

Содержание

1	Асимптотические оптимальные оценки	2
2	Проверка статистических гипотез	10
2.1	Критерий Фишера (F -критерий) в Гауссовской линейной регрессии	15
2.2	Критерий согласия Хи-квадрат Пирсона. Проверка простой гипотезы в схеме Бернулли.	20
3	Введение в робастное оценивание	25

1 Асимптотические оптимальные оценки

Пусть сл. векторы $\xi_n, \xi \in \mathbb{R}^K$, и определены на (Ω, \mathcal{F}, P) . Пусть функция распределения ξ_n есть $F_n(x)$, хар. ф-ция есть $\phi_n(t)$, а распределение есть Q_n . Для вектора ξ функцию распределения, хар. ф-цию и распределение обозначим $F(x)$, $\phi(t)$, Q соответственно.

Опр. 1. Функция распределения $F_n(x)$ сходится к $F(x)$ при $n \rightarrow \infty$ в основном (пишем $F_n(x) \Rightarrow F$), если $F_n(x) \rightarrow F(x) \forall x \in C(F)$

Опр. 2. Распределение Q_n сходится к распределению Q слабо (пишем $Q_n \xrightarrow{w} Q$), если \forall непрерывной и ограниченной $g : \mathbb{R}^K \rightarrow \mathbb{R}^1$

$$\int_{\mathbb{R}^K} g(x) Q_n(dx) \rightarrow \int_{\mathbb{R}^K} g(x) Q(dx)$$

или, эквивалентно, $Eg(\xi_n) \rightarrow Eg(\xi)$.

Теорема 1.

Следующие условия эквивалентны:

1. $F_n(x) \Rightarrow F$
2. $Q_n \xrightarrow{w} Q$
3. $\phi_n(t) \rightarrow \phi \forall t \in \mathbb{R}^K$

Если выполнено любое из условий 1 – 3, будем писать $\xi_n \xrightarrow{d} \xi$ и говорить, что ξ_n сходится к ξ по распределению.

Теорема 2 (О наследовании сходимости).

Пусть сл. векторы $\xi_n, \xi \in \mathbb{R}^K$, $H : \mathbb{R}^K \rightarrow \mathbb{R}^1$ Тогда:

1. Если $\xi_n \xrightarrow{d} \xi$, то $H(\xi_n) \xrightarrow{d} H(\xi)$
2. Если $\xi_n \xrightarrow{P} \xi$, то $H(\xi_n) \xrightarrow{P} H(\xi)$

Лемма Слуцкого

Пусть $\xi_n, \xi, \eta_n, a \in \mathbb{R}^1$, $\xi_n \xrightarrow{d} \xi$, а $\eta_n \xrightarrow{P} a$. Тогда:

1. $\xi_n + \eta_n \xrightarrow{d} \xi + a$
2. $\xi_n \eta_n \xrightarrow{d} a\xi$

Доказательство. Достаточно показать, что вектор

$$(\xi_n, \eta_n)^T \xrightarrow{d} (\xi, a)^T \quad (1)$$

Действительно, если (1) верно, то при $H(x, y) = x + y$ в силу Теоремы 2 получаем пункт 1 леммы, а при $H(x, y) = xy$ - пункт 2.

Для доказательства 1, проверим, что хар. ф-ция вектора $(\xi_n, \eta_n)^T$ сходится к хар. функции вектора $(\xi, \eta)^T$. Имеем:

$$|Ee^{it\xi_n + is\eta_n} - Ee^{it\xi + isa}| \leq |Ee^{it\xi_n + is\eta_n} - Ee^{it\xi_n + isa}| + |Ee^{it\xi_n + isa} - Ee^{it\xi + isa}| = \alpha_n + \beta_n$$

$$\alpha_n \leq E |e^{it\xi_n}(e^{it\eta_n+isa})| = E |e^{it\eta_n+isa}| = Eg(\eta_n), \quad g(x) \stackrel{\text{def}}{=} |e^{isx} - e^{isa}|$$

Ф-ция $g(x)$ непрерывна и ограничена, а т.к. $\eta_n \xrightarrow{d} a$, то в силу Теоремы 2 $Eg(\eta_n) \rightarrow Eg(a) = 0$.
Итак, $\alpha \rightarrow 0$.

$$\beta_n = |Ee^{isa}(e^{it\xi_n} - e^{it\xi})| = |e^{isa}E(e^{it\xi_n} - e^{it\xi})| = |E(e^{it\xi_n} - e^{it\xi})| \rightarrow 0$$

т.к. $\xi_n \xrightarrow{d} \xi$ и $\phi_n(t) \rightarrow \phi(t)$. □

Пусть наблюдаемые $X \sim P_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^K$, а $\hat{\theta}_n$ - оценка θ

Опр. 3. Если $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma(\theta)) \quad \forall \theta \in \Theta$ и ковариационная матрица $0 < \Sigma(\theta) < \infty$, то $\hat{\theta}_n$ называется асимптотической нормальной оценкой.

Опр. 4. Если $\hat{\theta}_n \xrightarrow{P} \theta \quad \forall \theta \in \Theta$, то $\hat{\theta}_n$ называется состоятельной оценкой.

Замечание 1. Далее $\theta \in \Theta \subseteq \mathbb{R}^1$, то есть θ и $\hat{\theta}_n$ - скаляры.

Если $\hat{\theta}_n$ - состоятельная оценка θ , то при больших n $\hat{\theta}_n \approx \theta$ с вероятностью, близкой к единице.

Если $\hat{\theta}_n$ - асимптотическая нормальная оценка θ (так как θ и $\hat{\theta}_n$ скаляры:

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)) \quad 0 < \sigma^2 < \infty, \quad \forall \theta \in \Theta$), то:

1. $\hat{\theta}_n$ - состоятельная оценка θ , так как $\hat{\theta}_n - \theta = n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{P} 0$ в силу п. 2 леммы Случко.
2. Скорость сходимости $\hat{\theta}_n$ к θ есть $O(\sqrt{n})$
3. При больших n со сл. в. $\sqrt{n}(\hat{\theta}_n - \theta)$ можно обращаться (с осторожностью!) как с Гауссовской величиной.

Например, пусть дисперсия предельного Гауссовского закона $\sigma^2(\theta)$ будет непрерывной ф-цией θ . Тогда

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} = \underbrace{\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)}}_{\xrightarrow{P} 1} \xrightarrow{d} \eta \sim N(0, 1)$$

в силу п. 2 леммы Случко. Значит,

$$P_\theta \left(\left| \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} \right| < \xi_{1-\alpha/2} \right) \rightarrow P(|\eta| < \xi_{1-\alpha/2}) = 1 - \alpha$$

То есть примерно с вероятностью $1 - \alpha$ выполнено неравенство, или эквивалентно раскроем по модулю

$$\underbrace{\hat{\theta}_n - n^{-1/2}\sigma(\hat{\theta}_n)\xi_{1-\alpha/2} < \theta < \hat{\theta}_n + n^{-1/2}\sigma(\hat{\theta}_n)\xi_{1-\alpha/2}}_{\text{Асимптотический доверительный интервал уровня } 1 - \alpha}$$

4. Асимптотические Гауссовские оценки можно сравнивать между собой:

Если $\sqrt{n}(\hat{\theta}_{i,n} - \theta) \xrightarrow{d} N(0, \sigma_i^2(\theta))$, $i = 1, 2, \dots$, то можно посчитать асимптотическую относительную эффективность (АОЭ):

$$e_{1,2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$$

Напомним, $e_{1,2} = \lim_{n \rightarrow \infty} \frac{n'(x)}{n(x)}$, где $\sqrt{n}(\hat{\theta}_{1,n} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta))$ и $\sqrt{n}(\hat{\theta}_{2,n} - \theta) \xrightarrow{d} N(0, \sigma_2^2(\theta))$.

Вопрос: Есть ли такая оценка θ_n^* , что АОЭ $e_{\theta_n^*, \hat{\theta}_n}(\theta) \geq 1 \forall \hat{\theta}_n$ и всех $\theta \in \Theta$, то есть эффективнее всех остальных?

Если да, то θ_n^* требует не больше наблюдений, чем любая $\hat{\theta}_n$, чтобы достичь одинаковой с $\hat{\theta}_n$ точности. Ясно, что пределная дисперсия $\sqrt{n}(\theta_n^* - \theta)$ должна быть не больше асимптотической дисперсии $\sqrt{n}(\hat{\theta}_n - \theta)$ для любой асимптотической Гауссовской оценки $\hat{\theta}_n$. Но какова самая маленькая асимптотическая дисперсия у $\sqrt{n}(\hat{\theta}_n - \theta)$?

Теорема Бахадура

Пусть X_1, \dots, X_n - н. о. р. сл. в., X_1 имеет плотность вероятности $f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^1$, по мере ν . Пусть выполнены следующие условия:

1. Θ - интервал.
2. Носитель $N_f = \{x : f(x, \theta) > 0\}$ не зависит от θ .
3. $\forall x \in N_f$ плотность $f(x, \theta)$ дважды непрерывно дифференцируема по θ
4. Интеграл $\int f(x, \theta) \nu(dx)$ можно дважды дифференцировать по θ , внося знак дифференцирования под знак интеграла.
5. Информация Фишера $0 < i(\theta) < \infty \forall \theta \in \Theta$
6. $\left| \frac{\partial^2}{\partial \theta^2} \ln(f(x, \theta)) \right| \leq M(x) \forall x \in N_f, \theta \in \Theta, E_\theta M(X_1) < \infty$

Тогда, если $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$, то $\sigma^2(\theta) \geq \frac{1}{i(\theta)}$ всюду за исключением множества Лебеговой меры нуль.

Замечание 2. Если вдобавок $\sigma^2(\theta)$ и $i(\theta)$ непрерывны, то $\sigma^2(\theta) \geq \frac{1}{i(\theta)}$ при всех $\theta \in \Theta$.

Доказательство. Без доказательства. □

Опр. 5. Если $\theta, \hat{\theta}_n \in \mathbb{R}^1$ и $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{i(\theta)})$, $n \rightarrow \infty, \forall \theta \in \Theta$, причем $0 < i(\theta) < \infty$, то $\hat{\theta}_n$ называется **асимптотически эффективной оценкой**.

Вопрос: Вообще можно ли найти такую оценку $\hat{\theta}_n$? Да

Дальше $X = (X_1, \dots, X_n)$, $X \sim P_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^1$. Условие (A):

1. Θ - интервал, $P_{\theta_1} \neq P_{\theta_2}$ при $\theta_1 \neq \theta_2$.
2. X_1, \dots, X_n - независимые одинаково распределенные случайные величины
3. X_1 имеет плотность вероятности $f(x, \theta)$ по мере ν
4. Носитель $N_f = \{x : f(x, \theta) > 0\}$ не зависит от θ .
5. Плотность вектора X есть $p(x, \theta) = \prod_{i=1}^n f(x_i, \theta)$.

