Statistical Inference

HW Author: *Farshad Hessami* Instructor: *Mohammadreza A. Dehaqani*

Fall 2025

Homework 1

- If you have any questions about the homework, don't hesitate to drop an email to the HW Author.
- Feel free to use the class group to ask questions — our TA team will do their best to help out!
- Please consult the course page for important information on submission guidelines and delay policies to ensure your homework is turned in correctly and on time.
- This course aims to equip you with the skills to tackle all problems in this domain and encourages you to engage in independent research. Utilize your learnings to extend beyond the classroom teachings where necessary.

Problem 1: Covariance

A box contains n tickets, labeled $1, \dots, n$. Two tickets are drawn without replacement. Let X be the first number drawn and Y be the second. Show that $Cov(X, Y) = \frac{-(n+1)}{12}$.

Problem 2: Variance and Covariance

Let $Y = 5X + \epsilon$, where ϵ follows a standard normal distribution ($\mathcal{N}(0, 1)$) and X follows a uniform distribution on the interval $(-1, 1)$. Assume that X and ϵ are independent.
(a) Determine the mean and variance of Y.
(b) Calculate the expected value of Y^2.
(c) Find the conditional expected value of Y given $X = x$.
(d) Determine the expected value of Y^3.
(e) Calculate the covariance between ϵ and ϵ^2. Are ϵ and ϵ^2 independent?
(f) Prove that for random variables X and Y, the covariance between the linear transformations $n + mX$ and $h + gY$ is equal to $mg\,\mathrm{Cov}(X, Y)$.
(g) Find an upper bound for |Cov(X, Y)| using an inequality you learned in class.

Problem 3: Cumulative Distribution Function

Let X be a non-negative random variable with cumulative distribution function (CDF) F. Show that

$$\mathbb{E}[X] = \int_0^{\infty} (1 - F(t))\, dt.$$

Hint: Consider expressing X as $X = \int_0^1 1_{\{t<X\}}\, dt$, and then assume the possibility of interchanging the order of expectation and integral.

Problem 4: The Blue Taxi Problem (Bayesian Inference)

In a city with one hundred taxis, one is painted blue, while the other 99 are green. During a hit-and-run incident at night, a witness claims to have seen a blue taxi leaving the scene and identifies it as the one involved. Consequently, the police arrested the blue taxi driver on duty that night. The driver asserts his innocence and has sought your legal representation to defend him in court. To build a case for reasonable doubt, you decide to conduct a test involving a scientist to assess the witness's ability to differentiate between blue and green taxis under conditions similar to the night of the accident. The collected data indicates that the witness correctly perceives blue cars as blue 99 percent of the time but misidentifies green cars as blue 2 percent. Your task is to deliver a brief speech to the jury, to give them sufficient doubt regarding your client's guilt. Your speech should be concise and clear, as most jurors may not have a background in this field. Using an illustrative table rather than complex formulas may aid their understanding.
Note: Articulate a compelling defense for the accused using clear and convincing language. This bonus question makes up 10% of your grade and can cover mistakes and shortcomings so write a well-presented argument, emphasize reasonable doubt in your narrative, avoiding complex wording to ensure the jury's comprehension.

Problem 5: Poisson Distribution (Extra Point)

Let X be a Poisson random variable with parameter λ. Show that the maximum of $P(X = i)$ occurs at $\lfloor \lambda \rfloor$, where $\lfloor \lambda \rfloor$ is the greatest integer less than or equal to λ.

Hint: Let p be the probability mass function of X. Prove that

$$p(i) = \frac{\lambda}{i} p(i-1).$$

Use this to find the values of i at which p is increasing and the values of i at which it is decreasing.

Problem 6: The Lifetime of a Device

Let X be the natural lifetime of a device that also fails if a catastrophe such as a shock occurs. Let Y be the time until the next catastrophe. Suppose that X and Y are independent exponential random variables with parameters λ and μ, respectively. Given that $X < Y$, find the expected lifetime of the device.
Warning: Since $X < Y$, we might fallaciously think that the expected lifetime of the device is $1/\lambda$.

Problem 7: Expectation Is Always The Best

Let X and Y be random variables. We wish to predict Y using a function of X, denoted as $g(X)$. The quality of a predictor is measured by its **mean squared error (MSE)**, defined as:

$$\text{MSE} = E[(Y - g(X))^2]$$

Prove that the conditional expectation, $g(X) = E[Y|X]$, is the "best" possible predictor by showing that it minimizes the mean squared error.

