\sum\limits_{j=1}^{k} f_{i_j}'(x)$

В SGD мы убрали подсчет большой суммы при помощи семплирования, в этом и заключается идея метода Монте-Карло

GD $\to$ SGD - метод Монте-Карло

Рассмотрим теперь другую задачу

<u>Задача:</u>

Нужно оценить $\theta = \mathbb{E}f(X), X \sim P$

$\hat{\theta} = \frac{1}{k}\sum\limits_{j=1}^{k} f(X_j)$

Причем, по УЗБЧ есть сходимость почти наверное: $\hat{\theta} \underset{k\to\infty}{\overset{\text{п.н.}}{\to}} \theta$

При помощи семплирования избавились от подсчета интеграла

Рассмотрим серию примеров:

<u>Примеры:</u>

1) $X \sim U\{1, ..., n\}$

   $f(i) = f_i$

   $\theta(x) = \mathbb{E}f(X)(x) = \frac{1}{n}\sum\limits_{i=1}^{n} f_i(x)$

$\hat{\theta} = \frac{1}{k}\sum\limits_{i=1}^{k} f_{i_j}(x),\ i_1, ..., i_k \sim U\{1, ..., n\}$

2) $f(\theta) = \mathbb{E}g(X, \theta)$, где распределение X не зависит от θ

   $f(\theta) = \int\limits_{\mathbb{R}} g(x, \theta)p(x)dx$

   $f'(\theta) = \int\limits_{\mathbb{R}} \frac{\partial g}{\partial \theta}(x, \theta)p(x)dx = \mathbb{E}\frac{\partial g}{\partial \theta}(x, \theta)$

Замечание

Многие используемые в ML метрики имеют вид $\mathbb{E}g(x, \theta)$, где θ - параметр модели. Второй пример показывает, что можно взять производную и применить метод Монте-Карло

3) $I = \int\limits_a^b f(x)dx = (b-a)\int\limits_a^b \frac{f(x)}{b-a}dx = (b-a)\int\limits_{\mathbb{R}} f(x)p(x)dx = (b-a)\mathbb{E}f(X)$

   $p(x) = \frac{1}{b-a}\mathbb{I}\{x \in [a, b]\}$

   X имеет распределение с плотностью $p(x)$, то есть $U[a, b]$

   $\hat{I} = \frac{b-a}{k}\sum\limits_{j=1}^{k} f(x_j), X_1, ..., X_k \sim U[a, b]$

4) Скорость сходимости метода

   $\theta = \mathbb{E}f(X), X \sim P$

   Оценка $\hat{\theta} = \frac{1}{k}\sum\limits_{j=1}^{k} f(X_j), X_1, ..., X_k \sim P$

   В силу ограниченности дисперсии можем по ЦПТ записать следующее

   $\sqrt{n}\frac{\hat{\theta}-\theta}{\sqrt{\mathbb{D}f(X)}} \xrightarrow{d} \mathcal{N}(0, 1)$

   $P\left(\sqrt{n}\frac{|\hat{\theta}-\theta|}{\sqrt{\mathbb{D}f(X)}} < 3\right) \approx 0.99$

   То есть порядок сходимости(погрешности) в МК: $\frac{3\sqrt{\mathbb{D}f(X)}}{\sqrt{n}} \sim \frac{1}{\sqrt{n}}$

5) Оценка интеграла методом прямоугольников

   $I = \int\limits_a^b f(x)dx$

   $\hat{I} = \sum\limits_{i=1}^{n} f\left(\frac{x_i + x_{i-1}}{2}\right)(x_i - x_{i-1})$

   $\{x_i\}$ - сетка точек

Теорема (о методе прямоугольников)

Если f - дважды дифференцируема, то равномерная сетка дает оценку:

$\left|I - \hat{I}\right| \leq \frac{M(b-a)^3}{24n^2}, M = \max_{[a,b]}|f''(x)|$

То есть у метода прямоугольников скорость сходимости $O\left(\frac{1}{n^2}\right)$

Замечание

1) Что нужно каждому методу:
   - Для МК: $\mathbb{D}f(X) < \infty$
   - Для МП: $f''(x) < \infty$
2) Для одномерных задач лучше использовать метод прямоугольников, так как он сходится быстрее
3) У МП экспоненциально растет сетка при росте размерности, так что в многомерном случае он будет работать очень долго, при $d \geq 5$ лучше использовать МК
4) МК эффективно работает с матожиданием

I - 2 Повторение теории вероятностей

I - 2.1 Случайные величины и их свойства

Опр

$(\Omega, \mathcal{F}, P)$ - вероятностное пространство
- Ω - пространство элементарных исходов
- $\mathcal{F}$ - σ-алгебра на Ω
- P - вероятностная мера

Опр

Функцией распределения $F(x)$ называется вероятность $P((-\infty, x])$

Свойства функции распределения:
- Неотриацтельная, ограниченная
- Неубывающая
- Непреывная справа
- $\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to \infty} F(x) = 1$

Первые 3 свойства очевидны, последнее следует из непрерывности меры

Опр

ξ(случайная величина) - измеримая функция $\Omega \to \mathbb{R}(\mathbb{R}^n)$

$F_\xi(x) = P(\xi \leq x)$ - функция распределения случайной величины

Задача(квантили экспоненциального распределения)

$\xi \sim \mathrm{Exp}(\theta)$

Найдем функцию, обратную к функции распределения

$F_\xi(x) = y = 1 - e^{-\theta x}$

$e^{-\theta x} = 1 - y \Rightarrow -\theta x = \ln(1 - y)$

$x = -\frac{\ln(1-y)}{\theta}$

Ответ: $x = -\frac{\ln(1-y)}{\theta}$

Опр

Матожиданием случайной величины называется $\mathbb{E}\xi = \int\limits_{\mathbb{R}} xp(x)dx$, если распределение абсолютно непрерывное и у него есть плотность

Дисперсия случайной величины: $\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$

Замечание

Конечно, в общем случае матожидание тоже можно определить, как $\mathbb{E}\xi = \int\limits_{\mathbb{R}} xP_\xi(dx)$

Задача(экспонента равномерного распределения)

Исследуем распоеделений $e^\xi, \xi \sim U[0, 1]$

$F_{e^\xi}(x) = P\left(e^\xi \leq x\right) = P(\xi \leq \ln(x)) = F_\xi \circ \ln(x)$ при $x \in [1, e]$

$p_{e^\xi}(x) = F'_{e^\xi}(x) = \frac{1}{x}\mathbb{I}\{x \in [1, e]\}$

$\mathbb{E}e^\xi = \int\limits_{\mathbb{R}} x \cdot \frac{1}{x}\mathbb{I}\{x \in [1, e]\}dx = \int\limits_1^e dx = e - 1$

Задача(распределение среднего)

$X_1, ..., X_n \sim P$

$\mathbb{E}X_i = a, \mathbb{D}X_i = \sigma^2$

$\mathbb{E}\overline{X} = \frac{n}{n}\mathbb{E}X_i = a$

$\mathbb{D}\overline{X} = \frac{1}{n^2}\mathbb{D}\left(\sum\limits_{i=1}^n X_i\right) = \frac{n}{n^2}\mathbb{D}X_i = \frac{\sigma^2}{n}$

$$\mathbb{E}\left(\overline{X}\right)^2 = \frac{1}{n^2}\left(\overbrace{\mathbb{E}X_1^2}^{\sigma^2+a^2} + ... + \mathbb{E}X_n^2 + 2\overbrace{\mathbb{E}X_1X_2}^{a^2}...\right) = \frac{\sigma^2}{n} + a^2$$

Задача(моменты нормального распределения)

$\xi \sim \mathcal{N}(0, \sigma^2)$

$\mathbb{E}|\xi|^n = \int\limits_{\mathbb{R}} |x|^n e^{-\frac{x^2}{2\sigma^2} \cdot \frac{1}{\sigma\sqrt{2\pi}}} dx 2 = \int\limits_0^{+\infty} \frac{x^n \exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$

Вспомним свойства Гамма-функции

$\Gamma(x) = \int\limits_0^{+\infty} e^{-t} t^{x-1} dx,\ \mathrm{Re}(x) > 0$

$\Gamma(x+1) = x\Gamma(x)$

Теперь сделаем замену $u = \frac{x^2}{2\sigma^2}$. Тогда $x = \sigma\sqrt{2u},\ dx = \frac{\sigma}{\sqrt{2u}}$

$\mathbb{E}|\xi|^n = 2 \int\limits_0^{+\infty} \frac{e^{-u}\left(\sigma\sqrt{2u}\right)^n}{\sqrt{2\pi}\sigma} \cdot \left(\frac{\sigma}{\sqrt{2u}}\right) du = \frac{\left(\sigma\sqrt{2}\right)^n}{\sqrt{\pi}} \int\limits_0^{+\infty} u^{\frac{n-1}{2}} e^{-u} du =$

$= \frac{\left(\sigma\sqrt{2}\right)^n}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right)$

<u>**Ответ:**</u> $\mathbb{E}|\xi|^n = \frac{\left(\sigma\sqrt{2}\right)^n}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right)$

Если n – нечетное, то $\mathbb{E}|\xi|^n = \frac{\sigma^n}{\sqrt{\pi}}(n-1)!!$. Если четное, то $\sigma^n(n-1)!!$

Задача(поиск больных)

Нужно провести тест на поиск больных, действуем следующим образом

1) Делим людей на группы по k человек
2) В каждой группе берем кровь на анализы и смешиваем(вероятность заболеть у конкретного человек равна p)
   - Если в крови есть антитела, то проверяем каждого из группы
   - Если антител нет, то все из группы здоровы

Пусть T - общее число тестов, нужно найти такое k, что $\mathbb{E}T \to \min$

T_i - количество тестов в i-ой группе.

$\mathbb{E}T = \sum\limits_{i=1}^{\lfloor \frac{n}{k} \rfloor} \mathbb{E}T_i$

$\mathbb{E}T_i = (1-p)^k + \left(1 - (1-p)^k\right)(k+1)$

$\mathbb{E}T = \lfloor \frac{n}{k} \rfloor \left((1-p)^k + \left(1-(1-p)^k\right)(k+1)\right) = (1-p)^k \left(\frac{n}{k} - \frac{n(k+1)}{k}\right) + \frac{(k+1)n}{k}$

$(1-p)^k \sim 1 - pk$

$\mathbb{E}T \sim (1-pk)\left(\frac{n}{k} - \frac{(k+1)n}{k}\right) + \frac{(k+1)n}{k} = \frac{n}{k} - pn - \frac{(k+1)n}{k} + pn(k+1) + \frac{(k+1)n}{k} =$

$= \frac{n}{k} + pnk$

$\frac{n}{k} + pnk \to \min \Leftrightarrow \frac{1}{k} + pk \to \min$

$-\frac{1}{k^2} + p = 0 \Leftrightarrow k = \frac{1}{\sqrt{p}}$

Ответ: $k = \frac{1}{\sqrt{p}}$

Опр

$A_1, ..., A_n$ независимы, если $\forall \{i_1, ..., i_k\} \subset \{1, ..., n\}$

$P\left(A_{i_1} \bigcap ... \bigcap A_{i_k}\right) = P\left(A_{i_1}\right) \cdot ... \cdot P\left(A_{i_k}\right)$

Для случайных величин:

$$P\left(\xi_{i_1} \in A_{i_1}, ..., \xi_{i_k} \in A_{i_k}\right) = \prod_{j=1}^{k} P\left(\xi_{i_j} \in A_{i_j}\right)$$

Теорема (о наследовании независимости)

Если f, g - борелевские функции, а ξ и η независимы, то $f(\xi)$ и $g(\eta)$ тоже независимы

▲

$P(f(\xi) \in A, g(\eta) \in B) = P(\xi \in f^{-1}(A), \eta \in g^{-1}(B)) =$

$= P(f(\xi) \in A) \cdot P(g(\eta) \in B)$

■

I - 2.2 Сходимости

1) Сходимость почти наверное

   $X_n \overset{\text{п.н.}}{\to} X \Leftrightarrow P(\{\omega \mid X_n(\omega) \nrightarrow X(\omega)\}) = 0$

2) Сходимости по вероятности

   $X_n \overset{P}{\to} X \Leftrightarrow \forall \varepsilon > 0 \;\; P(|X_n - X| > \varepsilon) \overset{n\to\infty}{\to} 0$

3) Сходимость по распределению

   $X_n \overset{d}{\to} X$, если $\forall f$ - ограниченной непрерывной функции $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.

   Или, равноценное условие, $\forall a$ - точка непрерывности F_X есть сходимость $F_{X_n}(a) \to F_X(a)$

п.н $\Rightarrow P \Rightarrow d$

Свойства сходимостей:

1) $X_n \xrightarrow{d} X = C \Rightarrow X_n \xrightarrow{P} X$
2) $X_n \xrightarrow{P} X \Rightarrow \exists \{X_{n_k}\} : X_{n_k} \xrightarrow{\text{п.н.}} X$

Задача

X_n - независимые случайные величины

$X_n \sim \mathcal{N}(0, \frac{1}{n}) \Rightarrow \sqrt{n} X_n \sim N(0,1)$

Нужно показать, что $X_n \xrightarrow{P} 0$

$F_{X_n}(a) = P(X_n \leq a) = F_{\xi(\sqrt{n} a)}, \xi \sim \mathcal{N}(0,1)$

$\lim\limits_{n\to\infty} F_{X_n}(a) = \lim\limits_{n\to\infty} F_\xi(a\sqrt{n}) = \begin{cases} 1,\ a>0 \\ \frac{1}{2},\ a=0 \\ 0,\ a<0 \end{cases}$

$F_0(X) = P(0 < x) = \begin{cases} 1,\ x>0 \\ 0,\ x\leq 0 \end{cases}$

Теорема (УЗБЧ)
$X_1, ..., X_n$ - НОРСВ, $\mathbb{E}|X_i| < +\infty$

Тогда $\overline{X} \xrightarrow{\text{п.н.}} \mathbb{E}X_1$

Теорема (ЦПТ)
$X_1, ..., X_n$ - НОРСВ, $\mathbb{E}|X_i| < +\infty, \mathbb{E}X_i^2 < +\infty$

Тогда $\sqrt{n}\frac{\overline{X} - \mathbb{E}X_1}{\sqrt{\mathbb{D}X_1}} \xrightarrow{d} \mathcal{N}(0,1)$

Неравенство Чебышева

$P(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{D}X}{a^2}$

Задача(распределение Бернулли)

$X_1, ..., X_n \sim \mathrm{Bern}(\frac{1}{2})$

С какого n будет выполнено $P\left(0.4 \leq \overline{X} \leq 0.6\right) \geq 0.7$?

$P\left(\left|\frac{X_1 + ... + X_n}{n} - 0.5\right| > 0.1\right) < \frac{\mathbb{D}\frac{X_1 + ... + X_n}{n}}{0.01} = \frac{25}{n}$

$P\left(\left|\frac{X_1 + ... + X_n}{n} - 0.5\right| \leq 0.1\right) > 1 - \frac{\mathbb{D}\frac{X_1 + ... + X_n}{n}}{0.01} = 1 - \frac{25}{n}$

Задача(распределение Пуассона)

$\xi_1, ..., \xi_{125} \sim \mathrm{Pois}(5)$

$\mathbb{E}\xi_i = 5, \mathbb{D}\xi_i = 5$

$P\left(\overline{\xi} \leq 5.5\right) = ?$

Воспользуемся ЦПТ:

$\sqrt{125}\frac{\overline{\xi}-5}{\sqrt{5}} \xrightarrow{d} \mathcal{N}(0,1) \Rightarrow \overline{\xi}-5 \xrightarrow{d} \mathcal{N}\left(0,\frac{1}{25}\right) \Rightarrow \overline{\xi}-5.5 \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2},\frac{1}{25}\right)$

Дальше нужно просто подставить 0 в функцию распределения $\mathcal{N}\left(-\frac{1}{2},\frac{1}{25}\right)$

I - 3 Статистики и оценки

I - 3.1 Вероятностно-статистическая модель

Опр
$(\mathcal{X},\mathfrak{B}(\mathcal{X}),\mathcal{P})$ - вероятностно-статистическая модель

$\mathcal{X}$ - выборочное пространство, $\mathfrak{B}(\mathcal{X})$ - σ-алгебра на нем

$\mathcal{P}=\{P_\theta \mid \theta\in\Theta\}$, Θ - множество параметров

Наша задача – оценить θ по данной выборке $X=(X_1,...,X_n)$ из $P\in\mathcal{P}$

Опр
Пусть (E,ε) - измеримое пространство. Тогда измеримая функция $S:\mathcal{X}\to E$ называется статистикой

Замечание
Смысл в том, что статистика - какая-то функция от выборки

Опр
Если $E=\Theta$, то такая статистика называется оценкой

Пример:

$\tau\sim\mathrm{Exp}(\lambda)$

$\tau(\theta)=\frac{1}{\theta}$

Примеры статистик:

1) $\overline{X},\overline{X^2},...,\overline{X^k}$ - выборочный k-ый момент
   $g(X)=\frac{1}{n}\sum_{i=1}^{n}g(X_i)$
2) $S^2=\overline{X^2}-\left(\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ - выборочная дисперсия
3) Порядковые статистики

   $X_{(1)},...,X_{(n)}$ - вариационный ряд

   $X_{(k)}$ – k-ая порядковая статистика

I - 3.2 Свойства оценок

Замечание
Для распределения P_θ имеются следующие обозначения : $\mathbb{E}_\theta, \mathbb{D}_\theta, P_\theta$ п.н., d_θ

Перейдем к свойствам оценок

$X = (X_1, ..., X_n)$ - выборка из распределения $P_\theta \in \mathcal{P}$

$\Theta \subseteq \mathbb{R}^d$

Опр
Оценка $\hat{\theta}(X)$ называется несмещенной оценкой параметра θ, если
$\mathbb{E}_\theta \hat{\theta}(X) = \theta \ \ \forall \theta \in \Theta$

Несмещенной относительно $\tau(\theta)$, если $\mathbb{E}_\theta \hat{\theta}(X) = \tau(\theta) \ \ \forall \theta \in \Theta$

Примеры:

1) $\hat{\theta}(X) = 1 : \mathbb{E}_\theta \hat{\theta}(X) = 1 \Rightarrow$ несмещенная при $\theta = 1$
2) $\hat{\theta}(X) = \overline{X} : \mathbb{E}_\theta \overline{X} = \mathbb{E}_\theta \frac{1}{n} \sum\limits_{i=1}^{n} X_i = \frac{1}{n} \mathbb{E}_\theta X_1 = \tau(\theta)$
3) $\hat{\theta}(X) = X_1 : \mathbb{E}_\theta \hat{\theta}(X) = \mathbb{E}_\theta X_1 \Rightarrow$ несмещенная относительно $\tau(\theta) = \mathbb{E}_\theta X_1$

Замечание
Смысл несмещенности: при бесконечном повторении эксперимента в среднем будем получать $\theta(\tau(\theta))$

Асимптотические свойства:

Пусть $X = (X_1, ..., X_n, ...,)$ - выборка неограниченного размера из распределения $P \in \mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}$

Опр

1) Оценка $\hat{\theta}(X)$ называется состоятельной оценкой параметра θ, если
$\hat{\theta}(X) \underset{n \to \infty}{\overset{P_\theta}{\to}} \theta \ \ \forall \theta \in \Theta$
2) Оценка $\hat{\theta}(X)$ называется сильно состоятельной оценкой параметра θ, если
$\hat{\theta}(X) \underset{n \to \infty}{\overset{P_\theta \text{ п.н.}}{\to}} \theta \ \ \forall \theta \in \Theta$
3) Оценка $\hat{\theta}(X)$ называется асимптотически нормальной оценкой параметра θ, если

$$\sqrt{n}\left(\hat{\theta}(X)-\theta\right) \xrightarrow[n\to\infty]{d_\theta} \mathcal{N}(0, \Sigma(\theta))$$

В таком случае $\Sigma(\theta)$ называется асимптотической матрицей ковариаций, в случае единичной размерности $\Sigma(\theta) = \sigma^2(\theta)$ - асимптотическая дисперсия

Замечание

$$\sqrt{n}\left(\hat{\theta}(X)-\theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta)) \Rightarrow \hat{\theta}(X) \overset{d_\theta}{\sim} \frac{1}{\sqrt{n}}\mathcal{N}(0, \sigma^2(\theta)) + \theta \overset{d_\theta}{=} \mathcal{N}\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$$

Смысл:

1) Состоятельность

При больших n оценка находится близко к θ, но нет численной характеристики, насколько

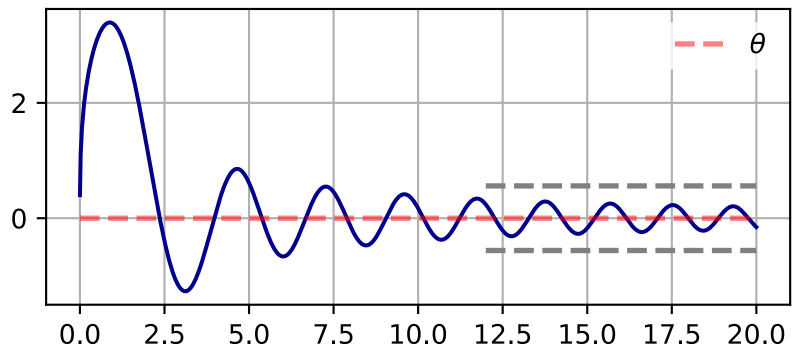


График 1: Иллюстрация состоятельной оценки

2) Асимптотическая нормальность

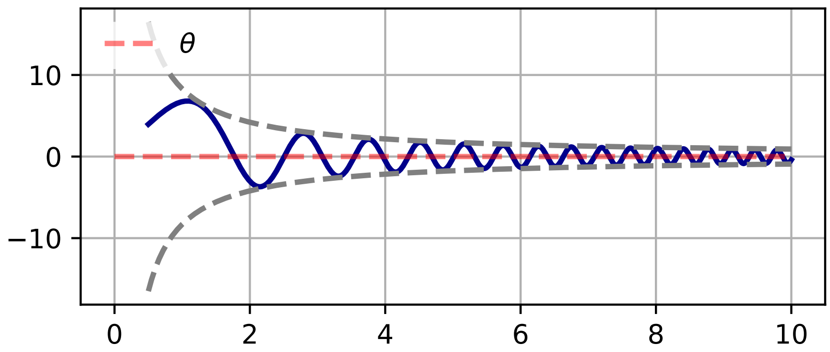


График 2: Иллюстрация асимптотически нормальной оценки

Здесь уже можем как-то оценить ширину, то есть численная характеристика есть

3) Сильная состоятельность

Имеет смысл для последовательно поступающих данных

<u>Пример:</u>

$X_1, ..., X_n \sim$ Laplace со сдвигом θ

$p_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$

$\mathbb{E}_\theta X_1 = \theta, \mathbb{D}_\theta X_1 = 2$

Исследуем $\hat{\theta} = \overline{X}$

1) $\mathbb{E}_\theta \overline{X} = \theta \Rightarrow$ оценка несмещенная
2) $\hat{\theta} = \overline{X} \underset{n\to\infty}{\overset{P_\theta}{\to}} \theta = \mathbb{E}_\theta X_1 \Rightarrow$ оценка состоятельная(по УЗБЧ)
3) Сильная состоятельность также следует из УЗБЧ
4) $\sqrt{n}\frac{\overline{X} - \overbrace{\mathbb{E}_\theta X_1}^{\theta}}{\sqrt{\mathbb{D}_\theta X_1}} \overset{d_\theta}{\to} \mathcal{N}(0,1)$ – ЦПТ

   $\sqrt{n}\left(\overline{X} - \theta\right) \overset{d_\theta}{\to} \mathcal{N}(0, \mathbb{D}_\theta X_1 = 2)$

   Значит оценка асимптотически нормальная с асимптотической дисперсией $\sigma^2(\theta) = 2$
5) Доверительный интервал

   $\overline{X} \overset{d_\theta}{\approx} \mathcal{N}\left(\theta, \frac{2}{n}\right)$

   По правилу 3σ:

   $P\left(\theta - 3\sqrt{\frac{2}{n}} < \overline{X} < \theta + 3\sqrt{\frac{2}{n}}\right) \approx 0.998$

   При $n = 200$ и $\overline{X} = 1$ с хорошей вероятностью $\theta \in (0.7, 1.3)$

<u>Утв</u>

Из асимптотической нормальности следует состоятельность(из сильной состоятельности тоже следует состоятельностьё)

I - 3.3 Наследование свойств

В этом параграфе наша цель заключается в том, чтобы получить оценку $\tau(\theta)$, обладающую некоторыми свойствами, если есть оценка $\psi(\theta)$, обладающая этими свойствами

<u>Пример:</u>

$X_1, ..., X_n \sim \text{Exp}(\theta)$

$\overline{X}$ - асимптотически нормальная оценка $\frac{1}{\theta}$, а нам нужна оценка на θ

Теорема (о наследовании сходимостей)
Пусть $\{\xi_n\}, n \in \mathbb{N}$ - последовательность случайных величин, ξ - случайная величина размерности d.

1) $\xi_n \xrightarrow{P} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная на множества B функция, причем $P(\xi \in B) = 1$, то $h(\xi_n) \xrightarrow{P} h(\xi)$
2) $\xi_n \xrightarrow{\text{п.н.}} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная на множества B функция, причем $P(\xi \in B) = 1$, то $h(\xi_n) \to \text{п.н. } h(\xi)$
3) Если $\xi_n \xrightarrow{d} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная, то $h(\xi_n) \xrightarrow{d} h(\xi)$

Следствие:
Если выполнены следующие условия:

1) $\hat{\theta}$ - (сильно) состоятельная оценка θ
2) Функция $\tau : \Theta \to \Theta$ непрерывна на Θ

Тогда $\tau\left(\hat{\theta}\right)$ - (сильно) состоятельная оценка $\tau(\theta)$

Теорема (дельта-метод)
Пусть:

1) $\hat{\theta}$ - асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}^d$ с асимптотической матрицей ковариаций $\Sigma(\theta)$
2) $\tau : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывно дифференцируема

Тогда $\tau\left(\hat{\theta}\right)$ - асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $D(\theta)\Sigma(\theta)D^T(\theta)$, где $D(\theta) = \frac{\partial \tau}{\partial \theta}(\theta)$

Вернемся к пример с экспоненциальным распределением

По ЦПТ асимптотическая дисперсия оценки $\overline{X}$ равна $\sigma^2(\theta) = \mathbb{D}_\theta X_1 = \frac{1}{\theta^2}$

Возьмем $\tau(x) = \frac{1}{x}$, тогда $\frac{\partial \tau}{\partial \theta}(x) = -\frac{1}{x^2}$

Подставим $\frac{1}{\theta}$ в дельта-метод: $\sigma^{2'}(\theta) = \frac{\partial \tau}{\partial \theta}\left(\frac{1}{\theta}\right) \cdot \frac{1}{\theta^2} \cdot \frac{\partial \tau}{\partial \theta}\left(\frac{1}{\theta}\right) = \frac{1}{\theta^2} \cdot \theta^4 = \theta^2$

То есть по дельта-методу $\frac{1}{\overline{X}}$ - асимптотически нормальная оценка θ с асимптотической дисперсией θ^2

I - 3.4 Доверительные интервалы

Опр
Пусть $X = (X_1, \ldots, X_n)$ - выборка из $P_\theta \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$

Если $\Theta \subseteq \mathbb{R}$, то пара статистик $(T_1(X), T_2(X))$ называется доверительным интервалом уровня доверия α, если $\forall \theta\ P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq \alpha$

Если выполняется равенство, то такой доверительный интервал называют точным.

Опр

Пусть $X = (X_1, ..., X_n, ...)$ - неограниченная выборка из $P_\theta \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$

Последовательность статистик $\left(T_1^{(n)}(X), T_2^{(n)}(X)\right)$ называется асимптотическим доверительным интервалом уровня доверия α, если $\liminf_{n \to \infty} P_\theta\left(T_1^{(n)}(X) \leq \theta \leq T_2^{(n)}(X)\right) \geq \alpha$

Если выполняется равенство, то такой доверительный интервал называют точным.

Замечание

Смысл: $(T_1(X), T_2(X))$ накрывает θ с большой вероятностью ($\geq \alpha$)

Алгоритм построения доверительного интервала:

1) Пусть $\hat{\theta}$ - асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$

   $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta)) \Rightarrow \sqrt{n}\frac{\hat{\theta}-\theta}{\sigma(\theta)} \xrightarrow{d_\theta} \mathcal{N}(0, 1)$

   $P\left(\sqrt{n}\frac{\left|\hat{\theta}-\theta\right|}{\sigma(\theta)} < Z_{\frac{1+\alpha}{2}}\right) \to \alpha$

   Z_α – это α квантиль стандартного нормального распределения

   В питоне считается так:

```
z = sps.norm.ppf((1 + alpha) / 2)
```

2) $\sigma(\theta)$ может не дать выразить интервал

   Пусть $\hat{\sigma}(\theta)$ – состоятельная оценка $\sigma(\theta)$

$$\sqrt{n}\frac{\hat{\theta}-\theta}{\hat{\sigma}} = \sqrt{n}\underbrace{\frac{\hat{\theta}-\theta}{\sigma(\theta)}}_{\xrightarrow{d_\theta}\mathcal{N}(0,1)} \cdot \underbrace{\frac{\sigma(\theta)}{\hat{\sigma}(\theta)}}_{\xrightarrow{P_\theta}1} \xrightarrow{d_\theta} \mathcal{N}(0,1)$$

Последний переход совершен по лемме Слуцкого

$\theta \in \left(\hat{\theta} - Z_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta} + Z_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right)$ - наш доверительный интервал

3) $\hat{\sigma}(\theta) = \sigma\left(\hat{\theta}\right)$, если $\sigma(\theta)$ непрерывна(по теореме о наследовании сходимости)

$\hat{\theta}$ - а.н.о. $\Rightarrow \hat{\theta}$ - состоятельная $\Rightarrow \sigma\left(\hat{\theta}\right)$ - состоятельная оценка $\sigma(\theta)$

<u>Пример:</u>

$X_1, ..., X_n \sim \text{Pois}(\theta)$

$\mathbb{E}_\theta X_1 = \theta, \mathbb{D}_\theta X_1 = \theta$

$\hat{\theta} = \overline{X}$ - а.н.о. $\mathbb{E}_\theta X_1$ с а.д. $\mathbb{D}_\theta X_1$

По ЦПТ: $\sqrt{n}\left(\overline{X} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \mathbb{D}_\theta X_1 = \theta)$

$\sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\theta}} \xrightarrow{d_\theta} \mathcal{N}(0,1) \Rightarrow P\left(\sqrt{n}\frac{\left|\overline{X}-\theta\right|}{\sqrt{\theta}} < Z_{\frac{1+\alpha}{2}}\right) \to \alpha$

$\sigma(\theta) = \sqrt{\theta}$ - непрерывная на $\mathbb{R}_+ \Rightarrow \sigma\left(\hat{\theta}\right) = \sqrt{\overline{X}}$ – состоятельная оценка $\sigma(\theta)$

$P\left(\sqrt{n}\frac{\left|\overline{X}-\theta\right|}{\sqrt{\overline{X}}} < Z_{\frac{1+\alpha}{2}}\right) \to \alpha$

Теперь выразим наш интервал:

$-Z_{\frac{1+\alpha}{2}} < \sqrt{n}\overline{X} - \frac{\theta}{\sqrt{\overline{X}}} < Z_{\frac{1+\alpha}{2}}$

$\overline{X} - Z_{\frac{1+\alpha}{2}}\sqrt{\frac{\overline{X}}{n}} < \theta < \overline{X} + Z_{\frac{1+\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}$

I - 4 Статистики и оценки. Семинар

I - 4.1 Оценки и сходимости

В вероятностно-статистическом пространстве $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ будем рассматривать $\mathcal{X} = \mathbb{R}, \mathfrak{B}(\mathcal{X}) = \mathfrak{B}(\mathbb{R})$

<u>Пример:</u> $\mathcal{P} = \{U[0,\theta] \mid \theta \geq 0\}, \Theta = [0, +\infty]$

<u>Опр</u>

Пусть (E, ε) - измеримое пространство. Тогда измеримая функция $S : \mathcal{X} \to E$ называется статистикой

Если $E = \Theta$, то такая статистика называется оценкой

Опр

Оценка называется несмещенной, если $\mathbb{E}_\theta \hat{\theta} = \theta \ \ \forall \theta \in \mathbb{R}$

Пример:

$X \sim U[0, \theta], \theta \in \mathbb{R}$

$\hat{\theta} = 20$ – идеальна при $\theta = 20$, смещенная при любом $\theta \neq 20$

Задача

$X_1, ..., X_n \sim \mathrm{Exp}(\theta), \theta > 0$

$\overline{X}$ - оценка $\frac{1}{\theta}$

Если посчитать матожидание $\frac{1}{\overline{X}}$, то поймем, что $\frac{1}{\overline{X}}$ не является несмещенной оценкой θ

Задача

$X_1, ..., X_n \sim \mathrm{Bern}(\theta), \theta \in (0, 1)$

Нужно показать, что $\nexists \hat{\theta}(X)$ оценки для $\frac{1}{\theta}$, такой что $\hat{\theta}$ – несмещенная оценка

Докажем от противного, если такая оценка существует, то

$\mathbb{E}_\theta \hat{\theta}(X) = \hat{\theta}(1) P_\theta(X_1 = 1) + \hat{\theta}(0) P_\theta(X_1 = 0) = c_1 \theta + c_2 (1 - \theta) \neq \frac{1}{\theta}$

То есть несмещенной оценки нет

Опр

1) Оценка $\hat{\theta}(X)$ называется состоятельной оценкой параметра θ, если
$\hat{\theta}(X) \underset{n \to \infty}{\overset{P_\theta}{\to}} \theta \ \ \forall \theta \in \Theta$
2) Оценка $\hat{\theta}(X)$ называется сильно состоятельной оценкой параметра θ, если
$\hat{\theta}(X) \underset{n \to \infty}{\overset{P_\theta \text{ п.н.}}{\to}} \theta \ \ \forall \theta \in \Theta$

Задача

$X_1, ..., X_n \sim U[0, \theta]$

$\hat{\theta}_1 = (n + 1) X_{(1)}$

$\hat{\theta}_2 = X_{(n)}$

Провери эти оценки на состоятельность и несмещенность

$\mathbb{E}_\theta \hat{\theta}_1 = (n + 1) \mathbb{E}_\theta X_{(1)} = \theta \cdot \frac{n+1}{n+1} = \theta \Rightarrow \hat{\theta}_2$ – несмещенная оценка

$\mathbb{E}_\theta \hat\theta_2 = \theta \cdot \frac{n}{n+1} \Rightarrow \hat\theta_2$ смещенная оценка

Проверим на состоятельность

$$P_\theta\Big(\big|X_{(n)} - \theta\big| > \varepsilon\Big) = P_\theta\Big(X_{(n)} < \theta - \varepsilon\Big) = P_\theta(\forall i : X_i < \theta - \varepsilon) =$$

$$= \prod_{i=1}^n P_\theta(X_i < \theta - \varepsilon) = \left(\frac{\theta-\varepsilon}{\theta}\right)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n \overset{n\to\infty}{\to} 0$$

Первый переход был сделан, так как $X_{(n)}$ всегда $< \theta$

Итого, $X_{(n)}$ – это состоятельная оценка

Исследуем $X_{(1)}$

$$P_\theta\Big(X_{(1)} \geq t\Big) = \prod_{i=1}^n P(X_i \geq t) = \left(\frac{\theta-t}{\theta}\right)^n = \left(1 - \frac{t}{\theta}\right)^n$$

$$P_\theta\Big(X_{(1)} < t\Big) = 1 - \left(1 - \frac{t}{\theta}\right)^n$$

$$P_\theta\Big(\big|(n+1)X_{(1)} - \theta\big| > \varepsilon\Big) = P_\theta\Big(X_{(1)} \leq \frac{\theta-\varepsilon}{n+1}\Big) + P_\theta\Big(X_{(1)} > \frac{\theta+\varepsilon}{n+1}\Big) =$$

$$= 1 - \left(1 - \frac{\theta-\varepsilon}{(n+1)\theta}\right) + \rho_+ \overset{n\to\infty}{\to} 1 - e^{-\frac{\varepsilon}{\theta}} = \rho_+ \not\to 0$$

То есть вторая оценка не является состоятельной

<u>Опр</u>

Оценка $\hat\theta(X)$ называется асимптотически нормальной оценкой параметра θ, если

$$\sqrt{n}\Big(\hat\theta(X) - \theta\Big) \underset{n\to\infty}{\overset{d_\theta}{\to}} \mathcal{N}(0, \Sigma(\theta))$$

В таком случае $\Sigma(\theta)$ называется асимптотической матрицей ковариаций, в случае единичной размерности $\Sigma(\theta) = \sigma^2(\theta)$ - асимптотическая дисперсия

При больших n имеем $\hat\theta \overset{d_\theta}{\approx} \mathcal{N}\left(\theta, \frac{\Sigma(\theta)}{n}\right)$

<u>Задача</u>

$X_1, ..., X_n \sim U[\theta - 1, \theta + 1]$

$\mathbb{E}_\theta X_1 = \theta, \mathbb{D}_\theta X_1 = \frac{1}{3}$

По ЦПТ получаем следующую сходимость:

$$\sqrt{n}\frac{\overline{X} - \theta}{\underbrace{\sqrt{\mathbb{D}_\theta X_1}}_{=\sqrt{\frac{1}{3}}}} \overset{d_\theta}{\to} \mathcal{N}(0, 1)$$

$$\sqrt{n}\Big(\overline{X} - \theta\Big) \overset{d_\theta}{\to} \mathcal{N}\left(0, \frac{1}{3}\right) \Rightarrow \overline{X} - \theta \overset{d_\theta}{\to} \mathcal{N}\left(0, \frac{1}{3n}\right)$$

$\hat\theta(X) = \overline{X}$ - а.н.о. θ с а.д. $\frac{1}{3}$

Также можем записать следующую вероятность попадания θ в доверительный интервал:

$$P\left(\overline{X} - \sqrt{\frac{3}{n}} \leq \theta \leq \overline{X} + \sqrt{\frac{3}{n}}\right) = 0.997$$

Какие свойства оценок важны?