Опр. 6. Функция $p(X, \theta)$ как функция θ при фиксированном X называется **правдоподобием** функции.

$$L_n(X, \theta) = \ln p(X, \theta) = \sum_{i=1}^n \ln f(X_i, \theta)$$

называется логарифмическим правдоподобием.

Пусть θ_0 будет истинное значение параметра.

Лемма 1 (Неравенство Йенсена). Пусть $g(x)$ выпукла книзу борелевская функция, $E|\xi| < \infty$, $E|g(\xi)| < \infty$. Тогда $g(E\xi) \leq Eg(\xi)$. Если ξ не является почти наверное константой и g строго выпукла, то неравенство строгое.

Теорема 3 (Экстремальное свойство правдоподобия).

Пусть выполнено Условие (A). Пусть $E_{\theta_0} |\ln f(X_1, \theta)| < \infty$, $\forall \theta \in \Theta$. Тогда

$$P_{\theta_0}(p(X, \theta_0) > p(X, \theta)) \rightarrow 1, \quad n \rightarrow \infty, \quad \theta_0 \neq \theta$$

Доказательство.

$$p(X, \theta_0) > p(X, \theta) \Leftrightarrow \ln p(X, \theta_0) > \ln p(X, \theta) \Leftrightarrow$$

$$\eta_n \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n \ln \left(\frac{f(X_i, \theta)}{f(X_i, \theta_0)} \right) < 0$$

То есть надо показать, что $P_{\theta_0}(\eta_n < 0) \rightarrow 1$. Но по слабому закону больших чисел:

$$\eta_n = n^{-1} \sum \ln \left(\frac{f(X_i, \theta)}{f(X_i, \theta_0)} \right) \xrightarrow{P} E_{\theta_0} \ln \left(\frac{f(X_1, \theta)}{f(X_1, \theta_0)} \right)$$

Возьмем функцию $-\ln x$ - строго выпукла вниз и $\frac{f(X_1, \theta)}{f(X_1, \theta_0)}$ не является п.н. константой (так как иначе если плотности п.н. совпадают, то и распределения при разных значениях совпадают, что противоречит Условию(A)(1)).

В силу неравенства Йенсена:

$$E_{\theta_0} \ln \frac{f(X_1, \theta)}{f(X_1, \theta_0)} < \ln E_{\theta_0} \frac{f(X_1, \theta)}{f(X_1, \theta_0)} = \ln \int_{N_f} \frac{f(x, \theta)}{f(x, \theta_0)} f(x, \theta_0) \nu(dx) = \ln 1 = 0$$

Но если η_n сходится по вероятности к отрицательному числу, то $P_{\theta_0}(\eta_n < 0) \rightarrow 1$ □

В силу теоремы 3 естественно брать оценкой то значение θ , которое максимизирует $p(X, \theta)$ при данном X

Опр. 7. Случайная величина $\hat{\theta}_n \in \Theta$ называется **оценкой максимального правдоподобия (о.м.п.)**, если $p(X, \hat{\theta}_n) = \max_{\theta \in \Theta} p(X, \theta)$, или эквивалентно $L_n(X, \hat{\theta}_n) = \max_{\theta \in \Theta} L_n(X, \theta)$

Итак, о.м.п $\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(X, \theta)$.

Если в $\forall \theta \in \Theta$ максимум не достигается, то о.м.п. не существует.

Если Θ - интервал, $L_n(X, \theta)$ - гладкая по θ функция, то θ удовлетворяет уравнению правдоподобия

$$\frac{\partial}{\partial \theta} L_n(X, \theta) = 0 \quad (2)$$

Теорема 4 (О состоятельности решения уравнения правдоподобия).

Пусть выполнено Условие (А). Пусть $\forall x \in N_f \exists$ непрерывная производная $f'_\theta(x, \theta)$. Тогда уравнение (2) с вероятностью, стремящейся к 1 при $n \rightarrow \infty$ имеет решение $\in \Theta$. При этом среди всех таких решений есть такой корень $\hat{\theta}_n$, что он является состоятельной оценкой θ_0

Доказательство. Пусть $S_n = \{\omega\}$, при которых уравнение (2) имеет решение для $\theta \in \Theta$. Тогда теорема 4 утверждает:

1. $P_{\theta_0}(S_n) \rightarrow 1$.
2. Существует такое решение $\hat{\theta}_n \in \Theta$, что

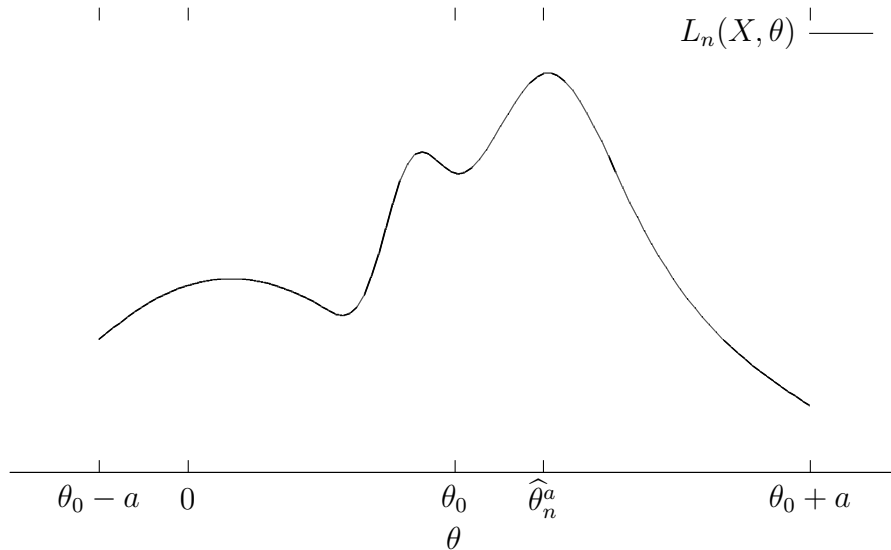
$$P_{\theta_0} \left(\left| \hat{\theta}_n - \theta_0 \right| < \varepsilon, S_n \right) \rightarrow 1, \quad n \rightarrow \infty, \quad \forall \varepsilon > 0$$

Докажем пункт 1: Выберем малое $a > 0$ так, что на $(\theta_0 - a, \theta_0 + a) \subseteq \Theta$. Пусть

$$S_n^a = \{\omega : L_n(X, \theta_0) > L_n(X, \theta_0 - a), L_n(X, \theta_0) > L_n(X, \theta_0 + a)\}$$

В силу теоремы 3 $P_{\theta_0}(S_n^a) \rightarrow 1$

При $\omega \in S_n^a$ функция $L_n(X, \theta)$ имеет локальный максимум $\hat{\theta}_n^a$ на интервале $(\theta_0 - a, \theta_0 + a)$



Значит, $\frac{\partial}{\partial \theta} L_n(X, \hat{\theta}_n^a) = 0$. Тогда $P_{\theta_0}(S_n) \geq P_{\theta_0}(S_n^a) \rightarrow 1$, так как $S_n^a \subseteq S_n$, и пункт 1 доказан.

Докажем пункт 2: $\forall n$ при $\omega \in S_n$ может существовать целое множество корней $\{\theta_n^*\}$. Выберем в этом множестве корень $\hat{\theta}_n$, ближайший к θ_0 . Это можно сделать, так как функция $\frac{\partial}{\partial \theta} L_n(x, \theta)$ непрерывна по θ , и последовательность корней есть корень. Этот корень $\hat{\theta}_n$ и есть состоятельная оценка θ . Покажем это:

\forall малого $\varepsilon > 0$:

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) \geq P_{\theta_0}(|\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon, S_n^\varepsilon) \quad (3)$$

Так как $S_n^\varepsilon \subseteq S_n$, $(\omega : |\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon) \subseteq (\omega : |\hat{\theta}_n - \theta_0| < \varepsilon)$

Но $P_{\theta_0}(|\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon, S_n^\varepsilon) \underset{\text{т.к. события из } S_n^\varepsilon \text{ лежат в } |\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon}{=} P_{\theta_0}(S_n^\varepsilon) \rightarrow 1$, значит в силу (3)

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) \rightarrow 1$$

□

Замечание 3. Пусть

$$\theta_n^* = \begin{cases} \text{сост. корню уравнения правдоподобия, если он суц.} \\ \theta', \theta' \in \Theta, \text{ иначе} \end{cases}$$

Тогда случайная величина θ_n^* всегда определена, и $\theta_n^* \xrightarrow{P} \theta_0$, так как

$$P(|\theta_n^* - \theta_0| < \varepsilon) = P(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) + P(|\theta' - \theta_0| < \varepsilon, \bar{S}_n) \rightarrow 1$$

Ясно, что

$$\frac{\partial}{\partial \theta} L_n(X, \theta_n^*) = \bar{o}_p(1) \quad (4)$$

Так как производная отлична от нуля только на \bar{S}_n .

Будем называть θ_n^* **обобщенным состоятельным корнем уравнения правдоподобия**

Теорема 5 (Об асимптотической эффективности состоятельности решения).

Пусть $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р. сл.в., и удовлетворяются предположения Теоремы Бахадура, в которых условия 3 и 6 заменены на предположения о третьей, а не второй производной. То есть

$$\left| \frac{\partial^3}{\partial \theta^3} \ln f(x, \theta) \right| \leq M(x) \quad \forall x \in N_f, \quad \forall \theta \in \Theta, \quad E_{\theta_0} M(X_1) < \infty$$

Тогда, если θ_n^* - обобщенный состоятельный корень из теоремы 4, то

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, \frac{1}{i(\theta_0)})$$

То есть θ_n^* - асимптотическая эффективная оценка.