Specifically, prove that for any arbitrary function g:

$$E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$$

Problem 8: Prediction

Let the point (X, Y) be uniformly distributed over the half disk $x^2 + y^2 \leq 1$, where $y \geq 0$. If you observe X, what is the best prediction for Y? If you observe Y, what is the best prediction for X? For both questions, "best" means having the minimum mean squared error.

Problem 9: Signal Quantizer (Extra Point)

In digital signal processing, raw continuous analog data X must be quantized, or discretized, in order to obtain a digital representation. In order to quantize the raw data X, an increasing set of numbers $a_i, i = 0, \pm 1, \pm 2, \ldots$, such that $\lim_{i \to \infty} a_i = \infty$ and $\lim_{i \to -\infty} a_i = -\infty$ is fixed, and the raw data are then quantized according to the interval $(a_i, a_{i+1}]$ in which X lies. Let Y denote the observed discretized value—that is,

$$Y = y_i \quad \text{if } a_i < X \leq a_{i+1}$$

The distribution of Y is given by

$$P(Y = y_i) = F_X(a_{i+1}) - F_X(a_i)$$

Suppose now that we want to choose the values $y_i, i = 0, \pm 1, \pm 2, \ldots$ so as to minimize $E[(X - Y)^2]$, the expected mean square difference between the raw data and their quantized version.

1. Find the optimal values $y_i, i = 0, \pm 1, \ldots$.
2. For the optimal quantizer Y, show that $E[Y] = E[X]$, so the mean square error quantizer preserves the input mean.
3. Show that $\text{Var}(X) = \text{Var}(Y) + E[(X - Y)^2]$.

Problem 10: Different Distribution Contexts

(a) A fire station will be located along a finite-length road A. Where should the station be located to minimize the expected distance from the fire if fires occur at points uniformly chosen on $(0, A)$? In other words, choose a to minimize $E[|X - a|]$, where X is uniformly distributed over $(0, A)$.
(b) suppose the road is infinite, stretching from point 0 outward to ∞. Where should the fire station be located if the fire distance from point 0 is exponentially distributed with rate λ? We want to minimize $E[|X - a|]$, where X is now exponential with rate λ.

Problem 11: Inequalities

(a) Chebyshev Inequality: According to Chebyshev's inequality, for a given random variable X with an average value of μ and a variance of σ^2, the following holds when $k = 2$:

$$P(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}.$$

This inequality indicates an upper bound on the probability, not the exact value, meaning the actual probability can significantly differ. For the specified distributions below, calculate the precise probability that $|X - \mu| \geq 2\sigma$.

- Let X be a continuous random variable whose probability density function (pdf) is defined as $f(x) = \frac{1}{2}$ within the interval $0 \leq x < 2$, and $f(x) = 0$ for all other values.

- X is a discrete random variable with a probability mass function (pmf) given by $p(1) = p(-1) = \frac{1}{2}$, and 0 everywhere else.

(b) Markov's inequality: Consider a random variable X for which the fourth moment about the mean, μ, denoted as $E[(X-\mu)^4]$, is defined. Prove that for any positive constant c, the following inequality holds:

$$P(|X-\mu| > c) \leq \frac{E[(X-\mu)^4]}{c^4}.$$

(c) (Extra Point) Jensen's inequality: Consider two distributions, P and Q, defined on the same measurable space X. Assume there exists a measure μ such that both P and Q are absolutely continuous with respect to μ (a common choice for μ could be $P+Q$). Let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ represent the Radon-Nikodym derivatives of P and Q with respect to μ, serving as the respective density functions. The Kullback-Leibler divergence from P to Q is given by

$$D_{\mathrm{KL}}(P\|Q) = \int_X p(x) \log\left(\frac{p(x)}{q(x)}\right) d\mu(x).$$

Demonstrate that $D_{\mathrm{KL}}(P\|Q)$ is non-negative, and it equals zero if and only if the distributions P and Q are identical. [Hint: Use Jensen's inequality, acknowledging that a convex function f satisfies $f''(t) > 0$ for almost every t, which indicates strict convexity.]

