Московский государственный университет
имени М. В. Ломоносова

Механико-математический факультет

Конспект курса
Математическая статистика и приложения

Весенний семестр 2021, 3-ий курс, эконом. поток.

Выполнил:
студент 3 курса 332 группы
Шерстобитов Андрей Сергеевич

Лектор:
Кандидат физико-математических наук,
доцент кафедры теории вероятностей
Болдин Михаил Васильевич

Москва
2021

Аннотация

В данном документе представлен конспект курса по предмету "Математическая статистика и приложения"за 6 семестр.

Авторы данного конспекта выражают благодарность лектору, кандидату физико-математических наук и доценту Болдину Михаилу Васильевичу за прочитанный годовой курс по предмету "Математическая статистика и приложения".

Также авторы выражают благодарность следующим людям за помощь в обнаружении опечаток и ошибок:

1. Османкин Евгений (333 группа)

2. Акимов Иван (332 группа)

3. Галкина Александра (333 группа)

4. Сакаев Рамиль (332 группа)

5. Ладыкова Екатерина (333 группа)

При обнаружении ошибок, опечаток, неточностей и пр. авторы просят обращаться к ним по почте andrei.sherstobtiov@math.msu.ru

Содержание

1 Асимптотические оптимальные оценки

1.1 Определения слабой сходимости (по распределению), сходимости по вероятности и п.н., теоремы о наследовании сходимости. Лемма Слуцкого.

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

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

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

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

или, эквивалентно, $\mathrm{E}g(\xi_n) \to \mathrm{E}g(\xi)$.

Теорема 1

Следующие условия эквивалентны:

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

Если выполненое 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 $\xi_n,\ \xi \in \mathbb{R}^K,\ H : \mathbb{R}^K \to \mathbb{R}^1$ непрерывная. Тогда:

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

Теорема: Лемма Слуцкого

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

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

Доказательство. Достаточно показать, что вектор

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

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:

$$\left|\mathrm{E}e^{it\xi_n + is\eta_n} - \mathrm{E}e^{it\xi + isa}\right| \leq \left|\mathrm{E}e^{it\xi_n + is\eta_n} - \mathrm{E}e^{it\xi_n + isa}\right| + \left|\mathrm{E}e^{it\xi_n + isa} - \mathrm{E}e^{it\xi + isa}\right| = \alpha_n + \beta_n$$

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

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

$$\beta_n = \left|\mathrm{E}e^{isa}(e^{it\xi_n} - e^{it\xi})\right| = \left|e^{isa}\mathrm{E}(e^{it\xi_n} - e^{it\xi})\right| = \left|\mathrm{E}(e^{it\xi_n} - e^{it\xi})\right| \to 0$$

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

1.2 Состоятельные и асимптотически нормальные оценки. Теорема Бахадура и асимптотически эффективные оценки.

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

Опр. 3. *Если $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma(\theta)) \; \forall\theta \in \Theta$ и $0 < \Sigma(\theta) < \infty$, то $\widehat{\theta}_n$ называется* ***асимптотически нормальной оценкой****.*

Опр. 4. *Если $\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta \; \forall\theta \in \Theta$, то $\widehat{\theta}_n$ называется* ***состоятельной оценкой****.*

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

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

Если $\widehat{\theta}_n$ - асимптотическая нормальная оценка θ (так как θ и $\widehat{\theta}_n$ скаляры: $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)), \; 0 < \sigma^2 < \infty, \; \forall\theta \in \Theta$), то:

1. $\widehat{\theta}_n$ - состоятельная оценка θ, так как $\widehat{\theta}_n - \theta = n^{-1/2}\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathrm{P}} 0$ в силу п. (2) леммы Слуцкого.
2. Скорость сходимости $\widehat{\theta}_n$ к θ есть $O(\sqrt{n})$

3. При больших n со сл. в. $\sqrt{n}(\widehat{\theta}_n - \theta)$ можно обращаться (с осторожностью!) как с Гауссовской величиной.

   Например, пусть дисперсия предельного Гауссовского закона $\sigma^2(\theta)$ будет непрерывной ф-цией θ. Тогда
$$\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma(\widehat{\theta}_n)} = \underbrace{\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma(\theta)}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sigma(\theta)}{\sigma(\widehat{\theta}_n)}}_{\xrightarrow{\mathrm{P}} 1} \xrightarrow{d} \eta \sim N(0,1)$$

   в силу п. 2 леммы Слуцкого. Значит,
$$\mathrm{P}_\theta\left(\left|\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma(\widehat{\theta}_n)}\right| < \xi_{1-\alpha/2}\right) \to \mathrm{P}(|\eta| < \xi_{1-\alpha/2}) = 1 - \alpha$$

   То есть примерно с вероятностью $1-\alpha$ выполнено неравенство, или эквивалентно раскроем по модулю
$$\underbrace{\widehat{\theta}_n - n^{-1/2}\sigma(\widehat{\theta}_n)\xi_{1-\alpha/2} < \theta < \widehat{\theta}_n + n^{-1/2}\sigma(\widehat{\theta}_n)\xi_{1-\alpha/2}}_{\text{Асимптотический доверительный интервал уровня } 1-\alpha}$$

4. Асимптотические Гауссовские оценки можно сравнивать между собой:
   Если $\sqrt{n}(\widehat{\theta}_{i,n} - \theta) \xrightarrow{d} N(0, \sigma_i^2(\theta))$, $i = 1, 2, \ldots$, то можно посчитать асимптотическую относительную эффективность (АОЭ):
$$e_{1,2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$$

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

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

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

Теорема: 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. $\left|\frac{\partial^2}{\partial\theta^2} \ln(f(x, \theta))\right| \leq M(x)\ \forall x \in N_f,\ \theta \in \Theta,\ \mathrm{E}_\theta M(X_1) < \infty$

Тогда, если $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$, 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 $\theta, \widehat{\theta}_n \in \mathbb{R}^1$ и $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{i(\theta)}),\ n \to \infty,\ \forall \theta \in \Theta$, причем $0 < i(\theta) < \infty$, то $\widehat{\theta}_n$ называется **асимптотически эффективной оценкой**.*

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

1.3 Теорема об экстремальном свойстве правдоподобия как мотивация к построению оценок максимального правдоподобия.

Дальше $X = (X_1, \dots, X_n),\ X \sim \mathrm{P}_\theta,\ \theta \in \Theta \subseteq \mathbb{R}^1$.

Опр. 6. ***Условие (A)***

1. *Θ - интервал, $\mathrm{P}_{\theta_1} \neq P_{\theta_2}$ 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. 7. *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 $g(x)$ выпуклая книзу борелевская функция, $\mathrm{E}\,|\xi| < \infty$, $\mathrm{E}\,|g(\xi)| < \infty$. Тогда $g(\mathrm{E}\xi) \leq \mathrm{E}g(\xi)$. 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

$$\mathrm{P}_{\theta_0}(p(X, \theta_0) > p(X, \theta)) \to 1,\ n \to \infty,\ \theta_0 \neq \theta$$

Доказательство.

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

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

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

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

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 пункту 1 Условия(A)).

В силу неравенства Йенсена:

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

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

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

Опр. 8. *Случайная величина $\widehat{\theta}_n \in \Theta$ называется **оценкой максимального правдоподобия (о.м.п.)**, если $p(X, \widehat{\theta}_n) = \max_{\theta\in\Theta} p(X, \theta)$, или эквивалентно $L_n(X, \widehat{\theta}_n) = max_{\theta\in\Theta} L_n(X, \theta)$*

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

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

1.4 Теорема о существовании состоятельного решения уравнения правдоподобия.

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

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

Теорема 4: О состоятельности решения уравнения правдоподобия

Пусть выполнено Условие (A). Пусть $\forall x \in N_f$ $\exists$ непрерывная производная $f'_\theta(x, \theta)$. Тогда уравнение (2) с вероятностью, стремящейся к 1 при $n \to \infty$ имеет решение $\in \Theta$. При этом среди всех таких решений есть такой корень $\widehat{\theta}_n$, 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. $P_{\theta_0}(S_n) \to 1$.
2. Существует такое решение $\widehat{\theta}_n \in \Theta$, что

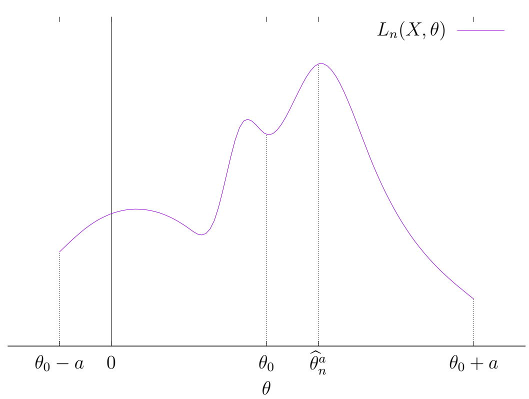
$$P_{\theta_0}\left(\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon, S_n\right) \xrightarrow[n\to\infty]{} 1, \ \forall \varepsilon > 0$$

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 $\mathrm{P}_{\theta_0}(S_n^a) \to 1$

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



Значит, $\frac{\partial}{\partial \theta} L_n(X, \widehat{\theta}_n^a) = 0$. Тогда $\mathrm{P}_{\theta_0}(S_n) \geq \mathrm{P}_{\theta_0}(S_n^a) \to 1$, 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 $\widehat{\theta}_n$, 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 $\widehat{\theta}_n$ и есть состоятельная оценка θ. Покажем это:

$\forall$ малого $\varepsilon > 0$:

$$\mathrm{P}_{\theta_0}\left(\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon, S_n\right) \geq \mathrm{P}_{\theta_0}\left(\left|\widehat{\theta}^{\varepsilon}_n - \theta_0\right| < \varepsilon, S^{\varepsilon}_n\right) \tag{3}$$

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

Но $\mathrm{P}_{\theta_0}\left(\left|\widehat{\theta}^{\varepsilon}_n - \theta_0\right| < \varepsilon, S^{\varepsilon}_n\right) \underset{\text{т.к. события из } S^{\varepsilon}_n \text{ лежат в } \left|\widehat{\theta}^{\varepsilon}_n - \theta_0\right| < \varepsilon}{=} \mathrm{P}_{\theta_0}(S^{\varepsilon}_n) \to 1$, значит в силу (3)

$$\mathrm{P}_{\theta_0}\left(\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon, S_n\right) \to 1$$

$\square$

Замечание 3. *Пусть*

$$\theta^*_n = \begin{cases} \textit{сост. корень уравнения правдоподобия, если он сущ.} \\ \theta', \ \theta' \in \Theta, \textit{иначе} \end{cases}$$

Тогда случайная величина θ^*_n *всегда определена, и* $\theta^*_n \xrightarrow{\mathrm{P}} \theta_0$*, так как*

$$\mathrm{P}\left(|\theta^*_n - \theta_0| < \varepsilon\right) = \mathrm{P}\left(\left|\widehat{\theta}_n - \theta_0\right| < \varepsilon, S_n\right) + \mathrm{P}\left(|\theta' - \theta_0| < \varepsilon, \overline{S}_n\right) \to 1$$

Ясно, что

$$\frac{\partial}{\partial\theta} L_n(X, \theta^*_n) = \overline{\overline{o}}_p(1) \tag{4}$$

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

Будем называть θ^*_n ***обобщенным состоятельным корнем уравнения правдоподобия***

1.5 Теорема об асимптотической эффективности состоятельного решения уравнения правдоподобия.

Теорема 5: Об асимптотической эффективности состоятельности решения

Пусть $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - 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

$$\left|\frac{\partial^3}{\partial\theta^3}\ln f(x,\theta)\right| \le M(x)\ \forall x \in N_f,\ \forall\theta\in\Theta,\ \mathrm{E}_{\theta_0}M(X_1) < \infty$$

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

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

То есть θ_n^* - асимптотическая эффективная оценка.

Доказательство. Будем обозначать $\frac{\partial}{\partial\theta}L_n(X,\theta), \frac{\partial^2}{\partial\theta^2}L_n(X,\theta), \ldots$ через $L_n'(\theta), L_n^{(2)}(\theta), \ldots$.

Для фиксированного X в силу формулы Тейлора и последнего замечания:

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

Отсюда,

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

<u>Рассмотрим числитель (5)</u> и покажем, что

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

Действительно,

$$\mathrm{E}_{\theta_0}\frac{f_{\theta_0}'(X_1,\theta_0)}{f(X_i,\theta_0)} = \int_{N_f}\frac{f_\theta'(x,\theta_0)}{f(x,\theta_0)}f(x,\theta_0)\nu(dx) = 0$$

$$\mathrm{D}_{\theta_0}\frac{f_{\theta_0}'(X_1,\theta_0)}{f(X_i,\theta_0)} = \mathrm{E}_{\theta_0}\left(\frac{\partial}{\partial\theta}\ln f(X_1,\theta_0)\right)^2 - \underbrace{\left(\mathrm{E}_{\theta_0}\frac{f_{\theta_0}'(X_1,\theta_0)}{f(X_i,\theta_0)}\right)^2}_{=0} \underset{\text{по опр.}}{=} i(\theta_0)$$

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

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

Теперь <u>рассмотрим знаменатель (5)</u>:

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

Действительно, в силу ЗБЧ

$$n^{-1}\sum_{i=1}^{n}\frac{f_\theta^{(2)}(X_i,\theta_0)}{f(X_i,\theta_0)}\xrightarrow{\mathrm{P}}\mathrm{E}_{\theta_0}\frac{f_\theta^{(2)}(X_1,\theta_0)}{f(X_1,\theta_0)}=\int_{N_f}\frac{f_\theta^{(2)}(x,\theta_0)}{f(x,\theta_0)}f(x,\theta_0)\nu(dx)=0$$

$$n^{-1}\sum_{i=1}^{n}\left(\frac{f'_\theta(X_i,\theta_0)}{f(X_i,\theta_0)}\right)^2\xrightarrow{\mathrm{P}}E_{\theta_0}\left(\frac{\partial}{\partial\theta}\ln f(X_1,\theta_0)\right)^2=i(\theta)$$

Применяя лемму Слуцкого, получим (7).

Далее рассмотрим второе слагаемое в знаменателе (5)

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

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 $\frac{1}{i(\theta_0)}\xi\sim N(0,\frac{1}{i(\theta_0)})$ $\square$

1.6 Построение асимптотически эффективной оценки по выборке н. о. р. случайных величин из гауссовского закона.

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

Тогда логарифмическое правдоподобие имеет вид

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

Система уравнений правдоподобия

$$\frac{\partial L_n(X,\theta)}{\partial\theta_i}=0,\ i=1,2,\dots,k \tag{9}$$

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

1. С вероятностью, стремящейся к единице при $n\to\infty$, система уравнений (9) имеет такое решение $\widehat{\theta}_n\in\Theta$, что $\widehat{\theta}_n$ сходится к истинному значению θ_0.

2. Соответствующая оценка θ_n^* асимптотически нормальна. А именно

$$\sqrt{n}(\theta_n^*-\theta_0)\xrightarrow[n\to\infty]{d}N(0,J^{-1}(\theta_0))$$

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

$$J(\theta)=(J_{ij}(\theta))=\left(\mathrm{E}_\theta\left\{\frac{\partial\ln f(X,\theta)}{\partial\theta_i}\cdot\frac{\partial\ln f(X,\theta)}{\partial\theta_j}\right\}\right)$$

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

Здесь $p(x,\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2\right\}$, *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 $\overline{X}$, *причем в т.* $\theta = \overline{X}$ $L_n(X,\theta)$ *достигает максимума, так как* $\frac{\partial^2 L_n(X,\overline{X})}{\partial\theta^2} = -\frac{1}{\sigma^2} < 0$

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

$$\theta_n^* = \begin{cases} \overline{X}, \ a < \overline{X} < b \\ \frac{a+b}{2}, \ \overline{X} \notin (a,b) \end{cases} \tag{10}$$

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

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0,\sigma^2) \tag{11}$$

Напомним, что в этой модели $i(\theta) = \frac{1}{\sigma^2}$. *Справедливость (11) с* θ_n^* *из (10) 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.*

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

Но на границах теряется асимптотическая Гауссовость.

2 Проверка статистических гипотез

2.1 Задача проверки параметрических гипотез - основные определения.

$X = (X_1, \dots, X_n)$ 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 $\Theta_1 \in \Theta \backslash \Theta_0$*

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 $X \in \overline{S} = X \backslash S$, то H_0 принимается.

