MSAI Statistics Home Assignment 3-4
soft deadline: 04/04/2024 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 $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$. Find:

1. (1 point) the method of moments estimator of λ
2. (1 point) the maximum likelihood estimator of λ
3. (1 bonus point) and the Fisher information $I(\lambda)$

Problem 2. Let $X_1, \dots, X_n \sim \text{Uniform}(a, b)$ where a and b are unknown parameters such that $a < b$.

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

Problem 3. (1 point) Prove that KL-divergence is non-negative: that for any two probability densities $p(x)$ and $q(x)$, $\text{KL}(p||q) \geqslant 0$. Hint: Jensen's inequality.

Problem 4. Gamma distribution

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

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 5. Laplace distribution

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

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 6. Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$. Let $\theta = e^\mu$ and let $\hat{\theta} = e^{\overline{X}}$ be the MLE. Create a dataset (using $\mu = 5$ and `numpy.random.seed(42)`) consisting of $n = 100$ observations. Use:

1. (1 bonus point) delta method
2. (1 bonus point) parametric bootstrap
3. (2 bonus point) nonparametric bootstrap

to get $\hat{\text{se}}$ and from this a 95-percent confidence interval for θ.