

## Домашнее задание 3

Дедлайн: 2025-02-27, 23:59.

Оцениваемые задачи:

1. В одном тропическом лесу водятся удавы и питоны. Длина удавов имеет нормальное распределение  $\mathcal{N}(\mu_X, \sigma_X^2)$ . По выборке из 10 удавов оказалось, что  $\sum X_i = 20$  метрам, а  $\sum X_i^2 = 1000$ . Длина питонов имеет нормальное распределение  $\mathcal{N}(\mu_Y, \sigma_Y^2)$ . По выборке из 20 питонов оказалось, что  $\sum Y_i = 60$  метрам, а  $\sum Y_i^2 = 4000$ . Все наблюдения независимы между собой.
  - а) Постройте точечные оценки для  $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2$ .
  - б) Постройте двусторонний 95%-й доверительный интервал для  $\sigma_X^2/\sigma_Y^2$ .
  - в) Проверьте гипотезу  $H_0: \sigma_X^2 = \sigma_Y^2$  против альтернативной  $H_1: \sigma_Y^2 > \sigma_X^2$  на уровне значимости 5%. Укажите точное  $p$ -значение.
  - г) Постройте примерный двусторонний 95%-й доверительный интервал для разницы  $\mu_X - \mu_Y$  с помощью статистики Уэлча.
  - д) Проверьте гипотезу  $H_0: \mu_X = \mu_Y$  против альтернативной  $H_1: \mu_Y > \mu_X$  на уровне значимости 5% с помощью теста Уэлча. Укажите точное  $p$ -значение.
2. Априорное распределение параметра  $\theta$  является треугольным на отрезке  $[0; 40]$  с модой в точке 30. Наблюдаемая величина  $X$  — это индикатор того, что  $\theta > 20$ . Оказалось, что  $X = 1$ .
  - а) Найдите апостериорную плотность  $\theta$ .
  - б) Найдите апостериорное математическое ожидание  $\theta$ .
  - в) Найдите апостериорную медиану  $\theta$ .
  - г) Постройте 94% байесовский интервал наивысшей плотности для  $\theta$ .
  - д) Постройте 94% симметричный по вероятности байесовский интервал для  $\theta$ .

Определение треугольного распределения можно найти, например, на википедии :)

Неоцениваемые задачи в удовольствие:

3. Величины  $X_1, \dots, X_n$  независимы и одинаково распределены с функцией плотности

$$f(x) = \begin{cases} \theta x^{\theta-1}, & \text{при } x \in [0; 1] \\ 0, & \text{иначе.} \end{cases}$$

- а) Оцените значение  $\theta$  с помощью метода максимального правдоподобия.
- б) Оцените дисперсию оценки  $\hat{\theta}_{ML}$  метода максимального правдоподобия.
- в) Как примерно распределена  $\hat{\theta}_{ML}$ ?
- г) Оцените значение  $\theta$  с помощью метода моментов.
- д) Оцените дисперсию оценки  $\hat{\theta}_{MM}$  метода моментов.
- е) Как примерно распределена  $\hat{\theta}_{MM}$ ?

4. Цыганка Роза ничего не понимает в статистике, но у неё всегда с собой колода из 36 карт.

Помогите цыганке Розе построить точный 95%-й доверительный интервал для неизвестной вероятности  $p$  того, что клиента ждёт дальняя дорога и казённый дом.

5. Величины  $X_1, \dots, X_n$  независимы и одинаково распределены с функцией плотности

$$f(x) = \begin{cases} \frac{\theta \exp(-\theta^2/2x)}{\sqrt{2\pi x^3}} & \text{при } x \in [0; +\infty), \\ 0, & \text{иначе.} \end{cases}$$

