Математическая статистика, конспекты

Паненко Семён по конспектам Троешестовой Лидии

6 ноября 2022 г.

Содержание

1 Основы математической статистики

Предмет математической статистики: из экспериментальных данных понять природу явления. Пусть заданы следующие величины:

- $X_1, ..., X_n$ - некоторые числовые характеристики многократного эксперимента
- $\mathcal{X}$ - выборочное пространство, множество всевозможных значений чиселнны характеристик
- $\mathcal{B}(\mathcal{X})$ - σ-алгебра на X, нужна, чтобы ввести распределения на X

Определение 1.1. *Вероятностно-статистической моделью называется тройка $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P})$, удовлетворяющая написанным выше условиям*

Вероятностно-статистическая модель нужна, чтобы математически моделировать случайный эксперимент. Если нам нужна выборка размера n, то можно взять ВСМ $(\mathcal{X}^n, \mathcal{B}(X^n), \mathcal{P})$. Если нужно промоделировать бесконечную выборку, то стоит взять $(\mathcal{X}^\infty, \mathcal{B}(\mathcal{X}^\infty), \mathcal{P})$, в которой сигма-алгебра порождена всевозможными конечными произведениями борелевский множеств в каждом из сомножителей декартового произведения.

Определение 1.2. *Выборкой называется вектор независимых, одинаково распределённых случайных величин.*

1.1 Методы вывода

Существует два основных метода вывода:

1. Параметрический: неизвестное распределение может быть параметризовано некоторым параметром $\theta \in \Theta$, то есть
$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$
Наша задача в этом случае заключается в том, чтобы определить как можно более точно значение параметра θ

2. Непараметрический: неизвестные распределения не могут быть или не параметризуются естественным образом. Наша цель в таком случае определить функцию распределения или плотность неизвестного неизвестного распределения.

Также существуют два возможных способа вывода:

1. Частотный: Построение суждений о неизвестном распределении P строятся на предеположении, что все распределения из $\mathcal{P}$ равноправны.

2. Байесовский: Мы считаем, что на множестве всех распределений задано некоторое распределение, которое мы называем "априорным знанием"

1.2 Методы численного вычисления интегралов

1.2.1 Метод прямоугольников

Мы хотим вычислить $I = \int_a^b f(x)dx$. Разобъём интервал интегрирования:
$$a = x_0 < x_1 < ... < x_n = b$$
Тогда в качестве оценки интеграла рассмотрим величину:
$$\hat{I} = \sum_{i=1}^{n} f\left(\frac{x_{i-1} + x_i}{2}\right)(x_i - x_{i-1})$$
Например, в качестве разбиения можно рассмотреть равномерную сетку:
$$x_i = x_0 + id, \quad d = \frac{b-a}{n}$$
Тогда оценка интеграла запишется в виде:
$$\hat{I} = \sum_{i=1}^{n} \left(\frac{b-a}{n}\right) f\left(a + \left(\frac{i-1}{2}\right)\frac{b-a}{n}\right)$$

Теорема *Пусть f - дважды непрерывно дифференцируемая функция. Тогда при использовании равномерной сетки выполнена оценка:*

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

Доказательство: Разложим $f(x)$ по формуле Тейлора с остаточным членом:

$$f(x) = f(r_i) + f'(x - r_i) + \frac{1}{2} f''(z(x))(x - r_i)^2$$

Где $r_i = \frac{1}{2}(x_{i-1} + x_i)$. Заметим, что по формуле интегрирования по частям, выполнено равенство:

$$\int_{x_{i-1}}^{x_i} f_1^X, X_2, ..., X_n(r_i)(x - r_i)dx = 0$$

Тогда:

$$R_i = \int_{x_{i-1}}^{x_i} f(x) - f(r_i)dx = \frac{1}{2} \int_{x_{i-1}}^{x_i} f''(z(x))(x - r_i)^2 dx$$

Тогда из предположений теоремы:

$$|R_i| \leq \frac{M}{2} \frac{(x - r_i)^3}{3} \leq \frac{M(b-a)^3}{24n^2}$$

Таким образом, для достаточно "хороших"функций погрешность метода имеет асимптотику $\frac{1}{n^2}$

1.2.2 Метод Монте-Карло

Пытаемся вычислить интеграл $I = \int_a^b f(x)dx$. Рассмотрим также случайную величину $\xi \sim U[a, b]$. Заметим, что:

$$I = \int_a^b f(x)dx = (b-a) \left[\frac{1}{b-a} \int_a^b f(x)dx \right] = \frac{1}{b-a} \mathbb{E} f(\xi)$$

Рассмотрим выборку $x_1, ..., x_n$ из распределения $U[0, 1]$. В качестве оценки интеграла выберем:

$$\hat{I} = (b-a)\overline{f(X)}$$

По УЗБЧ эта оценка почти-наверное сходится к истинному значению интеграла. В предположении конечного $\mathbb{E}(f(\xi))^2$, то по ЦПТ:

$$\sqrt{n} \frac{I - \hat{I}}{(b-a)\sqrt{\mathbb{D} f(\xi)}} \xrightarrow{d} \eta \sim \mathcal{N}(0,\, 1)$$

По свойствам нормального распределения $P(|\eta| \leq 3) \approx 0$. Поэтому при большиъ n оценка имеет порядок малости $\frac{1}{\sqrt{n}}$ с большой вероятностью

1.2.3 Сравнение методов

1. Погрешности: Метод прямоугольников - $\frac{1}{n^2}$, метод Монте-Карло - $\frac{1}{\sqrt{n}}$
2. Ошибка метода Монте-Карло вероятностная, зависит от $\mathbb{D}f(\xi)$.
3. Для интегралов малой кратности лучше метод прямоугольников
4. Для интегралов большой кратности необходимо слишком много точек для сходимости метода, поэтому лучше использовать метод Монте-Карло

2 Точечные и интервальные оценки

Пусть задана вероятностно статистическая модель $(\mathcal{X}^n, \mathcal{B}(\mathcal{X}^\backslash), \mathcal{P})$, причём $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, то есть мы имеем дело с параметрическим подходом. Пусть также $X = (X_1, X_2, ..., X_n)$ - выборка из неизвестного распределения этого класса.

Определение 2.1. *Пусть (E, ρ) - измеримое пространство. Измеримое отображение $S : \mathcal{X} \to E$ называется статистикой. В случае $E = \Theta$ оно также называется оценкой.*

> Оценивать можно не только сам параметр, но и измеримые функции от него: $\tau(\theta)$

Примеры статистик

1. $\overline{X}, \overline{X^2}, \overline{X^k} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ - выборочный k-ый момент
2. $S^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ - выборочная дисперсия
3. Вариационный ряд: $(X_{(1)}, X_{(2)}, X_{(3)}, ...X_{(n)})$ - упорядоченная по неубыванию выборка.
4. $X_{(k)} - k$-ая порядковая статистика

2.1 Свойства оценок

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из неизвестного распределения $P \in \{P_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$.

Обозначим через $\mathbb{E}_\theta, \mathbb{D}_\theta, \xrightarrow{d_\theta}, P_\theta$ - п.н - соответственно математиматическо ожидание, дисперсия, сходимость по распределению для распределения P_θ

Определение 2.2. *Оценка $\hat{\theta}$ называется несмещённой оценкой параметра $\tau(\theta)$, если выполнено:*

$$\forall \theta \in \Theta : \quad \mathbb{E}_\theta \hat{\theta}(X_1, X_2, ..., X_n) = \tau(\theta)$$

Смысл этого определения заключается в том, что при многократном повторении эксперимента в среднем мы будем получать истинное значение параметра.

Примеры:

1. $\hat{\theta}_1 = \overline{X}$, $\hat{\theta}_1 = X_1$ - несмещённые оценки для $\tau(\theta) = \mathbb{E}_\theta(X_1)$
2. $\mathcal{P} = \{Bern(\theta), \theta \in [0, 1]\}$. Тогда $\overline{X}, X_1$ - несмещённые оценки параметра θ.
3. $\mathcal{P} = \{Exp(\theta), \theta > 0\}$. Тогда $\overline{X}$ - несмещённая оценка параметра $\frac{1}{\theta}$
4. Рассмотрим оценку $S^2 = \overline{X^2} - \overline{X}^2$. Посчитаем её математиматическое ожидание:

$$\begin{aligned}
\mathbb{E}_\theta S^2 &= \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) - \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \\
&= \mathbb{E}_\theta X_1^2 - \frac{1}{n^2} \left(n\mathbb{E}_\theta X_1^2 - (n^2 - n)\mathbb{E}_\theta(X_1 X_2) \right) \\
&= \mathbb{E}_\theta X_1^2 - \frac{1}{n}\mathbb{E}_\theta X_1^2 - \left(1 - \frac{1}{n}\right)(\mathbb{E}_\theta X_1)^2 \\
&= \frac{n-1}{n}\left(\mathbb{E}_\theta X_1^2 - (\mathbb{E}_\theta X_1)^2\right) \\
&= \frac{n-1}{n}\mathbb{D}_\theta X_1
\end{aligned} \tag{1}$$

Таким образом, S^2 - смещённая оценка дисперсии, а $\frac{n}{n-1}S^2$ - несмещённая.

асимптотические свойства

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из неизвестного распределения $P \in \{P_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$.

Определение 2.3. *Оценка (последовательность оценок) $\hat{\theta}$ называется состоятельной оценкой параметра $\tau(\theta)$, если выполнено:*

$$\forall \theta \in \Theta : \quad \hat{\theta}(X_1, X_2, ..., X_n) \xrightarrow{P} \tau(\theta)$$

Определение 2.4. *Оценка (последовательность оценок) $\hat{\theta}$ называется сильно состоятельной оценкой параметра $\tau(\theta)$, если выполнено:*

$$\forall \theta \in \Theta : \quad \hat{\theta}(X_1, X_2, ..., X_n) \xrightarrow{P_\theta \text{ a.e.}} \tau(\theta)$$

Определение 2.5. *Оценка (последовательность оценок) $\hat{\theta}$ называется асимптотически нормальной оценкой параметра $\tau(\theta)$, если выполнено:*

$$\forall \theta \in \Theta : \quad \sqrt{n}(\hat{\theta}(X_1, X_2, ..., X_n) - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta))$$

При этом матрица $\Sigma(\theta)$ называется асимптотической матрицей ковариаций. В случае $d = 1$ это число, которое иногда называют асимптотической дисперсией оценки

Смысл:

1. Состоятельность - с ростом количества наблюдений вероятность большого отклонения оценки от истинного значения параметра мала.

