1 Асимптотические оптимальные оценки

Пусть сл. векторы $\xi_n, \xi \in \mathbb{R}^K$, и определены на $(\Omega, \mathcal{F}, \mathrm{P})$. Пусть функция распределения ξ_n есть $F_n(x)$, хар. ф-ция есть $\phi_n(t)$, а распределение есть Q_n. Для вектора ξ функцию распределения, хар. ф-цию и 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 $\xi_n, \xi \in \mathbb{R}^K, H : \mathbb{R}^K \to \mathbb{R}^1$ 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 $H(\xi_n) \xrightarrow{\mathrm{P}} H(\xi)$*

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:

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

$$\alpha_n \leq \mathrm{E}|e^{it\xi_n}(e^{it\eta_n+isa})| = \mathrm{E}|e^{it\eta_n+isa}| = \mathrm{E}g(\eta_n),\ g(x) := |e^{isx} - e^{isa}|$$

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, $\alpha \to 0$.

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

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

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

Опр. 3. *Если $n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma(\theta))\ \forall\theta \in \Theta$ и коварицонная матрица $0 < \Sigma(\theta) < \infty$, то $\hat{\theta}_n$ называется асимптотической нормальной оценкой.*

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

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

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

Если $\hat{\theta}_n$ - асимптотическая нормальная оценка θ (так как θ и $\hat{\theta}_n$ скаляры: $n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))\ 0 < \sigma^2 < \infty,\ \forall\theta \in \Theta$), то:

1. $\hat{\theta}_n$ - состоятельная оценка θ, так как $\hat{\theta}_n - \theta = n^{-1/2}n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{P}} 0$ в силу п. 2 леммы Слуцкого.

2. Скорость сходимости $\hat{\theta}_n$ к θ есть $O(n^{1/2})$

3. При больших n со сл. в. $n^{1/2}(\hat{\theta}_n - \theta)$ можно обращаться (с осторожностью!) как с Гауссовской величиной.

   Например, пусть дисперсия предельного Гауссовского закона $\sigma^2(\theta)$ будет непреревной ф-цией θ. Тогда
   $$\frac{n^{1/2}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} = \underbrace{\frac{n^{1/2}(\hat{\theta}_n - \theta)}{\sigma(\theta)}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)}}_{\xrightarrow{\mathrm{P}} 1} \xrightarrow{d} \eta \sim N(0,1)$$
   в силу п. 2 леммы Слуцкого. Значит,
   $$\mathrm{P}_\theta(|\frac{n^{1/2}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)}| < \xi_{1-\alpha/2}) \to \mathrm{P}(|\eta| < \xi_{1-\alpha/2}) = 1 - \alpha$$

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

4. Асимптотические Гауссовские оценки можно сравнивать между собой:
   Если $n^{1/2}(\hat{\theta}_{i,n} - \theta) \xrightarrow{d} N(0, \sigma_i^2(\theta)),\ i = 1, 2, \ldots$, 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 $n^{1/2}(\hat{\theta}_{1,n} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta))$ и $n^{1/2}(\hat{\theta}_{2,n'} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta))$.

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

Если да, то θ_n^* требует не больше наблюдений, чем любая $\hat{\theta}_n$, чтобы достичь одинаковой с $\hat{\theta}_n$ точности. Ясно, что пределеная дисперсия $n^{1/2}(\theta_n^* - \theta)$ должна быть не больше асимптотической дисперсии $n^{1/2}(\hat{\theta}_n - \theta)$ для любой асимптотической Гауссовской оценки $\hat{\theta}_n$. Но какова самая маленькая асимптотическая дисперсия у $n^{1/2}(\hat{\theta}_n - \theta)$?

Теорема Бахадура
Пусть $X_1, \ldots, X_n$ - 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. *$|\frac{\partial^2}{\partial \theta^2} \ln(f(x, \theta))| \leq M(x)\ \forall x \in N_f,\ \theta \in \Theta,\ \mathrm{E}_\theta M(X_1) < \infty$*

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

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

Доказательство. Без доказательства. □

Опр. 5. *Если $\theta, \hat{\theta}_n \in \mathbb{R}^1$ и $n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{i(\theta)}),\ n \to \infty,\ \forall \theta \in \Theta$, причем $0 < i(\theta) < \infty$, то $\hat{\theta}_n$ называется **асимптотически эффективной оценкой**.*

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

Дальше $X=(X_1,\dots,X_n),\ X\sim \mathrm{P}_\theta,\ \theta\in\Theta\subseteq\mathbb{R}^1$. 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. 6. *Функция $p(X,\theta)$ как функция θ при фиксированном X называется* ***правдоподобием*** *функции.*

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

называется логарифмическим правдоподобием.

