

Математическая статистика.

Андрей Тищенко @AndrewTGk

2024/2025

Семинар 10 января

Задача 1

$x_1, \dots, x_n \sim F_\xi(x)$, найти функцию распределения для $X_{(n)}, X_{(1)}$
 $F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_{(1)} \leq x, \dots, X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) =$
 $= P(X_1 \leq x) \dots P(X_n \leq x) = (F_\xi(x))^n$
 $F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, \dots, X_n > x) =$
 $= 1 - P(X_1 > x, \dots, X_n > x) = 1 - P(X_1 > x) \dots P(X_n > x) = 1 - (1 - F_\xi(x))^n$

Задача 2

$x_1, \dots, x_n \sim R(0, 1)$. Найти $EX_{(n)}, EX_{(1)}$.

$$F_{X_{(n)}}(x) = (F_\xi(x))^n$$

$$f_{X_{(n)}}(x) = (F_{X_{(n)}}(x))' = n(F_\xi(x))^{n-1} \cdot f_\xi(x)$$

$$F_\xi(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0, 1] \\ 1, & x > 1 \end{cases}$$

Подставим в предыдущее уравнение:

$$f_{X_{(n)}} = \begin{cases} 0, & x < 0 \\ nx^{n-1}, & x \in [0, 1] \\ 0, & x > 1 \end{cases}$$

$$EX_{(n)} = \int_{-\infty}^{+\infty} x f_{X_{(n)}}(x) dx = \int_0^1 x n x^{n-1} dx = n \int_0^1 x^n dx = \frac{n}{n+1}$$

Посчитаем для $X_{(1)}$:

$$F_{X_{(1)}}(x) = 1 - (1 - F_\xi(x))^n$$

$$f_{X_{(1)}}(x) = (F_{X_{(1)}}(x))' = n(1 - F_\xi(x))^{n-1} (F_\xi(x))' = n(1 - F_\xi(x))^{n-1} f_\xi(x) = \begin{cases} 0, & x < 0 \\ n(1-x)^{n-1}, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases}$$

$$EX_{(1)} = \int_0^1 x n (1-x)^{n-1} dx = n \int_0^1 x (1-x)^{n-1} dx = \left\langle \begin{matrix} t = 1-x \\ x = 1-t \end{matrix} \right\rangle = -n \int_1^0 (1-t) t^{n-1} dt = n \int_0^1 (1-t) t^{n-1} dt =$$
$$= n \int_0^1 t^{n-1} dt - n \int_0^1 t^n dt = 1 - \frac{n}{n+1}$$

Задача 3

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$E\bar{x} = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = Ex_i$$

$$\mathcal{D}(\bar{x}) = \mathcal{D}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathcal{D}x_i = \frac{\mathcal{D}x_1}{n}$$

Посчитаем выборочную дисперсию:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$ES^2 = E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right) = \frac{1}{n} \sum_{i=1}^n E(x_i - \bar{x})^2 = \mathcal{D}(x_1 - \bar{x}) = \mathcal{D}(x_1) + \mathcal{D}(\bar{x}) - 2 \operatorname{cov}(x_1, \bar{x}) = \frac{(n+1)\mathcal{D}(x_1)}{n} - 2 \operatorname{cov}(x_1, \bar{x})$$

$$\operatorname{cov}(x_1, \bar{x}) = \operatorname{cov}\left(x_1, \frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \operatorname{cov}(x_1, \sum_{i=1}^n x_i) = \frac{1}{n} \operatorname{cov}(x_1, x_1) = \frac{\mathcal{D}(x_1)}{n}$$

Тогда

$$ES^2 = \frac{(n+1)\mathcal{D}(x_1)}{n} - \frac{2\mathcal{D}(x_1)}{n} = \mathcal{D}(x_1) \left(1 - \frac{1}{n}\right)$$

Несмещённая выборочная дисперсия (её математическое ожидание равняется дисперсии x_1):

$$\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Посчитаем дисперсию S^2 :

$$\begin{aligned} \mathcal{D}\left(x_1 - \frac{1}{n} \sum_{i=1}^n x_i\right) &= \mathcal{D}\left(\frac{(n-1)x_1}{n}\right) + \mathcal{D}\left(\frac{1}{n} \sum_{i=2}^n x_i\right) = \frac{(n-1)^2}{n^2} \mathcal{D}(x_1) + \frac{n-1}{n^2} \mathcal{D}(x_1) = \\ &= \mathcal{D}(x_1) \left(\frac{(n-1)(n-1+1)}{n^2}\right) = \mathcal{D}(x_1) \frac{n-1}{n} \end{aligned}$$

Семинар 17 января.

$$T(x_1, x_2, \dots, x_n) = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^n |x_i - m|, x_i \sim N(m, \theta^2)$$

$$ET(x_1, x_2, \dots, x_n) = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^n E|x_i - m| = \sqrt{\frac{\pi}{2}} E|x_1 - m| = \sqrt{\frac{\pi}{2}} \int_{-\infty}^{+\infty} |x - m| \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{(x-m)^2}{2\theta^2}} dx$$