2. асимптотическая нормальность - даёт численную характеристику степени отклонения оценки от истинного значения параметра

3. Сильная состоятельность имеет смысл для данных потсупающих последовательно. С вероятностью 1, пока поступает всё больше и больше объектов выборки, мы приближаемся к истинному значению параметра.

Пример: Пусть $X_1, X_2, ..., X_n$ - выборка из распределения Лапласа со сдвигом θ. Его плотность:

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

1. Согласно закону больших чисел, $\overline{X}$ - сильно состоятельная оценка $\mathbb{E}_\theta X_1 = \theta$

2. Согласно центральной предельной теореме, $\overline{X}$ - а.н.о параметра θ с асмпт. дисперсией $\sigma^2 = \mathbb{D}_\theta X_1 = 2$

> Из сильной состоятельности или асимптотической нормальности следует обычная состоятельность. Другие следствия в общем случае неверны.

2.2 Наследование свойств оценок

Пусть есть оценка параметра $\psi(\theta)$ с какими-то свойствами. Наша цель: получить оценку параметра $\varphi(\psi(\theta))$ с такими же свойствами.

Теорема (О наследовании сходимостей) *Пусть $\xi_n \xrightarrow{a.e.} \xi$ - случайные вектора из $\mathbb{R}^m$. Пусть также $h : \mathbb{R}^m \to \mathbb{R}^k$ - непрерывная почти всюду относительно распределения ξ функция. Тогда $h(\xi_n) \xrightarrow{a.e.} h(\xi)$*

Теорема (О наследовании сходимостей) *Пусть $\xi_n \xrightarrow{P} \xi$ - случайные вектора из $\mathbb{R}^m$. Пусть также $h : \mathbb{R}^m \to \mathbb{R}^k$ - непрерывная почти всюду относительно распределения ξ функция. Тогда $h(\xi_n) \xrightarrow{P} h(\xi)$*

Теорема (О наследовании сходимостей) *Пусть $\xi_n \xrightarrow{d} \xi$ - случайные вектора из $\mathbb{R}^m$. Пусть также $h : \mathbb{R}^m \to \mathbb{R}^k$ - непрерывная **всюду** относительно распределения ξ функция. Тогда $h(\xi_n) \xrightarrow{d} h(\xi)$*

Пример: Пусть ξ_n - н.о.р.с.в, причём $\mathbb{E}\xi_n = a \neq 0$. Тогда по УЗБЧ:

$$\frac{\sum\limits_{i=1}^{n} X_i}{n} \xrightarrow{a.e.} a = \xi$$

Применяя теорему о наследовании сходимостей (первый вариант) к последовательности $\xi_n = \frac{S_n}{n}$, и функции $h(x) = \frac{1}{x}$, получаем, что:

$$\frac{n}{\sum\limits_{i=1}^{n} X_i} \xrightarrow{a.e.} h(a) = \frac{1}{a}$$

Теорема (Следствие из теоремы о наследовании сходимостей) *Пусть $\hat{\theta}$ - (сильно) состоятельная оценка параметра $\tau(\theta)$. Пусть также f непрерывна на Θ. Тогда $f(\hat{\theta})$ - (сильно) состоятельная оценка параметра $f(\tau(\theta))$.*

Теорема (Лемма Слуцкого) *Пусть $\{\xi_n\}, \{\eta_n\}$ - последовательности н.о.р.с.в., причём $\xi_n \xrightarrow{d} \xi, \eta_n \xrightarrow{d} c = const$. Тогда выполнено:*

1. $\xi_n + \eta_n \to \xi + c$
2. $\xi_n \times \eta_n \to c\xi$

> Лемма Слуцкого не следует из теоремы о наследовании сходимостей, так как из сходимости по распределению компонент вектора не следует сходимость самого вектора по распределению.

Теорема (О производной) *Пусть выполнены следующие условия:*

1. $\{\xi_n\}$ *- случайные векторы из* $\mathbb{R}^d$
2. $\xi_n \xrightarrow{d} \xi$
3. $h(x): \mathbb{R}^d \to \mathbb{R}^k$ *- непрерывно дифференцируемая в точке a функция*
4. $b_n \to 0$ *- числовая последовательность*

Тогда выполнена следующая сходимость:

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} h'(a)\xi$$

Пример: Пусть ξ_n - н.о.р.с.в., причём $\mathbb{E}\xi_n = a, \mathbb{D}\xi_n = \sigma^2$. Тогда по ЦПТ:

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

Применим теорему о производной к последовательности $\xi_n = \sqrt{n}\left(\frac{S_n}{n} - a\right)$, функции $h(x) = \frac{1}{x}$ и числовой последовательности $b_n = \frac{1}{\sqrt{n}}$. Тогда получим:

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

Получили, что $\frac{S_n}{n}$ - а.н.о. $\frac{1}{a}$ с асимптотической дисперсией $\frac{\sigma^2}{a^4}$

Теорема (Дельта-метод) *Пусть выполнены следующие условия:*

1. $\hat{\theta}$ - а.н.о параметра $\tau(\theta)$ с асмпт. матрицей ковариаций $\Sigma(\theta)$

2. f - непрерывно дифференцируемая функция на Θ

Тогда $f(\theta)$ - а.н.о. параметра $f(\tau(\theta))$ с асмпт. матрицей ковариаций

$$\Sigma'(\theta) = D(\theta)\Sigma(\theta)D(\theta)^T$$

Где $D(\theta) = \nabla f_{\tau(\theta)}$

Пример: Пусть $X_1, X_2, ..., X_n$ - выборка из экспоненциального распределения. По ЦПТ $\overline{X}$ - а.н.о. параметра $\frac{1}{\theta}$, причём её асимптотическая дисперсия: $\sigma^2 = \frac{1}{\theta^2}$. Применим δ-метод к функции $f(\theta) = \frac{1}{x}$. Получаем, что $\frac{1}{\overline{X}}$ - а.н.о. для θ с асимптотической дисперсией $(\sigma')^2 = \sigma^2 \times (-\frac{1}{x^2})_{x=\frac{1}{\theta}} = \theta^2$

2.3 Доверительные интервалы

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из неизвестного распределения $P \in \mathcal{P} = \{P_\theta, \theta \in \Theta\}$.

Определение 2.6. *В случае, если $\Theta \subseteq \mathbb{R}$, пара статистик $T_1(X), T_2(X)$ называются доверительным интервалом уровня доверия α, если*

$$\forall\theta \in \Theta: \quad P_\theta(\theta \in [T_1(X), T_2(X)]) \geq \alpha$$

Определение 2.7. *В случае, если $\Theta \subseteq \mathbb{R}^d$, область $S(X)$ называется доверительной областью уровня доверия α, если*

$$\forall\theta \in \Theta: \quad P_\theta(\theta \in S(X)) \geq \alpha$$

Определение 2.8. *В случае, если $\Theta \subseteq \mathbb{R}$, пара статистик $T_1(X), T_2(X)$ называются ассимптотическим доверительным интервалом уровня доверия α, если*

$$\forall\theta \in \Theta: \quad \varliminf_{n\to\infty} P_\theta(\theta \in [T_1(X), T_2(X)]) \geq \alpha$$

Зачем нужны доверительные интервалы? Если неизвестное распределение - абсолютно непрерывное, то вероятность того, что наша оценка в точности совпадёт со значением оцениваемого параметра заведомо равна нулю. Доверительные интервалы позволяют обойти эту проблему. Кроме того, стоит понимать, что неправильно говорить "θ лежит в доверительном интервале с вероятностью α потому что θ - некоторое число, правда, нам не известное. Корректнее говорить "θ накрывается доверительным интервалом с веротяностью α".

Обычно в качестве уровня доверия выбирается первая магическая стат. константа: $\alpha = 0.95$

Пример: Пусть $X_1, X_2, ..., X_n \sim Pois(\theta)$. Необходимо построить доверительный интервал уровня доверия α.
Для этого воспользуемся неравенством Чебышева:

$$P_\theta(|\sum_{i=1}^{n} X_i - n\mathbb{E}_\theta X_1| \geq n\varepsilon) \leq \frac{n\mathbb{D}X_i}{\varepsilon^2 n^2} = \frac{\theta}{n\varepsilon^2}$$

В таком случае также выполнено:

$$P_\theta(|\overline{X} - \theta| \leq \varepsilon) \leq 1 - \frac{\theta}{n\varepsilon^2}$$

Мы хотим оценить эту вероятность снизу числом α. Для этого следует взять:

$$\alpha = \sqrt{\frac{\theta}{n(1-\alpha)}}$$

Таким образом, с вероятностью α:

$$\theta - \sqrt{\frac{\theta}{n(1-\alpha)}} \leq \overline{X} \leq \theta + \sqrt{\frac{\theta}{n(1-\alpha)}}$$

Разрешая это неравенство относительно θ, получаем:

$$\theta \in [\sqrt{\overline{X} + \frac{1}{4n(1-\alpha)}} - \sqrt{\frac{1}{n\alpha(1-\alpha)}}, \sqrt{\overline{X} + \frac{1}{4n(1-\alpha)}} + \sqrt{\frac{1}{n\alpha(1-\alpha)}}]$$

Определение 2.9. *Пусть P - некоторое распределение на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Тогда p-квантилью этого распределения называется число:*

$$u_p = \inf\{x : \quad F(x) \geq p\}$$

Где $F(x)$ - функция распределения неизвестного распределения.

Теперь обсудим общие подходы к построению доверительных интервалов

2.4 Центральный интревал (Метод центральной функции)

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из неизвестного распределения $P \in \mathcal{P} = \{P_\theta, \theta \in \Theta\}$.