- Состоятельность – очень важно
- Асимптотическая нормальность – желательно, чтобы понимать предполагаемое распределение
- Несмещенность – не очень важно

Теорема (о наследовании состоятельности)

Если выполнены следующие условия:

1) $\hat{\theta}$ - (сильно) состоятельная оценка θ
2) Функция $\tau : \Theta \to \Theta$ непрерывна на Θ

Тогда $\tau\left(\hat{\theta}\right)$ - (сильно) состоятельная оценка $\tau(\theta)$

▲

Сразу следует из теоремы о наследовании сходимостей

■

Пример:

$X_1, ..., X_n \sim \mathrm{Exp}(\theta)$

$\overline{X}$ – сильно состоятельная оценка $\frac{1}{\theta}$(по УЗБЧ)

Тогда по теореме $\frac{1}{\overline{X}}$ – сильно состоятельная оценка θ

Теорема (дельта-метод)

Пусть:

1) $\hat{\theta}$ - асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}^d$ с асимптотической матрицей ковариаций $\Sigma(\theta)$
2) $\tau : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывно дифференцируема

Тогда $\tau\left(\hat{\theta}\right)$ - асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $D(\theta)\Sigma(\theta)D^T(\theta)$, где $D(\theta) = \frac{\partial \tau}{\partial \theta}(\theta)$

Задача

$X_1, ..., X_n \sim \mathrm{Exp}(\theta)$

Нужно доказать, что $\frac{1}{\overline{X}}$ – это а.н.о. θ с а.д. θ^2

ИЗ ЦПТ получаем, что $\overline{X}$ - а.н.о. $\frac{1}{\theta}$ с а.д. $\frac{1}{\theta^2}$

Возьмем функцию $\tau(x) = \frac{1}{x}, \frac{\partial\tau}{\partial\theta}(x) = -\frac{1}{x^2}$

Применим дельта-метод: ${\sigma'}^2(\theta) = \frac{\partial\tau}{\partial\theta}\left(\frac{1}{\theta}\right) \cdot \frac{1}{\theta^2} \cdot \frac{\partial\tau}{\partial\theta}\left(\frac{1}{\theta}\right) = \frac{1}{\theta^2} \cdot \theta^4 = \theta^2$

<u>Задача</u>

$X_1, ..., X_n \sim \operatorname{Exp}(\theta)$

Нужно доказать, что $\hat{\theta} = \frac{\overline{X^2}}{2\left(\overline{X}\right)^3}$ – а.н.о и найти ее а.д.

Из ЦПТ знаем, что $\overline{X}$ – а.н.о. $\frac{1}{\theta}$ с а.д. $\frac{1}{\theta^2}$

$\overline{X^2}$ - а.н.о. $\frac{2}{\theta^2}$

Опять же, из ЦПТ следует, что для случайных величин $Z_i = \begin{pmatrix} X_i \\ X_i^2 \end{pmatrix}$ величина $\overline{Z} = \begin{pmatrix} \overline{X} \\ \overline{X^2} \end{pmatrix}$ является а.н.о.

Посчитаем для этой оценки матрицу ковариаций

$\mathbb{D}Z_1 = \begin{pmatrix} \mathbb{D}X_1 & \operatorname{cov}(X_1, X_1^2) \\ \operatorname{cov}(X_1, X_1^2) & \mathbb{D}X_1^2 \end{pmatrix}$

$\mathbb{D}X_1 = \frac{1}{\theta^2}$

$\operatorname{cov}(X_1, X_1^2) = \mathbb{E}X_1^3 - \mathbb{E}X_1 \cdot \mathbb{E}X_1^2 = \frac{6}{\theta^3} - \frac{2}{\theta^3} = \frac{4}{\theta^3}$

$\mathbb{D}X_1^2 = \mathbb{E}X_1^4 - \left(\mathbb{E}X_1^2\right)^2 = \frac{20}{\theta^4}$

То есть матрица коварИций $\Sigma_Z(\theta) = \begin{pmatrix} \frac{1}{\theta^2} & \frac{4}{\theta^3} \\ \frac{4}{\theta^3} & \frac{20}{\theta^4} \end{pmatrix}$

В качестве функции возьмем $\tau(x, y) = \frac{y}{2x^3}$

$\frac{\partial\tau}{\partial x} = -\frac{3y}{2x^4}, \quad \frac{\partial\tau}{\partial y} = \frac{1}{2x^3}$

$\tau(\theta) = \tau\left(\frac{1}{\theta}, \frac{2}{\theta^2}\right) = \frac{\frac{2}{\theta^2}}{2 \cdot \frac{1}{\theta^3}} = \theta$

То есть $\tau\left(\hat{\theta}\right)$ - а.н.о. $\tau(\theta)$, где $\hat{\theta} = \frac{\overline{X^2}}{2\left(\overline{X}\right)^3}$

$$\sigma^2(\theta) = D(\theta)\Sigma_Z(\theta)D(\theta)^T = \begin{pmatrix} \frac{\partial\tau}{\partial x} & \frac{\partial\tau}{\partial y} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\theta^2} & \frac{4}{\theta^3} \\ \frac{4}{\theta^3} & \frac{20}{\theta^4} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial\tau}{\partial x} \\ \frac{\partial\tau}{\partial y} \end{pmatrix} = \frac{\theta^4}{4} - \frac{51}{8}\theta^2 + \frac{9}{20}$$

I - 4.2 Метод моментов

Пусть $X_1, ..., X_n \sim P_\theta \in \left\{P_\theta \,\middle|\, \theta \in \Theta \subseteq \mathbb{R}^d\right\}$

Рассмотрим набор борелевских функций $g_1, ..., g_d$, таких что $\mathbb{E}_\theta|g_i(X_1)| < +\infty$

Составим систему:

$$\begin{cases} \mathbb{E}_\theta g_1(X) = \overline{g_1(X)} \\ \dots \\ \mathbb{E}_\theta g_d(X) = \overline{g_d(X)} \end{cases}$$

Тогда решение этой системы будет называться оценкой θ по обобщенному методу моментов

Задача:

$X_1, \ldots, X_n \sim \mathcal{N}(a, \sigma^2)$

$\mathbb{E}_{a,\sigma^2} X_1 = a$

$\mathbb{E}_{a,\sigma^2} X_1^2 = \sigma^2 + a^2$

> **Замечание**
> Часто за g в методе моментов берут $g(x) = x$ или $x^2, \ldots, x^k$

Возьмем $g_1(x) = x, g_2(x) = x^2$

$$\begin{cases} a = \overline{X} \\ \sigma^2 + a^2 = \overline{X^2} \end{cases} \Rightarrow \begin{cases} a = \overline{X} \\ \sigma^2 = \overline{X^2} - a^2 = \overline{X^2} - \left(\overline{X}\right)^2 = S^2 \end{cases}$$

Тогда $\hat{\theta} = \left(\hat{a}, \widehat{\sigma^2}\right) = \left(\overline{X}, S^2\right)$ - оценка метода моментов

Ответ: $\hat{\theta} = \left(\hat{a}, \widehat{\sigma^2}\right) = \left(\overline{X}, S^2\right)$

I - 5 Статистики и оценки. Доказательства теорем

I - 5.1 Линейная регрессия

Рассмотрим задачу линейной регресии

$y(x) = x^T \theta, \theta \in \mathbb{R}^d$

$Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} = x\theta + \varepsilon$

$Y \in \mathbb{R}^n$ - отлики, они случайны; $\varepsilon \in \mathbb{R}^n$ - шум, тоже случаен; $x \in \mathbb{R}^{n \times d}, \theta \in \mathbb{R}^d$

Тогда оценкой параметра по МНК будет $\hat{\theta}_{\text{MSE}} = \left(x^T x\right)^{-1} x^T Y$

Утв

1) Если $\mathbb{E}\varepsilon = 0$, то $\hat{\theta}_{\text{MSE}}$ – несмещенная оценка θ

   $\mathbb{E}\left(x^T x\right)^{-1} x^T Y = \mathbb{E}\left(x^T x\right)^{-1} x^T x \theta = \theta$

2) Если $\mathbb{E}\varepsilon = 0, \mathbb{D}\varepsilon = \sigma^2 I_n$, то $\widehat{\sigma^2} = \frac{\left\|Y - x\hat{\theta}_{\text{MSE}}\right\|}{n-d}$ – несмещенная оценка σ^2

$$\mathbb{E}\widehat{\sigma^2} = (n-d)\mathbb{E}\sum_{i=1}^{n}\left(Y_i - x_i^T\hat{\theta}_{\text{MSE}}\right)^2$$

Так как $\mathbb{E}Y_i = \mathbb{E}x_i^T\theta = \mathbb{E}x_i^T\hat{\theta}_{\text{MSE}}$, то

$$\mathbb{E}\widehat{\sigma^2} = \frac{1}{n-d}\sum_{i=1}^{n}\mathbb{D}\left(Y_i - x_i^T\hat{\theta}_{\text{MSE}}\right) = \frac{1}{n-d}\operatorname{Tr}\left(\mathbb{D}\left(Y - x\hat{\theta}_{\text{MSE}}\right)\right)$$

$$\mathbb{D}\left(Y - x\hat{\theta}_{\text{MSE}}\right) = \mathbb{D}\left(Y - x\left(x^Tx\right)^{-1}x^TY\right) = \mathbb{D}\left(\left[I_n - \underbrace{x\left(x^Tx\right)^{-1}x^T}_{A}\right]Y\right) =$$

$$= \mathbb{D}[(I_n - A)Y] = (I_n - A)\mathbb{D}Y(I_n - A)^T = (I_n - A)\sigma^2 I_n(I_n - A)^T =$$

$$= \left(I_n - 2A + \underbrace{A^TA}_{A}\right)\sigma^2 = (I_n - A)\sigma^2$$

$$\operatorname{Tr}(I_n) = n, \ \operatorname{Tr}(A) = d$$

Значит, подставляя это, получим:

$$\mathbb{E}\widehat{\sigma^2} = \frac{1}{n-d}\operatorname{Tr}\left(\sigma^2(I_n - A)\right) = \sigma^2$$

I - 5.2 Наследование свойств. Продолжение

Теорема (о наследовании сходимостей)

Пусть $\{\xi_n\}, n \in \mathbb{N}$ - последовательность случайных величин, ξ - случайная величина размерности d.

1) $\xi_n \xrightarrow{P} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная на множества B функция, причем $P(\xi \in B) = 1$, то $h(\xi_n) \xrightarrow{P} h(\xi)$
2) $\xi_n \xrightarrow{\text{п.н.}} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная на множества B функция, причем $P(\xi \in B) = 1$, то $h(\xi_n) \to$ п.н. $h(\xi)$
3) Если $\xi_n \xrightarrow{d} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывная, то $h(\xi_n) \xrightarrow{d} h(\xi)$

▲

2) $\xi_n \xrightarrow{\text{п.н.}} \xi$

$$P(h(\xi_n) \to h(\xi)) \stackrel{?}{=} 1$$

$$P\left(\lim_{n\to\infty} h(\xi_n) = h(\xi)\right) \geq P\left(\lim_{n\to\infty} \xi_n = \xi, \xi \in B\right) = 1$$

То есть получили требуемое

1) $h(\xi_n) \xrightarrow{P} h(\xi) \Leftrightarrow \forall \varepsilon > 0 \ \ P(|h(\xi_n) - h(\xi)|) \to 0 \Leftrightarrow$

$$\Leftrightarrow \forall \varepsilon > 0 \, \forall \delta > 0 \, \exists N \in \mathbb{N} \, \forall n \geq N \ \ P(|h(\xi_n) - h(\xi)| > \varepsilon) < \delta$$

Докажем от противного: $\exists \varepsilon, \delta > 0 \; \exists$ подпоследовательность $\left\{\xi_{n_k}\right\}_{k=1}^{\infty}$: $P(|h(\xi_n) - h(\xi)| > \varepsilon) > \delta$

Но по свойствам сходимости по вероятности найдется еще одна подпоследовательность $\left\{\xi_{n_{k_s}}\right\}_{s=1}^{\infty}$ такая, что на этой подпоследовательность есть сходимость почти наверное. Теперь существование $\left\{\xi_{n_k}\right\}$ противоречит пункту 2 и получено противоречие

3) Возьмем функцию $f : \mathbb{R}^d \to \mathbb{R}$ ограниченную и непрерывную.

Рассмотрим $f(h(x))$, так как $h(x)$ непрерывна, то и композиция будет непрерывна, а так как $f(x)$ ограничена, то и композиция будет ограничена. Тогда по определению сходимости по распределению имеем

$\mathbb{E} f(h(\xi_n)) \to \mathbb{E} f(h(\xi))$, так как $\xi_n \xrightarrow{d} \xi$

Значит, $h(\xi_n) \xrightarrow{d} h(\xi)$

■

Теорема (лемма Слуцкого)

Пусть $\{\xi_n\}, \{\eta_n\}, n \in \mathbb{N}$ - две последовательности случайных величин, причем есть сходимость $\xi_n \xrightarrow{d} \xi, \eta_n \xrightarrow{d} C$, тогда

$\xi_n + \eta_n \xrightarrow{d} \xi + C, \xi_n \cdot \eta_n \to \xi \cdot C$

Замечание

Неправильное доказательство: $(\xi_n, \eta_n) \to (\xi, \eta)$, возьмем $h(x, y) = x + y$, тогда $\xi_n + \eta_n \xrightarrow{d} \xi + \eta$.

Покомпонентная сходимость по распределению не дает сходимость вектора(хотя в случае сходимости по вероятности и почти наверное дает), поэтому нельзя применить теорему о наследовании сходимостей

Теорема (о производной)

Пусть $\{\xi_n\}, \xi_n \in \mathbb{R}^d, n \in \mathbb{N}$, - последовательность случайных величин. Выполнены 3 условия:

1) $\xi_n \xrightarrow{d} \xi$

2) $h : \mathbb{R}^d \to \mathbb{R}$ - непрерывно дифференцируема в точке $a \in \mathbb{R}^d$

3) $b_n > 0, b_n \to 0$ - числовая бесконечно малая последовательность

Тогда $\frac{h(a+\xi_n b_n)-h(a)}{b_n} \xrightarrow{d} \frac{\partial h}{\partial x}\big|_a \cdot \xi$

▲

Докажем для одномерного случая

Рассотрим функцию

$$H(x) = \begin{cases} \frac{h(x+a)-h(a)}{x}, & x \neq 0 \\ h'(a), & \text{иначе} \end{cases}$$

Легко видеть, что $H(x)$ – непрерывная

По лемме Слуцкого $\xi_n \cdot b_n \xrightarrow{d} 0 \Rightarrow \xi_n b_n \xrightarrow{P} 0$

Тогда по теореме о наследовании сходимостей $H(\xi_n \cdot b_n) \xrightarrow{P} H(0) = h'(a)$

И еще раз применив лемму Слуцкого получаем

$\xi_n \cdot H(\xi_n \cdot b_n) \xrightarrow{d} \xi \cdot h'(a)$

■

Пример:

$\xi_1, ..., \xi_n$ - НОРСВ

$\mathbb{E}\xi_k = a \neq 0, \mathbb{D}\xi_k = \sigma^2, S_n = \sum_{i=1}^{n} \xi_i$

$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{d} ?$

По ЦПТ имеем сходимость:

$$\underbrace{\sqrt{n}\left(\frac{S_n}{n} - a\right)}_{\zeta_n} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

То есть $\zeta_n \xrightarrow{d} \zeta \sim \mathcal{N}(0, \sigma^2)$

Возьмем $h(x) = \frac{1}{x}$ и $b_n = \frac{1}{\sqrt{n}}$

$$\frac{h(a+b_n\zeta_n)-h(a)}{b_n} = \sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{d} -\frac{1}{a^2} \cdot \zeta \sim \mathcal{N}\left(0, \frac{\sigma^2}{a^4}\right)$$

Теорема (дельта-метод)

Пусть:

1) $\hat{\theta}$ - асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}^d$ с асимптотической матрицей коварииаций $\Sigma(\theta)$

2) $\tau : \mathbb{R}^d \to \mathbb{R}^m$ - непрерывно дифференцируема

Тогда $\tau\left(\hat{\theta}\right)$ - асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $D(\theta)\Sigma(\theta)D^T(\theta)$, где $D(\theta) = \frac{\partial \tau}{\partial \theta}(\theta)$

▲

$\hat{\theta}$ - а.н.о. $\Leftrightarrow \sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \Sigma(\theta))$

Применим теорему о производной

$a = \theta, h(x) = \tau(x), \xi_n = \sqrt{n}\left(\hat{\theta} - \theta\right)$

$\xi \sim \mathcal{N}(0, \Sigma(\theta)), b_n = \frac{1}{\sqrt{n}}$

$\frac{h(a+\xi_n \cdot b_n) - h(a)}{b_n} = \sqrt{n}\left(\tau\left(\hat{\theta}\right) - \tau(\theta)\right) \xrightarrow{d_\theta} \frac{\partial h}{\partial x} |_\theta \cdot \xi = D(\theta) \cdot \xi \sim$

$\sim \mathcal{N}(0, D(\theta)\Sigma(\theta)D^T(\theta))$

■

I - 5.3 Доверительные интервалы

Опр

1) Если $\Theta \subseteq \mathbb{R}$, то пара статистик $(T_1(X), T_2(X))$ называется доверительным интервалом уровня доверия α, если

   $P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq \alpha \; \forall \theta \in \Theta$

2) Если $\Theta \subseteq \mathbb{R}^d$, то статистика $S(X)$ называется доверительной областью уровня доверия α, если $P_\theta(\theta \in S(X)) \geq \alpha \; \forall \theta \in \Theta$

Замечание

Обычно $\alpha = 0.95$

I - 5.4 Метод центральной функции

Пусть $G(X, \theta)$ – функция, распределение которой не зависит от θ

Опр

$p \in (0, 1), p$ —квантилем называетс число $u_p = \min\{x : F(x) \geq p\}$

Возьмем $\alpha_1, \alpha_2 \in (0, 1) : \alpha_2 - \alpha_1 = \alpha$

g_j – квантиль α_j распределения $G(X, \theta)$

Введем $S(X) = \{\theta \in \Theta \,|\, g_1 \leq G(X, \theta) \leq g_2\}$

$P_\theta(\theta \in S(X)) = P_\theta(g_1 \leq G(X, \theta) \leq g_2) \geq \alpha_2 - \alpha_1 = \alpha$

Пример:

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$, где σ^2 – известна

Построем доверительный интервал уровня доверия α методом центральной функции

$X_i - a \sim \mathcal{N}(0, \sigma^2)$

$\overline{X} - a \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$

$\sqrt{n}\frac{\overline{X}-a}{\sigma} \sim \mathcal{N}(0,1) \Rightarrow G(X, a) = \sqrt{n}\frac{\overline{X}-a}{\sigma} \sim \mathcal{N}(0,1)$ – центральная функция

Возьмем $\alpha_2 = \frac{1+\alpha}{2}, \alpha_1 = \frac{1-\alpha}{2} \Rightarrow \alpha_2 - \alpha_1 = \alpha$

$Z_{\alpha_1} = -Z_{\alpha_2}$

$P_\theta\left(Z_{\frac{1-\alpha}{2}} \leq \sqrt{n}\frac{\overline{X}-a}{\sigma} \leq Z_{\frac{1+\alpha}{2}}\right) = \alpha$

Тогда доверительный интервал для a имеет вид

$a \in \left(\overline{X} - Z_{\frac{1+\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{1+\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$

I - 5.5 Точные доверительный интервалы в нормальной модели

В этом параграфе $X = (X_1, ..., X_n), X_i \sim \mathcal{N}(a, \sigma^2)$

1) Интервал для a, если σ известна

   $a \in \left(\overline{X} - Z_{\frac{1+\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{1+\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$

2) Интервал для σ, если a известно

   $\xi_i = \frac{X_i - a}{\sigma} \sim \mathcal{N}(0,1) \Rightarrow \sum_{i=1}^{n} \frac{(X_i-a)^2}{\sigma^2} \sim \chi_n^2$

   $P\left(\chi^2_{n,\frac{1-\alpha}{2}} \leq \sum_{i=1}^{n} \frac{(X_i-a)^2}{\sigma^2} \leq \chi^2_{n,\frac{1+\alpha}{2}}\right) = \alpha$

   $\sum_{i=1}^{n} \xi_i^2 \sim \chi_n^2$

   $\sigma \in \left(\sqrt{\frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{n,\frac{1-\alpha}{2}}}}, \sqrt{\frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{n,\frac{1+\alpha}{2}}}}\right)$

   Опр

   $\xi \sim \mathcal{N}(0,1), \eta \sim \chi_k^2$

   $\zeta = \frac{\xi}{\sqrt{\frac{\eta}{k}}} \sim T_k$ – распределение Стьюдента

3) Для a, если σ неизвестна

Теорема

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$

1) $\overline{X}$ и S^2 независимы
2) $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$
3) $\sqrt{n-1}\frac{\overline{X}-a}{S} \sim T_{n-1}$

▲

3) $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}, \sqrt{n}\frac{\overline{X}-a}{\sigma} \sim \mathcal{N}(0,1)$

Значит, по определению, $\frac{\sqrt{n}\frac{\overline{X}-a}{\sigma}}{\sqrt{\frac{nS^2}{\sigma^2(n-1)}}} = \sqrt{n-1}\frac{\overline{X}-a}{S} \sim T_{n-1}$

■

$$P\left(T_{n-1,\frac{1-\alpha}{2}} \leq \sqrt{n-1}\frac{\overline{X}-a}{S} \leq T_{n-1,\frac{1+\alpha}{2}}\right) = \alpha$$

Из симметрии распределения Стьюдента: $T_{n-1,\frac{1-\alpha}{2}} = -T_{n-1,\frac{1+\alpha}{2}}$

$$a \in \left(\overline{X} - \frac{T_{n-1,\frac{1+\alpha}{2}}}{\sqrt{n-1}} \cdot S, \overline{X} + \frac{T_{n-1,\frac{1+\alpha}{2}}}{\sqrt{n-1}} \cdot S\right)$$

Асимтотический доверительный интервал: $\left(\hat{\theta} - Z_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta} + Z_{\frac{1+\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right)$

4) Интервал для σ, если a неизвестно

   $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

   $P\left(\chi^2_{n-1,\frac{1-\alpha}{2}} \leq \frac{nS^2}{\sigma^2} \leq \chi^2_{n-1,\frac{1+\alpha}{2}}\right) = \alpha$

   $\sigma \in \left(\sqrt{\frac{nS^2}{\chi^2_{n-1,\frac{1+\alpha}{2}}}}, \sqrt{\frac{nS^2}{\chi^2_{n-1,\frac{1-\alpha}{2}}}}\right)$

5) Доверительная область для двух параметров

   В домашнем задании

Теорема (о разложении гауссовского вектора)

Пусть $\xi = (\xi_1, ..., \xi_n) \sim \mathcal{N}(a, \sigma^2 I_n)$

Пусть $\mathbb{R}^n = \alpha_1 \oplus ... \oplus \alpha_k$ – разложение в прямую сумму ортогональных подпространств

$\eta_j = \text{proj}_{\alpha_j}\xi$

Тогда

1) $\eta_1, ..., \eta_k$ – независимы в совокупности

2) $\mathbb{E}\eta_j = \mathrm{proj}_{\alpha_j}(a)$

3) $\frac{1}{\sigma^2}\left\|\eta_j - \mathbb{E}\eta_j\right\|^2 \sim \chi^2_{d_j}, d_j = \dim(\alpha_j)$

▲

Выберем ОНБ в $\mathbb{R}^n$ $(e_1, ..., e_n)$

Обозначим I_j – индексы базиса, соответствующие α_j

$B = (e_1, ..., e_n) \in \mathbb{R}^{n\times n}$

$\zeta_i = \langle \xi, e_i \rangle = e_i^T \xi$

$\zeta = \begin{pmatrix} \zeta_1 \\ \cdots \\ \zeta_n \end{pmatrix} = \begin{pmatrix} e_1^T \xi \\ \cdots \\ e_1^T \xi \end{pmatrix} = B^T \xi, \xi = B\zeta$(в силу ортогональности матрицы)

$\mathbb{E}\zeta = \mathbb{E}B^T\xi = B^T\mathbb{E}\xi = B^T a$

$\mathbb{D}\zeta = \mathbb{D}B^T\xi = B\mathbb{D}\xi B^T = B\sigma^2 I_n B^T = \sigma^2 I_n$

Значит, ζ - гауссовский вектор с независимыми компонентами

$\eta_j = \mathrm{proj}_{\alpha_j}(\xi) = \sum\limits_{i\in I_j} \langle \xi, e_i \rangle e_i = \sum\limits_{i \in I_j} \zeta_i \cdot e_i \Rightarrow$ компоненты ζ в разных η не пересекаются, значит, $\eta_1, ..., \eta_k$ независимы в совокупности

$\mathbb{E}\eta_j = \sum\limits_{i\in I_j} \mathbb{E}\langle \xi_i, e_i \rangle e_i = \mathrm{proj}_{\alpha_j}(a)$

$$\frac{1}{\sigma^2}\left\|\eta_j - \mathbb{E}\eta_j\right\|^2 = \frac{1}{\sigma^2}\left\|\sum_{i\in I_j} \underbrace{\langle \xi - a,\ e_i \rangle}_{\zeta_i - \mathbb{E}\zeta_i} e_i\right\|^2 = \frac{1}{\sigma^2}\sum_{i\in I_j}(\zeta_i - \mathbb{E}\zeta_i)^2 \sim \chi^2_{d_j}$$

■

Теорема

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$

1) $\overline{X}$ и S^2 независимы

2) $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

3) $\sqrt{n-1}\frac{\overline{X}-a}{S} \sim T_{n-1}$

▲

$\mathbb{R}^n = \alpha \oplus \alpha^\perp, \alpha = \mathrm{span}\begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix}$

$\mathrm{proj}_\alpha(X) : \arg\min_{c\in\mathbb{R}} \left\|X - \begin{pmatrix} c \\ \cdots \\ c \end{pmatrix}\right\|^2 \Rightarrow c = \overline{X} = \mathrm{proj}_\alpha(X) = \begin{pmatrix} \overline{X} \\ \cdots \\ \overline{X} \end{pmatrix}$

1.По предыдущей теореме $\begin{pmatrix} \overline{X} \\ \dots \\ \overline{X} \end{pmatrix}$ и $\begin{pmatrix} X_1 - \overline{X} \\ \dots \\ X_n - \overline{X} \end{pmatrix}$ независимы, значит и $\overline{X}$ и S^2 независимы

2.$\frac{1}{\sigma^2}\|\mathrm{proj}_{\alpha^\perp} - \mathbb{E}\mathrm{proj}_{\alpha^\perp} X\|^2 = \frac{1}{\sigma^2}\left\|\begin{pmatrix} X_1 - \overline{X} \\ \dots \\ X_n - \overline{X} \end{pmatrix}\right\|^2 = n\frac{S^2}{\sigma^2}$

$x = \mathrm{proj}_\alpha(x) = \mathrm{proj}_{\alpha^\perp}(x)$

■

I - 6 Доверительные интервалы. Семинар

I - 6.1 Доверительные интервалы и области

Опр

Интервал из статистик $(T_1(X), T_2(X))$ называется доверительным интервалом уровня доверия α для параметра θ, если $P_\theta(\theta \in (T_1(X), T_2(X))) \geq \alpha$

Опр

Последовательность статистик $\left(T_1^{(n)}(X), T_2^{(n)}(X)\right)$ называется асимптотическим доверительным интервалом уровня доверия α, если

$\liminf_{n\to\infty} P_\theta\left(T_1^{(n)}(X) \leq \theta \leq T_2^{(n)}(X)\right) \geq \alpha$

Если выполняется равенство, то такой доверительный интервал называют точным.

Задача

$X_1, ..., X_n \sim \mathrm{Exp}^*(\theta)$

$p_\theta(x) = e^{-(x-\theta)}\mathbb{I}\{x > \theta\}$

$X_i - \theta \sim \mathrm{Exp}(1)$

$P_\theta\left(\theta \in \left[X_{(1)} - c_\alpha, X_{(1)}\right]\right) \geq \alpha \Leftrightarrow P_\theta\left(X_{(1)} - c_\alpha \leq \theta \leq X_{(1)}\right) \geq \alpha$

$P_\theta\left(X_{(1)} \leq \theta + c_\alpha\right) \geq \alpha \Leftrightarrow P_\theta\left(X_{(1)} \geq \theta + c_\alpha\right) \leq 1 - \alpha$

Распишем по независимости:

$\prod_{i=1}^n P_\theta(X_i \geq \theta + c_\alpha) \leq 1 - \alpha \Leftrightarrow \prod_{i=1}^n \left(1 - P_\theta\left(X_{(i)} \leq \theta + c_\alpha\right)\right) \leq 1 - \alpha$

Теперь перепишем последнее неравенство через функцию распределения

$$\prod_{i=1}^{n}(1-(1-e^{-c_\alpha})) \leq 1-\alpha \Leftrightarrow \overbrace{\prod_{i=1}^{n} e^{-c_\alpha}}^{e^{-n\cdot c_\alpha}} \leq 1-\alpha \Leftrightarrow c_\alpha \geq -\frac{\ln(1-\alpha)}{n}$$

Если же $c_\alpha = -\frac{\ln(1-\alpha)}{n}$, то доверительный интервал будет точным

Ответ: $\theta \in \left[X_{(1)} - \frac{\ln(1-\alpha)}{n}, X_{(1)}\right]$

I - 6.2 Метод центральных функций

Опр
Функция $G(X,\theta) : \mathcal{X} \times \Theta \to \mathbb{R}$ называется центральной, если ее распределение не зависит от θ

Рассмотрим $\alpha_2, \alpha_1 \in [0,1], \alpha_2 > \alpha_1$

Пусть $g_{\alpha_1}, g_{\alpha_2}$ - квантили распределения $G(X,\theta)$ уровней α_1 и α_2

$D = \left\{\theta : g_{\alpha_1} \leq G(X,\theta) \leq g_{\alpha_2}\right\}$ - доверительная область уровня α для θ

Мы хотим, чтобы $\left[g_{\alpha_1}, g_{\alpha_2}\right]$ был как можно уже, но $\alpha_2 - \alpha_1 = \alpha$

$P_\theta(\theta \in D) = P_\theta\left(g_{\alpha_1} \leq G(X,\theta) \leq g_{\alpha_2}\right) = \alpha_2 - \alpha_1 = \alpha$

Задача

$X_1, ..., X_n \sim \mathrm{Exp}^*(\theta)$

$\xi_i = X_i - \theta \sim \mathrm{Exp}(1)$

$\sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n} X_i - n\theta \sim \Gamma(1,n)$

$g_{\frac{1-\alpha}{2}}, g_{\frac{1+\alpha}{2}}$ – квантили $\Gamma(1,n)$

Тогда можно следующим образом записать доверительный интервал:

$P_\theta\left(g_{\frac{1-\alpha}{2}} \leq n\overline{X} - n\theta \leq g_{\frac{1+\alpha}{2}}\right) = \alpha$

Ответ: $\theta \in \left[\overline{X} - \frac{g_{\frac{1+\alpha}{2}}}{n}, \overline{X} - \frac{g_{\frac{1-\alpha}{2}}}{n}\right]$

Итого, мы рассмотрели два подхода построения интервалов:

$\theta \in \left[\overline{X} - \frac{g_{\frac{1+\alpha}{2}}}{n}, \overline{X} - \frac{g_{\frac{1-\alpha}{2}}}{n}\right]$, $\theta \in \left[X_{(1)} - \frac{\ln(1-\alpha)}{n}, X_{(1)}\right]$

I - 6.3 Доверительный интервал Вильсона

Рассматриваем распределение $\mathrm{Bern}(\theta)$

- Что было в интервале Вальда:

$\overline{X} \approx \mathcal{N}\left(\theta, \frac{\mathbb{D}X_i}{n}\right), \mathbb{D}X_i \sim \overline{X}\left(1 - \overline{X}\right)$

По ЦПТ имеем:

$\sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\mathbb{D}X_i}} \xrightarrow{d} \mathcal{N}(0,1)$

$P_\theta\left(-Z_{\frac{1+\alpha}{2}} \leq \sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\mathbb{D}X_i}} \leq Z_{\frac{1+\alpha}{2}}\right) = \alpha$

В интервале Вальда мы заменяли дисперсию на ее состоятельную оценку, $\widehat{\mathbb{D}X_i} = \overline{X}\left(1 - \overline{X}\right)$

$\sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\widehat{\mathbb{D}X_i}}} = \sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\mathbb{D}X_i}} \cdot \frac{\sqrt{\mathbb{D}X_i}}{\sqrt{\widehat{\mathbb{D}X_i}}} \xrightarrow{d} \mathcal{N}(0,1)$(по лемме Слуцкого)

- Что будем делать сейчас?

