Домашнее задание №2 по курсу «Математическая Статистика в Машинном Обучении»

Федор Ерин

Задачи

Задача 1 [4 балла]

Пусть дана некоторая случайная величина X. Определим случайную величину Y, зависящую от X, следующим образом.

$$Y = \begin{cases} 1, & \text{если } X > 0; \\ 0, & \text{если } X \leq 0; \end{cases}$$

т.е. $Y \sim Bernoulli(\psi)$, $\psi = \mathsf{P}(Y = 1)$. Предположим, что $X^n \sim N(\theta, 1)$ и нам дана либо только выборка Y^n, либо X^n (из которой однозначно определяется выборка Y^n). Требуется по наблюдаемой выборке (X^n или Y^n) оценить параметр ψ.

a) **Оценка по выборке X^n.** Найдите MLE-оценку ψ_{MLE} для ψ, выразив ее через MLE-оценку θ_{MLE} для θ.

b) Для оценки ψ_{MLE}, полученной в предыдущем пункте, найдите приближенный 95% доверительный интервал, воспользовавшись дельта-методом.

c) **Оценка по выборке Y^n.** Пусть дана только выборка Y^n. Пусть $\tilde{\psi} = \overline{Y^n} = n^{-1} \sum_{i=1}^{n} Y_i$. Докажите, что $\tilde{\psi}$ является состоятельной оценкой для ψ.

d) Пусть дана выборка X^n и, соответственно, выборка Y^n. Подсчитайте асимптотическую относительную эффективность оценки $\tilde{\psi} = \tilde{\psi}(Y^n)$ по сравнению с оценкой $\psi_{MLE} = \psi_{MLE}(X^n)$. Стандартную ошибку для ψ_{MLE} возьмите из пункта b).

e) Допустим, что на самом деле $X^n \not\sim N(\theta, 1)$. Покажите, что в таком случае ψ_{MLE}, полученная в пункте не является состоятельной оценкой для $P(Y = 1)$. Будет ли, и если ответ "да", то к чему, сходится при $n \to \infty$ оценка ψ_{MLE} в смысле какой-нибудь сходимости?

В ответах можно использовать функцию распределения $\Phi(x)$ стандартной нормальной случайной величины:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Решение

a)

$$\psi = P(Y = 1) = P(X > 0) = 1 - P(X \leq 0) = 1 - P(X - \theta \leq -\theta) = 1 - \Phi(-\theta)$$

Так как $X^n \sim N(\theta, 1)$, то известна MLE-оценка матожидания нормального распределения: $\theta_{MLE} = \overline{X^n}$. Так как MLE-оценка не зависит от параметризации, получаем:

$$\psi_{MLE} = 1 - \Phi(-\theta_{MLE}) = 1 - \Phi(-\overline{X^n}).$$

b) Рассмотрим $\psi = g(\theta) = 1 - \Phi(-\theta)$, где $g'(\theta) = -f(-\theta)$ и $\psi_{MLE} = g(\theta_{MLE})$. Также есть оценка:

$$\hat{se}(\theta_{MLE}) = \hat{se}(\overline{X^n}) = \frac{1}{\sqrt{n}}$$

Согласно дельта-методу:

$$\frac{\hat{\psi_n} - \psi}{\hat{se}(\hat{\psi_n})} \sim N(0,1),$$

где $\hat{\psi_n} = g(\hat{\theta_n}), \hat{se}(\hat{\psi_n}) = |g'(\hat{\theta})| \cdot \hat{se}(\hat{\theta_n})$. Откуда получаем:

$$\hat{\psi}_{MLE} \pm \frac{z_{\alpha/2} f(-\theta_{MLE})}{\sqrt{n}}, \text{ где } \hat{se}(\hat{\psi_n}) = \frac{f(-\theta_{MLE})}{\sqrt{n}}$$

Модуль при $|f(x)|$ можем опустить, так как плотность неотрицательна. При $\alpha = 0.05$ получаем приближенный 95%-доверительный интервал:

$$\big(1 - \Phi(-\overline{X^n}) - \frac{2f(-\overline{X^n})}{\sqrt{n}}, 1 - \Phi(-\overline{X^n}) + \frac{2f(-\overline{X^n})}{\sqrt{n}}\big).$$

c) Покажем, что $\tilde{\psi} = \overline{Y^n}$ - это MLE-оценка параметра ψ распределения Бернулли:

$$L_n(\psi, y) = \prod_{i=1}^{n} \psi^{y_i}(1-\psi)^{1-y_i}$$

$$l_n(\psi, y) = \sum_{i=1}^{n} y_i \cdot \log(\psi) + (1 - y_i) \cdot \log(1 - \psi)$$

$$\frac{\partial l_n(\psi, y)}{\partial \psi} = \frac{\sum_{i=1}^n y_i}{\psi} - \frac{n - \sum_{i=1}^n y_i}{1-\psi} = n\big(\frac{\overline{Y^n}}{\psi} - \frac{1 - \overline{Y^n}}{1-\psi}\big) = 0$$

$$\psi_{MLE} = \overline{Y^n}$$

MLE-оценка обладает свойством состоятельности, следовательно, $\tilde{\psi} = \psi_{MLE}$ состоятельна.

d) Найдем дисперсии оценок:

$$\mathbb{V}\psi_{MLE} = se^2(\psi_{MLE}) = \frac{f^2(-\overline{X^n})}{n}$$

$$\mathbb{V}\tilde{\psi} = \mathbb{V}\frac{\sum_{i=1}^n Y_i}{n} = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}Y_i = \frac{\tilde{\psi}(1-\tilde{\psi})}{n} = \frac{\overline{Y^n}(1-\overline{Y^n})}{n}$$

