Параметрическое оценивание

а) Метод моментов

Параметрические модели:

$$\mathfrak{F} = \left\{ f(x;\theta) : \ \theta \in \Theta \right\}$$

1

$\Theta \subset \mathbb{R}^k$ **- пространство параметров,** $\theta = (\theta_1, \ldots, \theta_k)$ **- вектор параметров**

$T(\theta)$ **- значение, которое надо оценить**

<u>Пример</u>: $N(\mu, \sigma^2)$**,** $\theta = (\mu, \sigma)$**. Если** μ **необходимо оценить, то** $\mu = T(\theta)$**, а** σ **- мешающий параметр**

<u>Пример</u>: **пусть** $X_1, \dots, X_n \sim \text{Normal}(\mu, \sigma^2)$**.** $\theta = (\mu, \sigma)$ **- параметр,** $\Theta = \{(\mu, \sigma) :\ \mu \in \mathbb{R},\ \sigma > 0\}$ **- пространство параметров. Допустим, что измерения** X_i **- интегральная характеристика теста по исследованию крови. Задача состоит в том, чтобы** τ **- долю наблюдений, для которых характеристика превышает 1. Пусть** Z **- стандартная**
2 **нормальная случайная величина. Тогда**

$$\tau \;=\; \mathbb{P}(X > 1) = 1 - \mathbb{P}(X < 1) = 1 - \mathbb{P}\left(\frac{X-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right)$$

$$= \;1 - \mathbb{P}\left(Z < \frac{1-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{1-\mu}{\sigma}\right)$$

$\tau = T(\mu, \sigma) = 1 - \Phi((1-\mu)/\sigma)$ **- параметр, который необходимо оценить**

<u>Пример</u>: пусть X **имеет** $\mathrm{Gamma}(\alpha, \beta)$ **распределение, то есть**

$$f(x;\, \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0$$

,

3 **где** $\alpha, \beta > 0$ **и**

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy$$

обозначает гамма-функцию. $\theta = (\alpha, \beta)$ **- параметр. Гамма-распределение обычно используют для моделирования времени жизни. Если задача состоит в том, чтобы оценить среднее время жизни, то**

$$T(\alpha, \beta) = \mathbb{E}_{\theta}(X_1) = \alpha\beta$$

Метод моментов – не оптимален; прост в использовании; полученные с помощью этого метода оценки могут использоваться в качестве начальных значений для более «тонких» алгоритмов

Пусть $\theta = (\theta_1, \ldots, \theta_k)$ **- параметр. Для** $1 \leq j \leq k$ **определим** j **-й момент согласно формуле**

$$\alpha_j \equiv \alpha_j(\theta) = \mathbb{E}_\theta(X^j) = \int x^j dF_\theta(x)$$

и j **-й выборочный момент согласно формуле**

$$\widehat{\alpha}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

$\widehat{\theta}_n$ **- оценка параметра** $\theta = (\theta_1, \ldots, \theta_k)$ **на основе метода моментов, если**

$$
\begin{aligned}
\alpha_1(\widehat{\theta}_n) &= \widehat{\alpha}_1 \\
\alpha_2(\widehat{\theta}_n) &= \widehat{\alpha}_2 \\
\vdots \quad & \vdots \quad \vdots \\
\alpha_k(\widehat{\theta}_n) &= \widehat{\alpha}_k
\end{aligned}
$$

<u>Пример</u>: $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$**. Тогда** $\alpha_1 = \mathbb{E}_p(X) = p$ **и** $\widehat{\alpha}_1 = n^{-1} \sum_{i=1}^{n} X_i$**, откуда**

$$\widehat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

<u>Пример</u>: $X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma^2)$**. Тогда**

$$\alpha_1 = \mathbb{E}_\theta(X_1) = \mu$$

$$\alpha_2 = \mathbb{E}_\theta(X_1^2) = \mathbb{V}_\theta(X_1) + (\mathbb{E}_\theta(X_1))^2 = \sigma^2 + \mu^2$$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

6

$$\widehat{\sigma}^2 + \widehat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Решая систему линейных уравнений, получаем, что

$$\widehat{\mu} = \overline{X}_n$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

<u>Теорема</u>. Пусть $\widehat{\theta}_n$ **- оценка параметра** θ **с помощью метода моментов. Тогда (при определенных предположениях на распределение выборки):**

1. $\widehat{\theta}_n$ **существует с вероятностью 1**
2. $\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta$ **при** $n \to \infty$
3. **Оценка асимптотически нормальна, то есть**

7

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightsquigarrow N(0, \Sigma),$$

где $\Sigma = g\mathbb{E}_\theta(YY^T)g^T$, $Y = (X, X^2, \ldots, X^k)^T$,

$g = (g_1, \ldots, g_k)$ **и** $g_j = \partial\alpha_j^{-1}(\theta)/\partial\theta$.

<u>Замечание</u>: Последний пункт теоремы можно использовать для нахождений стандартных ошибок и доверительных интервалов

b) Метод максимального правдоподобия и его свойства

Пусть задана i.i.d. выборка $X_1, \ldots, X_n \sim F$**, при этом у распределения имеется плотность** $f(x;\theta)$

Функция правдоподобия задается формулой

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta)$$

Логарифмическая функция правдоподобия задается формулой

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta)$$

Будем рассматривать правдоподобие как функцию параметра

$$\mathcal{L}_n : \Theta \to [0, \infty)$$

Оценка максимального правдоподобия (ОМП оценка) определяется как такое значение $\widehat{\theta}_n$ **параметра** θ**, которое максимизирует** $\mathcal{L}_n(\theta)$

<u>Пример</u>: пусть $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$**.**

Распределение определяется по формуле $f(x;p) = p^x(1-p)^{1-x}$ **при** $x = 0, 1$**.**

9

$$\mathcal{L}_n(p) = \prod_{i=1}^{n} f(X_i;\, p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^S(1-p)^{n-S}$$ **,**

где $S = \sum_i X_i$**.**

$$\ell_n(p) = S\log p + (n-S)\log(1-p)$$

Откуда получаем, что ОМП оценка равна $\widehat{p}_n = S/n$**.**

<u>Пример</u>: пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$**.** $\theta = (\mu, \sigma)$ **- параметр. Функция правдоподобия (без учета констант) имеет вид**

$$\begin{aligned}
\mathcal{L}_n(\mu, \sigma) &= \prod_i \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(X_i - \mu)^2\right\} \\
&= \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (X_i - \mu)^2\right\} \\
&= \sigma^{-n} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{X} - \mu)^2}{2\sigma^2}\right\}
\end{aligned}$$

,

где $\overline{X} = n^{-1}\sum_i X_i$**,** $S^2 = n^{-1}\sum_i (X_i - \overline{X})^2$**. Последнее равенство следует из равенства** $\sum_i (X_i - \mu)^2 = nS^2 + n(\overline{X} - \mu)^2$**, которое легко следует из** $\sum_i (X_i - \mu)^2 = \sum_i (X_i - \overline{X} + \overline{X} - \mu)^2$

$$\ell(\mu, \sigma) = -n \log \sigma - \frac{nS^2}{2\sigma^2} - \frac{n(\overline{X} - \mu)^2}{2\sigma^2}$$