Пусть θ_0 будет истинное значение параметра.

Лемма 1 (Неравенство Йенсена)**.** *Пусть $g(x)$ выпукла книзу борелевская функция,* $\mathrm{E}|\xi|<\infty$, $\mathrm{E}|g(\xi)|<\infty$. *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.

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

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

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 Условию(A)(1)).

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

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

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

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

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

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 такиъ решений есть такой корень $\hat{\theta}_n$, что он является состоятельнаой 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 $\hat{\theta}_n \in \Theta$, что

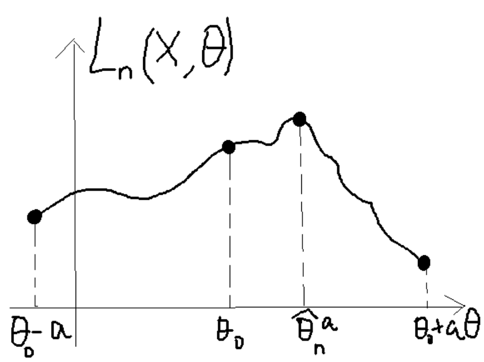
$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \epsilon, S_n) \to 1,\ n \to \infty,\ \forall\epsilon > 0$$

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 $\hat{\theta}_n^a$ на интервале $(\theta_0 - a, \theta_0 + a)$



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

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

$\forall$ малого $\epsilon > 0$:

$$\mathrm{P}_{\theta_0}(|\hat{\theta}_n - \theta_0| < \epsilon, S_n) \geq \mathrm{P}_{\theta_0}(|\hat{\theta}_n^\epsilon - \theta_0| < \epsilon, S_n^\epsilon) \tag{3}$$

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

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

$$\mathrm{P}_{\theta_0}(|\hat{\theta}_n - \theta_0| < \epsilon, S_n) \to 1$$

□

Замечание. *Пусть*

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

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

$$\mathrm{P}(|\theta_n^* - \theta_0| < \epsilon) = \mathrm{P}(|\hat{\theta}_n - \theta_0| < \epsilon, S_n) + \mathrm{P}(|\theta' - \theta_0| < \epsilon, \overline{S}_n) \to 1$$

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***

Теорема 5 (Об асимптотической эффективности состоятельности решения)**.**
Пусть $X = (X_1, \dots, X_n),\ \{X_i\}$ *- н.о.р. сл.в., и удовлетворяются предположения Теоремы Бахадура, в которых условия 3 и 6 заменены на предположения о третьей, а не второй производной. То есть*

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

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

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

То есть θ_n^* *- асимптотическая эффективная оценка.*

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

Для фиксированного X в силу формулы Тейлора и последнего замечания:

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

Отсюда,

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

Рассмотрим числитель (5) и покажем, что

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

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 знаменете (5)

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

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

Оценки максимального правдоподобия для векторого параметра

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

Тогда логарифмические правдоподобие имеет вид

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

Система уравнений правдоподобия

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

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

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

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

$$n^{1/2}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)),\ n \to \infty$$

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

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

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

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

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

Уравнение правдоподобия имеет вид

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

Его решение существует и единственно, это $\overline{X}$, 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. сущесвтует и равна $\overline{X}$, в противном случае о.м.п. не существует. Если положить

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

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

$$n^{1/2}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, \sigma^2) \tag{10}$$

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

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

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

Но на границах теряется асимптотическая Гауссовость.

2 Проверка статистических гипотез

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

Опр. 8. *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. 9. *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. 10. *Множество S называется критическим множеством или критерием, $\overline{S}$ - область принятия гипотезы.*

Опр. 11. ***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. 12. ***Мощность критерия*** *S называется функция $W(S, \theta) = W(\theta) := \mathrm{P}_\theta(X \in S)$ (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. 13. *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. 14. *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:

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

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

Введем множество

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

Теорема 6 (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. $X \in S(\lambda) \Leftrightarrow \frac{p_1(x)}{p_0(x)} > \lambda$. *Так как $p_1(X)$ и $p_0(X)$ - правдоподобие, то критерий называется критерием отношения правдоподобия Неймана-Пирсона.*

Замечание. *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 свойствоназыается 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 $\mathrm{P}_1(X \in S) \geq \mathrm{P}_0(X \in S)$ $\square$

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*

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

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

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

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

Итак,

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

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 $\mathrm{P}_{\theta_0}(X \in S^*_\alpha) = \alpha$, 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^+

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

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

О связи между доверительным оцениванием и проверкой гипотез