Problem 12: What Are AutoEncoders Doing Here? (Extra Point)

Autoencoders are a group of neural networks used for feature extraction. The autoencoder consists of an encoder and a decoder. The encoder is a function of input images. Let X represent input images drawn from $N(0,1)$, and let the encoder function be $\mathrm{En}(X)$. It is known that the latent space of this autoencoder, which compacts all image information, is a sample from $\exp(\mathrm{En}(X))$. What is the average of the latent vector of these images?

Problem 13: Gambler's Ruin

Simulate this problem with python, preferebly in an ipynb file.
The Gambler's Ruin is a classical problem in probability theory that explores the concept of random walks and the likelihood of a gambler losing their entire stake over time. In this problem, you will create a computer program to simulate the Gambler's Ruin scenario and analyze the results. Here is the task:
• Create a program that simulates the Gambler's Ruin scenario. The scenario is as follows:
1: A gambler starts with an initial stake (a certain amount of money).
2: The gambler repeatedly bets a fixed amount on a game with a win probability of p (e.g., 0.5 for a fair coin toss).
3: If the gambler wins a bet, their stake increases by the bet amount. If they lose, their stake decreases by the bet amount.
4: The game continues until the gambler reaches a desired target amount or loses their entire stake (goes broke).
• Implement the following steps in your program:
1: Simulate the gambling scenario for a specified number of rounds (e.g., 1,000 rounds).
2: Track the gambler's stake after each round.
3: Calculate and display statistics, such as the probability of reaching the target amount before going broke.
4: Report the mean value for more iterations and use an appropriate graph to visualize this metric.
• Extend your program to allow for different initial stakes, bet amounts, win probabilities, and target amounts.
• Visualize the results using charts, such as line plots showing the gambler's stake over time.

Problem 14: Sorting Algorithm Analysis

Simulate this problem with python, preferebly in an ipynb file.

Consider the 4 sotring algorithms: bubble sort, insertion sort, merge sort, and quick sort.
randomly generate a list of 1000 integers (make sure it's shuffled) and sort them using the above algorithms, each 100 times (reshuffle each time). each time record the time it took to sort the list.

(a) Plot the duration distribution for each method. Include the mean, median, standard deviation in the plot.
Sort the algorithms based on the mean duration.
Looking at the standard deviation, which algorithm is the most stable? Which one is the most volatile? Why do you think so?
(b) Are any of the distributions skewed?
Calculate the Kurtosis of the distributions. Which algorithm has the least kurtosis? What does this say about the sorting-algorithm's consistency in terms of duration?
(c) Draw the boxplots for the durations of each algorithm.
Which algorithm has the most outliers?

Note: Feel free to use built-in sorting functions—no need to reinvent the wheel! Let's focus on the fun part: analyzing the results!

Problem 15: Cookie Monster

Simulate this problem with python, preferebly in a ipynb file
Salt and pepper is a type of noise sometimes seen on digital images, similar to the one in **Figure 1**:
each noise pixel is either complete white or complete black.
now an image (**Figure 2**) consider 3 types noise: uniform, gaussian and exponential.
You're going to apply these noises on the image.

(a) For a noise ratio of %10, %15, %20, %30, %40, %50 of all pixels apply the uniform noise across the whole image, apply the gaussian noise with a mean and standard deviation that are randomly sampled (uniformly), and the exponential noise with $\lambda = 1$.
Plot the results of the above in three rows of 6 columns (for each noise ratio).
Note: remember, the noise color may only be black, or white.

(b) Alright, let's say that the face of the cookie monster is the region of interest. if at least the 20% of the pixels of his face are corrupted by noise, consider it as corrupted.
For each of the noise ratios, %20, %30, %40, %50, and the three types of noise, apply the noise to the image 100 times and count the times that the face is corrupted.
Report the average count of the corrupted faces for each noise ratio and type of noise.
Are the power of the noise types the same? If not, which one is the most detrimental? Why do you think so?

(c) One of the effective ways of salt-and-pepper noise removal is the median filter(don't freak out, it's just a filter that replaces the pixel value with the median of the neighboring pixels, you can use a library for this). apply the noise ratios of the section (b) with the 3 noise types (just once, not 100 times), and then apply the median filter to the image.
Plot the results of the above in three rows of 4 columns (for each noise ratio). Do the images look good now? depends? why ,if so?

Figure 1: Example of a Salt-and-pepper noisy image

Figure 2: Subject image (Download)