Домашнее задание № 1.

Тема: Оценивание параметров. Экспоненциальные семейства распределений

Крайний срок сдачи: 11 октября 2020 г. (до конца дня).

Домашнее задание состоит из пяти теоретических задач T1-T4, T5, четырёх вычислительных заданий N1-N4. Максимальный балл за T1 - T4, N1 - N4 равен 1.25, а за бонусную задачу T5* - 2 балла. Баллы выставляются с шагом 0.25 (то есть, можно получить 0.25 / 0.5 / ... баллов за одну задачу). Таким образом, оценка за домашнюю работу является (возможно, нецелым) числом от 0 до 12.*

Итоговая оценка за курс вычисляется по формуле

$$\min\Big(\mathrm{median}(x_1, ..., x_n), 10\Big),$$

где $x_1, ..., x_n$ *- оценки за домашние работы.*

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.

1

Два игрока играют в игру "камень-ножницы-бумага". Первый игрок выбрасывает камень, ножницы или бумагу с вероятностями $p, q, (1-p-q)$ соответственно ($p, q \in [0,1]$-константы, $p+q \leq 1$).

Второй игрок сначала сначала разыгрывает две i.i.d. величины η_1, η_2 с распределением P на [0,1] и затем выбрасывает камень, ножницы или бумагу с вероятностями $\eta_{(1)}, \eta_{(2)} - \eta_{(1)}, 1 - \eta_{(2)}$ соответственно, где

$$\eta_{(1)} = \min(\eta_1, \eta_2), \qquad \eta_{(2)} = \max(\eta_1, \eta_2).$$

Предполагается, что перед началом игры было сыграно несколько раундов для разминки.

T1 Вычислите значения параметров p и q, которые максимизируют вероятность победы первого игрока. Объясните, каким образом он может оценить эти значения на основе результатов разминочных раундов.

N1 Просимулируйте игру для случаев, когда P является равномерным распределением на $[0,1]$ и бета распределением с параметрами $\alpha = 1, \beta = 2$, а p_1, q_1 меняется от 0 до 1 с шагом $1/3$. Для каждого набора распределения P и каждого набора p_1, q_1 повторите игру 100 раз и сравните вероятности победы первого игрока.

2

T2 Пусть $X_1, ..., X_n$ - выборка из равномерного распределения на отрезке $[0,1]$. Обозначим $r-$ую порядковую статистику через $X_{(r)}$. Докажите, что

(i) $\mathbb{E}X_{(r)} = r/(n+1)$;

(ii) $\mathbb{E}X^2_{(r)} = r(r+1)/((n+1)(n+2))$;

(iii) мода (максимальное значение функции плотности) величины $X_{(r)}$ равно $(r-1)/(n-1)$.

N2 Известная теорема об асимптотической нормальности выборочных квантилей гласит, что

$$\sqrt{n}\Big(X_{(\lfloor \alpha n \rfloor + 1)} - x_\alpha\Big) \xrightarrow{Law} \mathcal{N}\Big(0, \frac{\alpha(1-\alpha)}{p^2(x_\alpha)}\Big), \qquad n \to \infty, \tag{1}$$

где $\alpha \in (0,1)$, $x_\alpha-$ 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 $p(x_\alpha) > 0$, см. [Лагутин М.Б. "Наглядная математическая статистика" , 2007, стр.88].

Пусть X имеет экспоненциальное распределение с функцией распределения

$$F(x) = 1 - e^{-\lambda x}, \qquad x > 0,$$

с параметром $\lambda > 0$. Перед проведением численного эксперимента, описанного ниже, зафиксируйте параметры α и λ.

(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 Дана выборка из распределения Лапласа с плотностью распределения

$$p(x,\theta) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \qquad x \in \mathbb{R}.$$

где $\mu \in \mathbb{R}, \sigma > 0$ - параметры. Найдите оценки параметров α, β

(a) методом максимального правдоподобия;

(b) методом моментов.

N3 Зафиксируйте значения μ, σ и просимулируйте случайную величину, имеющую распределение Лапласа. Повторите симуляции $M = 100$ раз, и по каждой выборке оцените параметры μ и σ методом моментов и методом максимума правдоподобия. Постройте диаграммы размаха, показывающие, какой метод лучше.

4

T4 Обозначим семейство распределений

$$P_\theta = \Big\{ Law(\xi^2), \qquad \xi \sim \mathcal{N}(0, \theta) \Big\}.$$

(a) Докажите, что данное семейство является экспоненциальным.

(b) Используя только свойства экспоненциальных семейств:
- найдите математическое ожидание и дисперсию величины X;
- предполагая, что дана выборка $x_1, ..., x_n$, найдите оценку параметра θ методом максимального правдоподобия и методом моментов.

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

$$Y_k = \log\big(X_k / X_{k-1}\big), \qquad k = 1, 2, ...$$

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

T5* Пусть $X_1, X_2, ...X_n$– последовательность i.i.d. случайных величин с равномерным распределением на отрезке $[a, b]$. Используя понятие достаточной статистики, докажите, что набор величин

$$Z_i = \frac{X_{(i)} - X_{(1)}}{X_{(n)} - X_{(1)}}, \qquad i = 2, ..., (n-1)$$

и вектор $(X_{(1)}, X_{(n)})$ являются независимыми.