

математическая статистика. Билеты

alexander.veselyev

13 марта 2019 г.

Содержание

1	Вероятностно-статистическая модель. Наблюдение и выборка. Задача оценивания параметров. Статистики и оценки. Свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Взаимосвязь между свойствами оценок. Выборочные характеристики и порядковые статистики. Свойства выборочного среднего.	4
2	Теорема о наследовании сходимостей. Лемма Слуцкого с доказательством и ее применение. Наследование асимптотической нормальности. Многомерный случай. Наследование состоятельности.	7
3	Непараметрическая задача. Эмпирическое распределение и эмпирическая функция распределения. Теорема Гливенко-Кантелли.	11
4	Выборочные квантили. Асимптотическая нормальность выборочной квантили. Медиана, выборочная медиана и ее асимптотическая нормальность (б/д). Примеры	13
5	Методы нахождения оценок: метод подстановки, метод моментов и метод максимального правдоподобия. Примеры. Состоятельность и асимптотическая нормальность оценки по методу моментов. Функция плотности в дискретном случае.	19
6	Свойства оценки максимального правдоподобия: экстремальное свойство и состоятельность.	22
7	Способы сравнения оценок.	24
8	Эффективные оценки. Информация Фишера. Неравенство Рао-Крамера. Пример для бернуллиевского распределения.	25
9	Эффективность и асимптотическая эффективность оценки максимального правдоподобия.	28
10	Байесовские оценки: определение, типы байесовских оценок, наилучшая оценка в байесовском подходе, сопряженные распределения, примеры.	29
11	Достаточные статистики. Критерий факторизации. Теорема Блэкуэлла-Колмогорова-Рао.	31
12	Полные статистики. Оптимальные оценки. Алгоритм нахождения оптимальных оценок. Примеры.	33

13 Теорема об экспоненциальном семействе.	34
14 Доверительные интервалы, центральная функция, примеры нахождения. Асимптотические доверительные интервалы, алгоритм нахождения.	36
15 Задача линейной регрессии. Метод наименьших квадратов. Несмещенность оценки по методу наименьших квадратов. Матрица ковариаций оценки по методу наименьших квадратов. Оценка дисперсии. Свойства оценки по методу наименьших квадратов.	38
16 Гауссовская линейная модель. Хи-квадрат распределение. Теорема об ортогональном разложении гауссовского вектора. Оптимальность оценки.	40
17 Распределения Стьюдента и Фишера. Доверительные интервалы для параметров гауссовской линейной модели. Доверительная область для оценки МНК. Пример нахождения доверительных интервалов для параметров нормального распределения по выборке.	43
18 Проверка статистических гипотез. Основная гипотеза и альтернатива. Критерий. Ошибка первого и второго рода. Уровень значимости, размер критерия и функция мощности. Равномерно наиболее мощные критерии и несмещенные критерии. Лемма Неймана-Пирсона.	46
19 Монотонное отношение правдоподобий. Примеры. Другие способы нахождения наиболее мощных критериев.	49
20 Проверка гипотез в гауссовской линейной модели.	50
21 Состоятельность критерия. Критерий хи-квадрат Пирсона: доказательство сходимости к хи-квадрат закону и доказательство состоятельности.	51
22 Критерий Колмогорова-Смирнова (без доказательства), удобная формула для вычисления значения статистики. Критерий Крамера-Мизеса-Смирнова (без доказательства), удобная формула для вычисления значения статистики.	54

1 Вероятностно-статистическая модель. Наблюдение и выборка. Задача оценивания параметров. Статистики и оценки. Свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Взаимосвязь между свойствами оценок. Выборочные характеристики и порядковые статистики. Свойства выборочного среднего.

Напоминание

Определение Алгебра \mathcal{X} - семейство подмножеств $A \subset 2^X$, т.ч.

1. $\emptyset \in A$
2. $x, y \in A \Rightarrow x \cup y \in A$
3. $x \in A \Rightarrow X \setminus x \in A$

Определение Сигма-алгебра - алгебра, замкнутая относительно счетного объединения

Определение Борелевская сигма-алгебра - минимальная сигма алгебра, содержащая все открытые подмножества пространства (также содержит и все замкнутые)

Один из способов порождения: лучи $(-\infty, x]$

Определение Пусть (Ω, \mathcal{F}, P) - вероятностное пространство.

$\xi : \Omega \rightarrow \mathbb{R}^n$ — **случайный вектор**, если $\forall B \in \mathcal{B}(\mathbb{R}^n) \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F}$

Определение Выборка (X_1, \dots, X_n) - набор независимых одинаково распределенных величин (векторов)

Определение Все возможные исходы эксперимента (одного) образуют **выборочное пространство** \mathcal{X}

Определение Вероятностно-статистическая модель - тройка $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$, где

1. \mathcal{X} - выборочное пространство (как правило $\subset \mathbb{R}^n$)

2. $\mathcal{B}_{\mathcal{X}}$ - борелевская сигма-алгебра на \mathcal{X}
3. \mathcal{P} - семейство распределений на $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$

Положим $X(x) = x$ - **Наблюдение**

1. с одной стороны $x \in \mathcal{X}$ - числовая природа
2. с другой, x - реализация случайного вектора X

Наблюдение - вектор $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}) \rightarrow (\mathcal{X}, \mathcal{B}_{\mathcal{X}})$

Пусть $P \in \mathcal{P}$, X - случайная величина (вектор) на $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P)$

$$P_X(B) = P(X \in B) = P(x : X(x) \in B) = P(x : x \in B) = P(B)$$

Зададим корректно выборку в рамках некой вероятностно-статистической модели

$$(\mathcal{X}^n, \mathcal{B}_{\mathcal{X}}^n, P^n)$$

1. $\mathcal{X}^n = \mathcal{X} \times \dots \times \mathcal{X}$
2. $\mathcal{B}_{\mathcal{X}}^n = \sigma(\mathcal{B}_1 \times \dots \times \mathcal{B}_n, \mathcal{B}_i \in \mathcal{B}_{\mathcal{X}})$
3. $P^n = P_1 \otimes \dots \otimes P_n$ - продолжение прямого произведения мер с полукольца прямоугольников

$B = B_1 \times \dots \times B_n$ - элемент полукольца

$$P^n(B) = \prod_{i=1}^n P(B_i)$$

Тогда $X_i = X(x_1, \dots, x_n) = x_i$ - случайная величина с распределением P , причем X_1, \dots, X_n - независимы

Определение Пусть $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$ - вероятностно-статистическая модель, X - наблюдение на ней. (E, \mathcal{E}) - измеримое пространство. Пусть $S : \mathcal{X} \rightarrow E$ - измеримое отображение (прообраз любого измеримого из E множества измерим в \mathcal{X}). Тогда $S(X)$ - **статистика** от наблюдения X

Бывает так, что \mathcal{P} допускает параметризацию $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^n$

Если $S(X)$ - статистика со значениями в Θ , то S называется **оценкой** параметра θ

Свойства оценок

Оценка θ^* называется **несмещенной** оценкой параметра θ , если $\forall \theta \in \Theta : \mathbf{E}_\theta \theta^* = \theta$

Оценка θ^* называется **состоятельной** (точнее, последовательность оценок) оценкой параметра θ , если $\forall \theta \in \Theta : \theta^* \xrightarrow{P_\theta} \theta$

Оценка θ^* называется **сильно состоятельной** (точнее, последовательность оценок) оценкой параметра θ , если $\forall \theta \in \Theta : \theta^* \xrightarrow{P_\theta \text{ п.н.}} \theta$

Оценка θ^* называется **асимптотически нормальной** (точнее, последовательность оценок) оценкой параметра θ , если

$$\sqrt{n}(\theta^* - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta))$$

$\sigma^2(\theta)$ - **асимптотическая дисперсия**

Пусть X_1, \dots, X_n - выборка с распределением из параметризованного семейства \mathcal{P} . Пусть $\mathbf{E}_\theta X_1 = \theta \forall \theta \in \Theta$

Тогда оценка $\frac{1}{n} \sum_{i=1}^n X_i$ - является несмещенной (следует из линейности математического ожидания), сильно состоятельной (следует из УЗБЧ) и асимптотически нормальной (следует из ЦПТ) оценкой параметра θ

Заметим, что из сильной состоятельности следует состоятельность (т.к. сходимость п.н. влечет сходимость по вероятности)

Также, из асимптотической нормальности следует состоятельность оценки (следует из леммы Слуцкого, примененной к $\xi_n = \sqrt{n}(\theta^* - \theta)$ и $\eta_n = \frac{1}{\sqrt{n}}$, а также того факта, что сходимость по распределению к константе влечет сходимость по вероятности)

Пусть $g(x)$ - борелевская функция на \mathbb{R} Тогда

$$\overline{g(X)} = \frac{1}{n} \sum_{i=1}^n g(X_i)$$

называется **выборочная характеристика** от функции g

Пример выборочной статистики - выборочное среднее.

Также, рассматриваются функции от выборочных статистик:

$$S(X) = h(g_1(X), \dots, g_k(X))$$

где все функции - борелевские.

Пример: выборочная дисперсия $s^2 = \overline{X^2} - (\overline{X})^2$

Замечание: $s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (такая функция называется выборочным 2 моментом)

Пусть X_1, \dots, X_n - выборка. Упорядочим её элементы по неубыванию. Полученный ряд

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

называется **вариационным рядом**, а k -й член вариационного ряда называется **k -й порядковой статистикой**

Замечание: пусть выборка размера n из распределения P , с функцией распределения F и плотностью p . Тогда плотность k -й порядковой статистики вычисляется по формуле:

$$p_{X_{(k)}}(x) = n C_{n-1}^{k-1} p(x) F^{k-1}(x) (1 - F(x))^{n-k}$$

2 Теорема о наследовании сходимостей. Лемма Слуцкого с доказательством и ее применение. Наследование асимптотической нормальности. Многомерный случай. Наследование состоятельности.

Напоминание

Виды сходимостей случайных векторов

1. $\xi_n \xrightarrow{п.н.} \xi$, если $P(\omega : \lim \xi_n(\omega) = \xi(\omega)) = 1$
2. $\xi_n \xrightarrow{P} \xi$, если $\forall \varepsilon > 0 \ P(\|\xi_n - \xi\|_2 > \varepsilon) \rightarrow 0$
3. $\xi_n \xrightarrow{L_p} \xi$, если $\mathbf{E}(\|\xi_n - \xi\|_p)^p \rightarrow 0$
4. $\xi_n \xrightarrow{d} \xi$, если для любой непрерывной ограниченной функции f : $\mathbf{E}(f(\xi_n)) \rightarrow \mathbf{E}(f(\xi))$

Взаимосвязь

Из сходимости почти наверное следует сходимость по вероятности

Из сходимости в L_p следует сходимость по вероятности

Из сходимости по вероятности следует сходимость по распределению

Теорема(наследование сходимостей)

- 1) Пусть $\xi_n \xrightarrow{\text{п.н.}} \xi$ - случайные векторы размера m
 Пусть $h : \mathbb{R}^m \rightarrow \mathbb{R}^k$ - функция, непрерывная почти всюду относительно распределения ξ (т.е. $\exists B \in \mathcal{B}(\mathbb{R}^m)$ т.ч. $P(\xi \in B) = 1$ и h непрерывна на B .
 Тогда $h(\xi_n) \xrightarrow{\text{п.н.}} h(\xi)$
- 2) Аналогичное верно, если заменить сходимость п.н. на сходимость по вероятности
- 3) Пусть $\xi_n \rightarrow \xi$ по распределению - случайные векторы размера m
 Пусть $h : \mathbb{R}^m \rightarrow \mathbb{R}^k$ - функция, непрерывная на множестве значений ξ
 Тогда $h(\xi_n) \xrightarrow{d} h(\xi)$.

▲

- 1) $P(\lim h(\xi_n) = h(\xi)) \geq P(\{\lim \xi_n = \xi\} \cap B) = 1$ (т.к. справа такие точки, что они принадлежат B и на них сходится последовательность сл. в. из непрерывности h на B будет следовать сходимость. Пересечение событий с вероятностью 1 дает вероятность 1)
- 2) Предположим, что это не так. Тогда существует подпоследовательность $\{\xi_{n_m}\}$ т.ч. $\exists \varepsilon > 0 \exists \varepsilon_1 > 0 :$

$$P(\|h(\xi_{n_m}) - h(\xi)\|_2 > \varepsilon) > \varepsilon_1$$

Из последовательности $\xi_{n_m} \rightarrow \xi$ сходящуюся по вероятности можно выбрать подпоследовательность $\{\xi_{n_{m_r}}\}$ сходящуюся п.н.
 $\Rightarrow h(\xi_{n_{m_r}}) \xrightarrow{\text{п.н.}} h(\xi)$ по пункту 1. Противоречие

- 3) $\xi_n \xrightarrow{d} \xi \Leftrightarrow \forall f$ - непрерывной и ограниченной $\mathbf{E}(f(\xi_n)) \rightarrow \mathbf{E}(f(\xi)) \Rightarrow \forall g$ $\mathbf{E}(g(h(\xi_n))) \rightarrow \mathbf{E}(g(h(\xi)))$ т.к. $g(h)$ - непрерывная и ограниченная

▲

Лемма(Слуцкого)

Пусть последовательности случайных величин $\{\xi_n\}$ и $\{\eta_n\}$ тако-

вы, что $\xi_n \xrightarrow{d} \xi$ и $\eta_n \xrightarrow{d} C$

Тогда $\xi_n + \eta_n \xrightarrow{d} \xi + C$ и $\xi_n \eta_n \xrightarrow{d} \xi C$

▲

(В доказательстве используется другое, эквивалентное определение сходимости по распределению: последовательность функций распределения стремится к предельной функции распределения для любой точки непрерывности предельной функции распределения)

Пусть t - точка непрерывности $F_{\xi+\eta}$ пусть $F_{\xi+\eta}$ непрерывна в точках $t \pm \varepsilon$

$$\begin{aligned} F_{\xi_n+\eta_n}(t) &= P(\xi_n+\eta_n \leq t) = P(\xi_n+\eta_n \leq t, \eta_n < C-\varepsilon) + P(\xi_n+\eta_n \leq t, \eta_n \geq C-\varepsilon) \\ &\leq P(\eta_n < C) + P(\xi_n \leq t - C + \varepsilon) = P(\eta_n < C) + P(\xi_n + C \leq t + \varepsilon) \end{aligned}$$

Первая вероятность стремится к 0 (сходимость по распределению к константе влечет сходимость по вероятности)

Из теоремы о наследовании: $\xi_n + C \rightarrow \xi + C$

$$\Rightarrow P(\xi_n + C \leq t + \varepsilon) = F_{\xi_n+C}(t + \varepsilon) \rightarrow F_{\xi+C}(t + \varepsilon)$$

$$\Rightarrow P(\eta_n < C) + P(\xi_n + C \leq t + \varepsilon) \rightarrow F_{\xi+C}(t + \varepsilon)$$

$$\begin{aligned} F_{\xi_n+\eta_n}(t) &= P(\xi_n+\eta_n \leq t) = P(\xi_n+\eta_n \leq t, \eta_n > C+\varepsilon) + P(\xi_n+\eta_n \leq t, \eta_n \leq C+\varepsilon) \\ &\geq P(\xi_n + \eta_n \leq t, \eta_n \leq C + \varepsilon) \geq P(\xi_n + C + \varepsilon \leq t, \eta_n \leq C + \varepsilon) \\ &= P(\xi_n + C + \varepsilon \leq t) - P(\xi_n + C + \varepsilon \leq t, \eta_n > C + \varepsilon) \\ &\geq P(\xi_n + C + \varepsilon \leq t) - P(\eta_n > C + \varepsilon) \end{aligned}$$

Второе слагаемое, аналогично, стремится к 0. первое слагаемое ($= F_{\xi_n+C}(t - \varepsilon)$) аналогично стремится к $F_{\xi+C}(t - \varepsilon)$

Значит:

$$F_{\xi+C}(t - \varepsilon) \leq \underline{\lim} F_{\xi_n+\eta_n} \leq \overline{\lim} F_{\xi_n+\eta_n} \leq F_{\xi+C}(t + \varepsilon)$$

Получается $\lim F_{\xi_n+\eta_n}(t) = F_{\xi+\eta}(t)$

Доказательство для произведения аналогично.