Заменяем $\frac{x-m}{\theta}$ на y

$$\frac{\theta}{2} \int_{-\infty}^{+\infty} |y| \cdot e^{-\frac{y^2}{2}} dy = \theta \int_0^{+\infty} y \cdot e^{-\frac{y^2}{2}} dy = \theta(1 - 0) = \theta$$

$$\frac{1}{n} \sum_{i=1}^n \sqrt{\frac{\pi}{2}} |x_i - m| \xrightarrow[n \rightarrow +\infty]{\text{п. н.}} E \sqrt{\frac{\pi}{2}} |x_i - m|$$

Задача

$$X = (X_1, \dots, X_n), X_i \sim R(0, \theta)$$

$$\hat{\theta} = X_{(n)}, \text{ доказать } \lim_{n \rightarrow \infty} EX_{(n)} = \theta$$

$$F_{X_{(n)}}(x) = (F_{X_i}(x))^n = \left(\frac{x}{\theta}\right)^n$$

$$f_{X_{(n)}}(x) = \frac{dF_{X_{(n)}}}{dx} = \frac{nx^{n-1}}{\theta^n}$$

$$EX_{(n)} = \int_0^\theta \frac{nx^n}{\theta^n} dx = \frac{nx^{n+1}}{(n+1)\theta^n} \Big|_0^\theta = \frac{n}{n+1} \theta \xrightarrow[n \rightarrow \infty]{} \theta. \text{ То есть смещённая, но асимптотически несмещённая.}$$

Докажем состоятельность, хотим:

$$\forall \varepsilon > 0 \quad P(|\hat{\theta} - \theta| < \varepsilon) \xrightarrow[n \rightarrow \infty]{} 1$$

$$P(-\varepsilon < X_{(n)} - \theta < \varepsilon) = F_{X_{(n)}}(\varepsilon + \theta) - F_{X_{(n)}}(\theta - \varepsilon) = 1 - \left(\frac{\theta - \varepsilon}{\theta}\right)^n \xrightarrow[n \rightarrow \infty]{} 1$$

Задача

$$I_n(\theta) = E \left(\frac{\delta \ln f(x, \theta)}{\delta \theta} \right)^2, \quad I_n(\theta) = n I_1(\theta), \quad x_1, \dots, x_n \sim N(\theta, \sigma^2).$$

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

$$\ln f(x, \theta) = \ln \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \right) = -\frac{(x-\theta)^2}{2\sigma^2} + \ln \frac{1}{\sqrt{2\pi}\sigma}$$

$$\frac{\delta \ln f(x, \theta)}{\delta \theta} = -\frac{2(x-\theta)}{2\sigma^2} \cdot (-1) = \frac{x-\theta}{\sigma^2}$$

$$E \left(\frac{x-\theta}{\sigma^2} \right)^2 = \frac{1}{\sigma^4} E(x-\theta)^2 = \frac{1}{\sigma^4} \sigma^2 = \frac{1}{\sigma^2} = I_1(\theta)$$

$$\mathcal{D}\hat{\theta} \geq \frac{1}{n I_1(\theta)} = \frac{\sigma^2}{n} = \mathcal{D}\bar{x}$$

Семинар 24 января

Задача 4 ДЗ

$$\hat{K}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X} + E x_1 - E x_1)(y_i - \bar{Y} + E y_1 - E y_1)$$

$$E \hat{K}_{xy} = E \frac{1}{n} \sum_{i=1}^n ((x_i - E x_1) - (\bar{X} - E x_1)) ((y_i - E y_1) - (\bar{Y} - E y_1)) =$$

$$= E ((x_i - E x_1) - (\bar{X} - E x_1)) \cdot ((y_i - E y_1) - (\bar{Y} - E y_1)) = E ((x_1 - E x_1)(y_1 - E y_1) + (x_1 - E x_1)(\bar{Y} - E y_1) +$$

$$= \text{cov}(x, y) - \frac{1}{n} \text{cov}(x, y) - \frac{1}{n} \text{cov}(x, y) + \frac{1}{n} \text{cov}(x, y)$$

Задача 5 ДЗ

Решал у доски, всем gl.

Задача 1

$X_1, \dots, X_n \sim \Pi(\theta)$. Проверить, что оценка $\hat{\theta} = \bar{X}$ является R-эффективной.

$$E \hat{\theta} = E \frac{1}{n} \sum_{i=1}^n x_i = E x_1 = \theta$$

$$\mathcal{D} \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \theta$$

$$P(\xi = x_1) = \frac{e^{-\theta} \theta^{x_1}}{x_1!}. \text{ Логарифмируем:}$$

$$\ln \frac{e^{-\theta} \theta^{x_1}}{x_1!} = -\theta + x_1 \ln \theta - \ln x_1!$$

Возьмём частную производную:

$$\frac{\delta(-\theta + x_1 \ln \theta - \ln x_1!)}{\delta \theta} = -1 + \frac{x_1}{\theta}$$

Возьмём матожидание квадрата этой величины:

$$E(-1 + \frac{x_1}{\theta})^2 = \frac{1}{\theta^2} E(x_1 - \theta)^2 = \frac{\mathcal{D} x_1}{\theta^2} = \frac{1}{\theta} \Rightarrow I_n(\theta) = \frac{n}{\theta}$$