Будем использовать честную дисперсию и уже из всего этого выражать θ и получать доверительный интервал

$\overline{X} \approx \mathcal{N}\left(\theta, \frac{\mathbb{D}X_i}{n}\right)$

$\mathbb{D}X_i = \theta(1-\theta)$

По ЦПТ:

$\sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\theta(1-\theta)}} \xrightarrow{d} \mathcal{N}(0,1)$

$\left|\frac{\overline{X}-\theta}{\sqrt{\theta(1-\theta)}}\right| \leq Z_{\frac{1+\alpha}{2}}$

Решаем квадратное уравнение

$n\left(\overline{X}-\theta\right)^2 \leq Z^2_{\frac{1+\alpha}{2}}\theta(1-\theta)$

Ответ: $\theta \in \left[\frac{2n\overline{X}+Z^2_{\frac{1+\alpha}{2}}}{2n+2Z^2_{\frac{1+\alpha}{2}}} \pm \frac{1}{2n+2Z^2_{\frac{1+\alpha}{2}}}\sqrt{Z^4_{\frac{1+\alpha}{2}} + 4nZ^2_{\frac{1+\alpha}{2}}\overline{X}\left(1-\overline{X}\right)}\right]$

Замечание

В случае интервала Вальда замена дисперсии на ее оценку приводит к тому, что при малых n доверительный интервал широкий. В интервале Вильсона дисперсия считается честно и ошибка на малых n меньше, интервал уже

Главное отличие интервала Вильсона – не заменяем дисперсию на ее состоятельную оценку

I - 7 Метод максимального правдоподобия

I - 7.1 Метод максимального правдоподобия

$X_1, ..., X_n \sim \mathrm{Exp}(\theta)$

Посмотрим как мы раньше искали оценки

1) $\hat{\theta}_1 = \frac{1}{\overline{X}}$ - а.н.о. θ с а.д. $\sigma_1^2(\theta) = \theta^2$
2) Найдем оценку по методу поментов с $g(x) = \mathbb{I}\{x > 1\}$

   $\mathbb{E}_\theta \mathbb{I}\{X_1 > 1\} = \int\limits_1^{+\infty} p_{\theta(x)} dx = e^{-\theta}$

   По УЗБЧ: $\overline{\mathbb{I}\{X > 1\}} \overset{P_\theta \text{ п.н.}}{\to} e^{-\theta}$

   По ЦПТ: $\sqrt{n}\left(\overline{\mathbb{I}\{X > 1\}} - e^{-\theta}\right) \overset{d_\theta}{\to} \mathcal{N}(0, \sigma_2^2(\theta))$

   $\mathbb{D}_\theta \mathbb{I}\{X_1 > 1\} = \mathbb{E}_\theta \mathbb{I}^2\{X_1 > 1\} - (\mathbb{E}_\theta \mathbb{I}\{X_1 > 1\})^2 = \mathbb{E}_\theta \mathbb{I}\{X_1 > 1\} - e^{-2\theta} =$
   $= e^{-\theta} - e^{-2\theta}$

   Теперь применим дельта-метод с $\tau(x) = -\ln(x)$

   $-\ln\left(\overline{\mathbb{I}\{X > 1\}}\right)$ - а.н.о. θ с а.д. $\sigma_2^2(\theta) = \left(\frac{\partial - \ln(x)}{\partial x} \mid_{e^{-\theta}}\right)^2 \cdot \left(e^{-\theta} - e^{-2\theta}\right) = e^{\theta} - 1$

   Итого, $\hat{\theta}_2 = -\ln\left(\overline{\mathbb{I}\{X > 1\}}\right)$ а.н.о. θ с а.д. $\sigma_2^2(\theta) = e^{\theta} - 1$

Сравнивания дисперсии этих мы видим, что скорее всего $\hat{\theta}_1$ лучше $\hat{\theta}_2$. Но теперь возникает две задачи:

1) Нужно минимизировать дисперсию оценки
2) Нужно каким-то образом научиться сравнивать оценки

Предложим другой способ генерации оценки

Пусть $X = (X_1, ..., X_n)$ – выборка из распределения P

$P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, где все P_θ либо дискретные, либо абсолютно непрерывные

> **Опр**
> Семейство распределений, в котором либо все распределения дискретные, либо все абсолютно непрерывные называется доминируемым семейством

Чтобы и в дискретном случае работали все выкладки с плотностью, определим ее следующим образом $p_\theta(x) = P_\theta(X_1 = x)$

> **Опр**
> $\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i)$ - функция правдоподобия(функция от θ)

$\ell_X(\theta) = \ln(\mathcal{L}_X(\theta)) = \sum\limits_{i=1}^{n} \ln(p_\theta(X_i))$ - логарифм правдоподобия

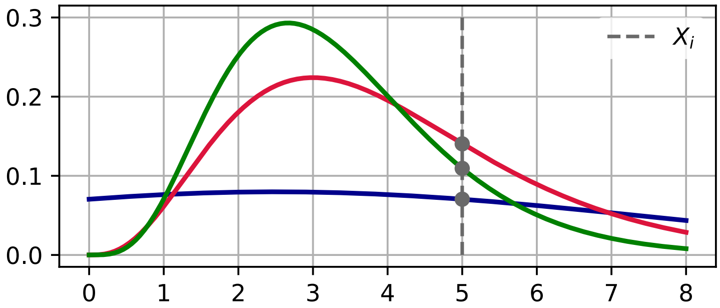


График 3: Иллюстрация метода правдоподобия

Идейно мы понимаем, что у “настоящей” плотности значения в точках где в окрестности много элементов выборки большие, то есть среди плотностей хочется выбрать ту, у которой значение в таких точках максимальное

Замечание

$\mathcal{L}_X(\theta)$ - степень правдоподобия θ для данной выборки X, значит, разумно брать ее максимум

Опр

$\hat{\theta}_{\text{MLE}}(X) = \arg\max_{\theta\in\Theta} \mathcal{L}_X(\theta) = \arg\max_{\theta\in\Theta} \ell_X(\theta)$ - оценка максимального правдоподобия (ОМП, MLE)

Утв:

ОМП не зависит от способа параметризации.

Пусть $\hat{\theta}$ - ОМП для θ, а функция $\tau : \Theta \to \Psi$ – биекция, тогда $\tau\left(\hat{\theta}\right)$ – ОМП для $\tau(\theta)$

Пример:

$X_1, ..., X_n \sim \text{Bern}(\theta)$. Нужно найти ОМП для θ и для $\ln\left(\frac{\theta}{1-\theta}\right)$

$p_\theta(x) = \theta^x (1-\theta)^{1-x}$

$\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum\limits_{i=1}^{n} X_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} X_i} \to \max_\theta$

$\ell_X(\theta) = \left(\sum\limits_{i=1}^{n} X_i\right)\ln(\theta) - \left(n - \sum\limits_{i=1}^{n} X_i\right)\ln(1-\theta)$

$\ell'_X(\theta) = \frac{\sum_{i=1}^{n} X_i}{\theta} - \frac{n-\sum_{i=1}^{n} X_i}{1-\theta} = 0 \Rightarrow \hat{\theta}_{\text{MLE}} = \overline{X}$

Тогда по утверждению $\ln\left(\frac{\overline{X}}{1-\overline{X}}\right)$ – ОМП для $\ln\left(\frac{\theta}{1-\theta}\right)$

Теорема

Пусть $\forall n \forall X_1, ..., X_n$ уравнение правдоподобия $\frac{\partial \ell_X}{\partial \theta} = 0$ имеет единственное решение

Тогда:

1) Если выполнены условия регулярности [L1-L5], то ОМП – состоятельная оценка θ
2) Если выполнены условия регулярности [L1-L9], то ОМП – асимптотически нормальная оценка θ с асимптотической матрицей ковариаций $i^{-1}(\theta)$, где $(i(\theta))_{jk} = \mathbb{E}_\theta \frac{\partial \ell_{X_1}}{\partial \theta_j} \cdot \frac{\partial \ell_{X_1}}{\partial \theta_k}$, здесь ℓ_{X_1} логарифм функции правдоподобия при $n = 1$

$i(\theta)$ называется информационной матрицей Фишера

Применим эту теореме для оценки для распределения Бернулли

$i(\theta) = \mathbb{E}_\theta \left(\frac{\partial \ell_{X_1}}{\partial \theta}\right)^2 = \mathbb{E}_\theta \left(\frac{X_1}{\theta} - \frac{1-X_1}{1-\theta}\right)^2 = \frac{1}{\theta^2 (1-\theta)^2} \mathbb{E}_\theta (X_1 - \theta)^2 =$

$= \frac{1}{\theta^2 (1-\theta)^2} \mathbb{D}_\theta X_1 = \frac{1}{\theta(1-\theta)}$

И дисперсия оценки $\sigma^2(\theta) = i^{-1}(\theta) = \theta(1 - \theta)$

Пример:

$X_1, ..., X_n \sim \text{Exp}(\theta)$

Нужно найти ОМП и ее дисперсию

$p_\theta(x) = \theta e^{-\theta x} \mathbb{I}\{x > 0\}$

$\mathcal{L}_X(\theta) = \prod_{i=1}^n p_\theta(X_i) = \theta^n e^{-\theta \sum_{i=1}^n X_i} \mathbb{I}\left\{X_{(1)} > 0\right\}$

Теперь найдем логарифм правдоподобия

$\ell_X(\theta) = n \ln(\theta) - \theta \sum_{i=1}^n X_i$

$\frac{\partial \ell_X}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n X_i = 0 \Leftrightarrow \theta = \frac{1}{\overline{X}}$

Тогда $\hat{\theta}_{\text{MLE}} = \frac{1}{\overline{X}}$

Посчитаем дисперсию

$i(\theta) = \mathbb{E}_\theta \left(\frac{1}{\theta} - X_1\right)^2 = \mathbb{D}_\theta X_1 = \frac{1}{\theta^2}$

$i^{-1}(\theta) = \theta^2$

I - 7.2 Задача про γ-котиков

На высоте 1 метр от точки θ находится источник γ-излучения, который излучает безмассовых γ-котиков, причем направления траекторий γ-котиков случайны, т.е. равномерно распределены по полуокружности.

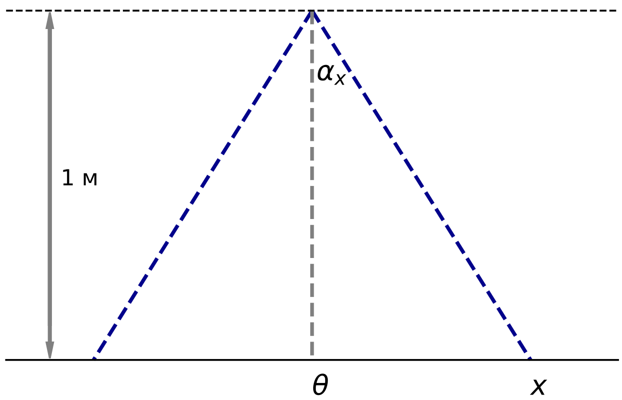


График 4: Иллюстрация модели измерений

x – точка пересечения с осью, α_x – угол

Запишем функцию распределения случайной величины в точке x

$F_\theta(x) = P_\theta(X \leq x) = \frac{\frac{\pi}{2}+\alpha_x}{\pi}$ – в предположении, что $\alpha_x > 0$

$F_\theta(x) = \frac{1}{2} + \frac{\alpha_x}{\pi} = \frac{1}{2} + \frac{1}{\pi}\arctan(x-\theta)$

Это нечто иное, как распределение Коши со сдвигом θ

$F'_\theta(x) = p_\theta(x) = \frac{1}{\pi\left(1+(x-\theta)^2\right)}$

Что если мы хотим оценить θ?

1) $\hat{\theta} = \overline{X}$

   Но у распределения Коши нет матожидания и дисперсии, поэтому мы не можем пользоваться ЦПТ

   Покажем, что в случае распределения Коши вообще нет смысла усреднять

   $\varphi_X(t) = \mathbb{E}e^{itX}$ - характеристическая функция, для распределения Коши равна $e^{-|t|}$

   $$\varphi_{\overline{X}}(t) = \mathbb{E}e^{it\frac{1}{n}\cdot\sum\limits_{i=1}^{n}X_i} = \prod_{i=1}^{n}\mathbb{E}e^{it\frac{X_1}{n}} = \prod_{i=1}^{n}\varphi_X\left(\frac{t}{n}\right) = \left(e^{-\frac{|t|}{n}}\right)^n = e^{-|t|}$$

То есть $\overline{X} \overset{d}{=} X_1$ и усреднение не дает ничего хорошего

2) Попробуем найти ОМП

$\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} \frac{1}{\pi\left(1+(X_i-\theta)^2\right)} = \frac{1}{\pi^n} \prod\limits_{i=1}^{n} \frac{1}{1+(X_i-\theta)^2}$

$\ell_X(\theta) = -n\ln(\pi) - \sum\limits_{i=1}^{n} \ln\left(1+(X_i-\theta)^2\right)$

В общем случае после дифференцирования нельзя будет решить уравнение правдоподобия

3) Попробуем оценить θ медианой по выборке

I - 7.3 Выборочные квантили

<u>Опр</u>
Пусть P – распределение на $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, $p \in (0, 1)$. Тогда p-квантилем называется $u_p = \min\{x \mid F(x) \geq p\}$

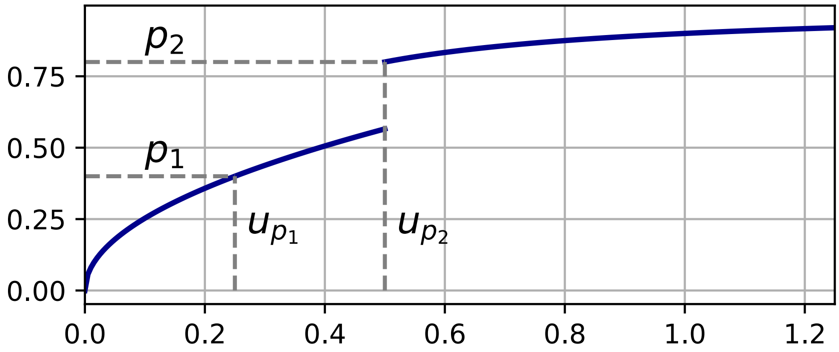


График 5: Иллюстрация квантилей

<u>Опр</u>
$\frac{1}{2}$-квантиль называется медианой, $\frac{1}{4}$-квантиль называется квартилем

$\widehat{u_\alpha} = X_{(\lceil n\alpha \rceil)}$ – выборочный α-квантиль

$\hat{\mu} = \begin{cases} X_{(k+1)}, & n=2k+1 \\ \frac{X_{(k)}+X_{(k+1)}}{2}, & n=2k \end{cases}$ – выборочная медиана

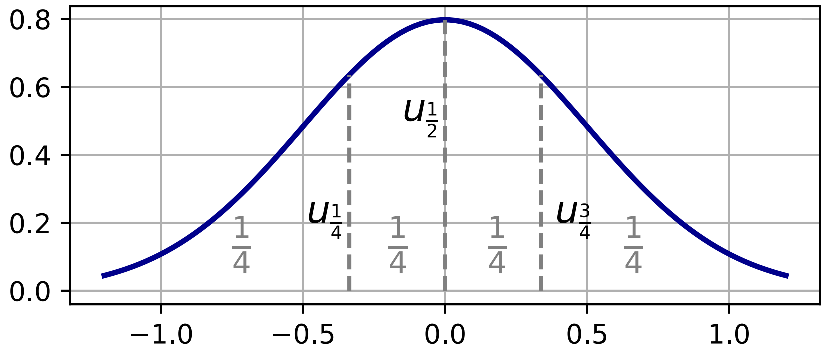


График 6: Квантили нормального распределения

Теорема (о выборочном квантиле)

Пусть $(X_n, n \in \mathbb{N})$ - выборка неограниченного размера из распределения P с плотностью $p(x)$, число $\alpha \in (0,1) : p(x)$ непрерывна в окрестности u_α и $p(u_\alpha) > 0$. тогда

1) $\sqrt{n}(\widehat{u_\alpha} - u_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{p^2(u_\alpha)}\right)$

2) $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4p^2(\mu)}\right)$

Для распределения Коши:

$\hat{\theta} = \hat{\mu}$ - а.н.о. θ с а.д. $\sigma^2 = \frac{1}{4\left(\frac{1}{\pi}(1+0)^2\right)^2} = \frac{\pi^2}{4} \approx 2.47$

Но если бы была ОМП, то ее дисперсия была бы $i(\theta) = \frac{1}{2} \Rightarrow \sigma^2(\theta) = 2$

I - 7.4 Достаточные статистики

Опр

Пусть $X = (X_1, ..., X_n)$ – выборка из неизвестного распределения $P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$

$S(X)$ – достаточная статистика для семейства $\mathcal{P}$, если условное распределение $P_\theta(X \in B \mid S(X))$ не зависит от θ $\forall B \in \mathfrak{B}\left(\mathbb{R}^d\right)$

Замечание

Вся информация о θ, которая есть в выборке, хранится в достаточной статистике

Если данные поступают последовательно, то можно только пересчитывать $S(X)$

Пример:

$X_1, ..., X_n \sim \mathrm{Bern}(\theta)$

Какая вообще есть информация в выборке?

1) Количество единиц $S(X) = \sum\limits_{i=1}^{n} X_i$

2) Порядок нулей и единиц

Покажем, что $S(X) = \sum\limits_{i=1}^{n} X_i$ – достаточная статистика

$$P_\theta\left(X_1 = x_1, ..., X_n = x_n \,\middle|\, \sum_{i=1}^{n} X_i = S\right) = \frac{P_\theta(X_1 = x_1, ..., X_n = x_n)}{P_\theta\left(\sum\limits_{i=1}^{n} X_i = S\right)} =$$

$$= \frac{\theta^{\sum\limits_{i=1}^{n} X_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} X_i}\mathbb{I}\left\{\sum\limits_{i=1}^{n} X_i = S\right\}}{\binom{n}{S}\theta^S(1-\theta)^{n-S}} = \frac{1}{\binom{n}{S}}\mathbb{I}\left\{\sum_{i=1}^{n} X_i = S\right\}$$

Значит, $S(X)$ – достаточная статистика

Теорема (критерий факторизации Неймана-Фишера)
Пусть $X = (X_1, ..., X_n)$ – выборка из распределения $P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, $\mathcal{P}$ – доминируемое распределение. Тогда

$S(X)$ – достаточная статистика $\Leftrightarrow p_\theta(x) = \Psi(S(x), \theta)h(x)$

h, Ψ – измеримые функции

▲
Рассмотрим только дискретный случай

$\Rightarrow$

$S(X)$ – достаточная статистика
$\Rightarrow p_\theta(x) = P_\theta(X = x) =$

$$= P_\theta(X = x, S(X) = S(x)) = \underbrace{P_\theta(X = x \mid S(X) = S(x))}_{h(x)} \cdot \underbrace{P_\theta(S(X) = S(x))}_{\Psi(S(x),\theta)}$$

$\Leftarrow$

$P_\theta(X = x \mid S(X) = S(x))$ не зависит от θ

Если $S(X) \neq S(x)$, то вероятность равна 0

$$P_\theta(X = x \mid S(X) = S(x)) = \frac{P_\theta(X = x, S(X) = S(x))}{P_\theta(S(X) = S(x))} = \frac{P_\theta(X = x)}{\sum\limits_{y: S(y) = S(x)} P_\theta(X = y)} =$$

$$= \frac{h(x)\Psi(S(x), \theta)}{\sum\limits_{y: S(y) = S(x)} \Psi(S(x) = S(y), \theta)h(y)} = \frac{h(x)}{\sum\limits_{y: S(y) = S(x)} h(y)}$$

Итоговое выражение не зависит от θ, значит, $S(X)$ – достаточная

■

Пример:

$X_1, ..., X_n \sim \Gamma(\alpha, \beta), \theta = (\alpha, \beta)$

$p_\theta(x) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x}$

$p_\theta(x_1, ..., x_n) = \frac{\alpha^{n\beta}}{\Gamma^n}(\beta) \left(\prod\limits_{i=1}^n x_i\right)^{\beta-1} e^{-\alpha \sum\limits_{i=1}^n x_i}$

Тогда по критерию факторизации Неймана-Фишера $\left(\prod\limits_{i=1}^n X_i, \sum\limits_{i=1}^n X_i\right)$ – достаточная статистика

I - 8 Метод максимального правдоподобия. Семинар

I - 8.1 Метод максимального правдоподобия

Опр

Семейство распределений, в котором либо все распределения дискретные, либо все абсолютно непрерывные называется доминируемым семейством

Чтобы и в дискретном случае работали все выкладки с плотностью, определим ее следующим образом $p_\theta(x) = P_\theta(X_1 = x)$

Опр

$\mathcal{L}_X(\theta) = \prod\limits_{i=1}^n p_\theta(X_i)$ - функция правдоподобия(функция от θ)

$\ell_X(\theta) = \ln(\mathcal{L}_X(\theta)) = \sum\limits_{i=1}^n \ln(p_\theta(X_i))$ - логарифм правдоподобия

Пример:

Оригинальне часы отстают на $\mathcal{N}(0, 1)$ мс, а поддельные – на $\mathcal{N}(0, 100^2)$. Мы померили и получили, что замедление равно 2 мс. Настоящие ли часы?

$P_{\mathcal{N}(0,1)}(|X| > 2) \leq 0.05$

Значит, видимо, часы поддельные

Посчитаем правдоподобие

$\mathcal{L}_{\mathcal{N}(0,1),X}(2) = 0.054$
$\mathcal{L}_{\mathcal{N}(0,100^2),X}(2) = 0.004$

Но функция правдоподобия у настоящих при таком отклонении больше в 13.5 раз. Тут никаких проблем нет, просто в данном случае отклонение 2 более маловероятно для $\mathcal{N}(0, 100^2)$, чем для $\mathcal{N}(0, 1)$

<u>Задача</u>

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$

Нужно найти ОМП для:

a. $\theta = (a, \sigma^2)$

b. $\theta = a$, если σ^2 – известна

c. $\theta = \sigma^2$, если a – известно

a. $\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (X_i - a)^2\right)$

$\ell_X(\theta) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma^2) - \frac{\sum\limits_{i=1}^{n} (X_i - a)^2}{2\sigma^2}$

$\frac{\partial \ell_X}{\partial a} = \frac{1}{\sigma^2} \sum\limits_{i=1}^{n} (X_i - a) = 0 \Rightarrow \hat{a} = \overline{X}$

$\frac{\partial \ell_X}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum\limits_{i=1}^{n} (X_i - a)^2 \Rightarrow \hat{\sigma} = \frac{1}{n} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2$

b.

$\frac{\partial \ell_X}{\partial a} = \frac{1}{\sigma^2} \sum\limits_{i=1}^{n} (X_i - a) = 0 \Rightarrow \hat{a} = \overline{X}$

c.

$\frac{\partial \ell_X}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum\limits_{i=1}^{n} (X_i - a)^2 \Rightarrow \hat{\sigma} = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - a)^2$

<u>Задача</u>

$X_1, ..., X_n \sim \mathrm{Geom}(\theta)$

$p_\theta(x) = (1 - \theta)^{x-1}\theta, x \in \mathbb{N}$

Посчитаем функцию правдоподобия

$\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} (1 - \theta)^{X_i - 1}\theta = \theta^n (1 - \theta)^{\sum\limits_{i=1}^{n} X_i - n}$

$\ell_X(\theta) = n \ln(\theta) + \left(\sum\limits_{i=1}^{n} X_i - n\right) \ln(1 - \theta)$

Решим уравнение правдоподобия

$\frac{\partial \ell_X}{\partial \theta} = \frac{n}{\theta} - \frac{\sum\limits_{i=1}^{n} X_i - n}{1 - \theta} = \frac{n - n\theta - \theta \cdot \sum\limits_{i=1}^{n} X_i + n\theta}{\theta(1 - \theta)} = 0 \Rightarrow \hat{\theta} = \frac{1}{\overline{X}}$

$\boxed{\textbf{Ответ: } \hat{\theta}_{\text{MLE}} = \frac{1}{\overline{X}}}$

Задача

$X_1, ..., X_n \sim U[0, \theta]$

$p_\theta(x) = \frac{1}{\theta}\mathbb{I}\{0 \leq x \leq \theta\}$

Найдем функцию правдоподобия

$\mathcal{L}_X(\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{I}\{0 \leq X_i \leq \theta\} = \frac{1}{\theta^n}\mathbb{I}\left\{0 \leq X_{(1)}\right\} \cdot \mathbb{I}\left\{X_{(n)} \leq \theta\right\}$

Здесь уже искать логарифм правдоподобия нет особого смысла. Мы хотим максимизировать функцию правдоподобия, $\frac{1}{\theta^n}$ убывает по θ. Значит нам нужно наименьшее θ, при котором произведение индикаторов не равно нулю

То есть $\hat{\theta} = X_{(n)}$

$\boxed{\textbf{Ответ: } \hat{\theta} = X_{(n)}}$

Основные свойства ОМП:

1) Если $\hat{\theta}$ – ОМП для θ, а фунция $\tau : \Theta \to \Psi$ – биекция, то $\tau\left(\hat{\theta}\right)$ – ОМП для $\tau(\theta)$
2) Если выполнены условия регулярности L1-L5, то единственность решения уравнения правдоподобия делает это решение ОМП с вероятностью $\to 1$
3) Если выполнены условия регулярности L1-L5, то ОМП – состоятельная оценка
4) Если выполнены условия регулярности L1-L9, то ОМП асимптотически нормальная оценка θ с асимптотической дисперсией

   асимптотической матрицей ковариаций $i^{-1}(\theta)$, где $(i(\theta))_{jk} = \mathbb{E}_\theta \frac{\partial \ell_{X_1}}{\partial \theta_j} \cdot \frac{\partial \ell_{X_1}}{\partial \theta_k}$. Причем у этой оценки наименьшая дисперсия

Пример:

$X_1, ..., X_n \sim \mathcal{N}(0, 1)$

$\overline{X}$ – ОМП для θ, тогда при $\tau(x) = e^x$ имеем, что $\tau\left(\overline{X}\right)$ – ОМП e^θ

Задача

$X_1, ..., X_n \sim U[a, b]$

$\theta = (a, b)$

Нужно найти ОМП для $\mathbb{E}_\theta X_1 = \frac{a+b}{2}$

$p_\theta(x) = 1(b - a)\mathbb{I}\{a \leq x \leq b\}$

$\mathcal{L}_X(\theta) = \frac{1}{(b-a)^n}\mathbb{I}\left\{a \le X_{(1)}\right\} \cdot \mathbb{I}\left\{X_{(n)} \le b\right\}$

$\hat{a} = X_{(1)}, \hat{b} = X_{(n)}$

Тогда теперь легко получить ответ

Задача

Нужно найти асимптотическую дисперсию для следующих оценок

1) $X_1, ..., X_n \sim \text{Geom}(\theta), \hat{\theta} = \frac{1}{\overline{X}}$

$$i(\theta) = \mathbb{E}_\theta\left(\frac{\partial \ell_{X_1}}{\partial \theta}\right)^2 = \mathbb{E}_\theta\left(\frac{1}{\theta} - \frac{X_1 - 1}{1-\theta}\right)^2 = \mathbb{E}_\theta\frac{(\theta X_1 - 1)^2}{\theta^2(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)^2}\mathbb{E}_\theta(\theta X_1 - 1)^2 =$$

$$= \frac{1}{(1-\theta)^2}\mathbb{E}_\theta\left(X_1 - \frac{1}{\theta}\right)^2 = \frac{1}{(1-\theta)^2}\mathbb{D}_\theta X_1 = \frac{1}{(1-\theta)^2} \cdot \frac{1-\theta}{\theta} = \frac{1}{(1-\theta)\theta}$$

Ответ: $i^{-1}(\theta) = \theta(1-\theta)$

2) $X_1, ..., X_n \sim \mathcal{N}(\theta, \sigma^2)$, σ^2 – известна

$$i(\theta) = \mathbb{E}_\theta\left(\frac{\partial \ell_{X_1}(\theta)}{\partial \theta}\right)^2 = \mathbb{E}_\theta\left(\frac{(X_1 - \theta)^2}{\sigma^2}\right) = \frac{1}{\sigma^4}\mathbb{E}_\theta(X_1 - \theta)^2 = \frac{1}{\sigma^4}\mathbb{D}X_1 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Ответ: $i^{-1}(\theta) = \sigma^2$

I - 8.2 Оценки для распределения Коши

Пусть $X_1, ..., X_n \sim \text{Cauchy}(\theta)$

Из лекции $\hat{\mu}$ - а.н.о. с а.д. $\frac{\pi^2}{4}$

Теорема (о выборочном квантиле)

Пусть $(X_n, n \in \mathbb{N})$ - выборка неограниченного размера из распределения P с плотностью $p(x)$, число $\alpha \in (0,1) : p(x)$ непрерывна в окрестности u_α и $p(u_\alpha) > 0$. тогда

1) $\sqrt{n}(\widehat{u_\alpha} - u_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{p^2(u_\alpha)}\right)$

2) $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4p^2(\mu)}\right)$

Пусть $X_1, ..., X_n \sim \text{Cauchy}(\theta)$

Причем θ не сдвиг, а масштаб

$p_\theta(x) = \frac{1}{\pi\theta\left(1 + \left(\frac{x}{\theta}\right)^2\right)}$

Найдем оценку для θ

$F_\theta(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x}{\theta}\right)$

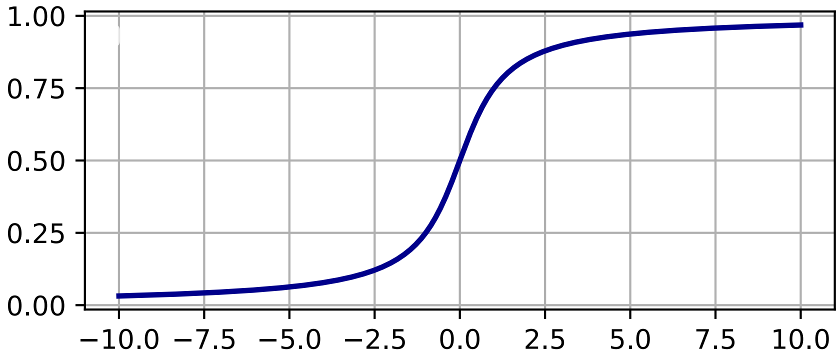


График 7: График функции распределения $\mathrm{Cauchy}(\theta)$

Рассмотрим выборку $|X_1|, ..., |X_n|$ и найдем ее медиану

$F_{|X|,\theta}(x) = \frac{2}{\pi} \arctan\left(\frac{x}{\theta}\right)$

Для медианы $F_{|X|,\theta}(\mu) = \frac{1}{2}$

Тогда решаем уравнения $\frac{2}{\pi} \arctan\left(\frac{\mu}{\theta}\right) = \frac{1}{2}$

$\frac{\mu}{\theta} = 1 \Rightarrow \mu = \theta$

Свели задачу к задаче из лекции

В качестве оценки возьмем $\hat{\mu}(|X_1|, ..., |X_n|)$

I - 9 Сравнение оценок

I - 9.1 Экспонецниальное семейство распределений

Опр

Семейство распределений $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ принадлежит экспоненциальному классу распределений, если

$p_\theta(x) = \frac{g(x)}{h(\theta)} \cdot e^{(a(\theta))^T u(x)}$

Где g, u – борелевские функции, а $h(\theta) = \int\limits_{\mathcal{X}} g(x) e^{(a(\theta))^T u(x)} dx$ – нормировочная компонента

Если $a(\theta) = \theta$, то параметризация называется естественной

Пример:

$\mathcal{P} = \left\{\mathcal{N}(a, \sigma^2) \,\middle|\, a \in \mathbb{R}, \sigma > 0\right\}, \theta = (a, \sigma)$

$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{xa}{\sigma^2} - \frac{a^2}{2\sigma^2}\right) =$

$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{a^2}{2\sigma^2}\right) \cdot e^{(a(\theta))^T u(x)}$

$a(\theta) = \begin{pmatrix} \frac{1}{2\sigma^2} \\ \frac{a}{\sigma^2} \end{pmatrix}, u(x) = \begin{pmatrix} -x^2 \\ x \end{pmatrix}$

Перейдем к естественной параметризации

$\theta_1 = \frac{1}{2\sigma^2}, \theta_2 = \frac{a}{\sigma^2} \Rightarrow \sigma^2 = \frac{1}{\theta_1}, a = \frac{\theta_2}{2\theta_1}$

$p_\theta(x) = \overbrace{\sqrt{\frac{\theta_1}{\pi}} \exp\left(-\frac{\theta_2^2}{2\theta_1}\right)}^{h^{-1}(\theta)} e^{\theta^T u(x)}$

Найдем достаточную статистику для распределения из экспоненциально класса

$X_1, ..., X_n : p_\theta(x_1, ..., x_n) = \prod\limits_{i=1}^{n} \frac{g(x_i)}{h(\theta)} e^{(a(\theta))^T u(x_i)} =$

$= h^{-n}(\theta) \left(\prod\limits_{i=1}^{n} g(x_i)\right) \exp\left((a(\theta))^T \sum\limits_{i=1}^{n} u(x_i)\right)$

$S(X) = \sum\limits_{i=1}^{n} u(X_i)$ – достаточная статистика по критерию Неймана-Фишера

Замечание

Как мы видим, достаточная статистика имеет фиксированную размерность

Теорема

$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ – семейство распределений, такое что $p_\theta(x)$ непрерывно дифференцируема и носитель не зависит от θ. Пусть $S(X)$ – достаточная статистика с фиксированной размерностью.

Тогда $\mathcal{P}$ принадлежит экспоненциальному классу.

