Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Расчётно-графическая работа №1

По дисциплине «Математическая статистика»

Михайлов Дмитрий Андреевич
P3206
368530
Медведев Владислав Александрович
P3206
368508

Санкт-Петербург
2025 год

Содержание

Задача №1

Условие задачи.

В файле cars93.csv представлены данные об автомобилях, проданных в некотором автосалоне за 93 год. Какие типы автомобилей представлены в датасете? Какой тип наиболее распространён, какой - менее? Рассчитайте выборочное среднее, выборочную дисперсию, выборочную медиану и межквартильный размах мощности для всей совокупности автомобилей и отдельно для американских и не американских авто. Построить график эмпирической функции распределения, гистограмму и box-plot мощности для всей совокупности и отдельно для каждого типа авто.

Задача №2

Условие задачи.

Предположите, какому вероятностному закону соответствует рапределения показателя, рассмотренного (расчёт выборочных характеристик и визуализация) в задании №1. Оцените параметры данного распределения методом максимального правдоподобия или методом моментов (**математическое обоснование оценки строго обязательно**). Какими статистическими свойствами обладает найденная оценка (**обосновать**)? Найти **теоретические** смещение, дисперсию, MSE (или хотя бы написать теоретические формулы, по которым данные показатели вычисляются, если в итоге получается «очень сложный» интеграл/ряд), информацию Фишера (если определена для вашей модели).

Решение.

Решение этой задачи можно представить аналитически. На основе представленных в первом задании данных можно заметить, что значение мощности автомобиля варьируется в определённом диапазоне, не мало важно заметить следующий факт, что значения мощностей автомобилей распределены симметрично вокруг среднего. Соответственно, принимая мощность автомобиля за случайную величину, можно предположить, что данная случайная величина имеет нормальное распределение.

$$X \sim N(\mu, \sigma^2) \tag{1}$$

Следовательно, оценивая параметры θ_1, θ_2 методом моментов для нормального распределения, вычисляем сперва теоретические моменты $E[X], E[X^2]$.

$$\begin{cases} E[X] = \mu \\ E[X^2] = \mu^2 + \sigma^2 \end{cases}$$

Далее, нам необходимо оценить выборочные моменты $\hat{m}_1, \hat{m}_2$.

$$\begin{cases} \hat{m}_1 = \frac{1}{n}\sum X_i \\ \hat{m}_2 = \frac{1}{n}\sum X_i^2 \end{cases}$$

В конечном итоге, после приравния получаем следующую систему, откуда можно найти оценки наших параметров нормального распределения (1):

$$\begin{cases} \hat{\mu} = \hat{m}_1 \\ \hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2 \end{cases}$$

Таким образом, оценки параметров будут следующие:

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = S^2 \end{cases}$$

Проверим свойство несмещенности у полученных оценок:

$$\begin{cases} E[\hat{\mu}] = \mu \\ E[\hat{\sigma}^2] = \sigma^2 \end{cases}$$

После того, как мы убедились в наличии этого свойства, так же стоит проверить на наличие свойства дисперсии оценок:

$$\begin{cases} Var(\hat{\mu}) = \frac{\sigma^2}{n} \\ Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} \end{cases}$$

Последняя проверка на наличие свойства о средней квадратичной ошибке:

$$E\|\hat{\theta} - \theta\|^2 = Var(\hat{\theta}) + \|bias(\hat{\theta})\|^2 \tag{2}$$

Или же:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + \|bias(\hat{\theta})\|^2 \tag{3}$$

Зная, что для несмещённой оценки $\|bias(\hat{\theta})\|^2 = 0$, получаем итоговую систему для наших оценок:

$$\begin{cases} MSE(\hat{\mu}) = \frac{\sigma^2}{n} \\ MSE(\hat{\sigma}^2) = \frac{2\sigma^4}{n} \end{cases}$$

Информация Фишера $I(\theta)$ определяется как: $I(\theta) = -E\left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right]$.
Для нормального распределения:

$$\begin{cases} I(\mu) = \frac{n}{\sigma^2} \\ I(\sigma^2) = \frac{n}{2\sigma^4} \end{cases}$$

Тогда дисперсия оценок, определяемая через информацию Фишера:

$$\begin{cases} Var(\hat{\mu}) = \frac{1}{I(\mu)} \\ Var(\hat{\sigma}^2) = \frac{1}{I(\sigma^2)} \end{cases}$$

Задача №3

Условие задачи.

Пусть P_θ - выбранное в предыдущем задании распределение, параметризующееся вектором θ (пример - равномерное распределение на $[-2\theta; 4\theta]$, $\hat{\theta} = \bar{X}$), $\hat{\theta}$ - оценка параметра θ, полученная в предыдущем упражнении. Проведите численный эксперимент по следующей схеме:

- Зафиксируйте конкретное значение $\theta = \theta_0$
- Заведите массив $n_1, ..., n_k$ объёмов выборки
- Сгенерируйте из распределения P_{θ_0} достаточно большое количество M выборок объёма n, где n принимает значения из массива $n_1, ..., n_m$. Для каждой сгенерированной выборки вычислите оценку $\hat{\theta}$.
- Эмпирически рассмотреть поведение оценки $\hat{\theta}$ в зависимости от объёма выборки (можно для каждого объёма выборки n_i вывести описательные статистики для оценок, изобразить гистограмму, box-plot, violin-plot).

Приложения

Задача №1

Ссылка на исходник с кодом программы, решающей эту задачу на языке Python. [1]

Задача №3

Ссылка на исходник с кодом программы, решающей эту задачу на языке Python. [2]

Список использованных источников

[1] Задача №1. *URL*: Исходник с кодом, решающий задачу №1.

[2] Задача №3. *URL*: Исходник с кодом, решающий задачу №3.