MSAI Statistics Home Assignment 1-4
soft deadline: 30/03/2025 23:59 AOE
hard deadline: 10/04/2025 19:00 Moscow Time

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 6, solving all homeworks' bonus problems correctly will give you additional 2 points to the score.

Soft deadline is the intended deadline for this homework. Hard deadline is the date and time of homework discussion webinar, where we will discuss solutions to this homework. After the solutions are released, no more homeworks are accepted.

Handwritten 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.

Using AI assistants is allowed, but must be clearly indicated in the solution file and a prompt must be provided. If your homework looks AI-generated but is not indicated as such, you will not get 6 + 2 points and instead you will have to take the exam to demonstrate your knowledge.

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

Problem 2. Let $X_1, \dots, X_n \sim U[0, \theta]$ 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?

Problem 3. There are two groups of people, A and B, who are ill. There are 100 people in each group. People in group A received a standard drug to treat the disease, and 90 people recovered in this group. People in group B are given a new drug to treat the disease, and 85 people recovered in this group. Let p_1 be the probability of recovery under the standard treatment, let p_2 be the probability of recovery under the new treatment, and let $\theta = p_1 - p_2$

1. (1 point) Provide an estimate $\hat{\theta}$ for θ
2. (3 points) Compute its bias and standard error

Problem 4* . In the context of the previous problem:

1. (1 bonus point) Use CLT to find the limiting distribution of $\hat{\theta}$
2. (1 bonus point) Find a 95% normal confidence interval for θ (see definition below)

Def 1 (Normal confidence interval). Suppose that $\hat{\theta} \sim \mathcal{N}\left(\theta, \hat{\text{se}}^2\right)$. Let Φ, be the CDF of a standard Normal and let $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2))$. Let

$$C_n = (\hat{\theta} - z_{\alpha/2}\hat{\text{se}},\ \hat{\theta} + z_{\alpha/2}\hat{\text{se}})$$

Then, $\mathbb{P}\left(\theta \in C_n\right) \to 1 - \alpha$.

Problem 5. Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Find:

1. (1 point) the method of moments estimator of λ
2. (1 point) the maximum likelihood estimator of λ

Problem 6. Let $X_1, \ldots, X_n \sim \text{Uniform}(a, b)$ where a and b are unknown parameters such that $a < b$.

1. (2 point) Find the method of moments estimators for a and b
2. (3 point) Find the maximum likelihood estimators for a and b

Problem 7. (1 point) Prove that KL-divergence is nonnegative:that for any two probability densities $p(x)$ and $q(x)$, $\text{KL}(p||q) \geqslant 0$. Hint: use Jensen's inequality.

Problem 8. Gamma distribution has PDF:

$$f_X(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}$$

1. (1 point) Show that Gamma distribution belongs to the exponential family
$$f_X(x) = h(x) \exp\left(\sum_i \eta_i(\theta) T_i(x) - A(\eta)\right)$$
2. (1 point) Find the sufficient statistics T_i and natural parameters η_i.
3. (2 bonus points) Find the maximum likelihood estimate for θ (assume you know k).
4. (1 bonus point) Is your maximum likelihood estimate for θ biased or not?

Problem 9. Laplace distribution has PDF:

$$f_X(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

1. (1 point) Show that Laplace distribution belongs to the exponential family.
$$f_X(x) = h(x) \exp\left(\sum_i \eta_i(\theta) T_i(x) - A(\eta)\right)$$
2. (1 point) Find the sufficient statistics T_i and natural parameters ν_i.

Problem 10* . You are Morpheus from the Matrix, but instead of giving Neo the chance to choose from blue and red pill, you offer this choice to 100 copies of Agent Smith. 90 agents take the blue pill, and 10 take the red pill. Let p_1 be the probability of an agent to take the blue pill.

1. (1 bonus point) Compute theoretical standard error for $\hat{p_1}$
2. (1 bonus point) Obtain normal confidence interval for p_1
3. (2 bonus points) Using Python, implement the bootstrap with 100 bootstrap samples
4. (1 bonus point) Obtain a normal bootstrap confidence interval (simply use standard error from bootstrap) for p_1
5. (1 bonus point) Compare the intervals and observe that bootstrap confidence interval is slightly narrower. Why do you think that happens?