Асимптотическая относительная эффективность оценки $\tilde{\psi}$ по сравнению с ψ_{MLE} равна:

$$ARE(\tilde{\psi}, \psi_{MLE}) = \frac{\mathbb{V}\psi_{MLE}}{\mathbb{V}\tilde{\psi}} = \frac{f^2(-\overline{X^n})}{\overline{Y^n}(1-\overline{Y^n})}.$$

Ответ:

a) $1 - \Phi(-\overline{X^n})$,

b) $\big(1 - \Phi(-\overline{X^n}) - \frac{2f(-\overline{X^n})}{\sqrt{n}}, 1 - \Phi(-\overline{X^n}) + \frac{2f(-\overline{X^n})}{\sqrt{n}}\big)$,

c) $\tilde{\psi}$ состоятельна,

d) $\frac{f^2(-\overline{X^n})}{\overline{Y^n}(1-\overline{Y^n})}$.

Задача 2 [4 балла]

Пусть n_1 — количество людей, которые получили лечение по методике 1, а n_2 — количество людей, которые получили лечение по методике 2. Обозначим через X_1 — количество людей, получивших лечение по методике 1, на которых эта методика повлияла положительно. Аналогично, обозначим через X_2 — количество людей, получивших лечение по методике 2, на которых эта методика повлияла положительно. Предположим, что $X_1 \sim Binomial(n_1, p_1)$ и $X_2 \sim Binomial(n_2, p_2)$. Положим $\psi = p_1 - p_2$.

(a) Найдите MLE-оценку ψ_{MLE} для параметра ψ.

(b) Найдите информационную матрицу Фишера $I(p_1, p_2)$.

(c) Используя многопараметрический дельта-метод найдите асимптотическую стандартную ошибку для ψ_{MLE}.

(d) Допустим, что $n_1 = n_2 = 200$, и конкретные значения случайных величин X_1 и X_2 равны 160 и 148 соответственно. Чему в этом случае равна оценка ψ_{MLE}. Найдите приблизительный (асимптотический) 90%-ый доверительный интервал для ψ, используя (а) многопараметрический дельта-метод и (б) параметрический бутстреп.

Решение

a) Известна MLE-оценка для параметра p распределения $Binominal(n, p)$:

$$\hat{p} = \frac{X}{n}, \text{ где } X \text{ - кол-во успехов, } n \text{ - кол-во испытаний.}$$

Рассмотрим функцию $\psi = g(p_1, p_2) = p_1 - p_2$. MLE-оценка не зависит от параметризации, следовательно:

$$\psi_{MLE} = g(\hat{p_1}, \hat{p_2}) = \hat{p_1} + \hat{p_2} = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

b) Распишем логарифм правдоподобия:

$$l_n(p, x) = \sum_{i=1}^{2} \log \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}$$

Так как будем брать производную по p, биноминальный коэффициент, зависящий от n_i и x_i, можно сразу опустить:

$$l_n(p, x) = \sum_{i=1}^{2} x_i \log p_i + (n_i - x_i) \log(1 - p_i)$$

Посчитаем компоненты матрицы Фишера:

$$H_{jj} = \frac{\partial^2 l_n}{\partial p_j^2} = -\frac{x_j}{p_j^2} - \frac{n_j - x_j}{(1 - p_j)^2}, \text{ j=\{1,2\}}$$

$$H_{jk} = \frac{\partial^2 l_n}{\partial p_j \partial p_k} = \frac{\partial}{\partial p_j}\left(\frac{x_k}{p_k} - \frac{n_k - x_k}{1 - p_k}\right) = 0, \text{ j,k=\{1,2\}}$$

$$\mathbb{E}_{p_j}(H_{jj}) = -\frac{n_j p_j}{p_j^2} - \frac{n_j - n_j p_j}{(1 - p_j)^2} = -\frac{n_j}{p_j} - \frac{n_j}{1 - p_j} = -\frac{n_j}{p_j(1 - p_j)}$$

Откуда получаем:

$$I(p_1, p_2) = -\begin{bmatrix} \mathbb{E}_{p_1 p_2}(H_{11}) & \mathbb{E}_{p_1 p_2}(H_{12}) \\ \mathbb{E}_{p_1 p_2}(H_{21}) & \mathbb{E}_{p_1 p_2}(H_{22}) \end{bmatrix} = \begin{bmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{bmatrix}.$$

c) Рассмотрим функцию $\psi = g(p_1, p_2) = p_1 - p_2$:

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial p_1} \\ \frac{\partial g}{\partial p_2} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$J_n = I^{-1}(p_1, p_2) = \begin{bmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{bmatrix}.$$

