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

Школа Анализа Данных

Задачи

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:
ОМП не зависит от параметризации, поэтому найдем сначала ОМП для θ и используем её для нахождения ОМП ψ.

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

$$\frac{\partial l(\theta)}{\partial \theta} = -\sum_{i=1}^{n} (X_i - \theta) = 0 => \theta_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

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

Ответ: $\psi_{MLE} = \Phi(\frac{1}{n} \sum\limits_{i=1}^{n} X_i)$

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

$$\frac{\Phi(\theta_{MLE}) - \Phi(\theta)}{|\Phi'(\theta_{MLE})| se(\theta_{MLE})} \to N(0, 1)$$

$$|\Phi'(\theta_{MLE})| se(\theta_{MLE}) = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2} \frac{1}{\sqrt{n}}$$

95% доверительный интервал:

$$(\Phi(\frac{1}{n} \sum_{i=1}^{n} X_i) - z_{\alpha/2} \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2}, \Phi(\frac{1}{n} \sum_{i=1}^{n} X_i) + z_{\alpha/2} \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2})$$

$$(\Phi(\frac{1}{n} \sum_{i=1}^{n} X_i) - 2 \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2}, \Phi(\frac{1}{n} \sum_{i=1}^{n} X_i) + 2 \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2})$$

Ответ: $(\Phi(\frac{1}{n} \sum\limits_{i=1}^{n} X_i) - 2 \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2}, \Phi(\frac{1}{n} \sum\limits_{i=1}^{n} X_i) + 2 \frac{1}{\sqrt{2\pi n}} e^{-(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)^2/2})$

c) Пусть $\tilde{\psi} = \langle \boldsymbol{Y}^n \rangle = n^{-1} \sum_{i=1}^{n} Y_i$. Доказать, что $\tilde{\psi}$ является состоятельной оценкой для ψ.
Решение:
$\mathbb{V}\tilde{\psi} = \frac{1}{n} \mathbb{V} Y_i = \frac{1}{n} P(Y_i = 1)(1 - P(Y_i = 1)) \to 0$ при $n \to \infty$ - оценка состоятельна

d) Подсчитать асимптотическую относительную эффективность оценки $\tilde{\psi}$ 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 $\tilde{\psi}$.
Решение:

$$ARE(\hat{\psi}, \psi_{MLE},) = \frac{se^2(\psi_{MLE})}{\mathrm{se}^2(\hat{\psi})} = \frac{e^{-\theta^2_{MLE}}}{2\pi n}\frac{n}{\Phi(\theta_{MLE})\Phi(-\theta_{MLE})} = \frac{e^{-\theta^2_{MLE}}}{2\pi\Phi(\theta_{MLE})\Phi(-\theta_{MLE})} \leq 1,$$

т.к. ОМП имеет наименьшую асимптотическую дисперсию (утверждение из лекции).

e) Допустим, что случайные величины $X_1, \ldots, X_n$ 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?
Решение:
1) Предположим, что случайные величины X_i имеют математическое ожидание M, тогда по закону больших чисел $\frac{1}{n}\sum\limits_{i=1}^{n} X_i \to M$ по вероятности. По теореме о наследовании $\Phi(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) \to \Phi(M)$ по вероятности. Таким образом, $\psi_{MLE} = \Phi(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)$ сходится по вероятности к $\Phi(M)$, где M - настоящее мат ожидание случайных величин X_i.
2) Покажем от противного, что ψ_{MLE} не является состоятельой оценкой. Предположим, что оценка является состоятельной, тогда $\psi_{MLE} \to P(Y = 1)$ по вероятности, тогда $P(Y = 1) = \Phi(M)$, что невозможно по распределению, получили противоречие. Таким образом, ψ_{MLE} не является состоятельой оценкой.