Попробуем самостоятельно подогнать оценку:

$$U(x, \theta) = \sum_{i=1}^n -1 + \frac{x_i}{\theta} = \frac{1}{\theta} \sum_{i=1}^n (x_i - \theta) = \frac{1}{\theta} (-n\theta + \sum_{i=1}^n \frac{x_i}{n}) = \frac{n}{\theta} (\sum_{i=1}^n (\frac{x_i}{n}) - \theta)$$

$$\hat{\theta} - \theta = a(\theta) U(x, \theta) \Rightarrow a(\theta) = \frac{\theta}{n}, \quad \hat{\theta} = \sum_{i=1}^n \frac{x_i}{n}$$

ДЗ

Задача 1

$$X_1, \dots, X_n \sim N(\theta, \sigma^2) \Rightarrow \forall i = \overline{1, n} \quad f(x_i, \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

$$\ln f(x_i, \theta) = \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{(x-\theta)^2}{2\sigma^2} = \ln \frac{1}{\sqrt{2\pi}\theta} - \frac{x^2}{2\sigma^2} + \frac{\theta x}{\sigma^2} - \frac{\theta^2}{2\sigma^2} \Rightarrow \frac{\delta}{\delta \theta} f(x_i, \theta) = \frac{x}{\sigma} - \frac{\theta}{\sigma^2}$$

$$U(x, \theta) = \sum_{i=1}^n \left(\frac{x_i}{\sigma} - \frac{\theta}{\sigma^2} \right)$$

По критерию эффективности хотим:

$$\hat{\theta} - \theta = \alpha(x)U(x, \theta)$$

Преобразуем: $U(x, \theta) = \left(\sum_{i=1}^n \frac{x_i}{\sigma} \right) - \frac{n\theta}{\sigma^2} \Rightarrow \underbrace{\frac{\sigma^2}{n}}_{\alpha(\sigma)} U(x, \theta) = \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \sigma x_i \right)}_{\hat{\theta}} - \theta$

Задача 2

$$X_1, \dots, X_n \sim N(m, \theta) \Rightarrow f(x_i, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-m)^2}{2\theta}}$$

$$\ln f(x, \theta) = \ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2} \ln \theta - \frac{(x-m)^2}{2\theta} \Rightarrow \frac{\delta}{\delta \theta} f(x, \theta) = -\frac{1}{2\theta} + \frac{(x-m)^2}{2\theta^2}$$

Применим критерий эффективности:

$$\begin{aligned} U(x, \theta) &= \sum_{i=1}^n \left(\frac{(x-m)^2}{2\theta^2} - \frac{1}{2\theta} \right) = \sum_{i=1}^n \left(\frac{(x-m)^2 - \theta}{2\theta^2} \right) = \frac{1}{2\theta^2} \sum_{i=1}^n ((x-m)^2 - \theta) = \\ &= \frac{1}{2\theta^2} \left(\sum_{i=1}^n ((x-m)^2) - n\theta \right) = \frac{n}{2\theta^2} \left(\frac{1}{n} \sum_{i=1}^n ((x-m)^2) - \theta \right) \Rightarrow \underbrace{\frac{2\theta^2}{n}}_{\alpha(\theta)} U(x, \theta) = \underbrace{\left(\frac{1}{n} \sum_{i=1}^n (x-m)^2 \right)}_{\hat{\theta}} - \theta \end{aligned}$$

Задача 3

$X_1, \dots, X_n \sim G(\theta) \Rightarrow Ex = \frac{1}{\theta}$. Проверить оценку $\hat{\theta} = \frac{1}{\bar{X}}$ на несмещённость.

Хотим $E\hat{\theta} = \theta$. Попробуем по определению:

$$E\hat{\theta} = E \frac{n}{\sum_{i=1}^n x_i} = nE \frac{1}{\sum_{i=1}^n x_i}?$$

Для $k = 1$ Попробуем решить через функцию правдоподобия:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P(\xi = x_i, \theta) = \prod_{i=1}^n (1-\theta)^{x_i-1} \theta \approx f(x, \theta)$$

$$\ln f(x_i, \theta) = \ln ((1-\theta)^{x_i-1} \theta) = (x_i-1) \ln(1-\theta) + \ln \theta$$

$$\frac{\delta}{\delta \theta} \ln f(x, \theta) = \frac{1}{\theta} - \frac{x_i-1}{1-\theta} = \frac{1-\theta-\theta x_i+\theta}{\theta-\theta^2} = \frac{1-\theta x_i}{\theta-\theta^2}$$

Применим критерий эффективности:

$$U(x, \theta) = \sum_{i=1}^n \frac{1-\theta x_i}{\theta-\theta^2} = \frac{1}{\theta-\theta^2} \left(n - \theta \sum_{i=1}^n x_i \right) = \frac{n}{\theta-\theta^2} \left(1 - \frac{\theta}{n} \sum_{i=1}^n x_i \right) = \frac{n\bar{X}}{\theta-\theta^2} \left(\frac{1}{\bar{X}} - \theta \right)$$

Значит $\frac{1}{\bar{X}}$ является R-эффективной, то есть несмещённой.

Задача 4

$X_1, \dots, X_n \sim Bi(k, \theta)$. Показать, что $\hat{\theta} = \frac{\bar{X}}{k}$ R-эффективная.