Доказательство. Будем обозначать $\frac{\partial}{\partial \theta} L_n(X, \theta), \frac{\partial^2}{\partial \theta^2} L_n(X, \theta), \dots$ через $L'_n(\theta), L_n^{(2)}(\theta), \dots$
 Для фиксированного X в силу формулы Тейлора и последнего замечания:

$$\bar{\bar{o}}_p(1) = L'_n(\theta_n^*) = L'_n(\theta_0) + L_n^{(2)}(\theta_0)(\theta_n^* - \theta_0) + \frac{1}{2} L_n^{(3)}(\tilde{\theta}_n)(\theta_n^* - \theta_0)^2, \quad \tilde{\theta}_n \in (\theta_0, \theta_n^*)$$

Отсюда,

$$\sqrt{n}(\theta_n^* - \theta_0) = - \frac{n^{-1/2} L'_n(\theta_0) + \bar{\bar{o}}_p(1)}{n^{-1}(L_n^{(2)}(\theta_0) + \frac{1}{2} L_n^{(3)}(\tilde{\theta}_n)(\theta_n^* - \theta_0))} \quad (5)$$

Рассмотрим числитель (5) и покажем, что

$$n^{-1/2} L'_n(\theta_0) = n^{-1/2} \sum_{i=1}^n \frac{f'_\theta(X_i, \theta_0)}{f(X_i, \theta_0)} \xrightarrow{d} \xi \sim N(0, i(\theta_0)) \quad (6)$$

Действительно,

$$\begin{aligned} E_{\theta_0} \frac{f'_{\theta_0}(X_1, \theta_0)}{f(X_1, \theta_0)} &= \int_{N_f} \frac{f'_\theta(x, \theta_0)}{f(x, \theta_0)} f(x, \theta_0) \nu(dx) = 0 \\ D_{\theta_0} \frac{f'_{\theta_0}(X_1, \theta_0)}{f(X_1, \theta_0)} &= E_{\theta_0} \left(\frac{\partial}{\partial \theta} \ln f(X_1, \theta_0) \right)^2 - \underbrace{\left(E_{\theta_0} \frac{f'_{\theta_0}(X_1, \theta_0)}{f(X_1, \theta_0)} \right)^2}_{=0} \stackrel{\text{по опр.}}{=} i(\theta_0) \end{aligned}$$

Так как f, f' - борелевские функции, то случайные величины $\left\{ \frac{f'_\theta(X_i, \theta_0)}{f(X_i, \theta_0)}, i = 1, \dots, n \right\}$ - н.о.р.,
 соотношение (6) следует из Центр. пред. Теоремы.

В силу Леммы Слуцкого числитель (5) $\xrightarrow{P} N(0, i(\theta_0))$

Теперь рассмотрим знаменатель (5):

$$n^{-1} L_n^{(2)}(\theta_0) = n^{-1} \sum_{i=1}^n \left[\frac{f_\theta^{(2)}(X_i, \theta_0)}{f(X_i, \theta_0)} - \left(\frac{f'_\theta(X_i, \theta_0)}{f(X_i, \theta_0)} \right)^2 \right] \xrightarrow{P} -i(\theta) \quad (7)$$

Действительно, в силу ЗБЧ

$$\begin{aligned} n^{-1} \sum_{i=1}^n \frac{f_\theta^{(2)}(X_i, \theta_0)}{f(X_i, \theta_0)} &\xrightarrow{P} E_{\theta_0} \frac{f_\theta^{(2)}(X_1, \theta_0)}{f(X_1, \theta_0)} = \int_{N_f} \frac{f_\theta^{(2)}(x, \theta_0)}{f(x, \theta_0)} f(x, \theta_0) \nu(dx) = 0 \\ n^{-1} \sum_{i=1}^n \left(\frac{f'_\theta(X_i, \theta_0)}{f(X_i, \theta_0)} \right)^2 &\xrightarrow{P} E_{\theta_0} \left(\frac{\partial}{\partial \theta} \ln f(X_1, \theta_0) \right)^2 = i(\theta) \end{aligned}$$

Применяя лемму Слуцкого, получим (7).

Далее рассмотрим второе слагаемое в знаменате (5)

$$\left| \frac{1}{2n} L_n^{(3)}(\tilde{\theta}_n)(\theta_n^* - \theta_0) \right| \leq \frac{1}{2} |\theta_n^* - \theta_0| n^{-1} \sum_{i=1}^n M(X_i) \xrightarrow[\text{л. Слуцкого}]{P} 0 \quad (8)$$

В силу (7) и (8) и Леммы Слуцкого знаменатель (5) сходится по вероятности к $-i(\theta_0)$
 Значит, что вся дробь (5) сходится по распределению к $\frac{1}{i(\theta_0)} \xi \sim N(0, \frac{1}{i(\theta_0)})$ \square

Оценки максимального правдоподобия для векторного параметра

Пусть $X = (X_1, \dots, X_n)$ - н.о.р., $X_1 \sim f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, Θ - открытое множество

Тогда логарифмическое правдоподобие имеет вид

$$L_n(X, \theta) = \sum_{i=1}^n \ln f(X_i, \theta)$$

Система уравнений правдоподобия

$$\frac{\partial L_n(X, \theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

При условиях регулярности, похожих на условия теоремы 5, показываются:

1. С вероятностью, стремящейся к единице при $n \rightarrow \infty$, система уравнений (1) имеет такое решение $\hat{\theta}_n \in \Theta$, что $\hat{\theta}_n$ сходится к истинному значению θ_0 .
2. Соответствующая оценка θ_n^* асимптотически нормальна. А именно

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)), \quad n \rightarrow \infty$$

Здесь $I(\theta) > 0$ - матрица информации Фишера, то есть

$$I(\theta) = (I_{ij}(\theta)), \quad I_{ij}(\theta) = E_{\theta} \left\{ \frac{\partial \ln f(X, \theta)}{\partial \theta_i} \cdot \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right\}$$

Пример. $X = (X_1, \dots, X_n)$, где $\{X_i\}$ - н.о.р., $X_1 \sim N(0, \sigma^2)$, $a < \theta < b$, a и b - известные конечные числа, дисперсия σ^2 известна. Построим асимптотически эффективную оценку θ_n^* для θ .

Здесь $p(x, \theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}$, значит

$$L_n(X, \theta) = \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta)^2$$

Уравнение правдоподобия имеет вид

$$\frac{\partial L_n(X, \theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \theta) = 0$$

Его решение существует и единственно, это \bar{X} , причем в т. $\theta = \bar{X}$ $L_n(X, \theta)$ достигает максимума, так как $\frac{\partial^2 L_n(X, \bar{X})}{\partial \theta^2} = -\frac{1}{\sigma^2} < 0$

Таким образом, если $a < \bar{X} < b$, то о.м.п. существует и равна \bar{X} , в противном случае о.м.п. не существует. Если положить

$$\theta_n^* = \begin{cases} \bar{X}, & a < \bar{X} < b \\ \frac{a+b}{2}, & \bar{X} \notin (a, b) \end{cases} \quad (9)$$

То в силу теоремы 5 (её условия выполнены, проверьте сами), θ_n^* - асимптотически эффективная оценка, то есть

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, \sigma^2) \quad (10)$$

Напомним, что в этой модели $i(\theta) = \frac{1}{\sigma^2}$. Справедливость (10) с θ_n^* из (9) легко проверить непосредственно.

Пример. Если Θ - компакт (то есть отрезок $[a, b]$), то о.м.п. существует всегда, так как непрерывная функция на отрезке всегда достигает своего максимума. Значит значение о.м.п.

$$\theta_n^* = \begin{cases} \bar{X}, & a < \bar{X} < b \\ a, & \bar{X} < a \\ b, & \bar{X} > b \end{cases}$$

Но на границах теряется асимптотическая Гауссовость.

2 Проверка статистических гипотез

$X = (X_1, \dots, X_n)$ имеет плотность вероятности $p(X, \theta)$ по мере μ , $\theta \in \Theta \subseteq \mathbb{R}^1$

Опр. 1. Предположение вида $H_0 : \theta \in \Theta_0$, где $\Theta_0 \in \Theta$, называется параметрической гипотезой. Альтернатива $H_1 : \theta \in \Theta_1$, где $\Theta_1 \in \Theta \setminus \Theta_0$

Опр. 2. Если $\Theta_0(\Theta_1)$ состоит из одной точки, то гипотеза H_0 (альтернатива H_1) называется простой. В противном случае $H_0(H_1)$ - сложная

Постановка задачи:

Необходимо построить правило (статистический критерий - *test*), который позволяет заключить, согласуется ли наблюдение X с H_0 или нет.

Правило.

Выберем в множестве значений x вектора X (у нас либо $x = \mathbb{R}^n$, либо $x = N_p \subseteq \mathbb{R}^n$ - носитель плотности) подмножество S . Если $X \in S$, то H_0 отвергается и принимается H_1 . Если $X \in \bar{S} = X \setminus S$, то H_0 принимается.

Опр. 3. Множество S называется критическим множеством или критерием, \bar{S} - область принятия гипотезы.

Опр. 4. *Ошибка 1-го рода* - принять H_1 , когда верна H_0 . Вероятность ошибки 1-го рода $\alpha = P(H_1|H_0)$ (это условная запись, а не условная вероятность). *Ошибка 2-го рода* - принять H_0 , когда верна H_1 . Вероятность ошибки 2-го рода $\beta = P(H_0|H_1)$.

Опр. 5. *Мощность критерия* S называется функция $W(S, \theta) = W(\theta) \stackrel{\text{def}}{=} P_\theta(X \in S)$ (вероятность отвергнуть H_0 , когда значение параметра есть θ).

Тогда

$$\begin{aligned} \alpha &= \alpha(\theta) = W(\theta), \quad \theta \in \Theta_0; \\ \beta &= \beta(\theta) = 1 - W(\theta), \quad \theta \in \Theta_1 \end{aligned}$$

Опр. 6. Обычно H_0 более важна. Поэтому рассматривают критерии такие, что

$$\alpha(\theta) = W(\theta) = P_\theta(X \in S) \leq \alpha \quad \forall \theta \in \Theta_0$$

Число α называют **уровнем значимости критерия**. Пишут S_α - критерий уровня α . Обычно α - маленькое число, которое мы задаем сами.

Опр. 7. Если критерий $S_\alpha^* \in \{S_\alpha\}$ и $\forall \theta \in \Theta_1$ и $\forall S_\alpha W(S_\alpha^*, \theta) \geq W(S_\alpha, \theta)$, то критерий S_α^* называется **РНМ-критерием (равномерно наиболее мощным)**.

Если $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$ (то есть H_0 и H_1 - простые), то задача отыскания РНМ-критерия заданного уровня α имеет вид:

$$\begin{aligned} P_{\theta_0}(X \in S_\alpha^*) &\leq \alpha, \\ P_{\theta_1}(X \in S_\alpha^*) &\geq P_{\theta_1}(X \in S_\alpha) \quad \forall S_\alpha \end{aligned}$$

Положим для краткости: $p_0(x) \stackrel{\text{def}}{=} p(x, \theta_0)$, $E_0 = E_{\theta_0}$, $p_1(x) = p(x, \theta_1)$, $E_1 = E_{\theta_1}$
Введем множество

$$S(\lambda) = \{x : p_1(x) - \lambda p_0(x) > 0\}, \lambda > 0$$

Теорема 1 (Лемма Неймана-Пирсона).

Пусть для некоторого $\lambda > 0$ и критерия R (когда X попадает в R , то H_0 отвергается) выполнено:

$$1. P_0(X \in R) \leq P_0(X \in S(\lambda))$$

Тогда:

$$2. P_1(X \in R) \leq P_1(X \in S(\lambda))$$

$$3. P_1(X \in S(\lambda)) \geq P_0(X \in S(\lambda))$$

Замечание 1. $X \in S(\lambda) \Leftrightarrow \frac{p_1(x)}{p_0(x)} > \lambda$. Так как $p_1(X)$ и $p_0(X)$ - правдоподобие, то критерий называется критерием отношения правдоподобия Неймана-Пирсона.