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:
Распишем логарифм правдоподобия:

$$\log L = \log\left(C_{n_1}^{X_1} p_1^{X_1}(1-p_1)^{n_1-X_1} C_{n_2}^{X_2} p_2^{X_2}(1-p_2)^{n_2-X_2}\right)$$

$$\frac{\partial \log L}{\partial p_1} = 0 => \hat{p_1} = \frac{X_1}{n_1} \qquad \frac{\partial \log L}{\partial p_2} = 0 => \hat{p_2} = \frac{X_2}{n_2}$$

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

Ответ: $\psi_{MLE} = \frac{X_1}{n_1} - \frac{X_2}{n_2}$

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

$$\frac{\partial \log L}{\partial p_1} = \frac{X_1}{p_1} - \frac{n_1 - X_1}{1-p_1}, \quad \frac{\partial^2 \log L}{\partial {p_1}^2} = -\frac{X_1}{p_1^2} - \frac{n_1 - X_1}{(1-p_1)^2}, \quad \frac{\partial^2 \log L}{\partial p_1 \partial p_2} = 0$$

$$\frac{\partial \log L}{\partial p_2} = \frac{X_2}{p_2} - \frac{n_2 - X_2}{1-p_2}, \quad \frac{\partial^2 \log L}{\partial {p_2}^2} = -\frac{X_2}{p_2^2} - \frac{n_2 - X_2}{(1-p_2)^2}, \quad \frac{\partial^2 \log L}{\partial p_1 \partial p_2} = 0$$

$$-E\left(\frac{\partial^2 \log L}{\partial p_1^2}\right) = \frac{n_1}{p_1(1-p_1)}, \quad -E\left(\frac{\partial^2 \log L}{\partial p_2^2}\right) = \frac{n_2}{p_2(1-p_2)}$$

$I(p_1,p_2)_{ij} = -\mathbb{E}(\frac{\partial^2 \log L}{\partial p_i \partial p_j})$, поэтому информационная матрица Фишера $I(p_1, p_2) = \begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}$

Ответ:

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

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

$$\psi = p_1 - p_2 => \nabla\psi(p_1, p_2) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

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

$$\hat{se}(\psi_{MLE}) = \sqrt{(\hat{\nabla}\psi)^T \hat{J}(\hat{\nabla}\psi)} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{X_1(n_1 - X_1)}{n_1^3} + \frac{X_2(n_2 - X_2)}{n_2^3}}$$

Ответ: $\hat{se}(\psi_{MLE}) = \sqrt{\frac{X_1(n_1-X_1)}{n_1^3} + \frac{X_2(n_2-X_2)}{n_2^3}}$

(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: Решение см. в приложенном ноутбуке
Ответ: а) (-0.009044678043395274, 0.12904467804339537) б) (-0.0085253402998786, 0.12852534029987872)

Задача 3 [1 балл]

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

Решение:

Оценка максимума правдоподобия для экспоненциального распределения:

$$L(\theta) = \log \prod_{i=1}^{n} \theta e^{-\theta X_i} = n \log\theta - \theta \sum_{i=1}^{n} X_i \qquad \frac{\partial L(\theta)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} X_i = 0 => \theta_{MLE} = \frac{1}{\langle X \rangle}$$

$$\lambda = 2\log\frac{\sup_{H1} L(\theta)}{L(\theta_0)} = 2\log\frac{L(\theta_{MLE})}{L(\theta_0)} = -2n(\log\langle X\rangle + 1 + \log\theta_0 - \langle X\rangle\theta_0),$$

если $\theta_{MLE} > \theta_0$, иначе $\lambda = 0$.

Задача 4 [3 балла]

Пусть $\boldsymbol{X}^n \sim \text{Uniform}(\theta, \theta+1)$. 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, $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(a) Найдите функцию мощности для данного теста.
Решение:

Будем рассматривать случаи при $c \in (0,1)$, т.к. остальные случаи не интересны

$$\beta(\theta) = P_{\theta>0}(X \in \text{критич. обл}) = P_{\theta>0}(X_{(n)} \geq 1 \text{ или } X_{(1)} \geq c) =$$

$$= P_{\theta>0}(X_{(n)} \geq 1) + P_{\theta>0}(c \leq X_{(1)} \text{ и } X_{(n)} < 1) =$$

$$= \begin{cases} 1 - (1-\theta)^n + (1-c)^n, & \text{если } 0 \leq \theta < c \\ 1, & \text{иначе} \end{cases}$$

