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

Кузина Е.М.

18 апреля 2020 г.

Задача 1 [5 баллов]

Пусть $\boldsymbol{X}^n = \{X_1, \ldots, X_n\} \sim \mathcal{N}(\theta, 1)$. Определим случайную величину Y, зависящую от $X \sim \mathcal{N}(\theta, 1)$ следующим образом.

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

Далее мы наблюдаем выборку $\boldsymbol{Y}^n$, по которой требуется оценить параметр $\psi = \mathsf{P}(Y = 1)$ распределения случайной величины Y.

a) Записать MLE-оценку ψ_{MLE} для ψ, в предположении знания выборки $\boldsymbol{X}^n$ и семейства породившего ее распределения $\mathcal{N}(\theta, 1)$.

Решение:

Оценим θ методом максимального правдоподобия:

$l(\theta) = log \prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}} exp(-\frac{(X_i-\theta)^2}{2}) = -nlog(\sqrt{2\pi}) - \sum\limits_{i=1}^{n} \frac{(X_i-\theta)^2}{2}$

$\frac{\partial l(\theta)}{\partial \theta} = \sum\limits_{i=1}^{n} (X_i - \theta) = 0 => \hat{\theta} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$

$\psi_{MLE} = \hat{\mathsf{P}}(Y = 1) = \hat{\mathsf{P}}(X > 0) = 1 - \hat{\mathsf{P}}(X \leq 0) = 1 - \int\limits_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} exp(-\frac{(x - \frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2}{2}) dx$

b) Найти приближенный 95% доверительный интервал для ψ, воспользовавшись дельта-методом.

Решение:

$g(\theta) = 1 - \int\limits_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-\theta)^2}{2}) dx => g'(\theta) = -\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} exp(-\frac{(x-\theta)^2}{2})(x-\theta) dx = \frac{1}{\sqrt{2\pi}} exp(-\frac{\theta^2}{2}) \neq 0$

$\hat{se}(\psi_{MLE}) = |g'(\hat{\theta})| \hat{se}(\hat{\theta}) = \frac{1}{\sqrt{2n\pi}} exp(-\frac{\hat{\theta}^2}{2})$, тк $\hat{se}(\hat{\theta}) = \sqrt{\frac{\sigma^2}{n}}$

Приближенный 95% доверительный интервал для ψ: $\psi_{MLE} \pm 2\hat{se}(\hat{\theta}) = \psi_{MLE} \pm \frac{2}{\sqrt{2n\pi}} exp(-\frac{\hat{\theta}^2}{2})$

c) Пусть $\hat{\psi} = \langle \boldsymbol{Y}^n \rangle = n^{-1} \sum_{i=1}^{n} Y_i$. Доказать, что $\hat{\psi}$ является состоятельной оценкой для ψ.

Решение:

$\mathbb{V}\hat{\psi} = \frac{1}{n} \mathbb{V} Y_i = \frac{1}{n} P(Y_i = 1)(1 - P(Y_i = 1)) \to 0$ при $n \to \infty$ - оценка состоятельна

d) Подсчитать асимптотическую относительную эффективность оценки $\hat{\psi}$ по сравнению с оценкой ψ_{MLE}. Стандартную ошибку оценки максимума правдоподобия для ψ_{MLE} предлагается взять из пункта b), где для ее нахождения использовался дельта-метод. После чего надо подсчитать стандартное отклонение величины $\hat{\psi}$.

Решение:

$$ARE(\psi_{MLE}, \hat{\psi}) = \frac{\mathrm{se}^2(\hat{\psi})}{se^2(\psi_{MLE})} = \frac{\frac{1}{n}(1 - \Phi(\frac{0-\hat{\theta}}{1}))\Phi(\frac{0-\hat{\theta}}{1})}{\frac{1}{2n\pi} exp(-\hat{\theta}^2)} = \frac{2\pi(1 - \Phi(-\hat{\theta}))\Phi(-\hat{\theta})}{exp(-\hat{\theta}^2)} \to \frac{2\pi(1 - \Phi(-\theta))\Phi(-\theta)}{exp(-\theta^2)}, \text{ при } n \to \infty$$

Асимптотическая относительная эффективность зависит от величины θ. Если $ARE(\psi_{MLE}, \hat{\psi}) \leq 1$, то $\hat{\psi}$ имет меньшую асимптотическую дисперсию, поэтому она эффективна/асимптотически нормальна. Иначе эффективной оценкой будет ψ_{MLE}.

e) Допустим, что случайные величины $X_1, \dots, X_n$ на самом деле не распределены нормально. Показать, что в таком случае ψ_{MLE} не является состоятельной оценкой. Будет ли, и если ответ "да", то к чему, сходится при $n \to \infty$ оценка ψ_{MLE} в смысле какой-нибудь сходимости?