Следствие

Если плотность "хорошая", то только экспоненциальный класс распределений допускает сжатие с помощью достаточной статистики

Пример:

1) Распределение Коши с $p_\theta(x) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$ не лежит в экспоненциальном классе, значит у него нет достаточной статистики
2) $U[0, \theta], S(X) = X_{(n)}$. Но здесь не выполнено условие на плотность, поэтому теорема не работает

Далее потребуем следующие условия

1) $a(\theta) = \theta$, то есть работаем с естественной параметризацией
2) $g(x), u(x)$ – непрерывные

3) Выполнено условие равномерной сходимости интеграла

$$\forall S\ \forall j \le k\ \exists \varphi(x) : \forall \theta \in \Theta \left| g(x) u_s^j(x) e^{\theta^T u(x)} \right| \le \varphi(x)$$

Следствие

1) $h(\theta)$ – непрерывно дифференцируема k раз

2) $p_\theta(x)$ непрерывно дифференцируема по θ k раз

3) Можно менять местами $\frac{\partial}{\partial \theta}$ и $\int$

Теорема

1) $\mathbb{E}_\theta u(X_1) = \nabla \ln(h(\theta))$

2) $\mathbb{D}_\theta u(X_1) = \left(\frac{\partial^2 \ln(h(\theta))}{\partial \theta_j \partial \theta_k}\right)_{jk}$

▲

1)

$$\frac{\partial h(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \int\limits_{\mathcal{X}} g(x) \exp(\theta^T u(x)) dx = \int\limits_{\mathcal{X}} g(x) u_j(x) e^{\theta^T u(x)} dx =$$

$$= \int\limits_{\mathcal{X}} g(x) u_j(x) e^{\theta^T u(x)} dx = h(\theta) \int\limits_{\mathcal{X}} u_j(x) \underbrace{\frac{g(x)}{h(\theta)} e^{\theta^T u(x)}}_{p_\theta(x)} dx =$$

$$= h(\theta) \int\limits_{\mathcal{X}} u_j(x) p_{\theta(x)} dx = h(\theta) \mathbb{E}_\theta u_j(X_1)$$

$$\frac{1}{h(\theta)} \frac{\partial}{\partial \theta_j} h(\theta) = \mathbb{E}_\theta u(X_1) \Rightarrow \frac{\partial \ln(h(\theta))}{\partial \theta_j} = \mathbb{E}_\theta u(X_1)$$

■

Теорема

Если Θ выпуклое, то ОМП существует и единственна

▲

$$\mathcal{L}_X(\theta) = h^{-n}(\theta) \left(\prod_{i=1}^n g(X_i) \right) e^{\theta^T \sum\limits_{i=1}^n u(X_i)}$$

$$\ell_X(\theta) = -n \ln h(\theta) + \sum_{i=1}^n \ln(g(X_i)) + \theta^T \left(\underbrace{\sum_{i=1}^n u(X_i)}_{\text{линейная функция}} \right)$$

$\nabla^2 \ln(h(\theta)) = \mathbb{D}_\theta u(X_1) \ge 0 \Rightarrow$ по вогнутости получаем существование и единственность

■

Теорема
Если Θ – выпуклое открытое множество, то условия L1-L9 выполняются

▲
L1-L7 выполняются из следствий и определений

L8:

$$(i(\theta))_{ij} = \mathbb{E}_\theta \frac{\partial \ell_{X_1}(\theta)}{\partial \theta_i \partial \theta_j}$$

$$\ell_{X_1}(\theta) = -\frac{\partial}{\partial \theta_j} \ln(h(\theta)) + u(X_1)$$

$$i(\theta) = \mathbb{E}_\theta \left(u(X_1) - \frac{1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta}\right)^2$$

L9:

$$\frac{\partial^3 \ln(p_\theta(x))}{\partial \theta^3} = - \underbrace{\frac{\partial^2}{\partial \theta^2} \frac{\frac{\partial h}{\partial \theta}}{h(\theta)}}_{\text{не зависит от X}}$$

■

I - 9.2 Сравнение оценок

Опр
Функция $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$, которая характеризует степень отличия оценки $\hat{\theta}$ от $\tau(\theta)$ называется функцией потерь (Loss function)

Примеры:

$d = 1$

1) $L(x, y) = |x - y|$

2) $L(x, y) = (x - y)^2$

3) $L(x, y) = \ln(1 + |x - y|)$

$d > 1$

$L(x, y) = (x - y)^T A (x - y)$

Причем, если $A = I_d$, то $L(x, y) = \sum_{i=1}^{n} (x_i - y_i)^2$, то есть простое MSE

$L\left(\hat{\theta}, \tau(\theta)\right)$ – штраф при оценивании $\tau(\theta)$ оценкой $\hat{\theta}$

Замечание
Оценка строится по выборке, штраф считается по оценке. То есть получается, что функция штрафа будет выдавать случайные значения

Опр

Функцией риска оценки $\hat{\theta}$ величины $\tau(\theta)$ называется $R_{\hat{\theta},\tau} = \mathbb{E}_\theta L\left(\hat{\theta}, \tau(\theta)\right)$

Если $\tau(\theta) = \theta$, то будем обозначать $R_{\hat{\theta}}$

Примеры:

1) $L(x, y) = (x - y)^2$

   $L\left(\hat{\theta}, \tau(\theta)\right) = \left(\hat{\theta} - \tau(\theta)\right)^2 \Rightarrow R_{\hat{\theta},\tau} = \mathbb{E}_\theta\left(\hat{\theta} - \tau(\theta)\right)^2 = \mathrm{MSE}_{\hat{\theta},\tau}(\theta)$

2) $L(x, y) = |x - y|$

   $\mathrm{MAE}_{\hat{\theta},\tau}(\theta) = \mathbb{E}_\theta\left|\hat{\theta} - \tau(\theta)\right|$

Пример:

Посчитаем MSE для двух разных оценок

$\tau(\theta) = \mathbb{E}_\theta X_1, \hat{\theta}_1 = X_1, \hat{\theta}_2 = \overline{X}$

$R_{\hat{\theta}_1} = \mathbb{E}_\theta\left(\hat{\theta}_1 - \mathbb{E}_\theta X_1\right)^2 = \mathbb{D}_\theta X_1$

$R_{\hat{\theta}_2} = \mathbb{E}_\theta\left(\hat{\theta}_2 - \mathbb{E}_\theta X_1\right) = \mathbb{D}_\theta \overline{X} = \frac{\mathbb{D}_\theta X_1}{n}$

Усреднение уменьшает риск в n раз

Рассмотрим разные подходы к сравнению оценок

1. Равномерный подход

Опр

1) $\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\forall \theta \in \Theta : R_{\hat{\theta}_1,\tau}(\theta) \leq R_{\hat{\theta}_2,\tau}(\theta)$
2) $\hat{\theta}_1$ лучше $\hat{\theta}_2$, если кроме того $\exists \theta \in \Theta : R_{\hat{\theta}_1,\tau}(\theta) < R_{\hat{\theta}_2,\tau}(\theta)$

Опр

Пусть K - класс оценок. $\hat{\theta}$ называется наилучшей в классе K, если она лучше всех в K

Если функция потерь – это $L(x, y) = (x - y)^2$, то такой подход называется среднеквадратичным

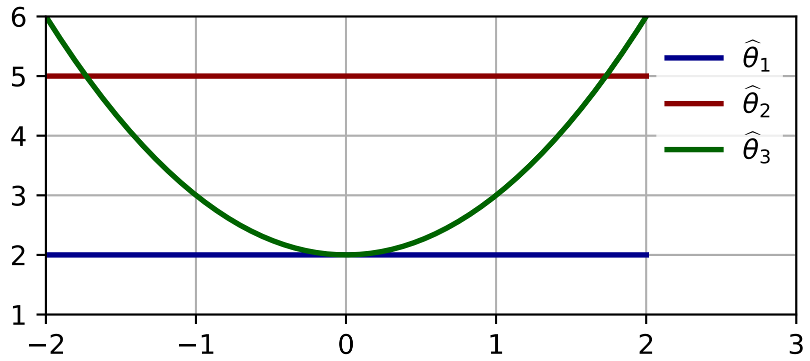


График 8: Иллюстрация среднеквадратичного подхода

По изображенным на графике функциям риска мы можем видеть, что $\hat{\theta}_1$ лучше $\hat{\theta}_2$, а также $\hat{\theta}_1$ лучше $\hat{\theta}_3$. Но вот уже $\hat{\theta}_2$ и $\hat{\theta}_3$ сравнить нельзя

Утв

Наилучшей оценки может не существовать

$\Theta = \mathbb{R}, \mathrm{K} = \left\{\hat{\theta}_1 = 1, \hat{\theta}_2 = 2\right\}, \tau(\theta) = \theta$

$\mathrm{MSE}_{\hat{\theta}_1}(\theta) = (1 - \theta)^2$

$\mathrm{MSE}_{\hat{\theta}_2}(\theta) = (2 - \theta)^2$

Графики этих фукнций риска – это две параболы, они пересекаются и сравнить их нельзя

Теорема

Справедливо bias-variance разложение

$$\mathrm{MSE}_{\hat{\theta},\tau} = \overbrace{\mathbb{D}_\theta \hat{\theta}}^{\text{variance}} + \overbrace{\left(\mathbb{E}_\theta \hat{\theta} - \tau(\theta)\right)^2}^{\text{bias}^2}$$

▲

$$\mathrm{MSE}_{\hat{\theta},\tau} = \mathbb{E}_\theta\left(\hat{\theta} - \tau(\theta)\right)^2 = \mathbb{E}_\theta\left(\left(\hat{\theta} - \mathbb{E}_\theta\hat{\theta}\right) + \left(\mathbb{E}_\theta\hat{\theta} - \tau(\theta)\right)\right)^2 =$$

$$= \mathbb{E}_\theta\left[\left(\hat{\theta} - \mathbb{E}_\theta\hat{\theta}\right)^2 + 2\left(\hat{\theta} - \mathbb{E}_\theta\hat{\theta}\right)\left(\mathbb{E}_\theta\hat{\theta} - \tau(\theta)\right) + \left(\mathbb{E}_\theta\hat{\theta} - \tau(\theta)\right)^2\right] =$$

$$= \mathbb{D}_\theta\hat{\theta} + \left(\mathbb{E}_\theta\hat{\theta} - \tau(\theta)\right)^2$$

Среднее слагаемое равно нулю, так как вторая скобка – константа, а матожидание $\mathbb{E}_\theta\left(\hat{\theta} - \mathbb{E}_\theta\hat{\theta}\right) = 0$

■

Следствие

В классе несмещенных оценок самая лучшая оценка – оценка с наименьшей дисперсией

2. Байесовский подход

Опр
Пусть Q распределение на Θ

$\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $\mathbb{E}_Q R_{\widehat{\theta}_1,\tau} \le \mathbb{E}_Q R_{\widehat{\theta}_2,\tau}$.

Если неравенство строгое, то оценка $\widehat{\theta}_1$ лучше

Замечание
Основная идея в том, что мы уже что-то знаем о θ, это знание и отражает распределение Q

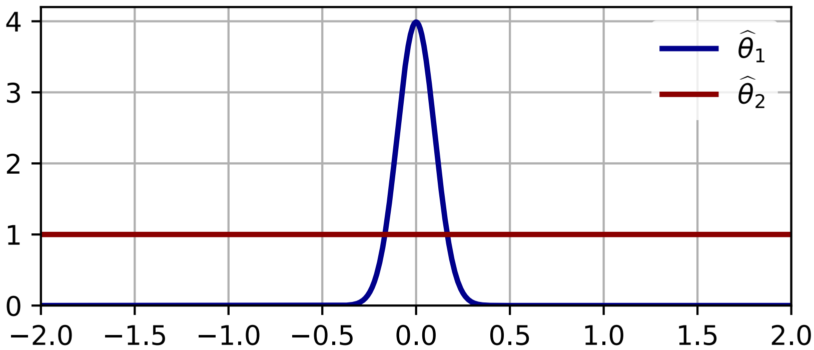


График 9: Иллюстрация байесовского подхода

Пусть $Q = U[-2, 2]$. В равномерном подходе мы бы не смогли сравнить такие оценки, но тут явно видим, что площадь под функцией риска $\widehat{\theta}_2$ больше, значит $\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$

3. Минимаксный подход

Опр
$\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $\sup\limits_{\theta\in\Theta} R_{\widehat{\theta}_1,\tau} \le \sup\limits_{\theta\in\Theta} R_{\widehat{\theta}_2,\tau}$.

Если неравенство строгое, то оценка $\widehat{\theta}_1$ лучше

В предудущем примере в таком подходе уже оценка $\widehat{\theta}_2$ лучшая, ведь супремум ее риска меньше

4. Асимптотический подход

Опр
$\hat{\theta}_1, \hat{\theta}_2$ - а.н.о. θ с а.д. $\sigma_1^2(\theta), \sigma_2^2(\theta)$

1) $\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\forall \theta \in \Theta : \sigma_1(\theta) \leq \sigma_2(\theta)$.
2) Если существует θ, в котором неравенство строгое, то оценка $\hat{\theta}_1$ лучше $\hat{\theta}_2$
3) $\mathrm{ARE}_{\hat{\theta}_1, \hat{\theta}_2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$ – относительная асимптотическая эффективность

Опр
Оценка $\hat{\theta}$ – асимптотически эффективная оценка $\tau(\theta)$, если у нее наименьшая а.д. среди всех а.н.о. с $\tau(\theta)$ с непрерывной дисперсией

Утв

Если L1-L9 выполнено, то ОМП – асимптотически эффективная оценка

Пример:

$X_1, ..., X_n \sim \mathcal{N}(\theta, 1)$

$\hat{\theta}_1 = \overline{X}$ - а.н.о. θ с а.д. $\sigma_1^2(\theta) = 1$

$\hat{\theta}_2 = \hat{\mu}$ - а.н.о. θ с а.д. $\sigma_2^2(\theta) = \frac{\pi}{2}$ (по теореме о выборочном квантиле)

$\mathrm{ARE}_{\hat{\theta}_1, \hat{\theta}_2} \approx 1.57$

То есть выборочная медиана оказалась хуже, так как оценка средним – ОМП

I - 10 Сравнение оценок. Семинар

$X_1, ..., X_n \sim P \in \{P_\theta \,|\, \theta \in \Theta\}$

Задача – оценить параметр θ

Общий принцп решения:

1) Построить мат модель
2) Выбрать множество оценок K
3) Задать численный критерий качества
4) Выбрать наилучшую по критерию оценку из K

Опр
$L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ - функция потерь

$L\left(\hat{\theta}, \tau(\theta)\right)$ – штраф при оценке $\tau(\theta)$ с помощью $\hat{\theta}$

$R_{\hat{\theta},\tau}(\theta) = \mathbb{E}_\theta L\left(\hat{\theta}, \tau(\theta)\right)$ – функция риска

$\text{MSE}_{\hat{\theta},\tau}(\theta) = \mathbb{E}_\theta \left(\hat{\theta} - \tau(\theta)\right)^2$

$\text{MAE}_{\hat{\theta},\tau}(\theta) = \mathbb{E}_\theta \left|\hat{\theta} - \tau(\theta)\right|$

Пример:

$X_1, ..., X_n \sim \text{Bern}(\theta), \hat{\theta} = \overline{X}$

$\text{MSE}_{\hat{\theta}}(\theta) = \mathbb{E}_\theta \left(\overline{X} - \theta\right)^2 = \mathbb{D}_\theta \overline{X} = \frac{\theta(1-\theta)}{n}$

Мы видим, что если реальное θ близко к концам отрезка $[0, 1]$, то функция риска меньше, чем при θ, близких к $\frac{1}{2}$

I - 10.1 Равномерный подход

Опр

1) $\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\forall \theta \in \Theta : R_{\hat{\theta}_1,\tau}(\theta) \leq R_{\hat{\theta}_2,\tau}(\theta)$

2) $\hat{\theta}_1$ лучше $\hat{\theta}_2$, если кроме того $\exists \theta \in \Theta : R_{\hat{\theta}_1,\tau}(\theta) < R_{\hat{\theta}_2,\tau}(\theta)$

Задача

$X_1, ..., X_n \sim U[0, \theta], \theta > 0$

$\text{K} = \left\{cX_{(n)} \,\middle|\, c \in \mathbb{R}\right\}$

$\mathbb{E}_\theta X_{(n)} = \frac{n}{n+1}\theta$

$\mathbb{E}_\theta X_{(n)}^2 = \int\limits_0^\theta x^2 \frac{nx^{n-1}}{\theta^n} dx = \int\limits_0^\theta \frac{nx^{n+1}}{\theta^n} dx = \frac{n}{n+2} \frac{x^{n+2}}{\theta^n} \ \Big|_0^\theta = \frac{n}{n+2}\theta^2$

Теперь можем выразить MSE:

$\text{MSE}_{cX_{(n)}}(\theta) = \mathbb{E}_\theta \left(cX_{(n)} - \theta\right)^2 = c^2 \mathbb{E}_\theta X_{(n)}^2 - 2c\theta \mathbb{E}_\theta X_{(n)} + \theta^2 =$

$= c^2 \frac{n}{n+2}\theta^2 - 2c\theta^2 \frac{n}{n+1} + \theta^2 = \theta^2 \left(\frac{n}{n+2}c^2 - \frac{2n}{n+1}c + 1\right) \to \min_c$

Просто найдем минимум как у квадратичной функции, получим:

$c_{\min} = \frac{2n}{(n+1)n} \cdot \frac{n+2}{2} = \frac{n+2}{n+1}$

Лучшая оценка в равномерном подходе по MSE: $\frac{n+2}{n+1}X_{(n)}$

Причем, как мы знаем, $\frac{n}{n+1}X_{(n)}$ – единственная несмещенная оценка такого вида, а $X_{(n)}$ – ОМП

Запишем еще bias-variance разложение

$\text{MSE}_{\hat{\theta}}(\theta) = \mathbb{D}_\theta \hat{\theta} + \left(\mathbb{E}_\theta \hat{\theta} - \theta\right)^2$

$\text{bias}^2 = \left(c\mathbb{E}_\theta X_{(n)} - \theta\right)^2 = \theta^2\left(c\frac{n}{n+1} - 1\right)^2$

$\text{variance} = \mathbb{D}_\theta c X_{(n)} = c^2 \frac{\theta^2 n}{(n+1)^2(n+2)}$

Замечание

Регулируя смещение мы можем получить меньшее или большее отклонение

Задача

$X_1, ..., X_n \sim \mathcal{N}(0, \sigma^2)$

$\text{K} = \left\{ \frac{1}{c} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \middle| c \in \mathbb{R}, c \neq 0 \right\}$

Нужно найти наилучшую оценку в с/к подходе:

1) $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1} \Rightarrow \frac{\mathbb{D}_\sigma n S^2}{\sigma^2} = 2(n-1)$

2) $\mathbb{E}_\sigma \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = (n-1)\sigma^2$

3) $\mathbb{D}_\sigma \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = (n-1)\sigma^4$

$\text{bias}^2 = \left[\mathbb{E}_\sigma \left(\frac{1}{c} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 - \sigma^2\right)\right]^2 = \left(\frac{n-1}{c} - 1\right)^2 \sigma^4$

$\text{variance} = \mathbb{D}_\sigma \frac{1}{c} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \frac{2(n-1)}{c^2}\sigma^4$

$\text{MSE}_{\hat{\sigma}}(\sigma) = \sigma^4 f(c), f(c) = \left(\frac{n-1}{c} - 1\right)^2 + \frac{2(n-1)}{c^2}$

$f'(c) = 2\left(\frac{n-1}{c} - 1\right) \cdot \left(-\frac{n-1}{c^2}\right) - \frac{4(n-1)}{c^3}$

Тогда $c_{\min} = n + 1$

Итого, $c = n + 1$ – лучшая оценка в с/к подходе, $c = n$ – ОМП, $c = n - 1$ – несмещенная

I - 10.2 Байесовский подход

Опр

Пусть Q распределение на Θ

$\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\mathbb{E}_Q R_{\hat{\theta}_1, \tau} \leq \mathbb{E}_Q R_{\hat{\theta}_2, \tau}$.

Если неравенство строгое, то оценка $\hat{\theta}_1$ лучше

Если Q – равномерное распределение, то просто сравниваем площадь под графиками риска

I - 10.3 Минимаксный подход

Опр

$\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\sup\limits_{\theta \in \Theta} R_{\hat{\theta}_1, \tau} \leq \sup\limits_{\theta \in \Theta} R_{\hat{\theta}_2, \tau}$.

Если неравенство строгое, то оценка $\hat{\theta}_1$ лучше

I - 10.4 Оценки в схеме Бернулли

$X_1, ..., X_n \sim \mathrm{Bern}(\theta)$

Будем сравнивать разные оценки

1) $\mathrm{K} = \{\text{все несмещенные оценки}\}$

   $\overline{X}$ – лучшая в с/к подходе и во всех других подходах

2) $\mathrm{K} = \{\text{все оценки}\}$

   В с/к подходе наилучших нет

3) Минимаксный подход

   $\tilde{\theta} = \overline{X} + \frac{1}{1+\sqrt{n}}\left(\frac{1}{2} - \overline{X}\right)$ – оценка Ходжеса-Лемана

   $\tilde{\theta} = \arg\min_{\hat{\theta}} \max_{\theta \in \Theta} \mathbb{E}_\theta \left(\hat{\theta} - \theta\right)^2$

   Рассмотрим $n = 9$:

   Доля $\theta \in [0, 1]$, при которых $\tilde{\theta}$ лучше $\overline{X}$ равна ≈ 0.66, то есть, если θ выбирается равномерно из $[0, 1]$, то в 66% случаев $\tilde{\theta}$ будет лучше

   Рассмотрим функцию $p(\theta) = P_\theta\left(\frac{\theta - d_n}{1 - 2d_n} \leq \overline{X} \leq \frac{1}{2}\right), d_n = \frac{1}{4\left(\sqrt{1+\sqrt{n}}\right)}, \theta \leq \frac{1}{2}$

   Во-первых, функция симметрична относительно $\theta = \frac{1}{2}$, во-вторых,

   $p(\theta) = P_\theta\left(\left|\tilde{\theta} - \theta\right| < \left|\overline{X} - \theta\right|\right)$

   То есть по ней мы можем понять, когда оценка Ходжеса Лемана скорее всего лучше, чем оценка средним

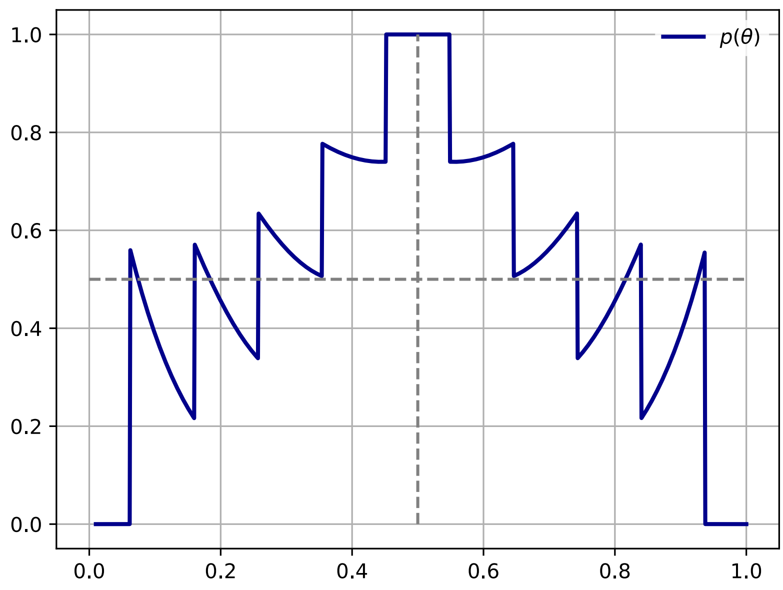


График 10: График $p(\theta)$

Если график ниже $\frac{1}{2}$, то настоящее значение ближе к $\overline{X}$, если выше – наоборот.

Замечание
$X_1, ..., X_n \sim \mathcal{N}(\theta, \sigma^2), \sigma$ – известна

Тогда $\overline{X}$ – наилучшая в с/к подходе и минимаксном подходе(при MSE). Но тем не менее $\exists \tilde{\theta}\ \forall \theta \in \mathbb{R}\ P_\theta\left(\left|\tilde{\theta} - \theta\right| < \left|\overline{X} - \theta\right|\right) > \frac{1}{2}$

$\tilde{\theta} = \overline{X} - \frac{\sigma}{2\sqrt{n}} \cdot \operatorname{sign}\left(\overline{X}\right) \min\left(\sqrt{n}\frac{|\overline{X}|}{\sigma}, \Phi_{\mathcal{N}(0,1)}\left(-\sqrt{n}\frac{|\overline{X}|}{\sigma}\right)\right)$

I - 10.5 Асимптотический подход

Опр
$\hat{\theta}_1, \hat{\theta}_2$ - а.н.о. θ с а.д. $\sigma_1^2(\theta), \sigma_2^2(\theta)$

1) $\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если $\forall \theta \in \Theta : \sigma_1(\theta) \leq \sigma_2(\theta)$.

2) Если существует θ, в котором неравенство строгое, то оценка $\hat{\theta}_1$ лучше $\hat{\theta}_2$

3) $\mathrm{ARE}_{\hat{\theta}_1, \hat{\theta}_2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$ – относительная асимптотическая эффективность

Замечание
Если $\mathrm{ARE}_{\hat{\theta}_1, \hat{\theta}_2} > 1$, то $\hat{\theta}_1$ лучше

Пример:

$X_1, ..., X_n \sim \mathcal{N}(\theta, \sigma^2), \sigma$ – известна

Сравним среднее и медиану в асимптотическом подходе

$\overline{X}$ – а.н.о. с а.д. $\sigma_1^2 = \mathbb{D}_\theta X_1 = \sigma^2$

По теореме о выборочном квантиле $\hat{\mu}$ – а.н.о. θ с а.д. $\sigma_2^2 = \frac{\pi\sigma^2}{2}$

$\text{ARE}_{\overline{X}, \hat{\mu}} = \frac{\pi}{2} > 1 \Rightarrow \overline{X}$ лучше $\hat{\mu}$

Пример:

Сравним еще среднее и медиану для распределения Лапласа

$X_1, ..., X_n \sim \text{Laplace}(\theta)$

$p_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$

$\overline{X}$ - а.н.о. θ с а.д. $\sigma_1^2 = 2$

$\hat{\mu}$ – а.н.о θ с а.д. $\sigma_2^2 = 1$

В данном случае медиана оказалось лучше

I - 11 Робастные оценки

I - 11.1 Приближенный поиска ОМП. Метод Ньютона

Пусть дана функция $f(x) : \mathbb{R} \to \mathbb{R}$ и надо решить уравнение $f(x) = 0$

Идея: приближение касательной

$y(x) = f(x_0) + f'(x_0)(x - x_0) = 0$

$x = x_0 - \frac{f(x_0)}{f'(x_0)}$

Исходя из этого запишем итерацию метода Ньютона

$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$

Применим теперь метод Ньютона для поиска ОМП:

Нам нужно получить $\ell_X(\theta) \to \max_\theta$

То есть найти экстремум, а это мы можем сделать, найдя ноль производной

$\frac{\partial \ell_X(\theta)}{\partial \theta} = 0 : \theta^{(k+1)} = \theta^{(k)} - \left(\nabla^2 \ell_X\left(\theta^{(k)}\right)\right)^{-1} \cdot \left(\nabla \ell_X\left(\theta^{(k)}\right)\right)$

> **Замечание**
> Пока неизвестно, какую брать начальную точку и насколько быстрая сходимость, то есть сколько итераций надо сделать

Теорема

В условиях L1-L9 если $\widehat{\theta}^{(0)}$ - а.н.о. θ, то

1) $\widehat{\theta}^{(1)}$ - а.н.о. θ с а.д. $i^{-1}(\theta)$
2) $\widehat{\theta}^{(1)}$ – асимптотически эквивалента ОМП, то есть

   $\sqrt{n}\left(\widehat{\theta}^{(1)} - \theta^*\right) \xrightarrow{P_\theta} 0, n \to \infty$, где θ^* – ОМП

▲

Без доказательства примем, что $\widehat{\theta}^{(1)} - \theta^* = \left(\widehat{\theta}^{(0)} - \theta^*\right)\varepsilon_n(\theta), \varepsilon_n(\theta) \xrightarrow{P_\theta} 0$

2) $\sqrt{n}\left(\widehat{\theta}^{(1)} - \theta^*\right) = \sqrt{n}\left(\widehat{\theta}^{(0)} - \theta\right)\varepsilon_n(\theta) + \sqrt{n}(\theta - \theta^*)\varepsilon_n(\theta) \to 0$

   По лемме Слуцкого и асимптотической нормальности каждое слагаемое сходится к нулю по распределению, а значит сходится к нулю и по вероятности

1) $$\sqrt{n}\left(\widehat{\theta}^{(1)} - \theta\right) = \overbrace{\sqrt{n}\left(\widehat{\theta}^{(1)} - \theta\right)}^{\xrightarrow{P_\theta} 0} + \overbrace{\sqrt{n}(\theta^* - \theta)}^{\xrightarrow{d_\theta} \mathcal{N}(0, i^{-1}(\theta))} \xrightarrow{d_\theta} \mathcal{N}\left(0, i^{-1}(\theta)\right)$$

■

Утв

Утверждение теоремы не изменится, если заменить $\nabla^2 \ell_X(\theta)$ на $\mathbb{E}_\theta \nabla^2 \ell_X(\theta) = -ni(\theta)$

$\widehat{\theta}^{(1)} = \widehat{\theta}^{(0)} + i^{-1}(\theta)\nabla\ell_X(\theta) \cdot \frac{1}{n}$

Пример:

Применим нашу теорему к распределению Коши и выборочной медиане

$\widehat{\theta}^{(0)} = \widehat{\mu}$ – выборочная медиана

$\widehat{\mu}$ – а.н.о. θ с а.д. $\sigma^2 = \frac{\pi^2}{4} \approx 2.47$, а у ОМП дисперсия была бы равна 2

Теперь выпишем одношаговую оценку

$$\widehat{\theta}^{(1)} = \widehat{\mu} + \frac{\sum_{i=1}^{n} \frac{X_i - \widehat{\mu}}{1 + (X_i - \widehat{\mu})^2}}{\sum_{i=1}^{n} \frac{1 - (X_i - \widehat{\mu}^2)}{\left(1 + (X_i - \widehat{\mu})^2\right)^2}}$$

Замечание

1) Если не выполнены L1-L9, то нужно сделать несколько итераций метода Ньютона

2) Если $\ell_X(\theta)$ не выпуклая, то можно запустить метод несколько раз, а результаты усреднить

I - 11.2 Робастность и симметричность

Пусть нам дана выборка $X = (X_1, ..., X_n) \sim \mathcal{N}(\theta, 1)$

$\hat{\theta} = \overline{X}$ – как мы знаем это хорошая оценка, асимптотически нормальная, ОМП и т.д. Но что если в данных будут выбросы – оценка испортится, хорошие свойства потеряются

Замечание

В реальных данных где-то 10% выбросов

Поэтому нужны методы детекции выбросов и оценки, которые умеют работать с этими выбросами.

Начнем с детекции выбросов. Рассмотрим график boxplot или "ящик с усами"

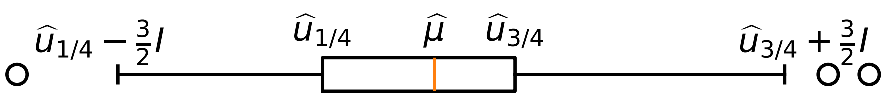


График 11: Иллюстрация boxplot

$\hat{\mu}$ – выборочная медиана, $\hat{u}_{\frac{1}{4}}, \hat{u}_{\frac{3}{4}}$ – соответствующие выборочные квантили, $I = \hat{u}_{\frac{3}{4}} - \hat{u}_{\frac{1}{4}}$

Выбросами в таком случае принято считать точки, которые не попадают в отрезок $\left[\hat{u}_{\frac{1}{4}} - \frac{3}{2}I, \hat{u}_{\frac{3}{4}} + \frac{3}{2}I\right]$

Теперь перейдем к робастным оценкам

I - 11.3 Робастные оценки

Пусть исследуемое семейство распределений – это $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. Причем P_0 имеет плотность $p_0(x)$, симметричную относительно 0, непрерывную, с носителем $(-c, c), c > 0$ и удовлетворяющую равенству $p_\theta(x) = p_0(x - \theta)$, то есть θ – это сдвиг

Мы хотим получить оценку, которая будет хорошо приближать θ, но также будет устойчивой к выбросам

1. Усеченное среднее

Опр

$\alpha \in \left(0, \frac{1}{2}\right), k = \lceil \alpha n \rceil$

$\overline{X}_\alpha = \frac{1}{n-2k} \sum\limits_{i=k+1}^{n-k} X_{(i)}$

Опр

Пусть $\hat{\theta} = f\left(X_{(1)}, ..., X_{(n)}\right)$. Определим K_n^* как наименьшее k, такое что:

1) Если $X_1, ..., X_{k+1} \to -\infty$, а $X_{k+2}, ..., X_n$ – фиксированные,то

   $f(X_1, ..., X_n) \to -\infty$

2) Если $X_{n-k}, ..., X_n \to -\infty$, а $X_1, ..., X_{n-k-1}$ – фиксированные,то

   $f(X_1, ..., X_n) \to -\infty$

Тогда величина $\tau_{\hat{\theta}} = \lim\limits_{n\to\infty} \frac{K_n^*}{n}$ называется асимптотической толерантностью оценки $\hat{\theta}$

Замечание

$\tau_{\hat{\theta}}$ – это наибольшая доля выбросов, которую выдержит оценка

Теорема (об усеченном среднем)

$\alpha \in \left(0, \frac{1}{2}\right), k = \lceil \alpha n \rceil$

$X_1, ..., X_n \sim P \in \mathcal{P}$ – для распределения выполнены условия выше

Тогда $\sqrt{n}\left(\overline{X}_\alpha - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma_\alpha^2)$

$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left[\int\limits_0^{u_{1-\alpha}} x^2 p_0(x) dx + \alpha u_{1-\alpha}^2\right]$

$u_{1-\alpha}$ – это $1-\alpha$ квантиль распределения P_0

Пример:

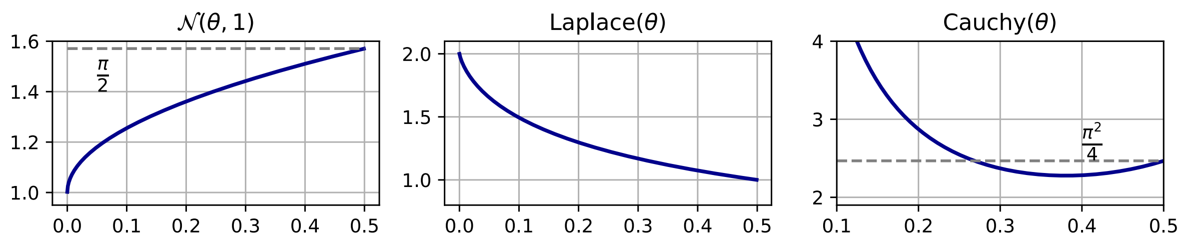


График 12: Графики дисперсий усеченного среднего

Также в качестве примера приведем сравнения среднего и усеченного среднего для $\mathcal{N}(\theta, 1)$

α	0	$\frac{1}{20}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{2}$
$\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}}$	1	0.99	0.94	0.84	0.74	0.64

Замечание

$\alpha = \frac{1}{8}$ защищает от 12.5% выбросов

Теорема

Если $\mathbb{D}_\theta X_1 < \infty$, то $\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}} \geq (1 - 2\alpha)^2$

▲

$\overline{X}_\alpha$ – а.н.о. θ с а.д. σ_α^2

$\overline{X}$ – а.н.о. θ с а.д. $\sigma^2 = \mathbb{D}_\theta X_1$

$$\frac{1}{2}\mathbb{D}_\theta X_1 = \frac{1}{2}\int\limits_{\mathbb{R}} x^2 p_0(x) dx = \int\limits_0^{+\infty} x^2 p_0(x) dx =$$

$$= \int\limits_0^{u_{1-\alpha}} x^2 p_0(x) dx + \int\limits_{u_{1-\alpha}}^{+\infty} x^2 p_0(x) dx \geq \int\limits_0^{u_{1-\alpha}} x^2 p_0(x) dx + \alpha u_{1-\alpha}^2 = \frac{\sigma_\alpha^2 (1-2\alpha)^2}{2}$$

Тогда $\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}} \geq (1 - 2\alpha)^2$

■

2. Медиана средних Уолша

Опр

$Y_{ij} = \frac{X_i + X_j}{2}$ – среднее Уолша

$W = \mathrm{median}\left(\left\{Y_{ij}, 1 \leq i \leq j \leq n\right\}\right)$

Пусть $X_1, ..., X_n$ – выборка из $P_\theta \in \mathcal{P}$(в условиях выше). Тогда $\sqrt{n}(W - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2)$

Где $\sigma^2 = \frac{1}{12}\left[\frac{1}{\int\limits_{\mathbb{R}} p_0^2(x) dx}\right]^2$

Утв

$\tau_W \approx 0.293$

То есть медиана Уолша может выдержать до 29.3% выбросов

$\mathrm{ARE}_{W, \overline{X}} \approx 0.955$ для $\mathcal{N}(\theta, 1)$

Замечание

Для общего случая есть теорема, которая доказывает, что $\text{ARE}_{W,\overline{X}} \geq \frac{108}{125} = 0.864$

То есть в худшем случае теряем 14% эффективности, причем плотность для этого худшего случая выглядит вот так: $p_0(x) = \frac{3\sqrt{5}}{100}(5 - x^2)\mathbb{I}\left\{|x| \leq \sqrt{5}\right\}$

I - 11.4 Эмпирическое распределение

Пусть $X = (X_1, ..., X_n)$ – выборка из распределения P, причем $\mathcal{P}$ – это все распределения на $\mathcal{X}$, то есть семейство распределений мы никак не ограничиваем

Опр

Эмпирическим распределением по X называется вероятностная мера, определяемая по правилу $\forall B \in \mathfrak{B}(\mathcal{X}) : \hat{P}_n(B) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}\{X_i \in B\}$

Свойства:

1) $\hat{P}_n(B)$ – случайная величина(доля элементов, попавших в B)
2) $\hat{P}_n(B)$ – дискретная вероятностная мера
3) $n \cdot \hat{P}_n(B) \sim \text{Binom}(n, P(B))$
4) Из УЗБЧ следует, что $\hat{P}_n(B) \overset{\text{п.н.}}{\to} P(B)$

Теперь перейдем в $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ и будем рассматривать случайные величины только там. Тогда справедливы следующие результаты

$\hat{P}_n(B)$ имеет эмпирическую функцию распределения

$\hat{F}_n(x) = \hat{P}_n((-\infty, x]) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}\{X_i \leq x\}$

И эта функция распределения тоже почти наверное сходится к настоящей, свойства этой сходимость устанавливают следующие теоремы

Теорема (Гливенко-Кантелли)

$D_n = \sup\limits_{x \in \mathbb{R}}\left|\hat{F}_n(x) - F(x)\right| \overset{\text{п.н.}}{\to} 0$

Заметим, что $D_n = \sup\limits_{B \in \mathcal{A}}\left|\hat{P}_n(B) - P(B)\right| \overset{\text{п.н.}}{\to} 0, \mathcal{A} = \{(-\infty, x]\}$

Теорема (Вапника-Червоненкиса)
$D_n = \sup\limits_{B \in \mathcal{A}} \left|\hat{P}_n(B) - P(B)\right| \overset{\text{п.н.}}{\to} 0 \Leftrightarrow$ конечна VC размерность при разбиении $\mathbb{R}^d$ множествами из $\mathcal{A}$

Причем, как видно, $\mathcal{A}$ набор множеств из $\mathbb{R}^d$, а не обязательно из $\mathbb{R}$

Замечание
VC-размерность не часть курса, но на всякий случай приведено определение из дискретного анализа

(X, R) – ранжированное пространство, $R \subseteq 2^X$.

