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

Школа Анализа Данных

Общие правила

- Построить тест размера α, значит представить правило вида $T(\boldsymbol{X}^n) > c(\alpha)$, где $T(\cdot)$ — заданная функция и $c(\alpha)$ — заданный критический порог.
- Найти критическую область для теста размера α, значит представить множество $R(\alpha)$ в пространстве $\mathcal{X}^n$, где $\mathcal{X}$ — область значений случайной величины X ($\boldsymbol{X}^n \in \mathcal{X}^n$). Примеры записи ответа об НКО:

$$R(\alpha) = \{\boldsymbol{X}^n \colon \prod_{i=1}^{n} X_i > c\},$$
$$R(\alpha) = \{\boldsymbol{X}^n \colon \sum_{i=1}^{n} |X_i| > c\},$$

  где c — найденный критический порог для уровня значимости α.
- Под аналитическим сравнением подразумевается доказательство эквивалентности/не эквивалентности тестов при $n \to \infty$.

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.

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

Случайная величина Y имеет распределение Бернулли. Далее по наблюдаемой выборке $\boldsymbol{X}^n$ требуется оценить параметр $\psi = \mathsf{P}(Y = 1)$ распределения случайной величины Y.

(a) Записать MLE-оценку ψ_{MLE} для параметра ψ.
(b) Найти приближенный 95%-ый доверительный интервал для ψ.
(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.
(d) Подсчитать асимптотическую относительную эффективность оценки $\tilde{\psi}$ по сравнению с оценкой ψ_{MLE}. Для этого предлагается использовать дельта-метод, чтобы оценить стандартную ошибки оценки максимума правдоподобия. После чего надо подсчитать стандартное отклонение величины $\tilde{\psi}$.
(e) Допустим, что случайные величины $\boldsymbol{X}^n = \{X_1, \ldots, X_n\}$ 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?

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.
(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.

Задача 3 [4 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \text{Uniform}(0, \theta)$, $Y = \max\{X_1, \dots, X_n\}$. Необходимо протестировать основную гипотезу $H_0 : \theta = 1/2$ против альтернативы $H_1 : \theta > 1/2$. В данном случае нельзя использовать тест Вальда, так как Y при $n \to \infty$ не сходится к нормальному распределению. Допустим, что мы будем использовать следующее правило: гипотеза H_0 отвергается, если $Y > c$.

(a) Найдите функцию мощности для данного теста.

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

(c) Каково значение p-value, если размер выборки $n = 20$ и $Y = 0.48$? Что можно сказать о гипотезе H_0?

(d) Каково значение p-value, если размер выборки $n = 20$ и $Y = 0.52$? Что можно сказать о гипотезе H_0?

Задача 4 [2 балла]

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

(a) Пусть $\lambda_0 > 0$. Построить критерий Вальда размера α для различения гипотез $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$.

(b) Пусть $\lambda_0 = 1$, $n = 20$ и $\alpha = 0.05$. Сгенерировать $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \text{Poisson}(\lambda_0)$ и применить критерий Вальда. Повторить эксперимент N раз и подсчитать долю $\hat{\mathsf{P}}_I$ от общего числа случаев, когда гипотеза H_0 была отклонена. Насколько получившаяся оценка $\hat{\mathsf{P}}_I$ вероятности ошибки первого рода оказалась близкой к 0.05? Для ответа на последний вопрос вновь воспользуйтесь критерием Вальда и найдите p-value для проверки гипотезы о том, что $\mathsf{P}_I = 0.05$. Что можно сказать о справедливости гипотезы $\mathsf{P}_I = 0.05$ для найденного значения p-value.
Количество экспериментов N выберете на свое усмотрение, но не менее 100.

Задача 5 [2 балла]

Найти наилучшую критическую область (НКО) и мощность критерия для проверки гипотезы $H_0 \colon a = a_0$, против гипотезы $H_1 \colon a = a_1$, $a_1 > a_0$ для выборки $\boldsymbol{X}^n \sim \mathcal{N}(a, \sigma^2)$ с известной дисперсией σ^2.

Задача 6 [2 балла]

В десятичной записи числа π среди первых 10002 знаков после запятой цифры $0, 1, \dots, 9$ 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 7 [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?