▲

Применение леммы Slutsky: доказательство факта, что из асимптотической нормальности следует сильная состоятельность (билет 1)

Теорема(δ -метод)

Пусть $\xi_n \xrightarrow{d} \xi$, числовая последовательность $b_n \rightarrow 0$, $b_n \neq 0$, функция $h : \mathbb{R} \rightarrow \mathbb{R}$ - дифференцируема в точке a . Тогда

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

▲

Определим функцию $H(x) = \frac{h(a+x)-h(a)}{x}$ и доопределим в 0 значением $h'(a)$. $H(x)$ непрерывна в точке 0.

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

Воспользуемся теоремой о наследовании сходимостей:

$$\frac{h(a + \xi_n b_n) - h(a)}{\xi_n b_n} = H(\xi_n b_n) \xrightarrow{P} H(0) = h'(a)$$

Значит, по лемме Slutsky:

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

▲

Следствие(наследование асимптотической нормальности)

Пусть θ^* - асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$, функция τ непрерывно дифференцируема на $\Theta \subset \mathbb{R}$.

Тогда $\tau(\theta^*)$ - асимптотически нормальная оценка параметра $\tau(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)[\tau'(\theta)]^2$

▲

$$\xi_n = \sqrt{n}(\theta^* - \theta), \quad b_n = \frac{1}{\sqrt{n}}, \quad a = \theta, \quad h = \tau$$

Осталось воспользоваться определением асимптотической нормальности и δ -методом.

▲

Утверждение

δ -метод работает и в многомерном случае: производная заменяется на матрицу частных производных.

Следствие(многомерный случай)

В многомерном случае на функцию τ накладывается ограничение на существование матрицы частных производных в каждой точке $\theta \in \Theta$. Тогда, если θ^* - асимптотически нормальная оценка

θ с асимптотической матрицей ковариации $\Sigma(\theta)$, то $\tau(\theta^*)$ - асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $(\tau'(\theta))^T \Sigma(\theta) \tau'(\theta)$

Утверждение(наследование состоятельности)

Если θ^* состоятельная (сильно состоятельная) оценка θ и функция τ непрерывна во всех точках $\theta \in \Theta$, то $\tau(\theta^*)$ - состоятельная (сильно состоятельная) оценка $\tau(\theta)$

▲

Утверждение - тривиальное следствие из теоремы о наследовании сходимостей.

▲

3 Непараметрическая задача. Эмпирическое распределение и эмпирическая функция распределения. Теорема Гливенко-Кантелли.

Определение Пусть X_1, \dots, X_n - выборка, $B \in \mathcal{B}(\mathbb{R})$
эмпирическое распределение:

$$P_n^*(B) = \frac{\sum_{i=1}^n I(X_i \in B)}{n}$$

По УЗБЧ: $P_n^*(B) \xrightarrow{\text{п.н.}} P(B)$

А значит, **эмпирическая функция распределения**

$F_n^*(x) = P_n^*((-\infty, x]) \xrightarrow{\text{п.н.}} F(x)$ (теорема Гливенко)

Теорема(Гливенко-Кантелли)

Пусть X_1, \dots, X_n - выборка с функцией распределения $F(x)$

Определим $D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|$

Тогда $D_n \xrightarrow{\text{п.н.}} 0$

▲

$F_n^*(x, \omega)$ - функция распределения \Rightarrow непрерывная справа \Rightarrow функция $\forall \omega |F_n^*(x, \omega) - F(x)|$ непрерывна справа \Rightarrow в D_n супремум можно брать не по всем действительным точкам, а по всем рациональным (потому что действительные точки можем приблизить

рациональными) \Rightarrow

$$D_n = \sup_{x \in \mathbb{Q}} |F_n^*(x, \omega) - F(x)|$$

является случайной величиной, как супремум счетного числа случайных величин.

Фиксируем $N \in \mathbb{N}$, положим $\forall k = \overline{1, N-1}$:

$x_k = \inf\{x \in \mathbb{R} : F(x) \geq \frac{k}{N}\}$ (такая величина называется квантиль)

$x_0 := -\infty, x_N = +\infty$

Пусть $x \in [x_k, x_{k+1})$, тогда

$$F_n^*(x, \omega) - F(x) \leq F_n^*(x_{k+1} - 0) - F(x_k) = (F_n^*(x_{k+1} - 0) - F(x_{k+1} - 0)) + (F(x_{k+1} - 0) - F(x_k))$$

во второй скобке уменьшаемое $\leq \frac{k+1}{N}$, а вычитаемое $\geq \frac{k}{N}$, значит разность $\leq \frac{1}{N}$

\Rightarrow

$$F_n^*(x, \omega) - F(x) \leq F_n^*(x_{k+1} - 0) - F(x_{k+1} - 0) + \frac{1}{N}$$

Аналогично

$$F_n^*(x, \omega) - F(x) \geq F_n^*(x_k) - F(x_k) - \frac{1}{N}$$

Значит

$$\forall x \in \mathbb{R} |F_n^*(x, \omega) - F(x)| \leq \max_{0 \leq k \leq N-1, 0 \leq l \leq N-1} \{|F_n^*(x_k) - F(x_k)| + |F_n^*(x_{l+1} - 0) - F(x_{l+1} - 0)|\} + \frac{1}{N}$$

ввиду сходимости эмпирической ф.р к истинной п.н. каждый член $|F_n^*(x_k) - F(x_k)| + |F_n^*(x_{l+1} - 0) - F(x_{l+1} - 0)|$ стремится к нулю. Соответственно, $\forall \varepsilon > 0$ выберем такое N , что $\frac{1}{N} < \varepsilon$

Это и будет означать, что $D_n \xrightarrow{\text{п.н.}} 0$

▲

4 Выборочные квантили. Асимптотическая нормальность выборочной квантили. Медиана, выборочная медиана и ее асимптотическая нормальность (б/д). Примеры

Определение Пусть $F(x)$ - функция распределения на \mathbb{R} . $\gamma \in [0, 1]$.

γ -квантиль функции распределения F - это $x_\gamma = \inf_{x \in \mathbb{R}} \{F(x) \geq \gamma\}$

Если функция распределения непрерывна, то $F(x_\gamma) = \gamma$

Если она еще и строго монотонная, то x_γ единственная.

Определение Пусть X_1, \dots, X_n - выборка. $X_{(1)}, \dots, X_{(n)}$ - вариационный ряд.

Тогда **выборочная квантиль уровня γ** -

$$z_\gamma^n = \begin{cases} X_{([n\gamma]+1)} & n\gamma \notin \mathbb{N} \\ X_{([n\gamma])} & n\gamma \in \mathbb{N} \end{cases}$$

Теорема (о выборочной квантили)

Пусть X_1, \dots, X_n - выборка из распределения P , с функцией распределения F и плотностью f .

Пусть z_p - p -квантиль функции распределения F . Пусть f непрерывно дифференцируема в окрестности точки z_p и $f(z_p) > 0$

Тогда $\sqrt{n}(X_{([np]+1)} - z_p) \xrightarrow{d} \mathcal{N}(0, \frac{p(1-p)}{f^2(z_p)})$

◀

Начнем с того, что выведем плотность k -й порядковой статистики (билет 1)

Заметим, что если $X_{(k)} \leq x$ то хотя бы k элементов выборки не больше x

Если $F_k(x)$ - функция распределения k -й порядковой статистики, то

$$F_k(x) = P(X_{(k)} \leq x) = \sum_{i=k}^n C_n^i F^i(x) (1 - F(x))^{n-i}$$

Продифференцируем $F_k(x)$ чтобы получить плотность $f_k(x)$

$$f_k(x) = \sum_{i=k}^n i C_n^i F^{i-1}(x) (1 - F(x))^{n-1} f(x) - \sum_{i=k}^n (n-i) C_n^i F^i(x) (1 - F(x))^{n-i-1} f(x)$$

Заметим, что

$$iC_n^i = \frac{n!}{(i-1)!(n-i)!} = nC_{n-1}^{i-1}$$

$$(n-i)C_n^i = \frac{n!}{i!(n-i-1)!} = nC_{n-1}^i$$

Тогда, заменив индекс суммирования в первой сумме на $s = i - 1$ получим

$$\sum_{s=k-1}^n nC_{n-1}^s F^s(x)(1-F(x))^{n-s-1} f(x) - \sum_{i=k}^n (n-i)C_n^i F^i(x)(1-F(x))^{n-i-1} f(x)$$

$$= nC_{n-1}^{k-1} F^{k-1}(x)(1-F(x))^{n-k} f(x)$$

Теперь покажем, что если мы возьмем последовательность случайных величин T_n , построенных по правилу

$$T_n = \frac{f(z_p)\sqrt{n}}{\sqrt{p(1-p)}}(X_{(k)} - z_p)$$

где $k = [np] + 1$

Получим $T_n \xrightarrow{d} \mathcal{N}(0, 1)$

Докажем один вспомогательный факт:

Если ξ - случайная величина с плотностью f_ξ , то случайная величина $\eta = a\xi + b$, $a > 0$, b - константы, имеет плотность $f_\eta = \frac{1}{a}f_\xi(\frac{x-b}{a})$

Это несложно понять, т.к.

$$F_\eta(x) = P(\eta \leq x) = P(a\xi + b \leq x) = P(\xi \leq \frac{x-b}{a}) = F_\xi(\frac{x-b}{a})$$

Осталось продифференцировать это равенство, чтобы получить требуемое.

Пусть $q_n(x)$ - плотность T_n . Тогда

$$q_n(x) = \frac{\sqrt{p(1-p)}}{f(z_p)\sqrt{n}} f_k(z_p + \frac{x\sqrt{p(1-p)}}{f(z_p)\sqrt{n}})$$

Обозначим $t_n = z_p + \frac{x\sqrt{p(1-p)}}{f(z_p)\sqrt{n}}$

Тогда

$$q_n(x) = nC_{n-1}^{k-1} f(t_n) F(t_n)^{k-1} (1 - F(t_n))^{n-k} \frac{\sqrt{p(1-p)}}{f(z_p)\sqrt{n}}$$

Обозначим $q_n(x) = A_1(n)A_2(n)A_3(n)$, где

$$\begin{aligned} A_1(n) &= \frac{f(t_n)}{f(z_p)} \\ A_2(n) &= nC_{n-1}^{k-1} \sqrt{\frac{p(1-p)}{n}} p^{k-1} (1-p)^{n-k} \\ A_3(n) &= \left(\frac{F(t_n)}{p}\right)^{k-1} \left(\frac{1-F(t_n)}{1-p}\right)^{n-k} \end{aligned}$$

Найдем пределы всех 3х выражений при $n \rightarrow +\infty$

A_1 :

Заметим, что $t_n \rightarrow z_p$ при $n \rightarrow +\infty$ и $f(x)$ непрерывна в окрестности z_p . Тогда

$$\frac{f(t_n)}{f(z_p)} \rightarrow 1 \Rightarrow A_1(n) \rightarrow 1, \quad n \rightarrow +\infty$$

A_2 :

Заметим, что

$$A_2(n) = kC_n^k \sqrt{\frac{p(1-p)}{n}} p^{k-1} (1-p)^{n-k}$$

Применив формулу Стирлинга

$$A_2(n) \sim k \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \sqrt{2\pi(n-k)} \left(\frac{n-k}{e}\right)^{n-k}} \sqrt{\frac{p(1-p)}{n}} p^{k-1} (1-p)^{n-k}$$

Учитывая, что $k = [np] + 1 \Rightarrow k \sim np$ и $n - k \sim n(1 - p)$

$$A_2(n) \sim np \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi np} \left(\frac{k}{e}\right)^k \sqrt{2\pi n(1-p)} \left(\frac{n-k}{e}\right)^{n-k}} \sqrt{\frac{p(1-p)}{n}} p^{k-1} (1-p)^{n-k}$$

Получим

$$A_2(n) \sim \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}$$

Теперь покажем, что $A_2(n) \rightarrow \frac{1}{\sqrt{2\pi}}$

$$A_2(n) \sim \frac{1}{\sqrt{2\pi}} \exp\left(k \ln \frac{np}{k} + (n-k) \ln \frac{n(1-p)}{n-k}\right)$$

Далее, k отличается от np не более, чем на 1. Значит, можем разложить логарифмы в ряд Тейлора в нуле

$$\ln \frac{np}{k} = \frac{np}{k} - 1 + O\left(\frac{1}{k^2}\right) = \frac{np-k}{k} + O\left(\frac{1}{n^2}\right)$$

$$\ln \frac{n(1-p)}{n-k} = \frac{n(1-p)}{n-k} - 1 + O\left(\frac{1}{(n-k)^2}\right) = \frac{k-np}{n-k} + O\left(\frac{1}{n^2}\right)$$

Значит

$$A_2(n) \sim \frac{1}{\sqrt{2\pi}} \exp\left(np - k + k - np + O\left(\frac{1}{n^2}\right)\right) \rightarrow \frac{1}{\sqrt{2\pi}}, n \rightarrow +\infty$$

A_3 :

Заметим, что $F(t_n) \rightarrow F(z_p) = p$ при $n \rightarrow +\infty$ Из этого делаем вывод, что

$$\left(\frac{F(t_n)}{p}\right)^{k-1} \sim \left(\frac{F(t_n)}{p}\right)^k$$

Далее

$$A_3(n) \sim \exp \left(k \ln \frac{F(t_n)}{p} + (n-k) \ln \frac{1-F(t_n)}{1-p} \right)$$

Теперь разложим $F(t_n)$ в ряд Тейлора в точке z_p , пользуясь непрерывной дифференцируемостью f

$$F(t_n) = F(z_p) + (t_n - z_p)f'(z_p) + \frac{1}{2}(t_n - z_p)^2 f''(z_p) + o((t_n - z_p)^2)$$

$$= p + \sqrt{\frac{p(1-p)}{n}}x + \frac{f'(z_p)}{f^2(z_p)} \frac{p(1-p)}{2n}x^2 + o((t_n - z_p)^2)$$

Тогда

$$\frac{F(t_n)}{p} = 1 + \sqrt{\frac{1-p}{np}}x + \frac{f'(z_p)}{f^2(z_p)} \frac{1-p}{2n}x^2 + o\left(\frac{1}{n}\right)$$

Если взять логарифм, то его можно разложить в ряд Тейлора

$$\ln \frac{F(t_n)}{p} = \sqrt{\frac{1-p}{np}}x + \frac{f'(z_p)}{f^2(z_p)} \frac{1-p}{2n}x^2 + o\left(\frac{1}{n}\right) - \frac{1}{2} \frac{1-p}{np}x^2 + o\left(\frac{1}{n}\right)$$

Тогда

$$k \ln \frac{F(t_n)}{p} = \sqrt{\frac{1-p}{np}}xk + \frac{f'(z_p)}{f^2(z_p)} \frac{k(1-p)}{2n}x^2 - \frac{1}{2} \frac{k(1-p)}{np}x^2 + o(1)$$

Аналогично

$$(n-k) \ln \frac{1-F(t_n)}{1-p} = -\sqrt{\frac{p}{n(1-p)}}x(n-k) - \frac{f'(z_p)}{f^2(z_p)} \frac{p(n-k)}{2n}x^2 - \frac{1}{2} \frac{p}{1-p} \frac{n-k}{n}x^2 + o(1)$$

Теперь воспользуемся тем, что $|k - np| \leq 1$, т.е. $k = np + O\left(\frac{1}{n}\right)$

$$k \ln \frac{F(t_n)}{p} = \dots = x\sqrt{np(1-p)} + \frac{p(1-p)}{2} \frac{f'(z_p)}{f^2(z_p)}x^2 - \frac{1-p}{2}x^2 + o(1)$$

$$(n-k) \ln \frac{1-F(t_n)}{1-p} = \dots = -x \sqrt{np(1-p)} - \frac{p(1-p)}{2} \frac{f'(z_p)}{f^2(z_p)} x^2 - \frac{p}{2} x^2 + o(1)$$