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

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = 1-(1-0)^n + (1-c)^n = (1-c)^n = 0.05 \Rightarrow \boxed{c = 1 - \sqrt[n]{0.05}}$$

Ответ: $c = 1 - 0.05^{\frac{1}{n}}$

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

$$1-(1-\theta_1)^n + (1-c)^n \geq 0.8$$

$$1-(1-\theta_1)^n + 0.05 \geq 0.8$$

$$(1-\theta_1)^n \leq 0.25$$

$$n \leq \log_{1-\theta_1} 0.25 = \log_{0.9} 0.25 \approx 13.16$$

$$\boxed{n \geq 14}$$

Ответ: при n ≥ 14

Задача 5 [2 балла]

Пусть $\boldsymbol{X}^n \sim \text{Pareto}(\theta, \nu)$, $\theta > 0$, $\nu > 0$, 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:

$$L(\theta, \nu) = \begin{cases} \prod_{i=1}^{n} \frac{\theta \nu^\theta}{X_i^{\theta+1}}, & \nu \leq X_{(1)} \\ 0, \text{иначе} \end{cases}$$

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

Минимум $\ln L(\theta,\nu)$ достигается при $\hat{\nu} = X_{(1)}$

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

Ответ: $\hat{\nu} = \min X_i$; $\quad \hat{\theta} = \frac{n}{\sum\limits_{i=1}^{n} \ln X_i - n ln \hat{\nu}}$

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

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln \frac{X_i}{\hat{\nu}}}$$

Известно, что если $X_i \sim Pareto(\theta, \nu)$ и $\hat{\nu} = \min X_i$, то $\ln \frac{X_i}{\hat{\nu}} \sim Exp(X_i, \hat{\nu})$

$E \ln \frac{X_i}{\hat{\nu}} = \frac{1}{\nu}$, $V \ln \frac{X_i}{\hat{\nu}} = \frac{1}{\nu^2}$.

Тогда $\dfrac{\frac{1}{n}\sum_{i=1}^{n} \ln \frac{X_i}{\hat{\nu}} - \frac{1}{\hat{\nu}}}{\frac{1}{\sqrt{n\nu^2}}} \sim N(0,1)$

Рассмотрим ф-ию $g(\theta) = \frac{1}{\theta}$, чтобы получить нужную нам статистику. Тогда согласно дельта-методу:

$$\frac{g\left(\frac{1}{n}\sum_{i=1}^{n} \ln \frac{X_i}{\nu}\right) - g\left(\frac{1}{\nu}\right)}{\left|g'\left(\frac{1}{\nu}\right)\right| \cdot se\left(\frac{1}{n}\sum_{i=1}^{n} \ln \frac{X_i}{\nu}\right)} = \frac{\hat{\theta} - \nu}{\frac{1}{\left(\frac{1}{\nu}\right)^2} \cdot \frac{1}{\sqrt{n\nu^2}}} = \frac{\sqrt{n}(\hat{\theta} - \nu)}{\nu} \to N(0,1) \Rightarrow$$

$$\sqrt{n}(\hat{\theta} - \nu) \to N(0, \nu^2)$$

Ответ: $\hat{\theta}$ асимптотически нормальна

Задача 6 [2 балла]

Пусть $\boldsymbol{X}^n$ — выборка 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 $H_1 : \theta \neq \theta_0$

Решение:

$$\lambda = 2\log \frac{\sup_{\theta \neq \theta_0} h(\theta, X)}{h(\theta_0, X)} = 2\log \frac{\sup_{\theta \neq \theta_0} \prod_{i=1}^n c(\theta) d(X_i) I[a \le X_i \le b(\theta)]}{\prod_{i=1}^n c(\theta_0) d(X_i) I[a \le X_i \le b(\theta_0)]} =$$

$$= 2\log \frac{c^n(\hat\theta) \prod_{i=1}^n d(X_i)}{c^n(\theta_0) \prod_{i=1}^n d(X_i) I[X_{(n)} \le b(\theta_0)]} \boxed{=}$$

Как получено последнее равенство?