Определение 2.10. *Пусть $G(\cdot, \cdot)$ - какая-то функция, такая, что распределение случайной величины $G(X, \theta)$ не зависит от θ. Тогда G называется центральной функции для семейства распределений $\mathcal{P}$.*

Теорема (О центральной функции) *Пусть выбраны $\alpha_1, \alpha_2, g_1, g_2$ и центральная функция $G(X, \theta)$ такие, что:*

$\alpha_i \in (0, 1), \quad \alpha_1 - \alpha_2 = \alpha$

$g_1, g_2 - \alpha_1, \alpha_2$ соответственно квантили распределения $G(X, \theta)$. Тогда точной областью уровня доверия α будет:

$$S(X) = \{\theta \in \Theta : \quad G(X, \theta) \in [g_1, g_2]\}$$

Доказательство этой теоремы немедленно следует из пристального взгляда на равенство:

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

Пример: Пусть $X_1, X_2, ..., X_n$ - выборка из распределения $\mathcal{N}(\theta,\, \sigma^2)$, причём σ^2 - известно. Требуется построить центральный доверительный интервал для θ уровня доверия α.
Рассмторим $G(X, \theta) = \sqrt{n}\frac{\overline{X} - \theta}{\sigma}$. Нетрудно вычислить, что $G(X, \theta) \sim \mathcal{N}(0,\, 1)$.
Таким образом, получаем доверительный интервал:

$$\theta : \quad \left(\overline{X} - z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \overline{X} + z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

1. Если $\mathcal{P} = \{P_\theta : \quad \theta \in \Theta\}$, причём θ - параметр сдвига, то есть $p_\theta(x) = p_0(x - \theta)$, то центральной функцией будет являться $G(X, \theta) = \overline{X} - \theta$

2. Если функция распределения F_θ - непрерывна, то в качестве центральной статистики можно взять $G(X, \theta) = -\sum\limits_{i=1}^{n} \ln F_\theta(X_i) \sim \Gamma(1, n)$, так как $F_\theta(X) \sim U[0, 1]$.

2.5 Ассимптотический доверительный интервал Вальда

Пусть $\hat{\theta}$ - а.н.о. θ с ассимптотической дисперсией $\sigma^2(\theta)$. Тогда из определения:

$$\sqrt{n}\frac{\hat{\theta}-\theta}{\sigma(\theta)} \to \mathcal{N}(0,\, 1)$$

Таким образом,

$$P_\theta(\sqrt{n}\frac{|\hat{\theta}-\theta|}{\sigma} \le z_{\frac{1+\alpha}{2}}) \to \alpha$$

Исходя из этого, можно получить следующий ассимптотический доверительный интервал

Определение 2.11. *Пусть $\hat{\theta}$ - а.н.о. θ с ассимптотической дисперсией $\sigma^2(\theta)$, а $\hat{\sigma}$ - состоятельная оценка σ. Тогда ассимптотическим доверительным интервалом Вальда уровня доверия α называется следующий интервал:*

$$\hat{\theta} \pm \frac{z_{\frac{1+\alpha}{2}}\hat{\sigma}}{\sqrt{n}}$$

В качестве $\hat{\sigma}$ можно всегда взять $\sigma(\hat{\theta})$.

Примеры:

1. $X_1, X_2, ..., X_n \sim \mathcal{N}(\theta,\, \sigma^2)$. По ЦПТ: $\overline{X}$−а.н.о θ а ас. дисперсией σ^2. Получаем интервал:
$$\overline{X} \pm \frac{z_{\frac{1+\alpha}{2}}\sigma}{\sqrt{n}}$$

2. $X_1, X_2, ..., X_n \sim Pois(\theta)$. По ЦПТ: $\overline{X}$−а.н.о θ а ас. дисперсией θ. $\sqrt{\overline{X}}$ - состоятельная оценка (корня) дисперсии. Получаем интервал:
$$\overline{X} \pm \frac{z_{\frac{1+\alpha}{2}}\sqrt{\overline{X}}}{\sqrt{n}}$$

3. $X_1, X_2, ..., X_n \sim U(0, \theta)$. По ЦПТ: $2\overline{X}$−а.н.о θ а ас. дисперсией $\frac{\theta^2}{3}$Получаем интервал:
$$2\overline{X} \pm \frac{2z_{\frac{1+\alpha}{2}}\overline{X}}{\sqrt{3n}}$$

2.6 Точные доверительные интервалы в нормальной модели

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из распределения $\mathcal{N}(a, \sigma^2)$.

1. Доверительный интервал для a, если σ известно:
$$\overline{X} \pm \frac{z_{\frac{1+\alpha}{2}}\sigma}{\sqrt{n}} \qquad (a,\ \sigma \text{ - known})$$

2. Доверительный интервал для σ, если a известно
Заметим, что велчина $\xi = \frac{X_i - a}{\sigma}$ распределена как $\mathcal{N}(0, 1)$. Тогда, по определению:
$$\sum_{i=1}^{n}\left(\frac{X_i - a}{\sigma}\right)^2 \sim \xi_n^2$$
Выберем левую часть в качестве центральной функции. Получаем, что:
$$P(\xi^2_{n,\frac{1-\alpha}{2}} \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - a)^2 \le \xi^2_{n,\frac{1+\alpha}{2}}) = \alpha$$
Откуда получаем доверительный интервал:
$$\left(\sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - a)^2}{\chi^2_{n,\frac{1-\alpha}{2}}}}, \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - a)^2}{\chi^2_{n,\frac{1+\alpha}{2}}}}\right) \qquad (\sigma,\ a \text{ - known})$$

3. Доверительный интервал для a, если σ неизвестно

Теорема *Пусть $X_1, X_2, ..., X_n \sim \mathcal{N}(\sigma, \theta^2)$. Тогда:*

- $\overline{X}, S^2$ - *независимы*
- $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$
- $\sqrt{n-1}\frac{\overline{X}-\theta}{\sigma} \sim T_{n-1}$

Эту теорему мы докажем позже, а пока, заметим, что функция $G(X, \theta) = \sqrt{n-1}\frac{\overline{X}-\theta}{S} \sim T_{n-1}$ является центральной. Поэтому получаем такой доверительный интервал:
$$\overline{X} \pm \frac{S \times T_{n-1,\frac{1+\alpha}{2}}}{\sqrt{n-1}} \qquad (\sigma,\ a \text{ - known})$$

4. Доверительный интервал для σ, если a - неизвестно.
   По теореме выше $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$. Получаем доверительный интервал:

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

Теорема (Об ортонормальном разложении Гауссовского вектора) *Пусть $\xi \sim \mathcal{N}(a, \sigma^2 I_n)$. Пусть также $\mathbb{R}^n = L_1 \bigoplus L_2 \bigoplus ... \bigoplus L_k$ - разложение пространства в прямую сумму ортогональных подпротсранств. Определим ν_i - проекция ξ на L_i. Тогда выполнено:*

1. *$\nu_1, \nu_2, ..., \nu_k$ - независимы в совокупности*
2. *$\mathbb{E}\nu_j = proj_{L_j}(a)$*
3. *$\frac{1}{\sigma^2}\|\nu_i - \mathbb{E}\nu_i\|^2 \sim \xi^2_{dimL_i}$*

Доказательство: Рассмотрим $e_1, e_2, ..., e_n$ - ортонормированный базис в $\mathbb{R}^n$, согласованный с $L_1, ..., L_k$. То есть:

- $e_1, e_2, ..., e_{dimL_1}$ - базис в L_1.
- $e_{dimL_1+1}, e_{dimL_1+2}, ..., e_{dimL_1+dimL_2}$ - базис в L_2
- ...
- $e_{n-dimL_k+1}, e_{n-dimL_k+2}, ..., e_n$ - базис в L_k

Рассмотрим также матрицу перехода $B = (e_1, e_2, ..., e_n)$. Посчитаем теперь матрицу средних и ковариаций ветворов $r_i = \langle e_i, \xi\rangle$ - проекций ξ на подпространство, порождённое i базисным веектором. Ясно, что $r = (r_1, ..., r_n) = B^T\xi$. Тогда выполнено:

1. $\mathbb{E}r = \mathbb{E}B^T\xi = B^T\mathbb{E}\xi = B^Ta$
2. $\mathbb{D}r = B^T\mathbb{D}\xi B = \sigma^2 B^T B = \sigma^2$ (так как B - ортогональная, в силу ортонормальности выбранного базиса)

Как видно, матрица ковариаций получилась диагональной. Значит, компоненты вектора независимы, так как мы знаем, из теоревера, что если нормальные величины не скоррелированы, то они независимы. Используем полученное разложение, чтобы получить:

$$\nu_i = \sum_{k=dimL_1+dimL_2+...+dimL_{k-1}+1}^{dimL_1+dimL_2+...+dimL_k} \langle\xi_k, e_k\rangle e_k = \sum_{k=dimL_1+dimL_2+...+dimL_{k-1}+1}^{dimL_1+dimL_2+...+dimL_k} e_k r_k$$

Повесим на это математическое ожидание:

$$\mathbb{E}\nu_i = \sum_{k=dimL_1+dimL_2+...+dimL_{k-1}+1}^{dimL_1+dimL_2+...+dimL_k} \langle \mathbb{E}\xi_k, e_k\rangle e_k = \sum_{k=dimL_1+dimL_2+...+dimL_{k-1}+1}^{dimL_1+dimL_2+...+dimL_k} a_k e_k = proj_{L_i}(a)$$

Теперь посчитаем последнюю величину:

$$\frac{1}{\sigma^2}\|\nu_i - \mathbb{E}\nu_i\|^2 = \frac{1}{\sigma^2}\|\sum_k \langle \xi_k - a, e_k\rangle e_k\|^2 = /\text{из независимости и ортогональности базиса}/ =$$

$$= \frac{1}{\sigma^2}\sum_k (\xi_k - \mathbb{E}\xi_k)^2 \sim \chi^2_{dimL_i}$$

Что и требовалось доказать. Теперь докажем теорему из пункта 3, она получется применением только что доказанной теоремы к подпространствам:

$$L_1 = \langle (1, ..., 1)\rangle, L_2 = L_1^{\perp}$$

В качестве Гауссовского вектора возьмём $X = (X_1, X_2, ..., X_n)$. Вектор средних и матрица ковариаций: $A = (a, ..., a), \Sigma = \sigma^2 I_n$. Посчитаем проекции на L_1, L_2:

$$X^1 = \langle X, (1, ..., 1)\rangle = n(\overline{X}, ..., \overline{X})$$

$$X^2 = X - X^1 = \begin{bmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ X_3 - \overline{X} \\ ...X_n - \overline{X} \end{bmatrix}$$