- Найдите оценку параметра  $\theta$  методом максимального правдоподобия, если по выборке из 100 наблюдений оказалось  $\sum 1/X_i = 12$ .
  - Найдите оценку параметра  $\theta$  методом максимального правдоподобия для произвольной выборки.
  - Найдите теоретическую информацию Фишера  $I(\theta)$ .
  - Пользуясь данными по выборке постройте оценку  $\hat{I}$  для информации Фишера.
  - Постройте 90% доверительный интервал для  $\theta$ . Подсказка:  $\mathbb{E}(1/X_i) = 1/\theta^2$ , интеграл берется, например, заменой  $x = \theta^2 a^{-2}$ .
6. Величины  $X_1$  и  $X_2$  независимы и распределены по Пуассону с интенсивностью  $a$ . Есть две гипотезы,  $H_0: a = 1$  и  $H_a: a = 2$ . Мальвина отвергает  $H_0$  в том случае, если  $X_1 + X_2 \geq 2$ . Найдите вероятность ошибок первого и второго рода.
7. Величины  $Y_1, \dots, Y_n$  независимы и имеют распределение Бернулли с неизвестным  $p$ ,  $\hat{p} = \bar{Y}$ .

- Постройте для неизвестного  $p$  доверительный интервал Вальда. Для этого вспомните про сходимость

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \xrightarrow{\text{dist}} \mathcal{N}(0; 1)$$

и решите неравенство

$$-z_{\text{cr}} \leq \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{\text{cr}}.$$

- Постройте для неизвестного  $p$  доверительный интервал Вильсона. Для этого воспользуйтесь сходимостью

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\text{dist}} \mathcal{N}(0; 1).$$

На этот раз потребуется решить (о ужас!) квадратное неравенство.

Обозначим центр интервала Вильсона с помощью  $\hat{p}_w$ .

- Докажите, что центр интервала Вильсона  $\hat{p}_w$  можно представить как средневзвешенное классической оценки  $\hat{p}$  и тривиальной оценки  $1/2$ ,

$$\hat{p}_w = u\hat{p} + (1-u)(1/2).$$

Найдите веса  $u$  и  $(1-u)$ .

- г) Докажите, что центр интервала Вильсона  $\hat{p}_w$  можно проинтерпретировать следующим образом: добавим  $f$  вымышленных единиц и  $f$  вымышленных нулей в выборку и посчитаем классическую оценку вероятности для выборки с вымышленными наблюдениями,

$$\hat{p}_w = \frac{\sum_{i=1}^n Y_i + f}{n + 2f}.$$

Какому целому числу примерно равно  $f$  для 95%-го доверительного интервала?

- д) Докажите, что интервал Вильсона можно записать в виде

$$\hat{p}_w \pm z_{\text{cr}} \cdot \sqrt{\frac{u\hat{p}(1-\hat{p}) + (1-u)(1/2)^2}{n_w}}.$$

Найдите  $n_w$ , а также веса  $u$  и  $(1-u)$ .

Таким образом, интервал Вильсона слегка корректирует число наблюдений и использует в качестве оценки дисперсии  $Y_i$  средневзвешенное между классической оценкой  $\hat{p}(1-\hat{p})$  и тривиальной оценкой  $1/4$ .

Доверительный интервал Агрести — Коуллы для уровня доверия 95% строится следующим образом. В выборку мысленно добавляют два наблюдения равных единице и два наблюдения равных нулю, считают оценку доли

$$\hat{p}_{ac} = \frac{\sum_{i=1}^n Y_i + 2}{n + 4},$$

а затем строят классический интервал Вальда, используя  $\hat{p}_{ac}$  вместо классической  $\hat{p}$ .

- е) Правда ли, что при уровне доверия 95% центры интервала Агрести — Коуллы и Вильсона совпадают?
- ж) Какой 95%-й интервал шире, Агрести — Коуллы или Вильсона?
- з) С помощью симуляций на компьютере сравните фактическую вероятность накрытия неизвестного параметра  $p$  интервалами Вальда, Вильсона и Агрести — Коуллы с номинальной 95%-й вероятностью. Для экспериментов возьмите  $n = 50$  и различные  $p$  от 0 до 1 с шагом 0.1.