Пусть $\dfrac{\partial \ell(\mu, \sigma)}{\partial \mu} = 0$ **и** $\dfrac{\partial \ell(\mu, \sigma)}{\partial \sigma} = 0$**,**

тогда $\widehat{\mu} = \overline{X}$ **и** $\widehat{\sigma} = S$**.**

<u>Пример</u>: пусть $X_1, \ldots, X_n \sim Unif(0, \theta)$**.**

$$f(x;\theta) = \begin{cases} 1/\theta, x \in (0,\theta) \\ 0, x \notin (0,\theta) \end{cases}$$

Рассмотрим фиксированное значение θ**. Допустим, что** $\theta < X_i$ **для некоторого** i**. Тогда** $f(X_i;\theta) = 0$**, поэтому** $\mathcal{L}_n(\theta) = \prod_i f(X_i;\theta) = 0$**.**

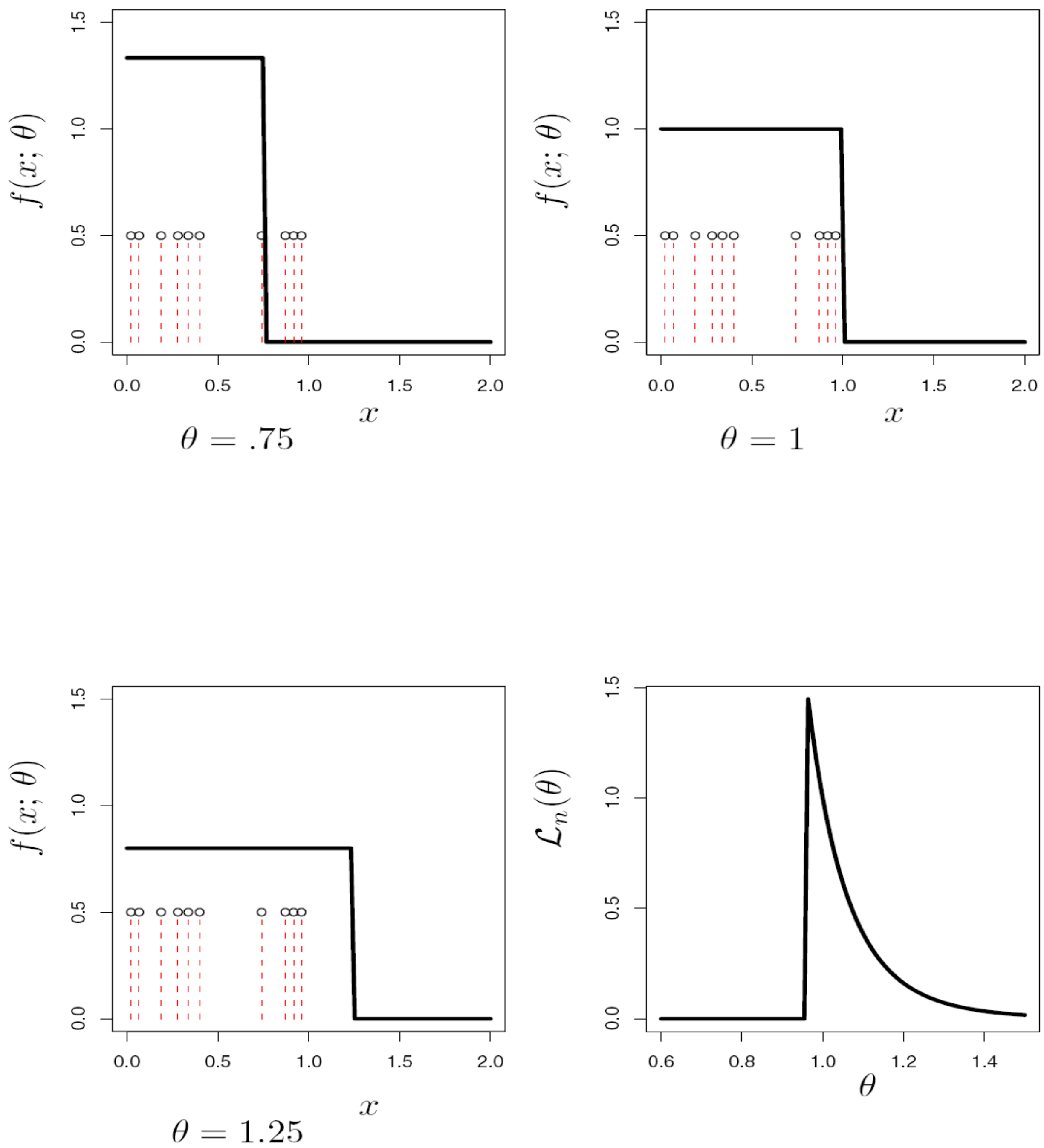
Таким образом, $\mathcal{L}_n(\theta) = 0$ **если** $X_i > \theta$ **хотя бы для одного** i**, то есть** $\mathcal{L}_n(\theta) = 0$ **при** $\theta < X_{(n)}$**, где** $X_{(n)} = \max\{X_1, \ldots, X_n\}$**.**

Рассмотрим произвольное $\theta \geq X_{(n)}$**. Тогда** $f(X_i;\theta) = 1/\theta$ **для любого** i **и** $\mathcal{L}_n(\theta) = \prod_i f(X_i;\theta) = \theta^{-n}$**. Таким образом,**

$$\mathcal{L}_n(\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \theta \geq X_{(n)} \\ 0 & \theta < X_{(n)} \end{cases}$$

Так как $\mathcal{L}_n(\theta)$ **строго убывающая функция параметра** θ **на интервале** $[X_{(n)}, \infty)$**, то** $\widehat{\theta}_n = X_{(n)}$

14



$f(x;\theta)$
x
$\theta = .75$
$f(x;\theta)$
x
$\theta = 1$
$f(x;\theta)$
x
$\theta = 1.25$
$\mathcal{L}_n(\theta)$
θ

Свойства функции правдоподобия:

1. **ОМП состоятельная, то есть $\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta_\star$, где $\theta_\star$ - реальное значение параметра θ**

2. **ОМП не зависит от параметризации, то есть $\widehat{\theta}_n$ - ОМП для θ, тогда $g(\widehat{\theta}_n)$ - ОМП для $g(\theta)$**

3. **ОМП асимптотически нормальна - $(\widehat{\theta} - \theta_\star)/\widehat{\mathsf{se}} \rightsquigarrow N(0,1)$**
4. **ОМП асимптотически оптимальна или эффективна (при достаточно большом объеме выборки ОМП имеет меньшую дисперсию)**
5. **ОМП приближенно совпадает с байесовской оценкой**

<u>Замечание</u>: вышеприведенные свойства ОМП имеют место, если функция $f(x;\theta)$ достаточно регулярная. В «слишком» сложных случаях ОМП оценка «теряет» эти свойства

Состоятельность ОМП

Расстояния Кульбака-Лейблера $D(f,g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$**,**

где f **и** g **- плотности распределений (это не расстояние в обычном смысле, так как функционал не семметричен)**

Можно показать, что $D(f,g) \geq 0$ $D(f,f) = 0$**. Будем писать** $D(\theta,\psi)$ **для** $\theta,\psi \in \Theta$ **вместо** $D(f(x;\theta), f(x;\psi))$**.**

