Занятие 7. Нижняя оценка Рао-Крамера.

Теорема 1. (Рао-Крамера)**.** *Пусть* $f_X(\beta, t)$ *- плотность распределения случайной величины* X *, где* β *- неизвестный параметр. Пусть* $b(X_1, \ldots, X_n)$ *несмещенная статистика оценки параметра* β *. (При определенных свойствах плотности, которые позволяют дифференцировать ее под знаком интеграла.)*

Тогда, во-первых, $\boxed{D[b(\vec{X}_n)] \geqslant \dfrac{1}{J(\beta)}}$*, где* $J(\beta) = M\left[\left(\dfrac{\partial}{\partial\beta}\ln(L(\beta, \vec{X}_n))\right)^2\right]$*,*
L *- функция правдоподобия распределения* X *;*

во-вторых, равенство достигается тогда и только тогда, когда выполняется равенство: $\pm J(\beta)[\beta - b(\vec{X}_n)] = \dfrac{\partial}{\partial\beta}\ln(L(\beta, \vec{X}_n))$.

Доказательство. 1) $L(\beta, \vec{x}_n) = \prod\limits_{k=1}^{n} f_X(\beta, x_k)$ - совместная плотность вероятности выборки $(X_1, \ldots, X_n)$.
Условие нормировки: $1 = \displaystyle\int_{\mathbb{R}^n} L(\beta, \vec{x}_n)\, dx_1 \ldots dx_n$.

При дифференцировании обеих частей равенства, получим:

$$0 = \frac{\partial 1}{\partial\beta} = \int_{\mathbb{R}^n} \frac{\partial L(\beta, \vec{x}_n)}{\partial\beta}\, dx_1 \ldots dx_n = \int_{\mathbb{R}^n} \frac{\partial\ln(L(\beta, \vec{x}_n))}{\partial\beta}\, L(\beta, \vec{x}_n)\, dx_1 \ldots dx_n =$$
$$= M\left(\frac{\partial\ln(L(\beta, \vec{X}_n))}{\partial\beta}\right).$$

2)Несмещенность означает $\beta = M\{b(\vec{X}_n)\}$. Продифференцируем это равенство по β:

$$1 = \frac{\partial\beta}{\partial\beta} = \int_{\mathbb{R}^n} \frac{\partial\left\{b(\vec{x}_n)L(\beta, \vec{x}_n)\right\}}{\partial\beta}\, dx_1 \ldots dx_n = \int_{\mathbb{R}^n} b(\vec{x}_n)\frac{\partial}{\partial\beta}L(\beta, \vec{x}_n)\, dx_1 \ldots dx_n =$$
$$= \int_{\mathbb{R}^n} b(\vec{x}_n)\frac{\partial lnL(\beta, \vec{x}_n)}{\partial\beta}\, L(\beta, \vec{x}_n)\, dx_1 \ldots dx_n = M\left(b(\vec{X}_n)\frac{\partial\ln(L(\beta, \vec{X}_n))}{\partial\beta}\right).$$

Получили два равенства

$$0 = M\left(\frac{\partial \ln(L(\beta, \vec{X}_n))}{\partial \beta}\right) \quad (1).$$

$$1 = M\left(b(\vec{X}_n)\frac{\partial \ln(L(\beta, \vec{X}_n))}{\partial \beta}\right) \quad (2).$$

3) Составим новое уравнение $(2) - \beta(1) = \ldots$

$$1 = M\left(\left\{b(\vec{X}_n) - \beta\right\}\frac{\partial \ln(L(\beta, \vec{X}_n))}{\partial \beta}\right) =$$

$$= \int_{\mathbb{R}^n} \{b(\vec{x}_n) - \beta\}\,\frac{\partial \ln(L(\beta, \vec{x}_n))}{\partial \beta}\,L(\beta, \vec{x}_n)\,dx_1 \ldots dx_n$$