Но тогда выборочное среднее - измеримая функция от X^1, а выборочная дисперсия - функция от X^2. Значит, они независимы, как функции от независимых случайных величин. Второй же пункт доказанной теоремы в точно утверждает, что $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

3 Методы поиска и сравнения оценок

Для мотивации последующего раздела рассмотрим следующую задачу: Пусть $X = X_1, X_2, ..., X_n$ - выборка из $Exp(\theta)$. Построим две оценки неизвестного параметра θ:

1. Метод моментов

$$\mathbb{E}_\theta X_1 = \overline{X} \Rightarrow \hat{\theta}_1 = \frac{1}{\overline{X}} \tag{2}$$

2. Обобщённый метод моментов с функцией $g(x) = I\{x > 1\}$

$$e^{-\theta} = P(X_1 > 1) = \mathbb{E}_\theta g(X_1) = \overline{I\{X > 1\}} \Rightarrow \hat{\theta}_2 = -\ln \overline{I\{X > 1\}} \quad (3)$$

Сравним их ассимптотические дисперсии. У $\hat{\theta}_1$ она равна (по ЦПТ) θ^2. Для подсчёта ас. дисперсии $\hat{\theta}_2$ заметим, что индикаторы распределены как Бернуллиевский случайные величины с параметром $p = e^{-\theta}$. Поэтому дисперсия $\overline{I\{X > 1\}}$ как оценки $e^{-\theta}$ равна в точности $e^{-\theta}(1 - e^{-\theta})$. Применим дельта-метод с функцией $\tau(x) = -\ln x$. Получим, что у $\hat{\theta}_2$ асимптотическая дисперсия будет равна $e^{-\theta}(1 - e^{-\theta}) \times e^2\theta = e^\theta - 1 > \theta^2$. Вывод: первая оценка "лучше"второй, так как ас. дисперсия лучше. А что такое лучше и почему мы сравниваем именно так?

3.1 Метод максимального правдоподобия

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из распределения $P \in \mathcal{P} = \{P_\theta \theta \in \Theta\}$.

Определение 3.1. *Семейство распределений $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ называется доминирующим, если*

1. *Все распределения P_θ дискретные*
2. *Все распределения P_θ абсолютно непрерывны.*

Пусть $\mathcal{P}$ - доминируеющее семейство.

Определение 3.2. *Функция $L_X(\theta)$:*

$$L_X(\theta) = \prod_{i=1}^{n} p_\theta(x_i)$$

Называется функцией правдоподобия для (реализации) выборки X и семейства распределений $\mathcal{P}$. Также рассматривают логарифмическую функцию правдоподобия:

$$l_X(\theta) = \ln L_X(\theta) = \sum_{i=1}^{n} \ln p_\theta(x_i)$$

По смыслу функция правдопобия - это то, насколько "вероятно"распределение с параметром θ произвело такую выборку

Определение 3.3. *Оценкой максимального правдоподобия параметра θ называется:*

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} L_X(\theta)$$

Если $\tau(\theta)$ - измеримая функция, то оценкой максимального правдоподобия для $\tau(\theta)$ называют величину:

$$\hat{\tau}(\theta) = \tau(\hat{\theta})$$

Это было сказано Давидом уже сильно после лекции и семинара. В контексте этого замечания следующая теорема кажется странной. Автор с этой странностью бороться не умеет, а ввязываться в ещё один длинный диалог об определениях и сигма-алгебрах не хочет. Поймите это как-нибудь для себя сами.

Теорема (О независимости от параметризации) *Пусть $\hat{\theta}$ - ОМП для θ. Пусть также $\psi : \Theta \to \Psi$ - биекция. Тогда $\tau(\hat{\theta})$ - ОМП для $\tau(\theta)$.*

Теорема (О свойствах ОМП) *Пусть $\forall n \forall X_1, X_2, ..., X_n$ уравнение $\frac{\partial}{\partial \theta} l_x(\theta) = 0$ имеет единственое решение. Тогда:*

1. *При выполнении условий регулярности $[L1 - L5]$ ОМП является состоятельной оценкой θ.*

2. *При выполнении условий регулярности $[L1 - L9]$ ОМП является ас. нормальной оценкой θ с ас. дисперсией $i^{-1}(\theta)$, где $i(\theta)$ - информационная матрица Фишера:*

$$i(\theta)_{\alpha,\beta} = \mathbb{E}_\theta \left[\frac{\partial l_{X_1}(\theta)}{\partial \theta_\alpha} \times \frac{\partial l_{X_1}(\theta)}{\partial \theta_\beta} \right]$$

Пример: $X = (X_1, X_2, ..., X_n) \sim Exp(\theta)$. Найти ОМП для $\theta, \frac{1}{\theta}$ и их ас. дисперсии. Запишем логарифм правдоподобия:

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

Для нахождения максимума найдём ноль производной:

$$\frac{\partial}{\partial \theta} l_X(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} X_i$$

Получаем, что ОМП для θ: $\hat{\theta} = \frac{n}{\sum\limits_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$. Для определения ас. дисперсии можно посчитать информационную матрицу Фишера:

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

Тогда ассимптотическая дисперсия равна $i^{-1}(\theta) = \theta^2$. Теперь, по свойству независимости от параметризации, получаем, Что $\overline{X}$ - ОМП для $\frac{1}{\theta}$. Ассимптотическая дисперсия в этом случае будет равна $\frac{1}{\theta^2}$.
Ещё один пример: Пусть $X = (X_1, X_2, ..., X_n) \sim Bern(\theta)$. Найти ОМП для $\theta, \ln \frac{\theta}{1-\theta}$. Запишем функцию правдоподобия:

$$L_X(\theta) = \prod_{i=1}^{n} p_\theta(x_i) = \theta^{\sum\limits_{i=1}^{n} X_i} (1-\theta)^{n - \sum\limits_{i=1}^{n} X_i}$$

Логарифмическая функция правдоподобия, стало быть:

$$l_X(\theta) = n \left[\overline{X} \ln \theta + (1 - \overline{X}) \ln(1-\theta) \right]$$

Приравниваем производную к нулю:

$$0 = \frac{\partial}{\partial \theta} l_X(\theta) = n \left[\frac{\overline{X}}{\theta} - \frac{(1 - \overline{X})}{1-\theta} \right]$$

Откуда получаем, что $\theta = \overline{X}$ - ОМП для θ. По свойству независимости от параметризации, $\ln \frac{\overline{X}}{1-\overline{X}}$ - ОМП для $\ln \frac{\theta}{1-\theta}$.
И последний пример: $X = (X_1, X_2, ..., X_n) \sim U[0, \theta]$, найти ОМП для θ. Запишем функцию правдоподобия:

$$L_X(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\{X_i \geq 0\} I\{X_i \leq \theta\} = \frac{I\{X_{(1)} \geq 0\}}{\theta^n} I\{X_{(n)} \leq \theta\}$$

Полученная функция 0 при $\theta < X_{(n)}$, а затем монотонно убывает. Значит, $\hat{\theta} = X_{(n)}$ - оценка максимального правдоподобия.

Я не знаю, зачем нужен следующий блок, потому что в нём мы попытаемся что-что сделать с Гамма-распределением, но у нас ничего не получится.

Задача: γ-котики. Есть две параллельные прямые, расположенные на расстоянии 1 друг от друга. На одной из прямых в точке θ стоит источник, который равновероятно по углам запускает частицу в сторону второй прямой. На второй прямой установлен датчик, который фиксирует положение прибывшей частицы. Требуется оценить параметр θ.
Найдём функцию распределения координаты частицы. Заметим, что плотность очевидно симметрична относительно $x = \theta$, поэтому при $x > \theta$:

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

Тогда получаем плотность:

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

Определение 3.4. *Полученное распределение называется распределением Коши со сдвигом θ*

θ является параметром сдвига для данного распределения.

1. Метод моментов неприменим, так как у полученного распределения нет моментов.

2. Попробуем вычислить ОМП:

$$l_X(\theta) = -n\ln\pi - \sum_{i=1}^{n}\ln\left(1 + (x_i - \theta)^2\right)$$

   Уравнение правдопобия:

$$\frac{\partial}{\partial\theta}l_X(\theta) = \sum_{i=1}^{n}\frac{2(x_i - \theta)}{1 + (x_i - \theta)^2}$$

   Вообще говоря, у этого уравнения может быть много решений

3. Попробуем взять $\hat{\theta} = \overline{X}$. Найдём её распределение:

$$\varphi_{\overline{X}}(\theta) = \mathbb{E}_\theta \exp\left(\frac{it}{n}\sum_{i=1}^{n}x_i\right) = \mathbb{E}_\theta\prod_{i=1}^{n}\exp\left(i\left[\frac{t}{n}\right]x_i\right) = \prod_{i=1}^{n}\mathbb{E}_\theta\exp\left(i\left[\frac{t}{n}\right]x_i\right) =$$

$$= \left[\varphi_{X_1}\left(\frac{t}{n}\right)\right]^n = e^{-|t|}$$

   А как мы знаем из ТВ, у распределения Коши такая же хар. функция. Поэтому $\overline{X} \sim Cauchy(\theta)$, и никакой сходимости нет, и всё плохо.