Будем говорить, что модель $\mathfrak{F}$ **идентифицируема, если из** $\theta \neq \psi$ **следует, что** $D(\theta,\psi) > 0$

Максимизация $\ell_n(\theta)$ **эквивалентна максимизации**

$$M_n(\theta) = \frac{1}{n}\sum_i \log\frac{f(X_i;\,\theta)}{f(X_i;\,\theta_\star)},$$

поскольку $M_n(\theta) = n^{-1}(\ell_n(\theta) - \ell_n(\theta_\star))$ **и** $\ell_n(\theta_\star)$ **- константа**

17
$$\begin{aligned} \mathbb{E}_{\theta_\star}\left(\log\frac{f(X_i;\,\theta)}{f(X_i;\,\theta_\star)}\right) &= \int \log\left(\frac{f(x;\,\theta)}{f(x;\,\theta_\star)}\right) f(x;\,\theta_\star)dx \\ &= -\int \log\left(\frac{f(x;\,\theta_\star)}{f(x;\,\theta)}\right) f(x;\,\theta_\star)dx \\ &= -D(\theta_\star,\theta) \end{aligned}$$

Таким образом, $M_n(\theta) \approx -D(\theta_\star,\theta)$ **принимает максимальное значение в точке** $\theta_\star$**, поскольку** $-D(\theta_\star,\theta_\star) = 0$ **и** $-D(\theta_\star,\theta) < 0$ **при** $\theta \neq \theta_\star$

<u>Теорема</u>. Пусть $\theta_\star$ **- реальное значение параметра** θ**. Обозначим через**

$$M_n(\theta) = \frac{1}{n}\sum_i \log\frac{f(X_i;\,\theta)}{f(X_i;\,\theta_\star)}$$

и $M(\theta) = -D(\theta_\star, \theta)$**. Допустим, что**

18

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathrm{P}} 0$$

и для каждого $\epsilon > 0$

$$\sup_{\theta:|\theta-\theta_\star|\geq\epsilon} M(\theta) < M(\theta_\star)$$

Пусть $\widehat{\theta}_n$ **обозначает ОМП. Тогда** $\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta_\star$

<u>Доказательство</u>. Так как $\widehat{\theta}_n$ **максимизирует** $M_n(\theta)$**, то** $M_n(\widehat{\theta}_n) \geq M_n(\theta_\star)$**. Следовательно,**

$$\begin{aligned} M(\theta_\star) - M(\widehat{\theta}_n) &= M_n(\theta_\star) - M(\widehat{\theta}_n) + M(\theta_\star) - M_n(\theta_\star) \\ &\leq M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n) + M(\theta_\star) - M_n(\theta_\star) \\ &\leq \sup_\theta |M_n(\theta) - M(\theta)| + M(\theta_\star) - M_n(\theta_\star) \\ &\xrightarrow{\mathrm{P}} 0 \end{aligned}$$

Таким образом, для любого $\delta > 0$

$$\mathbb{P}\left(M(\widehat{\theta}_n) < M(\theta_\star) - \delta\right) \to 0$$

Возьмем произвольное $\epsilon > 0$**. Согласно условию теоремы найдется** $\delta > 0$**,**

для которого из неравенства $|\theta - \theta_\star| \geq \epsilon$ **следует, что**

$$M(\theta) < M(\theta_\star) - \delta.$$

Значит,

$$\mathbb{P}(|\widehat{\theta}_n - \theta_\star| > \epsilon) \leq \mathbb{P}\left(M(\widehat{\theta}_n) < M(\theta_\star) - \delta\right) \to 0$$

<u>ОМП не зависит от параметризации</u>

<u>Теорема</u>. Пусть $\tau = g(\theta)$ **- функция параметра** θ**,** $\widehat{\theta}_n$ **- ОМП** θ**. Тогда** $\widehat{\tau}_n = g(\widehat{\theta}_n)$ **является ОМП для** $\tau = g(\theta)$**.**

<u>Доказательство</u>.

Обозначим через $h = g^{-1}$ **- обратную функцию к** g**. Тогда** $\widehat{\theta}_n = h(\widehat{\tau}_n)$**. Для любого** τ **будет выполнено, что**

$$\mathcal{L}(\tau) = \prod_i f(x_i; h(\tau)) = \prod_i f(x_i; \theta) = \mathcal{L}(\theta),$$

где $\theta = h(\tau)$**. Следовательно, для любого** τ

$$\mathcal{L}_n(\tau) = \mathcal{L}(\theta) \leq \mathcal{L}(\widehat{\theta}) = \mathcal{L}_n(\widehat{\tau})$$

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim N(\theta, 1)$**. ОМП для** θ **равна** $\widehat{\theta}_n = \overline{X}_n$**.**
Пусть $\tau = e^{\theta}$**. Тогда ОМП для** τ **равняется** $\widehat{\tau} = e^{\widehat{\theta}} = e^{\overline{X}}$**.**

<u>Асимптотическая нормальность ОМП</u>

Пусть $$s(X;\theta) = \frac{\partial \log f(X;\theta)}{\partial \theta}$$ **. Тогда информация Фишера равняется**

$$I_n(\theta) = \mathbb{V}_\theta \left(\sum_{i=1}^n s(X_i;\theta) \right)$$

$$= \sum_{i=1}^n \mathbb{V}_\theta \left(s(X_i;\theta) \right)$$

Лемма. $\mathbb{E}_\theta(s(X;\theta)) = 0$ **и** $\mathbb{V}_\theta(s(X;\theta)) = \mathbb{E}_\theta(s^2(X;\theta))$

Доказательство.

Очевидно, $1 = \int f(x;\theta)dx$

$$0 = \frac{\partial}{\partial\theta}\int f(x;\theta)dx = \int \frac{\partial}{\partial\theta} f(x;\theta)dx$$

$$= \int \frac{\frac{\partial f(x;\theta)}{\partial\theta}}{f(x;\theta)} f(x;\theta)dx = \int \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)dx$$

$$= \int s(x;\theta) f(x;\theta)dx = \mathbb{E}_\theta s(X;\theta)$$

<u>Теорема</u>. $I_n(\theta) = nI(\theta)$**, при этом**

$$\begin{aligned} I(\theta) &= -\mathbb{E}_\theta\left(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right) \\ &= -\int\left(\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2}\right) f(x;\,\theta)dx \end{aligned}$$

<u>Теорема</u> (асимптотическая нормальность ОМП). Пусть $\mathsf{se} = \sqrt{\mathbb{V}(\widehat{\theta}_n)}$**.**
При достаточно общих условиях регулярности на плотность распределения выборки будет выполнено, что

1. $\mathsf{se} \approx \sqrt{1/I_n(\theta)}$ **и**

25

$$\frac{(\widehat{\theta}_n - \theta)}{\mathsf{se}} \rightsquigarrow N(0,1)$$

2. Пусть $\widehat{\mathsf{se}} = \sqrt{1/I_n(\widehat{\theta}_n)}$**, тогда**

$$\frac{(\widehat{\theta}_n - \theta)}{\widehat{\mathsf{se}}} \rightsquigarrow N(0,1)$$

<u>Доказательство</u>.