Следовательно

$$A_3(n) \sim \exp\left(-\frac{x^2}{2} + o(1)\right) \rightarrow \exp\left(-\frac{x^2}{2}\right)$$

В итоге получили, что для всех x

$$\lim_{n \rightarrow +\infty} q_n(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Это означает, что $q_n(x)$ будет равномерно сходиться к плотности $\mathcal{N}(0, 1)$ на любом компакте. Из равномерной сходимости на отрезке $[a, b]$ следует, что

$$\lim_{n \rightarrow +\infty} (F_{T_n}(b) - F_{T_n}(a)) = \lim_{n \rightarrow +\infty} \int_a^b q_n(x) dx = \Phi(b) - \Phi(a)$$

Теперь нужно доказать, что $F_{T_n} \rightarrow \Phi(x)$ для всех x , где $\Phi(x)$ - стандартное нормальное распределение. Заметим, что для любого $a < b$

$$\overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(b) - \Phi(b)| \leq \overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(b) - F_{T_n}(a) + \Phi(b) - \Phi(a)| + \overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(a) - \Phi(a)|$$

Первый предел уходит в 0. Тогда

$$\overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(b) - \Phi(b)| \leq \overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(a) - \Phi(a)|$$

$$\overline{\lim}_{n \rightarrow +\infty} |F_{T_n}(a) - \Phi(a)| \leq \overline{\lim}_{n \rightarrow +\infty} (F_{T_n}(a) + \Phi(a)) = \Phi(a) + \overline{\lim}_{n \rightarrow +\infty} F_{T_n}(a)$$

Пользуясь тем, что $F_{T_n} < 1$

$$\Phi(a) + \overline{\lim}_{n \rightarrow +\infty} F_{T_n}(a) \leq \Phi(a) + \overline{\lim}_{n \rightarrow +\infty} (F_{T_n}(a) + F_{T_n}(-a) + 1) = \Phi(a) + \Phi(a) - \Phi(-a) + 1$$

Но эту сумму можно сделать сколь угодно малой, устремив a к $-\infty$. Значит $T_n \xrightarrow{d_\theta} \mathcal{N}(0, 1)$, что и требовалось



Определение Медиана - квантиль уровня $\frac{1}{2}$

Определение Выборочная медиана

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

Теорема (выборочная медиана, б/д)

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

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

Ц.П.Т: $\sqrt{n}(\bar{X} - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$

Теорема о выборочной медиане: $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, \frac{\pi}{2})$

Пример Распределение Коши

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

По теореме $\sqrt{n}(\hat{\mu} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{\pi^2}{4})$

5 Методы нахождения оценок: метод подстановки, метод моментов и метод максимального правдоподобия. Примеры. Состоятельность и асимптотическая нормальность оценки по методу моментов. Функция плотности в дискретном случае.

Метод подстановки

Пусть $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ таково, что для некоторого функционала G выполнено $\forall \theta \in \Theta \ G(P_\theta) = \theta$

Тогда оценка θ по методу подстановки называется $\theta^*(X_1, \dots, X_n) = G(P_n^*)$, где P_n^* - эмпирическое распределение, построенное по выборке X_1, \dots, X_n

Пример

1) $\mathcal{P} = \{\mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$, функционал - математическое ожидание.
 $\mathbf{E}(P_\theta) = \theta$

$$\mathbf{E}(P_n^*) = \int_{\mathbb{R}} x dP_n^* = \int_{\mathbb{R}} x dF_n^* = \sum_{i=1}^n X_i \frac{1}{n} = \bar{X}$$

2) Пусть $\mathcal{P} = \{\exp(\theta), \theta > 0\}$

$$F_\theta(x) = 1 - e^{\theta x} I\{x \geq 0\}$$

Тогда $G(F_\theta(x)) = -\ln(1 - F_\theta(1))$ т.ч. $\forall \theta \in \Theta: G(F_\theta(x)) = \theta$

Пусть P_n^* - эмпирическое распределение, тогда $G(P_n^*)$ - оценка θ методом подстановки:

$$G(P_n^*) = -\ln\left(1 - \frac{\sum_{i=1}^n I(X_i \leq 1)}{n}\right)$$

Метод моментов

Пусть X_1, \dots, X_n - выборка из неизвестного распределения $P \in \{P_\theta, \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$

Пусть борелевские функции $g_1(x), \dots, g_k(x)$ таковы, что набор значений

$m(\theta) = (m_1(\theta), \dots, m_k(\theta)) = (\mathbf{E}_\theta g_1(X_1), \dots, \mathbf{E}_\theta g_k(X_1))$ однозначно определяют параметр θ

Определение оценка параметра θ по **методу моментов** с пробными функциями $g_1(X), \dots, g_k(X)$ называется решение системы уравнений

$$\begin{cases} m_1(\theta^*) = \overline{g_1(X)} \\ \dots \\ m_k(\theta^*) = \overline{g_k(X)} \end{cases} \quad \text{Относительно } \theta^*$$

Иначе говоря $\theta^*(X_1, \dots, X_n) = m^{-1}(\overline{g_1(X)}, \dots, \overline{g_k(X)})$

Замечание: на практике обычно используются **стандартные** пробные функции: $g_i(x) = x^i$

Теорема (состоятельность ОММ)

Пусть $\theta \in \Theta \subset \mathbb{R}^k$. $\theta_n^* = m^{-1}(\overline{g_1(X)}, \dots, \overline{g_k(X)})$ - оценка θ по методу моментов. Пусть m биективно. Если m^{-1} непрерывна на множестве $m(\Theta)$ и $\forall i \mathbf{E}_\theta |g_i(X_1)| < +\infty$, то θ_n^* - сильно состоятельная оценка параметра θ

▲
 $\forall i = \overline{1, k} : \overline{g_i(X)} = \frac{1}{n} \sum_{m=1}^n g_i(X_m) \xrightarrow{P_{\theta \text{ п. н.}}} \mathbf{E}_{\theta} g_i(X_1) = m_i(\theta)$ (по
 УЗБЧ)

По теореме о наследовании сходимостей:

$$\theta_n^* = m^{-1}(\overline{g_1(X)}, \dots, \overline{g_k(X)}) \xrightarrow{P_{\theta \text{ п. н.}}} m^{-1}(m_1(\theta), \dots, m_k(\theta)) = m^{-1}(m(\theta)) = \theta$$

▲
Теорема (асимптотическая нормальность)
 $(k = 1)$ Если $\forall \theta \in \Theta \exists \mathbf{E}(g_1(x_1))^2 < \infty$ и m^{-1} дифференцируема,
 то по теореме о наследовании асимптотической нормальности
 θ_n^* - асимптотически нормальная с асимптотической дисперсией
 $([m^{-1}]' \mathbf{E}_{\theta} g_1(x_1))^2 \mathbf{D}_{\theta} g_1(x_1)$

Плотность дискретного распределения

Определение Считаящей мерой μ на \mathbb{Z} называется отображение

$\mu : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{Z} \cup \{+\infty\}$ Определенное по правилу

$$\mu(B) = \sum_{k \in \mathbb{Z}} I(k \in B)$$

Определение Интегралом от $f(x)$ по считающей мере называется $\int_{\mathbb{R}} f(x) \mu(dx) = \sum_{k \in \mathbb{Z}} f(k)$ при условии, что ряд сходится абсолютно

Замечание: такой интеграл обладает всеми свойствами интеграла Лебега (он им и является)

Определение Пусть ξ - дискретная случайная величина на \mathbb{Z}
 Её **плотностью** относительно считающей меры μ называется
 $p(x) = P(\xi = x), x \in \mathbb{Z}$

Корректность определения: ξ - дискретная случайная величина с плотностью $p(x) \Rightarrow \forall g(x) \mathbf{E} g(\xi) = \int g(x) p(x) d\mu = \sum_{k \in \mathbb{Z}} g(k) P(\xi = k)$

Определение Пусть X - наблюдение с неизвестным распределением $P \in \{P_{\theta}, \theta \in \Theta\}$, причем $\forall \theta \in \Theta$ у P_{θ} если плотность p_{θ} по одной и той же мере μ (либо Лебега, либо считающей). В этом случае семейство $\{P_{\theta}, \theta \in \Theta\}$ называется **доминируемым** относительно меры μ

Метод максимального правдоподобия

Определение Пусть X - наблюдение из неизвестного распределения $P \in \{P_{\theta}, \theta \in \Theta\}$ - семейство распределений, доминируемое относительно меры μ

Функцией правдоподобия $f_{\theta}(x)$ называется случайная вели-

чина $f_\theta(x) = p_\theta(x)$

Пример если $X = (X_1, \dots, X_n)$ - выборка с плотностью $p_\theta(x)$, то $f_\theta(X) = \prod_{i=1}^n p_\theta(X_i)$

Определение Пусть X - наблюдение с функцией правдоподобия $f_\theta(X)$ **Оценкой параметра θ по методу максимального правдоподобия** называется $\theta^* = \arg \max_{\theta \in \Theta} f_\theta(X)$

Определение $L_\theta(X) = \ln f_\theta(X)$ — **логарифмическая функция правдоподобия**

6 Свойства оценки максимального правдоподобия: экстремальное свойство и состоятельность.

Условия Регулярности модели

1) $\{P_\theta, \theta \in \Theta\}$ - семейство распределений, доминируемое относительно меры μ , $P_{\theta_1} \neq P_{\theta_2}$ при $\theta_1 \neq \theta_2$

2) $\mathcal{A} = \{x \in \mathcal{X}, p_\theta(x) > 0\}$ не зависит от θ

3) Наблюдение X - выбока из $P \in \{P_\theta, \theta \in \Theta\}$

Теорема (экстремальное свойство правдоподобия)

В условиях регулярности 1-3 $\forall \theta_0, \theta \in \Theta, \theta_0 \neq \theta$ выполнено:

$$P_{\theta_0}(f_{\theta_0}(X_1 \dots X_n) > f_\theta(X_1 \dots X_n)) \xrightarrow{n \rightarrow +\infty} 1$$

▲

$$f_{\theta_0}(X_1 \dots X_n) > f_\theta(X_1 \dots X_n) \Leftrightarrow \frac{1}{n} \ln \frac{f_\theta(X)}{f_{\theta_0}(X)} < 0$$

т.к. $f_\theta(X) = \prod_{i=1}^n p_\theta(X_i) \Rightarrow$ по УЗБЧ

$$\frac{1}{n} \ln \frac{f_\theta(X)}{f_{\theta_0}(X)} = \frac{1}{n} \sum_{i=1}^n \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} \xrightarrow{P_{\theta \text{ п.п.}}} \mathbf{E}_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)}$$

Тогда $P_{\theta_0}(f_{\theta_0}(X_1 \dots X_n) > f_\theta(X_1 \dots X_n)) \rightarrow P_{\theta_0}(\mathbf{E}_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} < 0)$

Вероятность справа 0 или 1. Покажем, что 1

$$\mathbf{E}_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \int_{\mathcal{A}} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) \mu(dx)$$

$\ln \frac{p_\theta(x)}{p_{\theta_0}(x)} = \ln(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1)$ и воспользуемся тем, что

$$\ln(1+x) \leq x, \quad x > -1$$

$$\int_{\mathcal{A}} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) \mu(dx)$$

$$\leq \int_{\mathcal{A}} \left(\frac{p_{\theta}(x)}{p_{\theta_0}(x)} - 1 \right) p_{\theta_0}(x) \mu(dx) = \int_{\mathcal{A}} p_{\theta}(x) \mu(dx) - \int_{\mathcal{A}} p_{\theta_0} \mu(dx) = 0$$

$$\int_{\mathcal{A}} p_{\theta}(x) \mu(dx) - \int_{\mathcal{A}} p_{\theta_0} \mu(dx) = 0$$

т.к. это плотности, и интегралы равны 1

Когда возможно равенство?

только когда $\frac{p_{\theta}(x)}{p_{\theta_0}(x)} = 1$, что противоречит 1му условию регулярности.

\Rightarrow неравенство строгое

▲

Условия регулярности (продолжение)

4) Θ - открытый интервал на \mathbb{R}

5) Плотность $p_{\theta}(x)$ - непрерывно дифференцируема по θ при $x \in \mathcal{A}$

Теорема

В условиях регулярности, если уравнение $\frac{\partial}{\partial \theta} f_{\theta} = 0$ имеет решение, то одно из его решений является состоятельной оценкой θ

◀

$\theta_0 \in \Theta, \varepsilon > 0$ т.ч. $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subset \Theta$

$P_{\theta_0} (f_{\theta_0}(X_1, \dots, X_n) > \max\{f_{\theta_0-\varepsilon}(X_1, \dots, X_n), f_{\theta_0+\varepsilon}(X_1, \dots, X_n)\}) \rightarrow 1$

$P_{\theta_0} \left(\text{на } (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \exists \text{ решение уравнения } \frac{\partial}{\partial \theta} f_{\theta} = 0 \right) \rightarrow 1$

\Rightarrow

$P_{\theta_0} (\text{ближайшее к } \theta_0 \text{ решение лежит в } (\theta_0 - \varepsilon, \theta_0 + \varepsilon)) \rightarrow 1$

Пусть θ^* ближайшее к θ_0 решение $\Rightarrow \theta^* \xrightarrow{P_{\theta_0}} \theta_0$

►

Следствие (состоятельность ОМП)

Пусть в условиях предыдущей теоремы $\forall n \forall X_1, \dots, X_n \exists!$ решение θ_n^* уравнения правдоподобия.

Тогда θ_n^* - состоятельная оценка параметра θ и с вероятностью, стремящейся к 1, является ОМП. (т.е. решение уравнения правдоподобия - максимум)

▲

Состоятельность следует из предыдущей теоремы.

Пусть θ - истинное значение, $[\theta - a, \theta + a] \subset \Theta \Rightarrow$ с вероятностью, стремящейся к 1, $f_\theta(X)$ достигает максимума на отрезке $[\theta - a, \theta + a]$ в точке θ_n^* , а других корней правдоподобия нет. Если $\hat{\theta}$ - ОМП (в ней не обязательно $\frac{\partial}{\partial \theta} \ln f_\theta(x) = 0$), не совпадающая с θ_n^* , то где-то между ними - точка минимума, в которой $\frac{\partial}{\partial \theta} \ln f_\theta(x) = 0$, что противоречит условию единственности решения уравнения.

▲

7 Способы сравнения оценок.

Определение Борелевская функция $g(x, y) \geq 0, \forall x, y$ называется **функцией потерь**

Примеры

- 1) $g = |x - y|$ (MAE)
- 2) $g = (x - y)^2$ (MSE)
- 3) Пусть A - неотрицательно определенная матрица,
 $g(\theta^*, \theta) = \langle A(\theta^* - \theta), \theta^* - \theta \rangle$

Определение Пусть $g(x, y)$ - функция потерь, тогда **функцией риска** оценки θ^* называется $R(\theta^*(X), \theta) = \mathbf{E}_\theta g(\theta^*(X), \theta)$

1. Равномерный подход

Определение Оценка θ^* **лучше** оценки $\hat{\theta}$, если $\forall \theta \in \Theta R(\hat{\theta}, \theta) \geq R(\theta^*, \theta)$ и хотя бы для одного θ неравенство строгое.

Определение Оценка θ^* называется **наилучшей** в классе \mathcal{K} , если она лучше любой другой оценки из \mathcal{K}

Замечание наилучшая оценка не всегда существует.

Средне квадратический подход.