Полученное справа выражение можно понимать как скалярное произведение по мере $L(\beta, \vec{x}_n) \geqslant 0$ пространства $\mathbb{R}^n$. Применим к полученному неравенство Коши-Буняковского: $(Y, Z)^2 \leqslant (Y, Y)(Z, Z)$. С учетом $1 = (Y, Z) = (Y, Z)^2$, получим

$$1 \leqslant D[b(\bar{X})]J(\beta).$$

Неравенство Коши-Буняковского превращается в равенство, когда сомножители в скалярном произведении линейно зависимы и в этом случае будет выполнено $D[b(\vec{X}_n)] = 1/J(\beta)$. Это происходит когда для некоторой постоянной величины α справедливо равенство

$$\alpha\{b(\vec{x}_n) - \beta\} = \frac{\partial \ln(L(\beta, \vec{x}_n))}{\partial \beta}$$

Используя это соотношение, можно определить, что равно $\alpha = \pm J(\beta)$.

□

7.1. Количество информации в статистике.

Количеством информации по Фишеру называют величину $J(\beta)$. Из приведенной выше теоремы следует, что «наилучшая» статистика для данного закона распределения, если она существует, обратно пропорциональна $J(\beta)$.

7.2. Количество информации в статистике биномиальной случайной величины

Положим, что проводятся испытания по схеме Бернулли и X_k индикатор успеха к k-ом опыте, а всего проводится n опытов. Предварительная задача состоит в том, чтобы написать формулу для величины L, при этом мы считаем вероятность успеха неизвестной величиной и обозначаем ее θ.

$$X_k = \begin{cases} 1, & \text{в } k\text{-ом опыте «успех»;} \\ 0, & \text{в } k\text{-ом опыте «неуспех».} \end{cases}$$

$$L(\theta, X_1, \ldots, X_n) = \prod_{k=1}^{n} \theta^{X_k}(1-\theta)^{1-X_k} = \theta^{\sum\limits_{k=1}^{n} X_k}(1-\theta)^{n-\sum\limits_{k=1}^{n} X_k}.$$

$$\ln[L(\theta, X_1, \ldots, X_n)] = \ln(\theta)\sum_{k=1}^{n} X_k + \ln(1-\theta)[n - \sum_{k=1}^{n} X_k].$$

$$\frac{\partial \ln[L(\theta, X_1, \ldots, X_n)]}{\partial\theta} = \frac{1}{\theta}\sum_{k=1}^{n} X_k - \frac{1}{1-\theta}[n - \sum_{k=1}^{n} X_k] = \frac{1}{\theta(1-\theta)}\sum_{k=1}^{n} X_k - \frac{n}{1-\theta}.$$

$$\left(\frac{\partial \ln[L(\theta, X_1, \ldots, X_n)]}{\partial\theta}\right)^2 = \frac{1}{\theta^2(1-\theta)^2}\left(\sum_{k=1}^{n} X_k\right)^2 - \frac{2n}{\theta(1-\theta)^2}\sum_{k=1}^{n} X_k + \frac{n^2}{(1-\theta)^2}.$$

$$M\left(\frac{\partial \ln[L(\theta, X_1, \ldots, X_n)]}{\partial\theta}\right)^2 = \frac{n\theta(1-\theta) + n^2\theta^2}{\theta^2(1-\theta)^2} - \frac{2n^2\theta}{\theta(1-\theta)^2} + \frac{n^2}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.$$

$$J(\theta) = \frac{n}{\theta(1-\theta)}.$$

Полученное выражение показывает, какой максимум информации может содержаться в статистике (зависящей от $\vec{X}_n$). Видно, что количество информации существенно растет при крайних значениях θ.

Замечание 1. *Мы говорим про какую-то статистику $b(\vec{X}_n)$, при этом подразумеваем, что эта статистика имеет пользу как оценка какого-то неизвестного параметра распределения. Поэтому «статистика» и «оценка» это синонимы.*

Определение 7.1. *Оценку $b(\vec{X}_n)$ (для параметра β), для которой неравенство Рао-Крамера переходит в равенство, называют эффективной. Если она существует, то оптимальна (т.е. не смещена и имеет минимальную дисперсию).*

7.3. ДЗ 1: найти, если это возможно, эффективную статистику для θ.

7.4. ДЗ 2: найти, если это возможно, эффективные статистики для $N(a, \sigma^2)$.