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

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

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

Тогда

$$\alpha = \alpha(\theta) = W(\theta),\ \theta \in \Theta_0;$$
$$\beta = \beta(\theta) = 1 - W(\theta),\ \theta \in \Theta_1$$

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

$$\alpha(\theta) = W(\theta) = \mathrm{P}_\theta(X \in S) \leq \alpha\ \forall \theta \in \Theta_0$$

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:

$$\mathrm{P}_{\theta_0}(X \in S^*_\alpha) \leq \alpha,$$
$$\mathrm{P}_{\theta_1}(X \in S^*_\alpha) \geq \mathrm{P}_{\theta_1}(X \in S_\alpha)\ \forall S_\alpha$$

2.2 Лемма Неймана-Пирсона

Положим для краткости: $p_0(x) \stackrel{\text{def}}{=} p(x, \theta_0),\ \mathrm{E}_0 \stackrel{\text{def}}{=} \mathrm{E}_{\theta_0},\ p_1(x) \stackrel{\text{def}}{=} p(x, \theta_1),\ \mathrm{E}_1 \stackrel{\text{def}}{=} \mathrm{E}_{\theta_1}$

Введем множество

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

Теорема: Лемма Неймана-Пирсона

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

1. $\mathrm{P}_0(X \in R) \leq \mathrm{P}_0(X \in S(\lambda))$

   Тогда:

2. $P_1(X \in R) \leq \mathrm{P}_1(X \in S(\lambda))$
3. $P_1(X \in S(\lambda)) \geq \mathrm{P}_0(X \in S(\lambda))$

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*

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

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

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

$$\mathrm{E}_0 I_R(x) \leq \mathrm{E}_0 I_S(x) \tag{1}$$

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 $x \in \mathbb{R}^n$:

$$\mathrm{E}_1 I_R(X) - \lambda \mathrm{E}_0 I_R(X) \leq \mathrm{E}_1 I_S(X) - \lambda \mathrm{E}_0 I_S(X)$$

$$\mathrm{E}_1 I_S(X) - \mathrm{E}_1 I_R(X) \geq \lambda \underbrace{[\mathrm{E}_0 I_S(X) - \mathrm{E}_0 I_R(X)]}_{\geq 0 \text{ по условию (1)}} \tag{3}$$

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

$$E_1 I_S(X) \geq E_1 I_{\mathbb{R}}(X)$$

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

$$\mathrm{P}_0(X \in S) = \int_{R^n} I_S(X) p_0(x) \mu(dx) \leq \int_{R^n} I_S(X) p_1(x) \mu(dx) = \mathrm{P}_1(X \in S)$$

То есть $\mathrm{P}(H_1 \,| H_0) \leq \mathrm{P}(H_1 |\, H_1)$

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

$$\mathrm{P}_1(X \in \overline{S}) = \int_{R^n} I_{\overline{S}}(X) p_1(x) \mu(dx) \leq \int_{R^n} I_{\overline{S}}(X) p_0(x) \mu(dx) = \mathrm{P}_0(X \in \overline{S})$$

То есть $1 - \mathrm{P}_1(X \in S) \leq 1 - \mathrm{P}_0(X \in S)$, откуда $\mathrm{P}_1(X \in S) \geq \mathrm{P}_0(X \in S)$ □

2.3 Оптимальная проверка гипотезы о среднем гауссовской выборки как пример применения леммы Неймана-Пирсона.

Пример. *$X = (X_1, \ldots, X_n)$, $\{X_i\}$ - 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*

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

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

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

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

Итак,

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

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

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

Преобразуем левую сумму в стандартную Гауссовскую величину. Тогда

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

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

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

3. *Положим* $S^*_\alpha=\{x:\sum_{i=1}^n x_i>\widetilde{\lambda}_\alpha\}$. *Тогда:*

$$\mathrm{P}_{\theta_0}(X\in S^*_\alpha)=\alpha \text{ и } \forall S_\alpha\ \mathrm{P}_{\theta_0}(X\in S_\alpha)\le\alpha=\mathrm{P}_{\theta_0}(X\in S^*_\alpha)$$

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

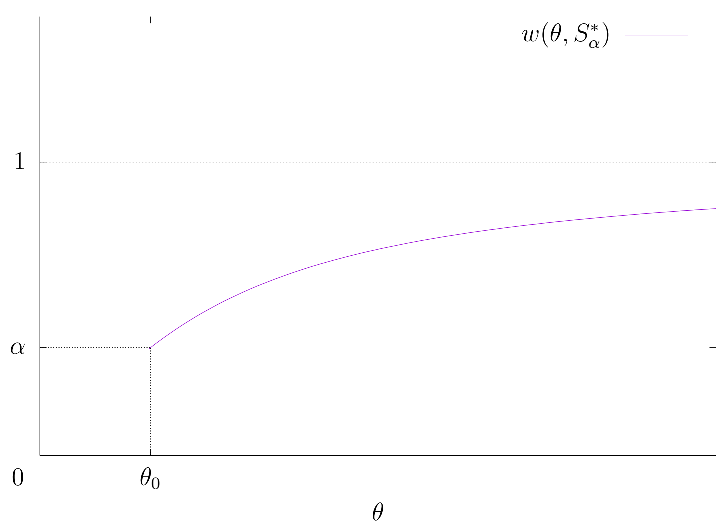
$$\mathrm{P}_{\theta_1}(X\in S_\alpha)\le\mathrm{P}_{\theta_1}(X\in S^*_\alpha)$$

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* $H_1^+:\theta>\theta_1$ *Мощность критерия* S^*_α *для* H_0 *при альт.* H_1^+

$$W(\theta,S^*_\alpha)=\mathrm{P}_\theta\left(\sum_{i=1}^n X_i>n\theta_0+\sqrt{n}\sigma\xi_{1-\alpha}\right)=$$

$$=\mathrm{P}_\theta\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n(X_i-\theta)>\frac{\sqrt{n}(\theta_0-\theta)}{\sigma}+\xi_{1-\alpha}\right)=1-\Phi\left(\xi_{1-\alpha}+\frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right)$$



2.4 Теорема о двойственности задач доверительного оценивания и проверки гипотез. Пример - проверка гипотезы о среднем гауссовской выборки при двухмерной альтернативе.

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

$$\mathrm{P}_\theta(\theta \in \Theta^*(X, \alpha)) \geq 1 - \alpha \ \forall \theta \in \Theta$$

Теорема 1

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 $\Theta^*(x, \alpha) = \{\theta : x \in \overline{S_\alpha}(\theta)\}$. 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 $\overline{S_\alpha}(\theta_0) = \{x : \theta_0 \in \Theta(x, \alpha)\}$ есть область принятия гипотезы H_0 (следовательно и критерий).

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

Доказательство.

$$1.\ \mathrm{P}_\theta(\theta \in \Theta^*(X, \alpha)) = \mathrm{P}_\theta(X \in \overline{S_\alpha}(\theta)) = 1 - \underbrace{\mathrm{P}_\theta(X \in S_\alpha(\theta))}_{\leq \alpha} \geq 1 - \alpha \ \forall \theta \in \Theta$$

$$2.\ \mathrm{P}_{\theta_0}(X \in S_\alpha(\theta_0)) = 1 - \mathrm{P}_{\theta_0}(X \in \overline{S_\alpha}(\theta_0)) = 1 - \underbrace{\mathrm{P}_{\theta_0}(\theta_0 \in \Theta^*(X, \alpha))}_{\geq 1-\alpha} \leq 1 - (1 - \alpha) = \alpha$$

□

Пример. *Пусть $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - 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 $\alpha,\ 0 < \alpha < 1$.*

Построим доверительное множество для θ уровня $1 - \alpha$. Пусть $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ - оптимальная оценка θ. Тогда

$$\frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} \sim N(0, 1)$$

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

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

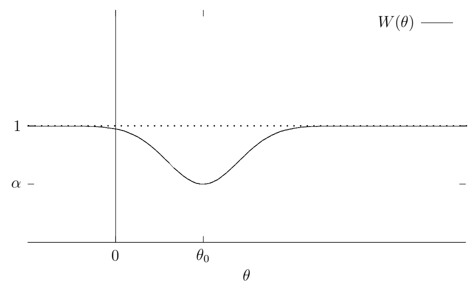
То есть $\Theta^*(X, \alpha) = \left\{\theta : \left|\frac{\sqrt{n}(\overline{X}-\theta)}{\sigma}\right| < \xi_{1-\alpha/2}\right\}$. В силу замечания к Теореме 1 $S_\alpha(\theta_0) = \left\{X : \left|\frac{\sqrt{n}(\overline{X}-\theta_0)}{\sigma}\right| \geq \xi_{1-}\right.$ есть критическое множество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

$$W(\theta) = \mathrm{P}_\theta(X \in S_\alpha(\theta_0)) = \mathrm{P}_\theta\left(\left|\frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma}\right| \geq \xi_{1-\alpha/2}\right) = 1 - \mathrm{P}_\theta\left(\left|\frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma}\right| < \xi_{1-\alpha/2}\right) =$$

$$= 1 - \mathrm{P}\left(-\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} < \frac{\sqrt{n}(\overline{X} - \theta)}{\sigma} < \xi_{1-\alpha} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right) =$$

$$= 1 - \left[\Phi\left(\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right) - \Phi\left(-\xi_{1-\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right)\right] =$$

$$= \left[\Phi\left(\xi_{\alpha/2} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right) + \Phi\left(\xi_{\alpha/2} + \frac{\sqrt{n}(\theta - \theta_0)}{\sigma}\right)\right]$$



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

2.5 Определения нецентральных распределений Стьюдента, хи-квадрат Пирсона и Фишера. Лемма о нецентральном распределении хи-квадрат.

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.*

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

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

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

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

имеет нецентральное распределение хи-квадрат Пирсона 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 нецентральное 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 $a_1, \ldots, a_k$. А именно*

$$\eta_k \stackrel{d}{=} (z_1 + \Delta)^2 + z_2^2 + \ldots + z_k^2, \text{ где } \{z_1, \ldots, z_k\} \text{ - н.о.р. } N(0,1) \text{ сл.в.}$$

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 $(\xi_1, \ldots, \xi_k)$ - н.о.р. $N(a_i, 1)$ сл.в. Пусть $\xi = (\xi_1, \ldots, \xi_k)^T$, ортогональная матрица

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

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

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

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

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

□

2.6 Лемма о стохастической упорядоченности нецентральных распределений Стьюдента, хи-квадрат Пирсона и Фишера.

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

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

Аналогично

$$\mathrm{P}(\eta_k(\Delta_2) > x) > \mathrm{P}(\eta_k(\Delta_1) > x), \Delta_2 > \Delta_1 \tag{5}$$

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

> *Нецентральные распределения Пирсона и Фишера стохастически упорядочены по параметру нецентральности.*

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

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

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

В силу (7)

$$\mathrm{P}(t_k(\mu_2) > x) = \mathrm{P}\left(\frac{\xi+\mu_2}{\sqrt{\frac{1}{k}\eta_k}} > x \right) = \mathrm{E}I\left(\xi > x\sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) =$$

$$= \mathrm{E}\left\{ 1 - I\left(\xi \le x\sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) \right\} = 1 - E\left\{ EI(\xi \le y)\Big|_{y = x\sqrt{\frac{\eta_k}{k}} - \mu_2} \right\} =$$

$$= 1 - \mathrm{E}\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) > 1 - \mathrm{E}\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_1 \right) = \mathrm{P}(t_k(\mu_1) > x)$$

$$\text{так как } \mathrm{E}\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_2 \right) < \mathrm{E}\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_1 \right) \text{ в силу возрастающей } \Phi(y)$$

□

2.7 Линейная гауссовская регрессия - доверительный эллипсоид для параметра $\beta = Ac$

Обратимся к линейной гауссовской модели

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

$$X = (X_1, \ldots, X_n)^T \text{ - наблюдения, } Z \text{ - } n \times p \text{ матрица регрессоров } p < n$$
$$\mathcal{E} \sim N(0, \sigma^2 \mathrm{E}_n),\ c = (c_1, \ldots, c_p)^T$$
$$c \text{ и } \sigma^2 \text{ неизвестны}$$

Рассмотрим новый вектор $\beta = Ac$, $A - k \times p$ матрица, $\mathrm{rk}\, A = k, k \le p$.

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

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

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

> Заметим, что $D > 0$, так как для $\alpha \in \mathbb{R}^k, \alpha \ne 0$,
>
> $$\alpha^T D\alpha = (A^t\alpha)^T(Z^TZ)^{-1}(A^T\alpha) > 0, \text{ т.к. } (Z^TZ)^{-1} > 0, A^T\alpha \ne 0 \text{ при } \mathrm{rk}\, A = k, \alpha \ne 0$$

В силу пункта 2 леммы 1

$$\frac{1}{\sigma^2}\left(\widehat{\beta}_n - \beta\right) D^{-1} \left(\widehat{\beta}_n - \beta\right) \sim \chi^2(k)$$

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

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

Значит,

$$\mathrm{P}_{\beta,\sigma^2}\left((\widehat{\beta}_n-\beta)^T D^{-1}(\widehat{\beta}_n-\beta) \le k\widehat{s}_n^2 f_{1-\alpha}(k,n-p)\right) = 1-\alpha$$

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

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

2.8 Критерий Фишера для линейной гипотезы $H_0 : \beta = \beta_0$ в гауссовской линейной регрессии. Его несмещенность и строгая монотонность мощности.

Рассмотрим проверку гипотезы $H_0 : \beta = \beta_0$ против $H_1 : \beta \ne \beta_0$. H_0 называют линейной гипотезой, так как $\beta = Ac$ 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:

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

То есть критическое множество (критерий уровня α):

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

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

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

Параметр нецентральности

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

Мощность F-критерия

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

Свойства мощности:

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

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

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

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

2.9 Определение порядка регрессии с помощью критерия Фишера.

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

$H_0: \ c_{(2)} = 0 \quad$ (*порядок не больше* m)
$H_1: \ c_{(2)} \neq 0$

Рассмотрим матрицу

$$A = \overbrace{\left(\begin{array}{ccc|ccc} 0 & & & 1 & & \\ & \ddots & & & \ddots & \\ & & 0 & & & 1 \end{array}\right)}^{p} \Rightarrow Ac = c_{(2)} \Rightarrow H_0 \Leftrightarrow Ac = 0$$
$$\qquad\qquad m \qquad\qquad p-m$$

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

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

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

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

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

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

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

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

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

2.10 Проверка однородности двух гауссовских выборок.

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.*

$H_0: \ a = b \quad$ (***гипотеза однородности***)
$H_1: \ a \neq b$

Замечание. *При* $\mathrm{D}X_1 \neq \mathrm{D}Y_1$ *эта задача называется* ***проблемой Беренса-Фишера***.

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

$$\widehat{X} \stackrel{def}{=} (X_1, \dots, X_m, Y_1, \dots, Y_n)^T \qquad c = (a, b)^T \qquad \mathcal{E}^T = (\varepsilon_1, \dots, \varepsilon_m, \widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n)^T$$

$$Z = \begin{pmatrix} m\begin{cases} 1 \\ \vdots \\ 1 \end{cases} & 0 \\ 0 & n\begin{cases} 1 \\ \vdots \\ 1 \end{cases} \end{pmatrix} \Rightarrow \underbrace{\widehat{X} = Zc + \mathcal{E}}_{\textbf{гаусс. лин. регрессия}} \qquad (12)$$

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

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

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

О.н.к. для вектора c - решение задачи

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

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

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

Тогда

$\widehat{\beta}_n = A\widehat{c}_n = \overline{X} - \overline{Y}$

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

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

F-критерий для H_0 *имеет вид*

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

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

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

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

1. *Если* $|a - b|$ *возрастает, то мощность F-теста возрастает*

2. *Если* $\sigma \to 0$ *или* $\frac{1}{n} + \frac{1}{m} \to 0$, *то мощность тоже возрастает*

2.11 Критерий согласия хи-квадрат Пирсона. Проверка для гипотезы о параметрах полиномиального распределения: опредлеление, поведение статистики Пирсона при альтернативах и теорема Пирсона. Проверка гипотезы о виде функции распределения.

Пусть проводятся n независимых испытаний, и в каждом испытании возможны $m \geq 2$ исходов $A_1, \dots, A_m$ таких, что $A_i A_j = \varnothing,\ i \neq j,\ \sum A_i = \Omega$, тогда $\mathrm{P}(A_j) = p_j > 0,\ \sum_{j=1}^m p_j = 1$. 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:

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

$H_1 :\ \ p_j \neq p_j^\circ$ хотя бы при одном j

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

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

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

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

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 $\frac{\nu_j}{n} \xrightarrow{\mathrm{P}} p_j$. Поэтому

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

Значит,

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

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

Теорема: Теорема Пирсона

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

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

Замечание. *Тогда*

$$\mathrm{P}(H_1|H_0) = \mathrm{P}(\chi^2_n > \chi_{1-\alpha}(m-1)|H_0) \to \alpha$$

$$\mathrm{P}(H_0|H_1) = \mathrm{P}(\chi^2_n \le \chi_{1-\alpha}(m-1)|H_1) \to 0$$

То есть

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

Вероятность принять правильную гипотезу близка к единице!

