

Домашнее задание № 1.

Тема: Оценивание параметров распределения

Крайний срок сдачи: 1 октября 2021 г., 18:00.

Домашнее задание состоит из пяти теоретических задач T1-T4, T5* и четырёх вычислительных заданий N1-N4. Оценка за домашнюю работу является числом от 0 до 12.

Решение нужно прислать через Ozon Masters bot @ozonm_bot в виде **одного PDF файла (в любом другом формате решения проверяться не будут)**. Этот PDF файл должен содержать

- решения теоретических задач T1-T4, T5* набранные в LaTeX, Word,... или написанные от руки и затем отсканированные;
- программный код для численных заданий N1-N4;
- графики, показывающие, что код работает корректно, и числовые результаты работы кода.

1

Дана выборка x_1, \dots, x_n из показательного закона с функцией распределения $F(x) = 1 - e^{-\theta x}$, $x \geq 0$, где $\theta > 0$ - параметр.

T1 (1.5 балла) Рассматриваются оценки параметра θ вида

$$\hat{\theta}_C = \frac{C}{x_1 + \dots + x_n},$$

где $C > 0$ может зависеть от n . При каком значении параметра C оценка $\hat{\theta}_C$ является несмещённой? При каком значении параметра C она является состоятельной?

Подсказка. Показательное (экспоненциальное) распределение является частным случаем гамма-распределения $\Gamma(k, \theta)$ с плотностью

$$p(x) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-\theta x}, \quad x > 0,$$

где $k, \theta > 0$, $\Gamma(k)$ — гамма функция. Известно, что сумма двух независимых случайных величин с распределениями $\Gamma(k_1, \theta)$ и $\Gamma(k_2, \theta)$ имеет распределение $\Gamma(k_1 + k_2, \theta)$.

N1 (1 балл) Известно, что функция надёжности

$$f(t) = e^{-\theta t}, \quad t > 0,$$

может быть несмещённо оценена при помощи

$$\hat{f}_n(t) = \left(1 - \frac{t}{x_1 + \dots + x_n}\right)^{n-1} \mathbb{I}\{x_1 + \dots + x_n > t\}.$$

Проверьте это утверждение эмпирически: зафиксируйте θ и смоделируйте выборку размера $n = 1000$ из показательного закона. Затем для значений t , выбранных по решётке из некоторого интервала, оцените $\hat{f}_n(t)$. Сравните график оценки $\hat{f}_n(t)$ с графиком функции надёжности. Как изменится визуальное впечатление от такого сравнения, если оценить функцию надёжности по нескольким (2-3) выборкам и взять среднее значение полученных оценок в каждой точке?

2

T2 (1 балл) Пусть X_1, \dots, X_n — i.i.d. случайные величины, имеющие равномерное распределение на отрезке $[0, 1]$. Обозначим r -ую порядковую статистику через $X_{(r)}$. Докажите, что

- (i) $\mathbb{E}X_{(r)} = r/(n+1)$;
- (ii) мода (точка максимума функции плотности) величины $X_{(r)}$ равна $(r-1)/(n-1)$.

N2 (1.5 балла) Известная теорема об асимптотической нормальности выборочных квантилей гласит, что

$$\sqrt{n}\left(X_{(\lfloor \alpha n \rfloor + 1)} - x_\alpha\right) \xrightarrow{Law} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{p^2(x_\alpha)}\right), \quad n \rightarrow \infty, \quad (1)$$

где $\alpha \in (0, 1)$, x_α — теоретическая квантиль (то есть решение уравнения $F(x) = \alpha$), $\mathcal{N}(0, \cdot)$ — нормальное распределение со средним 0 и дисперсией \cdot . Предполагается, что распределение является абсолютно непрерывным с плотностью p , а α выбрано таким образом, что $p(x_\alpha) > 0$.

Пусть X имеет гамма-распределение (см. задачу T1). Перед проведением численного эксперимента, описанного ниже, зафиксируйте α, k, θ .