Воспользуемся многопараметрическим дельта-методом:

$$\hat{se}(\hat{\psi}) = \sqrt{(\nabla \hat{g})^T \hat{J}_n (\nabla \hat{g})},$$

куда подставим $\hat{\psi} = \psi_{MLE}$, $\hat{J}_n = J_n(\hat{p})$, $\nabla \hat{g} = \nabla g(\hat{p})$:

$$\hat{se}(\psi_{MLE}) = \sqrt{\begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \begin{bmatrix} \frac{\hat{p_1}(1-\hat{p_1})}{n_1} & 0 \\ 0 & \frac{\hat{p_2}(1-\hat{p_2})}{n_2} \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} =$$

$$= \sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} = \sqrt{\frac{X_1(n_1 - X_1)}{n_1^3} + \frac{X_2(n_2 - X_2)}{n_2^3}}.$$

d) а) Согласно дельта-методу:

$$\frac{\hat{\psi_n} - \psi}{\hat{se}(\hat{\psi_n})} \sim N(0, 1)$$

Откуда получаем 90%-доверительный интервал ($\alpha = 0.1$):

$$\Rightarrow \psi_{MLE} \pm z_{\alpha/2} \cdot \hat{se}(\hat{\psi_n}) \Rightarrow$$

$$\Rightarrow (\frac{X_1}{n_1} - \frac{X_2}{n_2}) \pm 1.645 \cdot \sqrt{\frac{X_1(n_1 - X_1)}{n_1^3} + \frac{X_2(n_2 - X_2)}{n_2^3}} \Rightarrow$$

$$\Rightarrow 0.06 \pm 1.645 \cdot 0.042 \Rightarrow$$

$$\Rightarrow C_n = (-0.009, 0.129)$$

б) Сделаем расчет параметрического бутстрепа:

```python
n1 = n2 = 200
X1 = 160
X2 = 148
alpha = 0.1

# MLE оценки параметров распределений Bin(n1,p1) и Bin(n2,p2)
p1_est = X1 / n1
p2_est = X2 / n2

# Псевдовыборки из Bin(n1,p1_est) и Bin(n2,p2_est)
B = 1000
statistics = []
for i in range(B):
    p1_bootstrap = scipy.stats.binom.rvs(n1, p1_est, size=1)[0] / n1
    p2_bootstrap = scipy.stats.binom.rvs(n2, p2_est, size=1)[0] / n2
    statistics.append(p1_bootstrap - p2_bootstrap)
psi_boot = np.mean(statistics)
se_boot = np.sqrt(((np.array(statistics) - psi_boot)**2).sum() / B)
z = abs(stats.norm().ppf(alpha / 2))

print(f'Бутстреп оценка psi: {psi_boot:.4f}')
print(f'Бутстреп оценка se: {se_boot:.4f}')
print(f'Квантиль норм. распр. для alpha={alpha}: {z:.4f}')
print(f'{int(100*(1-alpha))}%-доверительный интервал: ({psi_boot - z*se_boot:.4f}, \
{psi_boot + z*se_boot:.4f})')
```

```
Бутстреп оценка psi: 0.0579
Бутстреп оценка se: 0.0406
Квантиль норм. распр. для alpha=0.1: 1.6449
90%-доверительный интервал: (-0.0090, 0.1247)
```

Ответ:

a) $\frac{X_1}{n_1} - \frac{X_2}{n_2}$,

b) $\begin{bmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{bmatrix}$,

c) $\sqrt{\frac{X_1(n_1-X_1)}{n_1^3} + \frac{X_2(n_2-X_2)}{n_2^3}}$,

d)

- а) $\hat{\psi}_\Delta = 0.06, C_n = (-0.009, 0.129)$,
- б) $\hat{\psi}_B = 0.058, C_n = (-0.009, 0.125)$.

Задача 3 [2 балла]

Пусть $X^n = \{X_1, \ldots, X_n\} \sim Poisson(\lambda)$.

- Постройте оценки параметра λ с помощью метода моментов с использованием пробных функций $g_1(x) = x$ и $g_2(x) = x^2$.
- Постройте оценку $\hat{\lambda}$ параметра λ с помощью метода максимального правдоподобия. Найдите информацию Фишера $I_X(\lambda)$. Является ли оценка $\hat{\lambda}$ эффективной?

Решение

- $\mathbb{E}g_1(x) = \mathbb{E}X = \lambda, \quad \overline{g_1(x)} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X^n} \quad \Rightarrow \quad \hat{\lambda} = \overline{X^n}$
  $\mathbb{E}g_2(x) = \mathbb{E}X^2 = \mathbb{V}X + (\mathbb{E}X)^2 = \lambda + \lambda^2, \quad \overline{g_2(x)} = \frac{1}{n}\sum_{i=1}^n X_i^2 = \overline{X^2}$
  Возьмем в качестве оценки $\hat{\lambda}$ решение уравнения:

$$\lambda^2 + \lambda - \overline{X^2} = 0$$

$$D = 1 + 4\overline{X^2}$$

$$\lambda = \frac{-1 \pm \sqrt{1 + 4\overline{X^2}}}{2}$$

$$\hat{\lambda} = \frac{\sqrt{4\overline{X^2} + 1} - 1}{2}$$

- Распишем логарифм правдободобия и найдем точку максимума:

$$l_n(\lambda, x) = \sum_{i=1}^{n} \log(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}) = \sum_{i=1}^{n}(-\lambda + x_i \log\lambda - \log x_i!) = -n\lambda + \log\lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log x_i!$$

$$\frac{\partial l_n(\lambda, x)}{\partial\lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

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

Это точка максимума, так как:

$$\frac{\partial^2 l_n(\lambda, x)}{\partial\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} \leq 0$$

Найдем информацию Фишера:

$$I_n(\lambda) = \sum_{i=1}^{n} \mathbb{V}\big(\frac{\partial \log f(\lambda, x)}{\partial\lambda}\big) = \sum_{i=1}^{n} \mathbb{V}\big(\frac{\partial}{\partial\lambda}(-\lambda + x\log\lambda - \log x!)\big) = \sum_{i=1}^{n} \mathbb{V}(-1 + \frac{x}{\lambda}) =$$

$$= \frac{1}{\lambda^2} \sum_{i=1}^{n} \mathbb{V}x = \frac{1}{\lambda^2}\lambda n = \frac{n}{\lambda} = nI_X(\lambda) \Rightarrow I_X(\lambda) = \frac{1}{\lambda}$$

Проверим, является ли оценка $\hat{\lambda}$ эффективной, для этого посчитаем $\mathbb{V}\hat{\lambda}$ и воспользуемся неравенством Рао-Крамера:

$$\mathbb{V}\hat{\lambda} = \mathbb{V}\overline{X^n} = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}X_i = \frac{1}{n^2}\lambda n = \frac{\lambda}{n}$$

$$\mathbb{V}\hat{\lambda} \geq \frac{1}{I_X(\lambda) \cdot n} \Rightarrow \frac{\lambda}{n} \geq \frac{1}{\frac{1}{\lambda} \cdot n} = \frac{\lambda}{n}$$

Равенство соблюдается, следовательно оценка эффективна.