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

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

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 $\nu = \sum_{i=1}^n X_i$

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

Здесь $\{X_i\}$ - н.о.р., $EX_1 = p,\ \operatorname{cov}(X_1, X_1) = \mathrm{E}(X_1 - p)(X_1 - p)^T = \mathrm{E}X_1X_1^T - pp^T = P - pp^T$. Поэтому соотношение (13) следует из соотноешния (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

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

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

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

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

$$U(E_m - zz^T)U^T = E_m - (Uz)(Uz)^T =$$

$$= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} = \widetilde{\mathrm{E}}_1$$

В силу (15) и теоремы о слабой сходимости

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

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

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

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 соотноешния (17) следует теорема Пирсона. □

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

$H_0: \ \ F(x) = F_0(x), \ (F_0$ *известна*$)$

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

Разобьем носитель X_1 *на непересекающиеся отрезки* $\Delta_1, \ldots, \Delta_m$, $m \geq 2$ *так, что* $X_1 \in \Delta_1 \cup \Delta_2 \cup \ldots \cup \Delta_m$

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

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

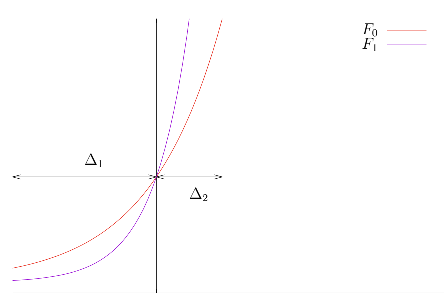
$$\chi_n^2 \stackrel{def}{=} \sum_{j=1}^{m} \frac{(\nu_j - np_j^\circ)^2}{np_j^\circ} \xrightarrow[H_0]{d} \chi^2(m-1)$$

Правило: H_0 будем отвергать, если $\chi_n^2 > \chi_{1-\alpha}(m-1)$. *($\alpha$ задано) Тогда* $\mathrm{P}(H_1|H_0) \xrightarrow[n\to\infty]{} \alpha$.

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

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

Замечание. *Если* $F_0 \not\equiv F_1$, *но* $p_j = p_j^\circ \ \forall j$, *то* $\mathrm{P}(H_0|H_1) = \mathrm{P}(H_0|H_0) \to 1 - \alpha \neq 0$. *Например:*



$m = 2$

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

2.12 Проверка сложной гипотезы в схеме испытаний Бернулли. Проверка независимости признаков.

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

Условия регулярности:

1. $\sum_{j=1}^m p_j(\theta) = 1, \ \theta \in \Theta$
2. $p_j(\theta) \geq c > 0 \ \forall j = 1, \dots, m$ и $\exists \ \frac{\partial p_j(\theta)}{\partial \theta_l}, \frac{\partial^2 p_j(\theta)}{\partial \theta_l \partial \theta_r}$
3. $\mathrm{rk} \underbrace{\frac{\partial p_j(\theta)}{\partial \theta_l}}_{m \times k} = k, \ \forall \theta \in \Theta$

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

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

логарифмического правдоподобия:

$$L_n(\nu, \theta) = \ln\left(\frac{n!}{\nu_1! \dots \nu_m!}\right) + \sum_{j=1}^m \nu_j \ln p_j(\theta)$$

оценки максимального правдоподобия (мультиномиальные):

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

Теорема: Теорема Фишера

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

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

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

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 $p_{ij} = \mathrm{P}(A_i B_j)$. 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

$$\widehat{p}_{i\bullet} = \frac{\nu_{i\bullet}}{n},\ \widehat{p}_{\bullet j} = \frac{\nu_{\bullet j}}{n},\ \text{где } \nu_{i\bullet} = \sum_{j} \nu_{ij},\ \nu_{\bullet j} = \sum_{i} \nu_{ij}$$

Статистика хи-квадрат имеет вид:

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

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

так как $m - k - 1 = sr - (s + r - 2) - 1 = (s-1)(r-1)$.
Правило: Если $\widehat{\chi}_n^2 > \chi_{1-\alpha}((s-1)(r-1))$, 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* $\boxed{H_0: \text{ признаки независимы}}$
E – *достаточно умный*
F – *способный*
G – *очень способный*

	Способности						
Как одевается	A и B	C	D	E	F	G	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 $15 = (s-1)(r-1) = (4-3)(6-1) \Rightarrow$ *Отвергаем* H_0

3 Введение в робастное оценивание

3.1 Схема засорений данных Мартина-Йохаи, функционал влияния, чувствительность к грубым выбросам. Теорема о вычислении функционала влияния. Эмпирическое среднее. М-оценка параметра сдвига.

Схема засорений Мартина-Йохаи имеет вид:

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

Здесь $\{u_t\}$ - "полезный сигнал"(временной ряд);
$\{z_t^\gamma\}$ - н.о.р. сл.в., $z_1^\gamma \sim Bin(1, \gamma)$ с $0 \le \gamma \le 1$ (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;
Последовательности $\{u_t\}, \{z_t^\gamma\}, \{\xi_t\}$ независимы между собой.
Пусть $y_1, \dots, y_n$ - наблюдения, и распределение вектора $Y_n = (y_1, \dots, y_n)$ зависит от неизвестного параметра β. Пусть $\hat{\beta}_n$ - некоторая оценка β

Основное предположение

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

$$\hat{\beta}_n \xrightarrow[n \to \infty]{\text{P}} \theta_\gamma;\ \theta_0 = \beta$$

Опр. 1. *Если существует предел*

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

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

Если функционал влияния существует, то

$$\theta_\gamma = \theta_0 + IF(\theta_\gamma, \mu_\xi)\gamma + \overline{\overline{o}}(\gamma),\ \gamma \to +0$$

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

Опр. 2. *Величина $GES(\theta_\gamma, M_\xi) \overset{def}{=} \sup_{\mu_\xi \in M_\xi} |IF(\theta_\gamma, \mu_\xi)|$ называется чувствительностью оценки $\widehat{\beta}_n$ 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 max при таких γ

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

Пример (Выборочное среднее).

$$\begin{cases} u_t = a + \varepsilon_t \\ y_t = u_t + z_t^\gamma \xi_t,\ t = 1, \dots, n \end{cases}$$

$\{\varepsilon_t\}$ - *н.о.р. сл.в.*, $\mathrm{E}\varepsilon_1 = 0$ *(тогда* $\mathrm{E}u_t = a$*)*, $\mathrm{E}|\xi_1| < \infty$
Возьмем оценкой a *эмпирическое среднее* $\overline{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$. *Тогда*

$$\overline{y} \xrightarrow{\mathrm{P}} E(u_1 + z_1^{\gamma}\xi_1) = a + \gamma \mathrm{E}\xi_1 = \theta_{\gamma}^{LS}$$

Функция θ_{γ}^{LS} *определена при всех* γ,

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

Если M_1 *- класс распределений с конечным первым моментом, то*

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

Оценка $\overline{y}$ *не B-робастна на классе* M_1*!*

3.2 Вычисление функционала влияния и чувствительности выборочной медианы, её робастность.

Пример (Выборочная медиана)**.** *Пусть*

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

$\varepsilon_1 \sim G(x)$, *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* $G(x) + G(-x) = 1\ \forall x$*), то автоматически* $G(0) = 1/2$. *Если вдобавок* $\mathrm{E}|\varepsilon_1| < \infty$, *то* $\mathrm{E}\varepsilon_1 = 0, \mathrm{E}u_1 = a$

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

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

Опр. 4. *Величина*

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

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:*

$$\sqrt{n}(\widehat{m}_n - a) \xrightarrow[n\to\infty]{d} N(0, \frac{1}{4g^2(0)})$$

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

$$\boxed{e_{\widehat{m}_n,\ \overline{X}} = 4g^2(0)\sigma^2}$$

<u>Изучим B-робастность выборочной медианы.</u> Пусть

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

Теорема 1

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

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

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

3. *Чувствительность выборочной медианы на классе всех возможных распределений M_ξ*

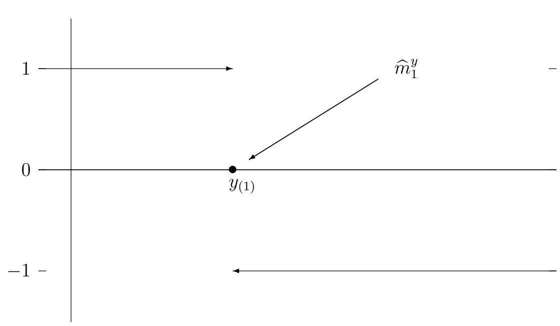
$$GES(\theta_\gamma^m, \mu_\xi) = \sum_{\mu_\xi \in M_\xi} |IF(\theta_\gamma^m, \mu_\xi)| = \frac{1}{2g(0)} < \infty$$

то есть <u>выборочная медиана B-робастна</u>.

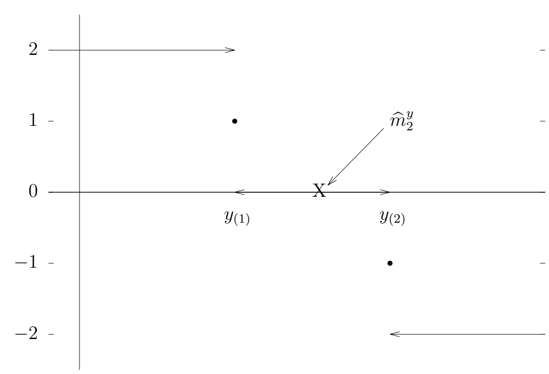
Доказательство. 1. Выборочная медиана $\widehat{m}_n^y$ удовлетворяет уравнению

$$l_n(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n} \operatorname{sgn}(y_t - \theta) = 0, \ \text{где} \ \operatorname{sgn}(x) = \begin{cases} -1, x < 0 \\ 0, x = 0 \\ 1, x > 0 \end{cases} \tag{2}$$

Справедливость формулы (2) легко понять из следующих рисунков:



$n = 1$



$n = 2$

Так бывает всегда: при нечетном n решение уравнения (2) всегда $\exists!$, это $\widehat{m}_n^y$; при четном n решений целый интервал и $\widehat{m}_n^y$ - его середина.

В силу Закона Больших Чисел при любом θ и любом $0 \le \gamma \le 1$

$$l_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \operatorname{sgn}(y_t - \theta) \xrightarrow{\text{P}} \mathrm{E}\operatorname{sgn}(y_1 - \theta) \stackrel{\text{def}}{=} \Lambda_M(\gamma, \theta)$$

Задача: Пусть ξ и η - независимые случайные векторы, причем η - дискретный вектор со значениями $\eta_1, \eta_2, \ldots$. Проверить, что

$$\mathrm{E}\phi(\xi,\eta) = \sum_{k\geq 1} \mathrm{E}\phi(\xi,\eta_k)\mathrm{P}(\eta=\eta_k) = \sum_{k\geq 1} \mathrm{E}(\phi(\xi,\eta_k)|H_k)\mathrm{P}(H_k), \text{ где гипотеза } H_k = (\eta=\eta_k)$$

Найдем удобный вид для $\Lambda_M(\gamma,\theta)$. Имеем

$$\Lambda_M(\gamma,\theta) = \mathrm{E}(1-2I(y_1-\theta\leq 0)) = 1-2\mathrm{E}I(\varepsilon_1 \leq \theta - a - z_1^\gamma \xi_1) = 1-2\mathrm{E}G(\theta-a-z_1^\gamma\xi_1) \quad (3)$$

так как $\operatorname{sgn} x = 1-2I(x<0)$, $x\neq 0$. Чтобы упростить (3), введем две гипотезы $H_1 = (z_1^\gamma = 0)$ и $H_2 = (z_2^\gamma = 1)$. Тогда, используя задачу, получаем из (3):

$$\Lambda_M(\gamma,\theta) = 1-2(1-\gamma)G(\theta-a) - 2\gamma\mathrm{E}G(\theta-a-\xi_1)$$

Функция $\Lambda_M(\gamma,\theta)$ определена при всех γ,θ.

2. Функция $\Lambda_M(\gamma,\theta)$ в окрестности точки $(0,a)$ удовлетворяет всем предположениям теормы о неявной функции. А именно:

   (a) $\Lambda_M(0,a) = 1-2G(0) = 0$

   (b) Существует и непрерывны по паре (γ,θ) функции $\frac{\partial\Lambda_M(\gamma,\theta)}{\partial\gamma}$ и $\frac{\partial\Lambda_M(\gamma,\theta)}{\partial\theta}$

   (c)
   $$\frac{\partial\Lambda_M(\gamma,\theta)}{\partial\theta} = -2g(0)\neq 0$$

   Значит, в окрестности точки $(0,a)$ определена функция $\theta_m(\gamma) = \theta_\gamma^m$ такая, что

   $$\Lambda_M(\gamma,\theta_\gamma^m) = 0$$

   Кроме того, $\theta_0^m = a$; $\theta_\gamma^m \to \theta_0$ при $\gamma\to 0$; Функция θ_0^m дифференцируема в точке $\gamma = 0$, и
   $$\left.\frac{d\theta_\gamma^m}{d\gamma}\right|_{\gamma=0} = -\left(\frac{\partial\Lambda_m(0,a)}{\partial\theta}\right)^{-1}\frac{\partial\Lambda_m(0,a)}{\partial\gamma} = \frac{1-2\mathrm{E}G(-\xi_1)}{2g(0)} \quad (4)$$

3. Покажем, что
   $$\widehat{m}_n^y \xrightarrow[n\to\infty]{\mathrm{P}} \theta_\gamma^m \quad (5)$$

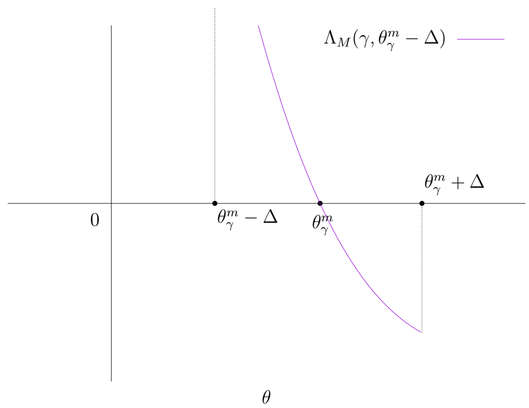
   Тогда из (4)-(5) будет слеовать, что функционал влияния выборочной медианы равен

   $$IF(\theta_\gamma^m,\mu_\xi) = \frac{1-2\mathrm{E}G(-\xi_1)}{2g(0)} \quad (6)$$

   Модуль числителя в (6) не больше единицы, причем если θ_1 неслучайно и $\theta_1\to\infty$, то числитель стремится к единице. Значит,

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

   то есть мы докажем теорему.



Докажем (5). Имеем при малых $\gamma,\ \xi,\ \theta$ (γ-фикс.) вблизи a:

$$\frac{\partial \Lambda_M(\gamma,\theta)}{\partial \theta} = -2(1-\gamma)g(\theta-a) - 2\gamma \mathrm{E} g(\theta - a - \xi_1) < 0$$

то есть $\underline{\Lambda_M(\gamma,\theta) \text{ убывает по } \theta}$. Значит, $\begin{cases} \Lambda_M(\gamma, \theta_\gamma^m - \Delta) > 0 \\ \Lambda_M(\gamma, \theta_\gamma^m + \Delta) < 0 \end{cases}$ Но

$$\begin{cases} l_n(\theta_\gamma^m - \Delta) \xrightarrow{\mathrm{P}} \Lambda_M(\gamma, \theta_\gamma^m - \Delta) > 0 \\ l_n(\theta_\gamma^m + \Delta) \xrightarrow{\mathrm{P}} \Lambda_M(\gamma, \theta_\gamma^m + \Delta) < 0 \end{cases} \tag{7}$$

Функция $l_n(\theta)$ монотонно убывает (точнее, не возрастает) по θ. В силу (7) с вероятностью сколь угодно близкой к единице при достаточно больших n <u>все</u> корни уравнения $l_n(\theta) = 0$ лежат в интервале $(\theta_\gamma^m - \Delta, \theta_\gamma^m + \Delta)$. А значит и выборочная медиана тоже! Поскольку $\Delta > 0$ любое, то получаем

$$\widehat{m}_n^y \xrightarrow[n\to\infty]{\mathrm{P}} \theta_\gamma^m$$

что и доказывает теорему.

$\square$

Как находить функционал влияния в общей ситуации?

Пусть оценка $\widehat{\beta}_n$ ищется как корень уравнения:

$$l_n(\theta) \overset{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n} \phi_t(J_n, \theta) = 0 \tag{8}$$

Пусть будут выполнены следующие условия

1. $$l_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \phi_t(J_n,\theta) \xrightarrow{\mathrm{P}} \Lambda(\gamma,\theta)$$

   при всех $|\theta - \beta| < \delta,\ 0 \le \gamma < \gamma_0$

2. $\Lambda(0, \beta) = 0$

3. Пусть $\Lambda(\gamma, \theta)$ можно продолжить на отрезок малых γ так, что при $|\theta - \beta| < \delta,\ |\gamma| < \gamma_0$ существуют и непрерывны по паре аргументов (γ, θ) частные производные $\frac{\partial \Lambda(\gamma,\theta)}{\partial \gamma}$, $\frac{\partial \Lambda(\gamma,\theta)}{\partial \theta}$.