Пусть $\ell(\theta) = \log \mathcal{L}(\theta)$**. Тогда**

$$0 = \ell'(\widehat{\theta}) \approx \ell'(\theta) + (\widehat{\theta} - \theta)\ell''(\theta)$$

Отсюда получаем, что

$$\widehat{\theta} - \theta = -\ell'(\theta)/\ell''(\theta)$$

$$\sqrt{n}(\widehat{\theta} - \theta) = \frac{\frac{1}{\sqrt{n}}\ell'(\theta)}{-\frac{1}{n}\ell''(\theta)} \qquad \text{(*)}$$

Пусть $Y_i = \partial \log f(X_i; \theta)/\partial\theta$**. Из леммы следует, что** $\mathbb{E}(Y_i) = 0$ **и** $\mathbb{V}(Y_i) = I(\theta)$**. Значит, для числителя в (*) выполняется, что**

$$n^{-1/2} \sum_i Y_i = \sqrt{n}\overline{Y} = \sqrt{n}(\overline{Y} - 0) \rightsquigarrow W \sim N(0, I(\theta))$$

Положим $A_i = -\partial^2 \log f(X_i;\theta)/\partial\theta^2$**. Тогда** $\mathbb{E}(A_i) = I(\theta)$ **и для знаменателя в (*) выполняется, что** $\overline{A} \xrightarrow{\text{P}} I(\theta)$**. Итак,**

$$\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \frac{W}{I(\theta)} \stackrel{d}{=} N\left(0, \frac{1}{I(\theta)}\right)$$

27 **Предположим, что** $I(\theta)$ **- непрерывная функция своего аргумента. Тогда** $I(\widehat{\theta}_n) \xrightarrow{\text{P}} I(\theta)$ **и**

$$\begin{aligned} \frac{\widehat{\theta}_n - \theta}{\widehat{\mathsf{se}}} &= \sqrt{n} I^{1/2}(\widehat{\theta}_n)(\widehat{\theta}_n - \theta) \\ &= \left\{ \sqrt{n} I^{1/2}(\theta)(\widehat{\theta}_n - \theta) \right\} \sqrt{\frac{I(\widehat{\theta}_n)}{I(\theta)}} \end{aligned}$$

Первый множитель стремится по распределению к $\mathrm{N}(0,1)$, второй множитель – к единице, ч.т.д.

<u>Теорема</u>. Пусть

$$C_n = \left(\widehat{\theta}_n - z_{\alpha/2}\,\widehat{\mathsf{se}},\ \widehat{\theta}_n + z_{\alpha/2}\,\widehat{\mathsf{se}}\right)$$

Тогда $\mathbb{P}_\theta(\theta \in C_n) \to 1 - \alpha$ **при** $n \to \infty$**.**

<u>Доказательство</u>. Обозначим через Z **- стандартно нормально распределенную случайную величину.**

$$
\begin{aligned}
\mathbb{P}_\theta(\theta \in C_n) &= \mathbb{P}_\theta\left(\widehat{\theta}_n - z_{\alpha/2}\,\widehat{\mathsf{se}} \le \theta \le \widehat{\theta}_n + z_{\alpha/2}\,\widehat{\mathsf{se}}\right) \\
&= \mathbb{P}_\theta\left(-z_{\alpha/2} \le \frac{\widehat{\theta}_n - \theta}{\widehat{\mathsf{se}}} \le z_{\alpha/2}\right) \\
&\to \mathbb{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha
\end{aligned}
$$

29 **Для** $\alpha = .05,\ z_{\alpha/2} = 1.96 \approx 2$**, поэтому**

$$\widehat{\theta}_n \pm 2\ \widehat{\mathsf{se}}$$

приближенно задают границы 95% доверительного интервала

__Пример__. Пусть $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$**. ОМП равна** $\widehat{p}_n = \sum_i X_i/n$**.**

$f(x;p) = p^x(1-p)^{1-x},\ \log f(x;p) = x\log p + (1-x)\log(1-p)$

$$s(X;p) = \frac{X}{p} - \frac{1-X}{1-p}$$

$$-s'(X;p) = \frac{X}{p^2} + \frac{1-X}{(1-p)^2}$$

30

$$I(p) = \mathbb{E}_p(-s'(X;p)) = \frac{p}{p^2} + \frac{(1-p)}{(1-p)^2} = \frac{1}{p(1-p)}$$

$$\widehat{\mathsf{se}} = \frac{1}{\sqrt{I_n(\widehat{p}_n)}} = \frac{1}{\sqrt{nI(\widehat{p}_n)}} = \left\{\frac{\widehat{p}(1-\widehat{p})}{n}\right\}^{1/2}$$

$$\widehat{p}_n \pm 2\left\{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}\right\}^{1/2}$$

приближенно задают границы 95% доверительного интервала

<u>**Пример**</u>**. Пусть** $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$**, где** σ^2 **- известно.**

$$s(X;\theta) = (X - \theta)/\sigma^2$$

$$s'(X;\theta) = -1/\sigma^2$$

$$I_1(\theta) = 1/\sigma^2$$

$$\widehat{\theta}_n = \overline{X}_n$$

$$\overline{X}_n \approx N(\theta, \sigma^2/n) \qquad \textbf{(*)}$$

На самом деле в (*) распределение в точности нормальное.

Пример. Пусть $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Тогда $\widehat{\lambda}_n = \overline{X}_n$ и $I_1(\lambda) = 1/\lambda$. Откуда

$$\widehat{\mathsf{se}} = \frac{1}{\sqrt{nI(\widehat{\lambda}_n)}} = \sqrt{\frac{\widehat{\lambda}_n}{n}}$$

Таким образом,

$$\widehat{\lambda}_n \pm z_{\alpha/2}\sqrt{\widehat{\lambda}_n/n}$$

приближенно задают границы 95% доверительного интервала

<u>Оптимальность и эффективность ОМП</u>

Пусть $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$. **ОПМ для** θ **равна** $\widehat{\theta}_n = \overline{X}_n$.

Обозначим через $\widetilde{\theta}_n$ **- выборочную медиану, которую также можно** **использовать для оценки** θ.

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightsquigarrow N(0, \sigma^2)$$

$$\sqrt{n}(\widetilde{\theta}_n - \theta) \rightsquigarrow N\left(0, \sigma^2 \frac{\pi}{2}\right)$$

В более общем случае, рассмотрим две оценки T_n **и** U_n. **Будем предполагать, что**

$$\sqrt{n}(T_n - \theta) \rightsquigarrow N(0, t^2)$$

$$\sqrt{n}(U_n - \theta) \rightsquigarrow N(0, u^2)$$

Определим асимптотическую относительную эффективность U_n **к** T_n **по формуле** $\mathrm{ARE}(U,T) = t^2/u^2$**. В рассмотренном случае** $\mathrm{ARE}(\widetilde{\theta}_n, \widehat{\theta}_n) = 2/\pi = .63$**. Интерпретация состоит в том, что асимптотическая относительная эффективность показывает долю данных,**

которые используются при оценке параметра.