Посчитаем функцию правдоподобия:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P(\xi = x_i, \theta) = \prod_{i=1}^n C_n^k \theta^{x_i} \cdot (1-\theta)^{k-x_i} \approx f(x, \theta)$$

$$\ln f(x_i, \theta) = \ln \frac{n!}{k!(n-k)!} + x_i \ln \theta + (k-x_i) \ln(1-\theta)$$

$$\frac{\delta}{\delta\theta} \ln f(x_i, \theta) = \frac{x_i}{\theta} + \frac{x_i - k}{1 - \theta} = \frac{x_i - \theta x_i + \theta x_i - \theta k}{\theta - \theta^2} = \frac{x_i - \theta k}{\theta - \theta^2}$$

$$I_1(\theta) = E \left(\frac{x_i - \theta k}{\theta - \theta^2} \right)^2 = \int_{-\infty}^{+\infty} \frac{(x - \theta k)^2}{(\theta - \theta^2)^2} C_n^k \theta^x (1 - \theta)^{k-x} dx$$

Задача 5

$$U(x, \theta) = \sum_{i=1}^n \frac{x_i - \theta k}{\theta - \theta^2} = \frac{1}{\theta - \theta^2} \left(-n\theta k + \sum_{i=1}^n x_i \right) = \frac{nk}{\theta - \theta^2} \left(\frac{1}{nk} \sum_{i=1}^n (x_i) - \theta \right) = \frac{nk}{\theta - \theta^2} \left(\frac{\bar{X}}{k} - \theta \right)$$

Получается, что $\frac{\bar{X}}{k}$ является R-эффективной

Семинар 31 января

Задача 1

$$X_1, \dots, X_n \sim f(x, \theta)$$

$$f(x, \theta) = \begin{cases} \frac{2}{\theta} x e^{-\frac{x^2}{\theta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Решал у доски.

Задача 2

$X_1, \dots, X_n \sim R(\theta_1, \theta_2)$, найти оценку максимального правдоподобия.

$$f(x, \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & x \in (\theta_1, \theta_2) \\ 0, & \text{иначе} \end{cases}$$

$$L(x, \theta_1, \theta_2) = \prod_{i=1}^n f(x_i, \theta) = \begin{cases} \left(\frac{1}{\theta_2 - \theta_1} \right)^n, & x_i \in (\theta_1, \theta_2) \\ 0, & \text{иначе} \end{cases}$$

Тогда $\hat{\theta}_1 = X_{(1)}$, $\hat{\theta}_2 = X_{(n)}$.

Попробуем по методу моментов:

$$\begin{cases} \hat{\mu}_1 = \mu_1 \\ \hat{\mu}_2 = \mu_2 = \frac{(\theta_2 - \theta_1)^2}{12} + (\mu_1)^2 \end{cases}$$

Распишем эту систему:

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n x_i = \frac{\theta_1 + \theta_2}{2} \\ \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{(\theta_2 - \theta_1)^2}{12} + \left(\frac{\theta_1 + \theta_2}{2} \right)^2 \end{cases} \Rightarrow \begin{cases} \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\theta_2^2 + \theta_1^2 - 2\theta_1\theta_2}{12} + \frac{\theta_1^2 + \theta_2^2 + 2\theta_1\theta_2}{4} = \frac{1}{3}(\theta_1^2 + \theta_2^2 + \theta_1\theta_2) \end{cases}$$

$$\begin{cases} 2\hat{\mu}_1 = \theta_1 + \theta_2 \\ 3\hat{\mu}_2 = \theta_1^2 + \theta_2^2 + \theta_1\theta_2 \end{cases}$$

Если решать эту систему до конца, можно получить

$$\begin{cases} \hat{\theta}_1 = \bar{X} - \sqrt{3}S \\ \hat{\theta}_2 = \bar{X} + \sqrt{3}S \end{cases}$$

Задача 3

$X_1, \dots, X_n \sim G(\theta)$. Найдём оценку по методу моментов и по методу максимального правдоподобия:

Сначала по методу моментов:

$$\hat{\mu}_1 = \mu_1 = \frac{1}{\theta} \Rightarrow \hat{\theta} = \frac{1}{\bar{X}}$$

Теперь по методу максимального правдоподобия:

$$L(x, \theta) = \prod_{i=1}^n P(\xi = x_i, \theta) = \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} = \theta^n (1 - \theta)^{\sum_{i=1}^n x_i - n}$$

$$\ln L(x, \theta) = n \ln \theta + \left(\sum_{i=1}^n x_i - n \right) \ln(1 - \theta)$$

$$\frac{\delta}{\delta \theta} L(x, \hat{\theta}) = \frac{n}{\hat{\theta}} - \frac{\sum_{i=1}^n x_i - n}{1 - \hat{\theta}} = 0 \Rightarrow \frac{n - n\hat{\theta} - \hat{\theta} \sum_{i=1}^n x_i + n\hat{\theta}}{\hat{\theta} - \hat{\theta}^2} = 0 \Rightarrow$$

$$\Rightarrow n - \hat{\theta} \sum_{i=1}^n x_i = 0 \Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$$