3.2 Выборочные квантили

Определение 3.5. *α-квантилью одномерного распределения с функцией распределения $F(x)$ называется величина:*

$$u_\alpha = \min\{x: \; F(x) \geq \alpha\}$$

Определение 3.6. *Выборочной α-квантилью выборки $X = (X_1, X_2, ..., X_n)$ называется величина:*

$$\hat{u}_\alpha = X_{(\lceil n\alpha \rceil)}$$

А вот выборочной медианой называют следующую вещь:

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

Теорема (О свойствах выборочной квантили) *Пусть выполнены следующие условия:*

1. *$X = (X_1, X_2, ..., X_n)$ - выборка из распределения с плотностью $f(X)$*
2. *$p \in (0, 1): \; f$ - непрерывно-дифференцируема в окрестности u_p*
3. *$p \in (0, 1): \; f(u_p) > 0$*

Тогда $\hat{u}_p$ является ас. нормальной оценкой u_p с ас. дисперсией $\frac{p(1-p)}{f(u_p)^2}$. Аналогично для медианы $\hat{\mu}$ является а.н.о. $u_{\frac{1}{2}}$ с ас. д. $\frac{1}{4f(u_{\frac{1}{2}})^2}$.

Выборочная медиана является а.н.о. сдвига в распределении Коши, но оптимальной ас. дисперсией не обладает. А именно, ассимптотическая дисперсия выборочный медианы: $\sigma_{\hat{\mu}} = \frac{\pi^2}{4} \approx 2.47$, в то время как ОМП имеет ас. дисперсию: $\sigma_{\hat{\theta}} = 2$.

3.3 Свойства оценок в модели линейной регрессии при несмещённости и гомоскедастичность шума

- Предполагаемая зависимость: $y(x) = \theta^T x$
- Наблюдаемая зависимость: $Y = X\theta + \varepsilon$
- Задача: Оценить θ

Рассмотрим $RSS(\theta) = \sum\limits_{i=1}^{n}(Y_i - \theta^T X_i)^2 = \|Y - X\theta\|^2$. Тогда МНК оценка:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, RSS(\theta)$$

Если $(X^T X)$ - невырождена, то $\hat{\theta} = (X^T X)^{-1} X^T Y$ - оценка МНК.

Рассмотрим некоторые свойства полученного резльутата

Теорема (О свойствах оценки МНК) *Пусть обозначения такие же, как выше. Тогда:*

1. *Если* $\mathbb{E}\varepsilon = 0$, *то* $\mathbb{E}\hat{\theta} = \theta, \mathbb{E}\hat{y}(x) = y(x), \quad \hat{y}(x) = \hat{\theta}^T x$
2. *Если* $\mathbb{D}\varepsilon = \sigma^2 I_n$, *то* $\mathbb{D}\hat{\theta} = \sigma^2 (X^T X)^{-1}, \mathbb{D}\hat{y}(x) = \sigma^2 X^T (X^T X)^{-1} x$

Докажим оба этих свойства.

1. $\mathbb{E}\hat{\theta} = \mathbb{E}\left[(X^T X)^{-1}(X^T Y)\right] = (X^T X)^{-1} X^T (X\theta) = \theta$
2. $\mathbb{D}\hat{\theta} = \mathbb{D}\left[(X^T X)^{-1}(X^T Y)\right] = (X^T X)^{-1} X^T \mathbb{D}\left[X\theta + \varepsilon\right] X (X^T X)^{-1} = (X^T X)^{-1} X^T \sigma^2 I_n X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$

Теорема (Об оценке дисперсии) *Если* $\mathbb{E}\varepsilon = 0, \mathbb{D}\varepsilon = \sigma^2 I_n$, *то оценка* $\hat{\sigma} = \frac{RSS(\hat{\theta})}{n-d}$ *- несмещённая оценка* σ^2

Доакжем это:

$$\mathbb{E}RSS(\hat{\theta}) = \mathbb{E}\sum_{i=1}^{n}(y_i - \hat{\theta}^T x_i)^2 = \mathbb{D}\sum_{i=1}^{n}(y_i - \hat{\theta} x_i) = tr\mathbb{D}(Y - X\hat{\theta})$$

Так как $\mathbb{E}(y_i - \hat{\theta} x_i) = 0$

$$\mathbb{D}(Y - X\hat{\theta}) = \mathbb{D}((I_n - X(X^T X)^{-1} X^T) Y) = /A = X(X^T X)^{-1} X^T/ = \mathbb{D}((I_n - A)Y) = (I_n - A)\mathbb{D}Y(I$$

$$tr\mathbb{D}(Y - X\hat{\theta}) = 2\sigma^2(n - tr(A)) = 2\sigma^2(n - d)$$

Потому что $tr(A) = tr(X(X^T X)^{-1} X^T) = tr((X^T X)(X^T X)^{-1}) = d$

3.4 Сравнение оценок

Пусть хотим оценить параметр $\tau(\theta) \in \mathbb{R}^d$. Введём несколько понятий:

- Функция $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ - количественная степень отклонения оценки от истинного $\tau(\theta)$, называется функцией потерь (loss functions). Примеры:

  1. $L(x,y) = (x-y)^2$
  2. $L(x,y) = |x-y|$
  3. $L(x,y) = (x-y)^T A(x-y)$, где A - положительно полуопределена.

  Если $\hat{\theta}$ - оценка $\tau(\theta)$, то при таком оценивании величина $L(\hat{\theta}, \tau(\theta))$ - штраф, который накладывается за ошибку. Недостаток такого подхода в том, что этот штраф случаен: при разных реализациях получаем разные штрафы

- Функция риска оценки $\hat{\theta}$ величины $\tau(\theta)$:

$$R_{\hat{\theta},\tau(\theta)}(\theta) = \mathbb{E}_\theta L(\hat{\theta}, \tau(\theta))$$

Подходы к сравнению оценок

3.4.1 Равномерный

Определение 3.7. *1. Оценка $\hat{\theta}_1$ не хуже оценки $\hat{\theta}_2$, если*

$$\forall \theta \in \Theta \quad R_{\hat{\theta}_1,\tau}(\theta) \le R_{\hat{\theta}_2,\tau}$$

2. Оценка $\hat{\theta}_1$ лучше оценки $\hat{\theta}_2$, если она не хуже, и, кроме этого:

$$\exists \theta \in \Theta \quad R_{\hat{\theta}_1,\tau}(\theta) < R_{\hat{\theta}_2,\tau}$$

3. Пусть K - некоторый класс оценок. Оценка $\hat{\theta}$ называется лучшей, если она лучше всех остальных оценок из этого класса.

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

В классе может не быть наилучшей оценки.

Теорема (О bias-variance разложении) *MSE допускает bias-variance разложение:*

$$MSE_{\hat{\theta},\tau}(\theta) = \mathbb{E}_\theta(\hat{\theta} - \tau(\theta))^2 = \mathbb{D}_\theta\hat{\theta} + \Big(\mathbb{E}_\theta\hat{\theta} - \tau(\theta)\Big)^2$$

Наилучшая оценка в классе несмещённых оценок, таким образом, это та, у которой наименьшая дисперсия.

3.4.2 Байесовский

Пусть задано Q - некоторое распределение на Θ.

Определение 3.8. • *$\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если*

$$\mathbb{E}_Q R_{\hat{\theta}_1,\tau(\theta)} \le \mathbb{E}_Q R_{\hat{\theta}_2,\tau(\theta)}$$

- *Определение "лучше"аналогично таковому в равномерном подходе, ровно как и лучшей оценки в классе.*

3.4.3 Min-Max подход

Определение 3.9. • *$\hat{\theta}_1$ не хуже $\hat{\theta}_2$, если*

$$\sup_{\theta in \Theta} R_{\hat{\theta}_1,\tau(\theta)} \le \sup_{\theta in \Theta} R_{\hat{\theta}_2,\tau(\theta)}$$

- *Определение "лучше"аналогично таковому в равномерном подходе, ровно как и лучшей оценки в классе.*

3.4.4 Ассимптотический (для класса а.н. оценок)

Определение 3.10. *Пусть $\hat{\theta}_1, \hat{\theta}_2$ - а.н.о. с ас. дисперсиями $\sigma_1^2(\theta), \sigma_2^2(\theta)$. Тогда $\hat{\sigma}_1$ не хуже $\hat{\sigma}_2$, если выполнено:*

$$\forall \theta \in \Theta \sigma_1(\theta) \le \sigma_2(\theta)$$

Определение "лучше"аналогично таковому во всех предыдущих подходах.

Определение 3.11. *Число*

$$ARE^{\tau}_{\hat{\theta}_1,\hat{\theta}_2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$$

Называется относительной ассимптотической эффективностью (assymptotic relative efficiency)

Определение 3.12. *Лучшая оценка в ассимптотическом подходе называется ассимптотически эффективной оценкой.*

Теорема (Об ассимптотической эффективности ОМП) *Если выполнены условия регулярности $[L1 - L9]$, то ОМП является ассимптотически эффективной с ас. дисперсией $i^{-1}(\theta)$, Где $i(\theta)$ - информация Фишера*

Примеры:

1. Пусть $X_1, X_2, ..., X_n \sim \mathcal{N}(\theta, 1)$. Тогда $\overline{X}, \hat{\mu}$ - среднее и медиана, явлвяются а.н.о. для θ с ас. дисперсиями $\sigma_1^2 = 1, \sigma_2^2 = \frac{\pi}{4}$. Ассимптотическая эффективность в данном случае:

$$ARE_{\overline{X},\hat{\mu}} = \frac{\pi}{2} \approx 1.57$$

2. Пусть $X_1, X_2, ..., X_n \sim Exp(\theta)$. Тогда $\frac{1}{\overline{X}}, \ln \overline{I\{X > 1\}}$ явлвяются а.н.о. для θ с ас. дисперсиями $\sigma_1^2 = \theta^2, \sigma_2^2 = e^{\theta} - 1$. Ассимптотическая эффективность в данном случае:

$$ARE_{\frac{1}{\overline{X}},\ln \overline{I\{X>1\}}} = \frac{e^{\theta} - 1}{\theta^2}$$

3.5 Достаточные статистики

Пусть $X = (X_1, X_2, ..., X_n) \sim P \in \mathcal{P} = \{P_\theta, \theta \in \Theta\}$

Определение 3.13. *Статистика $S(X)$ называется достаточной для $\mathcal{P}$, если условное распределение:*

$$P_\theta(X \in B \mid S(X))$$

Не зависит от $\theta, \forall B \in \mathcal{B}$.