$\mathrm{VC}(X, R) = \max\{m \mid \exists S \subset X : |S| = m, S \text{ дробится } m\}$

$S \subset X$ дробится R, если $\{S \cap r \mid r \in R\} = 2^S$

Теорема (Колмогорова-Смирнова)

$\sqrt{n} D_n \overset{d}{\to} \xi$ – распределение Колмогорова

$F_\xi(x) = \sum\limits_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2x^2)\mathbb{I}\{x \geq 0\}$

I - 11.5 Метод подстановки

Пусть $X = (X_1, ..., X_n)$ – выборка из произвольного распределения P с функцией распределения F

$\theta = G(P)$ – это то, что нам нужно оценить

$\theta = \int\limits_{\mathbb{R}} x \, dF(x)$

Чтобы получить оценку методом подстановки нужно просто исходное распределение заменить на эмпирическое $\hat{\theta} = G\left(\hat{P}_n\right)$

Пример:

1) $\theta = G(P) = \int\limits_{\mathcal{X}} f(x)\, dF(x) \Rightarrow \hat{\theta} = \int\limits_{\mathcal{X}} f(x)\, d\hat{F}_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} f(X_i) = \overline{f(X)}$

2) $\theta = G(P) = \mathbb{D}_\theta X_1 = \int\limits_{\mathcal{X}} x^2\, dF(x) - \left(\int\limits_{\mathcal{X}} x\, dF(x)\right)^2$

$\hat{\theta} = \overline{X^2} - \left(\overline{X}\right)^2 = S^2$

3) $\theta = \min\{x \mid F(x) \geq \alpha\}$

$\hat{\theta} = \min\left\{x \,\middle|\, \hat{F}_n(x) \geq \alpha\right\} = X_{(\lceil n\alpha \rceil)}$

I - 12 Достаточные статистики и робастные оценки. Семинар

I - 12.1 Достаточные статистики

Опр
$X_1, ..., X_n \sim P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. Тогда статистика $S(X)$ называется достаточной, если $\forall \theta \in \Theta \ \forall B \in \mathfrak{B}(\mathbb{R}^d) \ P_\theta(X \in B \mid S(X))$ не зависит от θ

Пример:

$X_1, ..., X_n \sim \mathrm{Bern}(\theta) \Rightarrow S(X) = \overline{X}$ или $S(X) = \sum_{i=1}^{n} X_i$ – достаточные статистики

Теорема (критерий факторизации Неймана-Фишера)
Пусть $X = (X_1, ..., X_n)$ – выборка из распределения $P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, $\mathcal{P}$ – доминируемое распределение. Тогда

$S(X)$ – достаточная статистика $\Leftrightarrow \mathcal{L}_X(\theta) = \Psi(S(X), \theta) h(X)$

h, Ψ – измеримые функции

Замечание
Совместную плотность можно записать через правдоподобие, как и было сделано, идейно от этого ничего не поменяется

Пример:

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2), \theta = (a, \sigma^2)$

Нужно найти достаточную статистику

$$\mathcal{L}_X(\theta) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - a)^2\right) =$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n} X_i^2}{\sigma^2} + \frac{a}{\sigma^2} \sum_{i=1}^{n} X_i - \frac{na^2}{2\sigma^2}\right]$$

Тогда по критерию факторизации Неймана-Фишера

$S(X) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$ – достаточная статистика

$h(x) = 1, \Psi(S(x), \theta) = \left(2\pi\theta_2\right)^{-\frac{n}{2}} \exp\left(\frac{1}{2\theta_2} S_1 + \frac{\theta_1}{\theta_2^2} S_2 + \frac{n\theta_1^2}{2\theta_2}\right)$

Если данные к нам приходят последовательно и в реальном времени, то легко записать формулу пересчета для порядковой статистики

$S_{n+1}(X) = \left(S_{n,1}(X_1, ..., X_n) + X_{n+1}^2, S_{n,2}(X_1, ..., X_n) + X_{n+1}\right)$

Пример:

$X_1, ..., X_n \sim \mathrm{Exp}^*(\theta)$ (экспоненциальное распределение со сдвигом θ)

Снова нужно найти достаточную статистику

$p_\theta(x) = e^{-(x-\theta)} I\{x \geq \theta\}$

$\mathcal{L}_X(\theta) = \mathbb{I}\left\{X_{(1)} \geq \theta\right\} \exp\left(-\sum_{i=1}^{n} X_i + n\theta\right) =$

$= \underbrace{\mathbb{I}\left\{X_{(1)} \geq \theta\right\} e^{n\theta}}_{\Psi(S(X),\theta)} \cdot \underbrace{\exp\left(-\sum_{i=1}^{n} X_i\right)}_{h(X)}$

$S(X) = X_{(1)}$ – достаточная статистика

Формула пересчета:

$S_{n+1}(X) = \min(S_n(X_1, ..., X_n), X_{n+1})$

I - 12.2 Робастные оценки

Опр
Пусть $\hat{\theta} = f\left(X_{(1)}, ..., X_{(n)}\right)$. Определим K_n^* как наименьшее k, такое что:

1) Если $X_1, ..., X_{k+1} \to -\infty$, а $X_{k+2}, ..., X_n$ – фиксированные,то

   $f(X_1, ..., X_n) \to -\infty$

2) Если $X_{n-k}, ..., X_n \to -\infty$, а $X_1, ..., X_{n-k-1}$ – фиксированные,то

   $f(X_1, ..., X_n) \to -\infty$

Тогда величина $\tau_{\hat{\theta}} = \lim\limits_{n\to\infty} \frac{K_n^*}{n}$ называется асимптотической толерантностью оценки $\hat{\theta}$

Пример:

$\hat{\theta} = \sum_{i=\lceil \ln(n) \rceil}^{\lfloor n - \ln(n) \rfloor} X_{(i)} \Rightarrow K_n^* = 2\ln(n)$. Тогда $\tau_{\hat{\theta}} = \lim\limits_{n\to\infty} \frac{K_n^*}{n} = 0$

Опр
Усеченным средним степени $\alpha \in \left(0, \frac{1}{2}\right)$ называется оценка

$\overline{X}_\alpha = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X_{(i)}, k = \lceil n\alpha \rceil$

Опр

Медиана средних Уолша – это оценка

$W = \text{median}\left(\left\{\frac{X_i+X_j}{2}, 1 \leq i \leq j \leq n\right\}\right)$

Замечание

Как уже отмечалось, $\tau_W \approx 0.293$

Теорема

Пусть P_0 – симметричное распределение с непрерывной плотностью $p_0(x)$, а P_θ – его сдвиг на θ. Тогда

1) $\overline{X}_\alpha$ – а.н.о. θ с а.д. $\frac{2}{(1-2\alpha)^2}\left[\int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \alpha u_{1-\alpha}^2\right]$

2) W - а.н.о. θ с а.д. $\frac{1}{12}\left[\left(\int\limits_{\mathbb{R}} p_0^2(x)dx\right)\right]^{-2}$

Пример:

$X_1, ..., X_n \sim U[\theta - 1, \theta + 1]$

Нужно посчитать а.д. для $\overline{X}_\alpha$ и W

$P_0 = U[-1, 1], p_0(x) = \frac{1}{2}I\{-1 \leq x \leq 1\}$

$P_\theta = U[\theta - 1, \theta + 1], p_\theta(x) = \frac{1}{2}\mathbb{I}\{\theta - 1 \leq x \leq \theta + 1\}$

То есть условия теоремы выше выполняются

$u_{1-\alpha} = 1 - 2\alpha$

$\sigma_{\overline{X}_\alpha}^2 = \frac{2}{(1-2\alpha)^2}\left[\int\limits_0^{1-2\alpha} \frac{x^2}{2}dx + \alpha(1-2\alpha)^2\right] = \frac{4\alpha+1}{3}$

$\sigma_W^2 = \frac{1}{12}\left[\left(\int\limits_{\mathbb{R}} \left(\frac{1}{2}\mathbb{I}\{x \in [-1, 1]\}\right)^2 dx\right)\right]^{-2} = \frac{1}{12}\left(\int\limits_{[-1,1]} \frac{1}{4}dx\right)^{-2} = \frac{1}{3}$

Также, как мы знаем, $\sigma_{\hat{\mu}}^2 = 1$

Можно сделать вывод, что часто мы жертвуем дисперсией оценки, чтобы получить устойчивость к выбросам

I - 12.3 Приближенный поиск ОМП

Как уже обсуждалось на лекции, будем искать ноль производной логарифма правдоподобия по методу Ньютона

$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \left(\nabla^2 \ell_X\left(\hat{\theta}^{(k)}\right)\right)^{-1} \cdot \left(\nabla \ell_X\left(\hat{\theta}^{(k)}\right)\right)$

Пример:

$X_1, ..., X_n \sim \mathrm{Exp}(\theta)$

$p_\theta(x) = \theta^{-\theta x}\mathbb{I}\{x \geq 0\}$

$\mathcal{L}_X(\theta) = \theta^n \mathbb{I}\left\{X_{(1)} \geq 0\right\} \exp\left(-\theta \sum\limits_{i=1}^{n} X_i\right)$

$\ell_X(\theta) = n\ln(n) - \theta \sum\limits_{i=1}^{n} X_i$

$\nabla \ell_X(\theta) = \frac{n}{\theta} - \sum\limits_{i=1}^{n} X_i$

$\nabla^2 \ell_X(\theta) = -\frac{n}{\theta^2}$

Теперь у нас есть все, чтобы записать итерацию метода Ньютона

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} - \frac{\frac{n}{\theta} - \sum\limits_{i=1}^{n} X_i}{-\frac{n}{\theta^2}} = \widehat{\theta}^{(k)} + \widehat{\theta}^{(k)}\left(1 - \widehat{\theta}^{(k)}\overline{X}\right)$$

И на вопросы о начальной точке и скорости сходимости отвечает теорема

Теорема

В условиях L1-L9 если $\widehat{\theta}^{(0)}$ - а.н.о. θ, то $\widehat{\theta}^{(1)}$ - а.н.о. θ с а.д. $i^{-1}(\theta)$(то есть с минимальной дисперсией)

I - 13 Бутстреп. Семинар

Лекция по бутстрепу была полностью по презентации, поэтому в конспект не вошла

I - 13.1 Эмпирическое распределение. Напоминание

Опр

$\widehat{P}_n(B) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}\{X_i \in B\}$ – эмпирическое распределение

$\widehat{F}_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}\{X_i \leq x\}$ – эмпирическая функция распределения

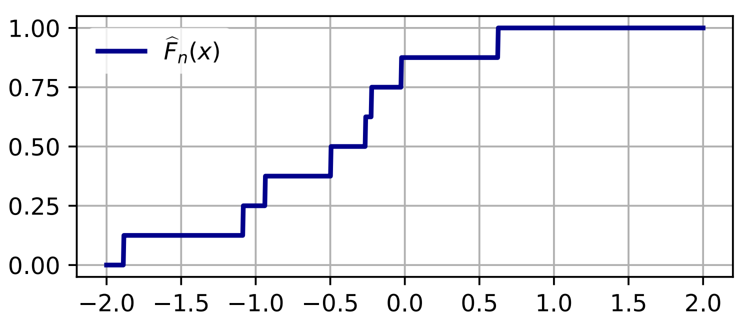


График 13: График эмпирической функции распределения

I - 13.2 Бутстреп

Дана выборка $X_1, ..., X_n \sim P$, нужно оценить $T(X)$, а также $G(T(X))$

__**Алгоритм:**__

1) Генерируем индексы $i_1, ..., i_n \sim U\{1, ..., n\}$. Теперь наша новая выборка – это $X^* = \left(X_{i_1}, ..., X_{i_n}\right)$
2) Считаем статистику $T\left(X_{i_1}, ..., X_{i_n}\right) = T(X^*)$
3) Повторяем шаги 1 и 2 B раз, получим B значений статистики
4) Считаем $G\left(T\left(X_j^*\right)\right)$ методом Монте-Карло(используем посчитанные на бутстрепных выборках значения)

X_j^* – j-ая бутстрепная статисика из B итераций

Какая точность у такого алгоритма? У метода Монте-Карло точность $O\left(\frac{1}{\sqrt{B}}\right)$, у метода подстановки, как мы знаем из лекции, $O\left(\frac{1}{\sqrt{n}}\right)$. Тогда нужно брать $B \geq n$, чтобы не уменьшить точность.

__**Пример:**__

Пусть дана выборка $3, 5, 7, 1, 9$. Нам нужно оценить дисперсию.

Мы сгенерируем как минимум 5 бутстрепных выборок. Пусть средние по этим выборкам равны: $4, 3.6, 3.8, 4.2, 5$.

Тогда дисперсию мы будем считать по формуле:

$$\frac{1}{B}\sum_{b=1}^{B}\left(T_b^*\right)^2 - \left(\frac{1}{B}\sum_{b=1}^{B} T_b^*\right)^2.$$

Где T_b^* – значение среднего по соответствующей бутстрепной выборке

> __**Замечание**__
> Таже можем строить доверительный интервалы при помощи бутстрепа, например, квантильный.
>
> $t = (T_1, ..., T_B)$ – бутстрепные статистики
>
> $\left(t_{\frac{1-\alpha}{2}}, t_{\frac{1+\alpha}{2}}\right)$ – квантильный доверительный интервал

Обсудим плюсы и минусы бутстрепа:

Плюсы:
- Прост в реализации
- Можно не использовать дельта-метод и другие сложные способы получения оценок

Минусы:

- Имеет более низкую точность, чем теоретические методы
- Может очень долго считаться. Например, при размере выборке 10^4 нужно сгенерировать 10^8 элементов.

II Проверка статистических гипотез

II - 1 Гипотезы и критерии

Пусть $\mathcal{X}$ - выборочное пространство, $\mathcal{P}$ - семейство распределений и $X = (X_1, ..., X_n)$

II - 1.1 Гипотезы в непараметрическом и параметрическом случаях

Опр
Рассмотрим высказывание такого вида: $H_0 : P \in \mathcal{P}_0$, где $\mathcal{P}_0 \subseteq \mathcal{P}$. Тогда H_0 - (основная) гипотеза

Опр
Пусть $\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}$ - параметрическое семейство распределений. Тогда $H_0 : \theta \in \Theta_0 \subseteq \Theta$ - гипотеза

Примеры:

1) $\mathcal{P}$ - семейство всех абсолютно непрерывных распределений

   $H_0 : P \in \{\mathcal{N}(a, \sigma^2) \,|\, a \in \mathbb{R}, \sigma \in \mathbb{R}_+\}$
2) $\mathcal{P} = \{\mathcal{N}(a, \sigma^2) \,|\, a \in \mathbb{R}, \sigma \in \mathbb{R}_+\}$ - семейство нормальных распределений
   $H_0 : a = 0(\theta = (a, \sigma^2) \in \{(0, \sigma^2) \,|\, \sigma \in \mathbb{R}_+\})$

II - 1.2 Проверка гипотез

Опр
Критерий для проверки гипотезы H_0 - такое множество $S \subseteq \mathcal{X}$ (подмножество реализаций выборки), для которого выполнено, что H_0 отвергается $\Leftrightarrow X \in S$

Замечание
Обычно критерий S можно представить в виде $S = \{x \in \mathcal{X} \,|\, T(x) > c\}$. c - какая-то константа, критическое значение критерия; $T(x)$ - статистика критерия

Результаты проверки гипотез

1) $X \in S \Rightarrow H_0$ отвергается и результат называется статистически значимым
2) $X \notin S \Rightarrow H_0$ не отвергается и результат не статистически значим

Смысл:

Мы можем утверждать с какой-то ошибкой, что H_0 не верна, но не можем утверждать, что H_0 верна. То есть можем отвечать только “нет” и “не знаю”.

Важно, что не можем говорить, что H_0 принимается.

Пример: Как проверить нормальность? Через моменты!

$\mathbb{E}(X_1 - \mu_1)^3 = 0, \mathbb{E}(X_1 - \mu_1)^4 = 3\sigma^4$

Увидим, что какой-то выборочный центральный момент не близок к теоретическом моменту нормального распределения $\Rightarrow$ можем с большой вероятностью сказать, что распределение не нормальное, но доказать нормальность не можем

Вывод: семейство распределений бесконечно, а выборка конечна. Можем увидеть различия в типичности выборки для данного распределения, но не можем утверждать, что выборка точно из этого семейства.

Аналогия с презумпцией невиновности: обвиняемый считается невиновным, пока его вина не доказана

Обвиняемый $\longleftrightarrow P$ - распределение

Невиновность $\longleftrightarrow H_0 : P \in \mathcal{P}_0$

Виновность $\longleftrightarrow P \notin \mathcal{P}_0$

Факты/улики $\longleftrightarrow T(X)$

Вердикт $\longleftrightarrow \begin{cases} H_0 \text{ отвергается } (X \in S) \\ H_0 \text{ не отвергается } (X \notin S) \end{cases}$

Верность Отвергаем H_0	H_0 верна	H_0 не верна
H_0 не отвергается	😐	Ошибка II рода
H_0 отвергается	Ошибка I рода	😆

Опаснее ошибка первого рода

Опр
Вероятность ошибки I рода: $P(I_S) = \sup\limits_{P \in \mathcal{P}_0} P(X \in S)$

Вероятность ошибки II рода: $P(\mathrm{II}_S) = P(X \notin S)$ для $P \notin \mathcal{P}_0$

Замечание

1) Можно встретить определение вероятности ошибки II рода по аналогии.
2) На практике в вероятности ошибки I рода sup часто опускают.

Задача на оптимизации на ошибки I, II рода

$$\begin{cases} P(I_S) \leq \alpha \text{ - ограничиваем опасную ошибку} \\ P(\mathrm{II}_S) \to \min_{P \notin \mathcal{P}_0} \text{ - минимизируем по остаточному признаку} \end{cases}$$

Опр

Число $\alpha \in (0, 1)$ - уровень значимости критерия S, если $P(I_S) \leq \alpha$

Замечание

На практике обычно $\alpha = 0.05$, то есть допускаем $\leq 5\%$ ошибок I рода

Вывод:

По уровню значимости отбираем допустимые критерии $(S : P(I_S) \leq \alpha)$ по min ошибке II рода берем наилучший

Опр

Мощность критерия $\mathfrak{P}_S(p) = 1 - P(\mathrm{II}_S) = P(X \in S)$ для $P \notin \mathcal{P}_0$

II - 1.3 Альтернативная гипотеза

Рассмотрим другую постановку задачи

$H_0 : P \in \mathcal{P}_0$ vs $H_1 : P \in \mathcal{P}_1$, где $\mathcal{P}_0, \mathcal{P}_1 \subseteq \mathcal{P}, \mathcal{P}_0 \bigcap \mathcal{P}_1 = \emptyset$

Опр

H_1 - альтернативная гипотеза

Рассмотрим задачу

$$\begin{cases} P(I_S) \leq \alpha \\ P(\mathrm{II}_S) \to \min_{P \in \mathcal{P}_1} \end{cases}$$

Замечание

Альтернативная H_1 помогает выбрать лучший критерий отвержения H_0 в пользу H_1 среди допустимых. Мы не можем отвергать или принимать H_1

Если $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ и $H_0 : \theta = \theta_0$, то рассматриваем следующие альтернативы:

1) $H_1 : \theta \neq \theta_0$ - двусторонняя

2) $H_1 : \theta > \theta_0$ - правосторонняя гипотеза
3) $H_1 : \theta < \theta_0$ - левосторонняя гипотеза

Задача 1

$X \sim \mathrm{Exp}(\theta), \theta$ - интенсивность

$H_0 : \theta = \theta_0$ - изменений нет vs $H_1 : \theta > \theta_0$

$\{x > c\}$ $\{x \in (c_1, c_2)\}$

$\{x < c\}$ $\{x \leq c_1\} \bigcup \{x \geq c_2\}$

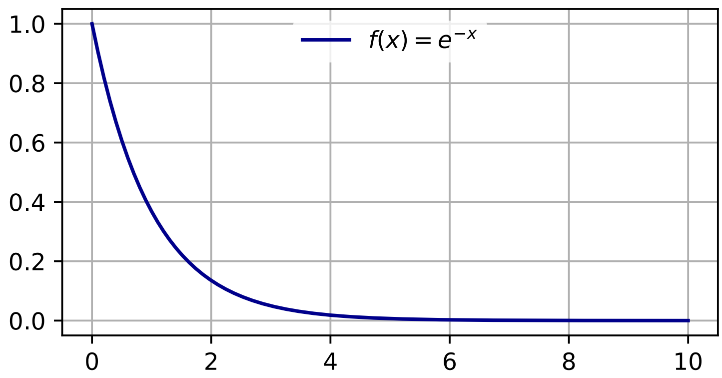


График 14: График плотности $\mathrm{Exp}(\theta)$

<u>**Соображение:**</u>

Интенсивность больше $\Rightarrow$ величина уменьшилась $\Rightarrow$ выгоднее взять $\{x \leq c_\alpha\}$

$P(I_S) = \sup\limits_{\theta = \theta_0} P(X \in S) = P_{\theta_0}(x < c_\alpha) = F_{\theta_0}(c_\alpha) = 1 - e^{-c_\alpha \theta_0} \leq \alpha$

$e^{-c_\alpha \theta_0} \geq 1 - \alpha \Rightarrow -c_\alpha \theta_0 \geq \ln(1 - \alpha) \Rightarrow c_\alpha \leq -\frac{1}{\theta_0} \ln(1 - c_\alpha)$

Выбираем c_α по max мощности:

$\mathfrak{P}_S(\theta) = P_\theta(x < c_\alpha) = 1 - e^{-c_\alpha \theta}$ - возрастает по c_α

Тогда $c_\alpha = -\frac{1}{\theta_0} \ln(1 - \alpha)$

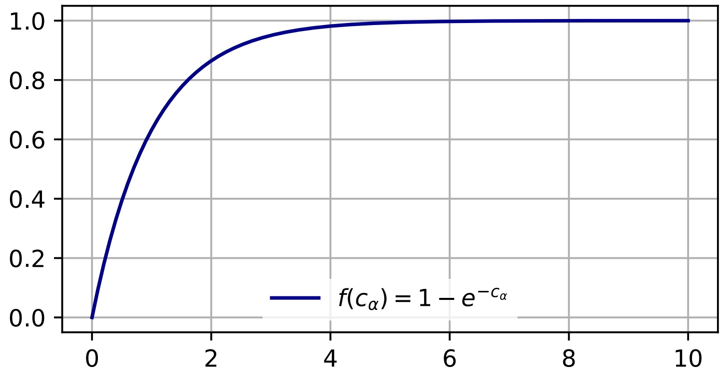


График 15: График мощности $\mathfrak{P}_S = 1 - e^{-c_\alpha \theta}$

Итого получаем критерий $S = \left\{ x \in \mathcal{X} \,\middle|\, x < -\frac{1}{\theta_0} \ln(1-\alpha) \right\}$

Его мощность: $\mathfrak{P}_S(\theta) = 1 - e^{-\frac{\theta}{\theta_0} \cdot \ln(1-\alpha)} = 1 - (1-\alpha)$

Пусть $\alpha = 0.05$, $\theta_0 = 1$ авт/час

$c \approx 0.051$ час $\Rightarrow c \approx 3$ мин

Замечание

c_α получился α-квантилем экспоненциального распределения

II - 1.4 Критерий Вальда

Опр

Критерий S - асимптотический критерий с уровнем значимости α, если $\lim\limits_{n\to\infty} \sup P(I_S) = \overline{\lim\limits_{n\to\infty}} P(I_S) \leq \alpha$

Замечание

При каком-то конечном размере выборки $P(I_S)$ может быть больше α, но в пределе сойдется к величине $\leq \alpha$

Пусть $\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}, \Theta \subseteq \mathbb{R}$. Пусть также $H_a : \theta = \theta_0$ - простая гипотеза. А $\hat{\theta}$ - асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$

Тогда $\theta = \theta_0$, $W(X) = \sqrt{n} \frac{\hat{\theta} - \theta_0}{\hat{\sigma}} \xrightarrow{d_{\theta_0}} \mathcal{N}(0,1)$

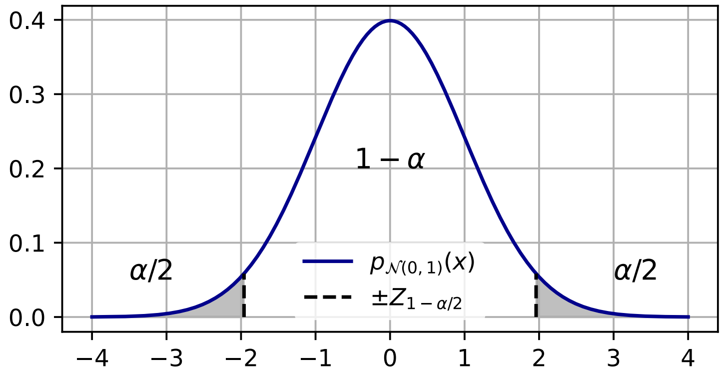


График 16: График квантилей нормального распределения

1) $H_1 : \theta \neq \theta_0$

   $S_1 = \left\{x \in \mathcal{X} \,\middle|\, |W(X)| \geq Z_{1-\frac{\alpha}{2}}\right\}$

2) $H_1' : \theta > \theta_0$

   $S_2 = \{x \in \mathcal{X} \mid W(X) \geq Z_{1-\alpha}\}$

3) $H_1'' : \theta < \theta_0$

   $S_3 = \{x \in \mathcal{X} \mid W(X) \leq Z_{\alpha}\}$

Покажем, что S_1 контролирует асимптотическую ошибку I рода:

$P_{\theta_0}\left(|W(X)| \geq Z_{1-\frac{\alpha}{2}}\right) = P_{\theta_0}\left(W(X) \leq Z_{\frac{\alpha}{2}}\right) + P_{\theta_0}\left(W(X) \geq Z_{1-\frac{\alpha}{2}}\right) \overset{n\to\infty}{\to}$
$\overset{n\to\infty}{\to} \frac{\alpha}{2} + \left(1 - \left(1 - \frac{\alpha}{2}\right)\right) = \alpha$

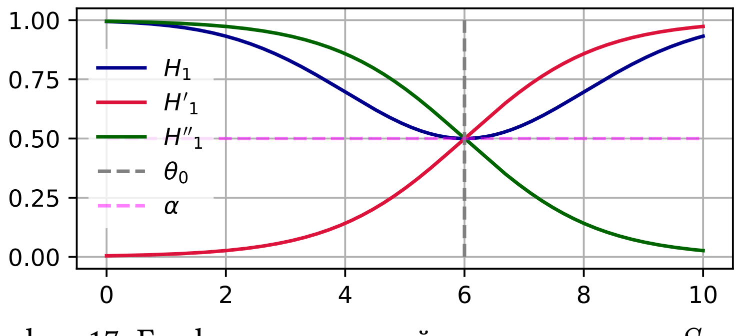


График 17: График мощностей для критериев S_1, S_2, S_3

Эквивалентный доверительный интервал

Возьмем $H_1 : \theta \neq \theta_0$

Для него $P(I_S) = P_{\theta_0}\left(|W(X)| \geq Z_{1-\frac{\alpha}{2}}\right) \overset{n\to\infty}{\to} \alpha$

$P_{\theta_0}\left(|W(X)| < Z_{1-\frac{\alpha}{2}}\right) \overset{n\to\infty}{\to} 1 - \alpha$

Тогда интервал Вальда $C = \left(\hat{\theta} - \frac{\hat{\sigma} Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \hat{\theta} + \frac{\hat{\sigma} Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)$ уровня доверия $1 - \alpha$

Но отвергаем $\Leftrightarrow \theta_0 \notin C \Leftrightarrow X \in S$

Замечание
В случаях (2) и (3) в H_0 можем поставить $\theta \leq \theta_0$ и $\theta \geq \theta_0$ соотвественно (так как супремум достигается в θ_0)

Задача 2

$X_1, ..., X_n \sim \text{Cauchy}(\theta)$, θ - сдвиг

$H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

$\hat{\mu}$ - асимптотически нормальная оценка с асимптотической дисперсией $\frac{\pi^2}{4}$

$W(X) = \sqrt{n} \frac{\hat{\mu} - \theta_0}{\frac{\pi}{2}}$ - статистика критерия

$S = \left\{X \in \mathcal{X} \,\middle|\, |W(X)| \geq Z_{1-\frac{\alpha}{2}}\right\}$

II - 2 Гипотезы и критерии. Семинар

Опр
$X_1, ..., X_n$ - выборка из распределения $P \in \mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}$. Есть множества $\Theta_1, \Theta_0 \subseteq \Theta$, такие что $\Theta_0 \bigcap \Theta_1 = \emptyset$ Тогда:
- $H_0 : \theta \in \Theta_0$ - основная гипотеза
- $H_1 : \theta \in \Theta_1$ - альтернативная гипотеза

Опр
Подмножество $S \subseteq \mathcal{X}$ - критерий для проверки H_0, если выполнено $X \in S \Leftrightarrow H_0$ отвергается

Опр
Вероятность ошибки I рода: $P(I_S) = \sup\limits_{P \in \mathcal{P}_0} P(X \in S)$
Вероятность ошибки II рода: $P(\text{II}_S) = P(X \notin S)$ для $P \notin \mathcal{P}_0$

Мы хотим минимизировать ошибки, но не можем сделать это одновременно

Решение:

$\begin{cases} P(I_S) \leq \alpha \\ P(\text{II}_S) \to \min_S \end{cases}$

α - уровень значимости / доверия, обычно $\alpha = 0.05$

Опр
$\mathfrak{P}_S(P) = P(X \in S) = 1 - P(\mathrm{II}_S), P \notin \mathcal{P}_0$

II - 2.1 Критерий Вальда

$X = (X_1, ..., X_n)$ - выборка из распределения $P \in \mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}$

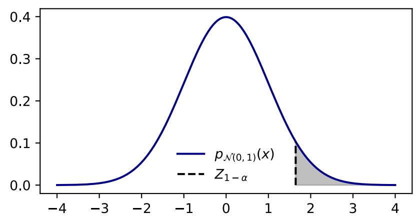
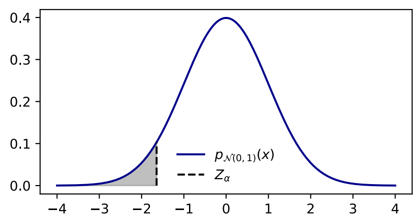
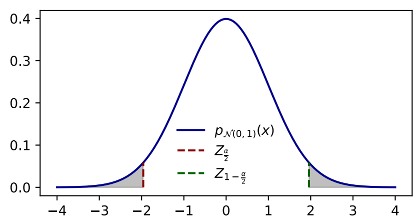
$\hat{\theta}$ - асимптотически нормальная оценка с асимптотической дисперсией σ^2

$\hat{\sigma}$ - состоятельная оценка σ

$W = \sqrt{n}\frac{\hat{\theta} - \theta_0}{\hat{\sigma}}$

$H_0 : \theta = \theta_0$
$H_1 : \theta > \theta_0$ - правосторонняя
$H_1 : \theta < \theta_0$ - левосторонняя
$H_1 : \theta \neq \theta_0$ - двусторонняя

$W \xrightarrow{d_{\theta_0}} \mathcal{N}(0, 1)$ (по лемме Слуцкого и ЦПТ)

$\theta > \theta_0$	$S_1 = \{W \geq Z_{1-\alpha}\}$	
$\theta < \theta_0$	$S_2 = \{W \leq Z_\alpha\}$	
$\theta_0 \neq \theta$	$S_3 = \left\{\|W\| \geq Z_{1-\frac{\alpha}{2}}\right\}$	

Задача

$X_1, ..., X_n \sim \mathrm{Pois}(\theta)$

Нужно построить асимптотический критерий Вальда уровня доверия α для $H_0 : \theta = \theta_0, H_1 : \theta > \theta_0$

$\hat{\theta} = \overline{X}$ - асимптотически нормальная оценка с асимптотической дисперсией θ

$\widehat{\sigma^2} = \overline{X}$

Теперь выразим статистику W

$W = \sqrt{n}\frac{\overline{X}-\theta_0}{\sqrt{\overline{X}}} \overset{d_{\theta_0}}{\to} \mathcal{N}(0,1)$
$S_1 = \left\{\sqrt{n}\frac{\overline{X}-\theta_0}{\sqrt{\overline{X}}} \geq Z_{1-\alpha}\right\}$
Для остальных критериев все аналогично

II - 2.2 Критерий отношения правдоподобия

Опр
$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\}$ - семейство распределений с плотностью p_θ

$\mathcal{L}_X(\theta)$ - функция правдоподобия

$\Lambda_{\theta_1,\theta_2}(X) = \frac{\mathcal{L}_X(\theta_1)}{\mathcal{L}_X(\theta_2)}$ - статистика отношения правдоподобия.

Замечание
Смысл: насколько θ_1 вероятнее, чем θ_2

1. $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ $(*)$. Причем $\theta_0 \neq \theta_1$

Теорема Неймана-Пирсона
В случае $(*)$, если $\exists c_\alpha : P_{\theta_0}\left(\Lambda_{\theta_1,\theta_0} > c_\alpha\right) = \alpha$, то критерий $S = \left\{\Lambda_{\theta_1,\theta_0} > c_\alpha\right\}$ имеет максимальную мощность среди критериев с уровнем доверия α

Задача 1 Часы отстают по времени на X - выборка из одного элемента, нужно поверить часы на оригинальность.