<u>**Теорема**</u>**. Пусть** $\widehat{\theta}_n$ **- ОМП,** $\widetilde{\theta}_n$ **- какая-то другая оценка. Тогда при соответствующих условиях регулярности (в предположении, что используемая параметрическая модель верна), будет выполнено, что**

$$\mathrm{ARE}(\widetilde{\theta}_n, \widehat{\theta}_n) \le 1$$

Таким образом, ОМП имеет наименьшую асимптотическую дисперсию, то есть ОМП эффективна/асимптотически оптимальна

c) Дельта-метод

Пусть $\tau = g(\theta)$**, где** g **- гладкая функция. ОМП для** τ **равняется** $\widehat{\tau} = g(\widehat{\theta})$**. Каково распределение** $\widehat{\tau}$**?**

35 **<u>Теорема</u>. Если** $\tau = g(\theta)$**, где** g **- дифференцируемая и** $g'(\theta) \neq 0$**, то**

$$\frac{(\widehat{\tau}_n - \tau)}{\widehat{\mathsf{se}}(\widehat{\tau})} \rightsquigarrow N(0,1),$$

где $\widehat{\tau}_n = g(\widehat{\theta}_n)$ **и** $\widehat{\mathsf{se}}(\widehat{\tau}_n) = |g'(\widehat{\theta})|\,\widehat{\mathsf{se}}(\widehat{\theta}_n)$**. Пусть**

$$C_n = \left(\widehat{\tau}_n - z_{\alpha/2}\,\widehat{\mathsf{se}}(\widehat{\tau}_n),\ \widehat{\tau}_n + z_{\alpha/2}\,\widehat{\mathsf{se}}(\widehat{\tau}_n)\right),$$

тогда $\mathbb{P}_\theta(\tau \in C_n) \to 1-\alpha$ **и** $n \to \infty$

Доказательство.

$$\widehat{\tau}_n = g(\widehat{\theta}_n) \approx g(\theta) + (\widehat{\theta}_n - \theta)g'(\theta) = \tau + (\widehat{\theta}_n - \theta)g'(\theta)$$

$$\sqrt{n}(\widehat{\tau}_n - \tau) \approx \sqrt{n}(\widehat{\theta}_n - \theta)g'(\theta)$$

$$\frac{\sqrt{nI(\theta)}(\widehat{\tau}_n - \tau)}{g'(\theta)} \approx \sqrt{nI(\theta)}(\widehat{\theta}_n - \theta)$$

Правая часть последней формулы стремится по распределению к $\mathrm{N}(0,1)$

Следовательно, $\dfrac{\sqrt{nI(\theta)}(\widehat{\tau}_n - \tau)}{g'(\theta)} \rightsquigarrow N(0,1)$**, откуда**

$$\widehat{\tau}_n \approx N\left(\tau, \mathsf{se}^2(\widehat{\tau}_n)\right),$$

где $\mathsf{se}^2(\widehat{\tau}_n) = \dfrac{(g'(\theta))^2}{nI(\theta)}$

Очевидно, что результат не изменится, если θ **заменить на** $\widehat{\theta}_n$ **.**

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$**,** $\psi = g(p) = \log(p/(1-p))$**.**

Информация Фишера равна $I(p) = 1/(p(1-p))$**. Оценка стандартной ошибки для** $\widehat{p}_n$ **равна**

$$\widehat{\mathsf{se}} = \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}$$

ОМП величины ψ **равна** $\widehat{\psi} = \log \widehat{p}/(1-\widehat{p})$**. Так как** $g'(p) = 1/(p(1-p))$**, то в соответствии с дельта-методом**

$$\widehat{\mathsf{se}}(\widehat{\psi}_n) = |g'(\widehat{p}_n)|\widehat{\mathsf{se}}(\widehat{p}_n) = \frac{1}{\sqrt{n\widehat{p}_n(1-\widehat{p}_n)}}$$

Таким образом, границы приближенного 95% доверительного интервала равны

$$\widehat{\psi}_n \pm \frac{2}{\sqrt{n\widehat{p}_n(1-\widehat{p}_n)}}$$

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$**. Допустим, что** μ **известно, а** σ **- неизвестно и необходимо оценить** $\psi = \log \sigma$**. Логарифм функции правдоподобия равен** $\ell(\sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2$**, значит**

$$\widehat{\sigma}_n = \sqrt{\frac{\sum_i (X_i - \mu)^2}{n}}$$

Для того, чтобы подсчитать стандартную ошибку, необходимо знать

информацию Фишера.

$$\log f(X; \sigma) = -\log \sigma - \frac{(X-\mu)^2}{2\sigma^2}$$

Вторая производная логарифма плотности равна

$$\frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}$$

Значит,

$$I(\sigma) = -\frac{1}{\sigma^2} + \frac{3\sigma^2}{\sigma^4} = \frac{2}{\sigma^2}$$

Таким образом, $\widehat{\mathsf{se}} = \widehat{\sigma}_n/\sqrt{2n}$**. Пусть** $\psi = g(\sigma) = \log\sigma$**, тогда** $\widehat{\psi}_n = \log\widehat{\sigma}_n$**. Так как** $g' = 1/\sigma$**, то**

$$\widehat{\mathsf{se}}(\widehat{\psi}_n) = \frac{1}{\widehat{\sigma}_n}\frac{\widehat{\sigma}_n}{\sqrt{2n}} = \frac{1}{\sqrt{2n}}$$

Итак, границы приближенного 95% доверительного интервала равны

$$\widehat{\psi}_n \pm 2/\sqrt{2n}$$

d) Случай векторного параметра (многопараметрический дельта-метод)

Пусть $\theta = (\theta_1, \dots, \theta_k)$ **и** $\widehat{\theta} = (\widehat{\theta}_1, \dots, \widehat{\theta}_k)$ **- ОМП для** θ**,** $\ell_n = \sum_{i=1}^{n} \log f(X_i; \theta)$

Обозначим $H_{jj} = \dfrac{\partial^2 \ell_n}{\partial \theta_j^2}$ **и** $H_{jk} = \dfrac{\partial^2 \ell_n}{\partial \theta_j \partial \theta_k}$

Информационная матрица Фишера имеет вид

$$I_n(\theta) = -\begin{bmatrix} \mathbb{E}_\theta(H_{11}) & \mathbb{E}_\theta(H_{12}) & \cdots & \mathbb{E}_\theta(H_{1k}) \\ \mathbb{E}_\theta(H_{21}) & \mathbb{E}_\theta(H_{22}) & \cdots & \mathbb{E}_\theta(H_{2k}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbb{E}_\theta(H_{k1}) & \mathbb{E}_\theta(H_{k2}) & \cdots & \mathbb{E}_\theta(H_{kk}) \end{bmatrix}$$

Пусть также $J_n(\theta) = I_n^{-1}(\theta)$

__Теорема__. **При соответствующих условиях регулярности**

$$(\widehat{\theta} - \theta) \approx N(0, J_n)$$

Также, если $\widehat{\theta}_j$ **- j-я компонента вектора** $\widehat{\theta}$**, то**

$$\frac{(\widehat{\theta}_j - \theta_j)}{\widehat{\mathsf{se}}_j} \rightsquigarrow N(0,1)$$

41 **где** $\widehat{\mathsf{se}}_j^2 = J_n(j,j)$ **- j-й диагональный элемент матрицы** J_n**. Приближенно ковариация между** $\widehat{\theta}_j$ **и** $\widehat{\theta}_k$ **равна** $\mathsf{Cov}(\widehat{\theta}_j, \widehat{\theta}_k) \approx J_n(j,k)$**.**

Пусть $\tau = g(\theta_1, \ldots, \theta_k)$ **- функция параметра,**

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial \theta_1} \\ \vdots \\ \frac{\partial g}{\partial \theta_k} \end{pmatrix}$$