Задача 4

$X_1 \sim Bi(12, p)$, $X_2 \sim Bi(12, p)$, $X_3 \sim Bi(15, p)$. По методу максимального правдоподобия построим оценку p :

$$L(x_1, x_2, x_3, p) = \prod_{i=1}^n P(X_i = x_i) = P(X_1 = 5)P(X_2 = 4)P(X_3 = 4) =$$

$$= C_{12}^5 p^5 (1-p)^7 \cdot C_{12}^4 p^4 (1-p)^8 \cdot C_{15}^4 p^4 (1-p)^{11} = C_{12}^5 \cdot C_{12}^4 \cdot C_{15}^4 \cdot p^{13} \cdot (1-p)^{26}$$

$$\ln L(x_1, x_2, x_3, p) = \ln(C_{12}^5 \cdot C_{12}^4 \cdot C_{15}^4) + 13 \ln p + 26 \ln(1-p)$$

$$\frac{\delta}{\delta p} L(x_1, x_2, x_3, p) = \frac{13}{p} - \frac{26}{1-p} \Rightarrow \frac{13}{\hat{p}} - \frac{26}{1-\hat{p}} = 0 \Rightarrow \hat{p} = \frac{1}{3}$$

ДЗ к семинару 7 января

Задача из учебника №14 стр. 203

Пусть $Z_n = (X_1, \dots, X_n)$ — выборка, соответствующая биномиальному распределению $Bi(10, \theta)$. Оценить неизвестный параметр θ методом максимального правдоподобия.

Построим функцию правдоподобия для вектора (X_1, \dots, X_n) :

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n C_n^{x_i} \cdot \theta^{x_i} (1 - \theta)^{n-x_i}$$

Логарифмируем и дифференцируем по θ полученное произведение:

$$\frac{\delta}{\delta \theta} \ln L(x_1, \dots, x_n, \theta) = \frac{\delta}{\delta \theta} \ln \left(\prod_{i=1}^n C_n^{x_i} \cdot \theta^{x_i} (1 - \theta)^{n-x_i} \right) = \frac{\delta}{\delta \theta} \sum_{i=1}^n \left(\ln (C_n^{x_i} \cdot \theta^{x_i} (1 - \theta)^{n-x_i}) \right) =$$

$$= \frac{\delta}{\delta \theta} \sum_{i=1}^n (\ln C_n^{x_i}) + \frac{\delta}{\delta \theta} \sum_{i=1}^n (x_i \ln \theta) + \frac{\delta}{\delta \theta} \sum_{i=1}^n ((n - x_i) \ln(1 - \theta)) =$$

$$= 0 + \frac{\delta}{\delta \theta} \ln \theta \sum_{i=1}^n x_i + \frac{\delta}{\delta \theta} \ln(1 - \theta) \sum_{i=1}^n (n - x_i) = \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{1}{1 - \theta} \sum_{i=1}^n (n - x_i) =$$

$$= \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{n^2}{1 - \theta} + \frac{1}{1 - \theta} \sum_{i=1}^n x_i = \frac{(1 - \theta)n\bar{x} - \theta n^2 + \theta n\bar{x}}{\theta - \theta^2} =$$

$$= \frac{n\bar{x} - \theta n\bar{x} - \theta n^2 + \theta n\bar{x}}{\theta - \theta^2} = \frac{n\bar{x} - \theta n^2}{\theta - \theta^2} = n \frac{\bar{x} - \theta n}{\theta - \theta^2}$$

Полученную производную стоит приравнять к 0 для поиска точки экстремума. Стоит заметить, что случаи $\theta = 0$ или $\theta = 1$ интереса не представляют и количество испытаний ненулевое, иначе оценивание параметра бессмысленно, поэтому достаточно приравнять к нулю только числитель:

$$n \frac{\bar{x} - \hat{\theta} n}{\hat{\theta} - \hat{\theta}^2} = 0 \Rightarrow \bar{x} - \hat{\theta} n = 0 \Rightarrow \hat{\theta} n = \bar{x} \Rightarrow \hat{\theta} = \frac{\bar{x}}{n}$$

Ответ: ОМП для θ является $\frac{\bar{x}}{n}$.

Задача 2

Выборка X_1, \dots, X_n порождена случайной величиной ξ с плотностью распределения

$$f_{\xi}(x, \theta) = \frac{1}{2} \exp(-|x - \theta|)$$

Построим оценки параметра θ по методу максимального правдоподобия и по методу моментов.

Метод максимального правдоподобия

Построим функцию правдоподобия:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_{\xi}(x_i, \theta) = \prod_{i=1}^n \frac{1}{2} \exp(-|x_i - \theta|) = \frac{1}{2^n} \exp\left(-\sum_{i=1}^n |x_i - \theta|\right)$$

Логарифмируем и продифференцируем по θ :

$$\begin{aligned} \frac{\delta}{\delta\theta} \ln L(x_1, \dots, x_n, \theta) &= \frac{\delta}{\delta\theta} \ln \frac{1}{2^n} \exp\left(-\sum_{i=1}^n |x_i - \theta|\right) = \frac{\delta}{\delta\theta} \ln \frac{1}{2^n} - \frac{\delta}{\delta\theta} \sum_{i=1}^n |x_i - \theta| = \\ &= -\frac{\delta}{\delta\theta} \sum_{i=1}^n |x_i - \theta| = -\sum_{i=1}^n \frac{\delta}{\delta\theta} |x_i - \theta| = -\sum_{i=1}^n g(x_i, \theta) \end{aligned}$$

Где $g(x, \theta) = \begin{cases} -1, & x > \theta \\ 0, & x = \theta, \text{ (производная модуля)} \\ 1, & x < \theta \end{cases}$.