Смысл данного определения заключается в том, что вся информация, которую в себе несёт выборка о неизвестном параметре θ заключена в достаточной статистике. Всегда существует тривиальная достаточная статистика: $S(X) = (X_1, X_2, ..., X_n)$. Следствием является то, что вместо всей выборкой достаточно хранить лишь достаточную статистику.

Важным частным случаем является ситуация, в которой размерность досаточной статистики меньше размера выборки.
Пример: Пусть $X_1, X_2, ..., X_n \sim Bern(p)$. Информация в выборке:

1. Кол-во успехов: $S(X) = \sum\limits_{i=1}^{n} X_i$
2. Порядок успех-неудача

Докажем, что $S(X)$ является достаточной статистикой:

$$\frac{P_\theta(X_1 = x_1, ..., X_n = x_n, S(X) = s)}{P_\theta(S(X) = s)} = \frac{I\{S(X) = s\}}{C_n^s}$$

Последнее выражение не зависит от θ.

Теорема (Критерий факторизации Неймана-Фишера) *Пусть* $X = (X_1, X_2, ..., X_n)$ *- выборка из* $P \in \mathcal{P}$ *- доминирующее семейство распределений. Тогда* $S(X)$ *является достаточной статистикой тогда и только тогда, когда справделиво разложение:*

$$p_\theta(X) = \psi(S(X), \theta) \times h(X)$$

Где $h(x)$ *не зависит от* θ.

Примеры:

1. $X_1, X_2, ..., X_n \sim \Gamma(\alpha, \beta), \theta = (\alpha, \beta)$.

$$p_\theta(X) = \frac{\alpha^{n\beta}}{\Gamma(\beta)^n} \left(\prod_{i=1}^{n} X_i\right)^{\beta - 1} \exp(-\alpha \sum_{i=1}^{n} X_i)$$

$$h(x) = 1, S(X) = \left(\sum_{i=1}^{n} X_i, \prod_{i=1}^{n} X_i\right). \text{ Но лучше взять } S_1(X) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} \ln X_i\right)$$

3.6 Экспоненциальный класс распределений

Определение 3.14. *Семейство $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ принадлежит экспоненциальному классу распределений, если его плотность представима в виде:*

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

Где $g(x) \geq 0, u(x)$ - борелевские. Если $a(\theta) = \theta$, то такая параметризация называется естественной.

Пример: Покажем, что $\mathcal{N} = \{\mathcal{N}(a,\, \sigma^2) :\ a \in \mathbb{R},\ \sigma > 0\}$ принадлежит эксп. классу.

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

Тогда можно выбрать:

1. $u(x) = (-x^2, x), \theta = \left(\frac{1}{2\sigma^2}, \frac{a}{\sigma^2}\right), g(x) = 1$

Найдём достаточные статистики. По критерию факторизации Неймана-Фишера:

$$S(X) = \sum_{i=1}^{n} u(x_i)$$

Всегда будет являться достаточной статистикой для семейства из экспоненциального класса

Теорема *Пусть $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ - семейство распределений, такое, что плотность непрерывно дифференцируема по x на своём носителе, который не зависит от θ. Пусть также $S(X)$ - достаточная статистика фиксированной размерности.*
Тогда это семейство лежит в экспоненциальном классе, причем $dimS(X) \geq dima(\theta)$

Следствием этой теоремы является то, что если плотность достаточно хороошая, то только у распределений из эксп. класса существуют достаточные статистики фиксированного размера.

Примеры:

1. Распределение Коши. Оно не принадлежит экспоненциальному классу, однако выполнены условия теоремы. Значит, для него не существует достаточной статистики фиксированного размера.

2. $U[0, \theta]$ - не принадлежит эксп. классу, так как носитель зависит от θ, однако существует достаточная статистика $X_{(n)}$,

Далее будем предполагать следующие условия на семейство распределений:

1. Естесственная параметризация
2. g, u - непрерывны
3. Условие равномерной сходимости интеграла по параметру:

$$\forall s \forall j \leq K \exists \phi(x): \quad \forall \theta |g(x) u_j(x) e^{\theta^T u(x)}| \leq \phi(x), \int_X \phi(x) dx - \text{сх-ся}$$

Следствия:

1. $h(\theta)$ непр. дифф. k раз
2. $p_\theta(x)$ непр. дифф. по θ k раз.
3. Можно переставлять местами $\frac{\partial}{\partial \theta}$ и $\int$

Теорема (О свйоствах эксп. класса)

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

$$\mathbb{D}_\theta U(X_1) = \nabla \nabla \ln h(\theta) \succeq 0$$

Доказательство производится путём прямой подстановки

Теорема Θ - *выпуклое множество в* $\mathbb{R}^d$, *тогда ОМП* $\exists !$

Доказательство: Запишем логарифм правдопобия:

$$l_X(\theta) = -n \ln h(\theta) + \sum_{i=1}^{n} \ln g(X_i) + \theta^T \sum_{i=1}^{n} U(X_i)$$

Видно, что полученная функция вогнута, поэтом оптимизация правдоподбия - задача оптимизации вогнутой функции на выпуклом множестве, а она имеет единственое решение.

Теорема Θ - *выпуклое, тогда выполнены* $[L1 - L9]$

L1-L5 - очевидно выполнены

L5-L7 - выполеннеы по следствию и предположениям

- L8:
$$i(\theta) = \mathbb{E}_\theta \left(\frac{\partial \ln p_\theta(X_i)}{\partial \theta} \right)^2, \frac{\partial \ln p_\theta(X_i)}{\partial \theta} = u(X) - \frac{1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta}$$

- L9:
$$\frac{\partial^3 \ln p_\theta(X)}{\partial^3 \theta} = \frac{\partial^2}{\partial^2 \theta} \left(u(X) - \frac{1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta} \right)$$
  Не зависит от x.

3.7 Приближённый поиск ОМП. Метод Ньютона для решения урванения $f(x) = 0$

Метод Ньютона это итеративный процесс, в котором

1. X_0 - стартовая точка
2. X_{k+1}- точка пересечения касательной в точке X_k с осью OX.
$$X_{k+1} = X_k - \frac{f(X_k)}{f'(X_k)}$$

У данного метода квадратичная скорость сходимости в общем случае. Попробуем применить его для приближённого нахождения ОМП.

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из $P \in \mathcal{P}$. При этом

- $p_\theta(x)$ - плотность
- $\Theta \subseteq \mathbb{R}^d, \theta^*$ - ОМП
- $\frac{\partial l_x(\theta)}{\partial \theta} = 0$ - уравнение правдоподобия

Если $\hat{\theta}_0$ - начальное приближение, то $\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{l'_x(\hat{\theta}_k)}{l''_X(\hat{\theta}_k)}$

Теорема (Об одношаговой оценке) *Пусть выполнены условия регулярности $[L1 - L9]$, причём $\hat{\theta}_0$ - а.н.о. для θ. Тогда $\hat{\theta}_1$ - а.н.о. θ с ас. дисперсией $i^{-1}(\theta)$, то есть ас. эффективна. При этом:*
$$\sqrt{n}(\hat{\theta}_1 - \theta^*) \xrightarrow{P} 0$$

Идея доказательства для $d = 1$:

Теорема (Без доказательства) $\hat{\theta}_1 - \theta^* = (\hat{\theta}_0 - \theta^*)\varepsilon_n(\theta)$, *где* $\varepsilon_n(\theta) \xrightarrow{P} 0$

Тогда:

1. Ас. эквивалентность:

$$\sqrt{n}(\hat{\theta}_1 - \theta^*) = \sqrt{n}(\hat{\theta}_0 - \theta^*)\varepsilon_n(\theta) = \sqrt{n}\varepsilon_n(\theta)(\hat{\theta}_0 - \theta) + \sqrt{n}\varepsilon_n(\theta)(\theta - \theta^*) \xrightarrow{P} 0$$

2. Ас. нормальность:

$$\sqrt{n}(\hat{\theta}_1 - \theta) = \sqrt{n}(\hat{\theta}_1 - \theta^*) + \sqrt{n}(\theta^* - \theta)$$

В данной теореме l''_X можно заменить на $\mathbb{E}_\theta l''_X = -ni(\theta)$

Пример: Наконец-то мы победим распределение Коши. Ассимптотически эффективная одношаговая оценка (полученная из оценки медианой):

$$\hat{\theta}_1 = \hat{\mu} + \frac{\sum\limits_{i=1}^{n} \frac{X_i - \hat{\mu}}{1 + (X_i - \hat{\mu})^2}}{\sum\limits_{i=1}^{n} \frac{1 - (X_i - \hat{\mu})^2}{(1 + (X_i - \hat{\mu})^2)^2}}$$

3.8 Robustness и симметричные распределения

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из $\mathcal{N}(\theta, \sigma^2)$. У оценки $\hat{\theta} = \overline{X}$ куча хороших свойств. Однако, если в выборке присутствуют выбросы, то все эти свойства теряются.

Определение 3.15. *Робастными называются оценки, допускающие отклонения от заданной модели.*

Определение 3.16. *Пусть оценка имеет вид* $\hat{\theta} = f(X_{(1)}, X_{(2)}, ..., X_{(n)})$. *Обозначим* K_n^* *- наименьшее число* k, *при котором*

1. *Если* $X_1, X_2, ..., X_{k+1} \to -\infty, X_{k+2}, ..., X_n - fix \Rightarrow f(X_1, ..., X_n) \to -\infty$

2. *Если* $X_{n-k}, X_{n-k+1}, ..., X_n \to +\infty, X_1, X_2, ..., X_{n-k-1} - fix \Rightarrow f(X_1, ..., X_n) \to +\infty$

Тогда ассимптотической толерантностью оценки $\hat{\theta}$ называется

$$\tau_{\hat{\theta}} = \lim_{n\to\infty} \frac{K_n^*}{n}$$

Пример: $\hat{\theta} = \overline{X} \Rightarrow \tau_{\hat{\theta}} = 0, \hat{\theta} = \hat{\mu} \Rightarrow K_n^* = \lfloor \frac{n}{2} \rfloor \Rightarrow \tau_{\hat{\theta}} = \frac{1}{2}$

Смысл данного определения заключается в том, что при сильном смещении элементов выборки оценка сама не смещается в сторону этих элементов.