<u>Теорема</u>. Допустим, что значение ∇g **в точке** $\widehat{\theta}$**, не равно** 0**. Положим** $\widehat{\tau} = g(\widehat{\theta})$**, тогда**

$$\frac{(\widehat{\tau} - \tau)}{\widehat{\mathsf{se}}(\widehat{\tau})} \rightsquigarrow N(0, 1)$$

, где

$$\widehat{\mathsf{se}}(\widehat{\tau}) = \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)}$$

$\widehat{J}_n = J_n(\widehat{\theta}_n)$ **и** $\widehat{\nabla} g$ **равняется значению** ∇g **в точке** $\theta = \widehat{\theta}$**.**

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\tau = g(\mu, \sigma) = \sigma / \mu$
Несложно показать, что

$$I_n(\mu, \sigma) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

Тогда

$$J_n = I_n^{-1}(\mu, \sigma) = \frac{1}{n} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix}$$

Очевидно, что

$$\nabla g = \begin{pmatrix} -\frac{\sigma}{\mu^2} \\ \frac{1}{\mu} \end{pmatrix}$$

Значит,

$$\widehat{\mathsf{se}}(\widehat{\tau}) = \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)} = \frac{1}{\sqrt{n}} \sqrt{\frac{1}{\widehat{\mu}^4} + \frac{\widehat{\sigma}^2}{2\widehat{\mu}^2}}$$

e) Параметрический бутстреп

В непараметрическом случае псевдовыборка $X_1^*, \ldots, X_n^*$ **генерировалась распределенной согласно распределению** $\widehat{F}_n$**. В данном случае псевдовыборка будет генерироваться распределенной с плотностью** $f(x;\ \widehat{\theta}_n)$

44 **<u>Пример</u>. Допустим, то в модели** $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ **получены ОМП. Чтобы подсчитать стандартные ошибки можно поступить следующим образом. Генерируется псевдовыборка** $X_1^*, \ldots, X_n^* \sim N(\widehat{\mu}, \widehat{\sigma}^2)$**, подсчитывается величина** $\widehat{\tau}^* = g(\widehat{\mu}^*, \widehat{\sigma}^*) = \widehat{\sigma}^* / \widehat{\mu}^*$**. Подсчеты повторяются** B **раз, подсчитываются величины** $\widehat{\tau}_1^*, \ldots, \widehat{\tau}_B^*$**. Оценка стандартной ошибки равна**

$$\widehat{\mathsf{se}}_{\mathrm{boot}} = \sqrt{\frac{\sum_{b=1}^{B} (\widehat{\tau}_b^* - \widehat{\tau})^2}{B}}$$

f) Достаточная статистика

Статистика $T(X^n)$ **- сигма-измеримая функция выборки**

Будем писать $x^n \leftrightarrow y^n$ **если** $f(x^n;\theta) = c\, f(y^n;\theta)$ **для некоторой константы** c **, которая может зависеть от** x^n **и** y^n **но не от** θ**. Статистика** $T(x^n)$ **называется достаточной, если из того, что** $T(x^n) \leftrightarrow T(y^n)$**,**

следует, что $x^n \leftrightarrow y^n$ **.**

Грубо говоря, статистика $T(X^n)$ **достаточная, если мы можем подсчитать функцию правдоподобия, зная только значение** $T(X^n)$

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$**. Тогда**

$$\mathcal{L}(p) = p^S(1-p)^{n-S}$$

и статистика $S = \sum_i X_i$ **достаточная**

<u>Пример</u>. Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma)$ **и** $T = (\overline{X}, S)$**. Тогда**

$$f(X^n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{nS^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{X}-\mu)^2}{2\sigma^2}\right\}$$

$$\begin{aligned} T_1(X^n) &= (X_1, \ldots, X_n) \\ T_2(X^n) &= (\overline{X}, S) \\ T_3(X^n) &= \overline{X} \\ T_4(X^n) &= (\overline{X}, S, X_3) \end{aligned}$$

Статистика T является достаточной статистикой минимальной размерности, если а) T - достаточная статистика б) T является функцией любой другой достаточной статистики

<u>Теорема</u>. T является достаточной статистикой минимальной размерности если выполнено: $T(x^n) = T(y^n)$ если и только если $x^n \leftrightarrow y^n$

Значения статистики разбивают множество исходов на классы. О достаточности статистики можно рассуждать в терминах этих классов

48

<u>Пример</u>. $X_1, X_2 \sim Bernoulli(\theta)$. **Пусть** $V = X_1,\ T = \sum_i X_i$ **и** $U = (T, X_1)$.

X_1	X_2	V	T	U
0	0	0	0	(0,0)
0	1	0	1	(1,0)
1	0	1	1	(1,1)
1	1	1	2	(2,1)

$$V \longrightarrow \{(0,0),(0,1)\},\ \{(1,0),(1,1)\}$$

$$T \longrightarrow \{(0,0)\},\ \{(0,1),(1,0)\},\ \{(1,1)\}$$

$$U \longrightarrow \{(0,0)\},\ \{(0,1)\},\ \{(1,0)\},\ \{(1,1)\}$$

V **- не достаточная статистика;** T **и** U **- достаточные статистики;** U **- статистика не минимальной размерности, поскольку если** $x^n = (1,0)$ **и** $y^n = (0,1)$**, то** $x^n \leftrightarrow y^n$**, однако** $U(x^n) \neq U(y^n)$

<u>Пример</u>. Для модели $N(\mu, \sigma^2)$ **статистика** $T = (\overline{X}, S)$ **- минимальная достаточная. Для пуассоновской и бернуллиевской моделей** $T = \sum_i X_i$ **- минимальная достаточная статистика. Статистика** $T = (\sum_i X_i, X_1)$ **- достаточная, но не минимальной размерности. Статистика** $T = X_1$ **- не достаточная.**

<u>Определение</u>. Статистика T **является достаточной, если распределение выборки при заданном значении** $T(X^n) = t$ **не зависит от** θ**, то есть** $f(x_1, \ldots, x_n | t; \theta) = h(x_1, \ldots, x_n, t)$

<u>Пример</u>. $T = \sum_i X_i$ **является достаточной для пуассоновского распределения**

<u>Теорема</u>. T **является достаточной iff найдутся функции** $g(t, \theta)$ **и** $h(x)$**, что** $f(x^n; \theta) = g(t(x^n), \theta) h(x^n)$**.**

<u>Пример</u>. $X = (X_1, X_2) \sim \mathrm{Bernoulli}(p)$, $T = X_1 + X_2$ - **достаточная статистика. Проверка «в лоб»:**