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 $X_1 \sim \text{Binomial}(n_1, p_1)$ и $X_2 \sim \text{Binomial}(n_2, p_2)$. Положим $\psi = p_1 - p_2$.

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

Решение:

Логарифмическая функция правдоподобия $l(p_1) = log(\prod\limits_{i=1}^{n1} p_1^y(1-p_1)^{1-y})$, где $y = \begin{cases} 0, & \text{методика не повлияла} \\ 1, & \text{иначе} \end{cases}$

$l(p_1) = log(p_1^{X_1}(1-p_1)^{n_1-X_1}) = X_1 log(p_1) + (n_1 - X_1) log(1-p_1)$

$\frac{\partial l(p_1)}{\partial p_1} = \frac{X_1}{p_1} - \frac{n_1 - X_1}{1-p_1} = 0 => \hat{p_1} = \frac{X_1}{n_1}$

Аналогичего $\hat{p_2} = \frac{X_2}{n_2}$

$\psi_{MLE} = \hat{p_1} - \hat{p_2} = \frac{X_1}{n_1} - \frac{X_2}{n_2}$

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

Решение:

$l(p_1, p_2) = log\left(\prod\limits_{i=1}^{n_1} p_1^{y_1}(1-p_1)^{1-y_1} \prod\limits_{i=1}^{n_2} p_2^{y_2}(1-p_2)^{1-y_2}\right) = \sum\limits_{i=1}^{n_1} X_i log(p_1) + \sum\limits_{i=1}^{n_1}(1-X_i) log(1-p_1) +$

$+ \sum\limits_{i=1}^{n_2} X_i log(p_2) + \sum\limits_{i=1}^{n_2}(1-X_i) log(1-p_2)$

$\frac{\partial l(p_1,p_2)}{\partial p_1} = \frac{\sum_{i=1}^{n_1} X_i}{p_1} - \frac{n_1 - \sum_{i=1}^{n_1} X_i}{1-p_1}, \quad \frac{\partial^2 l(p_1,p_2)}{\partial p_1^2} = -\frac{\sum_{i=1}^{n_1} X_i}{p_1^2} - \frac{n_1 - \sum_{i=1}^{n_1} X_i}{(1-p_1)^2}, \quad \frac{\partial^2 l(p_1,p_2)}{\partial p_1 \partial p_2} = 0$

$\frac{\partial l(p_1,p_2)}{\partial p_2} = \frac{\sum_{i=1}^{n_2} X_i}{p_2} - \frac{n_2 - \sum_{i=1}^{n_2} X_i}{1-p_2}, \quad \frac{\partial^2 l(p_1,p_2)}{\partial p_2^2} = -\frac{\sum_{i=1}^{n_2} X_i}{p_2^2} - \frac{n_2 - \sum_{i=1}^{n_2} X_i}{(1-p_2)^2}$

$-E(\frac{\partial^2 l(p_1,p_2)}{\partial p_1^2}) = \frac{n_1}{p_1(1-p_1)}, \quad -E(\frac{\partial^2 l(p_1,p_2)}{\partial p_2^2}) = \frac{n_2}{p_2(1-p_2)}$, так как X_i из распределения Бернулли

Тогда информационная матрица Фишера $I(p_1, p_2) = \begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}$

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

Решение:

$\psi = g(p_1, p_2) = p_1 - p_2$

$\nabla g(p_1, p_2) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad J = I^{-1}(p_1, p_2) = \begin{pmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{pmatrix}$

$\hat{se}(\psi_{MLE}) = \sqrt{(\hat{\nabla} g)^T \hat{J} (\hat{\nabla} g)} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

(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:

$\hat{\psi} = \hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2} = 0.06$

а) $\hat{se}(\psi_{MLE}) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.041976$, 90% доверительный интервал: (0.06 - 0.041976$z_{0.05}$, 0.06 + 0.041976$z_{0.05}$) => (0.06 - 0.041976$z_{0.05}$, 0.06 + 0.041976$z_{0.05}$) $\approx$ (-0.009, 0.129)

б) доверительный интервал: (-0.007, 0.127)

Задача 3 [3 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \text{Uniform}(\theta, \theta + 1)$. 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 $X_{(n)} \geq 1$ или $X_{(1)} \geq c$, где c — некоторая константа, $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$.