Ответ:

- $\hat{\lambda}_{g_1(x)} = \overline{X^n}, \hat{\lambda}_{g_2(x)} = \frac{\sqrt{4\overline{X^2}+1}-1}{2}$
- $\hat{\lambda} = \overline{X^n}, I_X(\lambda) = \frac{1}{\lambda}$, оценка эффективна.

Задача 4 [4 балла]

Пусть $X^n \sim Pareto(\theta, \nu)$, $\theta > 0$, $\nu > 0$, с функцией плотности

$$f_{\theta,\nu}(x) = \begin{cases} \frac{\theta \nu^\theta}{x^{\theta+1}}, & x \geq \nu, \\ 0, & x < \nu \end{cases}$$

a) Найдите MLE-оценки $\hat{\theta}$ и $\hat{\nu}$ для параметров θ и ν.

b) Пусть параметр ν известен. Найдите асимптотическое распределение оценки $\hat{\theta}$ с помощью дельта-метода.

c) Пусть параметр ν известен. Найдите истинные значения $\mathbb{E}_\theta[\hat{\theta}]$ и $\mathbb{V}_\theta[\hat{\theta}]$ как функции параметров θ, ν и размера выборки n. *Подсказка: следует использовать тот факт, что логарифм от случайной величины с распределением Парето, имеет экспоненциальное распределение.*

d) Пусть параметр ν известен. Найдите информацию Фишера $I_X(\lambda)$. Является ли MLE-оценка параметра $\hat{\theta}$ эффективной?

Решение

a) Распишем логарифм правдободобия и найдем оценки для θ и ν:

$$l_n(\theta, \nu, x) = \sum_{i=1}^{n} \log \theta + \theta \log \nu - (\theta + 1) \log x_i$$

$$\frac{\partial l_n(\theta, \nu, x)}{\partial \theta} = \frac{n}{\theta} + n \log \nu - \sum_{i=1}^{n} \log x_i = 0$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log x_i - n \log \nu}$$

Для оценки ν заметим, что $l_n(\theta, \nu, x)$ монотонно возрастает с ростом ν, убывает с ростом x, а $f_{\theta,\nu}(x)$ неотрицательна при $x \geq \nu$, откуда:

$$\hat{\nu} = X_{(1)}$$

b) Рассмотрим $\tau = g(\theta) = \theta$, согласно дельта-методу:

$$\frac{\hat{\tau} - \tau}{\hat{se}(\hat{\tau})} \sim N(0, 1),$$

причем

$$\hat{\tau} = g(\hat{\theta}) = \hat{\theta}$$

$$\hat{se}(\hat{\tau}) = |g'(\hat{\theta})| \cdot \hat{se}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}}$$

Тогда имеем:

$$\frac{\hat{\theta} - \theta}{\hat{\theta}/\sqrt{n}} \sim N(0, 1)$$

$$\hat{\theta} \sim N(\theta, \frac{\theta^2}{n})$$

c) Рассмотрим оценку:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log x_i - n \log \nu}$$

Так как $X \sim Pareto(\theta, \nu)$, заметим, что:

$$\log \frac{X}{\nu} \sim Exp(\frac{1}{\theta}) \sim \Gamma(1, \frac{1}{\theta}),$$

где у Гамма-распределения $shape = 1, scale = 1/\theta$. Тогда имеем:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log x_i - n \log \nu} \sim \text{Inv-}\Gamma(n, n\theta),$$

где Inv-$\Gamma(\alpha, \beta)$ = Inv-$\Gamma(n, n\theta)$ - обратное Гамма-распределение, у которого известны матожидание и дисперсия:

$$\mathbb{E}\hat{\theta} = \frac{\beta}{\alpha - 1} = \frac{n\theta}{n-1}$$

$$\mathbb{V}\hat{\theta} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

d) Найдем информацию Фишера $I_X(\theta)$:

$$\log f_{\theta,\nu}(x) = \log \theta + \theta \log \nu - (\theta + 1) \log x$$

$$\frac{\partial f_{\theta,\nu}(x)}{\partial \theta} = \frac{1}{\theta} + \log \nu - \log x$$

$$\frac{\partial^2 f_{\theta,\nu}(x)}{\partial \theta^2} = -\frac{1}{\theta^2}$$

$$I_X(\theta) = -\mathbb{E}(\frac{\partial^2 f_{\theta,\nu}(x)}{\partial \theta^2}) = -\mathbb{E}(-\frac{1}{\theta^2}) = \frac{1}{\theta^2}$$

Известно, что

$$\hat{\theta}_{MLE} \sim N\big(\theta, \frac{1}{nI_X(\theta)}\big) = N(\theta, \frac{\theta^2}{n}) \Rightarrow \mathbb{V}\hat{\theta}_{MLE} = \frac{\theta^2}{n}$$

Подставим $\mathbb{V}\hat{\theta}_{MLE}$ и $I_X(\theta)$ в неравенство Рао-Крамера:

$$\mathbb{V}\hat{\theta}_{MLE} \geq \frac{1}{I_X(\theta) \cdot n} \Rightarrow \frac{\theta^2}{n} \geq \frac{1}{\frac{1}{\theta^2} \cdot n} = \frac{\theta^2}{n}$$

Равенство соблюдается, следовательно оценка эффективна.