Приравняем производную к нулю:

$$-\sum_{i=1}^n g(x_i, \theta) = 0 \Rightarrow \sum_{i=1}^n g(x_i, \theta) = 0$$

Пусть $\begin{cases} G_{\theta} = \{x \mid x \in (x_1, \dots, x_n) \wedge x > \theta\} \\ E_{\theta} = \{x \mid x \in (x_1, \dots, x_n) \wedge x = \theta\} \\ L_{\theta} = \{x \mid x \in (x_1, \dots, x_n) \wedge x < \theta\} \end{cases}$, тогда

$$\begin{cases} \forall x \in G_{\theta} & g(x, \theta) = -1 \\ \forall x \in E_{\theta} & g(x, \theta) = 0 \\ \forall x \in L_{\theta} & g(x, \theta) = 1 \end{cases} \Rightarrow \sum_{i=1}^n g(x_i, \theta) = (-1) \cdot |G_{\theta}| + 0 \cdot |E_{\theta}| + 1 \cdot |L_{\theta}|$$

Преобразуем:

$$-|G_{\theta}| + 0|E_{\theta}| + |L_{\theta}| = 0 \Rightarrow |G_{\theta}| = |L_{\theta}|$$

То есть количество элементов больше параметра θ в выборке должно совпадать с количеством элементов меньше параметра θ .

$$\text{Получается } \hat{\theta} = \begin{cases} x_{(\lfloor n/2 \rfloor)}, & n \equiv 1 \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2}, & n \equiv 0 \end{cases}$$

Метод моментов

Напишем систему уравнений для моментов (поскольку неизвестный параметр θ единственный, должно хватить одного уравнения):

$$\hat{\mu}_1 = \mu_1(\theta) \Rightarrow \frac{1}{n} \sum_{i=1}^n x_i = E\xi$$

Посчитаем математическое ожидание случайной величины ξ :

$$\begin{aligned} E\xi &= \int_{-\infty}^{+\infty} x f_{\xi}(x, \theta) dx = \int_{-\infty}^{+\infty} x \frac{1}{2} \exp(-|x - \theta|) dx = \left\langle \begin{array}{l} a = x - \theta \\ da = dx \end{array} \right\rangle = \\ &= \frac{1}{2} \int_{-\infty}^{+\infty} (a + \theta) \exp(-|a|) da = \frac{1}{2} \underbrace{\int_{-\infty}^{+\infty} a \exp(-|a|) da}_{=0} + \frac{\theta}{2} \int_{-\infty}^{+\infty} e^{-|a|} da = \\ &= \theta \int_0^{+\infty} e^{-a} da = -\theta \int_0^{+\infty} e^{-a} d(-a) = -\theta e^{-a} \Big|_0^{+\infty} = -\theta(0 - 1) = \theta \end{aligned}$$

Итак, получаем уравнение:

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

Его даже решать не надо, получаем $\hat{\theta} = \overline{X}$.

Ответ:

По методу максимального правдоподобия: $\hat{\theta} = \begin{cases} x_{(\lfloor n/2 \rfloor)}, & n \equiv 1 \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2}, & n \equiv 0 \end{cases}$

По методу моментов: $\hat{\theta} = \overline{X}$

Задача 3

Выборка $X_1, \dots, X_n \sim \Pi(\theta) \Rightarrow \forall i \quad \begin{cases} P(X_i = k) = \frac{e^{-\theta} \theta^k}{k!} \\ EX_i = \theta \end{cases}$. Построим оценки ММ и МП для θ

Метод моментов

Снова неизвестный параметр только один, поэтому достаточно одного уравнения:

$$\frac{1}{n} \sum_{i=1}^n x_i = \theta \Rightarrow \hat{\theta} = \overline{X}$$

Метод максимального правдоподобия

Функция правдоподобия:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = e^{-n\theta} \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!}$$

Логарифм:

$$\ln L(x_1, \dots, x_n, \theta) = -n\theta + \sum_{i=1}^n (x_i \ln \theta - \ln x_i!) = -n\theta + n\overline{X} \ln \theta - \sum_{i=1}^n \ln x_i!$$

Производная по θ

$$\frac{\delta}{\delta \theta} L(x_1, \dots, x_n, \theta) = -n + \frac{n\overline{X}}{\theta}$$

Приравняем к нулю:

$$-n + \frac{n\overline{X}}{\hat{\theta}} = 0 \Rightarrow \hat{\theta} = \overline{X}$$

Ответ: оценки МП и ММ равны \overline{X}

Задача 4

Ученик и тренер стреляют в цель до первого попадания (геометрическое распределение). Известно, что тренер попадает в цель с вероятностью в два раза большей, чем ученик. В ходе соревнования тренер попал в цель при втором выстреле, а ученик — при пятом. Построить ОМП для вероятности попадания учеником в цель при единичном выстреле.

Пусть ξ — количество выстрелов, необходимых тренеру для попадания. Знаем $\xi \sim G(\theta_1)$.