- (i) Промоделируйте 100 выборок размера $n = 1000$ с этим распределением.
- (ii) Для каждой выборки, оцените левую часть (1).
- (iii) Постройте график квантиль-квантиль, сравнивающие эмпирические квантили в левой части (1) с теоретическими квантилями нормального распределения.
- (iv) Повторите шаги (i)-(iii) для $n = 10000$, $n = 100000$. Убедитесь, что с увеличением n распределение приближается к нормальному.
- (v) Оцените дисперсию выборок при каждом n . Убедитесь, что дисперсия приближается к значению дисперсии предельного закона.

3

T3 (1.5 балла) Дана выборка из распределения с плотностью

$$p(x, \alpha) = \frac{3}{\alpha^3} x^2 e^{-(x/\alpha)^3} \mathbb{I}\{x \geq 0\},$$

где $\alpha > 0$. Найдите оценку параметра α

- (i) методом моментов, используя третий момент;

- (ii) методом максимального правдоподобия;
- (iii) объясните, почему эти оценки совпадают, используя понятие экспоненциального семейства.

N3 (1.5 балла)

- (i) Зафиксируйте значения α и просимулируйте случайную величину, имеющую распределение с плотностью x_1, \dots, x_n . Рекомендуется воспользоваться для этого рассуждением из ТЗ (iii).
- (ii) По полученной выборке численно оцените параметр α методом максимального правдоподобия. Рекомендуется воспользоваться функцией `mle` из пакета `stats4`. Для получения справки наберите в R:
`install.packages("stats4")`
`library("stats4")`
`?mle`

4

T4 (1 балл) Обозначим семейство распределений

$$P_\theta = \left\{ Law(\xi^2), \quad \xi \sim \mathcal{N}(0, \theta) \right\}.$$

- (a) Докажите, что данное семейство является экспоненциальным.
- (b) Используя только свойства экспоненциальных семейств:
 - найдите математическое ожидание и дисперсию величины X ;
 - предполагая, что дана выборка x_1, \dots, x_n , найдите оценку параметра θ методом максимального правдоподобия и методом моментов.

N4 (1 балл) Пусть X_0, X_1, X_2, \dots - цены акций в моменты времени $0, 1, 2, \dots$. Предположим, что теоретические лог-доходности

$$Y_k = \log(X_k/X_{k-1}), \quad k = 1, 2, \dots$$

являются i.i.d. нормально распределёнными случайными величинами со средним 0 и неизвестной дисперсией θ . Параметр θ в этой модели называют волатильностью цены.

Рассмотрите цены акции некоторой компании (например, IBM - `data(ibm)` в пакете `waveslim`) и разделите всю временную шкалу на 10 примерно одинаковых по длине временных интервалов. Для каждого интервала оцените волатильность цены акции. Визуально проверьте, что резкие изменения в цене приводят к резким изменениям волатильности.

5

Т5* (2 балла) Дана выборка x_1, \dots, x_n из равномерного закона распределения на отрезке $[0, \theta]$. Рассматриваются две несмещённые оценки

$$\hat{\theta}_1(x_1, \dots, x_n) = \left(1 + \frac{1}{n}\right)x_{(n)}, \quad \hat{\theta}_2(x_1, \dots, x_n) = x_{(1)} + x_{(n)},$$

где $x_{(1)} = \min(x_1, \dots, x_n)$, $x_{(n)} = \max(x_1, \dots, x_n)$. Докажите, что

$$\frac{\text{Var}(\hat{\theta}_2(X_1, \dots, X_n))}{\text{Var}(\hat{\theta}_1(X_1, \dots, X_n))} = \frac{2n}{n+1}.$$

Комментарий. В данной задаче приводится пример двух несмещённых оценок, дисперсии которых для больших n различаются почти в 2 раза ($2n/(n+1) \rightarrow 2$ при $n \rightarrow \infty$).