42.510 Statistical Learning for Data Science FALL 2025

Assignment 1

Due: October 1, 23:59

Please submit solutions to Questions 1 to 10 via Gradescope. Entry code: `PGZE7D`

Probability review

1. A multiple choice quiz has 15 questions with 4 choices for each. If you guess every answer, what is the probability that you score at least 50% on the quiz?

2. An instructor is going to give the grades A, B, C, D, F according to the following scale

   (a) A: grade $> \mu + 1.5\sigma$;

   (b) B: $\mu + 0.5\sigma <$ grade $\leq \mu + 1.5\sigma$;

   (c) C: $\mu - 0.5\sigma <$ grade $\leq \mu + 0.5\sigma$;

   (d) D: $\mu - 2\sigma <$ grade $\leq \mu - 0.5\sigma$;

   (e) F: grade $\leq \mu - 2\sigma$.

   What percentage of students get each letter grade assuming that the grades follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$?

3. Let X be a Poisson random variable with mean 25. Define $p = \mathbb{P}(X \geq 32)$.

   (a) Use Markov's inequality to obtain an upper bound on p.

   (b) Use Chebyshev's inequality to obtain an upper bound on p.

   (c) Approximate p using the central limit theorem (also give a brief justification why Poisson may be approximated by a normal distribution).

4. Two types of coins are produced at a factory: a fair coin and a biased one that comes up heads 55 percent of the time. We have one of these coins, but do not know whether it is a fair coin or a biased one. In order to ascertain which type of coin we have, we shall perform the following statistical test: we shall toss the coin 1000 times. If the coin lands on heads 525 or more times, then we shall conclude that it is a biased coin, whereas if it lands on heads less than 525 times, then we shall conclude that it is a fair coin. If the coin is actually fair, what is the probability that we shall reach a false conclusion? What would it be if the coin were biased?

Statistics review

5. Let $X_n, n = 1, 2, \ldots$ be random variables with pdf f_θ for some parameter $\theta > 0$. For each of the following cases find the MLE of θ. You may assume $\theta > 0$.

   (a) $f_\theta(x) = \begin{cases} \theta x^{-(\theta+1)} & x \geq 1, \\ 0 & x < 1. \end{cases}$

   (b) $f_\theta(x) = \begin{cases} \sqrt{\theta} x^{\sqrt{\theta}-1} & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$

   (c) $f_\theta(x) = \begin{cases} \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right) & x \geq 0, \\ 0 & x < 0. \end{cases}$

6. Find bias, standard error and MSE of estimators of λ and θ below.

   (a) Let $X_1, \ldots, X_n$ be i.i.d. random variables $\sim$ Poisson(λ) and let $\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

   (b) Let $U_1, \ldots, U_n$ be i.i.d. random variables $\sim$ Unif$(0, \theta)$ and let $\hat{\theta}_n = \max\{U_1, \ldots, U_n\}$.

7. A person claims to be able to taste whether tea or milk was added first to a cup of English tea. To test her claim, 12 cups of visually indistinguishable tea are prepared, of which 6 of the cups are prepared tea-first, the other 6 milk-first. Being aware of this experimental setup, she would always try to pick 6 of the cups as tea-first, and the other 6 as milk-first. After tasting each cup of tea, she correctly identifies 5 of the tea-first cups (making 1 mistake), and 5 of the milk-first cups (also making 1 mistake).

   Compute the p-value, that is, the probability that one can do at least as well as her by guessing, and hence perform a hypothesis test at the $\alpha = 0.05$ level.

Statistical Decision theory

8. (KL-divergence) Let f, g be pdf's which are both supported on $A \subset \mathbb{R}$. A measure of "distance" (which is not actually a distance, mathematically) between the two distributions is given by

$$KL(f||g) = -\int_A f(x) \ln\left(\frac{g(x)}{f(x)}\right) \mathrm{d}x.$$

   This is called *Kullback-Leibler divergence*, or, *KL divergence* (or, also *relative entropy*) between f and g or X and Y where $X \sim f, Y \sim g$. Find the KL-divergence between the following distributions. Show computations.