4. Пусть $\lambda(\beta) \stackrel{\text{def}}{=} \frac{\partial \Lambda(0,\theta)}{\partial \theta} \neq 0$

Теорема 2

Пусть выполнены условия (1)-(4), и функции $\phi_t(J_n, \theta)$ непрерывны по θ. Тогда уравнение (8) с вероятностью, стремящейся к единице при $n \to \infty$, имеет при достаточно малых $\gamma > 0$ решение $\widehat{\beta}_n$, что соответствующая оценка $\widetilde{\beta}_n \xrightarrow{\text{P}} \theta_\gamma,\ \theta_0 = 0$, и существует функционал влияния:

$$\boxed{IF(\theta_\gamma, \mu_\xi) = -\frac{1}{\lambda(\beta)} \frac{\partial}{\partial \gamma} \Lambda(0, \beta)}$$

Пример (M-оценка медианы)**.** *Пусть*

$$\begin{cases} u_t = a + \varepsilon_t \\ y_t = u_t + z_t^\gamma \xi_t \end{cases} \quad \{\varepsilon_t\} \text{ - н.о.р.},\ \varepsilon_1 \sim g(x) = G'(x),\ g(x) = g(-x)$$

Тогда a - медиана функции распределения случайной величины u_1.

Будем искать оценку a, обозначим ее как $\widehat{a}_n$, как корень уравнения

$$\sum_{t=1}^{n} \psi(y_t - \theta) = 0 \tag{9}$$

*Тогда оценка называется M**-оценкой**. В частности, при $\psi(x) = x,\ \widehat{a}_n = \overline{y}$; при $\psi(x) = \operatorname{sgn}(x),\ \widehat{a}_n = \widehat{m}_n^y$.*

Пусть выполняются условия:

1. *$\psi(x)$ - нечетная строго возрастающая функция*

$$\lim_{x \to +\infty} \psi(x) = c_1 > 0,\ \lim_{x \to -\infty} \psi(x) = c_2 < \infty$$

2. *Существует непрерывная и ограниченая $\psi'(x)$, $\mathrm{E}\psi'(\varepsilon_1) \neq 0$*

Тогда уравнение (9) всегда имеет и притом <u>единственное</u> решение. Условия (1)-(2) выполнены, например, для $\psi(x) = arctan(x)$.

Найдем функционал влияния и чувствительность M-оценки, используя теорему 2. Проверим ее условия:

1. $$\frac{1}{n} \sum_{t=1}^{n} \psi(y_t - \theta) \xrightarrow{\text{P}} \mathrm{E}\psi(y_1 - \theta) \stackrel{def}{=} \Lambda(\gamma, \theta),\ \forall \theta,\ \forall 0 \le \gamma \le 1$$

   Введем гипотезы $H_1 = (z_1^\gamma = 0),\ H_2 = (z_2^\gamma = 0)$. Тогда

$$\Lambda(\gamma, \theta) = \sum_{i=1}^{2} \mathrm{E}(\psi(\underbrace{\varepsilon_1 + a + z_1^\gamma \xi_1}_{y_1} - \theta)|H_i)\mathrm{P}(H_i) = (1-\gamma)\mathrm{E}\psi(\varepsilon_1 + a - \theta) + \gamma \mathrm{E}\psi(\varepsilon_1 + \xi_1 + a - \theta)$$

2. $\Lambda(0,a) = \mathrm{E}\psi(\varepsilon_1) = 0$*, так как* $\psi(x)$ *- нечетная, а* $g(x)$ *- четная.*

3. Функция $\Lambda(\gamma,\theta)$ *определена при всех* γ *и* θ*. Частные производные* $\frac{\partial\Lambda(\gamma,\theta)}{\partial\gamma}$, $\frac{\partial\Lambda(\gamma,\theta)}{\partial\theta}$ *существуют при условиях (1)-(2) и непрерывны по паре* (γ,θ)*. В частности,*

$$\frac{\partial\Lambda(0,a)}{\partial\gamma} = -\mathrm{E}\psi(\varepsilon_1) + \mathrm{E}\psi(\varepsilon_1+\xi_1) = \mathrm{E}\psi(\varepsilon_1+\xi_1)$$

4. $\frac{\partial\Lambda(0,a)}{\partial\theta} = -\mathrm{E}\psi'(\varepsilon_1) \neq 0$

В силу теоремы 2

$$\widehat{a}_n \xrightarrow{\mathrm{P}} \theta_\gamma,\ \theta_0 = a$$

$$IF(\theta_\gamma,\mu_\xi) = \frac{\mathrm{E}\psi(\varepsilon_1+\xi_1)}{\mathrm{E}\psi'(\varepsilon_1)}$$

$$GES(\theta_\gamma,M_\xi) \leq \frac{\max\{|c_1|,|c_2|\}}{\mathrm{E}\psi'(\varepsilon_1)} < \infty,\ M_\xi \text{ - класс всех вер. распределений}$$

4 Статистистический анализ авторегрессионных моделей.

4.1 $AR(1)$ модель с коэффициентом из $\mathbb{R}^1$. Нахождение оценок максимального правдоподобия. Гауссовские и лапласовские инновации.

Пусть $\ldots, S_{-1}, S_0, S_1, \ldots$ - стоимости ценных бумаг, например, акций. Величины

$$u_t = \ln(S_t/S_{t-1}) = \ln S_t - \ln S_{t-1}$$

называются логарифмическими приращениями и для описания их поведения часто используют стохастические разностные уравнения. Например, $AR(p)$-уравнение имеет вид

$$u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_p u_{t-p} + \varepsilon_t,\ t \in \mathbb{Z}$$

Здесь $\{\varepsilon_t\}$ - н.о.р. сл.в., $\mathrm{E}\varepsilon_1 = 0$, $0 < \mathrm{E}\varepsilon_1^2 = \sigma^2 < \infty$; $\beta_1, \ldots, \beta_p \in \mathbb{R}^1$ $(\beta_p \neq 0)$ - это неизвестные коэффициенты авторегрессии.

Иногда удобно рассматривать $AR(p)$-уравнение для $t = 1, 2, \ldots$ при начальных условиях $u_{1-p}, \ldots, u_n$.

$ARCH(p)$-уравнение имеет вид:

$$u_t = \sigma_t \varepsilon_t,\ \text{где } \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2,\ t \in \mathbb{Z}$$

Здесь $\alpha_0 > 0,\ \alpha_k \geq 0,\ \alpha_p > 0,\ \{\varepsilon_t\}$ - н.о.р., $\mathrm{E}\varepsilon_1 = 0, \mathrm{E}\varepsilon_1^2 = 1$

4.2 Метод максимального правдоподобия и метод наименьших квадратов в авторегрессии.

$AR(1)$-модель.

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t = 1, 2, \ldots;\ u_0 = 0,\ \beta \in \mathbb{R}^1,\ \{\varepsilon_t\} \text{ - н.о.р. сл.в, } \mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 < \infty \qquad (1)$$

Тогда

$$u_t = \beta(\beta u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2 u_{t-2} = \ldots = \varepsilon_t + \beta\varepsilon_{t-1} + \ldots + \beta^{t-1}\varepsilon_1$$

1. Стационарный случай $|\beta| < 1$.

$$u_t \xrightarrow{\text{с.к.}} u_t^0 \stackrel{\text{def}}{=} \sum_{j \geq 0} \beta^j \varepsilon_{t-j}$$

и ряд среднеквадратично сходится, так как

$$\mathrm{E}(u_t - u_t^\circ)^2 = \mathrm{E}(\sum_{j \geq t} \beta^j \varepsilon_{t-j})^2 = \mathrm{E}\varepsilon_1^2 \sum_{j \geq t} \beta^{2j} = \underline{\underline{\mathcal{O}}}(\beta^{2t}) = \overline{\overline{o}}(1),\ t \to \infty$$

2. Критический случай (неустойчивая авторегрессия) $|\beta| = 1$

3. <u>Взрывающаяся авторегрессия $|\beta| > 1$</u>

$$\mathrm{D}u_t = \mathrm{D}\sum_{j=1}^{t-1} \beta^j \varepsilon_{t-j} = \mathrm{E}\varepsilon_1^2 \sum_{j=0}^{t-1} \beta^{2j} = \frac{\mathrm{E}\varepsilon_1^2 (1-\beta^{2t})}{1-\beta^2} = \underline{\underline{\mathcal{O}}}(\beta^{2t}) \to \infty,\ t \to \infty \text{ эксп. быстро}$$

Мы знаем: оптимальный среднеквадратичный прогноз u_{n+1} по $u_1, \ldots, u_n$ есть $\widetilde{u}_{n+1} = \beta u_n$. <u>Надо уметь оценивать β!</u>

Пусть $\varepsilon_1 \sim g(x)$ это плотность вероятности по мере Лебега. Положим

$$\mathcal{E} \overset{\text{def}}{=} (\varepsilon_1, \ldots, \varepsilon_n)^T,\ U = (u_1, \ldots, u_n)^T,\ B = \begin{pmatrix} 1 & & & 0 \\ -\beta & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -\beta & 1 \end{pmatrix}$$

Тогда из (1)

$$\mathcal{E} = BU \Rightarrow U = B^{-1}\mathcal{E} \tag{2}$$

Плотность вероятности вектора $\mathcal{E}$ есть $g_{\mathcal{E}}(x_1, \ldots, x_n) = \prod_{i=1}^n g(x_i)$. Тогда пл.в. вектора U есть в силу (2)

$$g_n(y, \beta) = \frac{1}{|\det(B^{-1})|} g_{\mathcal{E}}(By) = \left| By = \begin{pmatrix} y_1 - \beta * 0 \\ y_2 - \beta y_1 \\ \vdots \\ y_n - \beta y_{n-1} \end{pmatrix} \right| = \prod_{t=1}^n g(y_t - \beta y_{t-1}),\ \text{где } y = (y_1, \ldots, y_n)$$

О.м.п. для β - решение задачи

$$\ln g_U(U, \theta) = \sum_{t=1}^n \ln g(u_t - \theta u_{t-1}) \to \max_{\theta \in \mathbb{R}^1} \tag{3}$$

Для гладкой g уравнение максимального правдоподобия

$$\sum_{t=1}^n u_{t-1} \frac{g'(u_t - \theta u_{t-1})}{g(u_t - \theta u_{t-1})} = 0 \tag{4}$$

Пример $(\varepsilon_1 \sim N(0, \sigma^2))$**.** *Тогда* $g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$ *и задача* (3) *имеет вид*

$$\sum_{t=1}^n \ln \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(u_t - \theta u_{t-1})^2}{2\sigma^2}\right\} \to \max_{\theta \in \mathbb{R}^1}$$

Последняя задача эквивалентна следующей:

$$\sum_{t=1}^n (u_t - \theta u_{t-1})^2 \to \min_{\theta \in \mathbb{R}^1} \tag{5}$$

Решение задачи (5) *- о.м.п.*

$$\widehat{\beta}_{n,ML} = \frac{\sum_{t=1}^n u_{t-1} u_t}{\sum_{t=1}^n u_{t-1}^2} \tag{6}$$

Если мы не предполагаем гауссовость ε_1, то решение задачи (5) есть о.н.к.

$$\widehat{\beta}_{n,LS} = \frac{\sum_{t=1}^{n} u_{t-1}u_t}{\sum_{t=1}^{n} u_{t-1}^2} \tag{7}$$

Оценка $\widehat{\beta}_{n,ML}$ - параметрическая, а $\widehat{\beta}_{n,LS}$ - непараметрическая.

Пример $(\varepsilon_1 \sim Lap(\lambda))$**.** *Тогда $g(x) = \frac{\lambda}{2}\exp\{-\lambda\,|x|\}$, $\lambda > 0$. Задача (5) имеет вид:*

$$\sum_{t=1}^{n} \ln \frac{\lambda}{2} \exp\{-\lambda\,|u_t - \theta u_{t-1}|\} \to \max_{\theta\in\mathbb{R}^1}$$

что эквивалентно задаче

$$\sum_{t=1}^{n} |u_t - \theta u_{t-1}| \to \min_{\theta\in\mathbb{R}^1} \tag{8}$$

Решение (8) - о.м.п. $\widehat{\beta}_{n,ML}$. Если распредение ε_1 неизвестно, то решение (8) - о.н.к. $\widehat{\beta}_{n,LS}$.

Оценка $\widehat{\beta}_{n,LS}$ не выписывается явно!

4.3 Теорема о предельном распределении о.м.п. в $AR(1)$ при гауссовских инновациях.

Рассмотрим случай гауссовских $\{\varepsilon_t\}$, $\varepsilon_1 \sim N(0,1)$. Пусть

$$d_n^2(\beta) \stackrel{\text{def}}{=} \begin{cases} \frac{n}{1-\beta^2}, & |\beta| < 1 \\ \frac{n^2}{2}, & |\beta| = 1 \\ \frac{\beta^{2n}}{(\beta^2-1)^2}, & |\beta| > 1 \end{cases}$$

Покажем, что $d_n^2(\beta) \sim J_n(\beta)$, $n \to \infty$, где $J_n(\beta)$ - информации Фишера о параметре β, содержащаяся в $u_1, \dots, u_n$. Действительно, если $U = (u_1, \dots, u_n)$, $y = (y_1, \dots, y_n)$, то пл. вер.

$$g_U(y,\beta) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}(y_t - \beta y_{t-1})^2\right\}$$

и поэтому

$$J_n(\beta) = \mathrm{E}_\beta \left(\frac{\partial}{\partial\beta}\ln g_U(U,\beta)\right)^2 = \mathrm{E}_\beta\left(\frac{\partial}{\partial\beta}\left(-\frac{1}{2}\sum_{t=1}^{n}(u_t - \beta u_{t-1})^2\right)\right) =$$

$$= \mathrm{E}_\beta\left(\sum_{t=1}^{n} u_{t-1}(u_t - \beta u_{t-1})\right)^2 = \mathrm{E}_\beta\left(\sum_{t=1}^{n} u_{t-1}\varepsilon_t\right)^2 = \sum_{t=1}^{n}\mathrm{E}_\beta u_{t-1}^2 = \sum_{t=1}^{n-1}\mathrm{E}_\beta u_t^2$$

Но $u_t = \sum_{j=0}^{t-1}\beta^j\varepsilon_{y-j}$, и

$$\mathrm{E}u_t^2 = \mathrm{E}(\sum_{j=0}^{t-1}\beta^j\varepsilon_{t-j})^2 = \sum_{j=1}^{t-1}\beta^{2j} = \begin{cases} \frac{1-\beta^{2t}}{1-\beta^2}, & |\beta| \neq 1 \\ t, & |\beta| = 1 \end{cases}$$

Значит,

$$J_N(\beta) = \begin{cases} \frac{u-1}{1-\beta^2} - \frac{\beta^2(1-\beta^{2(n-1)})}{(1-\beta^2)^2}, & |\beta| \neq 1 \\ \frac{(n-1)(1+(n-1))}{2}, & |\beta| = 1 \end{cases}$$

Отсюда

$$J_n(\beta) \sim \begin{cases} \frac{u}{1-\beta^2}, & |\beta| < 1 \\ \frac{u^2}{2}, & |\beta| = 1 \\ \frac{\beta^{2n}}{(\beta^2-1)^2}, & |\beta| > 1 \end{cases} = d_n^2(\beta)$$

Распределение Коши с параметрами $(0,1)$ обозначается $K(0,1)$, $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
Пусть $W(s),\ s \in [0,1]$ - стандартный винеровский процесс. Обозначим $H(\beta),\ |\beta| = 1$, распределение сл. в. β

$$H(\beta) = \frac{W^2(1) - 1}{2^{3/2}\int_0^1 W^2(s)ds}$$

Теорема 1

Пусть $\{\varepsilon_t\}$-н.о.р. сл.в., $\varepsilon_1 \sim N(0,1)$. Тогда

$$d_n(\beta)(\widehat{\beta}_{n,ML} - \beta) \xrightarrow[n\to\infty]{d} \begin{cases} N(0,1), & |\beta| < 1 \\ H(0,1), & |\beta| = 1 \\ K(0,1), & |\beta| > 1 \end{cases}$$

Доказательство.

$$\widehat{\beta}_{n,ML} = \frac{\sum_{t=1}^n u_{t-1}u_t}{\sum_{t=1}^n u_{t-1}^2} = \frac{\sum_{t=1}^n u_{t-1}(\beta u_{t-1} + \varepsilon_t)}{\sum_{t=1}^n u_{t-1}^2} = \beta + \frac{\sum_{t=1}^n \varepsilon_t u_{t-1}}{\sum_{t=1}^n u_{t-1}^2}$$

Положим для краткости

$$M_n \overset{\text{def}}{=} d_n^{-1}(\beta)\sum_{t=1}^n \varepsilon_t u_{t-1},\ V_n \overset{\text{def}}{=} d_n^{-2}(\beta)\sum_{t=1}^n u_{t-1}^2$$