Мы хотим найти оценки:

1. Достаточно эффективные (с точки зрения ас. эффективности)
2. Робастные

Далее рассмотрим класс распределений $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, где

1. P_θ имеет плотность $p_\theta(x)$ - чётная, непрерывная, имеет носитель $(-c, c)$
2. $\theta \in \Theta = \mathbb{R}$ - параметр сдвига. $p_\theta(x) = p_0(x - \theta)$

3.8.1 Усечённое среднее

Определение 3.17. *Пусть $\alpha \in (0, \frac{1}{2}), K = \lceil \alpha n \rceil$, тогда усечённым средним с параметром α называется величина*

$$\overline{X_\alpha} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X_{(i)}$$

Важные частные случаи:

1. $\alpha = 0 \Rightarrow \overline{X_\alpha} = \overline{X}$
2. $\alpha = \frac{1}{2} \Rightarrow \overline{X_\alpha} = \hat{\mu}$

При этом заметим, что

$$\tau_{\overline{X_\alpha}} = \lim_{n\to\infty} \frac{\lceil n\alpha \rceil}{n} = \alpha$$

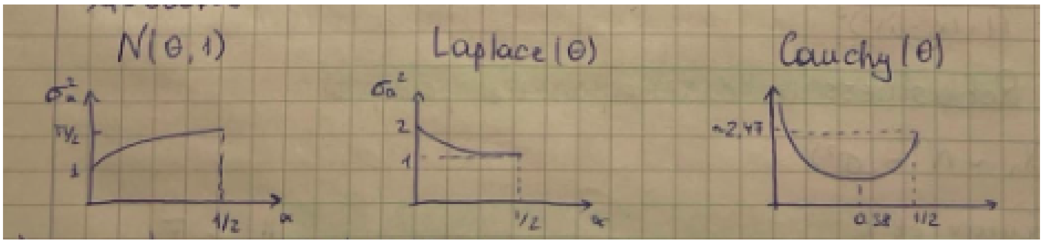

Рис. 1: Ас. эффективность усечённых средних

Теорема *Пусть $X_1, X_2, ..., X_n$ - выборка из $P \in \mathcal{P}$. Тогда $\overline{X_\alpha}$ - а.н.о θ, причём ас. дисперсия выражается как:*

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left(\int_0^{u_{1-\alpha}} x^2 p_0(x)dx + \alpha u_{1-\alpha}^2\right)$$

Где $u_\alpha - \alpha$ квантиль распределения P_0.

Примеры:
$\mathcal{N}(\theta,\, 1)$:

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

Теорема *Пусть $\mathbb{D}_\theta X_1 < +\infty$, тогда*

$$ARE_{\overline{X_\alpha},\overline{X}} \le (1-2\alpha)^2$$

Доказательство:

$$\frac{1}{2}\mathbb{D}_\theta X_1 = \frac{1}{2}\int_{\mathbb{R}} x^2 p_\theta(x)dx = \int_0^\infty x^2 p_\theta(x)dx = \left(\int_0^{u_{1-\alpha}} dx + \int^{u_{1-\alpha}^{\infty}} dx\right) x^2 p_\theta(x) \ge$$

$$\ge u_{1-\alpha}^2\alpha + \int_0^{u_{1-\alpha}} x^2 p_\theta(x)dx = \frac{(1-2\alpha^2)}{2}\sigma_\alpha^2$$

3.8.2 Медиана средних Уолша

Определение 3.18. *Обозначим $Y_{i,j} = \frac{X_i + X_j}{2}, 1 \le i \le j \le n$ - средние Уолша. Тогда медианой средних Уолша называется число W - медиана всех $Y_{i,j}$*

Теорема *Пусть $X_1, X_2, ..., X_n \sim P_\theta \in \mathcal{P}$, тогда медиана средних Уолша - а.н.о. для θ, причём ас. дисперсия выражается как:*

$$\sigma^2 = \frac{1}{12\left(\int_{\mathbb{R}} p_\theta^2(x)dx\right)^2}$$

Для $P_\theta \in \mathcal{P} \quad ARE_{W,\overline{X}} \geq \frac{108}{125} \approx 0.864$
$\tau_W \approx 0.293 \Rightarrow W$ устойчива к 29% выбросов

4 Проверка статистических гипотез

4.1 Гипотезы и критерии

Пусть $\mathcal{X}$ - выборочное пространство, $\mathcal{P}$ - семейство распределений на нём. Рассматриваются высказывания вида:

- $H_0: \quad P \in P_0$ - основная гипотеза
- $H_1: \quad P \in P_1$ - альтернативная гипотеза

Где $P_0, P_1 \in \mathcal{P},\ P_0 \bigcap P_1 = \emptyset$. Для параметрического подхода гипотезы имеют вид:

- $H_0: \quad \theta \in \Theta_0$
- $H_0: \quad \theta \in \Theta_1$

Где $\Theta_0, \Theta_1 \subseteq \Theta, \quad \Theta_0 \bigcap \Theta_1 = \emptyset$

Не обязательно, чтобы $\Theta_0 \bigcup \Theta_1 = \Theta$. Например, гипотезы могут быть:

1. $H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta > \theta_0$ - правосторонняя альтернатива
2. $H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta < \theta_0$ - левосторонняая альтернатива
3. $H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta \neq \theta_0$ - двусторонняя альтернатива

Истинное θ может не удовлетворять ни одной.

Примеры:

1. $\mathcal{P} = \{$все абс. непр. распределения$\}$
   $H_0: \ P \sim \mathcal{N}(a, \sigma^2),\ H_1: \ P \sim \exp(\lambda)$
2. $\mathcal{P} = \{\mathcal{N}(a, \sigma^2) : a \in \mathbb{R}, \sigma > 0\}$
   $H_0: \ a = 0,\ H_1 : a > 0$

Определение 4.1. *Подмножество $S \subseteq \mathcal{X}$ называется критерием проверки $H_0 vs. H_1$, если правило отвержения гипотезы H_0 при выборке X имеет вид:*

$$H_0 \textit{отвергается} \Leftrightarrow X \in S$$

Часто используется следующий частный случай: Берут некоторую статистику $T(X)$, а в качестве критерия выбирают множество:

$$S = \{X | T(x) > c\}$$

Для некоторого c, которое называется в данном случае критическим значением. Правило отвержения в этом случае будет записываться как:

$$H_0 \text{отвергается} \Leftrightarrow T(X) > c$$

Определение 4.2. *Пусть S - критерий для проверки $H_0 vs. H_1$. Тогда:*

- *$X \in S \Rightarrow H_0$ - отвергается, то результат тестирования* ***статистически значим***
-
- *$X \notin S \Rightarrow H_0$ - не отвергается, то результат тестирования* ***статистически незначим***

В общем случае неверно, что если H_0 отвергается, то принимается H_1. Этот приницип схож с приниципом презумпции невиновности: **Каждый обвиняемый в совершении преступления считается невиновным, пока его виновность не будет доказана в предусмотренном законом порядке и установлена вступившим в законную силу приговором суда** Аналогию с судом можно продолжить:

Обвиняемый	P - неизвестное распределение
Невиновность	$H_0 : P \in P_0$
Виновность	$H_1 : P \in P_1$
Преступление	$X = (X_1, ..., X_n)$ - выборка
Улики и доказательства	$T(X)$ - статистика критерия
Доказательство виновности	Справедливость $X \in S$

При проверке гипотез могут возникать ошибки. Ошибки бывают двух сортов: первого и второго.

	H_0 верна	H_0 неверна
H_0 не отв.	=)	II
H_0 отв.	I	=)

Определение 4.3. *Ошибка первого рода - ситуация, в которой мы отвергли верную гипотезу. Ошибка второго рода - ситуация, в которой мы не отвергли ложную гипотезу.*

Определение 4.4. *Вероятности ошибки (хотя это совсем не вероятность), принято называть следующие величины:*

1. *Вероятность ошибки первого рода:* $P(I_S) = \sup\limits_{P \in P_0} P(X \in S)$
2. *Вероятность ошибки второго рода:* $P(II_S) = \sup\limits_{P \in P_1} P(X \notin S)$

Ошибки первого рода считаются более опасными. Поэтому обычно решается задача:

$$\begin{aligned} P(I_S) \le \alpha \\ P(II_S) \to \min_S \end{aligned} \tag{4}$$

Определение 4.5. *Уровнем значимости критерия S называется такое число α, что $P(I_S) \le \alpha$ При этом число $\alpha_P = P(I_S)$ называется точным уровнем значимости.*

Обычно выбирают $\alpha = 0.05$ - первая магическая константа.
Обычно альтернативы имеют сложный вид:

- $H_0 : \theta = \theta_0 vs H_1 : \theta > \theta_0$
- $H_0 : P \sim \mathcal{N}(a, \sigma^2) vs H_1 : P \not\sim \mathcal{N}(a, \sigma^2)$

Значит, имеет смысл сравнивать критерии на разных распределениях из альтернативы. Для этого введём:

Определение 4.6. $\beta_S(P) = P(X \in S)$ *для* $P \in \mathcal{P}_1$ *- функция мощности.*

Пример: Пусть X - выборка из 1 элемента из распределения $Exp(\theta)$. Построить критерий уровня значимости α для проверки гипотез $H_0 : \theta = \theta_0 vs H_1 : \theta > \theta_0$. Заметим, что $\mathbb{E}_\theta X = \frac{1}{\theta} \Rightarrow$ чем больше θ, тем меньше X в среднем. Разумно брать критерии вида $S = \{X < c\}$ где c мы подберём из условия $P(I_S) \le \alpha$. Выпишем это условие:

$$P(I_S) = P_\theta(X < c) = 1 - e^{-\theta_0 c} \le \alpha \Rightarrow c \le -\frac{1}{\theta_0} \ln(1 - \alpha)$$

Опеределим мощность полученного критерия:

$$\beta_S(\theta) = 1 - \exp(\frac{\theta}{\theta_0}\ln(1-\alpha)) = 1 - (1-\alpha)^{\frac{\theta}{\theta_0}}$$