Заметим, что $c(\theta)$ - убывающая ф-ия, т.к. с ростом $b(\theta)$ расширяется диапазон X. Тогда оценка максимума правдоподобия достигает своего максимума при θ: $X_{(n)} = b(\theta) \Rightarrow \hat\theta = b^{-1}(X_{(n)})$

$$\boxed{=}\ 2\log \frac{c^n(\hat\theta)}{c^n(\theta_0) I[X_{(n)} \le b(\theta_0)]}$$

(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:

$d(x) = 2x$, $c(\theta) = \frac{1}{\theta^2}$, $b(\theta) = \theta$

$$\lambda = 2\log \frac{\frac{1}{(X_{(n)})^{2n}}}{\frac{1}{(\theta_0)^{2n}}} = 2\log \frac{(\theta_0)^{2n}}{(X_{(n)})^{2n}} = 4n \log \frac{\theta_0}{X_{(n)}} = -4n \log \frac{X_{(n)}}{\theta_0}$$

$$P_{\theta_0}(X_i \le t) = \frac{t^2}{\theta_0^2} I[0 \le X_i \le \theta_0] + I[\theta_0 \le X_i]$$

$$P_{\theta_0}\left(\frac{X_i}{\theta_0} \le t\right) = P_{\theta_0}(X_i \le t\theta_0) = t^2 I[0 \le t \le 1] + I[1 < t]$$

$$P_{\theta_0}\left(\max \frac{X_i}{\theta_0} \le t\right) = t^{2n} I[0 \le t \le 1] + I[1 < t]$$

$$P_{\theta_0}\left(-4n \ln \max \frac{X_i}{\theta_0} \le t\right) = P_{\theta_0}\left(\ln \max \frac{X_i}{\theta_0} \geqslant -\frac{t}{4n}\right) =$$

$$= P_{\theta_0}\left(\max \frac{X_i}{\theta_0} \geqslant e^{-t/(4n)}\right) = 1 - P_{\theta_0}\left(\max \frac{X_i}{\theta_0} < e^{-t/(4n)}\right) =$$

$$= 1 - e^{-t/2} I[0 \le e^{-t/(4n)} \le 1] + I[1 < e^{-t/(4n)}] \xrightarrow[n \to \infty]{} 1 - e^{-t/2} = F_{\chi^2_2}(t)$$

$$\boxed{\lambda \sim \chi^2_2}$$

Задача 7 [2 балла]

Найдите наилучшую критическую область (НКО) для проверки гипотезы $H_0\colon \text{Uniform}[-a, a]$ против гипотезы $H_1\colon \mathcal{N}(0, \sigma)$ по одному наблюдению ($n = 1$) при уровне значимости $\alpha = 0.1$. Найдите мощность полученного критерия.

Решение:

$\alpha = P(\text{отклон. } H_0 \mid H_0) = P_{H_0}(|x| > a) + P_{H_0}(|x| < b) =$

$= 0 + P_{H_0}(-b < x < b) = 0 + \frac{b+a}{2a} - \frac{-b+a}{2a} = \frac{b}{a} = 0.1 \Rightarrow b = 0.1a$

Наилучшая критическая область: $|x| > a$ или $|x| < 0.1a$

Мощность критерия $= P(H_1 \mid H_1) = P_{H_1}(|x| > a) + P_{H_1}(|x| < 0.1a) =$

$= 1 - \Phi\left(\frac{a}{\sigma}\right) + \Phi\left(\frac{-a}{\sigma}\right) + \Phi\left(\frac{0.1a}{\sigma}\right) - \Phi\left(-\frac{0.1a}{\sigma}\right)$

Задача 8 [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?

Задача 9 [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 можно сделать вывод исходя из найденных значений?
- Используйте критерий перестановок. Каково в этом случае значение p-value. Какой можно сделать вывод?

Задача 10 [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:

(a) Построить тест на основе критерия отношения правдоподобий для различения гипотез $H_0\colon p = p_0$ vs. $H_1\colon p \neq p_0$.

(b) Пусть $p_0 = \frac{1}{2}$. Сравнить (как аналитически, так и экспериментально) полученный 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.