Замечание 2. Утверждение 3 для $S(\lambda)$ означает, что

$$P(H_1 | H_1) \geq P(H_1 | H_0) \Leftrightarrow W(S(\lambda), \theta_1) \geq W(S(\lambda), \theta_0)$$

Это свойство называется несмещенностью критерия $S(\lambda)$

Доказательство. Далее для краткости $S(\lambda) = S$. Пусть $I_R(x) = \begin{cases} 1, x \in R \\ 0, x \notin R \end{cases}$, $I_S(x)$ определяем аналогично. Тогда Условие (А) имеет вид:

$$E_0 I_R(x) \leq E_0 I_S(x) \tag{1}$$

Докажем пункт 2: Верно неравенство

$$I_R(x)[p_1(x) - \lambda p_0(x)] \leq I_S(x)[p_1(x) - \lambda p_0(x)] \tag{2}$$

Действительно, если $(p_1(x) - \lambda p_0(x)) > 0$, то $I_S(x) = 1$ и (2) очевидно.

Если же $p_1(x) - \lambda p_0(x) \leq 0$, то правая часть (2) есть ноль, а левая \leq нуля.

Итак, (2) верно: интегрируем это неравенство по $x \in \mathbb{R}^n$:

$$\begin{aligned} E_1 I_R(X) - \lambda E_0 I_R(X) &\leq E_1 I_S(X) - \lambda E_0 I_S(X) \\ E_1 I_S(X) - E_1 I_R(X) &\geq \underbrace{\lambda [E_0 I_S(X) - E_0 I_R(X)]}_{\geq 0 \text{ по условию (1)}} \end{aligned} \tag{3}$$

В силу (1), (3) и условия $\lambda > 0$ получаем:

$$E_1 I_S(X) \geq E_1 I_R(X)$$

Докажем пункт 3: Пусть $\lambda \geq 1$. Из определения S $p_1(x) > p_0(x) \forall x \in S$. Отсюда

$$P_0(X \in S) = \int_{R^n} I_S(X) p_0(x) \mu(dx) \leq \int_{R^n} I_S(X) p_1(x) \mu(dx) = P_1(X \in S)$$

То есть $P(H_1 | H_0) \leq P(H_1 | H_1)$

Пусть $\lambda < 1$. Рассмотрим $\bar{S} = \{x : p_1(x) \leq \lambda p_0(x)\}$. При $\lambda < 1$ $p_1(x) < p_0(x)$ при $x \in \bar{S}$. Отсюда

$$P_1(X \in \bar{S}) = \int_{R^n} I_{\bar{S}}(X) p_1(x) \mu(dx) \leq \int_{R^n} I_{\bar{S}}(X) p_0(x) \mu(dx) = P_0(X \in \bar{S})$$

То есть $1 - P_1(X \in S) \leq 1 - P_0(X \in S)$, откуда $P_1(X \in S) \geq P_0(X \in S)$ □

Пример. $X = (X_1, \dots, X_n), \{X_i\}$ - н.о.р., $X_1 \sim N(\theta, \sigma^2)$, дисперсия σ^2 известна. Построим наиболее мощный критерий для проверки $H_0 : \theta = \theta_0$ против $H_1 : \theta = \theta_1$ (в случае $\theta_1 > \theta_0$). Уровень значимости возьмем α .

1. Имеем

$$p_0 = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0)^2 \right\}, \quad p_1 = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_1)^2 \right\};$$

$$S(\lambda) = \{x : p_1(x) - \lambda p_0(x) > 0\} \underset{\text{делим на } p_0}{\Leftrightarrow} \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \theta_1)^2 - (x_i - \theta_0)^2] \right\} > \lambda \Leftrightarrow$$

$$\Leftrightarrow \sum_{i=1}^n [(x_i - \theta_1)^2 - (x_i - \theta_0)^2] < \lambda_1 = -2\sigma^2 \ln \lambda \underset{\text{арифметика}}{\Leftrightarrow} (\theta_0 - \theta_1) \sum_{i=1}^n x_i \leq \lambda_2 \Leftrightarrow$$

$$\Leftrightarrow \sum_{i=1}^n x_i > \tilde{\lambda}, \quad \tilde{\lambda}(\lambda, n, \sigma^2, \theta_0, \theta_1)$$

Итак,

$$S(\lambda) = \left\{ x : \sum_{i=1}^n x_i > \tilde{\lambda} \right\} \text{ при некотором } \tilde{\lambda}$$

2. Определим $\tilde{\lambda} = \tilde{\lambda}_\alpha$ из уравнения

$$\alpha = P_{\theta_0}(X \in S(\tilde{\lambda}_\alpha)) = P_{\theta_0} \left(\sum_{i=1}^n X_i > \tilde{\lambda}_\alpha \right)$$

Преобразуем левую сумму в стандартную Гауссовскую величину. Тогда

$$\alpha = P_{\theta_0} \left(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \theta_0) > \frac{\tilde{\lambda}_\alpha - n\theta_0}{\sqrt{n}\sigma} \right) = 1 - \Phi \left(\frac{\tilde{\lambda}_\alpha - n\theta_0}{\sqrt{n}\sigma} \right)$$

так как $\frac{1}{\sqrt{n}\sigma} \sum_i (X_i - \theta_0) \sim N(0, 1)$ при H_0 .

Значит $\Phi \left(\frac{\tilde{\lambda}_\alpha - n\theta_0}{\sqrt{n}\sigma} \right) = 1 - \alpha$, $\left(\frac{\tilde{\lambda}_\alpha - n\theta_0}{\sqrt{n}\sigma} \right) = \xi_{1-\alpha}$ - квантиль станд. норм. закона уровня $1 - \alpha$. Окончательно, $\tilde{\lambda}_\alpha = n\theta_0 + \sqrt{n}\sigma \xi_{1-\alpha}$

3. Положим $S_\alpha^* = \{x : \sum_{i=1}^n x_i > \tilde{\lambda}_\alpha\}$ Тогда $P_{\theta_0}(X \in S_\alpha^*) = \alpha$, и:

$$\forall S_\alpha \quad P_{\theta_0}(X \in S_\alpha) \leq \alpha = P_{\theta_0}(X \in S_\alpha^*)$$

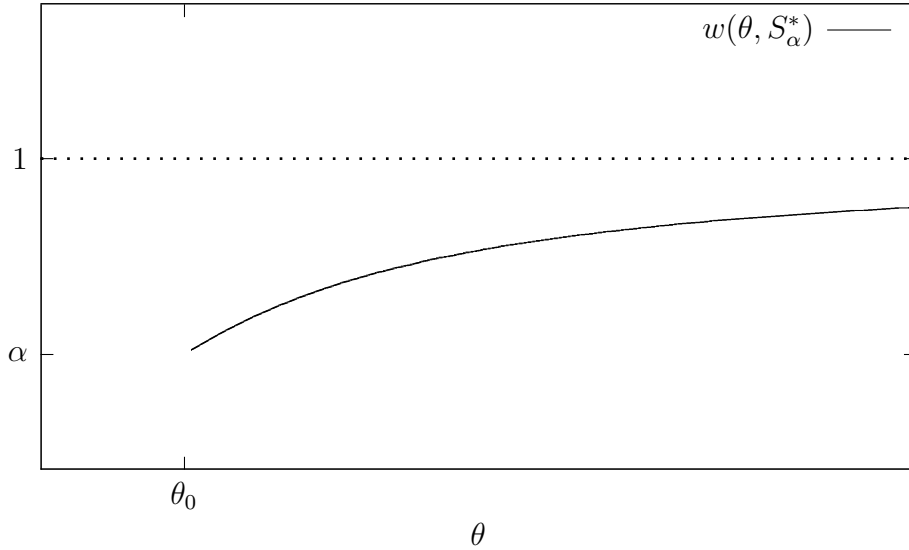
Значит, выполнено условие 1 Леммы Неймана-Пирсона, и в силу пункта 2 этой леммы

$$P_{\theta_1}(X \in S_\alpha) \leq P_{\theta_1}(X \in S_\alpha^*)$$

То есть S_α^* - наиболее мощный критерий уровня α .

Так как S_α^* не зависит от θ_1 , то S_α^* - РНМ-критерий для $H_0 : \theta = \theta_0$ против $H_1^+ : \theta > \theta_0$
Мощность критерия S_α^* для H_0 при альт. H_1^+

$$\begin{aligned} W(\theta, S_\alpha^*) &= P_\theta \left(\sum_{i=1}^n X_i > n\theta_0 + \sqrt{n}\sigma\xi_{1-\alpha} \right) = \\ &= P_\theta \left(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \theta) > \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + \xi_{1-\alpha} \right) = 1 - \Phi \left(\xi_{1-\alpha} + \frac{\sqrt{n}(\theta - \theta_0)}{\sigma} \right) \end{aligned}$$



О связи между доверительным оцениванием и проверкой гипотез

Опр. 8. Случайное подмножество $\Theta^* = \Theta^*(X, \alpha) \subseteq \Theta$ называется доверительным множеством уровня $1 - \alpha$, $0 < \alpha < 1$, если

$$P_\theta(\theta \in \Theta^*(X, \alpha)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

Теорема 2. 1. Пусть $\forall \theta_0 \in \Theta$ гипотеза $H_0 : \theta = \theta_0$ при альтернативе $H_1 : \theta \neq \theta_0$ имеет $S_\alpha(\theta_0)$ критерием уровня α . Пусть $\Theta^*(x, \alpha) = \{\theta : x \in \bar{S}_\alpha(\theta)\}$. тогда $\Theta^*(X, \alpha)$ - доверительное множество уровня $1 - \alpha$. (Если есть критерий, то можно по этому построить доверительное множество)

2. Если $\Theta^*(X, \alpha)$ - доверительное множество уровня $1 - \alpha$, то $\overline{S}_\alpha(\theta_0) = \{x : \theta_0 \notin \Theta(x, \alpha)\}$ есть область применения гипотезы H_0 (следовательно и критерий).

Замечание 3. Пункт 2 означает, что если θ_0 попало в доверительное множество, то H_0 надо применять.

Доказательство.