Ответ:

a) $\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log x_i - n \log \nu}, \hat{\nu} = X_{(1)}$

b) $\hat{\theta} \sim N(\theta, \frac{\theta^2}{n})$

c) $\mathbb{E}\hat{\theta} = \frac{n\theta}{n-1}, \mathbb{V}\hat{\theta} = \frac{n^2\theta^2}{(n-1)^2(n-2)}$

d) $I_X(\theta) = \frac{1}{\theta^2}$, MLE-оценка эффективна.

Задача 5 [2 балла]

Пусть $X^n \sim Uniform[0, \theta]$. Постройте байесовскую оценку параметра θ, если параметр θ имеет

a) плотность $q(t) = 1/t^2$ при $t > 1$;

b) равномерное распределение на отрезке $[0, 1]$.

Решение

a) Плотность распределения:

$$f(t, X_1, \ldots, X_n) = f_t(X_1, \ldots, X_n)q(t)$$

Байесовская оценка параметра θ:

$$\theta_n^* = \int_\Theta tq(t|X_1, \ldots, X_n)dt$$

Апостериорная плотность равна:

$$q(t|X_1, \ldots, X_n) = \frac{f_t(X_1, \ldots, X_n)q(t)}{\int_\Theta f_s(X_1, \ldots, X_n)q(s)ds}$$

Для $Uniform[0, \theta]$ имеем:

$$f_t(X_1, \ldots, X_n) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{I}[X_i \in [0, \theta]] = \frac{1}{\theta^n}\mathbb{I}[X_{(1)} > 0, X_{(n)} < \theta]$$

Плотность $q(t|X_1, \ldots, X_n)$ пропорциональна (как функция от t) произведению $q(t)f_t(X_1, \ldots, X_n)$, то есть пропорциональна:

$$\frac{1}{t^{n+2}}\mathbb{I}[X_{(1)} > 0, X_{(n)} < \theta]$$

Так как $t > 1$, плотность $q(t|X_1, \ldots, X_n)$ соответствует равномерному распределению $Uniform[0, \hat{\theta}]$ с параметром $\hat{\theta} = \frac{n+1}{n}max(X_{(n),1})$, искомая оценка равна:

$$\theta_n^* = \int_\Theta tq(t|X_1, \ldots, X_n)dt = \frac{n+1}{n}max(X_{(n),1})$$

Ответ:
a) $\frac{n+1}{n} \cdot max(X_{(n)}, 1)$

Задача 6 [1 балла]

Пусть $X^n \sim Exponential(\theta)$. Постройте тест на основе критерия отношения правдоподобий для проверки гипотезы $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.

Решение

- Функция правдоподобия равна:

$$L(\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

Откуда известна MLE-оценка для θ:

$$\hat{\theta} = \frac{1}{\overline{X_n}}$$

Статистика отношения правдоподобия:

$$\lambda = 2\log\frac{\sup_{\theta>\theta_0} L(\hat{\theta})}{\sup_{\theta=\theta_0} L(\hat{\theta}_0)} = 2\log\frac{\theta_0^n e^{-\theta_0\sum_{i=1}^n x_i}}{(\frac{1}{\overline{X_n}})^n e^{-\frac{1}{\overline{X_n}}\sum_{i=1}^n x_i}} = 2\log(\theta_0^n e^{-\theta_0\sum_{i=1}^n x_i}\cdot\overline{X_n}^n e^n) =$$

$$= 2\cdot(n\log\theta_0 - \theta_0\sum_{i=1}^n x_i + n\log\overline{X_n} + n) = 2n\cdot(\log\theta_0 - \theta_0\overline{X_n} + \log\overline{X_n} + 1)$$

При справедливости гипотезы H_0:

$$\lambda \sim \chi^2_1,$$

следовательно отклоняем H_0, если

$$\lambda = 2n\cdot(\log\theta_0 - \theta_0\overline{X_n} + \log\overline{X_n} + 1) > c,$$

где $c = \chi^2_{1,\alpha}$. Критерий имеет асимптотический уровень значимости α:

$$\mathbb{P}_{\theta=\theta_0}(\lambda > c) \le \alpha$$

Ответ: $\lambda = 2n\cdot(\log\theta_0 - \theta_0\overline{X_n} + \log\overline{X_n} + 1) > \chi^2_{1,\alpha}$

Задача 7 [3 балла]

Пусть $X^n \sim Uniform(\theta, \theta+1)$. Необходимо протестировать гипотезу $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$. В данном случае нельзя использовать тест Вальда, так как оценки θ при $n \to \infty$ не сходятся к нормальному распределению. Будем использовать следующее правило: гипотеза H_0 отвергается, если $X_{(n)} \ge 1$ или $X_{(1)} \ge c$, где c — некоторая константа, $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(a) Найдите функцию мощности для данного теста. *Внимание. В данной задаче вид функции мощности $\beta(\theta; c)$ зависит от значения c. Поэтому ответ должен быть полным в том смысле, что для любого $c \in \mathbb{R}$ в ответе должна быть указана соответствующая $\beta(\theta; c)$.*

(b) При каком значении параметра c размер теста будет равен 0.05?

