MSAI Statistics Home Assignment 1-2
soft deadline: 21/03/2024 23:59 AOE
hard deadline: 11/04/2024 19:00 GMT+3

As announced earlier, grading for HWs consists of points and bonus points. Solving bonus (indicated with a star) problems is not required, but recommended. Solving all homeworks' normal problems correctly will give you a score of 7, solving all homeworks' bonus problems correctly will give you additional 2 points to the score.

Hand-written solutions are accepted if the handwriting is clear enough and scanned with sufficient quality, but LaTeX is always preferable. This homework includes a python task, which can be solved in Google Colab or in a local Jupyter Notebook. It is thus handy to solve everything (both LaTeX and code) in a single Jupyter Notebook.

Solutions obtained with the use of ChatGPT and similar models can be accepted if the solution is clearly indicated as such, and model version and prompt is provided. If the solution is found to be from ChatGPT and similar models without indication and model/prompt detail, the teachers will evaluate the problem as zero points.

Problem 1. Let $X_1, \ldots, X_n \sim U[0, \theta]$ and let $\widehat{\theta} = \max\{X_1, \ldots, X_n\}$. Find:

- (2 points) bias of this estimator
- (2 points) standard error of this estimator
- (1 point) MSE of this estimator
- (1 point) Is this estimator consistent?
- (1 point) Is this estimator strongly consistent?

Problem 2. Let $X_1, \ldots, X_n \sim U[0, \theta]$ and let $\widehat{\theta} = 2\overline{X_n} = \frac{2}{n}\sum_{k=1}^{n} X_k$. Find:

- (2 points) bias of this estimator
- (2 points) standard error of this estimator
- (1 point) MSE of this estimator
- (1 point) Is this estimator consistent?
- (1 point) Is this estimator strongly consistent?

Problem 3. (3 points) Let $X_1, \ldots, X_n \sim Be(p)$ and let $Y_1, \ldots, Y_m \sim Be(q)$. Find the plug-in estimator, its bias and estimated standard error:

1. for p
2. for $p - q$

Problem 4. (3 points) Let $X_1, \ldots, X_n$ be distinct observations (no ties). Prove that there are exactly $\binom{2n-1}{n}$ possible distinct bootstrap samples.

Problem 5. (4 points, computer experiment) Generate $n = 100$ observations from $\mathcal{N}(0, 1)$. Compute the 95% confidence band for the CDF $F(\cdot)$ using DKW inequality. Repeat this $m = 1000$ times and report the fraction of times that the 95% confidence band contains:

1. the true CDF

2. the ECDF

Problem 6. (2 bonus points) Find $\mathbb{P}\left(|\widehat{F}(x) - F(x)| > \frac{t}{\sqrt{n}}\right)$. Hint: remember that ECDF is unbiased. Hint 2: you can choose between a lot of statistical instruments for this one. To name a few: CLT, Chebyshev inequality, Chernoof inequality, DKW theorem.

Problem 7. (2 bonus points) In Kolmogorov's theorem, $F(\cdot)$ is required to be continuous. What is the limit of $D_n = \sqrt{n} \sup_x \left|\widehat{F}(x) - F(x)\right|$ if $X_1, \ldots, X_n \sim Be(p)$ and $F(x)$ is Bernoulli CDF? Hint: CLT or its special case — de Moivre–Laplace theorem.