1. $P_\theta(\theta \in \Theta^*(X, \alpha)) = P_\theta(X \in \overline{S}_\alpha(\theta)) = 1 - \underbrace{P_\theta(X \in S_\alpha(\theta))}_{\leq \alpha} \geq 1 - \alpha \quad \forall \theta \in \Theta$
2. $P_{\theta_0}(X \in S_\alpha(\theta_0)) = 1 - P_{\theta_0}(X \in \overline{S}_\alpha(\theta_0)) = 1 - \underbrace{P_{\theta_0}(\theta_0 \in \Theta^*(X, \alpha))}_{\geq 1 - \alpha} \leq 1 - (1 - \alpha) = \alpha$

□

Пример. Пусть $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р. сл.в., $X_1 \sim N(0, \sigma^2)$, $\theta \in \mathbb{R}^1$. Построим критерий для $H_0 : \theta = \theta_0$ против $H_1 : \theta \neq \theta_0$. Уровень значимости пусть будет α , $0 < \alpha < 1$.

Построим доверительное множество для θ уровня $1 - \alpha$. Пусть $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$ - оптимальная оценка θ . Тогда $\frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} \sim N(0, 1)$,

$$P_\theta \left(\left| \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} \right| < \xi_{1-\alpha/2} \right) = 1 - \alpha$$

$$\Phi(\xi_{1-\alpha/2}) = 1 - \alpha/2$$

То есть $\Theta^*(X, \alpha) = \{\theta : \left| \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} \right| < \xi_{1-\alpha/2}\}$. В силу замечания к Теореме 2 $S_\alpha(\theta_0) = \{X : \left| \frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma} \right| \geq \xi_{1-\alpha}\}$ есть критическое множество для H_0 . Мощность

$$\begin{aligned} W(\theta) &= P_\theta(X \in S_\alpha(\theta_0)) = P_\theta \left(\left| \frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma} \right| \geq \xi_{1-\alpha/2} \right) = 1 - P_\theta \left(\left| \frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma} \right| < \xi_{1-\alpha/2} \right) = \\ &= 1 - P \left(-\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} < \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} < \xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} \right) = \\ &= 1 - \left[\Phi \left(\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} \right) - \Phi \left(-\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} \right) \right] = \\ &= \left[\Phi \left(\xi_{\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} \right) + \Phi \left(\xi_{\alpha/2} + \frac{\sqrt{n}(\theta - \theta_0)}{\sigma} \right) \right] \end{aligned}$$

***** TODO: вставить график стр 136*****

При $n \rightarrow \infty$ $W(\theta) \rightarrow 1 \quad \forall \theta \neq \theta_0$!. То есть $S_\alpha(\theta_0)$ состоятелен против любой фиксированной альтернативы.

2.1 Критерий Фишера (F -критерий) в Гауссовской линейной регрессии

Опр. 9. Если $\xi \sim N(0, 1)$, $\eta_k \sim \chi^2(k)$, ξ и η_k независимы, а константа $\mu \in \mathbb{R}^1$, то сл.в.

$$t_k(\mu) \stackrel{d}{=} \frac{\xi + \mu}{\sqrt{\frac{1}{k}\eta_k}} \sim S(k, \mu)$$

имеет нецентрального распределение Стьюдента с k степенями свободы и параметром нецентральности μ

Опр. 10. Если $\xi_i \sim N(a_i, 1)$, $i = 1, \dots, k$, и $\{\xi_1, \dots, \xi_k\}$ независимы, а $\Delta^2 = \sum_{j=1}^k a_j^2$, то сл. в.

$$\eta_k(\Delta) \stackrel{d}{=} \xi_1^2 + \dots + \xi_k^2 \sim \chi^2(k, \Delta^2)$$

имеет нецентрального распределение хи-квадрат с k степенями свободы и параметром нецентральности Δ^2

Опр. 11. Если $\eta_k \sim \chi^2(k, \Delta^2)$, $\nu_m \sim \chi^2(m)$, и η_k и ν_m независимы, то сл.в.

$$f_{k,m}(\Delta) \stackrel{d}{=} \frac{\frac{1}{k}\eta_k}{\frac{1}{m}\nu_m} \sim F(k, m, \Delta^2)$$

имеет нецентрального распределение Фишера с (k, m) степенями свободы и параметром нецентральности Δ^2

Лемма 1. 1. Распределение сл.в. $\eta_k \sim \chi^2(k, \Delta^2)$ зависит лишь от Δ , но не от a_1, \dots, a_k .
А именно

$$\eta_k \stackrel{d}{=} (z_1 + \Delta)^2 + z^2 + \dots + z^k, \text{ где } (z_1, \dots, z_k) - \text{н.о.р. } N(0, 1) \text{ сл.в.}$$

2. Если вектор $\xi \in \mathbb{R}^k$, $\xi \sim N(a, \Sigma)$, $\Sigma > 0$, то

$$\xi^T \Sigma^{-1} \xi \sim \chi^2(k, \Delta^2), \Delta^2 = a^T \Sigma^{-1} a$$

Доказательство. 1. По определению $\eta_k(\Delta) \stackrel{d}{=} \sum_{i=1}^k \xi_i^2$, где (ξ_1, \dots, ξ_k) - н.о.р. $N(a_i, 1)$ сл.в.

Пусть $\xi = (\xi_1, \dots, \xi_k)^T$, ортогональная матрица

$$C = \begin{pmatrix} \frac{a_1}{\Delta} & \cdots & \frac{a_k}{\Delta} \\ \vdots & \ddots & \vdots \end{pmatrix}, \nu = C\xi$$

Тогда $\eta_k \stackrel{d}{=} |\xi|^2 = |\nu|^2$, так как C - ортог. Но

$$\nu = C \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} + C \overset{\circ}{\xi} = \begin{pmatrix} \Delta \\ 0 \\ \vdots \\ 0 \end{pmatrix} + Z, \text{ где } \overset{\circ}{\xi} = \xi - E\xi, Z = C \overset{\circ}{\xi} \sim N(0, E_k)$$

Итак, $\eta_k \stackrel{d}{=} |\nu|^2 = (z_1 + \Delta)^2 + z_2^2 + \dots + z_k^2$

2. $\xi^T \Sigma^{-1} \xi = |\Sigma^{-1/2} \xi|^2$, причем $\Sigma^{-1/2} \xi \sim N(\Sigma^{-1/2} a, E_k)$. Отсюда $|\Sigma^{-1/2} \xi|^2 \sim \chi^2(k, \Delta^2)$ с $\Delta^2 = |\Sigma^{-1/2} a|^2 = a^T \Sigma^{-1} a$

□

Лемма 2. Случайная величина $t_k(\mu)$ обладает следующим свойством стохастической упорядоченности. при $\mu_2 > \mu_1$

$$P(t_k(\mu_2) > x) > P(t_k(\mu_1) > x) \text{ при всех } x \in \mathbb{R}^1 \quad (4)$$

Аналогично

$$P(\eta_k(\Delta_2) > x) > P(\eta_k(\Delta_1) > x), \Delta_2 > \Delta_1 \quad (5)$$

$$P(f_{k,m}(\Delta_2) > x) > P(f_{k,m}(\Delta_1) > x), \Delta_2 > \Delta_1 \quad (6)$$

Нецентральные распределения Пирсона и Фишера стохастически упорядочены по параметру нецентральности.

Доказательство. Докажем соотношение 4, 5 и 6 доказываются аналогично.

Заметим, что, если ξ и η - независимые случайные величины, и $E|\phi(\xi, \eta)| < \infty$, то

$$E\phi(\xi, \eta) = E \left\{ E\phi(\xi, \eta) |_{\xi=\eta} \right\} \quad (7)$$

В силу (7)

$$\begin{aligned} P(t_k(\mu_2) > x) &= P \left(\frac{\xi + \mu_2}{\sqrt{\frac{1}{k}\eta_k}} > x \right) = EI \left(\xi > x \sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) = \\ &= E \left\{ 1 - I \left(\xi \leq x \sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) \right\} = 1 - E \left\{ EI(\xi \leq y) \Big|_{y=x\sqrt{\frac{1}{k}\eta_k} - \mu_2} \right\} = \\ &= 1 - E\Phi \left(x \sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) > 1 - E\Phi \left(x \sqrt{\frac{1}{k}\eta_k} - \mu_1 \right) = P(t_k(\mu_1) > x) \\ \text{так как } E\Phi \left(x \sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) &< E\Phi \left(x \sqrt{\frac{1}{k}\eta_k} - \mu_1 \right) \text{ в силу возрастающей } \Phi(y) \end{aligned}$$

□

Обратимся к линейной гауссовской модели

$$X = Zc + \mathcal{E}$$

$X = (X_1, \dots, X_n)^T$ - наблюдения, Z - $n \times p$ матрица регрессоров $p < n$

$$\mathcal{E} \sim N(0, \sigma^2 E_n), \quad c = (c_1, \dots, c_p)^T$$

c и σ^2 неизвестны

Рассмотрим новый вектор $\beta = Ac$, A - $k \times p$ матрица, $rkA = k, k \leq p$.

Построим для β доверительное множество уровня $1 - \alpha$

Пусть \hat{c}_n - оценка наименьших квадратов (о.н.к.) для c , \hat{s}_n^2 - о.н.к. для σ^2 . Пусть $\hat{\beta}_n = A\hat{c}_n$.

$$\hat{c}_n \sim N(c, \sigma^2 (Z^T Z)^{-1}) \Rightarrow \underbrace{\hat{\beta}_n}_{\beta} \sim N(\underbrace{Ac}_{\beta}, \sigma^2 D), \text{ где } D = A(Z^T Z)^{-1} A^T$$

Заметим, что $D > 0$, так как для $\alpha \in \mathbb{R}^k, \alpha \neq 0$,

$$\alpha^T D \alpha = (A^t \alpha)^T (Z^T Z)^{-1} (A^T \alpha) > 0, \text{ т.к. } (Z^T Z)^{-1} > 0, A^T \alpha \neq 0 \text{ при } rk A = k, \alpha \neq 0$$

В силу пункта 2 леммы 1

$$\frac{1}{\sigma^2} (\hat{\beta}_n - \beta)^T D^{-1} (\hat{\beta}_n - \beta) \sim \chi^2(k)$$

так как $\frac{(n-p)\hat{s}_n^2}{\sigma^2} \sim \chi^2(n-p)$, $\hat{\beta}_n$ и \hat{s}_n^2 независимы, то

$$f_{k,n-p}(X, \beta) \stackrel{\text{def}}{=} \frac{\frac{1}{k} (\hat{\beta}_n - \beta)^T D^{-1} (\hat{\beta}_n - \beta) / \sigma^2}{\frac{1}{n-p} (n-p) \hat{s}_n^2 / \sigma^2} = \frac{(\hat{\beta}_n - \beta)^T D^{-1} (\hat{\beta}_n - \beta)}{k \hat{s}_n^2} \sim F(k, n-p)$$

Значит,

$$P_{\beta, \sigma^2} \left((\hat{\beta}_n - \beta)^T D^{-1} (\hat{\beta}_n - \beta) \leq k \hat{s}_n^2 f_{1-\alpha}(k, n-p) \right) = 1 - \alpha$$

$f_{1-\alpha}(k, n-p)$ - квантиль уровня $1 - \alpha$ $F(k, n-p)$.