Тогда

$$d_n(\beta)(\widehat{\beta}_{n,ML} - \beta) = \frac{M_n}{V_n}$$

Пусть $f_n(t,s)$ - совместная характеристическая функция M_n и V_n. Тогда (см. [RAO M.M. Statist, 1978, V.6, pp. 185-190])

$$f_n(t,s) \to f(t,s) = \begin{cases} \exp\left\{is - \frac{t^2}{2}\right\}, & |\beta| < 1 \\ (1 + t^2 - 2is)^{-1/2}, & |\beta| > 1 \end{cases} \tag{9}$$

1. $\underline{|\beta| < 1}$. Тогда $f(t,s)$ есть характеристическая функция вектора $(\xi,1)^T$, где $\xi \sim N(0,1)$. Действительно,

$$\phi(t,s) = \mathrm{E}\exp\{i(t\xi + s)\} = e^{is}\phi_\xi(t) = \exp\left\{is - \frac{t^2}{2}\right\}$$

Теорема: О наследовании сходимости

Пусть случайный вектор $S_n \xrightarrow[n\to\infty]{d} S,\ S_n, S \in \mathbb{R}^k,\ H : \mathbb{R}^k \to \mathbb{R}^1$ - борелевская функция, непрерывная на множестве A таком, что $\mathrm{P}(S \in A) = 1$. Тогда $H(S_n) \xrightarrow[n\to\infty]{d} H(S)$

В силу (9) $(M_n, V_n)^T \xrightarrow{d} (\xi, 1)^T$. Если $H(x, y) = \frac{x}{y}$, то $H(x, y)$ непрерывна при $y > 0$. Можно взять $A = \{y : y > 0\},\ \mathrm{P}((\xi, 1)^T \in A) = 1$. В силу теоремы о наследовании слабой сходимости

$$d_n(\beta)(\widehat{\beta}_{n,ML} - \beta) = \frac{M_n}{V_n} = H(M_n, V_n) \xrightarrow{d} H(\xi, 1) = \xi$$

2. $\underline{|\beta| > 1}$. Тогда $f(t, s)$ есть хар. функция вектора $(\xi\eta, \eta^2)^T$, где $\xi, \eta \sim N(0, 1)$, ξ и η независимы. Действительно,

$$\mathrm{E}\exp\left\{it(\xi\eta) + is\eta^2\right\} = \mathrm{E}\mathrm{E}\left(\exp\left\{\frac{it(\xi\eta) + is\eta^2}{2}\right\}\right) = \mathrm{E}\exp\left\{is\eta^2\right\}\mathrm{E}\left(\exp\left\{\frac{i(t\xi)\eta}{2}\right\}\right) =$$

$$= \mathrm{E}\exp\left\{is\eta^2\right\}\exp\left\{\frac{-t^2\eta^2}{2}\right\} = \mathrm{E}\exp\left\{i\left(s + \frac{it^2}{2}\right)\eta^2\right\} = \left|\mathrm{E}\exp\left\{ilx_1^2\right\} = (1 - 2il)^{-1/2}\right| =$$

$$= \left(1 - 2is + \frac{2t^2}{2}\right)^{-1/2} = (1 + t^2 - 2is)^{-1/2} = \phi(t, s)$$

Значит, $(M_n, V_n)^T \xrightarrow{d} (\xi\eta, \eta^2)^T$,

$$d_n(\beta)(\widehat{\beta}_{n,ML} - \beta) = \frac{M_n}{V_n} \xrightarrow{d} \frac{\xi\eta}{\eta^2} = \frac{\xi}{\eta} \sim K(0, 1)$$

3. $\underline{|\beta| = 1}$. Тогда

$$M_n = \frac{\sqrt{2}}{n}\sum_{t=1}^{n} \varepsilon_t u_{t-1},\ V_n = \frac{2}{n^2}\sum_{t=1}^{n} u_{t=1}^2$$

Далее, $u_t = u_{t-1} t\varepsilon_t = \varepsilon_1 + \ldots + \varepsilon_t$.

Введем киферовский последовательный процесс

$$w_n(s) \overset{\text{def}}{=} n^{-1/2}\sum_{i \le ns} \varepsilon_i,\ s \in [0, 1],\ w_n(s) = 0,\ 0 \le s \le 1/n$$

Тогда

$$n^{-1/2} u_{t-1} = w_n\left(\frac{t-1}{n}\right)$$

Пусть

$$\Delta w_n\left(\frac{t}{n}\right) \overset{\text{def}}{=} w_n\left(\frac{t}{n}\right) - w_n\left(\frac{t-1}{n}\right) = \frac{\varepsilon_t}{\sqrt{n}}$$

Тогда
$$M_n = \sqrt{2}\sum_{t=1}^{n} w_n\left(\frac{t-1}{n}\right)\Delta w_n\left(\frac{t}{n}\right),\ V_n = 2\sum_{t=1}^{n} w_n^2\left(\frac{t-1}{n}\right)\frac{1}{n}$$
Пусть
$$U_n \stackrel{\text{def}}{=} \left(w_n\left(\frac{1}{n}\right), w_n\left(\frac{2}{n}\right), \dots, w_n\left(\frac{n}{n}\right)\right)^T$$

Тогда $U_n = \left(\frac{\varepsilon_1}{\sqrt{n}}, \frac{\varepsilon_1+\varepsilon_2}{\sqrt{n}}, \dots, \frac{\varepsilon_1+\ldots+\varepsilon_n}{\sqrt{n}}\right)^T$ и это есть гауссовский вектор со средним ноль, $\mathrm{cov}\left(w_n\left(\frac{i}{n}\right), w_n\left(\frac{j}{n}\right)\right) = \frac{\min(i,j)}{n}$

Действительно,
$$U_n \begin{pmatrix} 1 & & 0 \\ \vdots & \ddots & \\ 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
Для $i \le j$
$$\mathrm{cov}\left(w_n\left(\frac{i}{n}\right), w_n\left(\frac{j}{n}\right)\right) = \mathrm{E}\left(\frac{1}{n}\sum_{t=1}^{i}\varepsilon_t \times \sum_{k=1}^{j}\varepsilon_k\right) = \frac{1}{n}\mathrm{E}\left(\sum_{t=1}^{i}\varepsilon_i\right)^2 = \frac{i}{n} = \frac{min(i,j)}{n}$$
Введем вектор $U \stackrel{\text{def}}{=} \left(w\left(\frac{1}{n}\right), w\left(\frac{2}{n}\right), \dots, w\left(\frac{n}{n}\right)\right)^T$, где $w(s)$ - стандартный винеровский. Это гауссовский вектор со средним 0, $\mathrm{cov}\left(w\left(\frac{i}{n}\right), w\left(\frac{j}{n}\right)\right) = \frac{\min(i,j)}{n}$. Значит,
$$U_n \stackrel{d}{=} U \Rightarrow \phi(U_n) \stackrel{d}{=} \phi(u),\ \text{где } \phi \text{ - } \forall \text{ бор.} \tag{10}$$

Действительно, пусть $\xi \stackrel{d}{=} \eta,\ \xi, \eta \in \mathbb{R}^k$. Тогда $f(\xi) \stackrel{d}{=} f(\eta)$, так как $\mathrm{P}(f(\xi) \in A) = \mathrm{P}(\xi \in f^{-1}(A)) = \mathrm{P}(\eta \in f^{-1}(A)) = \mathrm{P}(f(\eta) \in A)$

Пусть
$$\overline{M}_n = \sqrt{2}\sum_{t=1}^{n} w\left(\frac{t-1}{n}\right)\Delta w\left(\frac{t}{n}\right),\ \overline{V}_n = 2\sum_{t=1}^{n} w^2\left(\frac{t-1}{n}\right)\frac{1}{n}$$
Так как $M_n,\ V_n$ - борелевские функции от U_n, а $\overline{M}_n,\ \overline{V}_n$ - борелевские функции от U, то из (10) следует:
$$\frac{M_n}{V_n} \stackrel{d}{=} \frac{\overline{M}_n}{\overline{V}_n} \tag{11}$$
Но
$$\overline{M}_n \xrightarrow{\text{с.к.}} \sqrt{2}\int_0^1 w(s)dw(s),\ \overline{V}_n \xrightarrow{\text{с.к.}} 2\int_0^1 w^2(s)ds$$
Значит, $(\overline{M}_n, \overline{V}_n)^T \xrightarrow{d} \left(\sqrt{2}\int_0^1 w(s)dw(s), 2\int_0^1 w^2(s)ds\right)$, и, следовательно,
$$\frac{\overline{M}_n}{\overline{V}_n} \to \frac{\sqrt{2}\int_0^1 w(s)dw(s)}{2\int_0^1 w^2(s)ds} = \frac{w^2(1)-1}{2^{3/2}\int_0^1 w^2(s)ds} \tag{12}$$

Поскольку

$$d_n(\beta)(\widehat{\beta}_{n,ML} - \beta) = \frac{M_n}{V_n}$$

то соотношения (11)-(12) влекут утверждение теоремы.

□

4.4 Теорема о предельном распределении о.м.п. в $AR(1)$ при гауссовских инновациях при случайной нормировке.

Теорема 2

Пусть $\{\varepsilon_t\}$ - н.о.р. $N(0,1)$ сл.в. Тогда

$$\sqrt{\sum_{t=1}^{n} u_{t-1}^2}(\widehat{\beta}_{n,ML} - \beta) \xrightarrow{d} \begin{cases} N(0,1), & |\beta| \neq 1 \\ \widetilde{H}(\beta), & |\beta| = 1 \end{cases}$$

Здесь

$$\widetilde{H}(\beta) \text{ - распр. сл.в. } \frac{w^2(1) - 1}{2\sqrt{\int_0^1 w^2(s)ds}} = \frac{\int_0^1 w(s)dw(s)}{\sqrt{\int_0^1 w^2(s)ds}}$$

Доказательство.

$$\sqrt{\sum_{t=1}^{n} u_{t-1}^2}(\widehat{\beta}_{n,ML} - \beta) = \frac{M_n}{\sqrt{V_n}}$$

1. $\underline{|\beta| < 1}$: Тогда $(M_n, V_n)^T \xrightarrow{d} (\xi, 1)^T$, значит

$$\frac{M_n}{\sqrt{V_n}} \xrightarrow{d} \frac{\xi}{\sqrt{1}} \sim N(0,1)$$

2. $\underline{|\beta| > 1}$: Тогда $(M_n, V_n)^T \xrightarrow{d} (\xi\eta, \eta^2)^T$, значит

$$\frac{M_n}{\sqrt{V_n}} \xrightarrow{d} \frac{\xi\eta}{\sqrt{\eta^2}} = \xi \cdot \operatorname{sgn} \eta \sim N(0,1)$$

3. $\underline{|\beta| = 1}$: Тогда $(M_n, V_n)^T \xrightarrow{d} \left(\frac{1}{\sqrt{2}}(w^2(1) - 1), 2\int_0^1 w^2(s)ds\right)^T$, значит

$$\frac{M_n}{\sqrt{V_n}} \xrightarrow{d} \frac{w^2(1) - 1}{2\sqrt{\int_0^1 w^2(s)ds}}$$

□

4.5 Об оценке наименьших квадратов в авторегрессии

Если $\{\varepsilon_t\}$ - н.о.р. $N(0,1)$ сл.в. в $AR(1)$ ур-нии

$$u_t = \beta u_{t-1} + \varepsilon_t,\ u_0 = 0,\ t = 1, 2, \dots,\ \beta \in \mathbb{R}^1 \tag{13}$$

то о.м.п. - решение задачи

$$\sum_{t=1}^{n} (u_t - \theta u_{t-1})^2 \to \min_{\theta \in \mathbb{R}^1} \tag{14}$$

Если же $\{\varepsilon_t\}$ - н.о.р. сл.в. с неизвестным распределением, то задача (14) определяет о.н.к.

$$\widehat{\beta}_{n,LS} = \frac{\sum_{t=1}^{n} u_{t-1} u_t}{\sum_{t=1}^{n} u_{t-1}^2}$$

О.н.к. $\widehat{\beta}_{n,LS}$ - непараметрическая!

Теорема 3

Пусть $u_t = \beta u_{t-1} + \varepsilon_t,\ |\beta| < 1,\ t \in \mathbb{Z}$. Если $\{\varepsilon_t\}$ - н.о.р., $\mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 < \infty$, то

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) \underset{n\to\infty}{\xrightarrow{d}} N(0, 1 - \beta^2)$$

Замечание. *1. Если $|\beta| = 1$, то при $\mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 < \infty$,*

$$d_n(\beta)(\widehat{\beta}_{n,LS} - \beta) \xrightarrow{d} H(\beta)$$

2. Если $|\beta| > 1$, то в усл. п. (1)

$$d_n(\beta)(\widehat{\beta}_{n,LS} - \beta) \xrightarrow{d} \frac{\sum_{j\geq 1} \beta^{-j}\varepsilon_j}{\sum_{j\geq 1} \beta^{-j}\varepsilon'_j},\ \{\varepsilon_t\},\ \{\varepsilon'_t\} \text{ - \textit{нез. посл. с н.о.р. комп.}}$$

4.6 Теорема об $AR(1)$ модели с $|\beta| < 1$ - существование и свойства стационарного решения.

Рассмотрим стационарное $AR(1)$ уравнение

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},\ |\beta| < 1,\ \{\varepsilon_t\} \text{ - н.о.р.},\ \mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 = \sigma^2 < \infty \tag{15}$$

Опр. 1. *Любая последовательность $\{u_t\}$, для которой в (15) левая часть равно правой почти наверное, называется* ***решением уравнения*** (15)

Теорема 4

При $|\beta| < 1$ существует п.н. единственное строго стационарное решение уравнения (15). Оно имеет вид:

$$u_t = \sum_{j\geq 0} \beta^j \varepsilon_{t-j},\ \text{ряд с.к. сходится.} \tag{16}$$

Решение (16) является также стационарным в широком смысле, причем

$$\mathrm{E}u_t = 0,\ R(\tau) = \operatorname{cov}(u_t, u_{t+\tau}) = \frac{\sigma^2 \beta^{|\tau|}}{1-\beta^2}$$

Доказательство.

1. Сущестовавание предела

   Пусть $u_t^{(n)} \overset{\text{def}}{=} \sum_{j=0}^n \beta^j \varepsilon_{t-j}$ - частная сумма ряда (16). Ряд с.к. сходится, если для некоторой случайной величины S_t, $\mathrm{E}S_t^2 < \infty$, существует с.к. предел

   $$\lim_{n\to\infty} u_t^{(n)} = S_t,\ S_t - \text{ есть сумма ряда}$$

   То есть $\mathrm{E}\left|u_t^{(n)} - S_t\right|^2 \xrightarrow[n\to\infty]{} 0$. Известно (по критерию Коши), что эта с.к. сходимость с.к. фундаментальности, то есть

   $$\lim_{n,m\to\infty} \mathrm{E}\left|u_t^{(n)} - u_t^{(m)}\right|^2 = 0$$

   Пусть для краткости $l \overset{\text{def}}{=} \min(m,n)$, $k \overset{\text{def}}{=} \max(m,n)$. Тогда

   $$\mathrm{E}\left|u_t^{(n)} - u_t^{(m)}\right|^2 = \mathrm{E}\left|\sum_{j=l+1}^{k} \beta^j \varepsilon_{t-j}\right|^2 = \sigma^2 \sum_{j=l+1}^{k} \beta^{2j} \to 0,\ \text{т.к. } l,k \to \infty,\ |\beta| < 1$$

   Значит, что ряд (16) с.к. сходится. Имеем п.н.

   $$u_t = \sum_{j\geq 0} \beta^j \varepsilon_{t-j} = \varepsilon_t + \beta \sum_{j\geq 1} \beta^j \varepsilon_{t-j} = \varepsilon_t + \beta \sum_{s\geq 0} \beta^s \varepsilon_{t-s-1} = \varepsilon_t + \beta u_{t-1}$$

   Значит, $\{u_t\}$ из (16) есть решение (15).

2. Строгая стационарность

   Пусть $U(\tau) = (u_{t_1+\tau}, \dots, u_{t_k+\tau})$. Надо показать, что $U(\tau) \overset{d}{=} U(0)$.
   Пусть $U_n(\tau) \overset{\text{def}}{=} (u_{t_1+\tau}^{(n)}, \dots, u_{t_k+\tau}^{(n)})$

Задача. *Если* $\{\xi_t\}$ *- строго стац. посл., а* $\eta_t = f(\xi_t, \xi_{t-1}, \ldots, \xi_{t-k})$, *(f - бор.), то* $\{\eta_t\}$ *- строго стац. посл.*

В силу этой задачи $\{u_t^{(n)}\}$ - строго стационарна, то есть распределение вектора $U_n(\tau)$ вовсе не зависит от τ, но

$$U_n(\tau) \xrightarrow[n\to\infty]{d} U(\tau), \text{ т.к. } u_t^{(n)} \xrightarrow{\text{с.к.}} u_t \tag{17}$$

Значит, в силу (17), распределение $U(\tau)$ от τ не зависит.