Оригинальные отстают на $X \sim \mathcal{N}(0,1)$, а поддельные – на $X \sim \mathcal{N}(0,100)$

Рассмотрим распределение $\mathcal{N}(0,\theta)$. То есть рассматриваемый параметр – дисперсия

$H_0 : \theta = 1$
$H_1 : \theta = 100$

Посчитаем статистику отношения правдоподобия

$$\Lambda_{\theta_1,\theta_0} = \frac{p_{\theta_1}(X_1)}{p_{\theta_0}(X_1)} = \frac{\frac{1}{\sqrt{2\pi\cdot 100}}e^{-x^2\frac{1}{2\cdot 100}}}{\frac{1}{\sqrt{2\pi\cdot 1}}e^{-x^2\frac{1}{2\cdot 1}}} = \frac{e^{-x^2\overbrace{\left(\frac{1}{200}-\frac{1}{2}\right)}^{<0}}}{10}$$

$\Lambda_{\theta_1,\theta_0}(x)$ возрастает по $|x|$

$$P_{\theta_0}\left(\frac{e^{-x^2\left(\frac{1}{200}-\frac{1}{2}\right)}}{10} > c_\alpha\right) = \alpha$$

Учитывая монотонность, заменим выражение под знаком вероятности

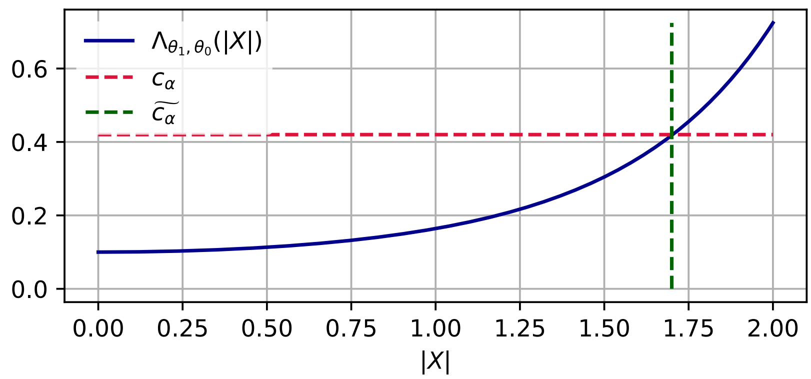


График 18: График $\Lambda_{\theta_1,\theta_0}(|X|)$

$$\exists\widetilde{c_\alpha}\ P_{\theta_0}\left(\frac{e^{-x^2\left(\frac{1}{200}-\frac{1}{2}\right)}}{10} > c_\alpha\right) = \alpha = P_{\theta_0}(|X| > \widetilde{c_\alpha})$$

$$P_{\theta_0}(|X| > \widetilde{c_\alpha}) = \alpha$$

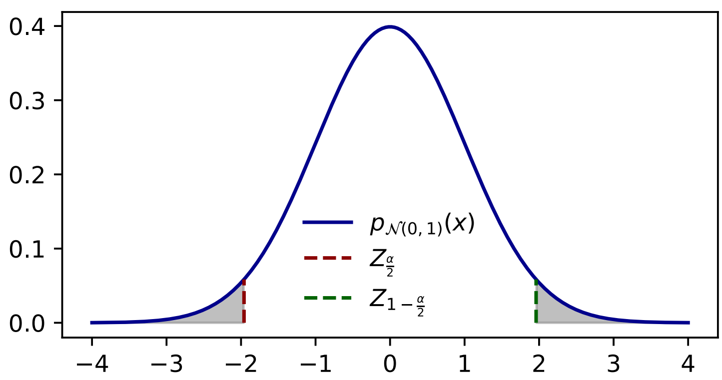


График 19: График квантилей нормального распределения

Новый критерий эквивалентен начальному (биекция между $|x|$ и $\Lambda(x)$)

Ответ:

$$S = \left\{|X| > Z_{1-\frac{\alpha}{2}}\right\}$$

$$S = \left\{\Lambda_{\theta_1,\theta_0}(x) > \Lambda_{\theta_1,\theta_0}(\widetilde{c_\alpha}), c_\alpha = Z_{1-\frac{\alpha}{2}}\right\}$$

2.

$H_0 : \theta = \theta_0$
$H_1 : \theta > \theta_0 (\clubsuit)$

Опр
Критерий S - равномерно наиболее мощный критерий (РНМК) уровня значимости α для проверки гипотез вида ($\clubsuit$), если $\forall \mathcal{R} \subseteq \mathcal{X}$ и $\forall \theta > \theta_0$ выполнено: $\mathfrak{P}_S(\theta) \geq \mathfrak{P}_{\mathcal{R}}(\theta)$

Теорема (о монотонном отношении правдоподобия)

Пусть $\theta_1 > \theta_0$

$\Lambda_{\theta_1, \theta_0}(X) = \frac{\mathcal{L}_X(\theta_1)}{\mathcal{L}_X(\theta_0)} = f_{\theta_1, \theta_0}\left(\underbrace{T(X)}_{\text{статистика}}\right)$, причем f возрастает по аргументу

Тогда $S = \{T(X) > c_\alpha\}$ - равномерно наиболее мощный критерий уровня значимости α для проверки ($\clubsuit$), где $c_\alpha : P(T(X) > c_\alpha) = \alpha$

Свойства:

1) В дискретном случае вместо α рассматриваем:
$\alpha_0 : \alpha_0 \leq \alpha, P_{\theta_0}(T(X) \geq c_\alpha) = \alpha_0$, α_0 - максимальное

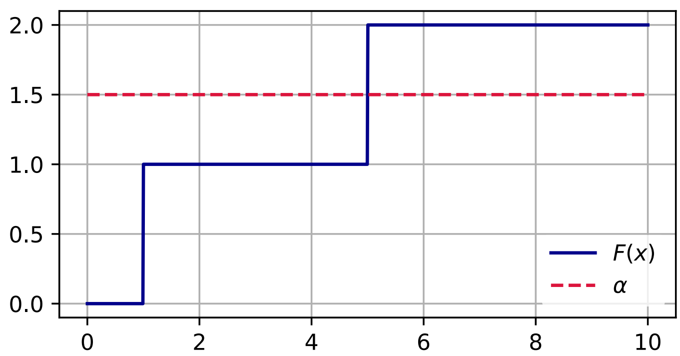


График 20: График дискретного распределения

2) Можно рассматривать $H_0 : \theta \leq \theta_0$

3) Если рассматривать случай $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$, то $S = \{T(X) < c_\alpha\}$

4) Если f_{θ_1, θ_0} из теоремы не возрастает, а убывает, то $S = \{T(X) < c_\alpha\}$

Пример:

$f_{\theta_1, \theta_0}(T(X))$ убывает по $T(X)$ и $H_0 : \theta \geq \theta_0, H_1 : \theta < \theta_0$

Тогда $S = \{T(X) > c_\alpha\}$

Задача 2

$X_1, ..., X_n \sim \mathrm{Exp}(\theta)$

Нужно построить РНМК уровня значимости α для проверки
$H_0 : \theta \leq \theta_0$
$H_1 : \theta > \theta_0$

$\theta_1 > \theta_0$

$$\Lambda_{\theta_1,\theta_0}(X) = \frac{(\theta_1)^n}{(\theta_0)^n} \cdot \frac{\exp\left(-\theta_1 \sum\limits_{i=1}^n X_i\right)}{\exp\left(-\theta_0 \sum\limits_{i=1}^n X_i\right)} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left((\theta_0 - \theta_1) \sum_{i=1}^n X_i\right)$$

$T(X) = \sum\limits_{i=1}^n X_i$

$\Lambda_{\theta_1,\theta_0}$ убывает по $T(X)$

$S = \left\{\sum\limits_{i=1}^n X_i < c_\alpha\right\}$

$c_\alpha : P_{\theta_0}(T(X) < c_\alpha) = \alpha$

И, так как $\sum\limits_{i=1}^n X_i = T(X) \sim \Gamma(\theta_0, n)$

$c_\alpha = \gamma_\alpha$ - α-квантиль $\Gamma(\theta_0, n)$

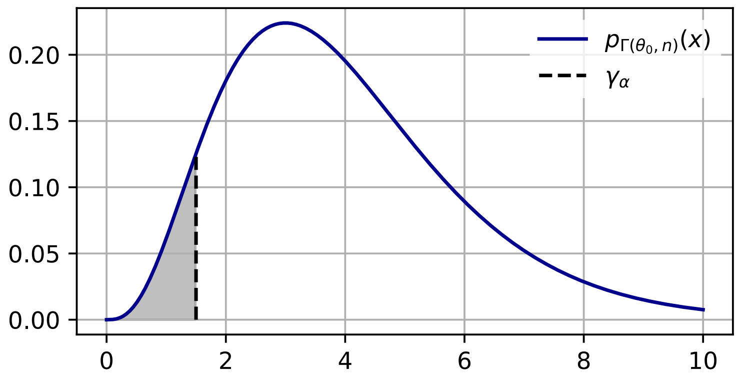


График 21: График квантилей $\Gamma(\theta_0, n)$

<u>Ответ:</u> $S = \left\{\sum\limits_{i=1}^n X_i < \gamma_\alpha\right\}$ **- РНМК**

В питоне:

```
c_alpha = sps.gamma.ppf(alpha, a = n, scale = 1 / theta_0)
```

Задача 3

$X_1, ..., X_n \sim \mathrm{Bern}(\theta)$

Построить РНМК уровня значимости α для проверки
$H_0 : \theta \geq \theta_0$
$H_1 : \theta < \theta_0$

$\theta_1 > \theta_2$
$\Lambda_{\theta_1,\theta_2}(X) = \frac{\theta_1^K(1-\theta_1)^{n-K}}{\theta_2^K(1-\theta_2)^{n-K}}, K = \sum\limits_{i=1}^{n} X_i$

$\Lambda_{\theta_1,\theta_2}(X) = \left(\frac{1-\theta_1}{1-\theta_2}\right)^n \left(\frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)}\right)^K$

Критерий будет иметь вид $S = \left\{\sum\limits_{i=1}^{n} X_i \overset{?}{<} c_\alpha\right\}$

$c_\alpha : P_{\theta_2}(T(X) < c_\alpha) = \alpha$

$T(X) = \sum\limits_{i=1}^{n} X_i \sim \mathrm{Binom}(n, \theta_0)$

Заменяем α на α_0, для которого выполняется условие $P_{\theta_2}\left(T(X) < c_{\alpha_0}\right) = \alpha_0$ и α_0 - максимальное $\alpha_0 < \alpha$

$c_{\alpha_0} = \begin{cases} u_\alpha\ , & \text{если } P_{\theta_0}\left(\sum\limits_{i=1}^{n} X_i \leq u_\alpha\right) = \alpha \\ u_\alpha - 1\ , & \text{иначе} \end{cases}$

Как найти c_{α_0} в питоне:

```
c_alpha0 = sps.binom(n = n, p = theta_0).ppf(alpha)
if sps.binom(n = n, p = theta_0).cdf(c_alpha0) > alpha:
    c_alpha0 -= 1
```

II - 3 Гипотезы и критерии. Повторение

II - 3.1 Гипотезы и критерии

__Опр__

Гипотеза:
$H_0 : P \in \mathcal{P}_0 \subseteq \mathcal{P}$
$H_0 : \theta \in \Theta_0 \subseteq \Theta$, а $P = P_\theta$

Критерий:
$S : \{x \in \mathcal{X} \mid x \in S \Leftrightarrow H_0 \text{ отвергается}\}$

Верность Отвергаем H_0	H_0 верна	H_0 не верна
H_0 отвергается	Ошибка I рода	😆
H_0 не отвергается	😳	Ошибка II рода

Опр

$P(I_S) = \sup\limits_{P \in \mathcal{P}_0} P(X \in S), \quad P(I_S) \leq \alpha$

$P(\mathrm{II}_S) = P(X \notin S)$ для $P \notin \mathcal{P}_0$, в случае использования альтернативной гипотезы $P \in \mathcal{P}_1$

Хотим ограничить опасную ошибку первого рода:

$\begin{cases} P(I_S) \leq \alpha \\ P(\mathrm{II}_S) \to \min_S \end{cases}$

$H_0 : \theta = \theta_0$ vs $\begin{cases} H_1 : \theta \neq \theta_0 \text{ двусторонняя} \\ H_1' : \theta > \theta_0 \text{ правосторонняя} \\ H_1'' : \theta < \theta_0 \text{ левосторонняя} \end{cases}$

$S_1 = \left\{ |W(X)| > Z_{1-\frac{\alpha}{2}} \right\}$
$S_2 = \{ W(X) > Z_{1-\alpha} \}$
$S_3 = \{ W(X) < Z_\alpha \}$

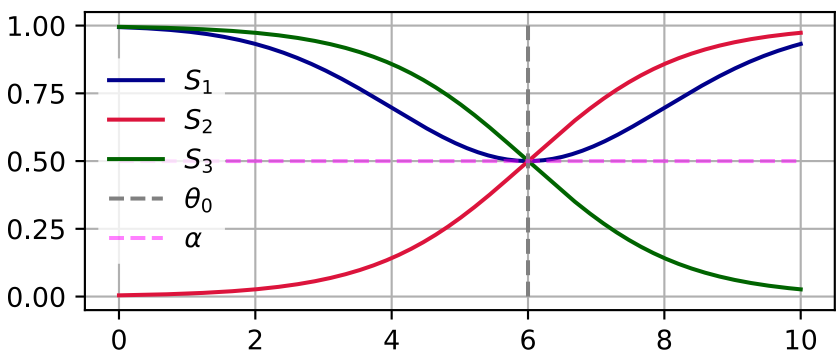


График 22: График мощностей для критериев S_1, S_2, S_3

Опр

Мощность критерия $\mathfrak{P}_S(p) = 1 - P(\mathrm{II}_S) = P(X \in S)$ для $P \notin \mathcal{P}_0$

II - 3.2 Доверительные интервалы

$W(X) = \sqrt{n}\frac{\hat{\theta} - \theta}{\hat{\sigma}} \xrightarrow{d} \mathcal{N}(0, 1)$

$\theta_0 \notin \left(\hat{\theta}_1, \hat{\theta}_2\right) \Leftrightarrow H_0$ отвергается

II - 4 Проверка гипотез. Семинар

II - 4.1 Оценка вероятности ошибки I рода

Опр

$P(I_S) = \sup\limits_{\theta \in \Theta_0} P_\theta(X \in S), X = (X_1, ..., X_n)$

Проблема:

Не всегда $\alpha = P(I_S)$, например:
- Критерий может быть асимптотическим
- На практике может не выполняться какое-то условие для применения критерия

Вывод:

Хотим оценивать $P(I_S)$

Пример:

$X_1, ..., X_n \sim \mathrm{Pois}(\theta)$

$H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

Запишем критерий Вальда:

$S = \left\{\sqrt{n}\frac{|\overline{X} - \theta_0|}{\sqrt{\overline{X}}} > Z_{1-\frac{\alpha}{2}}\right\}$

$P(I_S) = P_{\theta_0}\left(\sqrt{n}\frac{|\overline{X} - \theta_0|}{\sqrt{\overline{X}}} > Z_{1-\frac{\alpha}{2}}\right)$

Оценим вероятность ошибки I рода методом Монте-Карло:

1) Генерируем выборки $X^{(j)} = \left(X_1^{(j)}, ..., X_n^{(j)}\right) \sim \mathrm{Pois}(\theta_0)$
2) Проверяем критерий $I_j = \mathbb{I}\left\{X^{(j)} \in S\right\}$. Повторяем для $j = 1, ..., M$
   $I_j \sim \mathrm{Bern}\left(P_{I_S}\right)$
3) $\overline{I} =$ оценка $\mathbb{E}_{\theta_0}\mathbb{I}\{X \in S\} = P_{\theta_0}(X \in S)$

> **Замечание**
> В реальных задача важно понимать, насколько $P(I_S)$ близка к α.

Возникает вопрос: “Сколько итераций Монте-Карло нужно сделать?”

Ответим на него с использованием ЦПТ

$\gamma = P(I_S) \Rightarrow I_1, ..., I_M \sim \mathrm{Bern}(\gamma)$

Применим ЦПТ:

$\sqrt{M}\frac{\overline{I} - \gamma}{\sqrt{\gamma(1-\gamma)}} \xrightarrow{d} \mathcal{N}(0, 1)$

С вероятностью 95% γ лежит в интервале Вальда:

$\left(\overline{I} - Z_{0.975} \cdot \sqrt{\frac{\overline{I}\left(1-\overline{I}\right)}{M}}, \overline{I} + Z_{0.975} \cdot \sqrt{\frac{\overline{I}\left(1-\overline{I}\right)}{M}}\right)$

Посмотрим, что у нас получилось:

- $Z_{0.975} \leq 2$
- $\overline{I}\left(1 - \overline{I}\right) \leq \frac{1}{4}$

Значит, $Z_{0.975}\sqrt{\overline{I}\left(1 - \overline{I}\right)} \leq 1 \Rightarrow$ при $M \geq 10^6$ имеем, что $Z_{0.975}\frac{\sqrt{\overline{I}\left(1-\overline{I}\right)}}{\sqrt{M}} < 10^{-3}$

Замечание
Если γ не близка к $\alpha = 0.05$, то мы сможем это увидеть за $M \geq 10^6$ итераций.

Но если близка, то надо сделать сильно меньше итераций

Пример:

$\gamma \approx 0.05,\ M = 2 \cdot 10^5$. $\hat{\gamma} = \overline{I}$ будет лежать в интервале (0.049, 0.051) с вероятностью 0.95

Замечание
Если $\gamma = 0.1$, то при $M = 2 \cdot 10^5$ оценка $\hat{\gamma}$ с вероятностью 0.95 будет лежать в интервале $(0.086, 0.114)$

Замечание
Если H_0 не точечная($H_0 : \theta \in \Theta_0$), то можно взять несколько $\theta'_1, ..., \theta'_r \in \Theta_0$ и оценить P_{I_S} для них

Замечание
Аналогично можно оценивать мощность: берем $\theta \in \Theta_1$ и оцениваем вероятность отвергнуть критерий

II - 4.2 Критерий χ-квадрат

$X = (X_1, ..., X_n) \sim P$ с носителем $\mathcal{X}$

Рассмотрим разделение $\mathcal{X}$ на бины:

$\mathcal{X} = \bigsqcup_{i=1}^{k} B_j$

$\mu_j = \#\{i \mid X_i \in B_j\}$

$p_j^0 = P_0(X_1 \in B)$

Рассмотрим гипотезы:

$H_0 : P = P_0$ vs $H_1 : P \neq P_0$

Статистика: $\chi(x) = \sum_{j=1}^{n} \frac{\left(\mu_j - n \cdot p_j^0\right)^2}{n \cdot p_j^0} \xrightarrow{d_0} \chi_{k-1}^2$

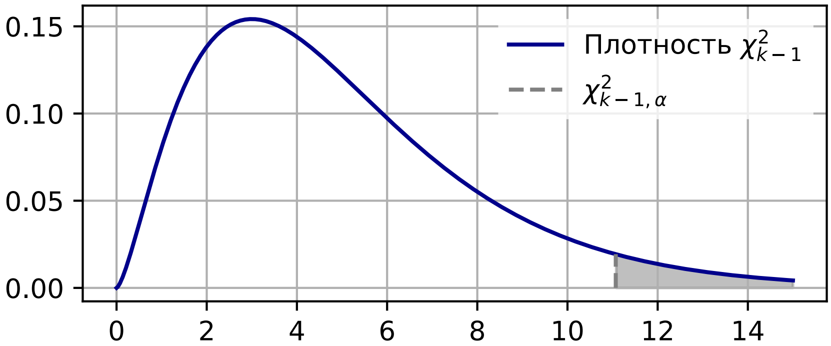


График 23: График плотности χ^2_{k-1}

Возьмем $\alpha = 0.05$

$S = \left\{\chi(X) > \chi^2_{k-1,1-\alpha}\right\}$

Задача

Есть генератор случайных чисел, нужно проверить, насколько он хороший. Числа генерируются из $\{1, 2, 3, 4\}$

$H_0 : P = U\{1, 2, 3, 4\}$ vs $H_1 : P \neq U\{1, 2, 3, 4\}$

$\mathcal{X} = \bigsqcup_{i=1}^{4} \{i\}$

$B_j = \{j\}$

$\#1 = 249, \#2 = 254, \#3 = 246, \#4 = 251$
$p_j^0 = \frac{1}{4}, \mu_j = \#j$

$\chi(x) = \sum\limits_{i=1}^{4} \frac{\left(\mu_j - n \cdot p_j^0\right)^2}{n \cdot p_j^0} = 0.136$

$\chi(X) \xrightarrow{d_0} \chi^2_3$

$\boxed{\textbf{Ответ: } S = \left\{\chi(X) > \chi^2_{3,1-\alpha}\right\}}$

В питоне считаеся так:

```
sps.chi2(df = 3).ppf(1 - alpha)
```

p-value:

$p(x) = P_{\chi^2_3}(\chi(X) > \xi) =$ `sps.chi2(df = 3).sf(xi)` $= 0.987$

$\xi = \chi(X)$ - статистика по реализации выборки

То есть мы не отвергаем нашу гипотезу

Замечание

Вообще, в таком случае наш генератор можно считать слишком хорошим и нам может захотеться отвергать такие случаи.

В таком случае нужно брать квантиль с другой стороны, важно чтобы вероятность попасть в такие области в сумме была не больше α

II - 4.3 Обобщенный критерий χ-квадрат

$X = (X_1, ..., X_n)$ из распределения P с носителем $\mathcal{X}$

$H_0 : P \in \mathcal{P}_0$ vs $H_1 : P \notin \mathcal{P}_0$

$\mathcal{P}_0 = \{P_\theta \mid \theta \in \Theta_0\}$

$\hat{\chi} = \sum_{i=1}^{k} \frac{\left(\mu_j - n \cdot p_j^0\left(\hat{\theta}\right)\right)^2}{n \cdot p_j^0}$

Задача

Есть районы 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Кол-во семян 🌱, попавших в районы 229, 211, 93, 35, 7, 0, 0, 1, 0

$H_0 : \overbrace{(X_1, ..., X_n)}^{X} \sim \mathrm{Pois}(\theta)$

$B_0 = \{0\}, B_1 = \{1\}, B_2 = \{2\}, B_3 = \{3\}, B_4 = \mathbb{N} \setminus \{0, 1, 2, 3\}$

1) $p_0^0(\theta) = \theta^0 e^{-\theta}$

$p_1^0(\theta) = \theta e^{-\theta}$

$p_2^0(\theta) = \frac{\theta^2}{2} e^{-\theta}$

$p_3^0(\theta) = \frac{\theta^3}{6} e^{-\theta}$

$p_4^0(\theta) = 1 - e^{-\theta}\left(1 + \theta + \frac{\theta^2}{2} + \frac{\theta^3}{6}\right)$

$\mu_j = \#j$

2) $\hat{\theta} = ?$

$\sum_{i=1}^{4} \mu_j \ln\left(p_j^0(\theta)\right) \to \max$

$\hat{\theta} = 0.93$

3) $\chi(X) = \sum_{j=0}^{4} \frac{\left(\mu_j - n \cdot p_j^0\left(\hat{\theta}\right)\right)^2}{n \cdot p_j^0\left(\hat{\theta}\right)} = 1.17$

p-value: 0.759 $\Rightarrow$ гипотеза не отвергается

II - 4.4 p-value, разбор ДЗ

Пусть $H_0 : \theta \in \Theta_0$ - основная гипотеза

S - критерий, $T(X)$ - статистика критерия

$x_1, ..., x_n$ – реализация выборки

$t = T(x_1, ..., x_n)$ - реализация статистики при данной реализациивыборки

Опр

p-value - вероятность при верности H_0 получить такое же значение статистики или более экстремальное

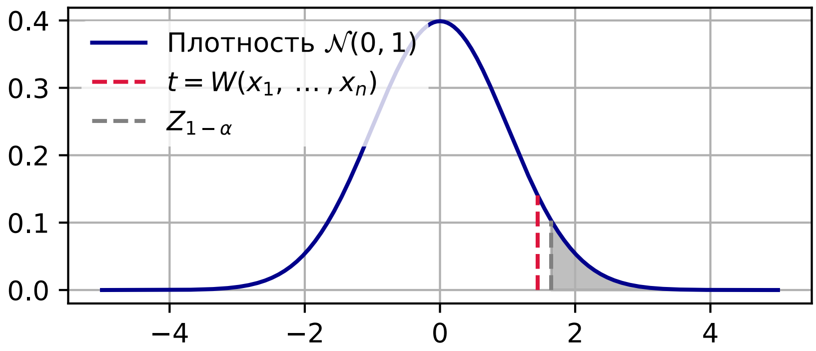


График 24: График нормального распределения и статистики W

$\theta > \theta_0$

$P_{\theta_0}(T(X) \geq c_\alpha)$

Утв:

H_0 отвергается $\Leftrightarrow$ p-value $\leq \alpha$

Задача 7(Автор решения: Хузин Э. Р.)

$T(X) = \overline{X}$

$S = \{T(X) \leq c_\alpha\}$

$P\left(\sqrt{n}\left(\overline{X} - \theta_0\right) \leq \sqrt{n}(c_\alpha - \theta_0)\right) = F_{\mathcal{N}(0,1)}(\sqrt{n}(c_\alpha - \theta_0)) = \alpha$

$c_\alpha = \theta_0 + \frac{Z_\alpha}{\sqrt{n}}$

$P_\theta(X \in S) = P_\theta\left(\overline{X} \leq \theta_0 + \frac{Z_\alpha}{\sqrt{n}}\right) + P_\theta\left(\sqrt{n}\left(\overline{X} - \theta\right) \leq \sqrt{n}(\theta_0 - \theta) + Z_\alpha\right)$

Задача 7(Решение семинариста)

$T(X) = \sum\limits_{i=1}^{n} X_i$

$S = \{T(X) < c_\alpha\}$

c_α - α -квантиль $\mathcal{N}(\theta_0 \cdot n, n)$

Посчитаем p-value:

$x_1, ..., x_n$ - реализация выборки

$t = \sum\limits_{i=1}^{n} x_i$ - реализация статистики

$p(t) = P\left(\sum\limits_{i=1}^{n} X_i \leq t\right)$

В питоне:

```
sps.norm(a = theta_0 * n, scale = np.sqrt(n).cdf(sample.limits(sum)())
```

Задача 8

$X_i \sim \Gamma(\theta, \beta)$

$\mathcal{L}_X(\theta) = \theta^{n\beta} \prod\limits_{i=1}^{n} X_i^{\beta-1} e^{-\theta \sum\limits_{i=1}^{n} X_i}$

$\sum\limits_{i=1}^{n} X_i \sim \Gamma(\theta, n\beta)$

$\theta_2 > \theta_1$

$T(X) = \sum\limits_{i=1}^{n} X_i$

$\Lambda_{\theta_2, \theta_1} = \left(\frac{\theta_2}{\theta_1}\right)^{n\beta} \exp\left(-\sum\limits_{i=1}^{n} X_i (\theta_2 - \theta_1)\right)$ - убывает по $T(X)$

$H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$

$S = \{T(X) > c_\alpha\}$ - РНМК

c_α – это $1 - \alpha$ квантиль $\Gamma(\theta_0, n\beta)$

$x_1, ..., x_n$ - реализация выборки

$t = \sum\limits_{i=1}^{n} x_i$

$p(t) = P_{\theta_0}\left(\sum\limits_{i=1}^{n} x_i \geq t\right) =$

```
= sps.gamma(a=n*b, scale = 1/theta_0).sf(sample.limits(sum)())
```

Задача

H_0 : часы оригинальные $P = \mathcal{N}(0, 1)$ vs H_1 : поддельные $P = \mathcal{N}(0, 100)$

$T(X) = |X|$

$S(X) = \left\{|X| > Z_{1-\frac{\alpha}{2}}\right\}$

$p(x) = P_{\theta_0}(|X| \geq 2) = P_{\mathcal{N}(0,1)}(|X| > 2) = 0.0453$

$x = 2$ - реализация

II - 5 МПГ. Семинар

Есть m выборок $\left(X_{1,1}, ..., X_{1,n_1}\right), ..., \left(X_{m,1}, ..., X_{m,n_m}\right)$

Есть m пар гипотез H_1 vs $H_1', ..., H_m$ vs H_m'

И есть m критериев $S_1, ..., S_m$

II - 5.1 Обобщенная ошибка первого рода

Опр

1) $V_{P,S}(X)$ – количество верных основных гипотез, которые были отвергнуты
2) $R_S(X)$ – количество отвергнутых основных гипотез

Опр

Групповая ошибка I рода $\mathrm{FWER} = P\left(V_{P,S}(X) > 0\right)$

Ожидаемая доля ложных отклонений $\mathrm{FDR} = \mathbb{E}\frac{V_{P,S}(X)}{\max(1, R_S(X))}$

Теорема

$\mathrm{FDR} \leq \mathrm{FWER}$

▲

$$\mathrm{FDR} = \mathbb{E}\frac{V_{P,S}(X)}{\max(1, R_S(X))} = \mathbb{E}\underbrace{\frac{V_{P,S}(X)}{\max(1, R_S(X))}}_{\leq 1} \cdot \mathbb{I}\left\{V_{P,S}(X) > 0\right\} \leq$$

$$\leq \mathbb{E}1 \cdot \mathbb{I}\left\{V_{P,S}(X) > 0\right\} = P\left(V_{P,S}(X) > 0\right) = \mathrm{FWER}$$

■

Задача

Пусть $X_j = \left(X_{j,1}, ..., X_{j,n_j}\right), j \in \{1, 2\}$

S_j – критерий для проверки гипотез H_j vs H_j'

Пусть мы знаем, что $P_j\left(I_{S,j}\right) = \alpha$

- Какие значения может принимать FWER(нужно привести примеры для крайних случаев)?
- Какое значение будет иметь FWER, если выборки являются независимыми?

1) $\mathrm{FWER} \geq \alpha$

$$\mathrm{FWER} = P\big(V_{P,S}(X) > 0\big) =$$
$$= P(\text{в 1 случае ошибка I рода} \vee \text{во 2 случае ошибка I рода}) \geq$$
$$\geq P(\text{ошибка I рода в 1 случае}) = \alpha$$

2) $\mathrm{FWER} = 2\alpha$

$X_1 = X_2, H_1 = H_2, S_1 = S_2$

$$\mathrm{FWER} \leq P(\text{в 1 случае ошибка I рода}) +$$
$$+ P(\text{во 2 случае ошибка I рода}) = 2\alpha$$

Если $X_1 = X_2, H_1 = H_2, S_1 \cap S_2 = \emptyset$, то $\mathrm{FWER} = 2\alpha$

3) Выборки независимы

$$\mathrm{FWER} = P_1\big(I_{S_1}\big) + P_2\big(I_{S_2}\big) -$$
$$- P(\text{в 1 случае ошибка I рода} \wedge \text{во 2 случае ошибка I рода}) =$$
$$= 2\alpha - P(\text{1 случай}) \cdot P(\text{2 случай}) = 2\alpha - \alpha^2$$

Если $\alpha = 0.05$, то $\mathrm{FWER} = 0.0975$

Вывод: случай независимости выборок – почти наихудший с точки зрения FWER, если не применить МПГ

> **Замечание**
> Цель МПГ – подобрать такие α_j, являющиеся уровнями значимости критериев(скорректированными значениями p-value), что $\mathrm{FWER} \leq \alpha$

II - 5.2 Методы, контролирующие FWER

1) Метод Бонферрони

Скорректируем α и p-value: $\alpha_j = \frac{\alpha}{m}, \tilde{p}_j = \min\big(1, m \cdot p_j\big)$

2) Метод Холма

- Сортируем гипотезы в порядке возрастания p-value
- Идем по возрастанию p-value и берем $\alpha_j = \frac{\alpha}{m-j+1}$

  Скорректированные p-value имеют вид
  $\tilde{p}_j = \min\big(1, \max\big(\tilde{p}_{j-1}, (m-j+1)p_j\big)\big)$

  p_j сраниваем с α_j или $\tilde{p}_j$ сравниваем с α
- Пока $p_j \leq \alpha_j$ ($\tilde{p}_j \leq \alpha$), отвергаем, все остальное – не отвергаем

Замечание
Если нет информации о зависимости / независимости, то метод неулучшаем(отвергаем больше всего с контролем FWER)

3) Метод Шидака

   Скорретируем $\alpha : \alpha_j = 1 - \sqrt[m]{1-\alpha}$

Теорема

Если $S_j = \{T_j(X) \geq c_\alpha^j\}$ и $P(T_1 < t_1, ..., T_m < t_m) \geq \prod_{j=1}^{m} P(T_j < t_j)$.

Тогда FWER $\leq \alpha$ в методе Шидака

▲

Пусть $1, ..., m_0$ – верные гипотезы. Знаем, что $m_0 \in \{0, 1, ..., m\}$

$$\text{FWER} = P(V_{P,S}(X) > 0) = P\left(\bigcup_{j=1}^{m_0} \left(T_j(X_j) \geq c_{\alpha,j}^j\right)\right) =$$

$$= 1 - P\left(\bigcap_{j=1}^{m_0} \left(T_j(X_j) \geq c_{\alpha,j}^j\right)\right) \leq 1 - \prod_{j=1}^{m} \left[1 - P\left(T_j(X_j) \geq c_{\alpha,j}^j\right)\right] \leq$$

$$\leq 1 - \prod_{i=1}^{m_0} (1 - \alpha_j) = 1 - \prod_{j=1}^{m_0} \left(1 - 1 + \sqrt[m]{1-\alpha}\right) = 1 - (1-\alpha)^{\frac{m_0}{m}} \leq$$

$$\leq 1 - (1 - \alpha) = \alpha$$

■

4) Методы Шидака-Холма

   Процедура аналогична процедуре в методе Холма, только

   $\alpha_j = 1 - \sqrt[m-j+1]{1-\alpha}$

Замечание
Метод Шидака можно применять не всегда, а метод Шидака-Холма работает в случае независимых статистик критерия

Задача

Дано, что $p_1 = 0.022, p_2 = 0.041, m = 2, \alpha = 0.05$

- Бонферрони: $\tilde{p}_1 = 0.044$ – отвергаем, $\tilde{p}_2 = 0.082$ – не отвергаем
- Холм: $\tilde{p}_1 = 0.044$ – отвергаем, $\tilde{p}_2 = 0.044$ – отвергаем

Метод Бонферрони отвергает 1 раз, а метод Холма – 2 раза

Задача

Пусть $X_1, ..., X_n$ – выборка из распределения P

$H_0 : P \in \mathcal{P}_0$ vs $H_1 P \in \mathcal{P}_1$

Было решено проверить гипотезы тремя различными критериями, получили следующие значения p-value : $0.0001, 0.0482, 0.7361$. Какое решение об отвержении или не отвержении принять для $\alpha = 0.05$?