Доверительное множество для β уровня $1 - \alpha$

$$\Theta^*(X, \alpha) = \left\{ \beta : (\hat{\beta}_n - \beta)^T D^{-1} (\hat{\beta}_n - \beta) < k \hat{s}_n^2 f_{1-\alpha}(k, n-p) \right\} =$$

$$= \{ \beta : f_{k,n-p}(X, \beta) < f_{1-\alpha}(k, n-p) \} - \text{доверительный эллипсоид}$$

Рассмотрим проверку гипотезы $H_0 : \beta = \beta_0$ против $H_1 : \beta \neq \beta_0$. H_0 называют линейной гипотезой, так как $\beta = A c$ получается линейным преобразованием c . В силу замечания 3 H_0 надо принимать, если $\beta_0 \in \Theta^*(X, \alpha)$, то есть область принятия H_0 :

$$\bar{S}_\alpha(\beta_0) = \{ x : f_{k,n-p}(x, \beta_0) \leq f_{1-\alpha}(k, n-p) \}$$

То есть критическое множество (критерий уровня α):

$$S_\alpha(\beta_0) = \{ x : f_{k,n-p}(x, \beta_0) > f_{1-\alpha}(k, n-p) \} \quad (8)$$

Критерий 8 называют **критерием Фишера** или **F-критерием**. $f_{k,n-p}(X, \beta_0)$ - статистика F-критерия.

Рассмотрим поведение F-критерия при альтернативе H_1 .

При H_1 в силу пункта 2 Леммы 1

$$f_{k,n-p}(X, \beta_0) = \frac{\overbrace{\frac{1}{k} (\hat{\beta}_n - \beta_0)^T D^{-1} (\hat{\beta}_n - \beta_0) / \sigma^2}^{\chi^2(k, \Delta^2)}}{\underbrace{\frac{1}{n-p} (n-p) \hat{s}_n^2 / \sigma^2}_{\chi^2(n-p)}} \sim F(k, n-p, \Delta^2)$$

Параметр нецентральности

$$\Delta^2 = \frac{1}{\sigma^2} (\beta - \beta_0)^T D^{-1} (\beta - \beta_0) \quad (9)$$

Мощность F-критерия

$$W(\beta, S_\alpha(\beta_0)) = P_{\beta, \sigma^2}(f_{k,n-p}(X, \beta_0) > f_{1-\alpha}(k, n-p)) = 1 - F_{k,n-p}(f_{1-\alpha}(k, n-p), \Delta^2)$$

Свойства мощности

1. Так как $\Delta = \Delta(\beta) = \Delta(\beta_0) > 0$ при $\beta \neq \beta_0$, то в силу соотношения 6

$$P_{\beta, \sigma^2}(f_{k, n-p}(X, \beta_0) > f_{1-\alpha}(k, n-p)) > P_{\beta_0, \sigma}(f_{k, n-p}(X, \beta_0) > f_{1-\alpha}(k, n-p)) = \alpha$$

То есть при $\beta \neq \beta_0$ $P(H_1|H_1) > P(H_1|H_0)$. То есть F-критерий несмещенный!

2. Мощность $W(\beta, S_\alpha(\beta_0))$ строго монотонна по Δ из соотношения 9

Пример (Определение порядка регрессии). $c_n^T = (\underbrace{c_{(1)n}^T}_{m\text{-вектор}}, \underbrace{c_{(2)n}^T}_{p-m\text{-вектор}})$, $1 \leq m \leq p$

$$H_0: c_{(2)} = 0 \quad (\text{порядок не больше } m)$$

$$H_1: c_{(2)} \neq 0$$

Рассмотрим матрицу

$$A = \begin{pmatrix} 0 & \cdots & 1 \\ & \ddots & \vdots & \ddots \\ & & 0 & \vdots & 1 \end{pmatrix} \Rightarrow Ac = c_{(2)} \Rightarrow H_0 \Leftrightarrow Ac = 0$$

$m \qquad p-m$

Пусть $\hat{c}_n^T = (\underbrace{\hat{c}_{(1)n}^T}_{m\text{-в-р}}, \underbrace{\hat{c}_{(2)n}^T}_{p-m\text{-в-р}})$. Тогда $\hat{\beta}_n = A\hat{c}_n = \hat{c}_{(2)n}$.

$$(Z^T Z)^{-1} = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right) \rightarrow D = A(Z^T Z)^{-1}A^T = B_{22} \Rightarrow$$

$$\Rightarrow f_{p-m, n-p}(X, 0) = \frac{\hat{c}_{(2)n}^T B_{22}^{-1} \hat{c}_{(2)n}}{(p-m)\hat{s}_n^2} \underset{H_0}{\sim} F(p-m, n-p)$$

H_0 отвергается, если $f_{p-m, n-p}(X, 0) > f_{1-\alpha}(p-m, n-m)$, то есть

$$S_\alpha(0) = \{x : \frac{\hat{c}_{(2)n}^T B_{22}^{-1} \hat{c}_{(2)n}}{(p-m)\hat{s}_n^2} > f_{1-\alpha}(p-m, n-m)\} \quad (10)$$

$$f_{p-m, n-p}(X, 0) \underset{H_1}{\sim} F(p-m, n-p, \Delta^2), \text{ где } \Delta^2 = \frac{c_{(2)}^T B_{22}^{-1} c_{(2)}}{\sigma^2} \quad (11)$$

Критерий 10 - несмещенный, то есть $P(H_1|H_1) > P(H_1|H_0) = \alpha$. Его мощность

$$W(c_{(2)}, S_\alpha(0)) = P_{c_{(2)}, \sigma^2}(f_{p-m, n-p}(X, 0) > f_{1-\alpha}(p-m, n-p)) = 1 - F_{p-m, n-p}(f_{1-\alpha}(p-m, n-p), \Delta^2)$$

строго возрастает по Δ^2 . Параметр нецентральности Δ^2 определен в 11.

Пример (Проверка однородности двух выборок). $X = (X_1, \dots, X_m)$, $Y = (Y_1, \dots, Y_n)$ - независимые гауссовские выборки. То есть $\{X_i\}$, $\{Y_i\}$ - н.о.р., $X_1 \sim N(a, \sigma^2)$, $Y_1 \sim N(b, \sigma^2)$. Совокупность $\{X_i\}$ и $\{Y_j\}$ независимы, $m+n > 2$.

Дисперсии DX_1 , DY_1 одинаковы ($= \sigma^2$), неизвестны, средние a и b неизвестны.

$$H_0: a = b \quad (\text{гипотеза однородности})$$

$$H_1: a \neq b$$

Замечание. При $DX_1 \neq DX_2$ эта задача называется **проблемой Беренса-Фишера**.

$$\begin{cases} X_i = a + \varepsilon_i, & i = 1, \dots, m, \quad \varepsilon_i = X_i - a \\ Y_j = b + \hat{\varepsilon}_j, & j = 1, \dots, n, \quad \hat{\varepsilon}_j = Y_j - b \end{cases} \Rightarrow \varepsilon_1, \dots, \varepsilon_m, \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n - \text{н.о.р. } N(0, \sigma^2) \text{ с.л.в.}$$

$$\begin{aligned} \hat{X} &\stackrel{\text{def}}{=} (X_1, \dots, X_m, Y_1, \dots, Y_n)^T \\ c &= (a, b)^T \\ \mathcal{E}^T &= (\varepsilon_1, \dots, \varepsilon_m, \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^T \end{aligned} \quad Z = \begin{pmatrix} m \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix} & 0 \\ 0 & n \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix} \end{pmatrix} \Rightarrow \hat{X} = Zc + \mathcal{E} \quad (12)$$

гаусс. лин. регрессия

Положим $A = (1, -1)$. Тогда $Ac = a - b = \beta$.

$$H_0 : Ac = a - b = \beta = 0 \quad (= \beta_0)$$

$$H_1 : Ac = a - b \neq 0 \quad (\beta \neq 0)$$

О.н.к. для вектора c - решение задачи

$$\sum_{i=1}^m (X_i - a)^2 + \sum_{j=1}^n (Y_j - b)^2 \rightarrow \min_{a,b} \Leftrightarrow \begin{cases} -2 \sum_i (X_i - a) = 0 \\ -2 \sum_j (Y_j - a) = 0 \end{cases}$$

Решением системы является $\hat{a}_n = \bar{X}$, $\hat{b}_m = \bar{Y}$ - оптимальные оценки a и b , $\hat{c}_n = (\bar{X}, \bar{Y})^T$ - оптимальная оценка для c . Оптимальная оценка для σ^2 :

$$\hat{S}_{m+n}^2 = \frac{1}{m+n-2} \left[\sum_i (X_i - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2 \right]$$

Тогда

$$\hat{\beta}_n = A\hat{c}_n = \bar{X} - \bar{Y}$$

$$Z^T Z = \begin{pmatrix} \overbrace{1 \dots 1}^m & 0 \\ 0 & \underbrace{1 \dots 1}_n \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} m & 0 \\ 0 & n \end{pmatrix} \Rightarrow f_{1,m+n-2}(X, 0) = \frac{(\bar{X} - \bar{Y})^2}{\left(\frac{1}{n} + \frac{1}{m}\right) \hat{S}_{m+n}^2}$$

$$D = A(Z^T Z)^{-1} A^T = (1 \quad -1) \begin{pmatrix} \frac{1}{m} & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n} + \frac{1}{m}$$

F -критерий для H_0 имеет вид

$$S_\alpha(0) = \{x \in \mathbb{R}^{m+n} : f_{1,m+n-2}(x, 0) > f_{1-\alpha}(1, m+n-2)\}$$

$$f_{1,m+n-2}(X, 0) \underset{H_0}{\sim} F(1, m+n-2)$$

$$f_{1,m+n-2}(X, 0) \underset{H_1}{\sim} F(1, m+n-2, \Delta^2),$$

$$\text{где параметр нецентральности } \Delta^2 = \Delta^2_{a-b}(\beta) = \frac{(a-b)^2}{\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)}$$

1. Если $|a - b|$ возрастает, то мощность F -теста возрастает
2. Если $\sigma \rightarrow 0$ или $\frac{1}{n} + \frac{1}{m} \rightarrow 0$, то мощность возрастает

2.2 Критерий согласия Хи-квадрат Пирсона. Проверка простой гипотезы в схеме Бернулли.