а) $T = 0$

$$P(X_1 = 0, X_2 = 0|t = 0) = 1, P(X_1 = 0, X_2 = 1|t = 0) = 0,$$

$$P(X_1 = 1, X_2 = 0|t = 0) = 0, P(X_1 = 1, X_2 = 1|t = 0) = 0;$$

б) $T = 1$

$$P(X_1 = 0, X_2 = 0|t = 1) = 0, P(X_1 = 0, X_2 = 1|t = 1) = \frac{1}{2},$$

$$P(X_1 = 1, X_2 = 0|t = 1) = \frac{1}{2}, P(X_1 = 1, X_2 = 1|t = 1) = 0;$$

с) $T = 2$

$$P(X_1 = 0, X_2 = 0|t = 2) = 0, P(X_1 = 0, X_2 = 1|t = 2) = 0,$$

$$P(X_1 = 1, X_2 = 0|t = 2) = 0, P(X_1 = 1, X_2 = 1|t = 2) = 1.$$

Проверка с помощью критерия факторизации

Пусть $t = x_1 + x_2$**, тогда**

$$\begin{aligned} f(x_1, x_2; \theta) &= f(x_1; \theta) f(x_2; \theta) \\ &= \theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2} \\ &= g(t,\theta) h(x_1, x_2) \end{aligned}$$

, где

51

$g(t,\theta) = \theta^t (1-\theta)^{2-t}$ **и** $h(x_1, x_2) = 1$

Пусть $\widehat{\theta}$ **- оценка параметра** θ**,** $R(\theta, \widehat{\theta}) = \mathbb{E}_\theta(\theta - \widehat{\theta})^2$ **- MSE**

<u>Теорема</u> (Блекуэлла-Рао-Колмогорова). Пусть $\widehat{\theta}$ **- оценка параметра** θ**,** T **- достаточная статистика. Получим новую оценку параметра** θ **по формуле**

$$\widetilde{\theta} = \mathbb{E}(\widehat{\theta}|T)$$

Тогда для каждого $\theta,\ R(\theta, \widetilde{\theta}) \leq R(\theta, \widehat{\theta})$

52

<u>Пример</u>. $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$**,** $T = \sum_i X_i$

$\widehat{\theta} = X_1$ **- несмещенная, но не состоятельная оценка**

$\widetilde{\theta} = \mathbb{E}(X_1|T) = n^{-1} \sum_i X_i$

g) Экспоненциальное семейство распределений

$\{f(x;\theta)\ :\theta \in \Theta\}$ **- однопараметрической экспоненциальное семейство распределений, если найдутся функции** $\eta(\theta)$**,** $B(\theta),\ T(x)$ **и** $h(x)$**, такие, что**

$$f(x;\theta) = h(x)e^{\eta(\theta)T(x)-B(\theta)}$$

Очевидно, что при этом $T(X)$ **- достаточная статистика**

<u>Пример</u>. Пусть $X \sim \mathrm{Poisson}(\theta)$**, тогда**

$$f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}e^{x\log\theta-\theta}$$

- принадлежит экспоненциальному семейству при $\eta(\theta) = \log\theta$**,** $B(\theta) = \theta$**,** $T(x) = x$

<u>Пример</u>. Пусть $X \sim \text{Binomial}(n, \theta)$

$$f(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{n}{x}\exp\left\{x\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\}$$

$$\eta(\theta) = \log\left(\frac{\theta}{1-\theta}\right), B(\theta) = -n\log(\theta)$$

$$T(x) = x, h(x) = \binom{n}{x}$$

Экспоненциальное семейство можно записать в виде

$$f(x;\eta) = h(x)e^{\eta T(x) - A(\eta)}$$

, где

$$A(\eta) = \log\int h(x)e^{\eta T(x)}dx$$

Пусть $X_1, \ldots, X_n$ - i.i.d. выборка из экспоненциального семейства. Тогда распределение выборки $f(x^n;\theta)$ также будет принадлежать экспоненциальному семейству

$$f(x^n;\theta) = h_n(x^n)h_n(x^n)e^{\eta(\theta)T_n(x^n)-B_n(\theta)}$$

$$h_n(x^n) \; = \; \prod_i h(x_i), \; T_n(x^n) \; = \; \sum_i T(x_i)$$

$$B_n(\theta) \; = \; nB(\theta)$$

Таким образом, статистика $\sum_i T(X_i)$ будет достаточной

Пример. Пусть $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$**. Тогда**

$$f(x^n; \theta) = \frac{1}{\theta^n} I(x_{(n)} \leq \theta)$$

, где $x_{(n)} = max\{x_1, \ldots, x_n\}$**. Таким образом,** $T(X^n) = max\{X_1, \ldots, X_n\}$ **- достаточная статистика. Так как** $T(X^n) \neq \sum_i T(X_i)$**, то равномерное распределение не принадлежит экспоненциальному семейству**

Теорема. Пусть плотность с.в. X **принадлежит экспоненциальному семейству. Тогда**

$$\mathbb{E}(T(X)) = A'(\eta), \ \mathbb{V}(T(X)) = A''(\eta)$$

Если $\theta = (\theta_1, \ldots, \theta_k)$ - вектор параметров, то плотность $f(x;\theta)$ принадлежит экспоненциальному семейству, если

$$f(x;\theta) = h(x) \exp\left\{ \sum_{j=1}^{k} \eta_j(\theta) T_j(x) - B(\theta) \right\}$$

Аналогично, векторная функция $T = (T_1, \ldots, T_k)$ является достаточной статистикой. Простая выборка объема n также будет иметь плотность из экспоненциального распределения, причем достаточная статистика будет иметь вид $\left(\sum_i T_1(X_i), \ldots, \sum_i T_k(X_i)\right)$

<u>Пример</u>. Рассмотрим нормальную плотность с двухмерным параметром $\theta = (\mu, \sigma)$**. В таком случае**

$$f(x;\theta) = \exp\left\{ \frac{\mu}{\sigma^2}x - \frac{x^2}{2\sigma^2} - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\right\}$$

$$\eta_1(\theta) = \frac{\mu}{\sigma^2},\ T_1(x) = x$$

$$\eta_2(\theta) = -\frac{1}{2\sigma^2},\ T_2(x) = x^2$$

$$B(\theta) = \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\right),\ h(x) = 1$$

Статистика $(\sum_i X_i, \sum_i X_i^2)$ **будет достаточной**

Как и раньше, плотность, принадлежащую экспоненциальному семейству, можно записать в виде

$$f(x;\eta) = h(x)\exp\left\{T^T(x)\eta - A(\eta)\right\}$$ **, где**

$$A(\eta) = \log \int h(x) e^{T^T(x)\eta} dx$$

Несложно показать, что

$$\mathbb{E}(T(X)) = \dot{A}(\eta) \quad \mathbb{V}(T(X)) = \ddot{A}(\eta)$$

h) Максимизация функции правдоподобия

Зачастую невозможно в явном виде подсчитать значение $\widehat{\theta}$, на котором правдоподобие достигает максимума => итеративные численные методы <=> последовательность $\theta^0, \theta^1, \ldots$, которая при определенных условиях сходится к $\widehat{\theta}$

<u>Метод Ньютона-Рафсона</u>: разложим по Тейлору функцию правдоподобия вблизи параметра θ^j

$$0 = \ell'(\widehat{\theta}) \approx \ell'(\theta^j) + (\widehat{\theta} - \theta^j)\ell''(\theta^j)$$

$$\widehat{\theta} \approx \theta^j - \frac{\ell'(\theta^j)}{\ell''(\theta^j)}$$

$$\widehat{\theta}^{j+1} = \theta^j - \frac{\ell'(\theta^j)}{\ell''(\theta^j)}$$

В многомерном случае итерационная схема принимает вид

$$\widehat{\theta}^{j+1} = \theta^j - H^{-1}\ell'(\theta^j)$$

$\ell'(\theta^j)$ - вектор первых производных, H - матрица вторых производных логарифма правдоподобия