(c) Найдите такое минимальное $n \ge 1$, что при $\theta = 0.1$ и размере теста 0.05 мощность критерия не меньше 0.8.

Решение

a) Функция мощности критерия с критической областью $R = \{x : T(x) > c\}$:

$$\beta(\theta; c) = \mathbb{P}_\theta(X \in R) = \mathbb{P}_\theta(X_{(n)} \ge 1, X_{(1)} \ge c) = 1 - \mathbb{P}_\theta(X_{(n)} < 1, X_{(1)} < c)$$

Так как $X^n \sim Uniform(\theta, \theta+1)$, то должно быть $c > \theta$ и $\theta < 1$, иначе $\beta = 1$ и H_0 всегда будет отвергаться. Следовательно, $\theta < min(c, 1)$.

$$\beta_{\theta<min(c,1)}(\theta; c) = 1 - \mathbb{P}_\theta(X_{(n)} < 1) + \mathbb{P}_\theta(c < X_{(1)}, \ldots, X_{(n)} < 1) =$$

$$= \begin{cases} 1 - (1-\theta)^n + (1-c)^n, & c \in (\theta, 1] \\ 1 - (1-\theta)^n, & c > 1 \end{cases}$$

Итого:

$$\beta(\theta; c) = \begin{cases} 1 - (1-\theta)^n + (1-c)^n, & c \in (\theta, 1] \\ 1 - (1-\theta)^n, & c > 1 \\ 1, & \text{otherwise} \end{cases}$$

b) Определим значение c, при котором размер теста равен 0.05:

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta=0} \beta(\theta) = (1-c)^n = 0.05$$

$$c = 1 - 0.05^{1/n}$$

c) Мощность критерия равна $1-\beta$. Составим набор условий для поиска нужного значения n:

$$\begin{cases} n \geq 1, \\ \theta = 0.1, \\ \sup_{\theta=0.1} \beta(\theta) = 0.05, \\ 1 - \beta \geq 0.8. \end{cases} \tag{1}$$

$$\begin{cases} n \geq 1, \\ \theta = 0.1, \\ c = 1 - 0.05^{1/n}, \\ 1 - 0.9^n + 0.05 \geq 0.8. \end{cases} \tag{2}$$

Откуда получаем:

$$0.9^n \leq 0.25 \Rightarrow n \geq \log_{0.9} 0.25 \approx 13.2 \Rightarrow n_{min} = 14$$

Ответ:

a) $\beta(\theta; c) = \begin{cases} 1 - (1-\theta)^n + (1-c)^n, & c \in (\theta, 1] \\ 1 - (1-\theta)^n, & c > 1 \\ 1, & \text{otherwise} \end{cases}$

b) $c = 1 - 0.05^{1/n}$,

c) $n_{min} = 14$.

Задача 8 [2 балла]

Пусть X^n — выборка н.о.р. с.в. со следующей функцией плотности:

$$f(x, \theta) = \begin{cases} c(\theta)d(x), & a \leq x \leq b(\theta) \\ 0, & \text{иначе} \end{cases}$$

где $b(\theta)$ — монотонно возрастающая функция одного аргумента.

(a) Построить статистику отношения правдоподобий λ для тестирования гипотезы $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

(b) Найти распределение статистики λ при выполнении H_0 для следующей функции плотности:

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 \leq x \leq \theta \\ 0, & \text{иначе} \end{cases}$$

Решение

b) Напишем логарифм правдоподобия с.в. с плотностью $f(x, \theta)$:

$$l_n(x, \theta) = \sum_{i=1}^{n} \log \frac{2x_i}{\theta^2} \qquad \Rightarrow \qquad \frac{\partial l_n(x, \theta)}{\partial \theta} = \sum_{i=1}^{n} -\frac{2}{\theta} < 0$$

Максимум правдоподобия достигается при $\hat{\theta} = X_{(n)}$. Получаем отношение правдоподобий:

$$\lambda = 2\log\frac{\theta_0^{2n}}{X_{(n)}^{2n}} = 4n(\log\theta_0 - \log X_{(n)})$$

Найдем плотность распределения $X_{(n)}$:

$$F_{X_{(n)}}(x) = F^n(x) = \Big(\int_0^x \frac{2t}{\theta_0^2}dt\Big)^n = \frac{x^{2n}}{\theta_0^{2n}}, \qquad 0 \le x \le \theta$$

$$f_{X_{(n)}}(x) = 2n\frac{x^{2n-1}}{\theta_0^{2n}}, \qquad 0 \le x \le \theta$$

Рассмотрим λ, как функцию $g(x)$, это монотонное обратимое преобразование:

$$g^{-1}(x) = \theta_0 e^{-\frac{x}{4n}}$$

Теперь найдем распределение λ:

$$f_\lambda(x) = |(g^{-1})'(x)|f_{X_{(n)}}(g^{-1}(x)) = \frac{\theta_0}{4n}e^{-\frac{x}{4n}}\frac{2n(\theta_0 e^{-\frac{x}{4n}})^{2n-1}}{\theta_0^{2n}} = \frac{1}{2}(e^{-\frac{x}{4n}})^{2n} = \frac{e^{-x/2}}{2}$$

Итого:

$$f_\lambda(x) = \begin{cases} \frac{e^{-x/2}}{2}, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$$

Ответ:

b) $f_\lambda(x) = \begin{cases} \frac{e^{-x/2}}{2}, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases}$

Задача 10 [2 балла]

В десятичной записи числа π среди первых 10002 знаков после запятой цифры $0, 1, \ldots, 9$ встречаются соответственно 968, 1026, 1021, 974, 1014, 1046, 1021, 970, 948, 1014 раз. Можно ли при уровне значимости 0.05 считать эти цифры случайными? При каком уровне значимости эта гипотеза отвергается?