Пусть проводятся n независимых испытаний, и в каждом испытании возможны $m \geq 2$ исходов A_1, \dots, A_m таких, что $A_i A_j = \emptyset$, $i \neq j$, $\sum A_i = \Omega$, тогда $P(A_j) = p_j > 0$, $\sum_{j=1}^m p_j = 1$. Пусть $\nu = (\nu_1, \dots, \nu_m)^T$, а ν_j - число появления A_j в n опытах, тогда $\sum_{j=1}^m \nu_j = n$. По вектору наблюдений ν необходимо проверить следующую гипотезу:

$$H_0: p_j = p_j^\circ, j = 1, \dots, m$$

$$H_1: p_j \neq p_j^\circ \forall j$$

Замечание. H_0 - простая гипотеза, т.к. полностью определяет распределение вектора ν .

$$P(\nu_1 = k_1, \dots, \nu_m = k_m) \stackrel{H_0}{=} \frac{n!}{k_1! \dots k_m!} (p_1^\circ)^{k_1} \dots (p_m^\circ)^{k_m}$$

Это полиномиальное распределение $\Pi(n, p_1^\circ, \dots, p_m^\circ)$. Статистика Хи-квадрат Пирсона:

$$\chi_n^2 \stackrel{\text{def}}{=} \sum_{j=1}^m \frac{(\nu_j - np_j^\circ)^2}{np_j^\circ}$$

Поведение при альтернативе: Очевидно

$$\chi_n^2 = n \sum_{j=1}^m \frac{(\nu_j/n - p_j^\circ)^2}{p_j^\circ}$$

В силу теоремы Бернулли $\frac{\nu_j}{n} \xrightarrow{P} p_j$. Поэтому

$$\sum_{j=1}^m \frac{(\nu_j/n - p_j^\circ)^2}{p_j^\circ} \xrightarrow[\text{т. о наслед. сход.}]{P} \sum_{j=1}^m \frac{(p_j - p_j^\circ)^2}{p_j^\circ} \underset{H_1}{>} 0$$

Значит,

$$\chi_n^2 \xrightarrow[H_1]{P} \infty, n \rightarrow \infty$$

Поэтому большие значения χ_n^2 часто свидетельствуют о том, что стоит отвергнуть H_0 . Но насколько "большие" значения?

Теорема Пирсона

$$\chi_n^2 \xrightarrow[H_0]{d} \chi^2(m-1), n \rightarrow \infty$$

Правило: Если $\chi_n^2 \leq \chi_{1-\alpha}(m-1)$, то принимаем H_0 , иначе принимаем H_1 .

Замечание. Тогда

$$P(H_1|H_0) = P(\chi_n^2 > \chi_{1-\alpha}(m-1)|H_0) \rightarrow \alpha$$

$$P(H_0|H_1) = P(\chi_n^2 \leq \chi_{1-\alpha}(m-1)|H_1) \rightarrow 0$$

То есть

$$\begin{cases} P(H_0|H_0) \rightarrow 1 - \alpha \\ P(H_1|H_1) \rightarrow 1 \end{cases}$$

Вероятность принять правильную гипотезу близка к единице!

Доказательство. Покажем, что вектор $\nu = (\nu_1, \dots, \nu_m)^T$ асимптотически нормален, то есть

$$\sqrt{n}(\nu/n - p) \xrightarrow{d} N(0, P - pp^T), \text{ где } P \stackrel{\text{def}}{=} \begin{pmatrix} p_1^\circ & & 0 \\ & \ddots & \\ 0 & & p_m^\circ \end{pmatrix} \quad (13)$$

Введем вектора X_1, \dots, X_n , где $X_i = (0, \dots, 0, \frac{1}{j}, 0, \dots, 0)^T$, если в i -ом испытании произошло A_j . Тогда $\nu = \sum_{i=1}^n X_i$

$$\sqrt{n}(\nu/n - p) = \sqrt{n} \sum_{i=1}^n (X_i - p) \quad (14)$$

Здесь $\{X_i\}$ - н.о.р., $EX_1 = p$, $\text{cov}(X_1, X_1) = E(X_1 - p)(X_1 - p)^T = EX_1X_1^T - pp^T = P - pp^T$. Поэтому соотношение (13) следует из соотношения (14) и ЦПТ.

Матрица $P - pp^T$ вырождена, так как сумма ее столбцов равна нулю: если $e = (1, \dots, 1)^T$, то $(P - pp^T)e = p - p(p^Te) = p - p = 0$

Пусть

$$P^{-1/2} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{1}{\sqrt{p_1^\circ}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{p_m^\circ}} \end{pmatrix}, \quad \xi_n \stackrel{\text{def}}{=} \sqrt{n}P^{-1/2}(\nu/n - p)$$

В силу теоремы о наследовании слабой сходимости и соотношения (13)

$$\xi_n \xrightarrow{d} N(0, P^{-1/2}(P - pp^T)(P^{-1/2})^T) = N(0, E_m - zz^T), \text{ где } z = (\sqrt{p_1^\circ}, \dots, \sqrt{p_m^\circ})^T \quad (15)$$

Пусть ортогональная матрица $U = \begin{pmatrix} \sqrt{p_1^\circ} & \dots & \sqrt{p_m^\circ} \\ \dots & \dots & \dots \end{pmatrix}$. Тогда

$$\begin{aligned} U(E_m - zz^T)U^T &= E_m - (Uz)(Uz)^T = \\ &= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} (1 \ 0 \ \dots \ 0) = \begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{pmatrix} = \tilde{E}_1 \end{aligned}$$

В силу (15) и теоремы о слабой сходимости

$$U\xi_n \xrightarrow{d} N(0, \tilde{E}_1) = (0, \eta_2, \dots, \eta_m)^T \quad (16)$$

где $\{\eta_2, \dots, \eta_m\}$ - независимые $N(0, 1)$ сл.в. Из (16) и теоремы о наследовании слабой сходимости следует:

$$|U\xi_n|^2 \xrightarrow{d} \eta_2^2 + \dots + \eta_m^2 \sim \chi^2(m-1) \quad (17)$$

Осталось заметить, что

$$|U\xi_n|^2 = |\xi_n|^2 = \sum_{j=1}^m \left[\frac{1}{\sqrt{p_j^\circ}} \sqrt{n}(\nu_j/n - p_j^\circ) \right]^2 = \sum_{j=1}^m \frac{(\nu_j - np_j^\circ)^2}{np_j^\circ} = \chi_n^2$$

Из этого равенства и соотношения (17) следует теорема Пирсона. \square

Пример (Проверка простой гипотезы о виде функции распределения). $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р., $X_1 \sim F(x)$.

$H_0: F(x) = F_0(x)$, (F_0 известна)

$H_1: F(x) \neq F_1(x)$, $F_1(x) \neq F_0(x)$

Разобъем носитель X_1 на непересекающиеся отрезки $\Delta_1, \dots, \Delta_m$, $m \geq 2$ так, что $X_1 \in \Delta_1 \cup \Delta_2 \cup \dots \cup \Delta_m$

$$p_j^\circ \stackrel{\text{def}}{=} P(X_1 \in \Delta_j | H_0) = \int_{\Delta_j} dF_0(x) > 0 \quad \forall j$$

Тогда $\sum_{j=1}^m p_j^\circ = 1$. С каждой величиной X_i свяжем испытание с исходами A_1, \dots, A_m , причем A_j происходит тогда и только тогда, когда $X_i \in \Delta_j$. При H_0 $P(A_j) = p_j^\circ$. Тогда наблюдения X_1, \dots, X_n порождают полиномиальную схему независимых испытаний. Пусть ν_j - число исхода A_j в этих испытаниях, то есть число наблюдений среди X_1, \dots, X_n , попавших в Δ_j . В силу теоремы Пирсона:

$$\chi_n^2 \stackrel{\text{def}}{=} \sum_{j=1}^m \frac{(\nu_j - np_j^\circ)^2}{np_j^\circ} \xrightarrow{H_0} \chi^2(m-1)$$

Правило: H_0 будем отвергать, если $\chi_n^2 > \chi_{1-\alpha}^2(m-1)$. (α задано) Тогда $P(H_1 | H_0) \rightarrow \alpha$, $n \rightarrow \infty$.

$$p_j \stackrel{\text{def}}{=} P(X_1 \in \Delta_j | H_1) = \int_{\Delta_j} dF_1(x)$$

Если верна H_1 и хоть при одном j $p_j \neq p_j^\circ$, то $P(H_0 | H_1) = P(\chi_n^2 < \chi_{1-\alpha}^2(m-1) | H_1) \rightarrow 0$

Замечание. Если $F_0 \neq F_1$, но $p_j = p_j^\circ \quad \forall j$, то $P(H_0 | H_1) = P(H_0 | H_0) \rightarrow 1 - \alpha \neq 0$. Например **** TODO вставить график стр 152 ****

Здесь $P(X_1 \in \Delta_1 | H_0) = F_0(0) = P(X_1 \in \Delta_1 | H_1) = F_1(0)$. Значит, и $P(X_1 \in \Delta_2 | H_0) = 1 - F_0(0) = P(X_1 \in \Delta_2 | H_1) = 1 - F_1(0)$.

Проверка сложной гипотезы в схеме исп. Бернулли

Пусть проводится n независимых испытаний, исходы A_1, \dots, A_m , $\nu = (\nu_1, \dots, \nu_m)^T$ - вектор наблюдений. Пусть $H_0: P(A_j) = p_j(\theta)$, $\theta \in \Theta \in \mathbb{R}^k$, $k < m-1$.

Условия регулярности

1. $\sum_{j=1}^m p_j(\theta) = 1$, $\theta \in \Theta$
2. $p_j(\theta) \geq c > 0 \quad \forall j = 1, \dots, m$ и $\exists \frac{\partial p_j(\theta)}{\partial \theta_l}, \frac{\partial^2 p_j(\theta)}{\partial \theta_l \partial \theta_r}$

$$3. \underbrace{\text{rank}\left(\frac{\partial p_j(\theta)}{\partial \theta_l}\right)}_{m \times k} = k, \forall \theta \in \Theta$$

В качестве оценки θ при H_0 будем использовать мультиномиальные оценки максимального правдоподобия:

$$P(\nu_1 = k_1, \dots, \nu_m = k_m) = \frac{n!}{k_1! \dots k_m!} p_1^{k_1}(\theta) \dots p_m^{k_m}(\theta)$$

логарифмического правдоподобия:

$$L_n(\nu, \theta) = \ln \left(\frac{n!}{k_1! \dots k_m!} \right) + \sum_{j=1}^m \nu_j \ln p_j(\theta)$$

оценки максимального правдоподобия (мультиномиальные):

$$L_n(\nu, \theta) \rightarrow \max_{\theta \in \Theta}$$

Теорема Фишера

Пусть выполнены условия регулярности, $\hat{\theta}_n$ - мульт. о.м.п. Тогда

$$\hat{\chi}_n^2 = \sum_{j=1}^m \frac{(\nu_j - np_j(\hat{\theta}_n))^2}{np_j(\hat{\theta}_n)} \xrightarrow[H_0]{d} \chi(m - k - 1)$$

Правило: Если $\hat{\chi}_n^2 \leq X_{1-\alpha}(m-k-1)$, то принимаем H_0 , иначе принимаем H_1 . Тогда $P(\overline{H_0} | H_0) \rightarrow \alpha$

Пример (Проверка независимости признаков). Пусть объект классифицирован по двум A и B , $A = \{A_1, \dots, A_s\}$, $B = \{B_1, \dots, B_r\}$, $s > 1$, $r > 1$. Проводится n опытов, и пусть ν_{ij} - число объектов, имеющих признаки $A_i B_j$.