Решим методом Холма, скорректированные p-value следующие:
$0.0003, 0.0964, 0.7361$

1 из 3 критериев отвергает H_0, однако H_0 одна на все случаи и $\mathrm{FWER} = P\big(V_{P,S}(X) > 0\big) \leq \alpha$, то есть вероятность ошибиться хотя бы 1 раз $\leq \alpha$

Принимаем решение 1 раз $\Rightarrow \mathrm{FWER} = P\big(S_{1,2,3}\big) \leq \alpha \Rightarrow$можем отвергать H_0

II - 6 Гауссовская линейная модель. Статистические свойства линейной регресии

II - 6.1 Свойства гауссовских случайных векторов

1) Пусть $\xi = (\xi_1, ..., \xi_d) \sim \mathcal{N}(a, \Sigma)$ – гауссовский вектор, имеет плотность

$p(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \exp\Big(-\frac{1}{2}(x-a)^T \Sigma^{-1} (x-a)\Big)$

2) $\mathbb{E}\xi = a, \mathbb{D}\xi = \Sigma$

3) Линейная комбинация $B\xi$ гауссовского вектора ξ тоже есть гауссовский вектор, причем

   $B\xi \sim \mathcal{N}\big(Ba, B\Sigma B^T\big)$

4) Если $\xi \sim \mathcal{N}(a, \Sigma)$, то $\Sigma^{-\frac{1}{2}}(\xi - a) \sim \mathcal{N}(0, I_d)$

5)

<u>Теорема (об ортогональном разложении гауссовского вектора)</u>
Пусть $\xi \sim \mathcal{N}\big(a, \sigma^2 I_n\big)$ – гауссовский случайный вектор размерности n с независимыми компонентами. Пусть также $L_1 \oplus ... \oplus L_k = \mathbb{R}^n$ – разложение в сумму ортогональных подпространств и $\eta_j = \mathrm{proj}_{L_j} \xi$ – проекция вектора ξ на L_j

Тогда $\eta_1, ..., \eta_k$ независимы в совокупности, причем

$\mathbb{E}\eta_j = \mathrm{proj}_{L_j}(a)$

$\frac{1}{\sigma^2}\big\|\eta_j - \mathbb{E}\eta_j\big\|^2 \sim \chi^2_d, d_j = \dim L_j$

II - 6.2 Гауссовская линейная модель

$Y = X\theta + \varepsilon, Y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times d}, \theta \in \mathbb{R}^d, \varepsilon \in \mathbb{R}^n$

Случайны: Y, ε

Константы: X

Параметры: θ

При этом $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$

То есть шум:

- нормальный
- несмещенный
- гомоскедастичный

Следствие: $Y \sim \mathcal{N}(X\theta, \sigma^2 I_n)$

Утв

Пусть $\hat{\theta} = \left(X^T X\right)^{-1} X^T Y$. тогда

1) $\hat{\theta}$ и $Y - X\hat{\theta}$ независимы

2) $\frac{1}{\sigma^2}\left\|X\theta - X\hat{\theta}\right\|^2 \sim \chi_d^2$

3) $\frac{1}{\sigma^2}\left\|Y - X\hat{\theta}\right\|^2 \sim \chi_{n-d}^2$

▲

$L(X) = \left\{X\theta \,\middle|\, \theta \in \mathbb{R}^d\right\}, \dim L(X) = d$

$\mathbb{R}^n = L(X) \oplus L^{\perp}(X)$

$\mathrm{proj}_{L(X)} Y = X\tilde{\theta}$, где $\tilde{\theta} = \underset{\theta \in \mathbb{R}^d}{\arg\min} \|Y - X\theta\|^2 = \hat{\theta}$, так как задача МНК

$\mathrm{proj}_{L^{\perp}(X)} Y = Y - X\hat{\theta}$

По теореме о разложении гауссовского вектора: $X\hat{\theta}$ и $Y - X\hat{\theta}$ независимы

Также $\hat{\theta} = \left(X^T X\right)^{-1}\left(X^T X\right)\hat{\theta} = \left[\left(X^T X\right)^{-1} X^T\right] X\hat{\theta}$, то есть, $\hat{\theta}$ – линейная комбинация $X\hat{\theta} \Rightarrow \hat{\theta}$и$Y - X\hat{\theta}$ независимы, что доказывает первый пункт

$\frac{1}{\sigma^2}\left\|X\hat{\theta} - \mathbb{E}X\hat{\theta}\right\| = \frac{1}{\sigma^2}\left\|X\hat{\theta} - X\theta\right\| \sim \chi_d^2$, что доказывает второй пункт

$\frac{1}{\sigma^2}\left\|Y - X\hat{\theta} - \mathbb{E}\left(Y - X\hat{\theta}\right)\right\|^2 = \frac{1}{\sigma^2}\left\|Y - X\hat{\theta}\right\|^2 \sim \chi_{n-d}^2$

■

II - 6.3 Доверительные интервалы и критерии для параметров GLM

Замечание
Далее α – уровень значимости критерия, а $1 - \alpha$ – уровень доверия интервала

1. Доверительный интервал для σ^2

Из утверждения $\frac{1}{\sigma^2}\left\|Y - X\hat{\theta}\right\|^2 \sim \chi^2_{n-d}$

Несмещенная оценка σ^2 равна $\hat{\sigma}^2 = \frac{1}{n-d}\left\|Y - X\hat{\theta}\right\|^2$

$\frac{(n-d)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-d}$

$P\left(\chi^2_{n-d,\frac{\alpha}{2}} \leq \frac{(n-d)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{n-d,1-\frac{\alpha}{2}}\right) = 1 - \alpha$

$P\left(\frac{(n-d)\hat{\sigma}^2}{\chi^2_{n-d,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-d)\hat{\sigma}^2}{\chi^2_{n-d,\frac{\alpha}{2}}}\right) = 1 - \alpha$

Тогда $\left(\frac{(n-d)\hat{\sigma}^2}{\chi^2_{n-d,1-\frac{\alpha}{2}}}, \frac{(n-d)\hat{\sigma}^2}{\chi^2_{n-d,\frac{\alpha}{2}}}\right)$ – доверительный интервал для σ^2 уровня доверия $1 - \alpha$

1. Доверительный интервал для θ_j и критерий для гипотезы $H_0 : \theta_j = 0$

Замечание
Такой критерий также называется критерием о незначимости коэффициента линейной регрессии

Утв

$\forall c \in \mathbb{R}^d \ \frac{c^T\left(\hat{\theta}-\theta\right)}{\hat{\sigma}\sqrt{c^T(X^TX)^{-1}c}} \sim T_{n-d}$

▲

- $\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\left(X^TX\right)^{-1}\right)$

  $Y \sim \mathcal{N}(X\theta, \sigma^2 I_n)$

  $\hat{\theta} = \left(X^TX\right)^{-1}X^TY \Rightarrow \mathbb{E}\hat{\theta} = \left(X^TX\right)^{-1}X^TX\theta = \theta$

  $$\mathbb{D}\hat{\theta} = \left(\left(X^TX\right)^{-1}\sigma^2 I_n\left(X\left(X^TX\right)^{-T}\right)\right) = \sigma^2 I_n$$
- $c^T\hat{\theta} \sim \mathcal{N}\left(c^T\theta, \sigma^2 c^T\left(X^TX\right)^{-1}c\right)$
- $\xi = \frac{c^T\left(\hat{\theta}-\theta\right)}{\sigma\sqrt{c^T(X^TX)^{-1}c}} \sim \mathcal{N}(0,1)$
- $\eta = \frac{(n-d)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-d}$

Осталось только показать, что $\xi \perp\!\!\!\perp \eta$

ξ зависит только от $\hat{\theta}$, η только от $Y - X\hat{\theta}$, а по утверждению они независимы

Тогда $\frac{\xi}{\sqrt{\frac{\eta}{n-d}}} = \frac{c^T\left(\hat{\theta}-\theta\right)}{\hat{\sigma}\sqrt{c^T(X^TX)^{-1}c}} \sim T_{n-d}$

■

Пусть $c_j = \left(0, ..., 0, \underbrace{1}_{i}, 0, ..., 0\right)$

Тогда $c_j^T\theta = \theta_j$

$\frac{c_j^T\left(\hat{\theta}-\theta\right)}{\hat{\sigma}\sqrt{c_j^T(X^TX)^{-1}c_j}} = \frac{\hat{\theta}_j-\theta_j}{\hat{\sigma}\left((X^TX)^{-1}_{jj}\right)} \sim T_{n-d}$ по утверждению

$P\left(\left|\frac{\hat{\theta}_j-\theta_j}{\hat{\sigma}\left((X^TX)^{-1}_{jj}\right)}\right| \leq t_{n-d,1-\frac{\alpha}{2}}\right) = 1 - \alpha$

Тогда интервал следующий:

$\left(\hat{\theta}_j - \hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}t_{n-d,1-\frac{\alpha}{2}}, \hat{\theta}_j + \hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}t_{n-d,1-\frac{\alpha}{2}}\right)$ – доверительный интервал уровня доверия $1 - \alpha$ для θ_j

Для $H_0 : \theta_j = 0$

Рассмотрим статистику $T_j^0(X,Y) = \frac{\hat{\theta}_j}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}}$

$T_j^0(X,Y) \overset{H_0}{\sim} T_{n-d}$

$P_0\left(\left|T_j^0(X,Y)\right| > t_{n-d,1-\frac{\alpha}{2}}\right) = \alpha$

$S = \left\{\left|T_j^0(X,Y)\right| > t_{n-d,1-\frac{\alpha}{2}}\right\}$ – критерий для проверки $H_0 : \theta_j = 0$ уровня значимости α

3. Доверительная область для $\theta \in \mathbb{R}^d$

> **<u>Опр</u>**
> Пусть $\xi \sim \chi^2_{k_1}$, $\eta \sim \chi^2_{k_2}$ и $\xi \perp\!\!\!\perp \eta$
>
> Тогда $\zeta = \frac{k_2\xi}{k_1\eta} = \frac{\frac{\xi}{k_1}}{\frac{\eta}{k_2}} \sim F_{k_1,k_2}$ – распределение Фишера со степенями свободы k_1, k_2

Имеем:

- $\xi = \frac{1}{\sigma^2}\left\|X\hat{\theta} - X\theta\right\|^2 \sim \chi^2_d$
- $\eta = \frac{1}{\sigma^2}\left\|Y - X\hat{\theta}\right\| \sim \chi^2_{n-d}$
- ξ зависит от $X\hat{\theta}$, $\hat{\theta}$ – ЛК $X\hat{\theta}$, η зависит от $Y - X\hat{\theta}$, $\hat{\theta}$ независима с $Y - X\hat{\theta} \Rightarrow$ $\xi \perp\!\!\!\perp \eta$

Тогда $\frac{(n-d)\left\|X\hat{\theta}-X\theta\right\|}{d\left\|Y-X\hat{\theta}\right\|^2} \sim F_{k_1,k_2}$

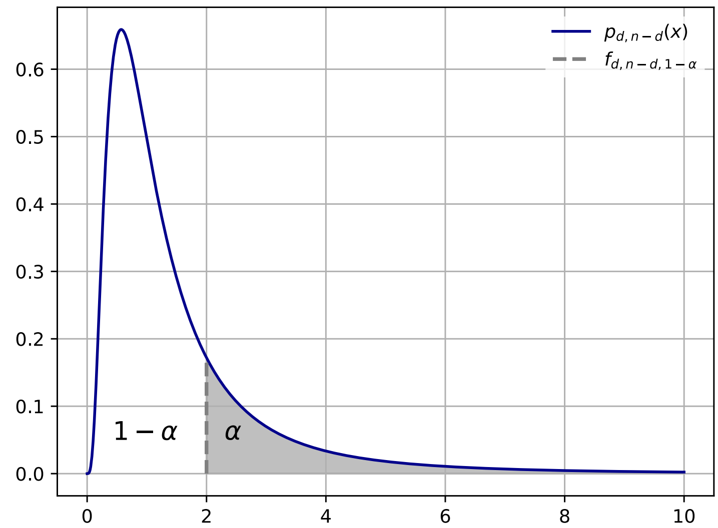


График 25: График квантилей распределения Фишера

$P\left(F(X,Y) \le f_{d,n-d,1-\alpha}\right) = 1-\alpha$

Тогда $\left\{\theta \in \mathbb{R}^d \,\middle|\, F(X,Y) \le f_{d,n-d,1-\alpha}\right\}$ – доверительный эллипсоид для θ уровня доверия $1-\alpha$

II - 7 Гауссовская линейная модель. Семинар

II - 7.1 Гауссовская линейная модель

Пусть $X \in \mathbb{R}^{n \times d}$ – признаки

Y – таргеты

Предпологаемая модель $Y = X\theta + \varepsilon$

При этом $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$

То есть шум:

- нормальный
- несмещенный
- гомоскедастичный

Заполним таблицу

	coeff	stderr	t	$P > \|t\|$	$[0.025, 0.975]$
X_j	$\hat{\theta}_j$	$\sqrt{\left(\hat{\sigma}^2 (X^T X)^{-1}\right)_{jj}}$	$T_j^0(X, Y)$	$P_{T_{n-d}}\left(\left\|T_j^0(X, Y)\right\| \geq T_{n-d, 1-\frac{\alpha}{2}}\right)$	$\hat{\theta}_j \pm T_{n-d,1-\frac{\alpha}{2}} \cdot \cdot \hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}$

1) $\hat{\theta}_j$

   $\hat{\theta} = (X^T X)^{-1} X^T Y, \hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2 (X^T X)^{-1}\right)$

2) Оценка дисперсии ошибки:

   $\hat{\sigma}^2 = \frac{1}{n-d} \|Y - X\theta\|_2^2$

3) Оценка дисперсии коэффициентов:

   $\underbrace{\frac{1}{n-d} \|Y - X\theta\|_2^2}_{\hat{\sigma}^2} (X^T X)^{-1}$

4) T-статистика для коэффициента $\hat{\theta}_j$

   $T_j^0(X, Y) = \frac{\hat{\theta}_j - \theta_j}{\hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}} \sim T_{n-d}$

   $H_0 : \theta_j = 0$

   $T_j^0(X, Y) \overset{H_0}{=} \frac{\hat{\theta}_j}{\hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}} \overset{H_0}{\sim} T_{n-d}$

5) Критерий для проверки H_0:

   $S = \left\{\left|T_j^0(X, Y)\right| > T_{n-d, 1-\frac{\alpha}{2}}\right\}$

6) p-value

   $p(t) = P_{T_{n-d}}\left(\left|T_j^0(X, Y)\right| \geq t\right)$

7) Доверительный интервал

   $P\left(\left|T_j^0(X, Y)\right| < T_{n-d, 1-\frac{\alpha}{2}}\right) = 1 - \alpha$

   $\left(\hat{\theta}_j - T_{n-d, 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}, \hat{\theta}_j + T_{n-d, 1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}\right)$

II - 7.2 Общий случай линейных гипотез

Опр

$H_0 : T\theta = \tau, T \in \mathbb{R}^{k \times d}, \tau \in \mathbb{R}^k, \tau = \text{const}, k \leq d$

Гипотезы такого вида называются линейными

Пример:

$$H_0 = \begin{cases} \theta_1 = 0 \\ \theta_2 = \theta_3 \Leftrightarrow \theta_2 - \theta_3 = 0 \end{cases}$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, T = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \end{pmatrix}, \tau = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

1) Строим критерий

$\hat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2 \left(X^T X\right)^{-1}\right)$

$$\hat{t} = T\hat{\theta} \sim \mathcal{N}\left(\underbrace{T\theta}_{\tau(H_0)}, \sigma^2 \underbrace{\left(T\left(X^T X\right)^{-1} T^T\right)}_{B}\right)$$

$\frac{B^{-\frac{1}{2}}\left(\hat{t}-\tau\right)}{\sigma} \overset{H_0}{\sim} \mathcal{N}(0, I_k)$

$\frac{\left(\hat{t}-\tau\right)B^{-1}\left(\hat{t}-\tau\right)}{\sigma^2} = \left(\frac{B^{-\frac{1}{2}}\left(\hat{t}-\tau\right)}{\sigma}\right)^T \cdot \frac{B^{-\frac{1}{2}}\left(\hat{t}-\tau\right)}{\sigma} = \sum_{i=1}^{k} z_i^2$, z_i – координаты вектора $\frac{B^{-\frac{1}{2}}\left(\hat{t}-\tau\right)}{\sigma}$

$\sum_{i=1}^{k} z_i^2 \overset{H_0}{\sim} \chi_k^2$, так как $z_i \overset{H_0}{\sim} \mathcal{N}(0, 1)$

$\sum_{i=1}^{k} z_i^2$ зависит от $\hat{\theta}$

Величина

$\frac{1}{\sigma^2} \|Y - X\theta\|_2^2 \sim \chi_{n-d}^2$ не зависит от $\hat{\theta}$

Тогда $F(X, Y) = \frac{\left(\hat{t}-\tau\right)^T B^{-1}\left(\hat{t}-\tau\right)}{\left\|Y - X\hat{\theta}\right\|_2^2} \cdot \frac{n-d}{k} \sim F_{k,n-d}$

Теперь построим критерий уровня значимости α, используя распределение Фишера

$S = \left\{F(X, Y) > f_{k,n-d,1-\alpha}\right\}$

Задача

$X_1, \dots, X_n \sim \mathcal{N}\left((\theta_1, \dots, \theta_1), \sigma^2 I_n\right)$

$Y_1, \dots, Y_m \sim \mathcal{N}\left((\theta_2, \dots, \theta_2), \sigma^2 I_m\right)$

$X_i = \theta_1 + \varepsilon_i$

$Y_i = \theta_2 + \varepsilon_i$

Мы хотим проверить гипотезу $H_0 : \theta_1 = \theta_2$

Сведем к GLM:

$$W = \begin{pmatrix} X_1 \\ \dots \\ X_n \\ Y_1 \\ \dots \\ Y_m \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

Размер матрицы $(n + m) \times 2$

$W = Z\theta + \varepsilon$ – GLM

$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$

$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_{n+m})$

$\boxed{k = 1}$

$T\theta = \tau$

$T = \begin{pmatrix} 1 & -1 \end{pmatrix}$

$\tau = 0$

GLM: $Y = X\theta + \varepsilon$

$Z^T Z = \begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix}$

$B = \frac{1}{n} + \frac{1}{m}$

$(Z^T Z)^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{m} \end{pmatrix}$

$\hat{t} = T\hat{\theta} = T(Z^T Z)^{-1} Z^T W = \overline{X} - \overline{Y}$

$\hat{\theta} = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}$

Выпишем статистики из предудыщей задачи:

$\frac{(\hat{t}-\tau) B^{-1} (\hat{t}-\tau)}{\sigma^2} = \frac{(\overline{X}-\overline{Y})^T (\frac{1}{n}+\frac{1}{m})^{-1} (\overline{X}-\overline{Y})}{\sigma^2} = \frac{(\overline{X}-\overline{Y})^2}{(\frac{1}{n}+\frac{1}{m})\sigma^2} \sim \chi_1^2$

$\left\|W - Z\hat{\theta}\right\|_2^2 = \left\|X - \hat{\theta}_1\right\|^2 + \left\|Y - \hat{\theta}_2\right\|^2 = n\hat{S}_X^2 + m\hat{S}_Y^2$

Распишем отдельно каждую из компонент:

$\frac{\left\|X-\hat{\theta}_1\right\|^2}{\sigma^2} \sim \chi_{n-1}^2$

$\frac{\left\|Y-\hat{\theta}_2\right\|^2}{\sigma^2} \sim \chi^2_{m-1}$

Значит, $\frac{\left\|W-Z\hat{\theta}\right\|^2}{\sigma^2} \sim \chi^2_{n+m-2}$

$$F(Z,W) = \frac{\left(\overline{X}-\overline{Y}\right)^2(n+m-2)}{\left(\frac{1}{n}+\frac{1}{m}\right)\left(n\hat{S}_X^2+m\hat{S}_Y^2\right)} \sim F_{1,n+m-2}$$

$$\boxed{S_F = \left\{F_{Z,W} > f_{1,n+m-2,1-\alpha}\right\}}$$

$$\sqrt{\frac{\left(\overline{X}-\overline{Y}\right)^2}{\left(\frac{1}{n}+\frac{1}{m}\right)\sigma^2}} \sim \mathcal{N}(0,1)$$

$$T(Z,W) = \frac{\left(\overline{X}-\overline{Y}\right)\sqrt{n+m-2}}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sigma\sqrt{n\hat{S}_X^2+m\hat{S}_Y^2}} \sim T_{n+m-2}$$

$$\boxed{S_T = \left\{|T(Z,W)| > t_{n+m-2,1-\frac{\alpha}{2}}\right\}}$$

II - 8 Подсчет информации Фишера для частных случаев GLM

II - 8.1 Логистическая регрессия

$X_1, ..., X_n \in \mathbb{R}^d$ – векторы признаков

$Y_1, ..., Y_n \in \{0,1\}$ – классы

$\mu_\theta(x) = \frac{1}{1+e^{x^\top \theta}}, Y_i \sim \mathrm{Bern}(\mu_\theta(X_i))$

Запишем функцию правдоподобия

$$\mathcal{L}_X(\theta) = \prod_{i=1}^n \mu_\theta(X_i)^{Y_i}(1-\mu_\theta(X_i))^{1-Y_i} = \prod_{i=1}^n \sigma\left(X_i^\top\theta\right)^{Y_i}\left(1-\sigma\left(X_i^\top\theta\right)\right)^{1-Y_i}$$

$$\ell_Y = \sum_{i=1}^n \left(Y_i \ln\left(X_i^\top\theta\right) + (1-Y_i)\ln\left(1-\sigma\left(X_i^\top\theta\right)\right)\right)$$

$$\begin{aligned}\nabla\ell_Y(\theta) &= \sum_{i=1}^n \frac{Y_i\sigma\left(X_i^\top\theta\right)\left(1-\sigma\left(X_i^\top\theta\right)\right)}{\sigma\left(X_i^\top\theta\right)}X_i + (1-Y_i)\frac{\sigma\left(X_i^\top\theta\right)\left(1-\sigma\left(X_i^\top\theta\right)\right)}{1-\sigma\left(X_i^\top\theta\right)}(-X_i) = \\ &= \sum_{i=1}^n \left(Y_iX_i - Y_iX_i\sigma\left(X_i^\top\theta\right) - X_i\sigma\left(X_i^\top\theta\right) + Y_iX_i\sigma\left(X_i^\top\theta\right)\right) = \sum_{i=1}^n X_i\left(Y_i - \sigma\left(X_i^\top\theta\right)\right) = \\ &= X^\top(Y - S(\theta))\end{aligned}$$

$$S(\theta) = \begin{pmatrix}\sigma\left(X_1^\top\theta\right) \\ \cdots \\ \sigma\left(X_n^\top\theta\right)\end{pmatrix}$$

Найдем гессиан

$$\nabla^2 \ell_Y(\theta) = \nabla \sum_{i=1}^{n} X_i (Y_i - \sigma(X_i^\top \theta)) = -\sum_{i=1}^{n} X_i \sigma(X_i^\top \theta)(1 - \sigma(X_i^\top \theta)) X_i^\top =$$
$$= X^T V(\theta) X$$

$V(\theta) = \operatorname{diag}\left(\left[\sigma(X_i^\top \theta), 1 - \sigma(X_i^\top \theta)\right]_{i=1}^{n}\right)$

$I_Y(\theta) = \left(-\mathbb{E}_\theta \frac{\partial^2 \ell_Y(\theta)}{\partial \theta_j \partial \theta_k}\right)_{j=1,k=1}^{d,d} = -\nabla^2 \ell_Y(\theta)$, так как $\nabla^2 \ell_Y(\theta)$ не зависит от Y, то есть константа

Доверительный интервал тогда будет иметь вид

$$\boxed{\theta_j \in \left(\hat{\theta}_j - Z_{1-\frac{\alpha}{2}} \sqrt{I_Y^{-1}(\theta)_{jj}}, \hat{\theta}_j + Z_{1-\frac{\alpha}{2}} \sqrt{I_Y^{-1}(\theta)_{jj}}\right)}$$

II - 8.2 Пуассоновская регрессия

$X_1, ..., X_n \in \mathbb{R}^d, Y_1, ..., Y_n \in \mathbb{N}_0$

$\mu_\theta(x) = e^{x^\top \theta}, Y_i \sim \operatorname{Pois}(\mu_\theta(X_i))$

Распишем функцию правдоподобия

$\mathcal{L}_Y(\theta) = \prod_{i=1}^{n} \frac{\mu_\theta(X_i)^{Y_i}}{Y_i} e^{-\mu_\theta(X_i)} = \prod_{i=1}^{n} \frac{e^{X_i^\top \theta Y_i}}{Y_i!} \exp\left(e^{-X_i^\top \theta}\right)$

$\ell_Y(\theta) = \sum_{i=1}^{n} \left(X_i^\top \theta Y_i - \ln(Y_{i!}) - e^{X_i^\top \theta}\right)$

$\nabla \ell_Y(\theta) = \sum_{i=1}^{n} Y_i X_i - e^{X_i^\top \theta} X_i = X^T (Y - E(\theta))$

$E(\theta) = \begin{pmatrix} e^{X_1^\top \theta} \\ \vdots \\ e^{X_n^\top \theta} \end{pmatrix}$

$\nabla^2 \ell_Y(\theta) = -\sum_{i=1}^{n} X_i e^{X_i^\top \theta} X_i = -X^T \operatorname{diag}(E(\theta)) X$

$I_Y(\theta) = -\nabla^2 \ell_Y(\theta)$, так как $\nabla^2 \ell_Y(\theta)$ не зависит от Y, то есть – константа

II - 8.3 Линейная регрессия

$\theta_j \in \left(\hat{\theta}_j \pm Z_{1-\frac{\alpha}{2}} \sqrt{I_Y^{-1}(\theta)_{jj}}\right)$ – асимптотический доверительный интервал

$\theta_j \in \left(\hat{\theta}_j \pm t_{n-d,1-\frac{\alpha}{2}} \sigma \sqrt{X^\top X_{jj}^{-1}}\right)$ – точный доверительный интервал

$I_Y(\theta) = \frac{1}{\sigma^2} X^T X$

II - 9 Калибровка. Семинар

II - 9.1 Калибровка. Определение

Решаем задачу бинарной классификации.

$X_1, \ldots, X_n \in \mathbb{R}^d$– признаки

$Y_1, \ldots, Y_n \in \{0, 1\}$ – классы

Пусть $\hat{p}$ – модель, которая выдает вероятность $\hat{p}(x)$ того, что объект x принадлежит классу 1

Опр

Модель $\hat{p}$ идеально скалиброванa, если выполнено $\forall \rho$, предсказанного моделью $P(Y = 1 \,|\, \hat{p}(x) = \rho) = \rho$

Замечание

Смысл: если модель дает объекту x вероятность $\frac{2}{3}$, то среди всех таких объектов x, что $\hat{p}(x) = \frac{2}{3}$ будут действительно из класса 1

II - 9.2 Калибровочная кривая

Замечание

В реальности модель разным x_i дает разные вероятности, поэтому есть некоторые проблемы с оценкой вероятности из определения(вероятность не будет ровно ρ)

Решение:

Рассмотрим разбиение $[0, 1] = \bigsqcup\limits_{j=1}^{k} B_j$

$B_j = \left[\xi_{j-1}, \xi_j\right), B_k = [\xi_{k-1}, 1],$

$0 < \xi_0 < \xi_1 < \ldots < \xi_n = 1$

$\hat{p}_j = \frac{1}{\# B_j} \sum\limits_{\hat{p}(x) \in B_j} \hat{p}(x)$ (выдача модели)

$p_j = \frac{1}{\# B_j} \# \left\{\hat{p}(x) \in B_j \,\middle|\, y = 1\right\}$ (реальные метки)

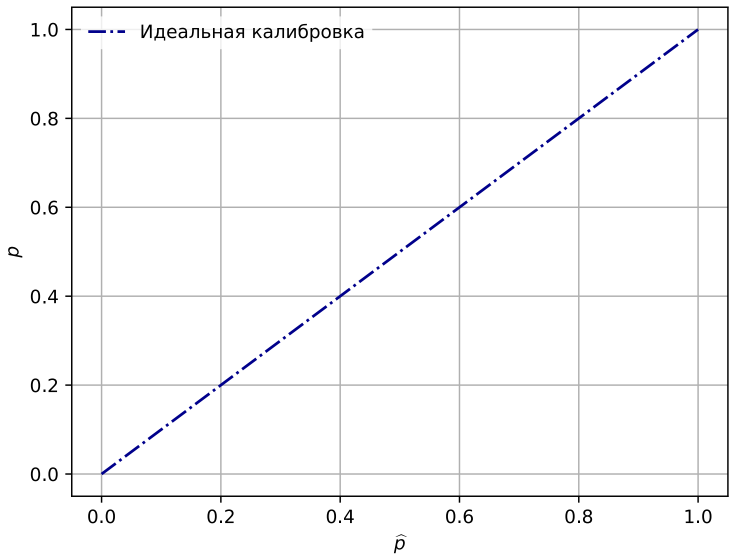


График 26: Модель с идеальной калибровкой

<u>Пример:</u>

$\boxed{i}$

$1, 2, 3, 4, 5, 6, 7, 8, 9, 10$

$\boxed{\hat{p}}$

$0.1, 0.15, 0.2, 0.4, 0.6, 0.8, 0.85, 0.9, 0.9, 0.95$

$\boxed{Y_i}$

$\underbrace{0,\ 0,\ 1}, \underbrace{0,\ 0,\ 1}, \underbrace{0,\ 1,\ 1,\ 1}$

$B_1 = [0, 0.4), B_2 = [0.4, 0.85), B_3 = [0.85, 1]$

$\hat{p}_1 = 0.15, \hat{p}_2 = 0.6, \hat{p}_3 = 0.9$

$p_1 = \frac{1}{3}, p_2 = \frac{1}{3}, p_3 = \frac{3}{4}$

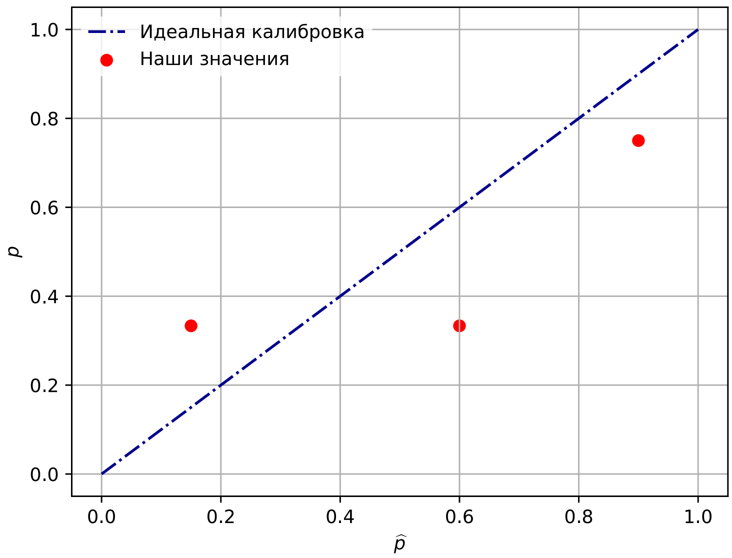


График 27: Калибровочная кривая №1

Пример:

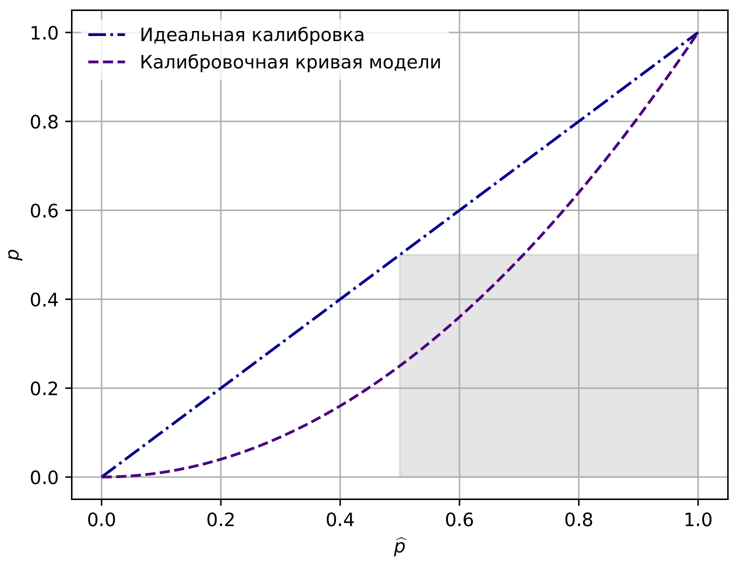


График 28: Калибровочная кривая №2

Делаем бинарную классификацию с порогом $\frac{1}{2}$

Пример:

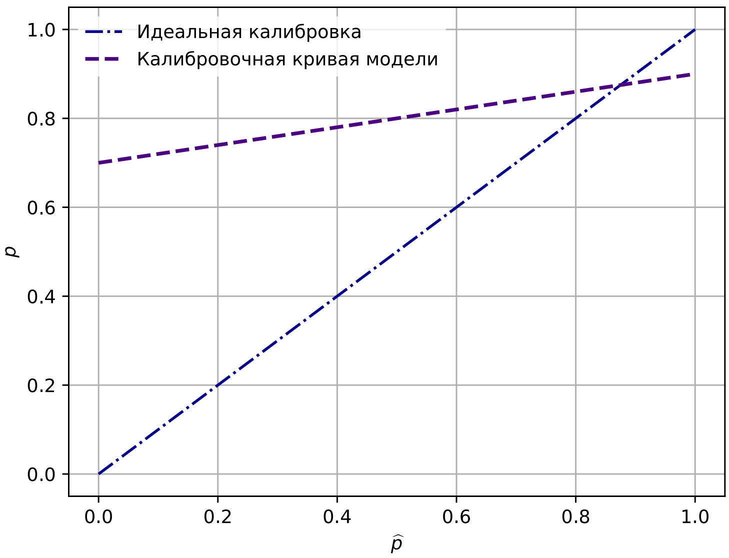


График 29: Калибровочная кривая №3

Слишком уверенный классификатор

<u>**Пример:**</u>

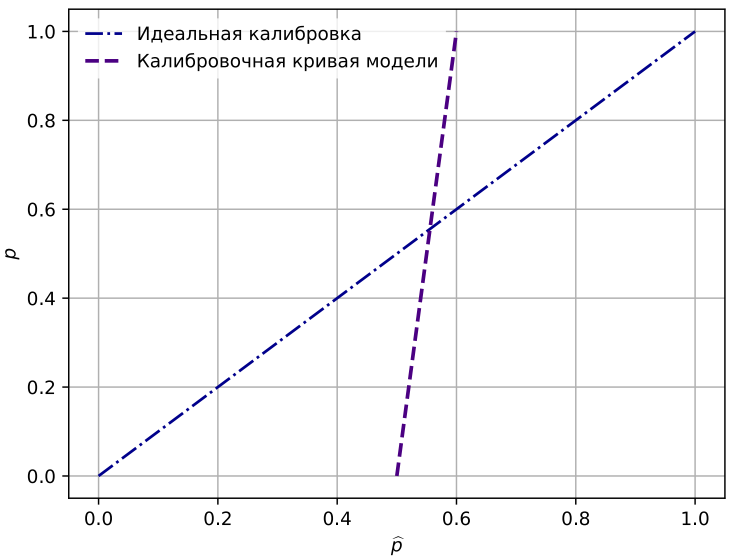


График 30: Калибровочная кривая №4

Слишком неуверенный классификатор

II - 9.3 Метрики калибровки

1) Expected calibration error (ECE)

$$\mathrm{ECE} = \sum_{j=1}^{k} \frac{\#B_j}{n} \left|p_j - \hat{p}_j\right|$$

2) Brier score (BS)

$$\mathrm{BS} = \sum_{i=1}^{n} \left(y_i - \hat{p}(x_i)\right)^2$$

$$\widetilde{\mathrm{BS}} = \sum_{i=1}^{n} y_i \ln \hat{p}(x_i) + (1 - y_i) \ln(1 - \hat{p}(x_i))$$

II - 9.4 Методы калибровки

1) Гистограммная калибровка
   - Разбиваем на бины
   - Меняем предсказания в конкретном бине B_j (то есть меняем $\hat{p}$ в бине)

$$\hat{p}_j = \underset{\rho}{\operatorname{argmin}} \sum_{\hat{p}(x) \in B_j} \left(y(x) - \rho\right)^2 = p_j$$

     - Проблема: использовали бины

2) Изотоническая регрессия
   - Аналогично гистограммной калибровке, только теперь подбираем трешхолды для бинов:

$$\min_{\substack{0 \le t_1 \le \ldots \le t_{k-1} \le 1 \\ 0 \le \rho_1 \le \ldots \le \rho_k \le 1}} \sum_{i=1}^{n} \left(y_i - g_{\rho_i, t}(x_i)\right)^2$$

   $t_1, \ldots, t_{k-1}$ – пороги для бинов

$$g_{\rho_i, t}(x_i) = \sum_{j=1}^{k} \rho_j \cdot \mathbb{I}\left\{\hat{p}(x_i) \in \left[t_{j-1}, t_j\right)\right\}$$

3) Калибровка Платта
   - Применим сигмоиды поверх нашей модели(то есть к выходам $\hat{p}$):

   $\rho(x_i) = \frac{1}{1 + e^{-a\hat{p}(x_i) - b}}$

   a, b получаем, минимизируя логлосс по выборке

II - 10 Информация Фишера

Пусть $X = (X_1, ..., X_n) \sim P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, где $\mathcal{P}$ – доминируемое семейство с плотностью $p_\theta(x)$

$\ell_X(\theta) = \ln \mathcal{L}_X(\theta) = \ln\left(\prod\limits_{i=1}^{n} p_\theta(X_i)\right) = \sum\limits_{i=1}^{n} \ln p_\theta(X_i)$

Опр

$U_X(\theta) = \frac{\partial \ell_X(\theta)}{\partial \theta}$ – вклад выборки X в оценку θ

Опр

$I_X(\theta) = \mathbb{D}_\theta U_X(\theta)$ – информация Фишера выборки X о параметре θ

Пример:

$X_1, ..., X_n \sim \mathrm{Bern}(\theta)$

$$\mathcal{L}_X(\theta) = \theta^{\sum\limits_{i=1}^{n} X_i} (1-\theta)^{n - \sum\limits_{i=1}^{n} X_i}$$

$$\ell_X(\theta) = \left(\sum_{i=1}^{n} X_i\right) \ln\theta + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-\theta)$$

$$U_X(\theta) = \frac{\partial \ell_X(\theta)}{\partial \theta} = \sum_{i=1}^{n} X_i \cdot \frac{1}{\theta} - \left(n - \sum_{i=1}^{n} X_i\right) \cdot \frac{1}{1-\theta} = \frac{\sum\limits_{i=1}^{n} X_i - n\theta}{\theta(1-\theta)}$$

$$I_X(\theta) = \mathbb{D}\left(\frac{\sum\limits_{i=1}^{n} X_i - n\theta}{\theta(1-\theta)}\right) = \frac{\sum\limits_{i=1}^{n} \mathbb{D}X_i}{\theta^2(1-\theta)^2} = \frac{n\theta(1-\theta)}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$$