3. Единственность
   Пусть $\{\widetilde{u}_t\}$ - любое строго стационарное решение (15). Тогда п.н.
   $$\widetilde{u}_t = \beta\widetilde{u}_{t-1} + \varepsilon_t = \underbrace{\varepsilon_t + \beta\varepsilon_{t-1} + \ldots + \beta^{k-1}\varepsilon_{t-k+1}}_{u_t^{(k)}} + \beta^k\widetilde{u}_{t-k}$$
   Имеем
   $$\mathrm{P}(\left|\beta^k\widetilde{u}_{t-k}\right| > \delta) = \mathrm{P}(\left|\beta^k\widetilde{u}_0\right| > \delta) \underset{|\beta|<1}{\to} 0, \ k\to\infty$$
   Знаем, что $u_t^{(k)} \xrightarrow{\text{с.к.}} u_t = \sum_{j\geq 0}\beta^j\varepsilon_{t-j}, \ \mathrm{E}u_t^2 < \infty$. Значит, $u_t^{(k)} + \beta^k\widetilde{u}_{t-k} \xrightarrow[k\to\infty]{\mathrm{P}} u_t$. Следовательно, п.н.
   $$\widetilde{u}_t = p\lim_{k\to\infty}(u_t^{(k)} + \beta^k\widetilde{u}_{t-k}) = u_t = \sum_{j\geq 0}\beta^j\varepsilon_{t-j}$$

4. Стационарность в широком смысле
   Последовательность $\{u_t\}$ из (16) стационарна в широком смысле так как она стационарная в узком смысле и есть момент до 2-го порядка. Тогда из (15)
   $$\mathrm{E}u_t = \beta\mathrm{E}u_{t-1} + \mathrm{E}\varepsilon_t; \ (1-\beta)\mathrm{E}u_0 = 0, \ \mathrm{E}u_0 = 0$$
   Для $\varepsilon > 0$ $\mathrm{E}u_{t+\tau}u_t = \beta\mathrm{E}u_{t+\tau-1}u_t + \mathrm{E}\varepsilon_{t+\tau}u_t$. Но $\mathrm{E}\varepsilon_{t+\tau}u_t = \mathrm{E}\varepsilon_{t+\tau}\mathrm{E}u_t = 0$, т.к. $\varepsilon_{t+\tau}$ и u_t нез.
   $$\mathrm{E}u_t^2 = \beta^2\mathrm{E}u_{t-1}^2 + \underbrace{2\beta\mathrm{E}(u_{t-1}\varepsilon_t)}_{0} + \mathrm{E}\varepsilon_t^2 \Rightarrow (1-\beta^2)\mathrm{E}\underbrace{u_0^2}_{R(0)} = \mathrm{E}\varepsilon_0^2 = \sigma^2 \Rightarrow R(0) = \frac{\sigma^2}{1-\beta^2}$$
   Получаем:
   $$R(\tau) = \beta R(\tau-1), \ R(0) = \frac{\sigma^2}{1-\beta^2} \Rightarrow R(\tau) = \frac{\sigma^2\beta^{|\tau|}}{1-\beta^2}, \ \forall\tau$$

□

4.7 Последовательности с сильным перемешиванием - определение и примеры. Формулировка предельных теорем - закона больших чисел и центральной предельной теоремы.

Опр. 2. *Пусть* $\{u_t\}, \ t\in\mathbb{Z}$, *строго стационарная последовательность. Если*

$$\alpha(\tau) \overset{def}{=} \sup_{\substack{A\in M_{-\infty}^0 \\ B\in M_\tau^\infty}} |\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)| \to 0, \ \tau\to\infty$$

То $\{u_t\}$ *удовлетворяет условию* ***сильного перемешивания с коэффициентом пер.*** $\alpha(\tau)$. *Здесь* $M_a^b = \sigma\{u_t, \ a\leq t\leq b\}$

Примеры. *1.* $\{\varepsilon_t\}$ *- н.о.р. сл.в. Здесь* $\alpha(\tau) = 0,\ \tau > 0$

2. $u_t = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \ldots + \alpha_q\varepsilon_{t-q}$, $\{\varepsilon_t\}$ *- н.о.р. Здесь* $\alpha(\tau) = 0,\ \tau > q$

3. $u_t = \beta_1 u_{t-1} + \ldots + \beta_q u_{t-p} + \varepsilon_t$, $\{\varepsilon_t\}$ *- н.о.р.,* ε_1 *имеет Лебегову пл. в.,* $\mathrm{E}\varepsilon_1 = 0, \mathrm{E}\varepsilon_1^2 < \infty$
Mokkadem (1988): стр. стац. реш. $\{u_t\}$ *удовлетворяет усл. с.п. с* $\alpha(\tau) \le c\lambda^\tau,\ 0 < \lambda < 1$

Задача. *Если* $\{u_t\}$ *удовлетворяет условию сильного перемешивания с коэффициентом* $\alpha(\tau)$*, а* $\eta_t = f(u_t, u_{t-1}, \ldots, u_{t-k})$*, то* $\{\eta_t\}$ *уд. усл. с.п. с* $\alpha_\eta(\tau) \le \alpha(\tau - k),\ \tau > k$

Теорема: Закон больших чисел (З.Б.Ч.)

Если $\{u_t\},\ t \in \mathbb{Z}$, строго стац. посл. с с.п., $\mathrm{E}\,|u_1| < \infty$, то

$$\frac{1}{n}\sum_{t=1}^{n} u_t \xrightarrow[n\to\infty]{\text{п.н.}} \mathrm{E}u_1$$

Теорема: Центральная предельная теорема (Ц.П.Т.)

Пусть $\{u_t\},\ t \in \mathbb{Z}$, стр. стац. посл. с с.п., $\mathrm{E}u_1 = 0,\ \mathrm{E}\,|u_1|^{2+\delta} < \infty$ при некотором $\delta > 0$. Пусть $\sum_{\tau\ge1}(\alpha(\tau))^{\frac{2}{2+\delta}} < \infty$. Тогда

1. Ряд $\Delta \stackrel{\text{def}}{=} \mathrm{E}u_0^2 + 2\sum_{\tau\ge1}\mathrm{E}u_0u_\tau$ сходится абсолютно
2. Если $\Delta > 0$, то $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \xrightarrow{d} N(0, \Delta^2)$

Следствие. *Если* $\{u_t\}$ *уд. с.п.,* $\mathrm{E}u_1 = m,\ \mathrm{E}\,|u_1 - m|^{2+\delta} < \infty,\ \sum_{\tau\ge1}(\alpha(\tau))^{\frac{2}{2+\delta}} < \infty,\ \overline{\Delta}^2 = \mathrm{D}u_t + 2\sum_{t\ge1}R(\tau)$*, то при* $\widetilde{\Delta} > 0$

$$\sup_x\left|\mathrm{P}(\sqrt{n}(\overline{u} - m) \le x) - \Phi\left(\frac{x}{\widetilde{\Delta}}\right)\right| \underset{n\to\infty}{\to} 0$$

4.8 Оценка наименьших квадратов в стационарной $AR(1)$ модели - теорема об асимптотической нормальности. Применения к доверительному оцениванию и проверке гипотез.

Мы рассматриваем $AR(1)$ модель.

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z} \tag{18}$$

в которой $\{\varepsilon_t\}$ - н.о.р. сл.в., $\mathrm{E}\varepsilon_1 = 0,\ 0 < \sigma^2 = \mathrm{E}\varepsilon_1^2 < \infty,\ |\beta| < 1$. Пусть функция распределения ε_1 есть $G(x),\ g(x) = G'(x)$, $G(x)$ и $g(x)$ неизвестны. Пусть наблюдения $u_0, \ldots, u_n$ - выборка из стационарного решения $AR(1)$ уравнения. Мы берем оценкой неизвестного параметра β о.н.к., которая получается решением задачи

$$\sum_{t=1}^{n} (u_t - \theta u_{t-1})^2 \to \min_{\theta}$$

Эту оценку обозначим $\widehat{\beta}_{n,LS}$, очевидно,

$$\widehat{\beta}_{n,LS} = \frac{\sum_{t=1}^{n} u_{t-1} u_t}{\sum_{t=1}^{n} u_{t-1}^2} \tag{18'}$$

Наша ближайшая цель - доказать теорему 3,

Вот что нам известно и что будем использовать

1. В силу доказанной теоремы 4 стационарное решение уравнения (18) при $|\beta| < 1,\ \mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 = \sigma^2 < \infty$ имеет вид

   $$u_t = \sum_{j \ge 0} \beta^j \varepsilon_{t-j},\ \text{ряд с.к. сход.} \tag{19}$$

   Последовательность $\{u_t\},\ t \in \mathbb{Z}$, и (19) является строго стационарной, а также стационарной в широком смысле

   $$\mathrm{E}u_t = 0,\ R(\tau) = \operatorname{cov}(u_t, u_{t+\tau}) = \frac{\sigma^2 \beta^{|\tau|}}{1 - \beta^2},\ \tau \in \mathbb{Z}$$

   Поскольку $u_{t-1} = u_{t-1}(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$, сл.в. ε_t не зависит от $\{u_{t-1}, u_{t-2}, \ldots\}$

2. Строго стационарная последовательность (19) удовлетворяет условию сильного перемешивания с коэффициентом перемешивания $\alpha(\tau) \le c\lambda^\tau,\ 0 < \lambda < 1$, если ε_1 имеет плотность вероятности $g(x)$.

3. З.Б.Ч. для посл. с с.п.

4. Ц.П.Т. для посл. с с.п.

5. Mokkadem (1988): Если $\{u_t\}$ удовлетворяет условию сильного перемешивания с коэффициентом $\alpha(\tau)$, а $\eta_t = f(u_t, u_{t-1}, \ldots, u_{t-k})$, то $\{\eta_t\}$ уд. усл. с.п. с $\alpha_\eta(\tau) \le \alpha(\tau - k),\ \tau > k$

Доказательство теоремы 3. Предположим дополнительно, что $\mathrm{E}\,|\varepsilon_1|^{2+\delta} < \infty$ при некотором $\delta > 0$. Пусть еще существует плотность вероятности $\varepsilon_1 \sim g(x)$ по мере Лебега.

1. Покажем, что ряд (19) сходится не только в L^2 (с.к.), но и в $L^{2+\delta}$, и, значит, $\mathrm{E}\,|u_1|^{2+\delta} < \infty$

   Справедливо неравенство Минковского: если $\mathrm{E}\,|\xi|^{2+\delta} < \infty,\ \mathrm{E}\,|\eta|^{2+\delta} < \infty$ при $\delta > 0$, то

$$\{\mathrm{E}\,|\xi+\eta|^{2+\delta}\}^{\frac{1}{2+\delta}} \le \{\mathrm{E}\,|\xi|^{2+\delta}\}^{\frac{1}{2+\delta}} + \{\mathrm{E}\,|\eta|^{2+\delta}\}^{\frac{1}{2+\delta}}$$

   Рассмотрим частную сумму $S_n \stackrel{\text{def}}{=} \sum_{j=0}^{n} \beta^j \varepsilon_{t-j},\ l \stackrel{\text{def}}{=} \min(m,n),\ k \stackrel{\text{def}}{=} \max(m,n),$

$$\{\mathrm{E}\,|S_n - S_m|^{2+\delta}\}^{\frac{1}{2+\delta}} = \left\{\mathrm{E}\left|\sum_{j=l+1}^{k} \beta^j \varepsilon_{t-j}\right|^{2+\delta}\right\}^{\frac{1}{2+\delta}} \le$$

$$\le \sum_{j=l+1}^{k} \left\{\mathrm{E}\left|\beta^j \varepsilon_{t-j}\right|^{2+\delta}\right\}^{\frac{1}{2+\delta}} = \mathrm{E}\left\{|\varepsilon_1|^{2+\delta}\right\}^{\frac{1}{2+\delta}} \sum_{j=l+1}^{k} |\beta|^j \xrightarrow[|\beta|<1,\ l,k\to\infty]{} 0$$

   Значит, последовательность $\{S_n\}$ частных сумм фундаментальна, и ряд $u_t = \sum_{j\ge 0} \beta^j \varepsilon_{t-j}$ сходится в $L^{2+\delta}$, $\mathrm{E}\,|u_1|^{2+\delta} < \infty$

2.

$$\widehat{\beta}_{n,LS} = \frac{\sum_{t=1}^{n} u_{t-1} u_t}{\sum_{t=1}^{n} u_{t-1}^2} \stackrel{\text{п.н.}}{=} \beta + \frac{\sum_{t=1}^{n} u_{t-1}\varepsilon_t}{\sum_{t=1}^{n} u_{t-1}^2}$$

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) = \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{t-1}\varepsilon_t}{\frac{1}{n} \sum_{t=1}^{n} u_{t-1}^2}$$

3. В силу результатов Mokkadem (1988) посл. $\{u_t\}$ удовлетворяет условию сильного перемешивания с коэффициентом $\alpha(\tau) \le c\lambda^\tau,\ 0 < \lambda < 1$. Последовательность $\{\varepsilon_t u_{t-1} = (u_t - \beta u_{t-1})u_{t-1}\}$ тоже удовлетворяет условию сильного перемешивания с эксп. убывающим коэфф. $\alpha'(\tau) \le c'\lambda^\tau$,

$$\sum_{\tau\ge 1} (\alpha'(\tau))^{\frac{\delta}{2+\delta}} \le \sum_{\tau \ge 1} (c'\lambda^\tau)^{\frac{\delta}{2+\delta}} = \frac{(c'\lambda)^{\frac{\delta}{2+\delta}}}{1-\lambda^{\frac{\delta}{2+\delta}}} < \infty$$

$$\mathrm{E}\varepsilon_t u_{t-1} = \mathrm{E}\varepsilon_t \mathrm{E}u_{t-1} = 0;\ \mathrm{E}\,|\varepsilon_t u_{t-1}|^{2+\delta} = \mathrm{E}\,|\varepsilon_1|^{2+\delta}\,\mathrm{E}\,|u_1|^{2+\delta} < \infty$$

   В силу Ц.П.Т. для посл. с с.п.

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t u_{t-1} \xrightarrow{d} N(0,\Delta^2),\ \text{где } \Delta^2 = \mathrm{E}(\varepsilon_1 u_0)^2 + \underbrace{2\sum_{\tau\ge 1} \mathrm{E}(\varepsilon_1 u_0 \varepsilon_{1+\tau} u_\tau)}_{0} = \mathrm{E}\varepsilon_1^2 \mathrm{E}u_0^2$$

4.

$$\frac{1}{n}\sum_{t=1}^{n} u_{t-1}^2 \xrightarrow{\text{п.н.}} \mathrm{E}u_0^2 \text{ в силу З.Б.Ч. для посл с с.п.}$$

5. Значит
$$\sqrt{n}(\widehat{\beta}_{n,LS}-\beta)\xrightarrow{d}\frac{1}{\mathrm{E}u_0^2}N(0,\mathrm{E}\varepsilon_1^2\mathrm{E}u_0^2)=N\left(0,\frac{\mathrm{E}\varepsilon_1^2\mathrm{E}u_0^2}{(\mathrm{E}u_0^2)^2}\right)=N\left(0,\frac{\mathrm{E}\varepsilon_1^2}{\mathrm{E}u_0^2}\right)=N(0,1-\beta^2)$$

□

Вот два важных вопроса:

1. Как построить непараметрические оценки, асимптотически гауссовские, и с меньшей ас. дисперсией, чем у о.н.к?
2. Будет ли о.н.к. $\widehat{\beta}_{n,LS}$ B-робастна?

Асимптотические доверительные интервалы

В силу (18')
$$\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta)}{\sqrt{1-\widehat{\beta}_{n,LS}^2}}=\underbrace{\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta)}{\sqrt{1-\beta^2}}}_{\xrightarrow{d}N(0,1)}\underbrace{\frac{\sqrt{1-\beta^2}}{\sqrt{1-\widehat{\beta}_{n,LS}^2}}}_{\xrightarrow{\mathrm{P}}1}\xrightarrow[n\to\infty]{d}N(0,1)\text{ по лемме Слуцкого}$$

Пусть $\xi_{1-\alpha/2}$ - квантиль уровня $1-\alpha/2$ ф.р. $\Phi(x)\sim N(0,1)$. Тогда
$$\mathrm{P}\left(\left|\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta)}{\sqrt{1-\widehat{\beta}_{n,LS}^2}}\right|<\xi_{1-\alpha/2}\right)\xrightarrow[n\to\infty]{}1-\alpha$$

То есть при больших n примерно с вероятностью $1-\alpha$
$$\widehat{\beta}_{n,LS}-\frac{\sqrt{1-\widehat{\beta}_{n,LS}^2}}{\sqrt{n}}\xi_{1-\alpha/2}<\beta<\widehat{\beta}_{n,LS}+\frac{\sqrt{1-\widehat{\beta}_{n,LS}^2}}{\sqrt{n}}\xi_{1-\alpha/2}$$

Получили доверительный интервал для β ас. уровня $1-\alpha$.