Пусть $p_{ij} = P(A_i B_j)$. Гипотеза независимости H_0 : $p_{ij} = p_{i\bullet} p_{\bullet j}$ для положительных $p_{i\bullet}$ и $p_{\bullet j}$ таких, что $\sum_{i=1}^s p_{i\bullet} = 1$, $\sum_{j=1}^r p_{\bullet j} = 1$.

При H_0 логарифмическое правдоподобие

$$L_n(\nu, p_{i\bullet}, p_{\bullet j}) = \ln \frac{n!}{\prod_{i,j} \nu_{ij}} + \sum_{i=1}^s \sum_{j=1}^r \nu_{ij} \ln(p_{i\bullet} p_{\bullet j})$$

Максимизируя эту функцию по $p_{i\bullet}$, $p_{\bullet j}$ при условиях, что $\sum_{i=1}^s p_{i\bullet} = 1$, $\sum_{j=1}^r p_{\bullet j} = 1$, находим оценки

$$\hat{p}_{i\bullet} = \frac{\nu_{i\bullet}}{n}, \hat{p}_{\bullet j} = \frac{\nu_{\bullet j}}{n}, \text{ где } \nu_{i\bullet} = \sum_j \nu_{ij}, \nu_{\bullet j} = \sum_i \nu_{ij}$$

Статистика Хи-квадрат имеет вид

$$\hat{\chi}_n^2 = \sum_{i=1}^s \sum_{j=1}^r \frac{(\nu_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}}$$

$$\hat{\chi}_n^2 \xrightarrow[H_0]{d} \chi((s-1)(r-1))$$

так как $m - k - 1 = sr - (s + r - 2) - 1 = (s-1)(r-1)$.

Правило: Если $\hat{\chi}_n^2 > \chi_{1-\alpha}^2((s-1)(r-1))$, то отвергаем H_0 . Асимптотический уровень теста есть α

Пример (W.H.Gilby. Biometrika, 8,94). 1725 школьников классифицировали в соответствии с их качеством одежды и в соответствии с умственными способностями. Использовали следующие градации:

A — умственно отсталый

B — медлительный и тупой

C — тупой

D — медлительный, но умный

E — достаточно умный

F — способный

G — очень способный

H_0 : признаки независимы

	Способности						
Как одевается	А и В	С	Д	Е	Ф	Г	Сумма
Очень хорошо	33	48	113	209	194	39	636
Хорошо	41	100	202	255	138	15	751
Сносно	39	58	70	61	33	4	256
Очень плохо	17	13	22	10	10	1	73
Сумма	130	219	407	535	375	59	1725

Здесь $\chi_n^2 = 174.92 > \chi_{0.999}(15) = 37.697$.

Здесь $15 = (s-1)(r-1) = (4-3)(6-1) \Rightarrow \text{Отвергаем } H_0$

3 Введение в робастное оценивание

Схема засорений Мартина-Йохаи имеет вид:

$$y_t = u_t + z_t^\gamma \xi_t, \quad t = 1, \dots, n$$

Здесь $\{u_t\}$ - "полезный сигнал" (временной ряд);

$\{z_t^\gamma\}$ - н.о.р. сл.в., $z_1^\gamma \sim \text{Bin}(1, \gamma)$ с $0 \leq \gamma \leq 1$ (γ - уровень засорения);

$\{\xi_t\}$ - н.о.р. сл.в. - грубые выбросы, ξ_1 - имеет распределение $\mu_\xi \in M_\xi$; Распределение μ_ξ неизвестно, а множество M_ξ известно;

Последовательность $\{u_t\}, \{z_t^\gamma\}, \{\xi_t\}$ независимы между собой.

Пусть y_1, \dots, y_n - наблюдения, и распределение вектора $Y_n = y_1, \dots, y_n$ висит от неизвестного параметра β . Пусть $\hat{\beta}_n$ - некоторая оценка β

Основное предположение

При любом $0 \leq \gamma \leq 1$ существует предел

$$\hat{\beta}_n \xrightarrow{P} \theta_\gamma, \quad n \rightarrow \infty; \quad \theta_0 = \beta$$

Опр. 1. Если существует предел

$$IF(\theta_\gamma, \mu_\xi) \stackrel{\text{def}}{=} \lim_{\gamma \rightarrow +0} \frac{\theta_\gamma - \theta_0}{\gamma}$$

то $IF(\theta_\gamma, \mu_\xi)$ называется **функционалом влияния оценки** $\hat{\beta}_n$

Если функционал влияния существует, то

$$\theta_\gamma = \theta_0 + IF(\theta_\gamma, \mu_\xi)\gamma + \bar{o}(\gamma), \quad \gamma \rightarrow +0$$

$IF(\theta_\gamma, \mu_\xi)$ характеризует главный линейный по γ член в разложении по γ асимптотического смещения $\theta_\gamma - \theta_0 = \theta_\gamma - \beta$

Опр. 2. Величина $GES(\theta_\gamma, M_\xi) \stackrel{\text{def}}{=} \sup_{\mu_\xi \in M_\xi} |IF(\theta_\gamma, \mu_\xi)|$ называется чувствительностью оценки $\hat{\beta}_n$ к засорениям (выбросам).

Если $GES(\theta_\gamma, M_\xi) < \infty$, то главный член по γ асимптотического смещения $IF(\theta_\gamma, \mu_\xi)\gamma$ равномерно по μ_ξ тах при таких γ

Опр. 3. Если $GES(\theta_\gamma, M_\xi) < \infty$, то оценка $\hat{\beta}_n$ называется **робастной по смещению**, или **B-робастной**.

Пример (Выборочное среднее).

$$\begin{cases} u_t = a + \xi_t \\ y_t = u_t + z_t' \xi_t, \quad t = 1, \dots, n \end{cases}$$

$\{\xi_t\}$ - н.о.р. сл.в., $E\xi_1 = 0$ (тогда $Eu_t = a$), $E|\xi_1| < \infty$

Возьмем оценкой a эмпирическое среднее $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$. Тогда $\bar{y} \xrightarrow{P} E(u_1 + z_1' \xi_1) = a + \gamma E\xi_1 = \theta_\gamma^{LS}$. Функция θ_γ^{LS} определена при всех γ ,

$$\frac{d\theta_\gamma^{LS}}{d\gamma} = E\xi_1 = IF(\theta_\gamma^{LS}, \mu_\xi)$$

Если M_1 - класс распределений с конечным первым моментом, то

$$GES(\theta_\gamma^{LS}, M_1) = \sup_{\mu_\xi \in M_1} |E\xi_1| = \infty$$

Оценка \bar{y} не B-робастна на классе M_1 !

Пример (Выборочная медиана). Пусть

$$u_t = a + \varepsilon_t, \quad t = 1, \dots, n, \quad \text{где } \{\varepsilon_t\} \text{ - н.о.р. сл.в.} \quad (1)$$

$\varepsilon_1 \sim G(x)$, функция распределения $G(x)$ неизвестна, $G(0) = 1/2$. Тогда функция распределения u_1 есть $F(x) = G(x - a)$, то есть $F(a) = 1/2$. Таким образом, медиана $G(x)$ это 0, медиана $F(x)$ - a .

Если ε_1 имеет симметричное относительно 0 распределение (то есть $\varepsilon_1 \stackrel{d}{=} -\varepsilon_1$, что для непрерывной $G(x)$ равносильно условию $G(x) + G(-x) = 1 \quad \forall x$), то автоматически $G(0) = 1/2$. Если вдобавок $E|\varepsilon_1| < \infty$, то $E\varepsilon_1 = 0, Eu_1 = a$

Итак, при сформулированных условия оценку медианы можно использовать как оценку математического ожидания.

Пусть $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(n)}$ будет вариационный ряд наблюдений u_1, \dots, u_n .

Опр. 4. Величина

$$\hat{m}_n = \begin{cases} u_{(k+1)} & n = 2k - 1 \\ \frac{u_{(k+1)} + u_{(k)}}{2} & n = 2k \end{cases}$$

называется **выборочной медианой** наблюдений u_1, \dots, u_n .

Мы знаем, что если $G(x)$ дифф. в нуле, и $g(0) = G'(0) > 0$, то для выборочной медианы справедлива асимптотическая нормальность:

$$\sqrt{n}(\hat{m}_n - a) \xrightarrow{d} N(0, \frac{1}{4g^2(0)}), \quad n \rightarrow \infty$$

Если в (1) $\{\varepsilon_t\}$ - н.о.р., $E\varepsilon_1 = 0$, $0 < E\varepsilon_1^2 = \sigma^2 < \infty$, то $\sqrt{n}(\hat{m}_n - a) \xrightarrow{d} N(0, \sigma^2)$. Значит асимптотическая оптимальная эффективность выборочной медианы относительно $\{\varepsilon_t\}$ равна

$$e_{\hat{m}_n, \bar{X}} = 4g^2(0)\sigma^2$$

Изучим B -робастность выборочной медианы. Пусть

$$\begin{cases} u_t = a + \varepsilon_t \\ y_t = u_t + z_t^\gamma \xi_t, \quad t = 1, \dots, n \end{cases} \quad \hat{m}_n^y = \begin{cases} y_{(k+1)}, & n = 2k - 1 \\ \frac{y_{(k)} + y_{(k+1)}}{2}, & n = 2k \end{cases}$$

Теорема 1.

Пусть $\exists g(x) = G'(x)$, $g(x)$ непрерывна и ограничена, $g(0) > 0$, $G(0) = 1/2$. Тогда:

1. $\hat{m}_n^y \xrightarrow{P} \theta_\gamma^m, \quad \theta_0 = a$
2. Существует функционал влияния выборочной медианы

$$IF(\theta_\gamma^m, \mu_\xi) = \frac{1 - 2EG(-\xi_1)}{2g(0)}$$

3. Чувствительность выборочной медианы на классе всех возможных распределений M_ξ

$$GES(\theta_\gamma^m, \mu_\xi) = \sum_{\mu_\xi \in M_\xi} |IF(\theta_\gamma^m, \mu_\xi)| = \frac{1}{2g(0)} < \infty$$

то есть выборочная медиана B -робастна.

Доказательство. стр 164

□