$g(x, y) = (x - y)^2, \mathcal{K}$ - класс несмещенных оценок параметра $\tau(\theta)$. Тогда задача сводится к поиску оценки с равномерно наименьшей дисперсией.

$$\begin{aligned} & \mathbf{E}_\theta(\theta^*(X) - \theta)^2 = \\ & = \mathbf{E}_\theta(\theta^*(X) - \mathbf{E}_\theta \theta^*(X))^2 + \mathbf{E}_\theta(\mathbf{E}_\theta \theta^*(X) - \theta)^2 = \mathbf{E}_\theta(\theta^*(X) - \mathbf{E}_\theta \theta^*(X))^2 + (\mathbf{E}_\theta \theta^*(X) - \theta)^2 \end{aligned}$$

Определение Оценка θ^* **допустимая** оценка θ в классе \mathcal{K} , если $\nexists \hat{\theta}(X) \in \mathcal{K}$ т.ч. $\hat{\theta}$ лучше θ^*

2. Минимаксный подход

Определение Оценка θ^* называется **наилучшей в минимаксном подходе**, если

$$\sup_{\theta \in \Theta} R(\theta^*(X), \theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\hat{\theta}(X), \theta)$$

3. Асимптотический подход

Пусть θ_1^* и θ_2^* - асимптотически нормальные оценки с асимптотическими дисперсиями $\sigma_1^2(\theta)$ и $\sigma_2^2(\theta)$

Тогда θ_1^* **лучше** θ_2^* в **асимптотическом подходе**, если

$$\forall \theta \in \Theta \quad \sigma_1^2(\theta) \leq \sigma_2^2(\theta)$$

Причем для некоторых θ неравенство строгое.

4. Байесовский подход

На множестве Θ задаем вероятностную меру Q и θ случайно выбирается из Θ по закону Q . Если $\theta^*(X)$ - оценка параметра θ , а $R(\theta^*(X), \theta)$ - функция риска, то положим $R(\theta^*(X)) = \mathbf{E}_Q R(\theta^*(X), \theta) = \int_{\Theta} R(\theta^*(X), \theta) dQ$

Определение Оценка $\theta^*(X)$ называется **наилучшей в байесовском подходе**, если $R(\theta^*(X)) = \min_{\hat{\theta}} R(\hat{\theta}(X))$

Утверждение Пусть θ^* - наилучшая оценка параметра θ в байесовском подходе. Тогда θ^* - допустимая оценка в равномерном подходе.

▲

Пусть это не так, т.к. $\exists \hat{\theta}(X)$ т.ч.

$$R(\hat{\theta}(X), \theta) \leq R(\theta^*(X), \theta) \Rightarrow$$

$$\mathbf{E}_Q R(\hat{\theta}(X), \theta) \leq \mathbf{E}_Q R(\theta^*(X), \theta) \Rightarrow$$

противоречие ▲

8 Эффективные оценки. Информация Фишера. Неравенство Рао-Крамера. Пример для бернуллиевского распределения.

Пусть семейство распределений $\mathcal{P} = \{P_{\theta}, \theta \in \Theta\}$ доминируемо относительно меры μ

Условия Регулярности

- 1) $\Theta \subset \mathcal{R}$ - открытый интервал
- 2) Множество $\mathcal{A} = \{x \in \mathcal{X} : p_\theta(x) > 0\}$ не зависит от θ
- 3) Для для любой статистики $S(x_1, \dots, x_n)$: если $\exists \mathbf{E}_\theta S^2(x_1, \dots, x_n) < \infty$ для $\forall \theta$, то

$$\begin{aligned} \frac{\partial}{\partial \theta} \mathbf{E}_\theta S(X_1, \dots, X_n) &= \frac{\partial}{\partial \theta} \int_{\mathcal{A}^n} S(x_1, \dots, x_n) p_\theta(x_1, \dots, x_n) d\mu = \\ &= \int_{\mathcal{A}^n} \frac{\partial}{\partial \theta} S(x_1, \dots, x_n) p_\theta(x_1, \dots, x_n) d\mu \\ &= \int_{\mathcal{A}^n} S(x_1, \dots, x_n) \frac{p'_\theta}{p_\theta} p_\theta d\mu = \\ &= \int_{\mathcal{A}^n} S(x_1, \dots, x_n) \frac{\partial \ln p_\theta}{\partial \theta} p_\theta d\mu = \mathbf{E} S(X_1, \dots, X_n) \left(\frac{\partial}{\partial \theta} \ln p_\theta(X_1, \dots, X_n) \right) \end{aligned}$$

(возможно дифференцирование под знаком интеграла)

Случайная величина $\mathcal{U}_\theta(X) = \frac{\partial}{\partial \theta} \ln p_\theta(X)$ называется **вкладом наблюдения X**

Величина $\mathcal{I}_X(\theta) = \mathbf{E}_\theta \mathcal{U}_\theta^2(X)$ называется **количеством информации (по Фишеру)** о параметре θ , содержащемся в наблюдении X

- 4) $\mathcal{I}_X(\theta)$ положительна и конечна $\forall \theta \in \Theta$

Теорема (неравенство Рао-Крамера)

Пусть выполнены условия регулярности 1-4 и $\widehat{\theta}(X)$ - несмещенная оценка параметра $\tau(\theta)$ с условием $\mathbf{E} \widehat{\theta}^2(X) < +\infty \quad \forall \theta \in \Theta$

Тогда:

$$\mathbf{D}_\theta \widehat{\theta}(X) \geq \frac{(\tau'(\theta))^2}{\mathcal{I}_X(\theta)}$$

◀

Из условия регулярности 3:

Положим $S(X) = 1 \Rightarrow 0 = \mathbf{E}_\theta \mathcal{U}_\theta(X)$

Положим $S(X) = \widehat{\theta} \Rightarrow \frac{\partial}{\partial \theta} \tau(\theta) = \tau'(\theta) = \mathbf{E}_\theta(\widehat{\theta}(X) \mathcal{U}_\theta(X)) (1)$

$0 = \mathbf{E}_\theta(\tau(\theta) \mathcal{U}_\theta(X)) (2)$

вычитая из (1) (2), получим:

$$\tau'(\theta) = \mathbf{E}_\theta[(\widehat{\theta} - \tau(\theta) \mathcal{U}_\theta(X))]$$

Из неравенства Коши-Буняковского:

$$[\tau'(\theta)]^2 = (\mathbf{E}_\theta[(\widehat{\theta} - \tau(\theta) \mathcal{U}_\theta(X))])^2 \leq \mathbf{E}_\theta(\widehat{\theta}(X) - \tau(\theta))^2 \mathbf{E}_\theta(\mathcal{U}_\theta(X))^2$$



Оценка, при которой достигается равенство в неравенстве Рао-Крамера называется **эффективной** оценкой параметра $\tau(\theta)$

Замечание Эффективная оценка является наилучшей оценкой в среднеквадратичном подходе, в классе несмещенных оценок.

Утверждение(аддитивности вероятности)

Если $X = (X_1, \dots, X_n)$ - выбоорка, то $\mathcal{I}_X(\theta) = ni(\theta)$, где $i(\theta)$ - информация, заключенная в одном члене выборки.

Теорема (критерий эффективности)

Пусть выполнены условия регулярности Рао-Крамера

Тогда $\hat{\theta}(X)$ - эффективная $\Leftrightarrow \hat{\theta}(X) - \tau(\theta) = c(\theta)\mathcal{U}_\theta(X)$, где $c(\theta) = \frac{\tau'(\theta)}{\mathcal{I}_X(\theta)}$



В доказательстве неравенства Рао-Крамера было:

$$[\tau'(\theta)]^2 = (\mathbf{E}_\theta[(\hat{\theta} - \tau(\theta)\mathcal{U}_\theta(X))])^2 \leq \mathbf{E}_\theta(\hat{\theta}(X) - \tau(\theta))^2 \mathbf{E}_\theta(\mathcal{U}_\theta(X))^2$$

Равенство в неравенстве Коши-Буняковского \Leftrightarrow случайные величины являются линейно зависимыми (почти наверное)

т.е. $\hat{\theta}(X) - \tau(\theta) = c(\theta)\mathcal{U}_\theta(X) + a(\theta)$ P_θ п.н. $\forall \theta$

Возьмем математическое ожидание от обеих частей, слева получим 0, т.к. оценка несмещенная, $\mathbf{E}_\theta(c(\theta)\mathcal{U}_\theta(X)) = 0$ (т.к. $\mathbb{E}_\theta(\mathcal{U}_\theta(X)) = 0$), значит $a(\theta) = 0$

Умножив обе части $\hat{\theta}(X) - \tau(\theta) = c(\theta)\mathcal{U}_\theta(X)$ на $\mathcal{U}_\theta(X)$ и взяв математическое ожидание, получим требуемое.



Пример Пусть X_1, \dots, X_n из $Bern(\theta)$

$$p_\theta = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

$$\mathcal{U}_\theta(X) = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1 - \theta} = \frac{\sum x_i - n\theta}{\theta(1 - \theta)} = \frac{n}{\theta(1 - \theta)}(\bar{X} - \theta)$$

(т.е. \bar{X} - эффективная оценка θ)

$$\hat{\theta} - \theta = c(\theta)\mathcal{U}_\theta$$

$$c(\theta) = \frac{\tau'(\theta)}{ni(\theta)} \Rightarrow ni(\theta) = \tau'(\theta)c^{-1}(\theta) = \frac{n}{\theta(1 - \theta)} \Rightarrow i(\theta) = \frac{1}{\theta(1 - \theta)}$$

9 Эффективность и асимптотическая эффективность оценки максимального правдоподобия.

Условия регулярности (продолжение) для ОМП

- 4) p_θ дифференцируема на Θ
- 5) Плотность $p_\theta(x)$ трижды непрерывно дифференцируема на Θ для $\forall x \in \mathcal{A}$
- 6) Интеграл $\int_{\mathcal{A}} p_\theta(x) \mu(dx)$ можно трижды непрерывно дифференцировать под знаком интеграла
- 7) Информация Фишера $i(\theta) = \mathbf{E}_\theta \left(\frac{\partial}{\partial \theta} \ln p_\theta(x) \right)^2$ одного наблюдения
- 8) $\forall \theta_0 \in \Theta \exists c > 0$ и $H(x)$ т.ч. $\forall x \in \mathcal{A} \forall \theta \in (\theta_0 - c; \theta_0 + c)$

$$\left| \frac{\partial^3}{\partial \theta^3} \ln p_\theta(X) \right| \leq H(X)$$

и $\mathbf{E}_\theta H(X_1) < +\infty$

Теорема (об асимптотической нормальности ОМП)

В условиях регулярности 1-8

- 1) Пусть $\hat{\theta}$ - состоятельная последовательность решений уравнения правдоподобия. Тогда $\hat{\theta}$ - асимптотически нормальная оценка θ с асимптотической дисперсией $\frac{1}{i(\theta)}$
- 2) Пусть $\hat{\theta}$ - какая-то асимптотически нормальноая оценка θ с асимптотической дисперсией $\sigma^2(\theta)$, причем $\sigma(\theta)$ непрерывна по θ . Тогда $\forall \theta \in \Theta : \sigma^2(\theta) \geq \frac{1}{i(\theta)}$

Вывод в условиях регулярности 1-8 ОМП является наилучшей в асимптотическом подходе среди всех асимптотически нормальных оценок с непрерывной асимптотической дисперсией.

Теорема (эффективность ОМП)

В условиях неравенства Рао-Крамера, если $\hat{\theta}$ - эффективная оценка θ , то $\hat{\theta}$ - ОМП.

◀

По критерию эффективности: $\hat{\theta} - \theta = \frac{1}{\mathcal{I}_X(\theta)} \mathcal{U}_\theta(X)$

$$\mathcal{U}_\theta(X) = \frac{\partial}{\partial \theta} \ln p_\theta(X) = \mathcal{I}_X(\theta)(\hat{\theta} - \theta)$$

Получаем: при $\theta < \hat{\theta}$: $\frac{\partial}{\partial \theta} \ln p_\theta(X) > 0$ (т.к. $\mathcal{I}_X(\theta) \geq 0$)

При $\theta > \hat{\theta}$: $\frac{\partial}{\partial \theta} \ln p_\theta(X) < 0$

\Rightarrow на $\hat{\theta}$ достигается максимум функции правдоподобия $\Rightarrow \hat{\theta}$ - ОМП.



10 Баясовские оценки: определение, типы байесовских оценок, наилучшая оценка в байесовском подходе, сопряженные распределения, примеры.

Напоминание (теорема Байеса)

Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, $\{D_n\}_{n=1}^\infty$ - разбиение Ω . $A \in \mathcal{F}$ - событие. Тогда

$$P(D_n|A) = \frac{P(A|D_n)P(D_n)}{\sum_{j=1}^\infty P(A|D_j)P(D_j)}$$

Терминология

1) A - произошедшее событие, является результатом эксперимента.

2) $P(D_n)$ - априорная вероятность, задается до эксперимента.

3) $P(D_n|A)$ - апостериорная вероятность, учитывает результат эксперимента.

Теорема (Байеса, в общем случае)

Пусть ξ, η - случайные вектора размерности n и m соответственно

$$P_{\eta|\xi}(y|x) = \frac{p_{\xi|\eta}(x|y)p_\eta(y)}{\int_{\mathbb{R}^m} p_{\xi|\eta}(x|s)p_\eta(s)ds}$$

Основы байесовского подхода

Пусть θ - случайный вектор со значениями из множества $\Theta \subset \mathbb{R}^d$, имеющий распределение Q с плотностью $q(t)$, которая называется **априорной**

Пусть $X = (X_1, \dots, X_n)$ - выборка из неизвестного распределения $P \in \{P_t, t \in \Theta\}$

Причем P_t имеет плотность $p_t(x)$

Зависимость Q и X определим через плотность вектора (Q, X)

$$f(t, x_1, \dots, x_n) = q(t)p_t(x_1) \dots p_t(x_n)$$

t - значение θ

x_1 - значение X_1

...

x_n - значение X_n

Реализация модели

1) Генерируем значение θ из распределения Q и получаем число

t

2) Генерируем выборку из распределения P_t

3) Способы оценки параметра:

1. Апостериорное распределение

$$q(t|x_1, \dots, x_n) = \frac{q(t)p_t(x_1) \dots p_t(x_n)}{\int_{\Theta} q(t)p_t(x_1) \dots p_t(x_n) dt}$$

2. Точные оценки параметров

(a)

$$\mathbf{E}(\theta|X) = \int_{\Theta} tq(t|x_1, \dots, x_n) dt$$

(b) $\arg \max_{t \in \Theta} q(t|x_1, \dots, x_n)$ - мода апостериорного распределения.

(c) медиана апостериорного распределения.

Задача $X_1, \dots, X_n \sim \mathcal{U}[0; \theta + 1]$, причем $\forall i : X_i \leq 2$
 $\theta \sim \text{Bern}(\frac{1}{2})$. Найти апостериорное распределение.

◀

$$p_t(x) = \frac{1}{t+1} I(x \leq t+1)$$

$$q(t) = \frac{1}{2}, t \in \{0, 1\}$$

$$q(0|X) = \frac{1}{z} q(0) p_0(x_1) \dots p_0(x_n) = \frac{1}{z} \frac{1}{2} I(X_{(n)} \leq 1)$$

$$q(1|X) = \frac{1}{z} q(1) p_1(x_1) \dots p_1(x_n) = \frac{1}{z} \left(\frac{1}{2}\right)^{n+1}$$

$$\text{Здесь везде } z = \frac{1}{2} I(X_{(n)} \leq 1) + \frac{1}{2^{n+1}}$$

▶

Байесовский подход к сравнению оценок

θ_1^* не хуже θ_2^* , если $\mathbf{E}_Q R_{\theta_1^*}(\theta) \leq \mathbf{E}_Q R_{\theta_2^*}(\theta)$

Теорема

Байесовская оценка $\hat{\theta} = \mathbf{E}(\theta|X)$ является наилучшей оценкой в байесовском подходе с квадратичной функцией потерь.