Проверка гипотез

Проверим гипотезу $H_0:\beta=\beta_0$ против альтернатив $H_1:\beta\neq\beta_0$. Критическое множество (критерий):
$$S_\alpha=\left\{u_0,\dots,u_n:\left|\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta_0)}{\sqrt{1-\beta_0^2}}\right|>\xi_{1-\alpha/2}\right\}$$

. Тогда, очевидно, $\mathrm{P}(H_1|H_0)\to\alpha$, а так как при H_1:
$$\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta_0)}{\sqrt{1-\beta_0^2}}=\frac{\sqrt{n}(\widehat{\beta}_{n,LS}-\beta)}{\sqrt{1-\beta_0^2}}+\frac{\sqrt{n}(\beta-\beta_0)}{\sqrt{1-\beta_0^2}}\xrightarrow[n\to\infty]{\mathrm{P}}\infty$$

то $\mathrm{P}(H_0|H_1)\to 0$. Значит
$$\begin{cases}\mathrm{P}(H_0|H_0)\to 1-\alpha\\ \mathrm{P}(H_1|H_1)\to 1\end{cases}\qquad n\to\infty$$

Вероятность принять правильную гипотезу близка к единице!

4.9 Функционал влияния о.н.к. в стационарной модели $AR(1)$ (двумя способами), неробастность о.н.к. на естественном классе засорений.

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},\ \{\varepsilon_t\} \text{ - н.о.р, } \mathrm{E}\varepsilon_1 = 0,\ 0 < \sigma^2 = \mathrm{E}\varepsilon_1^2 < \infty,\ |\beta| < 1$$

Пусть наблюдаются

$$y_t = u_t + z_t^\gamma \xi_t,\ t = 0, 1, \dots, n;$$

$\{u_t\}$ - стац.;
$\{z_t^\gamma\}$ - н.о.р., $z_1^\gamma \sim Br(\gamma)$ и $0 \le \gamma \le 1$;
$\{\xi_t\}$ - н.о.р., $\xi_1 \sim \mu_\xi,\ \mu_x\xi \in M_2$, т.е. $\mathrm{E}\xi_1^2 < \infty$;
Последовательности $\{u_t\}$, $\{z_t^\gamma\}$, $\{\xi_t\}$ независимы между собой. Пусть

$$\widehat{\beta}_{n,LS}^y = \frac{\sum_{t=1}^n y_{t-1}y_t}{\sum_{t=1}^n} y_{t-1}^2 \text{ - о.н.к., построенная по засоренным данным } \{y_t\}$$

Найдем ее функционал влияния.

Первый способ.

Предположим дополнительно, что у ε_1 существует плотность вероятности $g(x) = G'(x)$. Тогда последовательность $\{u_t\}$ удовлетворяет условиям с.п., а так как $\{z_t^\gamma \xi_t\},\ t \in \mathbb{Z}$, - посл. н.о.р. сл.в., которые не зависят от $\{u_t\}$, то $\{y_t\}$ - строго стационрная последовательность с с.п. Кроме того,

$$\mathrm{E}\,|y_1| < \infty,\ \text{т.к. } \mathrm{E}y_1^2 = \mathrm{E}(u_1 + z_1^\gamma\xi_1)^2 = \mathrm{E}u_1^2 + 2\underbrace{\mathrm{E}u_1}_{0}\mathrm{E}z_1^\gamma \mathrm{E}\xi_1 + \mathrm{E}(z_1^\gamma\xi_1)^2 = \mathrm{E}u_1^2 + \gamma\mathrm{E}\xi_1^2 < \infty$$

Значит, в силу З.Б.Ч. для посл. с с.п. имеем:

$$\widehat{\beta}_{n,LS}^y = \frac{\frac{1}{n}\sum_{t=1}^n y_{t-1}y_t}{\frac{1}{n}\sum_{t=1}^n y_t^2} \xrightarrow{\mathrm{P}} \theta_\gamma^{LS} = \frac{\mathrm{E}y_0y_1}{\mathrm{E}y_0^2} = \frac{\mathrm{E}(u_0 + z_0^\gamma\xi_0)(u_1 + z_1^\gamma\xi_1)}{\mathrm{E}(u_1 + z_1^\gamma\xi_1)^2} = \frac{\mathrm{E}u_0u_1 + \gamma^2(\mathrm{E}\xi_0)^2}{\mathrm{E}u_0^2 + \gamma\mathrm{E}\xi_0^2}$$

Отсюда

$$IF(\theta_\gamma^{LS}, \mu_\xi) = \left.\frac{d\theta_\gamma^{LS}}{d\gamma}\right|_{\gamma=0} = -\beta(1-\beta^2)\frac{\mathrm{E}\xi_0^2}{\mathrm{E}\varepsilon_1^2}$$

Если M_2 - множество распределений с конечным 2-ым моментом, то при $\beta \ne 0$

$$GES(\theta_\gamma^{LS}, M_2) = \sup_{\mu_\xi \in M_2}\left|IF(\theta_\gamma^{LS}, \mu_\xi)\right| = \infty$$

О.н.к. неробастна!

Второй способ.

Предположим опять, что ε_1 имеет плотность вероятности (Лебегову) $g(x)$. Тогда, как говорилось, последовательность $\{y_t\}$ удовлетворяет усл. с.п.

Оценка $\widehat{\beta}_{n,LS}^y$ - корень уравнения

$$l_{n,LS}(\theta) = \frac{1}{n}\sum_{t=1}^n y_{t-1}(y_t - \theta y_{t-1}) = 0$$

1.
$$l_{n,LS}(\theta)=\frac{1}{n}\sum_{t=1}^{n}y_{t-1}(y_t-\theta y_{t-1})\xrightarrow{\text{P}}\text{E}y_0(y_1-\theta y_0),\ \theta\in\mathbb{R},\ 0\leq\gamma\leq 1$$

То есть $\Lambda_{LS}(\gamma,\theta)=\text{E}y_0(y_1-\theta y_0)$. Пусть

$$\begin{array}{ll} H_{00}=(z_0^\gamma=0,z_1^\gamma=0); & H_{01}=(z_0^\gamma=0,z_1^\gamma=1);\\ H_{10}=(z_0^\gamma=1,z_1^\gamma=0); & H_{11}=(z_0^\gamma=1,z_1^\gamma=1); \end{array}$$

Тогда

$$\Lambda_{LS}(\gamma,\theta)=\sum_{i,j=0}^{1}\text{E}(y_0(y_1-\theta y_0)|H_{ij})\text{P}(H_{ij})=(1-\gamma)^2\text{E}u_0(u_1-\theta u_0)+$$

$$+\gamma(1-\gamma)\text{E}(u_0+\xi_0)(u_1-\theta u_0-\theta\xi_0)+\gamma^2\text{E}(u_0+\xi_0)(u_1+\xi_1-\theta u_0-\theta\xi_1)$$

Значит, функция $\Lambda_{LS}(\gamma,\theta)$ определена при всех γ и θ.

2.
$$\Lambda_{LS}(0,\beta)=\text{E}u_0(u_1-\beta u_0)=\text{E}u_0\xi_1=0$$

3. $\frac{\partial\Lambda_{LS}(\gamma,\theta)}{\partial\gamma}$, $\frac{\partial\Lambda_{LS}(\gamma,\theta)}{\partial\theta}$ существуют и непрерывны по паре (γ,θ) при $\gamma,\theta\in\mathbb{R}^1$

$$\frac{\partial\Lambda_{LS}(\gamma,\theta)}{\partial\gamma}=-\beta\text{E}\xi_0^2,\ \frac{\partial\Lambda_{LS}(\gamma,\theta)}{\partial\theta}=-\text{E}u_0^2$$

4.
$$\lambda(\beta)=-\text{E}u_0^2=-\frac{\text{E}\varepsilon_1^2}{1-\beta^2}<0$$

Так как $\phi_t(J_n,\theta)$ непр., то

$$IF(\theta_\gamma^{LS},\mu_\xi)=-\left(\frac{\partial\Lambda_{LS}(0,\beta)}{\partial\theta}\right)^{-1}\frac{\partial\Lambda_{LS}(0,\beta)}{\partial\gamma}=-\frac{(-\beta\text{E}\xi_0^2)(1-\beta^2)}{-\text{E}\varepsilon_1^2}=-\beta(1-\beta^2)\frac{\text{E}\xi_0^2}{\text{E}\varepsilon_1^2}$$

Очевидно, что при $\beta\neq 0$

$$GES(\theta_\gamma^{LS},M_2)=\infty$$

то есть $\widehat{\beta}_{n,LS}^y$ не B - робастна!

Задача.

$$u_t=\beta u_{t-1}+\varepsilon_t,\ t\in\mathbb{Z},\ \{\varepsilon_t\}\text{ - \textit{н.о.р}},\ \text{E}\varepsilon_1=0,\ 0<\sigma^2=\text{E}\varepsilon_1^2<\infty,\ |\beta|<1,\ \beta\neq 0$$

$$y_t=u_t+z_t^\gamma\xi_t$$

Оценка $\widehat{\beta}_n$ - ищется как корень уравнения

$$\sum_{t=1}^{n}y_{t-2}(y_t-\theta y_{t-1})=0$$

1. *Будет ли оценка $\widehat{\beta}_n$ B-робастной?*
2. *Чему равен фунционал влияния 2-го порядка?*

4.10 $AR(p)$ модель - формулировка теоремы о виде стационарного решения. Доказательство для $p = 1$.

Авторегрессия порядка p ($AR(p)$ - модель) описывается стохастическим разностным уравнением

$$u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_p u_{t-p} + \varepsilon_t, \ t \in \mathbb{Z} \tag{20}$$

Здесь $\{\varepsilon_t\}$ - н.о.р. сл.в., $\mathrm{E}\varepsilon_1 = 0, \ 0 < \mathrm{E}\varepsilon_1^2 = \sigma^2 < \infty$; $\beta_1, \ldots, \beta_p$ - неизвестные коэффициенты авторегрессии, $\beta_j \in \mathbb{R}$

Опр. 3. *Уравнение*

$$x^p = \beta_1 x^{p-1} + \ldots + \beta_p \tag{21}$$

называется характеристическим уравнением, соответствующим уравнению (20).

Теорема 5

Пусть корни характеристического уравнения (21) по модулю меньше единицы. Тогда уравнение (20) имеет почти наверное единственное строго стационарное решение

$$u_t = \sum_{j \geq 0} \gamma_j \varepsilon_{t-j}, \text{ ряд с.к. сходится (в } L_2) \tag{22}$$

Коэффициенты $\{\gamma_j\}$ определяются рекуррентным соотношением

$$\gamma_j = \beta_1 \gamma_{j=1} + \ldots + \beta_p \gamma_{j-p}, \ j = 1, 2, \ldots;$$

$$\gamma_0 = 1, \ \gamma_j = 0, j < 0;$$

$$|\gamma_j| \leq c\lambda^j, \ 0 < \lambda < 1$$

Ряд (16) определяет также стационарную в широком смысле последовательность с нулевым средним.

В случае $p = 1$ утверждение Теоремы 5 совпадает с утверждением доказанной нами ранее Теоремы 4. Доказательство теоремы 5 при $p > 1$ идейно не отличается от доказательства теоремы 4, но технически громоздко. Мы его опускаем.

Замечание. *Поскольку из* (22) $u_t = u_t(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, *то сл.в.* ε_{t+1} *не зависит от множества сл.в.* $\{u_t, u_{t-1}, \ldots\}$.

4.11 Оптимальный с.к. прогноз в $AR(p)$.

Пусть наблюдения $u_1, u_2, \ldots, u_n$ будут выборкой из стационарного решения (22). Пусть $n \geq p$. Оптимальный с.к. прогноз ненаблюдаемой величины u_{n+1} по наблюдениям $u_1, u_2, \ldots, u_n$ есть решение задачи

$$\mathrm{E}\left|u_{n+1} - \phi(u_1, u_2, \ldots, u_n)\right|^2 \to \min_{\substack{\text{бор. } \phi: \\ \mathrm{E}\phi(u_1, u_2, \ldots, u_n) < \infty}}$$

Мы знаем, что решение этой задачи есть

$$u_{n+1}^* = \phi^*(u_1, u_2, \ldots, u_n) = \mathrm{E}(u_{n+1} | u_1, u_2, \ldots, u_n)$$

Имеем:

$$\mathrm{E}(u_{n+1}|u_1,u_2,\ldots,u_n) = \mathrm{E}\left(\sum_{j=1}^{p}\beta_j u_{n+1-j} + \varepsilon_{n+1}|u_1,u_2,\ldots,u_n\right) =$$

$$= \mathrm{E}\left(\sum_{j=1}^{p}\beta_j u_{n+1-j}|u_1,u_2,\ldots,u_n\right) + \mathrm{E}\left(\varepsilon_{n+1}|u_1,u_2,\ldots,u_n\right) = \beta_1 u_n + \beta_2 u_{n-1} + \ldots + \beta_p u_{n+1-p}$$

так как $\mathrm{E}\left(\varepsilon_{n+1}|u_1,u_2,\ldots,u_n\right) = \mathrm{E}\varepsilon_{n+1} = 0$ в силу Замечания. Итак, опт. с.к. прогноз

$$\boxed{u^*_{n+1} = \beta_1 u_n + \ldots + \beta_p u_{n+1-p}}$$

Чтобы построить u^*_{n+1} надо оценить $\beta_1,\ldots,\beta_p$.

4.12 Оценка наименьших квадратов в $AR(p)$ - теорема об асимптотической нормальности.

Построим о.н.к. неизвестных $\beta_1,\ldots,\beta_p$ по наблюдениям $u_{n-p},\ldots,u_n$.

Положим $\widetilde{u}_{t-1} \stackrel{\mathrm{def}}{=} (u_{t-1},\ldots,u_{t-p})^T,\ \beta = (\beta_1,\ldots,\beta_p)^T$, тогда (20) имеет вид

$$u_t = \widetilde{u}^T_{t-1}\beta + \varepsilon_t,\ t \in \mathbb{Z}$$

Оценка наименьших квадратов вектора β - решение задачи

$$\sum_{t=1}^{n}(u_t - \widetilde{u}^T_{t-1}\theta)^2 \to \min_{\theta=(\theta_1,\ldots,\theta_p)^T\in\mathbb{R}^p} \tag{23}$$

Решение задачи (23) совпадает с решением системы уравнений

$$\sum_{t=1}^{n} u_{t-j}(u_t - \widetilde{u}^T_{t-1}\theta) = 0,\ j = 1,\ldots,p$$

Запишем эти уравнения в векторной форме

$$\sum_{t=1}^{n}\widetilde{u}_{t-1}(u_t - \widetilde{u}^T_{t-1}\theta) = 0 \tag{23'}$$

Решение этого векторного уравнения

$$\widehat{\beta}_{n,LS} = \left(\sum_{t=1}^{n}\widetilde{u}_{t-1}\widetilde{u}^T_{t-1}\right)^{-1}\sum_{t=1}^{m}\widetilde{u}_{t-1}u_t$$

Если $p\times p$ матрица $\sum_{t=1}^{n}\widetilde{u}_{t-1}\widetilde{u}^T_{t-1}$ вырождена, то полагаем $\widehat{\beta}_{n,LS} = 0$

Теорема 6

Пусть $\{\varepsilon_t\}$ - н.о.р. сл.в., $\mathrm{E}\varepsilon_1 = 0,\ 0 < \mathrm{E}\varepsilon_1^2 = \sigma^2 < \infty$. Пусть корни характеристического уравнения (21) по модулю меньше единицы. Тогда

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) \xrightarrow[n\to\infty]{d} N(0, \delta^2 K^{-1}),\ \text{где } K = \mathrm{E}\widetilde{u}_0\widetilde{u}_0^T > 0$$

Замечание 1. *В теореме 6 речь идет о сходимости по распределению векторов* $\xi_n \overset{def}{=} \sqrt{n}(\widetilde{\beta}_{n,LS} - \beta)$ *к гауссовскому вектору* $\xi \sim N(0, \sigma^2 K^{-1})$

Напомним, что если $\xi_n, \xi \in \mathbb{R}^p$ *случ. векторы, то* $\xi_n \xrightarrow{d} \xi$, *если*

$$\int_{\mathbb{R}^p} g(x)d\mathrm{P}_n(x) \to \int_{\mathbb{R}^p} g(x)d\mathrm{P}(x) \tag{24}$$

для любой непрерывной и ограниченной функции $g : \mathbb{R}^p \to \mathbb{R}^1$. *Здесь* P_n *и* P *- распределения векторов* ξ_n *и* ξ. *Можно проверить, что это определенеие равносильно следующему: для любой непрерывной и ограниченной* $g : \mathbb{R}^p \to \mathbb{C}$ *выполнено* (24). *Разумеется,* (24) *означает, что* $\mathrm{E}g(\xi_n) \to \mathrm{E}g(\xi)$. *Мы знаем, что если* $\xi_n \xrightarrow{d} \xi$, *то для любого постоянного вектора* $\lambda \in \mathbb{R}^p$

$$\lambda^T\xi_n \xrightarrow{d} \lambda^T\xi \tag{25}$$

Действительно, функция $\pi(x) = \lambda^T x$ *непрерывна, и* (25) *следует из теоремы о наследовании сходимости, но верно и обратное: если выполнено соотношение* (25), *то* $\xi_n \xrightarrow{d} \xi$.