ЕМ алгоритм: допустим, что имеется выборка Y, логарифм
61 правдоподобия $f(y;\theta)$ которой сложно максимизировать. Однако допустим, что можно найти другую случайную величину Z, что $f(y;\theta) = \int f(y,z;\theta)\,dz$ и что логарифм правдоподобия на основе $f(y,z;\theta)$ максимизировать просто

В таком случае говорят, что Y - наблюдения, Z - латентные (ненаблюдаемые, скрытые) данные => по сути надо восстановить пропущенные данные

<u>Пример</u> (смесь двух нормальных распределений). Пусть $\phi(y;\mu,\sigma)$ **- плотность нормального распределения с параметрами** μ **и** σ**. Плотность смеси двух нормальных распределений имеет вид**

$$f(y;\theta) = (1-p)\phi(y;\mu_0,\sigma_0) + p\phi(y;\mu_1,\sigma_1)$$

$\theta = (\mu_0, \sigma_0, \mu_1, \sigma_1, p)$ **- вектор параметров**

62 **Элемент выборки с вероятностью** p **принадлежит одному нормальному распределению, а с вероятностью** $1-p$ **- другому**

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} [(1-p)\phi(y_i;\mu_0,\sigma_0) + p\phi(y_i;\mu_1,\sigma_1)]$$

Допустим, что имеется «полная» информация $(Y_1, Z_1), \ldots, (Y_n, Z_n)$, где $\mathbb{P}(Z_i = 1) = p$, то есть $Z_i = 0$ означает, что соответствующее наблюдение имеет первую нормальную плотность, а $Z_i = 1$ - вторую

Описание EM-алгоритма:

1. Задать начальное значение θ^0. Для $j = 1, 2, \ldots$ повторять шаги 2 и 3 алгоритма
2. (E-шаг) Подсчитать математическое ожидание по пропущенным данным Z^n, считая что θ^i и Y^n - фиксированы

$$J(\theta|\theta^j) = \mathbb{E}_{\theta^j}\left(\log \frac{f(Y^n, Z^n; \theta)}{f(Y^n, Z^n; \theta^j)} \,\middle|\, Y^n = y^n\right)$$

3. (M-шаг) Найти значение θ^{j+1}, которое максимизирует $J(\theta|\theta^j)$

Каждый шаг EM-алгоритма увеличивает значение функции правдоподобия, т.е. $\mathcal{L}(\theta^{j+1}) \geq \mathcal{L}(\theta^j)$

$$
\begin{aligned}
J(\theta^{j+1}|\theta^j) &= \mathbb{E}_{\theta^j}\left(\log \frac{f(Y^n, Z^n; \theta^{j+1})}{f(Y^n, Z^n; \theta^j)} \,\middle|\, Y^n = y^n\right) \\
&= \log \frac{f(y^n; \theta^{j+1})}{f(y^n; \theta^j)} + \mathbb{E}_{\theta^j}\left(\log \frac{f(Z^n|Y^n; \theta^{j+1})}{f(Z^n|Y^n; \theta^j)} \,\middle|\, Y^n = y^n\right)
\end{aligned}
$$

$$
\frac{\mathcal{L}(\theta^{j+1})}{\mathcal{L}(\theta^j)} = \log \frac{f(y^n; \theta^{j+1})}{f(y^n; \theta^j)}
$$

$$
\begin{aligned}
&= J(\theta^{j+1}|\theta^j) - \mathbb{E}_{\theta^j}\left(\log \frac{f(Z^n|Y^n; \theta^{j+1})}{f(Z^n|Y^n; \theta^j)} \,\middle|\, Y^n = y^n\right) \\
&= J(\theta^{j+1}|\theta^j) + K(f_j, f_{j+1})
\end{aligned}
$$

,

где $f_j = f(y^n;\theta^j)$, $f_{j+1} = f(y^n;\theta^{j+1})$ и

$$K(f,g) = \int f(x)\log(f(x)/g(x))\,dx$$

Так как θ^{j+1} **максимизирует** $J(\theta|\theta^j)$**, то** $J(\theta^{j+1}|\theta^j) \geq J(\theta^j|\theta^j) = 0$

В силу свойства расхождения Кульбака-Лейблера $K(f_j, f_{j+1}) \geq 0$,

откуда $\mathcal{L}(\theta^{j+1}) \geq \mathcal{L}(\theta^j)$ **ч.т.д.**

<u>Пример</u> (продолжение)

Рассмотрим случай, когда $p = 1/2, \sigma_1 = \sigma_2 = 1$

$$f(y;\mu_1,\mu_2) = \frac{1}{2}\phi(y;\mu_0,1) + \frac{1}{2}\phi(y;\mu_1,1)$$

$$f(y_i|Z_i = 0) = \phi(y;\mu_0,1),\ f(y_i|Z_i = 1) = \phi(y;\mu_1,1)$$

$$f(y) = \sum_{z=0}^{1} f(y,z)$$

$$f(z,y) = f(z)f(y|z) = \frac{1}{2}\phi(y;\mu_0,1)^{1-z}\phi(y;\mu_1,1)^z$$

$$\prod_{i=1}^{n}\phi(y_i;\mu_0,1)^{1-z_i}\phi(y_i;\mu_1,1)^{z_i}$$

66

$$\widetilde{\ell} = -\frac{1}{2}\sum_{i=1}^{n}(1-z_i)(y_i-\mu_0) - \frac{1}{2}\sum_{i=1}^{n}z_i(y_i-\mu_1)$$

$$J(\theta|\theta^j) = -\frac{1}{2}\sum_{i=1}^{n}(1-\mathbb{E}(Z_i|y^n,\theta^j))(y_i-\mu_0) - \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}(Z_i|y^n,\theta^j))(y_i-\mu_1)$$

$$\mathbb{E}(Z_i|y^n,\theta^j) = \mathbb{P}(Z_i=1|y^n,\theta^j)$$

$$
\begin{aligned}
\mathbb{P}(Z_i = 1 | y^n, \theta^i) &= \frac{f(y^n | Z_i = 1; \theta^j)\mathbb{P}(Z_i = 1)}{f(y^n | Z_i = 1; \theta^j)\mathbb{P}(Z_i = 1) + f(y^n | Z_i = 0; \theta^j)\mathbb{P}(Z_i = 0)} \\
&= \frac{\phi(y_i; \mu_1^j, 1)\frac{1}{2}}{\phi(y_i; \mu_1^j, 1)\frac{1}{2} + \phi(y_i; \mu_0^j, 1)\frac{1}{2}} \\
&= \frac{\phi(y_i; \mu_1^j, 1)}{\phi(y_i; \mu_1^j, 1) + \phi(y_i; \mu_0^j, 1)} \\
&= \tau(i)
\end{aligned}
$$

Максимизируя $J(\theta|\theta^j)$ **по параметрам** μ_1 **и** μ_2**, получаем, что**

$$\widehat{\mu}_1^{j+1} = \frac{\sum_{i=1}^{n} \tau_i y_i}{\sum_{i=1}^{n} \tau_i}$$

$$\widehat{\mu}_0^{j+1} = \frac{\sum_{i=1}^{n} (1-\tau_i) y_i}{\sum_{i=1}^{n} (1-\tau_i)}$$