◀

$$R_{\hat{\theta}}(\theta) = MSE_{\hat{\theta}}(\theta) = \mathbf{E}_{\theta}(\hat{\theta} - \theta)^2$$

Пусть θ^* - произвольная оценка θ

$$\begin{aligned} \mathbf{E}_Q MSE_{\theta^*}(\theta) &= \int_{\Theta} MSE_{\theta^*}(t) q(t) dt = \int_{\Theta} \mathbf{E}_t(\theta^* - t)^2 q(t) dt = \\ &= \int_{\Theta} \left(\int_{\mathcal{X}} (\theta^*(x) - t)^2 p_t(x) dx \right) q(t) dt = \int_{\Theta} \int_{\mathcal{X}} (\theta^*(x) - t)^2 f(t, x) dt dx = \\ &= \int_{\Theta} \int_{\mathcal{X}} (f(t, x) - \text{совместная плотность}) = \mathbf{E}(\theta^* - \theta)^2 \rightarrow \min \\ &(\text{минимизируем по } \theta^*) \Rightarrow \text{минимум достигается на } \hat{\theta} \text{ по теореме} \\ &\text{о наилучшем приближении} \end{aligned}$$

►

Определение Пусть $X = (X_1, \dots, X_n)$ - выборка из распределения $P \in \mathcal{P}$, где $\mathcal{P} = \{P_t, t \in \Theta\}$ - некоторое семейство распределений. Пусть на множестве Θ задано семейство распределений $\mathcal{Q} = \{Q_\alpha, \alpha \in \mathcal{A}\}$

Семейство \mathcal{Q} называется **сопряженным** к семейству \mathcal{P} , если при взятии апериорного из \mathcal{Q} соответствующее апостериорное тоже лежит в \mathcal{Q}

т.е., если $X \sim P_t, \theta \sim Q_\alpha$, то $\theta|X \sim Q_\alpha$

Задача

$X_1, \dots, X_n \sim \text{Exp}(\theta)$. Найти сопряженное распределение, найти байесовскую оценку.

◀

Плотность выборки $p_t(x_1, \dots, x_n) = \prod_{i=1}^n t e^{-tx_i} = t^n e^{-t \sum x_i}$

Берем $q(t)$ пропорционально этому выражению

$q(t) \propto t^{\beta-1} e^{-\alpha t}$ - это $\Gamma(\alpha, \beta)$

Соответственно $q(t) = \frac{\alpha^\beta t^{\beta-1} e^{-\alpha t}}{\Gamma(\beta)}$

Покажем, что $\Gamma(\alpha, \beta)$ - сопряжение к $\text{Exp}(\theta)$, для этого найдем апостериорное распределение.

$q(t|X) \propto q(t) p_t(x_1) \dots p_t(x_n) \propto t^{\beta-1} e^{-\alpha t} t^n e^{-t \sum x_i} = t^{n+\beta-1} e^{-t(\alpha + \sum x_i)}$

- это $\Gamma(\alpha + \sum x_i, n + \beta)$

$\Rightarrow \Gamma(\alpha, \beta)$ - сопряженное к $\text{Exp}(\theta)$

Кроме того, $\Gamma(\alpha + \sum x_i, n + \beta)$ - апостериорное распределение

$\mathbf{E}(\theta|X) = \frac{n+\beta}{\alpha + \sum x_i}$

►

11 Достаточные статистики. Критерий факторизации. Теорема Блэкуэлла-Колмогорова-Рао.

Определение Пусть (Ω, \mathcal{F}, P) - вероятностное пространство, $P \in \mathcal{P}, \mathcal{P} = \{P_\theta, \theta \in \Theta\}, G \subset \mathcal{F}$, Если $\forall A \in \mathcal{F} P_\theta(A|G)$ не

зависит от θ , то G называется **достаточной σ -алгебра**

Пусть X - наблюдение, с распределением P_θ , тогда статистика $s(X)$ называется **достаточной**, если $\forall A \ p_\theta(A|s(X))$ не зависит от θ

Пример

$X_1, \dots, X_n \sim \text{Bern}(\theta)$

$s(X) = \sum_{i=1}^n X_i$ - достаточная статистика.

$\forall x \ p(X = x|s(X) = s(x))$ не зависит от θ т.к.

$$\begin{aligned} p(X = x|s(X) = s(x)) &= \frac{p_\theta(X = x)}{p_\theta(s(X) = s(x))} \text{ т.к. из } X=x \text{ следует } s(X)=s(x) \\ &= \frac{\theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}}{C_n^{\sum x_i} \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}} = \frac{1}{C_n^{\sum x_i}} \end{aligned}$$

Теорема (критерий факторизации Неймана-Фишера)

Пусть $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ либо целиком состоит из дискретных распределений, либо целиком состоит из абсолютно непрерывных распределений.

$s(x)$ - достаточная статистика $\Leftrightarrow p_\theta(x) = h(x)\gamma_\theta(s(x))$

◀ (только дискретный случай)

$(\leftarrow) \ p_\theta(x) = h(x)\gamma_\theta(x)$ Тогда

$$\begin{aligned} P(X = x|s(X) = s(x)) &= \frac{P(X = x)}{P(s(X) = s(x))} = \frac{P_\theta(X = x)}{\sum_{y:s(y)=s(X)} P_\theta(X = y)} = \\ &= \frac{h(x)\gamma_\theta(s(x))}{\sum_{y:s(y)=s(X)} h(y)\gamma_\theta(s(y))} = \frac{h(x)}{\sum_{y:s(y)=s(X)} h(y)} \end{aligned}$$

\Rightarrow не зависит от θ и $s(X)$ - достаточная.

(\rightarrow) Пусть $s(X)$ - достаточная статистика. Тогда

$$p_\theta(X = x) = \text{\textbackslash *по определению*} = P_\theta(X = x|s(X) = s(x)) * P_\theta(s(X) = s(x))$$

Первый множитель - $h(x)$, второй $\gamma_\theta(s(x))$



Теорема (Колмогоров-Блэкуэлл-Рао)

Пусть θ^* - несмещенная оценка $\tau(\theta)$ и $s(X)$ - достаточная статистика, тогда

1. $\mathbf{E}(\theta^*|s(x))$ - несмещенная оценка параметра $\tau(\theta)$
2. $\mathbf{D}(\mathbf{E}(\theta^*|s(x))) \leq \mathbf{D}(\theta^*) \ \forall \theta \in \Theta$

3. Равенство достигается $\Leftrightarrow \theta^*$ - является почти наверное $s(x)$ измеримой

◀

- 1) $\mathbf{E}(\mathbf{E}(\theta^*|s(X))) = \mathbf{E}(\theta^*) = \tau(\theta) \Rightarrow$ несмещенная оценка.
 2) Обозначим $\hat{\theta} = \mathbf{E}(\theta^*|s(X))$

$$\mathbf{D}\theta^* = \mathbf{E}_\theta(\theta^* - \tau(\theta))^2 = \mathbf{E}_\theta[(\theta^* - \hat{\theta}) + (\hat{\theta} - \tau(\theta))]^2 = \mathbf{E}_\theta(\theta^* - \hat{\theta})^2 + \mathbf{D}\hat{\theta} + 2\mathbf{E}_\theta[(\theta^* - \hat{\theta})(\hat{\theta} - \tau(\theta))]$$

Покажем, что $\mathbf{E}_\theta[(\theta^* - \hat{\theta})(\hat{\theta} - \tau(\theta))] = 0$

$$\begin{aligned} \mathbf{E}_\theta[(\theta^* - \hat{\theta})(\hat{\theta} - \tau(\theta))] &= \mathbf{E}_\theta(\mathbf{E}_\theta[(\theta^* - \hat{\theta})(\hat{\theta} - \tau(\theta))|s(X)]) = \mathbf{E}_\theta[(\hat{\theta} - \tau(\theta))\mathbf{E}_\theta((\theta^* - \hat{\theta})|s(X))] \\ &\quad \setminus * \hat{\theta} - \tau(\theta) - s(x)\text{-измерима} * \setminus \\ &= \mathbf{E}_\theta[(\hat{\theta} - \tau(\theta))\mathbf{E}_\theta((\theta^* - \hat{\theta})|s(X))] = 0 \end{aligned}$$

т.к.

$$\mathbf{E}_\theta(\theta^* - \hat{\theta}|s(X)) = \mathbf{E}_\theta(\theta^*|s(X)) - \hat{\theta} = \hat{\theta} - \hat{\theta} = 0$$

$$\Rightarrow \mathbf{D}\theta^* \geq \mathbf{D}\hat{\theta} \text{ т.к. } \mathbf{E}_\theta(\theta^* - \hat{\theta})^2 \geq 0$$

- 3) Равенство достигается, если $\mathbf{E}_\theta(\theta^* - \hat{\theta})^2 = 0 \forall \theta \in \Theta$, т.е. $\theta^* = \hat{\theta}$ п.н.

▶

12 Полные статистики. Оптимальные оценки. Алгоритм нахождения оптимальных оценок. Примеры.

Определение Наилучшая оценка в классе несмещенных оценок с квадратичной функцией потерь (т.е. с наименьшей дисперсией) называется **оптимальной**

Определение статистика $s(X)$ называется **полной** для $\{P_\theta, \theta \in \Theta\}$ если для любой борелевской функции f т.ч. $E_\theta f(s(X)) = 0$ следует, что $f(s(X)) = 0$ P_θ п.н. $\forall \theta \in \Theta$

Лемма

Пусть $s(X)$ - достаточная статистика $\{P_\theta, \theta \in \Theta\}$ Если $\hat{\theta}(X)$ - единственная несмещенная s -измеримая оценка $\tau(\theta)$, то $\hat{\theta}(X)$ - оптимальная оценка $\tau(\theta)$

◀

Пусть $\theta^*(X)$ - не s -измеримая несмещенная оценка $\tau(\theta)$ и лучше

$\hat{\theta}(X)$

Тогда $\theta^{**} = \mathbf{E}_{\theta}(\theta^*(X)|s(X))$ - лучше θ^* (К-Б-Р) и является s -измеримой. Тогда $\theta^{**} = \hat{\theta}$ и, значит, θ^* хуже $\hat{\theta}$. Противоречие

►

Теорема (Лемана-Шеффе об оптимальной оценке)

Пусть $s(X)$ - полная, достаточная статистика для $\{P_{\theta}, \theta \in \Theta\}$, а $\varphi(s(X))$ - несмещенная оценка $\tau(\theta)$. Тогда $\varphi(s(X))$ - оптимальная оценка $\tau(\theta)$

◄

Покажем, что $\varphi(s(X))$ - единственная несмещенная s -измеримая оценка $\tau(\theta)$

Пусть не так, т.е. $\exists \psi : \mathbf{E}_{\theta}\psi(s(X)) = \tau(\theta) \Rightarrow \mathbf{E}_{\theta}(\varphi(s(X)) - \psi(s(X))) = 0 \forall \theta \in \Theta$

т.к. $s(X)$ полная, то $\psi - \varphi = 0$ P_{θ} п.н. \Rightarrow с точностью до п.н. $\varphi(s(X))$ - единственная несмещенная s -измеримая оценка.

►

Пример

$X_i \sim \mathcal{U}[0, \theta]$. Найти оптимальную оценку θ

◄

$$f = I(X_{(n)} \leq \theta)I(X_{(1)} \geq 0)\frac{1}{\theta^n}$$

$X_{(n)}$ - достаточная статистика. (теорема о факторизации)

$$0 = \mathbf{E}f(X_{(n)}) = \int_0^{\theta} f(x)\frac{n}{\theta^n}x^{n-1}dx =$$

$$= \frac{n}{\theta^n} \int_0^{\theta} f(x)X^{n-1}dx = 0 \forall \theta > 0 \Rightarrow X_{(n)} - \text{полная}$$

$$\mathbf{E}_{\theta}X_{(n)} = \int_0^{\theta} x\frac{n}{\theta^n}x^{n-1}dx =$$

$$= \frac{n}{\theta^n} \int_0^{\theta} x^n dx = \frac{\theta n}{n+1}$$

$$\Rightarrow \mathbf{E}_{\theta} \left(\frac{n+1}{n} X_{(n)} \right) = \theta \Rightarrow \frac{n+1}{n} X_{(n)} - \text{оптимальная оценка } \theta$$

►

13 Теорема об экспоненциальном семействе.

Теорема

Пусть X - наблюдение с распределением $P \in \mathcal{P} = \{P_{\theta} : \theta \in \Theta \subset \mathbb{R}^k\}$

$$p_{\theta}(x) = h(x)\exp(\sum_{i=1}^k a_i(\theta)S_i(x) + b(\theta))$$

Тогда $(S_1(X), \dots, S_k(X))$ - полная достаточная статистика, если $(a_1, \dots, a_k)(\Theta) \supset k$ -мерный параллелепипед.

◀

Функция правдоподобия:

$$f_{\theta}(X_1, \dots, X_n) = h(X_1) \dots h(X_n) \exp(\sum_{i=1}^n \langle a(\theta), S(X_i) \rangle + nb(\theta))$$

\Rightarrow достаточность следует из критерия Неймана-Фишера

Докажем полноту.

$$\varphi = \varphi^+ - \varphi^-$$

$$\mathbf{E}_{\theta} \varphi(S(X)) = \int_{\mathbb{R}^n} \varphi(S(x)) h(x_1) \dots h(x_n) \exp(\langle a(\theta), S(x) \rangle + nb(\theta)) dP_{\theta} = 0$$

(обозначим $H(X) = h(x_1) \dots h(x_n)$)

\Rightarrow

$$\int_{\mathbb{R}^n} \varphi^+(S(x)) H(x) \exp(\langle a(\theta), S(x) \rangle) dP_{\theta} = \int_{\mathbb{R}^n} \varphi^-(S(x)) H(x) \exp(\langle a(\theta), S(x) \rangle) dP_{\theta}$$

Определим меру ν на \mathbb{R}^k : $\forall B \in \mathcal{B}(\mathbb{R}^k)$:

$$\nu(B) = \int_{x: S(x) \in B} H(x) dP_{\theta}$$

Эта мера будет σ -конечной

Напоминание σ -конечная мера — такая мера, что все пространство может быть представлено в виде счетного объединения измеримых множеств конечной меры.

$$\int_{\mathbb{R}^k} \varphi^+(S) \exp(\langle a(\theta), S \rangle) d\nu = \int_{\mathbb{R}^k} \varphi^-(S) \exp(\langle a(\theta), S \rangle) d\nu$$

(т.е. мы сделали замену переменных в интеграле Лебега)

Лемма(б/д)

Пусть μ_1, μ_2 - σ -конечные меры на $\mathcal{B}(\mathbb{R}^k)$ и $\forall a$ из некоторого k -мерного параллелепипеда:

$$\int_{\mathbb{R}^k} \exp(\langle a, s \rangle) d\mu_1 = \int_{\mathbb{R}^k} \exp(\langle a, s \rangle) d\mu_2$$

Тогда $\mu_1 = \mu_2$

(Продолжение док-ва)

$$\text{Пусть } \nu^+(B) = \int_B \varphi^+(S) d\nu, \nu^-(B) = \int_B \varphi^-(S) d\nu$$

Эти меры, опять же, σ -конечные. Т.е. имеем:

$$\int_{\mathbb{R}^k} \exp(\langle a(\theta), s \rangle) d\nu^+ = \int_{\mathbb{R}^k} \exp(\langle a(\theta), s \rangle) d\nu^-$$

\Rightarrow

$$\nu^+ = \nu^- \Rightarrow \varphi^+ = \varphi^- \text{ } \nu\text{-почти всюду}$$

$$0 = \int_{S: \varphi^+(S) \neq \varphi^-(S)} d\nu = \nu(\{S : \varphi^+(S) \neq \varphi^-(S)\}) = \int_{x: \varphi^+(S(x)) \neq \varphi^-(S(x))} H(x) dP_\theta$$

$H(x) > 0$ (Если $H(x) = 0$, то изначально вместо \mathbb{R}^n можно будет взять другой носитель)

\Rightarrow

$$\int_{x: \varphi^+(S(x)) \neq \varphi^-(S(x))} dP_\theta = 0$$

\Rightarrow

$$P_\theta(S : \varphi^+(S) \neq \varphi^-(S)) = 0 \quad \forall \theta \in \Theta$$

$\Rightarrow \varphi = 0$ P_θ п.н. $\forall \theta \in \Theta \Rightarrow S(X)$ - полная

►

14 Доверительные интервалы, центральная функция, примеры нахождения. Асимптотические доверительные интервалы, алгоритм нахождения.

