

Домашнее задание №1 по курсу «Математическая Статистика в Машинном Обучении»

Школа Анализа Данных

26.02.2023

Оценки и сходимости

Задача 1 [2 балла]

Пусть $\mathbf{X} = \{X_1, \dots, X_n\} \sim \text{Uniform}(0, \theta)$, $\hat{\theta} = 2\bar{\mathbf{X}}$. Найдите значения bias, se и MSE этой оценки. Является ли оценка несмещенной? Состоятельной?

Задача 2 [3 балла]

Пусть $\mathbf{X} = \{X_1, \dots, X_n\} \sim \text{Poisson}(\ln \lambda)$, $\hat{\lambda} = e^{\bar{\mathbf{X}}}$. Найдите bias, se и MSE этой оценки. Является ли оценка несмещенной? Состоятельной?

Эмпирическая функция распределения

Задача 3 [2 балла]

Пусть $\hat{F}(x)$ — эмпирическая функция распределения, построенная по выборке $\mathbf{X} = \{X_1, \dots, X_n\}$. Пусть $x, y \in \mathbb{R}$. Найдите ковариацию $\text{Cov}(\hat{F}(x), \hat{F}(y))$.

Задача 4 [2 балла]

Пусть $\mathbf{X} = \{X_1, \dots, X_n\} \sim F(x)$, и пусть $\hat{F}(x)$ — эмпирическая функция распределения. Для фиксированных чисел $a, b \in \mathbb{R}$, таких что $a < b$ определим статистический функционал $T(F) = F(b) - F(a)$. Пусть $\hat{\theta} = \hat{F}(b) - \hat{F}(a)$. Найдите оценку $\hat{\sigma}$ стандартного отклонения и $(1 - \alpha)$ -доверительный интервал.

Задача 5 [2 балла]

Скачайте данные о качестве красных вин. Постройте график для $\hat{F}(x; \mathbf{X})$ для уровня кислотности (pH). Для каждой точки x постройте:

- 95%-ый доверительный интервал на основе неравенства Дворецкого-Кифера-Вольфовица.
- Асимптотический нормальный 95%-ый доверительный интервал для значения $F(x)$.

По значениям уровня кислотности \mathbf{X} подсчитайте оценку $T(\mathbf{X})$ для функционала $T(F) = F(3.5) - F(3.4)$ и найдите оцените аналитически стандартное отклонение $\hat{\sigma}$ оценки $T(\mathbf{X})$. Постройте асимптотический нормальный 95%-ый доверительный интервал для $T(F)$.

Задача 6 [2 балла]

В процессе очистки питьевой воды выпадает значительный осадок. Для его уменьшения можно воздействовать на разные факторы, в т.ч. на количество микроорганизмов в жидкости, способствующих окислению органики. В группу из 261 очистительных установок был добавлен реагент, подавляющий активность микроорганизмов, а состав остальных 119 остался без изменений. Пусть θ — разность в средних значениях количества твердых частиц в этих двух группах установок. Оценить по данным `WaterTreatment` величину θ , оценить стандартную ошибку оценки, построить 95% и 99% доверительные интервалы. Какие выводы можно сделать на основе полученных результатов?

Бутстреп

Задача 7 [2 балла]

Провести моделирование, чтобы сравнить различные типы доверительных интервалов, построенных с помощью бутстрепа. Пусть $n = 50$, $T(F) = \int (x - \mu)^3 dF(x) / \sigma^3$ — коэффициент асимметрии, где F — распределение χ_k^2 с тремя степенями свободы ($k = 3$). Постройте 95% доверительные интервалы для $T(F)$ по выборке $\mathbf{X} = \{X_1, \dots, X_n\}$, используя три подхода на основе бутстрепа.

Задача 8 [3 балла]

Пусть $\mathbf{X} = \{X_1, \dots, X_n\} \sim \text{Exp}(\lambda)$, $\theta = e^{\frac{1}{\lambda}}$ и $\hat{\theta} = e^{\bar{\mathbf{X}}}$. Найдите аналитически плотность распределения $p_{\hat{\theta}}(x)$ оценки $\hat{\theta} = e^{\bar{\mathbf{X}}}$, математическое ожидание $\mathbb{E}(\hat{\theta})$, и дисперсию $\mathbb{V}(\hat{\theta})$, а также bias, se, MSE оценки $\hat{\theta}$. Является ли оценка $\hat{\theta}$ смещенной? Состоятельной?

Задача 9 [2 балла]

Пусть $\mathbf{X} = \{X_1, \dots, X_n\} \sim \text{Exp}(\lambda)$, $\theta = e^{\frac{1}{\lambda}}$ и $\hat{\theta} = e^{\bar{\mathbf{X}}}$. Сгенерируйте выборку \mathbf{X} из $n = 1000$ наблюдений для $\lambda = 0.25$. Нарисуйте гистограмму значений $\{\hat{\theta}_i^*\}_{i=1}^B$ бутстрепных оценок. Эта гистограмма является оценкой распределения $p_{\hat{\theta}}(x)$. Сравните ее с настоящим распределением $p_{\hat{\theta}}(x)$, вычисленным аналитически в предыдущем пункте. Используя бутстреп, подсчитайте величину se и постройте тремя способами 95%-ый доверительный интервал для θ .