(a) Найдите функцию мощности для данного теста.

Решение:

Рассмотрим случай, когда $c \geq 1$, тогда очевидно, что условие, того что гипотеза H_0 отвергается, можно переписать в виде $X_{(1)} \geq c$. В случае $c \leq 1$ условие отвержения гипотезы также можно переписать в виде $X_{(1)} \geq c$. Тогда вероятность ошибки первого рода:

$$P_\theta(H_1) = P_\theta(X_{(1)} \geq c) = P_\theta^n(X_i \geq c) = (1 - P_\theta^n(X_i \leq c))^n = (1 - \int\limits_\theta^c dx)^n = (1 - c + \theta)^n,$$

так как плотность распределения $X_i : f(x) = \begin{cases} 1, & x \in [\theta, \theta + 1] \\ 0, & \text{иначе} \end{cases}$

$W(\theta) = P_\theta(H_1) = (1 - c + \theta)^n$ - функция мощности для данного теста

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

Решение:

Размер теста $\alpha = \sup_{\theta \in F_0} P_\theta(H_1) = (1 - c)^n \Rightarrow c = 1 - \sqrt[n]{\alpha} \Rightarrow c = 1 - \sqrt[n]{0.05}$

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

Решение:

$(1 - c + \theta)^n = (1 - 1 + \sqrt[n]{0.05} + 0.1)^n \geq 0.8 \Rightarrow 0.1 \geq \sqrt[n]{0.8} - \sqrt[n]{0.05} \Rightarrow n \geq 27$

Задача 4 [2 балла]

Пусть $\{X_1, \dots, X_n\}$ — н.о.р.с.в со следующей функцией плотности:

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

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:

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

Решение:

Воспользуемся теоремой из лекции на 41 слайде (ссылка). Множество $f = \{\theta_0, \theta_1 \neq \theta_0\}$. Тогда при выполнении гипотезы H_0 статистика отношения правдоподобия λ асимптотически распределена по закону χ_1^2 с одной степенью свободы, так как множество $f_0 = \{\theta_0\}$.

Задача 5 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \text{Pareto}(\theta, \nu)$, $\theta > 0$, $\nu > 0$ с функцией плотности

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

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

Решение:

$$L(\theta,\nu) = \begin{cases} \prod\limits_{i=1}^{n} \frac{\theta\nu^\theta}{x^{\theta+1}}, & \nu \leq \min\limits_i X_i \\ 0, & \text{иначе} \end{cases}$$

$$ln(L(\theta,\nu)) = nln(\theta) + n\theta ln(\nu) - (\theta+1)\sum_{i=1}^{n} ln(X_i)$$

Понятно, что максимум будет достигаться при $\hat{\nu} = \min\limits_i X_i$.

$$\frac{\partial L(\theta.\nu)}{\partial\theta} = \frac{n}{\theta} + nln(\nu) - \sum_{i=1}^{n} ln(X_i) = 0 => \hat{\theta} = \frac{n}{\sum\limits_{i=1}^{n} ln(X_i) - nln(\hat{\nu})}$$

b) Используя дельта-метод найдите асимптотическое распределение для $\hat{\theta}$ при известном ν.

Решение:

Переименуем ОМП $\hat{\theta}$ в θ_{MLE}, чтобы не было путаницы.

Пусть $\theta_{MLE} = g(\theta) = \theta$, тогда $g'(\theta) = 1 \neq 0$.

Т.к. ОМП асимптотически нормальна, то $se(\theta) \approx \sqrt{\frac{1}{I_n(\theta)}}$, где $I_n(\theta)$ - информация Фишера.

$$I_n(\theta) = -E(\frac{\partial l^2(\theta)}{\partial\theta^2}) = -E\frac{n}{\theta^2} = \frac{n}{\theta^2}, \quad se(\theta) \approx \frac{\theta}{\sqrt{n}}, \quad \hat{se}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}}$$

Тогда $\theta_{MLE} = |g'(\theta)|\hat{se}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}}$

По теореме из лекции на слайде 39 (ссылка) $\theta_{MLE} : \frac{\sqrt{n}(\theta_{MLE} - \theta)}{\hat{\theta}} \sim N(0,1)$, т.е. оценка асимптотически нормальна.