Определение Пусть X наблюдение из $P \in \{P_\theta \mid \theta \in \Theta\}$

Пара статистик $(T_1(X), T_2(X))$ называется **доверительным интервалом** уровня доверия γ для θ , если $\forall \theta \in \Theta \quad P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq \gamma$.

Доверительный интервал **точный**, если вместо \geq стоит $=$

Определение Пусть $\Theta \subset \mathbb{R}^k$, $S(X) \subset \mathbb{R}^k$ называется **доверительной областью** уровня доверия γ , если $P_\theta(\theta \in S(X)) \geq \gamma \quad \forall \theta \in \Theta$.

Метод центральной функции

Пусть существует функция $F(x, \theta)$ такая, что её распределение Q не зависит от θ .

Пусть $\alpha_1, \alpha_2 \in (0, 1) : \alpha_2 - \alpha_1 = \gamma$, $z_{\alpha_1}, z_{\alpha_2}$ - квантили уровней α_1, α_2 соответственно распределения Q

$$P(z_{\alpha_1} \leq F(x, \theta) \leq z_{\alpha_2}) \geq \gamma$$

Надо найти такую $S(x) : \theta \in S(x) \Leftrightarrow z_{\alpha_1} \leq F(x, \theta) \leq z_{\alpha_2}$

$\Rightarrow P(\theta \in S(x)) \geq \gamma \Rightarrow S(x)$ - доверительная область

Утверждение ξ - случайная величина с непрерывной функцией распределения F . Тогда $-\ln F(\xi) \sim \text{Exp}(1)$

◀

$$\eta = F(\xi)$$

$$P(\eta \leq x) = \mathbb{P} * x \in [0, 1] * \mathbb{P} = P(F(\xi) \leq x) = P(\xi \leq \sup(y : F(y) \leq x)) = x \Rightarrow \eta \sim \mathcal{U}[0, 1]$$

$$P(-\ln \eta \leq x) = \mathbb{P} * x \in [0, +\infty) * \mathbb{P} = P(\ln \eta \geq -x) = P(\eta \geq e^{-x}) \sim \exp(1)$$

▶

$$\xi_1, \dots, \xi_n \sim \exp(1) \text{ и независимы } \Rightarrow \sum_{i=1}^n \xi_i \sim \Gamma(1, n)$$

$$X_1, \dots, X_n \sim P_\theta \text{ с непрерывной функцией распределения } F_\theta$$

$$\Rightarrow -\sum_{i=1}^n \ln F_\theta(X_i) - \text{центральная функция.}$$

Определение Пусть X_1, X_2, \dots - выборка (бесконечного объема)

$(T_{1,n}(X_1, \dots, X_n), T_{2,n}(X_1, \dots, X_n))$ - **асимптотический доверительный интервал** уровня γ если $\forall \theta \in \Theta$:

$$\lim_{n \rightarrow \infty} p_\theta(T_{1,n} \leq \theta \leq T_{2,n}) \geq \gamma$$

Если можно написать $\lim_{n \rightarrow \infty} p_\theta(T_{1,n} \leq \theta \leq T_{2,n}) = \gamma$ то асимптотический доверительный интервал называется **точным**

Пусть $\hat{\theta}_n$ - асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$, непрерывной на Θ

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\sigma^2(\theta)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\sigma^2(\hat{\theta}_n)}} = \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\sigma^2(\theta)}} * \sqrt{\frac{\sigma^2(\theta)}{\sigma^2(\hat{\theta}_n)}}$$

$\hat{\theta}_n$ - асимптотически нормальная оценка \rightarrow состоятельная

$$\rightarrow \sigma^2(\hat{\theta}_n) \xrightarrow{P} \sigma^2(\theta) \text{ (наследование)}$$

$$\rightarrow \sqrt{\frac{\sigma^2(\theta)}{\sigma^2(\hat{\theta}_n)}} \xrightarrow{P} 1$$

По Лемме Slutsky:

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\sigma^2(\hat{\theta}_n)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Т.о.

$$P_{\theta} \left(-u_{\frac{1+\gamma}{2}} \leq \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\sigma^2(\hat{\theta}_n)}} \leq u_{\frac{1+\gamma}{2}} \right) \rightarrow \gamma$$

u_{α} - α квантиль стандартного распределения

$$P_{\theta} \left(\hat{\theta}_n - \sqrt{\frac{\sigma^2(\hat{\theta}_n)}{n}} * u_{\frac{1+\gamma}{2}} \leq \theta \leq \hat{\theta}_n + \sqrt{\frac{\sigma^2(\hat{\theta}_n)}{n}} * u_{\frac{1+\gamma}{2}} \right) \rightarrow \gamma$$

15 Задача линейной регрессии. Метод наименьших квадратов. Несмещенность оценки по методу наименьших квадратов. Матрица ковариаций оценки по методу наименьших квадратов. Оценка дисперсии. Свойства оценки по методу наименьших квадратов.

Пусть $X \in \mathbb{R}^n$ - наблюдение, которое представляется в виде $X = l + \varepsilon$, где l - неслучайный вектор (измеримая величина), ε - случайный вектор (ошибка измерений)

Про ε известно, что $\mathbf{E}\varepsilon = 0$ $\mathbf{D}\varepsilon = \sigma^2 I_n$, σ^2 неизвестно.

Про l известно, что $l \in L$, L - линейное подпространство \mathbb{R}^n , $\dim(L) = k < n$, L известно.

Задача оценить l и σ^2

Пусть z_1, \dots, z_k - базис в L (вектор-столбцы)

$Z = (z_1 \dots z_k)$ - матрица $n \times k$

Тогда $l = \sum_{i=1}^k \theta_i z_i$ θ_i - неизвестные координаты вектора l в базисе (z_1, \dots, z_k)

Перейдем к оценке (θ, σ^2)

Метод наименьших квадратов

Определение $\hat{\theta}(X) = \arg \min_{\theta} \|X - Z\theta\|^2$ - оценка МНК параметра θ (минимальное расстояние на $Z\theta = \text{proj}_L X$)

Лемма

$$\hat{\theta} = (Z^T Z)^{-1} Z^T X$$

◀

$$\|X - Z\theta\|^2 = (X - Z\theta)^T (X - Z\theta) = X^T X - 2X^T Z\theta + \theta^T Z^T Z\theta$$

Берем производную по θ_i :

$$-2(X^T Z)_i + 2(\theta^T Z^T Z)_i = 0$$

$$\Rightarrow -X^T Z + \theta^T Z^T Z = 0 \Rightarrow Z^T Z \theta = Z^T X \Rightarrow \theta = (Z^T Z)^{-1} Z^T X$$

►

Утверждение

$$\mathbf{E}\hat{\theta} = \theta$$

◄

$\mathbf{E}X = Z\theta$ по линейности

из предыдущей леммы:

$$\mathbf{E}\hat{\theta} = \mathbf{E}(Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathbf{E}X = (Z^T Z)^{-1} Z^T Z \theta = \theta$$

►

Утверждение

$$\mathbf{D}\hat{\theta} = \sigma^2 (Z^T Z)^{-1}$$

◄

$\mathbf{D}X = \sigma^2 I_n$, I_n - единичная матрица

Получим:

$$\begin{aligned} \mathbf{D}\hat{\theta} &= \mathbf{D}(Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathbf{D}X [(Z^T Z)^{-1} Z^T]^T \\ &= (Z^T Z)^{-1} Z^T \sigma^2 I_n [(Z^T Z)^{-1} Z^T]^T = \sigma^2 I_n (Z^T Z)^{-1} = \sigma^2 (Z^T Z)^{-1} \end{aligned}$$

► **Теорема**(б/д)

Пусть $t = T\theta$, T - матрица $(m \times k)$.

Тогда $\hat{t} = T\hat{\theta}$ является оптимальной оценкой параметра t в классе линейных несмещенных оценок (т.е. оценок вида $B\bar{X}$, B - матрица)

Лемма

$$\mathbf{E}\|X - Z\hat{\theta}\|^2 = \sigma^2(n - k)$$

◄

т.к. $\mathbf{E}(X - Z\hat{\theta}) = 0$, то

$$\mathbf{E}\|X - Z\hat{\theta}\|^2 = \sum_{i=1}^n \mathbf{E}(X - Z\hat{\theta})_i^2 = \sum_{i=1}^n \mathbf{D}(X - Z\hat{\theta})_i = \text{tr}\mathbf{D}(X - Z\hat{\theta})$$

$$\mathbf{D}(X - Z\hat{\theta}) = \mathbf{D}((I_n - Z(Z^T Z)^{-1} Z^T)X)$$

Обозначим $A = Z(Z^T Z)^{-1} Z^T$

Заметим, что $A^T = A$, $A^2 = A$

$$\mathbf{D}(X - Z\hat{\theta}) = (I_n - A)\mathbf{D}X(I_n - A) = \sigma^2(I_n - A)^2 = \sigma^2(I_n - 2I_n A + A^2) = \sigma^2(I_n - A)$$

$\text{tr}(AB) = \text{tr}(BA)$ (напоминание из лин. алгебры)

$$\begin{aligned} \Rightarrow \mathbf{E}\|X - Z\hat{\theta}\|^2 &= \sigma \text{tr}(I_n - A) = \sigma^2(n - \text{tr}(Z(Z^T Z)^{-1} Z^T)) = \\ &= \sigma^2(n - \text{tr}(Z^T Z(Z^T Z)^{-1})) = \sigma^2(n - k) \end{aligned}$$



Следствие

$\frac{1}{n-k} \mathbf{E} \|X - Z\hat{\theta}\|^2$ - несмещенная оценка параметра σ^2

Замечание

$$X - Z\hat{\theta} = \text{proj}_{L^\perp} X$$



$$X = \text{proj}_L X + \text{proj}_{L^\perp} X = Z\hat{\theta} + \text{proj}_{L^\perp} X$$



16 Гауссовская линейная модель. Хи-квадрат распределение. Теорема об ортогональном разложении гауссовского вектора. Оптимальность оценки.

Определение Случайная величина ξ имеет распределение **Хи-квадрат с k степенями свободы**, если $\xi \sim \Gamma(\frac{k}{2}, \frac{1}{2})$

Свойство Пусть $\xi_1, \dots, \xi_n \sim \mathcal{N}(0, 1)$ и независимы $\Rightarrow \sum_{i=1}^n \xi_i^2 \sim \chi_k^2$

Теорема (ортогональное разложение гауссовского вектора)
Пусть $\xi \sim \mathcal{N}(a, \sigma^2 I_n)$ - гауссовский случайный вектор размерности n с независимыми компонентами. Пусть также $L_1 \oplus \dots \oplus L_k = \mathbb{R}^n$ - разложение в прямую сумму ортогональных подпространств. и $\eta_j = \text{proj}_{L_j} \xi$ - проекция вектора ξ на подпространство L_j .

Тогда η_1, \dots, η_k - независимы в совокупности, причем

$$\mathbf{E} \eta_j = \text{proj}_{L_j} a$$

$$\frac{1}{\sigma^2} \|\eta_j - \mathbf{E} \eta_j\|^2 \sim \chi_{d_j}^2$$

где $d_j = \dim L_j$



Пусть $e_1 \dots e_n$ - ортонормированный базис в \mathbb{R}^n , причем I_j - множество индексов, соответствующих базису в подпространстве L_j , т.е. $\{e_i, i \in I_j\}$ - базис L_j
обозначим $\gamma_i = \langle \xi, e_i \rangle$, $\gamma = (\gamma_1, \dots, \gamma_n)^T$.

Тогда $\gamma = B\xi$, $B = (e_1^T \dots, e_n^T)^T$ - ортогональная матрица.
Вектор γ также является гауссовским, как линейное преобразование гауссовского вектора ξ , причем.

$$\mathbf{E}\gamma = B\mathbf{E}\xi = Ba$$

$$\mathbf{D}\gamma = B\mathbf{D}\xi B^T = B\sigma^2 I_n B^T = \sigma^2 I_n, \quad BB^T = I_n$$

Получаем, что $\gamma \sim \mathcal{N}(Ba, \sigma^2 I_n)$, причем его компоненты независимы в совокупности, по критерию независимости компонент гауссовского вектора.

$$\text{Проекция имеет вид } \eta_j = \text{proj}_{L_j} \xi = \sum_{i \in I_j} \langle \xi, e_i \rangle e_i = \sum_{i \in I_j} \gamma_i e_i$$

Отсюда получаем, что векторы $\eta_1 \dots \eta_k$ - независимы в совокупности как функции от независимых компонент вектора ξ

$$\mathbf{E}\eta_j = \sum_{i \in I_j} \langle \mathbf{E}\xi, e_i \rangle e_i = \text{proj}_{L_j} \mathbf{E}\xi = \text{proj}_{L_j} a$$

Поскольку базис ортогональный, а величины $(\gamma_i - \mathbf{E}\gamma_i) \sim \mathcal{N}(0, \sigma^2)$ и независимы в совокупности, получаем:

$$\frac{1}{\sigma^2} \|\eta_j - \mathbf{E}\eta_j\|^2 = \frac{1}{\sigma^2} \left\| \sum_{i \in I_j} (\gamma_i - \mathbf{E}\gamma_i) e_i \right\|^2 = \frac{1}{\sigma^2} \sum_{i \in I_j} (\gamma_i - \mathbf{E}\gamma_i)^2 \sim \chi_{d_j}^2$$

►

Гауссовская регрессионная модель

Пусть ε - гауссовский вектор, т.е. $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$
 $X = l + \varepsilon$

$$p_X(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \frac{-\sum (x_i - l_i)^2}{2\sigma^2}$$

$$\begin{aligned} X &= \text{proj}_L X + \text{proj}_{L^\perp} X \\ \Rightarrow \sum (x_i - l_i)^2 &= \|x - l\|^2 = \|\text{proj}_L x - l\|^2 + \|\text{proj}_{L^\perp} x\|^2 \\ &(\text{Теорема Пифагора и факт, что } \text{proj}_{L^\perp} l = 0) \\ \Rightarrow p_X(x) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \frac{-\|\text{proj}_L x - l\|^2 - \|\text{proj}_{L^\perp} x\|^2}{2\sigma^2} \end{aligned}$$

По критерию Неймана-Фишера: $s(x) = (\text{proj}_L X, \|\text{proj}_L X - X\|^2)$
- достаточная статистика.

Теорема (б/д)

$(\text{proj}_L X, \|\text{proj}_L X - X\|^2)$ - полная.

Утверждение $\hat{\theta}$ - оптимальная оценка θ
 $Z\hat{\theta}$ - оптимальная оценка l

$\frac{1}{n-k} \|X - Z\hat{\theta}\|^2$ - оптимальная оценка σ^2



$$proj_{L^\perp} X = X - Z\hat{\theta} \quad proj_L X = Z\hat{\theta}$$

$$\hat{\theta} = (Z^T Z)^{-1} Z^T (Z\hat{\theta})$$

Получаем, что все оценки оптимальные, так как являются несмещенными и функциями от полных достаточных статистик.



Утверждение В гауссовской линейной модели $\hat{\theta}$ и $X - Z\hat{\theta}$ независимы, причем $\frac{1}{\sigma^2} \|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2$ и $\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2$



$$Z\hat{\theta} = proj_L X \Rightarrow X - Z\hat{\theta} = proj_{L^\perp} X$$

По теореме о разложении гауссовского вектора:

$$\frac{1}{\sigma^2} \|Z\hat{\theta} - \mathbf{E}proj_L l\|^2 = \frac{1}{\sigma^2} \|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2$$

$$\frac{1}{\sigma^2} \|(X - Z\hat{\theta}) - \mathbf{E}(X - Z\theta)\|^2 = (\mathbf{E}(X - Z\theta) = 0) = \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2$$

$Z\hat{\theta}$ и $X - Z\hat{\theta}$ независимы $\Rightarrow \hat{\theta} = (Z^T Z)^{-1} Z^T Z\hat{\theta}$ независима с $X - Z\hat{\theta} \Rightarrow \hat{\theta}$ и σ^2 независимы.