Пусть η — количество выстрелов, необходимых ученику для попадания. Знаем $\eta \sim G(\theta_2)$.

Также знаем, что $\theta_1 = 2\theta_2$.

(Неоднородная выборка???) (ξ, η) получила реализацию $(x_1, x_2) = (2, 5)$. Нужно построить оценку максимального правдоподобия для параметра θ_2 .

Функция правдоподобия:

$$L(x_1, x_2, \theta_1, \theta_2) = P(\xi = 2) \cdot P(\eta = 5) = (1 - \theta_1) \cdot \theta_1 \cdot (1 - \theta_2)^4 \cdot \theta_2 = 2(1 - 2\theta_2) \cdot (1 - \theta_2)^4 \cdot \theta_2^2$$

Логарифмируем:

$$\ln L(x_1, x_2, \theta_1, \theta_2) = \ln 2 + \ln(1 - 2\theta_2) + 4 \ln(1 - \theta_2) + 2 \ln \theta_2$$

Продифференцируем:

$$\begin{aligned} \frac{\delta}{\delta \theta_2} \ln L(x_1, x_2, \theta_1, \theta_2) &= -\frac{2}{1 - 2\theta_2} - \frac{4}{1 - \theta_2} + \frac{2}{\theta_2} = \frac{2(1 - 2\theta_2)(1 - \theta_2) - 4\theta_2 \cdot (1 - 2\theta_2) - 2\theta_2 \cdot (1 - \theta_2)}{(\theta_2 - 2\theta_2^2)(1 - \theta_2)} = \\ &= \frac{2(1 - 3\theta_2 + 2\theta_2^2) - 4(\theta_2 - 2\theta_2^2) - 2(\theta_2 - \theta_2^2)}{\theta_2 - 3\theta_2^2 + 2\theta_2^3} = \frac{2 - 6\theta_2 + 4\theta_2^2 - 4\theta_2 + 8\theta_2^2 - 2\theta_2 + 2\theta_2^2}{\theta_2 - 3\theta_2^2 + 2\theta_2^3} = \\ &= \frac{14\theta_2^2 - 12\theta_2 + 2}{2\theta_2^3 - 3\theta_2^2 + \theta_2} \end{aligned}$$

Приравняем к нулю:

$$\frac{14\hat{\theta}_2^2 - 12\hat{\theta}_2 + 2}{2\hat{\theta}_2^3 - 3\hat{\theta}_2^2 + \hat{\theta}_2} = 0 \Rightarrow 14\hat{\theta}_2^2 - 12\hat{\theta}_2 + 2 = 0 \Rightarrow 7\hat{\theta}_2^2 - 6\hat{\theta}_2 + 1 = 0 \Rightarrow \mathcal{D}' = 9 - 7 = 2 \Rightarrow \begin{cases} \hat{\theta}_2 = \frac{3 + \sqrt{2}}{7} \Rightarrow \theta_1 > 1 \\ \hat{\theta}_2 = \frac{3 - \sqrt{2}}{7} \end{cases}$$

Ответ: $\hat{\theta}_2 = \frac{3 - \sqrt{2}}{7} \approx 0.22654$

Задача 5

Выборка X_1, \dots, X_n порождена случайной величиной X с плотностью распределения:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, & x \in (0, 1) \\ 0, & x \notin (0, 1) \end{cases}$$

Построим оценку максимального правдоподобия для параметра θ и исследуем его на несмещённость. Построим функцию правдоподобия:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \theta^{-n} \prod_{i=1}^n x_i^{\frac{1-\theta}{\theta}}$$

Логарифмируем функцию правдоподобия:

$$\ln L(x_1, \dots, x_n, \theta) = \sum_{i=1}^n \left(\frac{1-\theta}{\theta} \ln x_i \right) - n \ln \theta = \frac{1}{\theta} \sum_{i=1}^n (\ln x_i) - \sum_{i=1}^n (\ln x_i) - n \ln \theta$$

Продифференцируем логарифм по θ :

$$\frac{\delta}{\delta \theta} \ln L(x_1, \dots, x_n, \theta) = -\frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^n \ln x_i = \frac{-n\theta - \sum_{i=1}^n \ln x_i}{\theta^2}$$

Приравняем к нулю:

$$-n\hat{\theta} - \sum_{i=1}^n \ln x_i = 0 \Rightarrow \hat{\theta} = -\frac{1}{n} \sum_{i=1}^n \ln x_i$$

Проверим на несмещённость:

$$\begin{aligned} E\hat{\theta} &= -E \ln x_1 = -\int_0^1 \ln(x) \cdot \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} dx = \left\langle a = x^{\frac{1}{\theta}}, \frac{d}{dx} x^{\frac{1}{\theta}} = \frac{1}{\theta} x^{\frac{1}{\theta}-1} \right\rangle = -\int_{0^{\frac{1}{\theta}}}^{\frac{1}{\theta}} \ln(a^{\theta}) da = \\ &= -\theta \int_0^1 \ln(a) da = -\theta (a \ln a - a) \Big|_0^1 = \theta \end{aligned}$$

Несмещённая.