Таким образом, полученный критерий:

$$S = \{X < -\frac{1}{\theta_0}\ln(1-\alpha)\}$$

Пример с числами: Пусть $\theta_0 = 1, a = 0.05 \Rightarrow c = 0.051$. Соответственно, если $X < 0.051$, то гпиотеза H_0 отвергается (результат стат. значим). Иначе - статистически незначим

4.2 Критерий Вальда

Определение 4.7. *Критерий S - асимптотический критерий уровня значимости α, если*

$$\overline{\lim_{n\to\infty}} P(I_S) \leq \alpha$$

Пусть $X = (X_1, ..., X_n)$ - выборка из неизвестного распределения $P \in \mathcal{P}, \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^d\}$. Раассмотрим гипотезы:

1. $H_0 : \theta = \theta_0$
2. $H_1 : \theta = \theta_1$

Пусть у нас есть $\hat{\theta}$ - ассимптотически нормальная оценка θ с ассимптотической дисперсией $\sigma(\theta)^2$, а также $\hat{\sigma}$ - состоятельная оценка σ. Тогда из свойств а.н.о следует, что:

$$W(X) = \sqrt{n}\frac{\hat{\theta} - \theta}{\hat{\sigma}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Таким образом, мы получаем следующий критерий

Определение 4.8. *Критерий $S = \{\sqrt{n}\frac{|\hat{\theta}-\theta|}{\hat{\sigma}} > z_{\frac{1-\alpha}{2}}\}$*

Ясно, что это ассимптотический критерий уровни значимости α. Изучим мощность данного критерия:

$$\beta_S(\theta) = P_\theta(W > z_{1-\frac{\alpha}{2}}) + P_\theta(W < z_{\frac{\alpha}{2}}) =$$

$$= P_\theta(\sqrt{n}\frac{\hat{\theta} - \theta}{\hat{\sigma}} > z_{1-\frac{\alpha}{2}} - \sqrt{n}\frac{\theta - \theta_0}{\hat{\sigma}}) + P_\theta(\sqrt{n}\frac{\hat{\theta} - \theta}{\hat{\sigma}} < z_{\frac{\alpha}{2}} - \sqrt{n}\frac{\theta - \theta_0}{\hat{\sigma}}) \to$$
$$\to 1 - \Phi(z_{1-\frac{\alpha}{2}} - W(\theta)) + \Phi(z_{\frac{\alpha}{2}} - W(\theta))$$

Замечания:

1. Кртерий Вальда обощается на случай $H_1 : \theta > \theta_0$:

$$S_r = \{\sqrt{n}\frac{\hat{\theta} - \theta_0}{\hat{\sigma}} > z_{1-\alpha}\}$$

(уроень значимости α)

2. Кртерий Вальда обощается на случай $H_1 : \theta < \theta_0$:

$$S_l = \{\sqrt{n}\frac{\hat{\theta} - \theta_0}{\hat{\sigma}} < z_{\alpha}\}$$

(уроень значимости α)

3. Можно рассмотреть доверительный интервал:

$$C = (\hat{\theta} \pm z_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}})$$

Он имеет уровень доверия $1 - \alpha$. Тогда H_0 отвергается тогда и только тогда, когда $\theta_0 \notin C$.

4.

4.3 Критерии отношения правдоподобия

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из распределения $P \in \mathcal{P}$, где $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ - доминирующее семейство. Попробуем в частных случаях решить задачу:

$$\begin{cases} P(I_S) \leq \alpha \\ P(II_S) \to \min \end{cases} \tag{5}$$

1. Простые гипотезы.
Рассмотрим случай: $H_0 : \theta = \theta_0 vs. H_1 : \theta = \theta_1$. Рассмотрим статистику:

$$\Lambda(X) = \frac{L_X(\theta_1)}{L_X(\theta_0)}$$

Теорема (Лемма Неймана-Пирсона) *Если* $\exists c_\alpha : \quad P_{\theta_0}(\Lambda(X) > c_\alpha) = \alpha$, *то критерий*

$$S(X) = \{\Lambda(X) > c_\alpha\}$$

является критерием уровня значимости α *и наибольшей мощностью из всех таких критериев.*

2. Сложные гипотезы

Определение 4.9. *Критерий S уровня значимости α называется равномерно наиболее мощным критерием (РНМК), если для любого другого критерия T того же уровня значимости выполнено неравенство:*

$$\forall P \in P_1 : \quad \beta_S(P) \geq \beta_T(P)$$

Теорема (О монотонном отношении правдоподобия) *Пусть $\forall \theta_1 > \theta_2$ отношение правдоподобия представимо в виде:*

$$\frac{L_X(\theta_1)}{L_X(\theta_2)} = f_{\theta_1, \theta_2}(T(X))$$

Где $T(X)$ - некоторая статистика, а f_{θ_1,θ_2} - монотонно возрастающая функция. Тогда критерий:

$$S = \{T(X) > c_\alpha\}$$

является РНМК с уровнем значимости α для проверки гипотез

$$H_0 : \theta = \theta_0 \ vs \ H_1 : \theta > \theta_0$$

Где c_α подбирается из условия $P_{\theta_0}(T(X) > c_\alpha) = \alpha$

Пример: Рассмотрим $(X_1, X_2, ..., X_n) \sim Exp(\theta)$. Рассмотрим также гипотезы $H_0 : \theta = \theta_0, H_1 : \theta > \theta_0$. Требуется найти РМНК. Пусть $\theta_1 > \theta_2$, выпишем отношение правдоподобия:

$$\Lambda(X) = \frac{L_X(\theta_1)}{L_X(\theta_2)} = \frac{\theta_1^n \exp(-\theta_1 \sum\limits_{i=1}^{n} X_i)}{\theta_2^n \exp(-\theta_2 \sum\limits_{i=1}^{n} X_i)} = \left(\frac{\theta_1}{\theta_2}\right)^n \exp\left[-(\theta_1 - \theta_2)\left(\sum_{i=1}^{n} X_i\right)\right]$$

Тогда для статистики $T(X) = \sum\limits_{i=1}^{n} X_i$ выполняются условия теормы о монотонном отношении правдоподобия. Полученный критерий:

$$S = \{T(X) < c_\alpha\}$$

Где c_α находится из условия $P_{\theta_0}(T(X) < c_\alpha) = \alpha$. Мы знаем из курса ТВ, что $T(X) \sim \Gamma(\theta_0, n)$. В таком случае, c_α - α квантиль распределения $\Gamma(\theta_0, n)$

5 Непараметрический подход

5.1 Эмперическое распределение

Пусть $X = (X_1, X_2, ..., X_n)$ - выборка из распределения $P \in \mathcal{P}$.

Определение 5.1. *Распределение $\hat{P}_n(X)$ - эмперическое распределение по выборке X, если для любого множества $B \in \mathcal{B}$:*

$$\hat{P}_n(B) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \in B\}$$

Заметим следующие свойства:

1. $\hat{P}_n(B)$ - случайная величина, доля элементов, попавших в B.
2. $\hat{P}_n$ - случайная вероятностная мера.
3. $n\hat{P}_n(B) \sim B(n, P(B))$
4. $\mathbb{E}\hat{P}_n(B) = P(B)$
5. $\mathbb{D}\hat{P}_n(B) = \frac{p(1-p)}{n}/$
6. $\hat{P}_n(B) \overset{a.e.}{\to} P(B)$

Рассмотрим случай $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Тогда $\hat{P}_n$ соответствует эмпирической функции распределения. Можно определить эмперичческую функцию распределения

Определение 5.2. *Эмпирической функцией распределения называется функция:*

$$\hat{F}_n(x) = \hat{P}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \le x\}$$

Тогда истинно утверждение: $\hat{F}_n(x) \overset{a.e.}{\to} F(x)$. Более того, верна следующая

Теорема (Гливенко-Кантелли) *Если F - функция распределения для $P, (X_1, X_2, ..., X_n) \sim P$, тогда верно:*

$$D_n = \sup_{x \in R} |\hat{F}_n(x) - F(x)| \overset{a.e.}{\to} 0$$

Заметим также, что $D_n = \sup_{B \in \mathcal{A}} |\hat{P}_n(B) - P(B)|$, $\mathcal{A} = \{(-\infty, x], x \in \mathbb{R}\}$. Эту теорему можно обобщить и на другие $\mathcal{A}$. Например: $\mathcal{A} = \{$Все объединения интервал
На любые конечные объединения обобщить нельзя.

Теорема (Вапника-Червоненкинса)

$$\sup_{B \in \mathcal{A}} |\hat{P}_n(B) - P(B)| \xrightarrow{a.e.} 0 \Leftrightarrow dim_{VC} < \infty$$

При разбиении $\mathbb{R}^d$ множествами из $\mathcal{A}$

Теорема (Колмогорова-Смирнова) *Рассмотрим $\sqrt{n}D_n$. Утверждается, что*

$$\sqrt{n}D_n \xrightarrow{d_\theta} \xi$$

Где $\xi \sim Kolmogorov$ - распределение Колмогорова, его функция распределения:

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

Скорость сходимости: $\sqrt{N}$

5.2 Метод подстановки

Определение 5.3. *Пусть $X = (X_1, X_2, ..., X_n) \in P, F$ - функция распределения. Пусть также $\theta = G(P)$ - параметр, значение которого хотим оценить. Тогда оценкой по методу подстановки называют $\hat{\theta} = G(\hat{p}_n)$ - оценка методом подстановки*

Примеры:

1. $\theta = G(P) = \mathbb{E}_P f(X_1) = \int_\mathbb{R} f(x)dF(x)$ - линейный функционал. Тогда $\hat{\theta} = \mathbb{E}_{\hat{p}_n} f(x_1) = \overline{f(x_i)}$

2. $G(P) = \mathbb{D}_P(x_1) = \mathbb{E}_P(x_1^2) - \mathbb{E}_p(x_1)^2$.Тогда получаем

$$\hat{\theta} = \overline{X^2} - \overline{X}^2 = S^2$$

3. $\theta = G(P) = F^{-1}(\alpha) = \min\{x | \hat{F}_n(x) \geq \alpha\} = X_{(\lceil n\alpha \rceil)}$