17 Распределения Стюдента и Фишера. Доверительные интервалы для параметров гауссовской линейной модели. Доверительная область для оценки МНК. Пример нахождения доверительных интервалов для параметров нормального распределения по выборке.

Определение Пусть $\xi \sim \mathcal{N}(0, 1)$, $\eta \sim \chi_k^2$ $\xi \perp \eta \Rightarrow \frac{\xi}{\sqrt{\frac{\eta}{k}}} \sim St(k)$ -
Распределение Стюдента

Определение Пусть $\xi \sim \chi_k^2$, $\eta \sim \chi_r^2$, $\xi \perp \eta \Rightarrow$

$$\frac{\xi \backslash k}{\eta \backslash r} \sim F_{k,r}$$

- **Распределение Фишера**

ДИ для σ^2

$$\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 = \frac{1}{\sigma^2} \|proj_{L^\perp} X\|^2 \sim \chi_{n-k}^2$$

Распределение χ_{n-k}^2 не зависит от $\sigma^2 \Rightarrow$ центральная статистика

Пусть $z_{1-\gamma}$ - $1 - \gamma$ квантиль χ_{n-k}^2

$$P_{\theta, \sigma^2} \left(\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 > z_{1-\gamma} \right) = \gamma$$

$$P_{\theta, \sigma^2} \left(0 \leq \sigma^2 \leq \frac{\|X - Z\hat{\theta}\|^2}{z_{1-\gamma}} \right) = \gamma$$

ДИ для θ_i

$\hat{\theta}$ — гауссовский вектор (как линейное преобразование X - гауссовского вектора)

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(Z^T Z)^{-1})$$

$$\text{Пусть } A = (Z^T Z)^{-1} \Rightarrow \hat{\theta}_i \sim \mathcal{N}(\theta_i, \sigma^2 A_{ii})$$

\Rightarrow

$$\frac{\hat{\theta}_i - \theta_i}{\sigma \sqrt{A_{ii}}} \sim \mathcal{N}(0, 1)$$

$$\text{Далее, } \hat{\sigma}^2 = \frac{1}{n-k} \|X - Z\hat{\theta}\|^2 - \text{оптимальная оценка } \sigma^2 \text{ и } \hat{\sigma}^2 \sim \chi_{n-k}^2$$

\Rightarrow

$$\frac{\hat{\theta}_i - \theta_i}{\sqrt{A_{ii} \frac{1}{n-k} \|X - Z\hat{\theta}\|^2}} \sim St(n-k)$$

$St(n-k)$ - центральная статистика

$$t_{1-\frac{\gamma}{2}} - 1 - \frac{\gamma}{2} \text{ квантиль для } St(n-k). \text{ Тогда ДИ для } \theta_i \text{ уровня } \gamma$$

$$\left(\hat{\theta}_i - t_{1-\frac{\gamma}{2}} \sqrt{\hat{\sigma}^2 A_{ii}}, \hat{\theta}_i + t_{1-\frac{\gamma}{2}} \sqrt{\hat{\sigma}^2 A_{ii}} \right)$$

Доверительная область для θ

Из теоремы об ортогональном разложении гауссовского вектора:

$$\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2, \quad \frac{1}{\sigma^2} \|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2$$

\Rightarrow

$$\frac{\|Z\hat{\theta} - Z\theta\|^2}{\|X - Z\hat{\theta}\|^2} \frac{n-k}{k} \sim F_{k, n-k}$$

$F_{k, n-k}$ - центральная статистика

u_γ - γ -квантиль распределения $F_{k, n-k}$, тогда

$$S(X) = \{\theta : \frac{\|Z\hat{\theta} - Z\theta\|^2}{\|X - Z\hat{\theta}\|^2} \frac{n-k}{k} < u_\gamma\}$$

доверительная область уровня γ для θ

Пример с нормальным распределением

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

$$X = (X_1, \dots, X_n)^T = (a, \dots, a)^T + \varepsilon^T, \text{ где } \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

(В этом случае $k=1$ и $Z = (1, \dots, 1)^T$)

$$\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-1}^2$$

$$\hat{\theta} = (Z^T Z)^{-1} Z^T X = \bar{X}$$

$$((Z^T Z)^{-1} = \frac{1}{n}, Z^T X = \sum X_i)$$

$$\|X - Z\hat{\theta}\|^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

Пусть $u_{1-\gamma}$ - $1 - \gamma$ квантиль для χ_{n-1}^2

$$P_{\theta, \sigma^2} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} > u_{1-\gamma} \right) = \gamma$$

$0 < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{u_{1-\gamma}}$ - доверительный интервал уровня γ

Далее

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 (Z^T Z)^{-1})$$

Уже получили, что $\bar{X} \sim \mathcal{N}\left(a, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\sqrt{n}(\bar{X}-a)}{\sqrt{\sigma^2}} \sim \mathcal{N}(0, 1)$

$$\hat{\theta} = \bar{X} \sim \mathcal{N}(0, 1) \perp \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 = \frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 \sim \chi_{n-1}^2$$

\Rightarrow

$$\frac{\frac{\sqrt{n}(\bar{X}-a)}{\sqrt{\sigma^2}}}{\sqrt{\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 \frac{1}{n-1}}} = \sqrt{\frac{n(n-1)}{\sum (X_i - \bar{X})^2}} (\bar{X} - a) \sim St(n-1)$$

Пусть z_γ - γ -квантиль $St(n-1)$ Тогда

$$P_{\theta, \sigma^2} \left(\bar{X} - z_{1+\frac{\gamma}{2}} \sqrt{s^2} < a < \bar{X} + z_{1+\frac{\gamma}{2}} \sqrt{s^2} \right) = \gamma$$

s^2 - выборочная дисперсия

18 Проверка статистических гипотез. Основная гипотеза и альтернатива. Критерий. Ошибка первого и второго рода. Уровень значимости, размер критерия и функция мощности. Равномерно наиболее мощные критерии и несмещенные критерии. Лемма Неймана-Пирсона.

Пусть \mathcal{P} - семейство распределений

$$\mathcal{P}_0, \mathcal{P}_1 \subset \mathcal{P}, \quad \mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$$

Определение Гипотезой называется утверждение вида $P \in \mathcal{P}_0$
т.е. $H_0 : P \in \mathcal{P}_0$

Пусть $X_1, \dots, X_n \sim P$

$S \in \mathcal{B}(\mathbb{R}^n)$ - **критерий** (т.е. критерий это любое борелевское множество)

Проверка гипотезы:

Если $(X_1, \dots, X_n) \in S$, то гипотеза H_0 отвергается (в пользу альтернативы, если она есть)

$$H_1 : P \in \mathcal{P}_1$$

Если $(X_1, \dots, X_n) \notin S$, то гипотеза H_0 не отвергается

Типы ошибок

Ошибка 1го рода — отвергнуть верную гипотезу

Ошибка 2го рода — не отвергнуть неверную гипотезу

Считается, что ошибка 1го рода более серьезная.

Определение Функция мощности $f(Q, S) = Q(x \in S)$,
 $S \in \mathcal{B}(\mathbb{R}^n)$; $Q \in \mathcal{P}$
 $P_0 \in \mathcal{P}_0$: $f(P_0, S)$ - для вероятности ошибки 1го рода
 $P_1 \in \mathcal{P}_1$: $1 - f(P_1, S)$ - для вероятности ошибки 2го рода

Неформальная постановка задачи: Сперва установим порог на ошибку первого рода, затем минимизируем ошибку 2го рода

ε — **уровень значимости** критерия S , если $\forall P \in \mathcal{P}_0 : f(P, S) \leq \varepsilon$

ε — **размер критерия** S , если $\varepsilon = \sup_{P \in \mathcal{P}_0} f(P, S)$

Пусть S, R - критерии уровня значимости ε
 S **мощнее** R , если $\forall P \in \mathcal{P}_1 : f(P, R) \leq f(P, S)$

S называется **равномерно наиболее мощным критерием** уровня значимости ε , если для любого критерия R уровня значимости ε , S мощнее R

Гипотеза H называется **простой**, если она имеет вид $P = P_0$

Пусть H_0, H_1 - простые гипотезы.
 $H_0 : P = P_0$
 $H_1 : P = P_1$

Либо P_1, P_0 оба дискретные, либо оба абсолютно непрерывные.

Пусть p_0, p_1 - их плотности.

Рассмотрим $S_\lambda = \{x : p_1(x) - \lambda p_0(x) \geq 0\}$, $\lambda \geq 0$

Определение Критерий S называется **несмещенным**, если

$$\sup_{P \in \mathcal{P}_0} f(P, S) \leq \inf_{P \in \mathcal{P}_1} f(P, S)$$

Лемма (Нейман-Пирсон)

1) Пусть R - критерий:

$$f(P_0, R) \leq f(P_0, S_\lambda)$$

Тогда $f(P_1, R) \leq f(P_1, S_\lambda)$

2) $f(P_0, S_\lambda) \leq f(P_1, S_\lambda)$

◀

1) Рассмотрим $P_1(x \in R) - \lambda P_0(x \in R)$

$$\begin{aligned} &= \int_R P_1(x) \mu(dx) - \lambda \int_R P_0(x) \mu(dx) = \int_R (P_1(x) - \lambda P_0(x)) \mu(dx) \\ &= \int_{\mathbb{R}} (P_1(x) - \lambda P_0(x)) I(x \in R) \mu(dx) \leq \int_{\mathbb{R}} (P_1(x) - \lambda P_0(x)) I(x \in R) I(x \in S_\lambda) \mu(dx) \\ &\leq \int_{\mathbb{R}} (P_1(x) - \lambda P_0(x)) I(x \in S_\lambda) \mu(dx) \\ &= \int_{S_\lambda} P_1(x) \mu(dx) - \lambda \int_{S_\lambda} P_0(x) \mu(dx) = P_1(x \in S_\lambda) - \lambda P_0(x \in S_\lambda) \end{aligned}$$

Получаем

$$P_1(x \in R) - P_1(x \in S_\lambda) \leq \lambda(P_0(X \in S_\lambda) - P_0(x \in R))$$

Справа ≤ 0 по условию

$$\Rightarrow P_1(x \in R) \leq P_1(x \in S_\lambda)$$

2) Пусть $\lambda \geq 0$

$$f(P_1, S_\lambda) = P_1(x \in S_\lambda) \geq \lambda P_0(x \in S_\lambda) \geq P_0(x \in S_\lambda)$$

Пусть $0 \leq \lambda < 1$

$$P_1(x \notin S_\lambda) = 1 - P_1(x \in S_\lambda)$$

$$\int p_1(x) I(x \notin S_\lambda) \mu(dx) = P_1(x \notin S_\lambda) \leq \lambda P_0(x \notin S_\lambda) = \lambda \int p_0(x) I(x \notin S_\lambda) \mu(dx)$$

$$\begin{aligned}
& \int p_1(x)I(x \notin S_\lambda)\mu(dx) - \lambda \int p_0(x)I(x \notin S_\lambda)\mu(dx) = \\
& = \int (p_1(x) - \lambda p_0(x))I(x \notin S_\lambda)\mu(dx) \leq 0 \\
& \Rightarrow P_1(x \notin S_\lambda) \leq \lambda P_0(x \notin S_\lambda) \leq P_0(x \notin S_\lambda) \\
& \Rightarrow 1 - P_1(x \in S_\lambda) \leq 1 - P_0(x \in S_\lambda) \rightarrow P_1(x \in S_\lambda) \geq P_0(x \in S_\lambda)
\end{aligned}$$

►

Следствие Пусть $P_0(x \in S_\lambda) = \varepsilon \Rightarrow S_\lambda$ - несмещенный р.н.м.к. уровня значимости ε

19 Монотонное отношение правдоподобий. Примеры. Другие способы нахождения наиболее мощных критериев.

$$H_0 : \theta \leq \theta_0 \quad (\theta = \theta_0)$$

$$H_1 : \theta > \theta_0$$

P_θ имеет плотность p_θ (либо все распределения дискретные, либо все абсолютно непрерывные)

$f_\theta(X_1, \dots, X_n)$ - функция правдоподобия

$$G(T(X)) = \frac{f_{\theta_2}(X_1, \dots, X_n)}{f_{\theta_1}(X_1, \dots, X_n)}, \quad \theta_2 > \theta_1$$

G - монотонная (одинаковая для всех $\theta_2 > \theta_1$)

Тогда семейство распределений обладает **монотонным отношением правдоподобия** по статистике $T(X)$

Теорема(б/д) Пусть $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ обладает возрастающим отношением правдоподобия. Пусть $P_{\theta_0}(T(X) \geq c) = \alpha$.

Тогда $\{T(X) \geq c\}$ - р.н.м.к. уровня α

Пример

Пусть $X_1, \dots, X_n \sim \text{Bern}(\theta)$

$$f_\theta = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

$$\frac{f_{\theta_2}}{f_{\theta_1}} = \left(\frac{\theta_2}{\theta_1}\right)^{\sum x_i} \left(\frac{1 - \theta_2}{1 - \theta_1}\right)^{n - \sum x_i} = G(\sum x_i)$$

G возрастает $\sum x_i \sim \text{Bin}(n, \theta_0)$

$$P_{\theta_0}(\sum x_i \geq c) = \alpha$$

Если у нас $H_0 : \theta \geq \theta_0$ ($\theta = \theta_0$), а $H_1 : \theta < \theta_0$, то сделаем замену:

$$\hat{\theta} = -\theta, \quad \hat{\theta}_0 = -\theta_0$$

$$\hat{P}_{\hat{\theta}} := P_{-\hat{\theta}} = P_{\theta}, \quad \hat{p}_{\hat{\theta}} = p_{\theta}, \quad \hat{f}_{\hat{\theta}} = f_{\theta}$$

Тогда получим:

$$\hat{H}_0 : \hat{\theta} \leq \hat{\theta}_0 \quad (\hat{\theta} = \hat{\theta}_0)$$

$$\hat{H}_1 : \hat{\theta} > \hat{\theta}_0$$

$$\hat{\theta}_2 > \hat{\theta}_1 \quad (\theta_2 < \theta_1) : \frac{\hat{f}_{\hat{\theta}_2}}{\hat{f}_{\hat{\theta}_1}} = \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = G(T(X)), \quad G \text{ возрастает}$$

Пример (другой способ нахождения р.н.м.к.)

$$X_1 \dots X_n \sim U[0; \theta]$$

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta < \theta_0$$

$$S = \{X_{(n)} \leq c\} \quad P_{\theta_0}\{X_{(n)} \leq c\} = \left(\frac{c}{\theta_0}\right)^n = \alpha \Rightarrow \sqrt[n]{\alpha}\theta_0 = c$$

Пусть: $R : P_{\theta_0}(X_1, \dots, X_n \in R) \leq \alpha$

Докажем, что $P_{\theta}(X_{(n)} \leq c) \geq P_{\theta}(X_1, \dots, X_n \in R) \quad \forall \theta < \theta_0$

(т.е. S - р.н.м.к. уровня значимости α)

1) $\theta \leq \sqrt[n]{\alpha}\theta_0 (= c) \Rightarrow P_{\theta}(X_{(n)} \leq c) = 1$ (с правее значений, куда мы попадаем)