Решение

– Пусть $X = (X_0, \ldots, X_9)$ — мультиномиальное распределение, где X_i - независимые одинаково распределённые случайные величины, такие, что их распределение задаётся функцией вероятности:

$$\mathbb{P}(X_i = j) = p_j,\ j = 0, \ldots, 9$$

Цифры в числе π можно считать случайными, если появление их равновероятно, то есть они имеют распределение X и $p_0 = (p_{00}, \ldots, p_{09})$, где $p_{00} = \cdots = p_{09} = 0.1$. Сформулируем критерий:

$$H_0 : p = p_0 \text{ vs. } H_1 : p \ne p_0$$

Применим χ^2 тест, в котором при гипотезе H_0 данные распределены с заданной частотой:

```
n_digits = 10
digits_occurencies = [968, 1026,1021, 974, 1014, 1046, 1021, 970, 948, 1014]
N = sum(digits_occurencies)
expected_occurencies = [N/n_digits for _ in range(n_digits)]
```

```
from scipy.stats import chisquare

chisquare(digits_occurencies, f_exp=expected_occurencies)
```

```
Power_divergenceResult(statistic=9.367726454709057, pvalue=0.40404520751503087)
```

Статистика χ^2 Пирсона равна 9.36, а $p-value$ равно 0.4, средовательно не отвергаем H_0 и считаем цифры случайными на уровне значимости 0.05. Гипотезу можем отвергнуть при $\alpha = 0.4$.

Ответ: цифры случайны при $\alpha = 0.05$, отвергнем H_0 при $\alpha = 0.4$.

Задача 11 [2 балла]

Предположим, что у нас есть 10 статей, написанных автором, скрывающемся под псевдонимом. Мы подозреваем, что эти статьи на самом деле написаны некоторым известным писателем. Чтобы проверить эту гипотезу, мы подсчитали доли четырехбуквенных слов в 8-и сочинениях подозреваемого нами автора:

.224 .261 .216 .239 .229 .228 .234 .216

В 10 сочинениях, опубликованных под псевдонимом, доли четырехбуквенных слов равны

.207 .204 .195 .209 .201 .206 .223 .222 .219 .200

– Используйте критерий Вальда. Найдите $p-value$ и 95%-ый доверительный интервал для разницы средних значений. Какой вывод можно сделать исходя из найденных значений?
– Используйте критерий перестановок. Каково в этом случае значение . Какой вывод можно сделать?

Решение

– Сделаем расчет для критерия Вальда:

```
alpha = 0.05
X = np.array([.224, .261, .216, .239, .229, .228, .234, .216])
Y = np.array([.207, .204, .195, .209, .201, .206, .223, .222, .219, .200])

theta = 0. # H0: различий нет
theta_est = X.mean() - Y.mean()
se_est = np.sqrt(np.std(X, ddof=0)**2/X.size + np.std(Y, ddof=0)**2/Y.size)
wald = abs((theta_est - theta) / se_est)
z_alpha = scipy.stats.norm.ppf(1 - alpha / 2)
conf_interval = (theta_est - z_alpha * se_est, theta_est + z_alpha * se_est)
p_value = 2 * (1 - scipy.stats.norm.cdf(wald))

print(f'p-value: {p_value:.6f}')
print(f'Точечная оценка разности средних: {theta_est:.3f}')
print(f'95%-доверительный интервал: ({conf_interval[0]:.3f}, {conf_interval[1]:.3f})')
```

```
p-value: 0.000075
Точечная оценка разности средних: 0.022
95%-доверительный интервал: (0.011, 0.033)
```

На уровне значимости 0.05 уверенно отвергаем нулевую гипотезу, при которой средние равны в двух выборках, средовательно стиль написания статей разный.

– Сделаем расчет для критерия перестановок:

```
permutations = 100_000
XY = np.hstack([X, Y])
m = X.size
differences = []

for _ in range(permutations):
    np.random.shuffle(XY)
    X_permuted, Y_permuted = XY[:m], XY[m:]
    T = abs(X_permuted.mean() - Y_permuted.mean())
    differences.append(T > theta_est)

p_value = np.mean(differences)
print(f'p-value: {p_value:.5f}')
```

```
p-value: 0.00073
```

Вывод аналогичный - средние стат. значимо различаются, H_0 отвергаем.

Ответ: на уровне значимости $\alpha = 0.05$ средние стат. значимо различаются по обоим критериям - Вальда и перестановок.

Задача 12 [2 балла]

Девочка каждый будний день путешествует в метро от станции A до станции B. Со станции A составы идут в двух направлениях: до станции B и до станции C. Если приходит поезд до станции C, Девочке приходится дожидаться следующего поезда. Девочке кажется, что ей очень везёт с поездами до станции B и что поезда до станции B ходят чаще, чем поезда до станции C, поэтому на протяжении двух месяцев записывает, до какой станции идут поезда, которые она успела увидеть, спустившись на станцию (таблица `trains.csv`). Необходимо проверить, ходят ли поезда до станции B значимо чаще, чем до станции C.

Обозначим реальную частоту поездов до станции B через p и будем считать, что поезда приходят случайно и независимо друг от друга. Необходимо проверить гипотезу, что $p = p_0 = \frac{1}{2}$. Для этого:

(a) Постройте и примените тест на основе критерия отношения правдоподобий для различения гипотез $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$.