Действительно, функция $g(x) = e^{ix},\ x \in \mathbb{R}^1$, *ограниченная непрерывная функция* $g : \mathbb{R}^1 \to \mathbb{C}$. *Тогда из* (25) *следует, что*

$$\mathrm{E}g(\lambda^T\xi_n) = \mathrm{E}e^{i\lambda^T\xi_n} \to \mathrm{E}g(\lambda^T\xi) = \mathrm{E}e^{i\lambda^T\xi},\ \forall\lambda \in \mathbb{R}^p$$

Последнее соотношение означает, что характеристическая функция вектора ξ_n $\mathrm{E}e^{i\lambda^T\xi_n}$ *при любом значении аргумента* λ *сходится к характеристической функциии* $\mathrm{E}e^{i\lambda^T\xi}$ *вектора* ξ. *Значит* $\xi_n \xrightarrow{d} \xi$.

Лемма 1 (Прием Крамера-Уолда)**.** *Если* $\xi_n, \xi \in \mathbb{R}^p$, *то*

$$\xi_n \xrightarrow{d} \xi \Leftrightarrow \lambda^T\xi_n \xrightarrow{d} \lambda^T\xi,\ \forall\lambda \in \mathbb{R}^p$$

Этот прием сводит сходимость по распределению векторов $\xi_n,\ \xi$ *к сходимости скаляров* $\lambda^T\xi_n,\ \lambda^T\xi$, *но при всех* λ

Замечание 2. *Пусть* $\{\xi_n\}$ *- сл. посл.,* $\xi_n \in \mathbb{R}^k$.
Будем писать $\xi_n = \overline{\overline{o}}_p(1),\ n \to \infty$, *если* $\xi_n \xrightarrow{\mathrm{P}} 0$.
Будем говорить, что последовательность $\{\xi_n\}$ *ограничена по вероятности и писать* $\underline{\underline{\mathcal{O}}}_p(1)$, *если*

$$\forall\varepsilon > 0\ \exists A = A(\varepsilon) :\ \sup_n \mathrm{P}(|\xi_n| > A) < \varepsilon$$

Задача. *Пусть* $\xi_n \in \mathbb{R}^k,\ \eta_n \in \mathbb{R}^1$.

1. Если $\xi_n = \underline{\underline{\mathcal{O}}}_p(1)$*, а* $\eta_n = \bar{\bar{o}}_p(1)$*, то* $\xi_n\eta_n = \bar{\bar{o}}_p(1)$

2. Если $\xi_n \xrightarrow{d} \xi$*, то* $\xi_n = \underline{\underline{\mathcal{O}}}_p(1)$.

Доказательство теоремы 6. Мы докажем теорему при дополнительных предположениях: $\mathrm{E}\left|\varepsilon_1\right|^{2+\delta} < \infty$ для некоторой $\delta > 0$; существует плотность вероятности $g(x)$ по мере Лебега.

1. Покажем, что матрица $K = \mathrm{E}\widetilde{u}_0\widetilde{u}_0^T > 0$. Для $\alpha \in \mathbb{R}^p,\ \alpha \neq 0$, имеем:

   $$\alpha^T K\alpha = \mathrm{E}(\alpha^T\widetilde{u}_0)(\widetilde{u}_0^T\alpha) = \mathrm{E}\left|\alpha\widetilde{u}_0\right|^2$$

   Но $u_t = \sum_{s\geq 0}\gamma_s\varepsilon_{t-s} = \varepsilon_t + \sum_{s\geq 1}\gamma_s\varepsilon_{t-s}$, ряд сходится в $L^{2+\delta}$.

   Значит,

   $$\alpha^T\widetilde{u}_0 = \alpha^T(u_{-1},\ldots,u_{-p})^T = \alpha_1\varepsilon_{-1} + \sum_{s\geq 1}\gamma_s\varepsilon_{-1} + \alpha_2 u_{-2} + \ldots + \alpha_p u_{-p}$$

   Случайная величина ε_{-1} абсолютно непрерывна и не зависит от остальных слагаемых. Значит, при $\alpha_1 \neq 0$ величина $\alpha^T\widetilde{u}_0$ абсолютно непрерывна, $\mathrm{P}(\alpha^T\widetilde{u}_0 \neq 0) = 1$, $\mathrm{E}\left|\alpha^T\widetilde{u}_0\right|^2 > 0$. Если $\alpha_1 = 0$, то повторяем рассуждения для первой ненулевой комп. вектора α

2. Покажем, что следующий вектор

   $$K^{-1}\sqrt{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t \xrightarrow{d} N(0,\sigma^2K^{-1})$$

   Для этого достаточно показать, что $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t \xrightarrow{d} N(0,\sigma^2K)$ и применить Теорему о слабой сходимости. В силу леммы 1 достаточно проверить, что $\forall\lambda\in\mathbb{R}^p$

   $$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\lambda^T\widetilde{u}_{t-1}\varepsilon_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t \xrightarrow{d} N(0,\sigma^2\lambda^TK\lambda) \qquad (26)$$

   Последовательность η_t есть функция от $\{u_t\}$, это строго стац. последовательность с сильным перемешиванием с коэффициентом перемешивания $\alpha(\tau) \leq c\lambda^T,\ 0 < \lambda < 1$.

   $$\mathrm{E}\eta_t = \mathrm{E}(\lambda^T\widetilde{u}_{t-1})\mathrm{E}\varepsilon_t = 0$$

   $$\mathrm{E}\left|\eta_t\right|^{2+\delta} = \mathrm{E}\left|\lambda^T\widetilde{u}_{t-1}\right|^{2+\delta}\underbrace{\mathrm{E}\left|\varepsilon_t\right|^{2+\delta}}_{<\infty\text{ по усл.}} < \infty,\ \text{т.к. } \mathrm{E}\left|u_t\right|^{2+\delta} < \infty \text{ (см. док-во Т. 3)}$$

   $$\left\{\mathrm{E}\left|\lambda^T\widetilde{u}_{t-1}\right|^{2+\delta}\right\}^{\frac{1}{2+\delta}} \leq \sum_{i=1}^{p}|\lambda_i|\left\{\mathrm{E}\left|u_1\right|^{2+\delta}\right\}^{\frac{1}{2+\delta}} < \infty \text{ в силу нер-ва Минковсокго}$$

   Кроме того, при $t < s$

   $$\mathrm{E}\eta_t\eta_s = \mathrm{E}\left\{(\lambda^T\widetilde{u}_{t-1}\varepsilon_t)(\lambda^T\widetilde{u}_{s-1})\right\}\mathrm{E}\varepsilon_s = 0$$

   Кроме того,

   $$\sum_{\tau\geq 1}(\alpha(\tau))^{\frac{2}{2+\delta}} < \infty$$

В силу Ц.П.Т. для последовательности с с.п.

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t \xrightarrow{d} N(0, \mathrm{E}\eta_0^2),\ \mathrm{E}\eta_0^2 = \sigma^2\lambda^T K\lambda$$

Т.о. доказали (26), поэтому

$$K^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t \xrightarrow{d} N(0, \sigma^2 K^{-1}) \tag{27}$$

3. Пусть $K_n \overset{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}\widetilde{u}_{t-1}^T$. Если $\det(K_n) > 0$, то

$$\widetilde{\beta}_{n,LS} = K_n^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}u_t = K_n^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}(\widetilde{u}_{t-1}^T\beta + \varepsilon_t) = \beta + K_n^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t$$

Значит, при невырожденной K_n

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) = K_n^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t$$

В силу З.Б.Ч. для посл. с с.п.

$$K_n \xrightarrow{\text{п.н.}} K = \mathrm{E}\widetilde{u}_0\widetilde{u}_0^T > 0,\ \det(K) > 0$$

Поэтому $\det(K_n) \xrightarrow{\text{п.н.}} \det(K) > 0$, и если $S_n = \{\omega : \det(K_n) > 0\}$, то $\mathrm{P}(S_n) \to 1$.

Напомним,

$$\widehat{\beta}_{n,LS} = \begin{cases} K_n^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{u}_{t-1}u_t,\ \omega \in S_n \\ 0,\ \omega \in \overline{S}_n \end{cases}$$

Покажем, что

$$\gamma_n \overset{\text{def}}{=} \sqrt{n}(\widehat{\beta}_{n,LS} - \beta) - K^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{u}_{t-1}\varepsilon_t \xrightarrow{\mathrm{P}} 0 \tag{28}$$

Из (27) и (28) следует, что

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) \xrightarrow{d} N(0, \sigma^2 K^{-1})$$

Действительно, если ξ_n и η_n - любые случайные векторы такие, что $\xi_n = \eta_n + \alpha_n$, и $\eta_n \xrightarrow{d} \xi$, а $\alpha_n = \overline{\overline{o}}_p(1)$, то $\lambda^T\xi_n = \lambda^T\eta_n + \lambda^T\alpha_n \xrightarrow{d} \lambda^T\xi$ в силу Леммы Слуцкого (т.к. $\lambda^T\eta_n \xrightarrow{d} \lambda^T\xi$, а $\lambda^T\alpha_n \xrightarrow{\mathrm{P}} 0$) $\forall\lambda \in \mathbb{R}^p$.

Значит, $\xi_n \xrightarrow{d} \xi$ в силу Леммы 1.

Пусть

$$S_n^{\Delta} = \left\{\omega :\ \det(K_n) \geq \Delta = \frac{1}{2}\det(K)\right\}$$

Тогда

$$\mathrm{P}(\overline{S_n^{\Delta}}) \leq \mathrm{P}(|\det(K_n) - \det(K)| > \Delta) \to 0 \Rightarrow \mathrm{P}(S_n^{\Delta}) \to 1$$

Далее $|\bullet|$ - Евклидова норма матрицы или вектора на S_n^Δ

$$\sqrt{n}(\widehat{\beta}_{n,LS} - \beta) = K_n^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{u}_{t-1}\varepsilon_t$$

Поэтому $\forall \delta > 0$:

$$\mathrm{P}(|\gamma_n| > \delta) = \mathrm{P}\left(\left|(K_n^{-1} - K^{-1}) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{u}_{t-1}\varepsilon_t\right| > \delta, S_n^\Delta\right) + \mathrm{P}\left(|\gamma_n| > \delta, \overline{S}_n^\Delta\right)$$

Вторая вероятность не больше $\mathrm{P}(\overline{S}_n^\Delta) \to 0$. Первая вероятность не больше

$$\mathrm{P}\left(\left|K_n^{-1}\right| |K - K_n| \left|K^{-1}\right| \left|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{u}_{t-1}\varepsilon_t\right| > \delta, S_n^\Delta\right) \tag{29}$$

В (29) $|K - K_n| \xrightarrow{\mathrm{P}} 0$, $|K^{-1}|$ - кон. число, $\left|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{u}_{t-1}\varepsilon_t\right| = \underline{\underline{\mathcal{O}}}_p(1)$.

Рассмотрим $|K_n^{-1}|$ на S_n^Δ. Пусть A_{ji}^n и A_{ji} - алгебр. доп. в K_n и K соответственно. Тогда:

$$K_n^{-1} = (a_{ij}^n) \stackrel{\mathrm{def}}{=} \left(\frac{A_{ji}^n}{\det(K_n)}\right)$$

На S_n^Δ:

$$\left|a_{ij}^n\right| \leq \frac{\left|A_{ij}^n\right|}{\Delta}$$

Пусть $B_n = (b_{ij}^n) \stackrel{\mathrm{def}}{=} \left(\frac{\left|A_{ji}^n\right|}{\Delta}\right)$, $B = (b_{ij}) \stackrel{\mathrm{def}}{=} \left(\frac{|A_{ji}|}{\Delta}\right)$ - $p \times p$ матрицы.

Поскольку $b_{ij}^n \xrightarrow{\text{п.н.}} b_{ij}$, то $|B_n| \xrightarrow{\text{п.н.}} |B|$, и поэтому $|K_n^{-1}| \leq |B_n| = \underline{\underline{\mathcal{O}}}_p(1)$.

Поэтому вероятность в (27) не больше

$$\mathrm{P}\left(|B_n| |K - K_n| \left|K^{-1}\right| \left|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{u}_{t-1}\varepsilon_t\right| > \delta\right) \to 0$$

Соотношение (28) доказано, а значит и Теорема.

$\square$

4.13 Проверка гипотез о порядке авторегрессии.

Пусть $\beta^T = (\underbrace{\beta_1^T}_{m \text{ в-р}}, \underbrace{\beta_2^T}_{p-m \text{ в-р}})$

$H_0: \quad \beta_2 = 0 \qquad$ (порядок авторегр. $\le m$)

$H_1: \quad \beta_2 \neq 0$

Лемма 2. *Пусть* $\xi_n, \xi \in \mathbb{R}^p$, $\xi_n \xrightarrow{d} \xi \sim N(a, \Sigma)$, $\Sigma > 0$.
Пусть оценки $\widehat{\Sigma}_n \xrightarrow{\mathrm{P}} \Sigma$, $\widetilde{\Sigma}_n = \begin{cases} \widehat{\Sigma}_n^{-1}, & \det \widehat{\Sigma} \neq 0 \\ E_p, & \det \widehat{\Sigma} = 0 \end{cases}$. *Тогда*

$$(\xi - a)^T \widehat{\Sigma}_n^{-1} (\xi - a) \xrightarrow{d} \chi^2(p)$$

Возьмем оценкой матрицы $K = \mathrm{E} \widetilde{u}_0 \widetilde{u}_0^T$ матрицу

$$K_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{t=1}^{n} \widetilde{u}_{t-1} \widetilde{u}_{t-1}^T$$

Тогда

$$K_n \xrightarrow{\text{п.н.}} K > 0$$

Пусть $\widehat{\beta}_{n,LS} = (\underbrace{\widehat{\beta}_{1n}^T}_{m \text{ в-р}}, \underbrace{\widehat{\beta}_{2n}^T}_{p-m \text{ в-р}})$. В силу Теоремы 6:

$$\sqrt{n}(\widehat{\beta}_{2n} - \beta_2) \xrightarrow{d} N(0, \sigma B_{22}), \text{ где } K^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

Тогда при $H_0 : \beta_2 = 0$ в силу Леммы 2 $\left(\widetilde{K}_n^{-1} = \begin{pmatrix} \widetilde{B}_{11} & \widetilde{B}_{12} \\ \widetilde{B}_{21} & \widetilde{B}_{22} \end{pmatrix} = \begin{cases} K_n^{-1}, & \det K_n \neq 0 \\ E_p, & \det K_n = 0 \end{cases} \right)$

$$\frac{n \widehat{\beta}_{2n}^T \widetilde{B}_{22}^{-1} \widehat{\beta_{2n}}}{\sigma^2} \xrightarrow{d} \chi(p - m)$$

Задача. *Пусть* $\widehat{s}_n^2 \stackrel{def}{=} \frac{1}{n} \sum_{t=1}^{n} (u_t - \widetilde{u}_{t-1}^T \widehat{\beta}_{n,LS})^2$. *Показать, что* $\widehat{s}_n^2 \xrightarrow{\mathrm{P}} \sigma^2$.

Тестовая статистика для H_0:

$$t_n \stackrel{\text{def}}{=} \frac{n \widehat{\beta}_2^T \widetilde{B}_{22}^{-1} \widehat{\beta}_2}{\widehat{s}_n^2}$$

При H_0 $t_n \xrightarrow{d} \chi^2(p - m)$; критическое множество $t_n > \chi_{1-\alpha}(p - m)$, $\chi_{1-\alpha}(p - m)$ - квантиль уровня $1 - \alpha$ распределения хи-квадрат с $p - m$ степенями свободы.

Тогда $\mathrm{P}(H_1 | H_0) \to \alpha$, $\mathrm{P}(H_0 | H_1) \to 0 \Rightarrow \begin{cases} \mathrm{P}(H_0 | H_0) \to 1 - \alpha \\ \mathrm{P}(H_1 | H_1) \to 1 \end{cases}$

Хороший критерий!

5 Литература

1. Болдин М.В. Коспект лекций (слайды) по курсу "Математическая статистика и приложения."Эконом. поток, 2021.

2. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М., Высшая школа, 1984.

3. Боровков А.А., Математическая статистика. М., Наука, 1984.

4. Болдин М.В., Симонова Г.И., Тюрин Ю.Н. Знаковый анализ линейных моделей. М., Наука, 1997.

5. Ширяев А.Н. Основы стохастической финансовой математики. Факты. Модели.М., Фазис, 1988.

6. Леман Э. Теория точечного оценивания. М., Наука, 1991.

7. Ширяев А.Н.. Вероятность-1, 5-ое изд. МЦНМО, М.,2011.