2) $\sqrt[n]{\alpha}\theta_0 < \theta < \theta_0$

$$P_{\theta}(X_{(n)} \leq c) = \left(\frac{c}{\theta}\right)^n = \alpha \left(\frac{\theta_0}{\theta}\right)^n$$

$$P_{\theta}(X_1, \dots, X_n \in R) = \int_{[0, \theta]^n} I(X_1, \dots, X_n \in R) \frac{1}{\theta^n} dX_1 \dots dX_n =$$

$$\frac{1}{\theta^n} \left(\frac{\theta_0}{\theta}\right)^n \int_{[0, \theta_0]^n} I(X_1, \dots, X_n \in R) dX_1 \dots dX_n$$

$$\leq \left(\frac{\theta_0}{\theta}\right)^n \int_{[0, \theta_0]^n} \frac{1}{\theta_0^n} I((x_1, \dots, x_n) \in R) dx_1 \dots dx_n$$

$$= \left(\frac{\theta_0}{\theta}\right)^n P_{\theta_0}((X_1, \dots, X_n) \in R) \leq \alpha \left(\frac{\theta_0}{\theta}\right)^n = P_{\theta}(X_{(n)} \leq c)$$

20 Проверка гипотез в гауссовской линейной модели.

$$X = l + \Sigma, \quad \Sigma \sim \mathcal{N}(0, \sigma^2 I_n)$$

$$l = Z\theta, \quad \theta \in \mathbb{R}^k, \quad k < n$$

$H : T\theta = t$, где T - матрица размера $m \times k$, ранка $m \leq k$, $t \in \mathbb{R}^m$

$$\begin{aligned}
\hat{\theta} &= (Z^T Z)^{-1} Z^T X, \quad \hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 (Z^T Z)^{-1}) \\
\hat{\theta} - \theta &\sim \mathcal{N}(0, \sigma^2 (Z^T Z)^{-1}) \\
\Rightarrow T(\hat{\theta} - \theta) &\sim \mathcal{N}(0, \sigma^2 T(Z^T Z)^{-1} T^T) \\
B &:= T(Z^T Z)^{-1} T^T, \quad B \in \text{Mat}_{m \times m}, \quad \text{rank} B = m \\
\sqrt{B^{-1}} T(\hat{\theta} - \theta) &\sim \mathcal{N}(0, \sigma^2 I_m) \\
\Rightarrow \frac{1}{\sigma^2} \|\sqrt{B^{-1}} T(\hat{\theta} - \theta)\|^2 &\sim \chi_m^2 \\
\frac{1}{\sigma^2} \|\sqrt{B^{-1}} T(\hat{\theta} - \theta)\|^2 &= \frac{1}{\sigma^2} (T\hat{\theta} - T\theta)^T B^{-1} (T\hat{\theta} - T\theta) \\
\hat{\theta} \perp\!\!\!\perp X - Z\hat{\theta} &\Rightarrow \frac{1}{\sigma^2} (T\hat{\theta} - T\theta)^T B^{-1} (T\hat{\theta} - T\theta) \perp\!\!\!\perp \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2 \\
\frac{n-k}{m} \frac{(T\hat{\theta} - T\theta)^T B^{-1} (T\hat{\theta} - T\theta)}{\|X - Z\hat{\theta}\|^2} &\sim F_{m, n-k}
\end{aligned}$$

$u_{1-\alpha}$ - квантиль распределения $F_{m, n-k}$

$$P_{\theta} \left(\frac{n-k}{m} \frac{(T\hat{\theta} - T\theta)^T B^{-1} (T\hat{\theta} - T\theta)}{\|X - Z\hat{\theta}\|^2} > u_{1-\alpha} \right) = \alpha$$

\leftarrow критерий уровня значимости α для проверки H

21 Состоятельность критерия. Критерий хи-квадрат Пирсона: доказательство сходимости к хи-квадрат закону и доказательство состоятельности.

X_1, \dots, X_n - выборка
 $P(X_1 = a_j) = p_j, \quad \sum p_j = 1$
 $H_0 : p_j = p_j^0, \quad j \in \{1, \dots, m\}$
 $\mu_j = \sum_{i=1}^n I(x_i = a_j)$
 Рассмотрим статистику

$$\hat{\chi}_n = \sum_{j=1}^m \frac{(\mu_j - np_j^0)^2}{np_j^0}$$

критерий хи-квадрат Пирсона
Теорема

В предположении, что гипотеза H_0 верна

$$\hat{\chi}_n \xrightarrow{d} \chi_{m-1}^2$$

Как применять?

Пусть $u_{1-\varepsilon}$ - $(1 - \varepsilon)$ квантиль распределения χ_{m-1}^2

$\{\hat{\chi}_n > u_{1-\varepsilon}\}$ - наш критерий

$$P(\hat{\chi}_n > u_{1-\varepsilon}) \rightarrow \varepsilon, \quad n \rightarrow +\infty$$

(ε - асимптотический уровень доверия)

$\forall j \in \{1, \dots, m\} \quad np_j^0 \geq 10 \Rightarrow$ точность достаточно высока для применения критерия

Определение Критерий называется **состоятельным**, если $\forall P \in \mathcal{P}_1 : P((X_1, \dots, X_n) \in S) \rightarrow 1$ (т.е. вероятность ошибки 2го рода $\rightarrow 0$)

Утверждение Критерий хи-квадрат Пирсона состоятельный



$$\hat{\chi}_n \xrightarrow{d} \chi_{m-1}^2$$

Если $\exists p_j \neq p_j^0$ (т.е. $p_j \in \mathcal{P}_1$) $\Rightarrow P(\hat{\chi}_n > u_{1-\varepsilon}) \rightarrow 1$ - это надо показать, тогда это и будет определением состоятельности. Заметим, что не важно, что стоит на месте $u_{1-\varepsilon}$ Будет $\rightarrow 1$ и при любой другой константе, стоящей на месте $u_{1-\varepsilon}$

$$\hat{\chi}_n^2 = \sum_{j=1}^m \frac{n \left(\frac{\mu_j}{n} - p_j^0 \right)^2}{p_j^0}$$

Возьмем это $p_j \neq p_j^0$

$$\text{Т.к. } \frac{\mu_j}{n} = \frac{\sum I(X_i=a_j)}{n} \text{ (У.З.Б.Ч)} \xrightarrow{\text{п.н.}} \mathbf{E}I(X_1 = a_j) = p_j$$

\Rightarrow

$$\hat{\chi}_n^2 \geq \frac{n \left(\frac{\mu_j}{n} - p_j^0 \right)^2}{p_j^0} \xrightarrow{\text{п.н.}} \infty$$

Пояснение:

$$\frac{\mu_j}{n} \xrightarrow{\text{п.н.}} p_j \Rightarrow \text{(теорема о наследовании)}$$

$$\frac{\left(\frac{\mu_j}{n} - p_j^0 \right)^2}{p_j^0} \xrightarrow{\text{п.н.}} \frac{(p_j - p_j^0)^2}{p_j^0}$$

С ростом n критерий стремится к бесконечности (п.н.). Значит, вероятность того, что он будет больше какой-то константы

стремится к 1.



Теперь можно доказать теорему о сходимости к распределению хи-квадрат



Обозначим $Y_i = (I(X_i = a_1), \dots, I(X_i = a_m))^T$, $i \in \{1, \dots, n\}$

т.е. в этом векторы все нули, кроме одной координаты, где стоит 1.

X_i независимы, значит их индикаторы тоже независимы, а значит Y_1, \dots, Y_n независимы

$$\mathbf{E}Y_i = (p_1^0, \dots, p_m^0)^T$$

$$\mathbf{D}I(X_1 = a_j) = p_j^0 - (p_j^0)^2$$

$$\text{cov}(I(X_1 = a_i), I(X_1 = a_j)) = \mathbf{E}(I(X_1 = a_i)I(X_1 = a_j)) - p_i^0 p_j^0 = -p_i^0 p_j^0$$

Обозначим Σ - матрица ковариаций (размер $m \times m$)

Тогда на i -м диагональном элементе будет $p_i^0 - (p_i^0)^2$

На i, j элементе (вне диагонали) $-p_i^0 p_j^0$

$$\Sigma = A - \widehat{\Sigma}, \text{ где } A - \text{диагональная матрица, } A_{ii} = p_i^0, \widehat{\Sigma}_{ij} = p_i^0 p_j^0$$

$$\widehat{\Sigma} = (p_1^0, \dots, p_m^0)^T (p_1^0, \dots, p_m^0)$$

Обозначим $\varpi^0 = (p_1^0, \dots, p_m^0)$

По многомерной Ц.П.Т.

$$\sqrt{n} \left(\frac{Y_1 + \dots + Y_n}{n} - (\varpi^0)^T \right) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

$$Y_1 + \dots + Y_n = (\mu_1, \dots, \mu_m) =: \mu$$

A диагональная, значит A^{-1} диагональная $\rightarrow \sqrt{A^{-1}}$ диагональная, с элементами на диагонали: $\sqrt{A^{-1}}_{ii} = \frac{1}{\sqrt{p_i^0}}$

$$\sqrt{A^{-1}} \sqrt{n} \left(\frac{\mu}{n} - (\varpi^0)^T \right) \xrightarrow{d} \mathcal{N}(0, I_m - \sqrt{A^{-1}} (\varpi^0)^T \varpi^0 \sqrt{A^{-1}})$$

$$\sqrt{A^{-1}} (\varpi^0)^T \varpi^0 \sqrt{A^{-1}} = (\sqrt{p_1^0}, \dots, \sqrt{p_m^0})^T (\sqrt{p_1^0}, \dots, \sqrt{p_m^0})$$

Возьмем такую ортогональную матрицу C размера $m \times m$, что её первый столбец равен $(\sqrt{p_1^0}, \dots, \sqrt{p_m^0})^T$

$$\Rightarrow (\sqrt{p_1^0}, \dots, \sqrt{p_m^0}) C = (1, 0, \dots, 0)$$

Также отметим, что $C^T(\sqrt{p_1^0}, \dots, \sqrt{p_m^0})^T(\sqrt{p_1^0}, \dots, \sqrt{p_m^0})C$
 $= ((\sqrt{p_1^0}, \dots, \sqrt{p_m^0})C)^T(\sqrt{p_1^0}, \dots, \sqrt{p_m^0})C = (1, 0, \dots, 0)^T(1, 0, \dots, 0) =:$
 M

— матрица $m \times m$, где везде нули, кроме $M_{11} = 1$

Тогда

$$C\sqrt{A^{-1}}\sqrt{n}\left(\frac{\mu}{n} - (\varpi^0)^T\right) \xrightarrow{d} \mathcal{N}(0, I_m - M)$$

$$\Rightarrow \|C\sqrt{A^{-1}}\sqrt{n}\left(\frac{\mu}{n} - (\varpi^0)^T\right)\|^2 \xrightarrow{d} \chi_{m-1}^2$$

(т.к. слева сумма квадратов стандартных нормальных величин, и их $m-1$ т.к. 1-я величина имеет отклонение 0)

$$\|C\sqrt{A^{-1}}\sqrt{n}\left(\frac{\mu}{n} - (\varpi^0)^T\right)\|^2 = \|\sqrt{A^{-1}}\sqrt{n}\left(\frac{\mu}{n} - (\varpi^0)^T\right)\|^2 =$$

(ортогональная матрица не дает вклад в норму по определению)

$$= \sum_{j=1}^m \left(\frac{\sqrt{n}}{\sqrt{p_j^0}} \left(\frac{\mu_j}{n} - p_j^0 \right) \right)^2 = \sum_{j=1}^m \frac{(\mu_j - np_j^0)^2}{np_j^0} = \hat{\chi}_n$$

►

22 Критерий Колмогорова-Смирнова (без доказательства), удобная формула для вычисления значения статистики. Критерий Крамера-Мизеса-Смирнова (без доказательства), удобная формула для вычисления значения статистики.

Критерий Колмогорова-Смирнова

Теорема(б/д)

Пусть $X_1, \dots, X_n \sim F$, F -непрерывная функция распределения,
 $F_n^*(x) \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ - эмпирическая функция распределения.
 Тогда

$$\sqrt{n} \sup_x |F_n^*(x) - F(x)| \xrightarrow{d} K$$

Где K - распределение Колмогорова с функцией распределения:

$$\begin{cases} \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Критерий

Пусть $u_{1-\varepsilon}$ - квантиль распределения K

$$\{\sqrt{n} \sup_x |F_n^*(x) - F(x)| > u_{1-\varepsilon}\}$$

ε - асимптотический уровень значимости.

$$\sup_x |F_n^*(x) - F(x)| = \max_{1 \leq i \leq n} \max(|F(X_{(i)}) - \frac{i}{n}|, |F(X_{(i)}) - \frac{i-1}{n}|)$$

Критерий Крамера-Мизеса-Смирнова (ω^2)

$X_1, \dots, X_n \sim P$, F_n^* - эмпирическая функция распределения

$$\omega_n = n \int_{-\infty}^{+\infty} (F_n^*(x) - F(x))^2 dP$$

Теорема(б/д)

$$\omega_n \xrightarrow{d} a1$$

(Полагаем, что $X_{(0)} = -\infty$, $X_{(n+1)} = +\infty$)

$$\begin{aligned} \int_{-\infty}^{+\infty} (F_n^*(x) - F(x))^2 dP &= \int_{-\infty}^{+\infty} \left(\sum_{k=0}^n \frac{k}{n} I(X_{(k)} \leq x < X_{(k+1)}) - F(x) \right)^2 dP = \\ &= \frac{1}{3} - 2 \sum_{k=0}^n \int_{X_{(k)}}^{X_{(k+1)}} \frac{k}{n} F(x) dP + \sum_{k=0}^n \frac{k^2}{n^2} (F(X_{(k+1)}) - F(X_{(k)})) \end{aligned}$$

Пояснение:

$$\frac{1}{3} = \int_{-\infty}^{+\infty} F^2(x) dP$$

$$\sum_{k=0}^n \frac{k^2}{n^2} (F(X_{(k+1)}) - F(X_{(k)})) = \int_{-\infty}^{+\infty} \left(\sum_{k=0}^n \frac{k}{n} I(X_{(k)} \leq x < X_{(k+1)}) \right)^2 dP$$

Продолжим:

$$= \frac{1}{3} - \sum_{k=0}^n \frac{k}{n} (F^2(X_{(k+1)}) - F^2(X_{(k)})) + \sum_{k=0}^n \frac{k^2}{n^2} (F(X_{(k+1)}) - F(X_{(k)}))$$

Теперь покажем, что это равняется:

$$\frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n \left(F(X_{(k)}) - \frac{2k-1}{2n} \right)^2$$

$$\begin{aligned} \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n \frac{(2k-1)^2}{4n^2} &= \frac{1}{12n^2} + \frac{1}{4n^3} \sum_{k=1}^n (4k^2 - 4k + 1) = \\ &= \frac{1}{12n^2} + \frac{1}{4n^3} \left(4 \frac{n(n+1)(2n+1)}{6} - 4 \frac{n(n+1)}{2} + n \right) = \frac{1}{3} \end{aligned}$$

Разберемся со 2м слагаемым

$$\begin{aligned} \sum_{k=0}^n (F^2(X_{(k+1)}) - F^2(X_{(k)})) &= \\ \frac{1}{n} (F^2(X_{(2)}) - F^2(X_{(1)}) + 2F^2(X_{(3)}) - 2F^2(X_{(2)}) + \dots + n(1 - F^2(X_{(n)}))) &= \\ \setminus * F^2(X_{(n+1)}) = 1 * \setminus & \\ = -\frac{1}{n} \sum_{k=1}^n F^2(X_{(k)}) + 1 & \end{aligned}$$

Не забываем про то, что в нашей сумме это слагаемое было со знаком минус

Разберемся с последним слагаемым

$$\sum_{k=0}^n \frac{k^2}{n^2} (F(X_{(k+1)}) - F(X_{(k)})) = -\frac{1}{n^2} \sum_{k=1}^n (2k-1)F(X_{(k)}) + \frac{1}{n^2} n^2$$

\setminus * Аналогично второму, только $(s-1)^2 - s^2 = -2s + 1$ * \setminus
 Что и будет удвоенным произведением суммы квадратов минус 1 той суммы, к которой сводим.