(b) Постройте и примените тест Вальда для различения гипотез $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$.

(c) Сравните (как аналитически, так и экспериментально) полученный LLR-тест с тестом Вальда для различения этих гипотез.

Примечание. Аналитическое сравнение тестов подразумевает доказательство их (асимптотической) эквивалентности или неэквивалентности, где под эквивалентностью понимается идентичность выносимых тестами решений.

Решение

a) Критерий отношения правдоподобий - отвергаем H_0, если:

$$T_\lambda(X^n) = 2 \log \frac{L(\hat{\theta})}{L(\hat{\theta}_0)} > \chi^2_{1,1-\alpha}$$

Так как случайная величина распределена по закону $X \sim Bernoulli(p)$: либо поезд до станции B - успех ($X = 1$), либо до C - не успех ($X = 0$), то имеем:

$$L(\hat{\theta}) = \hat{\theta}^{\sum x_i}(1-\hat{\theta})^{n-\sum x_i}$$

$$L(\hat{\theta}_0) = \hat{\theta}_0^n$$

где θ - доля поездов до станции B, $\theta_0 = 0.5$ - нулевая гипотеза. Реализуем расчет критерия, приняв уровень значимости $\alpha = 0.05$:

```
trains = pd.read_csv('trains.csv')
to_B = trains.loc[trains['train_to_B'] == 1].shape[0]
to_C = trains.loc[trains['train_to_B'] == 0].shape[0]
p_0 = 0.5
half = p_0 * total
total = to_B + to_C

print(f'to_B = {to_B}, to_C = {to_C}')
print(f'p_0 = {p_0}, p_B = {to_B/total:.4f}, p_C = {to_C/total:.4f}')

alpha = 0.05
lambda_ = to_B * np.log(to_B / total) + to_C * np.log(to_C / total) - total * np.log(p_0)
chi2_value = scipy.stats.chi2.ppf(1 - alpha, 1)
p_value = 1 - scipy.stats.chi2.cdf(lambda_, 1)

print(f'T_lambda = {lambda_:.4f}')
print(f'95% chi2 = {chi2_value:.4f}')
print(f'p-value = {p_value:.4f}')
```

```
to_B = 34, to_C = 12
p_0 = 0.5, p_B = 0.7391, p_C = 0.2609
T_lambda = 5.4824
95% chi2 = 3.8415
p-value = 0.0192
```

Значение $T_\lambda(X^n)$ позволяет отвергнуть H_0, $p - value$ также меньше 0.05.

b) Критерий Вальда - отвергаем H_0, если:

$$|W| = \left|\frac{\hat{\theta} - \theta_0}{\hat{se}}\right| > z_{\alpha/2}$$

Здесь $\theta_0 = 0.5$, MLE-оценка $\hat{\theta} = \overline{X^n}$, оценка $\hat{se} = \sqrt{\hat{\theta}(1-\hat{\theta})/n}$, выполним расчет:

```
theta = 0.5
theta_est = to_B / total
se_est = np.sqrt(theta_est * (1 - theta_est) / total)
wald = abs((theta_est - theta) / se_est)
z_alpha = scipy.stats.norm.ppf(1 - alpha / 2)
conf_interval = (theta_est - z_alpha * se_est, theta_est + z_alpha * se_est)
p_value = 2 * (1 - scipy.stats.norm.cdf(wald))

print(f'p-value: {p_value:.6f}')
print(f'Точечная оценка разности средних: {theta_est:.3f}')
print(f'95%-доверительный интервал: ({conf_interval[0]:.3f}, {conf_interval[1]:.3f})')
```

```
p-value: 0.000221
Точечная оценка разности средних: 0.739
95%-доверительный интервал: (0.612, 0.866)
```

Вывод аналогичный - H_0 уверенно отвергаем, поезда до станции B ездят чаще.

c) LLR-тест и тест Вальда дали одинаковые экспериментальные выводы, покажем также аналитически, что эти тесты асимптотически эквивалентны:

$$T_\lambda^{LLR}(X_n) = 2\log\frac{\hat{p}^{n_1}(1-\hat{p})^{n-n_1}}{p_0^{n_1}(1-p_0)^{n-n_1}} = 2\big(n_1\log\frac{\hat{p}}{p_0} + (n-n_1)\log\frac{1-\hat{p}}{1-p_0}\big) =$$

$$= 2n\big(\hat{p}\log\frac{\hat{p}}{p_0} + (1-\hat{p})\log\frac{1-\hat{p}}{1-p_0}\big) \approx -2n\big[\cancel{(p_0-\hat{p})} + \frac{(p_0-\hat{p})^2}{2\hat{p}} + \cancel{(\hat{p}-p_0)} + \frac{(\hat{p}-p_0)^2}{2(1-\hat{p})}\big] =$$

$$= n\frac{(\hat{p}-p_0)^2}{\hat{p}(1-\hat{p})} = \frac{(\hat{p}-p_0)^2}{\frac{\hat{p}(1-\hat{p})}{n}} > \chi^2_{1,1-\alpha}$$

$$T_\lambda^{Wald}(X_n) = \frac{|\hat{p}-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \frac{|\hat{p}-p_0|}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} > z_{\alpha/2}$$

Так как $\chi^2_n \sim \sum_{i=1}^n Z_i^2$, тесты эквивалентны.

Ответ:
a-b) поезда до станции B ходят значимо чаще согласно LLR-тесту и тесту Вальда (уровень значимости выбран $\alpha = 0.05$),
c) тесты асимпотически эквивалентны.