Семинар 7 февраля

$X_1, \dots, X_n \sim F(x, \theta)$. Считается, что $(T_1(x_1, \dots, x_n), T_2(x_1, \dots, x_n))$ является доверительным интервалом уровня $1 - \alpha$, если:

$$P(T_1(x_1, \dots, x_n) < \theta < T_2(x_1, \dots, x_n)) = 1 - \alpha$$

Например, для $X_1, \dots, X_n \sim N(m, \sigma^2)$, σ известна.

$\hat{m} = \bar{X}$, $\mathcal{D}\bar{X} = \frac{\sigma^2}{n} \Rightarrow \frac{\sqrt{n}(\bar{X}-m)}{\sigma} \sim N(0, 1)$. Для построения доверительного интервала нужно оценить вероятность попадания опорной статистики на интервал:

$$P\left(Z_{\alpha/2} < \frac{\sqrt{n}(\bar{X} - m)}{\sigma} < Z_{1-\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{X} - \frac{\sigma Z_{1-\alpha/2}}{\sqrt{n}} < m < \bar{X} + \frac{\sigma Z_{1-\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha$$

Если σ тоже неизвестна, то подставляем её оценку $\tilde{S} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$ и получаем распределение Стьюдента, значит стоит брать его квантили.

$$\frac{\sqrt{n}(\bar{X} - m)}{\tilde{S}} = \frac{\sqrt{n}(\frac{\bar{X}-m}{\sigma})}{\frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma}\right)^2}$$

То есть стандартное гауссовское делим на корень из χ^2 .

Итого:

$$P\left(\bar{X} - \frac{\tilde{S} t_{1-\alpha/2, n-1}}{\sqrt{n}} < m < \bar{X} + \frac{\tilde{S} t_{1-\alpha/2, n-1}}{\sqrt{n}}\right) = 1 - \alpha$$

Если математическое ожидание известно, но мы хотим интервал для дисперсии:

$$\sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^2} \sim \chi^2(n)$$

$$P\left(\chi_{n, 1-\alpha/2}^2 < \frac{\sum (x_i - m)^2}{\sigma^2} < \chi_{n, 1-\alpha/2}^2\right) = 1 - \alpha$$

$$P\left(\frac{\sum_{i=1}^n (x_i - m)^2}{\chi_{n, 1-\alpha/2}^2} < \sigma^2 < \frac{\sum_{i=1}^n (x_i - m)^2}{\chi_{n, 1-\alpha/2}^2}\right) = 1 - \alpha$$

Если неизвестны оба:

$$\sum_{i=1}^n \frac{(x_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P\left(\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\chi_{n-1, 1-\alpha/2}^2} < \sigma^2 < \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\chi_{n-1, 1-\alpha/2}^2}\right) = 1 - \alpha$$

Задача

Импортёр упаковывает чай в пакеты с номинальным весом 125 грамм. Известно, что упаковочная машина работает с известным среднеквадратическим отклонением 10 грамм. Выбрали 50 пакетов чая, выборочное среднее их веса оказалось равно 125,8.

То есть $n = 50$, $\bar{X} = 125,8$, $X_1, \dots, X_n \sim N(m, 100)$.

$$\bar{X} \sim N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\sqrt{n}(\bar{X}-m)}{\sigma} \sim N(0, 1) \Rightarrow$$

$$P\left(Z_{0,025} < \frac{\sqrt{n}(\bar{X}-m)}{\sigma} < Z_{0,95}\right) = 0,95$$

$$P\left(\bar{X} - \frac{\sigma Z_{0,95}}{\sqrt{n}} < m < \bar{X} + \frac{\sigma Z_{0,95}}{\sqrt{n}}\right) = 0,95$$

$$P(123,028 < m < 128,571) = 0,95$$

125 лежит в этом интервале, поэтому всё хорошо.

Длина интервала получается $\frac{2\sigma Z_{0,95}}{\sqrt{n}}$, хотим, чтобы это равнялось 2

$$\sqrt{n} = \sigma Z_{0,95} \Rightarrow n \approx 384$$

ДЗ 14 февраля

Задача 1

10 изделий сделано за 79, 74, 112, 95, 83, 96, 77, 84, 70, 90 минут. Построить ДИ уровня 0.95 для среднего времени сборки.

Получаем $X_i \sim N(m, \sigma)$, просят доверительный интервал для m . С прошлого семинара:

$$P\left(\bar{X} - \frac{\tilde{S}t_{1-\alpha/2, n-1}}{\sqrt{n}} < m < \bar{X} + \frac{\tilde{S}t_{1-\alpha/2, n-1}}{\sqrt{n}}\right) = 1 - \alpha$$

Здесь:

$$n = 10$$

$$\alpha = 0.05$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

$t_{1-\alpha/2, n-1}$ = так и не понял где посмотреть (квантиль распределения Стьюдента)

Задача 2

Теперь ДИ для дисперсии уровня 0.9, опять воспользуемся записями семинара:

$$P\left(\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\chi_{n-1, 1-\alpha/2}^2} < \sigma^2 < \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\chi_{n-1, 1-\alpha/2}^2}\right) = 1 - \alpha$$

Задача 3

Тоже построить ДИ для математического ожидания и дисперсии гауссовской величины, только с другими значениями. Из сложностей только $\tilde{S}^2 = \frac{n}{n-1} S^2$

Задача 4

Показать, что $S^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2$