	Количество пациентов	Количество осложнений
Плацебо	80	45
Хлорпромазин	75	26
Дименгидринат	85	52
Пентобарбитал (100 мг)	67	35
Пентобарбитал (150 мг)	85	37

Задача 8 [2 балла]

Был проведен эксперимент по оценке эффективности различных лекарств, используемых для уменьшения послеоперационных эффектов, и получены следующие результаты

(a) Протестировать отличие каждого из лекарств от плацебо на 5%-ом уровне значимости (отличие может быть как в положительную сторону, так и в отрицательную). Указать подсчитанные оценки вероятностей успешных исходов. Указать значение p-value для каждого из тестов «лекарство vs плацебо».

(b) Проделать эксперименты, аналогичные экспериментам предыдущего пункта, но использовать при этом методы Бонферрони и Benjamini-Hochberg.

Задача 9 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \mathcal{N}(\mu, \sigma^2)$, где параметр μ известен. Построить тест на основе критерия отношения правдоподобий для различения гипотез $H_0\colon \sigma = \sigma_0$ и $H_1\colon \sigma \neq \sigma_0$. Сравнить (как аналитически, так и экспериментально) полученный тест с тестом Вальда для различения этих гипотез.

Разбаловка

- Задача 1. **5 балла**.
  - **1 балл**. Найдена MLE-оценку ψ_{MLE} для ψ.
  - **1 балл**. Найти приближенный 95% доверительный интервал для ψ.
  - **1 балл**. Доказано, что $\tilde{\psi}$ является состоятельной оценкой для ψ.
  - **1 балл**. Подсчитана асимптотическая относительная эффективность (ARE) оценки $\tilde{\psi}$ по сравнению с оценкой ψ_{MLE}.
  - **1 балл**.
    - **0.5 балла** Показано, что если $\boldsymbol{X}^n \not\sim \mathcal{N}(\theta, 1)$, то ψ_{MLE} не является состоятельной оценкой.
    - **0.5 балла** за обоснованный ответ на вопрос «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**.
  - **1 балл**. Найдена MLE-оценку ψ_{MLE}.
  - **1 балл**. Вычислена информационная матрица Фишера $I(p_1, p_2)$.
  - **1 балл**. С помощью многопараметрического дельта-метода найдена асимптотическая ошибка для ψ_{MLE}.
  - **1 балл**. Найдены 90%-ые доверительные интервалы для ψ, с помощью многопараметрического дельта-метода и параметрического бутстрепа.
- Задача 3. **4 балла**.
  - **1 балл**. Найдите функцию мощности.
  - **1 балл**. Найдено критическое значение $c(\alpha)$ для $\alpha = 0.05$.
  - **1 балл**. Верно найдено p-value. Сделан вывод о справедливости гипотезы H_0.
  - **1 балл**. Верно найдено p-value. Сделан вывод о справедливости гипотезы H_0.
- Задача 4. **2 балла**.
  - **1 балл**. Построен критерий Вальда.
  - **1 балл**.
    - **0.5 балла**. Экспериментально найдена вероятность ошибки первого рода.
    - **0.5 балла**. Протестирована гипотеза о том, что $\mathsf{P}_I = 0.05$. Найдено p-value для данной гипотезы. Сделан вывод о справедливости гипотезы $\mathsf{P}_I = 0.05$.
- Задача 5. **2 балла**.
  - **1 балл**. Найдена наилучшая критическую область (НКО).
  - **1 балл**. Найдена мощность критерия.
- Задача 6. **2 балла**.

- **1 балл**. «Можно ли при уровне значимости 0.05 считать эти цифры случайными?»
- **1 балл**. «При каком уровне значимости эта гипотеза отвергается?»

- Задача 7. **2 балла**.
  - **1 балл**. Найдено p-value для критерия Вальда, и построен 95%-ый доверительный интервал. Сделан вывод исходя из найденных значений.
  - **1 балл**. Найдено p-value для критерия перестановок. Сделан вывод исходя из найденного значения.
- Задача 8. **2 балла**.
  - **1 балл**. Сделан вывод об "успешности" каждого из лекарств по сравнению с плацебо на 5% уровне значимости. Найдены оценки вероятностей успешных исходов.
  - **1 балл**. Проведено множественное тестирование с помощью методов Бонферрони (**0.5 балла**) и Benjamini-Hochberg (**0.5 балла**).
- Задача 9. **2 балла**.
  - **1 балл**. Построен тест для различения гипотез $H_0\colon \sigma = \sigma_0$ vs. $H_1\colon \sigma \neq \sigma_0$.
  - **1 балл**. Проведено сравнение полученного тест с тестом Вальда для различения гипотез H_0 и H_1.