   (a) What is the KL-divergence between $X \sim \mathcal{N}(a, \sigma^2)$ and $Y \sim \mathcal{N}(b, \sigma^2)$?

   (b) What is the KL-divergence between $X \sim$ Exp(a) and $Y \sim$ Exp(b)?

9. For a given loss function L, the risk R is given by the expected loss

$$R(f) = \mathbb{E}[L(Y, f(X))],$$

where $f = f(X)$ is a function of the random predictor variable X.

(a) Consider a regression problem and the squared error loss

$$L(Y, f(X)) = (Y - f(X))^2.$$

Derive the expression of $f = f(X)$ which minimizes the associated risk.

(b) What if we use the absolute (L1) loss instead:

$$L(Y, f(X)) = |Y - f(X)|.$$

What will be the minimizer?

10. Consider a binary classification problem where response $Y \in \{-1, 1\}$ and predictors $\mathbf{X} \in \mathbb{R}^p$. Consider the following imbalanced loss function: $L(-1, -1) = L(1, 1) = 0$, $L(-1, 1) = C_+ > 0$ (cost of false positive), $L(1, -1) = C_- > 0$ (cost of false negative). Compute the Bayes classifier/predictor at $\mathbf{X} = \mathbf{x}$.

Additional practice questions: NO NEED TO SUBMIT

11. Choose a number X at random from the set of numbers $\{1, 2, 3, 4, 5\}$. Now choose a number at random from the subset no larger than X, that is, from $\{1, ..., X\}$. Call this second number Y.

    (a) Find the joint mass function of X and Y.

    (b) Find the conditional mass function of X given that $Y = i$. Do it for $i = 1, 2, 3, 4, 5$.

    (c) Are X and Y independent? Why?

12. A complex machine is able to operate effectively as long as at least 3 of its 5 motors are functioning. If each motor independently functions for a random amount of time with density function
$$f(x) = xe^{-x}, \quad x > 0,$$
compute the density function of the length of time that the machine functions.

13. Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hit their target with probability 0.6, compute the expected number of ducks that are hit. Assume that the number of ducks in a flock is a Poisson random variable with mean 6. You may provide the final expression or an approximation to the solution.

14. Let $X_n, n = 1, 2, \ldots$ be random variables such that $\mathbb{P}\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2}$ and $\mathbb{P}\left(X_n = n\right) = \frac{1}{n^2}$. Does X_n converge in probability? Does it converge in L^2?

15. If $Z \sim \mathcal{N}(0, 1)$, find $\mathrm{Cov}(Z, Z^2)$.

16. Consider testing $\mathbf{H}_0 : \mu = 0$ vs $\mathbf{H}_\mathrm{A} : \mu \neq 0$ based on a random sample of size n from a $\mathcal{N}(\mu, 1)$ distribution.

    (a) Calculate the p-values for the following three cases:
    (i) $\overline{x} = 0.1$, $n = 100$; (ii) $\overline{x} = 0.1$, $n = 400$; (iii) $\overline{x} = 0.1$, $n = 900$.

    (b) Given the significance level $\alpha = 0.01$, conduct hypothesis tests for the three cases in (a).

17. A certain component is critical to the operation of an electrical system and must be replaced immediately upon failure. If the mean lifetime of this type of component is 100 hours and its standard deviation is 30 hours, how many of these components must be in stock so that the probability that the system is in continual operation for the next 2,000 hours is at least 0.95?

18. Consider non-negative integers a, b. Show that if $a \leq b$, then $a \leq \sqrt{ab}$. Now consider a two class classification problem where the elements of the i-th class are generated from a pdf $f_i, i = 1, 2$. For simplicity consider that the elements lie in $\mathbb{R}$. Suppose the decision regions (say, $\mathcal{R}_1$ and $\mathcal{R}_2$) for classification are chosen by minimizing the probability of misclassification. Show that this misclassification probability satisfies:
$$\mathbb{P}(\text{misclassification}) \leq \int_{-\infty}^{\infty} \sqrt{f_1(x) f_2(x)}\, \mathrm{d}x.$$