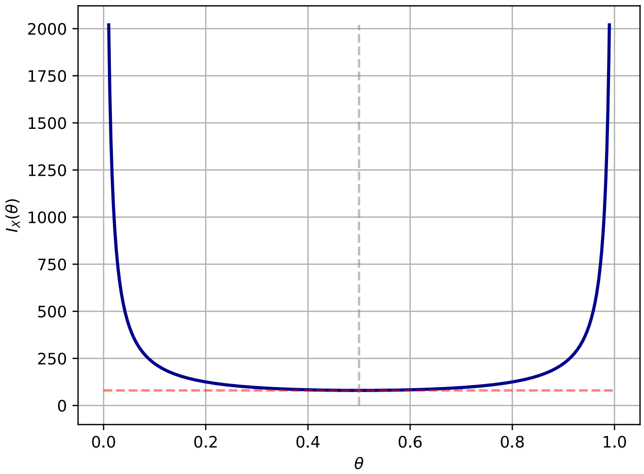


График 31: Информация Фишера

Теорема

В условиях регулярности E1-E4 выполняется следующее:

1) $\mathbb{E}_\theta U_X(\theta) = 0$

2) $I_X(\theta) = \mathbb{E}_\theta U_X^2(\theta)$

3) $I_X(\theta) = ni(\theta)$, где $i(\theta)$ – информация Фишера одного элемента

4) $I_X(\theta) = -\mathbb{E}\frac{\partial^2 \ell_X(\theta)}{\partial \theta^2}$

▲

1)

$$U_X(\theta) = \frac{\partial}{\partial \theta}\left(\sum_{i=1}^{n} p_\theta(X_i)\right) = \sum_{i-1}^{n} \frac{\ln p_\theta(X_i)}{\partial \theta} = \sum_{i=1}^{n} U_{X_i}(\theta)$$

$$\forall i = 1, \ldots, n\ \mathbb{E}_\theta = U_{X_i}(\theta) = \mathbb{E}\frac{\partial \ln p_\theta(X_i)}{\partial \theta} = \int\limits_{\mathbb{R}^d} p_\theta(x)\frac{\partial \ln p_\theta(x)}{\partial \theta}dx =$$

$$= \int_{\mathbb{R}^d} \frac{p_\theta(x)}{p_\theta(x)}\frac{\partial}{\partial \theta}p_\theta(x)dx = \int\limits_{\mathbb{R}^d} \frac{\partial}{\partial \theta}p_\theta(x)dx =$$

$$= \frac{\partial}{\partial \theta}\int\limits_{\mathbb{R}^d} p_\theta(x)dx = \frac{\partial}{\partial \theta}1 = 0$$

2)

$$I_X(\theta) = \mathbb{D}U_X(\theta) = \mathbb{E}U_X^2(\theta) - (\mathbb{E}U_X(\theta))^2 = \mathbb{E}U_X^2(\theta)$$

3)

$$I_X(\theta) = \mathbb{D}U_X(\theta) = \mathbb{D}\sum_{i=1}^{n} U_{X_i}(\theta) = \sum_{i=1}^{n} \mathbb{D}U_{X_i}(\theta) = n\mathbb{D}U_{X_1}(\theta) =$$
$$= nI_{X_1}(\theta) = ni(\theta)$$

4)

$$\frac{\partial^2 \ln p_\theta(x)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial p_\theta(x)}{\partial\theta}}{p_\theta(x)}\right) = \frac{\frac{\partial^2 p_\theta(x)}{\partial\theta^2}}{p_\theta(x)} - \frac{\left(\frac{\partial p_\theta(x)}{\partial\theta}\right)^2}{p_\theta^2(x)} \Rightarrow \frac{\partial^2 \ell_X(\theta)}{\partial\theta^2} =$$
$$= \sum_{i=1}^{n}\left(\frac{\frac{\partial^2 p_\theta(X_i)}{\partial\theta^2}}{p_\theta(X_i)} - \frac{\left(\frac{\partial p_\theta(X_i)}{\partial\theta}\right)^2}{p_\theta^2(X_i)}\right)$$

$$\mathbb{E}\frac{\partial^2 \ln p_\theta(x)}{\partial\theta^2} = \int_{\mathbb{R}^d}\left(\frac{\frac{\partial^2 p_\theta(x)}{\partial\theta^2}}{p_\theta(x)} - \frac{\left(\frac{\partial p_\theta(x)}{\partial\theta}\right)^2}{p_\theta^2(x)}\right)p_\theta(x)dx = \frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}^d} p_{\theta(x)}dx - I_{X_i}(\theta) =$$
$$= -I_{X_i}(\theta)$$

Тогда $\mathbb{E}\frac{\partial^2 \ln p_\theta(x)}{\partial\theta^2} = \mathbb{E}\frac{\partial^2 \sum_{i=1}^{n} \ln p_\theta(X_i)}{\partial\theta^2} = \sum_{i=1}^{n} \mathbb{E}\frac{\partial^2 \ln p_\theta(X_i)}{\partial\theta^2} = -\sum_{i=1}^{n} I_{X_i}(\theta) = -I_X(\theta)$

■

II - 11 Энтропия и дивергенция

II - 11.1 1. Дискретный случай

Рассмотрим случай с носителем $\{a_1, ..., a_k\}$ и вероятностями $\{p_1, ..., p_k\}$

Замечание

$0 \ln 0 = 0$

Опр

$H(P) = -\sum_{j=1}^{k} p_j \ln p_j$ – энтропия

$H(P, Q) = -\sum_{j=1}^{k} p_j \ln q_j$ – кросс-энтропия

$\mathrm{KL}(P, Q) = H(P, Q) - H(P) = \sum_{j=1}^{k} p_j \ln \frac{p_j}{q_j}$ – дивергенция Кульбака — Лейблера

Свойства энтропии:

1) $H(P) \geq 0$, при этом $H(P) = 0 \Leftrightarrow P$ – вырожденное

   $0 \leq p_j \leq 1 \Rightarrow \ln p_j \leq 0 \Rightarrow H(P) \geq 0$

   Равенство достигается тогда и только тогда, когда $\forall j \; p_j \ln p_j = 0$

   Так как p_j – это вероятности из категориального распределения, то $\exists j' p_{j'} = 1$, а остальные равны 0

   В обратную сторону очевидно: $\ln p_{j'} = 0$ иначе $p_j = 0$

2) $H(P) \leq \ln k$, равенство достигается при $p_1 = ... = p_k = \frac{1}{k}$

   Пусть ξ – случайная величина из распределения P

   Тогда

   $$H(P) = -\sum_{j=1}^{k} p_j \ln p_j = -\mathbb{E} \ln p(\xi) = \mathbb{E} \ln \tfrac{1}{p(\xi)} \leq \ln \mathbb{E} \tfrac{1}{p(\xi)} =$$

   $$= \ln \left(\sum_{j=1}^{k} p_j \frac{1}{p_j} \right) = \ln k$$

   Равенство достигается при $p_1 = ... = p_k = k$

II - 11.2 2. Общий случай

Пусть P, Q – распределения на одном и том же вероятностном пространстве $\xi \sim P$

Опр
$H(P) = -\mathbb{E} \ln p(\xi)$ – энтропия

$H(P, Q) = -\mathbb{E} \ln q(\xi)$ – кросс-энтропия

$\mathrm{KL}(P, Q) = H(P, Q) - H(P) = \mathbb{E} \ln \frac{p(\xi)}{q(\xi)}$ – дивергенция Кульбака–Лейблера

Свойства KL дивергенции:

1) $\mathrm{KL}(P, Q) \geq 0$ при этом $\mathrm{KL}(P, Q) = 0 \Leftrightarrow P = Q$

   $$-\mathrm{KL}(P, Q) = \mathbb{E} \ln \frac{q(\xi)}{p(\xi)} \overset{\text{Йенсен}}{\leq} \ln \mathbb{E} \frac{q(\xi)}{p(\xi)} = \ln \int_{\mathcal{X}} p(x) \frac{q(x)}{p(x)} dx = \ln 1 = 0$$

   Тогда $\mathrm{KL}(P, Q) \geq 0$

2) $\mathrm{KL}\ (P, Q) \neq \mathrm{KL}\ (Q, P)$ в общем случае

3) Пусть $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ – семейство дискретных распределений и $\widehat{P}_n$ – эмпирическая функция распределения P_θ

Тогда

$\mathrm{KL}\left(\hat{P}_n, P_\theta\right) = \mathbb{E}_{\hat{P}_n} = \mathbb{E}_{\hat{P}_n} \ln \frac{\hat{p}_n(X_i)}{p_\theta(X_i)} = \sum\limits_{i=1}^{n} \frac{1}{n} \ln \frac{\frac{1}{n}}{p_\theta(X_i)} =$

$= -\ln n - \frac{1}{n} \sum\limits_{i=1}^{n} \ln p_\theta(X_i) = -\frac{\ell_X(\theta)}{n} - \ln n$

Тогда $\mathrm{KL}\left(\hat{P}_n, P_\theta\right) \to \min\limits_\theta \Leftrightarrow \ell_X(\theta) \to \max\limits_\theta$

4) Пусть $\|\theta_1 - \theta_2\| \to 0$

Тогда $\mathrm{KL}\left(P_{\theta_1}, P_{\theta_2}\right) = \frac{1}{2}(\theta_1 - \theta_2)^\top i(\theta)(\theta_1 - \theta_2) + o\left(\|\theta_1 - \theta_2\|^2\right)$

Разложим KL в окрестности θ_1 по $\theta_2 - \theta_1$

$$\mathrm{KL}\left(P_{\theta_1}, P_{\theta_2}\right) = \mathrm{KL}\left(P_{\theta_1}, P_{\theta_1}\right) + (\theta_2 - \theta_1)\nabla_\delta \ \mathrm{KL}\left(P_{\theta_1}, P(\theta_1 + \delta)\right) \ |_{\delta=0} +$$
$$+(\theta_2 - \theta_1)^\top \nabla^2_\delta \ \mathrm{KL}\left(P_{\theta_1}, P_{\theta_1+\delta}\right)(\theta_2 - \theta_1) \ |_{\delta=0} + o\left(\|\theta_1 - \theta_2\|^2\right)$$

- $\mathrm{KL}\left(P_{\theta_1}, P_{\theta_2}\right) = 0$
- $\mathrm{KL}\left(P_{\theta_1}, P_{\theta_1+\delta}\right) = \nabla_{\theta_1} \ \mathrm{KL}\left(P_{\theta_1}, P_{\theta_1}\right) = 0$, так как

  $\forall f \in C^1(\mathcal{X})$ выполнено

  $$\frac{df(x+y)}{dy} \ |_{y=0} = [z = x + y] = \frac{df(z)}{dz} \ |_{z=x} \cdot \frac{dz}{dy} \ |_{y=0} = \frac{df(z)}{dz} \ |_{z=x}$$

- $$\nabla^2_\delta \ \mathrm{KL}\left(P_{\theta_1}, P_{\theta_1+\delta}\right) \ |_{\delta=0} = \left(\nabla^2_\delta \int\limits_{\mathcal{X}} p_{\theta_1}(x) \ln \frac{p_{\theta_1}(x)}{p_{\theta_1+\delta}(x)} dx\right) \ |_{\delta=0} =$$

  $$= \left(-\int\limits_{\mathcal{X}} p_{\theta_1}(x) \nabla^2_\delta \ln p_{\theta_1+\delta}(x) dx\right) \ |_{\delta=0} = -\mathbb{E}_{\theta_1} \nabla^2_{\theta_1} \ln p_{\theta_1}(x) = i(\theta_1)$$

  Тогда $\mathrm{KL}\left(P_{\theta_1}, P_{\theta_2}\right) = \frac{1}{2}(\theta_1 - \theta_2)^\top i(\theta)(\theta_1 - \theta_2) + o\left(\|\theta_1 - \theta_2\|^2\right)$

II - 12 Натуральный градиент

<u>Пример:</u>

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2), S = \sigma^2$

Нужно найти ОМП градиентным спуском

$\ell_X(\theta) = -\frac{n}{2} \ln(2\pi S) - \frac{1}{2S} \sum\limits_{i=1}^{n} (X_i - a)^2$

$\frac{\partial \ell_X(\theta)}{\partial a} = \frac{1}{S} \sum\limits_{i=1}^{n} (X_i - a) = \frac{1}{nS}\left(\overline{X} - a\right)$

$\frac{\partial \ell_X(\theta)}{\partial S} = -\frac{n}{2S} + \frac{1}{2S^2}\sum\limits_{i=1}^{n}(X_i - a)^2 = \frac{n}{2S^2}\left(-S + \sum\limits_{i=1}^{n}\frac{(X_i-a)^2}{n}\right)$

Шаг градиентного спуска

$$a_{t+1} = a_t + \eta_1 \frac{\overline{X} - a_t}{S_t}$$

$$S_{t+1} = S_t + \eta_2 \frac{\sum\limits_{i=1}^{n}\frac{(X_i-a_t)^2}{n} - S_t}{S_t^2}$$

Пусть $f : \mathbb{R}^d \to \mathbb{R}$, тогда $\nabla f = \left(\frac{\partial f}{\partial x_j}\right)_{j=1}^{d}$

Замечание
Градиент – направление наибольшего роста функции

Пусть $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ – доминируемое семейство распределений и $f(P_\theta)$ – некоторый функционал от распределения

Пример:

$\mathcal{P} = \left\{\mathcal{N}(a, \sigma^2) \,\middle|\, a \in \mathbb{R}, \sigma \in \mathbb{R}_+\right\}$

1) $\theta = (0, 1), \Delta\theta = (1, 0)$

2) $\theta = (0, 100^2), \Delta\theta = (1, 0)$

Схожесть в пространстве параметров не всегда отражает схожесть в пространстве распределений

Хотим получить новое определение градиента, которое отражает близость в пространстве распределений

$\nabla_N f(P_\theta) \sim \underset{\Delta\theta:\ \mathrm{KL}(P_\theta, P_{\theta+\Delta\theta}) \leq \varepsilon}{\arg\max} f(P_{\theta+\Delta\theta}) - f(P_\theta))$

$\begin{cases} f(P_{\theta+\Delta\theta}) - f(P_\theta) \to \max\limits_{\Delta\theta} \\ \mathrm{KL}(P_\theta, P_{\theta+\Delta\theta}) \leq \varepsilon \end{cases}$

$L(\theta, \Delta\theta, \lambda) = f(P_{\theta+\Delta\theta}) - f(P_\theta) - \lambda(\mathrm{KL}(P_\theta, P_{\theta+\Delta\theta}) - \varepsilon)$

$\frac{\partial L}{\partial \Delta\theta} = \nabla_\theta f(P_\theta) - \frac{\lambda}{2}(\nabla_{\Delta\theta}(\Delta\theta_i^\top(\theta)\Delta\theta)) = \nabla_\theta f(P_\theta) - \lambda i(\theta)\Delta\theta$

$\Delta\theta = \frac{1}{\lambda} i(\theta)^{-1} \nabla f(P_\theta)$

Опр
$\nabla_N f(P_\theta) = i(\theta)^{-1} \nabla f(P_\theta)$ – натуральный градиент

В примере тогда имеем $i(\theta) = \begin{pmatrix} \frac{1}{S} & 0 \\ 0 & \frac{1}{2S^2} \end{pmatrix}$

Тогда $i(\theta)^{-1} = \begin{pmatrix} S & 0 \\ 0 & 2S^2 \end{pmatrix}$

Тогда $\nabla_N \ell_X(\theta) = \begin{pmatrix} n\left(\overline{X} - a\right) \\ n\left(\widehat{S^2} - S\right) \end{pmatrix}$

Градиентный спуск записывается аналогично

Пример:

Логистическая регрессия

$\nabla \ell_X(\theta) = X^\top (Y - S(\theta)), S(\theta) = \left(\sigma\left(X_i^\top \theta\right)\right)_{i=1}^n$

$I_X(\theta) = X^\top V(\theta) X, V(\theta) = \operatorname{diag}\left(\left(\sigma\left(X_i^\top \theta\right)\left(1 - \sigma\left(X_i^\top \theta\right)\right)\right)_{i=1}^n\right)$

$\nabla_N \ell_X(\theta) = i(\theta)^{-1} \nabla \ell_X(\theta) = n\left(X^\top V(\theta) X\right)^{-1} X^\top (Y - S(\theta))$

$\theta_{t+1} = \theta_t - \eta \cdot \nabla_N \ell_X(\theta)$ – получился IRLS

II - 13 Свойства ОМП. Семинар

Теорема (о монотонном отношении правдоподобия)

$\forall \theta_0, \theta_1 \in \Theta$ в условиях выполненных L1, L2 имеем:

$\theta_0 \neq \theta_1 : P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_1)) \overset{n \to \infty}{\to} 1$

▲

$$\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_1) \Leftrightarrow \frac{1}{n} \ln \frac{\mathcal{L}_X(\theta_0)}{\mathcal{L}_X(\theta_1)} > 0$$

$$\frac{1}{n} \ln \frac{\mathcal{L}_X(\theta_0)}{\mathcal{L}_X(\theta_1)} = \frac{1}{n} \sum_{i=1}^n \ln \frac{P_{\theta_0}(X_i)}{P_{\theta_1}(X_i)} \overset{\text{п.н.}}{\to} \mathbb{E}_{\theta_0} \ln \ln \frac{P_{\theta_0}(X_i)}{P_{\theta_1}(X_i)} = \operatorname{KL}\left(P_{\theta_0}, P_{\theta_1}\right) > 0$$

Значит $P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_1)) \to P_{\theta_0}\left(\operatorname{KL}\ \left(P_{\theta_0}, P_{\theta_1}\right) > 0\right) = 1$

■

Следствие:

Если Θ конечная, то ОМП существует и является состоятельной оценкой

▲

$$\forall \theta_0 \neq \theta_1 P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_1)) \to 1$$

Возьмем пересечение конечного числа событий, вероятность каждого из которых $\to 1$, тогда и у всего пересечения вероятность $\to 1$

Тем самым с вероятностью $\to 1$ при больших n максимум достигается на θ_0

■

Теорема (о состоятельности ОМП)

В условиях выполненных L1-L5 имеем:

Уравнение правдоподобия $\frac{\partial \ell_X(\theta)}{\partial \theta} = 0$ имеет решение $\tilde{\theta}$ – состоятельная оценка

▲

Выберем интервал $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subseteq \Theta$ [L4]

Применим теорему о монотонном отношении правдоподобия к краям интервала

$$P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_0 - \varepsilon)) \to 1$$

$$P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_0 + \varepsilon)) \to 1$$

Тогда выполняется и для пересечения:

$P_{\theta_0}(\mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_0 - \varepsilon), \mathcal{L}_X(\theta_0) > \mathcal{L}_X(\theta_0 + \varepsilon)) \to 1$ [L5]

На интервале $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ возьмем $\tilde{\theta}$ – ближайший к θ_0

Это верно $\forall \varepsilon$ $P_{\theta_0}\left(\left|\tilde{\theta} - \theta_0\right| > \varepsilon\right) \to 0 \Rightarrow \tilde{\theta}$ – состоятельная оценка

■

Следствие:

Пусть $\forall n \ \forall x_1, ..., x_n$ – уравнение правдоподобия имеет единственное решение $\tilde{\theta}$. Тогда $\tilde{\theta}$ – состельная оценка θ и $P_{\theta_0}\left(\tilde{\theta} = \text{ОМП}\right) \to 1$ [L1-L5]

▲

Состоятельность следует из теоремы выше.

Вероятность того, что $\tilde{\theta}$ максимизирует $\mathcal{L}_X(\theta)$ стремится к 1(что есть максимум в интервале), значит $P_\theta\left(\tilde{\theta} = \text{ОМП}\right) = 1$

■

Теорема (об асимптотической нормальности) [L1] - [L9]

1) Пусть $\hat{\theta}$ - последовательность решений уравнения правдоподобия, т.ч. $\hat{\theta}$ - состоятельная оценка ($\exists$ по теореме о состоятельности оценки).

   Тогда $\hat{\theta}$ - а.н.о θ с асимптотической дисперсией $\frac{1}{i(\theta)}$(одномерный случай)

2) Пусть $\hat{\theta}$ – асимптотически нормальная оценка с непрерывной асимптотической дисперсией $\sigma^2(\theta)$, тогда $\sigma^2(\theta) \geq \frac{1}{i(\theta)} \forall \theta \in \Theta$

▲

Пусть $\tilde{\theta}$ – состоятельная последовательность решений уравнения правдоподобия. Рассмотрим вклад $U_X\left(\tilde{\theta}\right) = \frac{\partial \ell_X(\theta)}{\partial \theta} \mid_{\theta=\tilde{\theta}}$

Разложим вклад по Тейлору с остаточным членом в форме Лагранжа

$U_X\left(\tilde{\theta}\right) = U_X(\theta) + U_X'(\theta)\left(\tilde{\theta} - \theta\right) + \frac{1}{2}U_X''(\theta^*)\left(\tilde{\theta} - \theta\right)^2$

θ^* лежит на отрезке $\left[\tilde{\theta}, \theta\right]$

$\tilde{\theta}$ – состоятельная оценка θ и $\tilde{\theta} \xrightarrow{P_\theta} \theta \Rightarrow \theta^* \xrightarrow{P_\theta} \theta$

$\tilde{\theta}$ – решение уравнения правдоподобия

$U_X\left(\tilde{\theta}\right) = 0$

$$\sqrt{n}\left(\tilde{\theta} - \theta\right) = \sqrt{n}\frac{-U_X(\theta)\frac{1}{n}}{\frac{1}{n}U_X'(\theta) + \frac{1}{2}U_X''(\theta^*)\left(\tilde{\theta} - \theta\right)\frac{1}{n}}$$

1) $$\sqrt{n}\frac{1}{n}U_X(\theta) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U_{X_i}(\theta)\right) - \mathbb{E}_\theta U_{X_1}(\theta) \xrightarrow{\text{ЦПТ}} \mathcal{N}(0, \mathbb{D}U_X(\theta))$$

2) $$\frac{1}{n}U_X'(\theta) = \frac{1}{n}\sum_{i=1}^n U_{X_i}'(\theta) \xrightarrow{P_\theta} \mathbb{E}_\theta U_{X_i}'(\theta) = \mathbb{E}_\theta \frac{\partial^2 \ell_X(\theta)}{\partial \theta^2} = -i(\theta)$$

3) $$\left|\frac{1}{n}U_{X(\theta^*)}''\right| \leq \frac{1}{n}\sum_{i=1}^n \left|U_{X_i}''(\theta^*)\right| = \frac{1}{n}\sum_{i=1}^n \left|\frac{\partial^3 \ln p_{\theta^*}(X_i)}{\partial \theta^3}\right| \leq [\text{L9}] \leq$$
$$\leq \frac{1}{n}\sum_{i=1}^n H(X_i) \xrightarrow{\text{п.н.}} \mathbb{E}_\theta H(X_1) < +\infty \quad [\text{L9}]$$

Условие, что $\theta^* \in (\theta - c, \theta + c)$ будет выполнено с вероятностью $\to 1$, $\tilde{\theta} \xrightarrow{P_\theta} \theta$

$\left|\frac{1}{n}U_{X(\theta^*)}''\right|\left|\tilde{\theta} - \theta\right| \xrightarrow{P_\theta} 0$

4) Числитель $\to \mathcal{N}(0, i(\theta))$

Знаменатель $\xrightarrow{P_\theta} -i(\theta)$

По лемме Слуцкого

$$\sqrt{n}\left(\tilde{\theta} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}\left(0, \frac{1}{i(\theta)}\right)$$

■

Замечание
Если $\tilde{\theta}$ – единственное решение, то оно асимптотически нормальная оценка

Следствие:

ОМП – наилучшая оценка в условиях L1-L9 в асимптотическом подходе

Замечание
Если L1-L9 не выполнены, то могут быть и более хорошие оценки

Пример:

$X_1, ..., X_n \sim U(0, \theta)$

ОМП $= X_{(n)}$

$n\left(\theta - X_{(n)}\right) \xrightarrow{d_\theta} \mathrm{Exp}(1)$

$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}(...)$

II - 14 Информация Фишера. Семинар

II - 14.1 Многомерная информация Фишера

Опр
$U_X(\theta) = \frac{\partial \ell_X(\theta)}{\partial \theta}$ – вклад

Опр
$I_X(\theta) = \mathbb{D}_\theta U_X(\theta)$ – информация Фишера
$I_X(\theta) = \left(I_{jk}(\theta)\right)_{j,k=1}^{d,d}$, где $I_{jk}(\theta) = \mathbb{E}_\theta \frac{\partial \ell_X(\theta)}{\partial \theta_j} \cdot \frac{\partial \ell_X(\theta)}{\partial \theta_k} = -\mathbb{E}_\theta \frac{\partial^2 \ell_X(\theta)}{\partial \theta_j \partial \theta_k}$

Утв:

При справедливости L1-L9 ОМП $\hat{\theta}$ – асимптотически нормальная оценка с асимптотической матрицей ковариаций $I_{X_1}(\theta)^{-1}$

При этом если есть θ^* – асимптотически нормальная оценка с асимптотической матрицей ковариаций $\Sigma(\theta)$, такой что $\Sigma(\theta)$ непрерывна, То

$\Sigma(\theta) - I_{X_1}(\theta)^{-1}$ – неотрицательно определена

Задача

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$

$\mathcal{L}_X(\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i)$

$\ell_X(\theta) = \sum\limits_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i-a)^2}{2\sigma^2}\right)\right)$

$\frac{\partial \ell_X(\theta)}{\partial a} = \frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} 2(X_i - a)$

$\frac{\partial^2 \ell_X(\theta)}{\partial a^2} = -\frac{n}{\sigma^2}$

$\frac{\partial \ell_X(\theta)}{\partial \sigma^2} = -\frac{\sum\limits_{i=1}^{n} (X_i-a)^2}{\sigma^4} - \frac{n}{2\sigma^2}$

$\frac{\partial^2 \ell_X(\theta)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} + \frac{2\sum\limits_{i=1}^{n} (X_i-a)^2}{\sigma^6}$

$\frac{\partial^2 \ell_X(\theta)}{\partial \sigma^2 \partial a} = -\frac{2}{\sigma^4} \sum\limits_{i=1}^{n} (a - X_i)$

$\tilde{\theta} = \begin{pmatrix} \overline{X} \\ S^2 \end{pmatrix}$ – ОМП

$I_{X_1}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$

$I_{X_1}(\theta)^{-1} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$

<u>Задача</u>

$X_1, ..., X_n \sim \mathcal{N}(a, \sigma^2)$. Найти ОМП для $\frac{a}{\sigma}$ и ее асимптотическую дисперсию

Рассмотрим функцию $\tau(x, y) = \frac{x}{\sqrt{y}}$

$\frac{\partial \tau}{\partial x} = \frac{1}{\sqrt{y}}, \frac{\partial \tau}{\partial y} = -\frac{x}{2y^{\frac{3}{2}}}$

II - 14.2 Информация Фишера для обобщенных линейных моделей

1) Логистическая регрессия

   $\mu_\theta(x) = \sigma(\theta^\top x)$

   $I_X(\theta) = \left(-\mathbb{E}_\theta \frac{\partial^2 \ell_Y(\theta)}{\partial \theta_j \partial \theta_k}\right) = X^T V(\theta) X$

2) Пуассоновская регрессия

   $X_1, ..., X_n \in \mathbb{R}^d$ – признаки

   $Y_1, ..., Y_n \in \mathbb{N}$

   $\mu_\theta(x) = e^{\theta^\top x}, Y \sim \text{Pois}(\mu_\theta(X))$

   $I_X(\theta) = X^\top \operatorname{diag}(\mathbb{E}(\theta)) X$

II - 15 Доказательства теорем

II - 15.1 Критерий χ^2

Теорема (критерий хи-квадрат)

Пусть $X = (X_1, ..., X_n)$ – выборка из неизвестного распределения P и $H_0 : P = P_0$. Пусть также $\mathcal{X} = \bigsqcup\limits_{j=1}^{k} B_j$ – разбиение выборочного пространства на непересекающиеся бины. Обозначим $m_j = \#\{X_i \in B_j\} = \sum\limits_{i=1}^{n} \mathbb{I}\{X_i \in B_j\}$, а также $p_j^0 = P_0(X_i \in B_j)$.

Тогда $\chi(X) = \sum\limits_{j=1}^{k} \frac{\left(\mu_j - np_j^0\right)^2}{np_j^0} \xrightarrow{d_0} \chi^2_{k-1}$

▲

Пусть

$$Y_i = \begin{pmatrix} \mathbb{I}\{X_i \in B_1\} \\ \vdots \\ \mathbb{I}\{X_i \in B_k\} \end{pmatrix} \in \mathbb{R}^k$$

$$\mathbb{E}Y_i = \begin{pmatrix} p_1^0 \\ \vdots \\ p_k^0 \end{pmatrix} = p_0$$

$$\begin{aligned} \operatorname{cov}\left(\mathbb{I}\{X_1 \in B_j\}, \mathbb{I}\{X_1 \in B_l\}\right) &= \mathbb{E}\mathbb{I}\{X_1 \in B_j\} \cdot \mathbb{I}\{X_1 \in B_l\} - \\ &- \mathbb{E}\mathbb{I}\{X_1 \in B_j\} \cdot \mathbb{E}\mathbb{I}\{X_1 \in B_l\} = \\ &= p_j \delta_{jl} - p_j p_l \end{aligned}$$

δ_{jl} – символ Кронекера. Тогда получаем, что $\mathbb{D}Y_i = \operatorname{diag}(p_0) - p_0 p_0^\top$

Теперь воспользуемся ЦПТ для Y.

$$\sqrt{n}\left(\overline{Y} - p_0\right) \xrightarrow{d_0} \mathcal{N}\left(0, \operatorname{diag}(p_0) - p_0 p_0^\top\right)$$

Введем новую случайную величину, $\xi_n = \operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)\sqrt{n}\left(\overline{Y} - p_0\right)$. По теореме о наследовании сходимостей.

$$\xi_n \xrightarrow{d_0} \mathcal{N}\left(0, \operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)\left(\operatorname{diag}(p_0) - p_0 p_0^\top\right) \operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)\right)$$

$$B = \operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)\ \operatorname{diag}(p_0)\ \operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right) -$$

$$-\left(\operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)p_0\right)\left(\operatorname{diag}\left(\frac{1}{\sqrt{p_0}}\right)p_0\right)^\top =$$

$$= I - \sqrt{p_0}\left(\sqrt{p_0}\right)^\top$$

Возьмем ортогональную матрицу $V = \begin{pmatrix} \sqrt{p_1^0} & \dots & \sqrt{p_k^0} \\ \dots & \dots & \dots \\ \vdots & \vdots & \vdots \\ \dots & \dots & \dots \end{pmatrix} \in \mathbb{R}^{k\times k}$

По теореме о наследовании сходимости

$$V\xi_n \xrightarrow{d_0} \mathcal{N}\left(0, V\left(I_k - \sqrt{p_0}\sqrt{p_0}^\top\right)V^\top\right)$$

$$A = VV^\top - V\sqrt{p_0}\sqrt{p_0}^\top V^\top = I_k - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^\top = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix} = I'_k$$

По теореме о наследовании схолимости $\|V\xi_n\|^2 \sim \|\mathcal{N}(0, I'_k)\|^2 = \chi^2_{k-1}$

$$\|V\xi_n\|^2 = \|\xi_n\|^2 = \sum_{j=1}^{k}\left(\sqrt{n}\frac{1}{\sqrt{p_j^0}}\left(\overline{Y} - p_j^0\right)\right)^2 = \sum_{j=1}^{k} n\left(\frac{1}{p_j^0}\left(\frac{\mu_j}{n} - p_j^0\right)^2\right) =$$

$$= \sum_{j=1}^{k}\frac{\left(\mu_j - np_j^0\right)^2}{np_j^0} = \chi(X)$$

■

II - 15.2 Лемма Неймана-Пирсона

Теорема (Лемма Неймана-Пирсона)

Пусть X – выборка из неизвестного распределения $P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ – доминируемое семейство

Пусть также $\mathcal{L}_X(\theta)$ – функция правдоподобия, а $\Lambda_{\theta_0,\theta_1} = \frac{\mathcal{L}_X(\theta_1)}{\mathcal{L}_X(\theta_0)}$, а исследуемые гипотезы $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Тогда если $\exists c_\alpha : P_{\theta_0}\left(\Lambda_{\theta_0,\theta_1}(X) > c_\alpha\right) = \alpha$, то критерий

$S = \left\{\Lambda_{\theta_0, \theta_1}(x) > c_\alpha\right\}$ – РНМК среди критериев уровня значимости α

▲

S – критерий уровня значимости α, так как $P(I_S) = P_{\theta_0}\left(\Lambda_{\theta_0, \theta_1}(X) > c_\alpha\right) = \alpha$.

Покажем теперь, что он наиболее мощный. $\forall R$ – критерий уровня значимости α выполнено

$$P_{\theta_0}(X \in R) \leq \alpha = P_{\theta_0}(X \in S)$$

$$(\mathcal{L}_X(\theta_1) - c_\alpha \mathcal{L}_X(\theta_0))\mathbb{I}\{X \in R\} \leq$$
$$\leq (\mathcal{L}_X(\theta_1) - c_\alpha \mathcal{L}_X(\theta_0))\mathbb{I}\{X \in R\}\mathbb{I}\left\{\frac{\mathcal{L}_X(\theta_1)}{\mathcal{L}_X(\theta_0)} \geq c_\alpha\right\} \leq$$
$$\leq (\mathcal{L}_X(\theta_0) - c_\alpha \mathcal{L}_X(\theta_1))\mathbb{I}\{X \in S\}$$

Тогда

$$\int\limits_{\mathcal{X}} \left(\prod_{i=1}^n p_{\theta_1}(x_i) - c_\alpha \prod_{i=1}^n p_{\theta_0}(x_i)\right)\mathbb{I}\{x \in R\} dx_1 ... dx_n =$$
$$= P_{\theta_1}(X \in R) - c_\alpha P_{\theta_0}(X \in R) \leq P_{\theta_1}(X \in S) - c_\alpha P_{\theta_0}(X \in S)$$

Тогда разность мощностей

$$\mathfrak{P}_S(\theta_1) - \mathfrak{P}_R(\theta_1) = P_{\theta_1}(X \in S) - P_{\theta_1}(X \in R) \geq$$
$$\geq c_\alpha\left(P_{\theta_0}(X \in S) - P_{\theta_0}(X \in R)\right) \geq 0$$

Тогда $\forall R$ уровная значимости α $\mathfrak{P}_S(\theta) \geq \mathfrak{P}_R(\theta_1)$, значит S – наиболее мощный критерий уровня значимости α

■

<u>Утв</u>

В таком случае мощность будет не меньше ошибки первого рода.

$P(I_S) < \mathfrak{P}_S(\theta_1)$

▲

$S = \left\{\Lambda_{\theta_0, \theta_1}(x) > c_\alpha\right\}$

1) $c_\alpha \geq 1$.

   Тогда $\forall X \in S : \mathcal{L}_X(\theta_1) \geq \mathcal{L}_X(\theta_0)$

$$\int\limits_S \prod_{i=1}^{n} p_{\theta_1}(x_i)dx_1...dx_n = \underbrace{P_{\theta_1}(X \in S)}_{\mathfrak{P}_S(\theta_1)} \geq \underbrace{P_{\theta_0}(X \in S)}_{P(I_S)}$$

2) $c_\alpha < 1$

   Тогда $\forall X \in S : \mathcal{L}_X(\theta_1) < \mathcal{L}_X(\theta_0)$

   $$\int\limits_{\overline{S}} \prod_{i=1}^{n} p_{\theta_1}(x_i)dx_1...dx_n = 1 - P_{\theta_1}(X \in S) \leq 1 - P_{\theta_0}(X \in S)$$

   Значит, $\mathfrak{P}_S(\theta_1) = P_{\theta_1}(X \in S) > P_{\theta_0}(X \in S) = P(I_S)$

■