Statistics: Internals 2

Srikanth Pai, Asst. Professor, MSE

15 October 2025

Maximum Marks: 40 **Duration: 90 Minutes**

The total marks for this paper is well over 40. Attempt as many problems as you can. I will correct your paper and truncate the marks at 40.

1. (2 marks each) Answer the following questions. Give brief explanations for statistical puzzles.

   (a) Write the full form of MLE.

   (b) Compute the sample mean of -1,2,-3,4,-5.

   (c) What is a sufficient statistic for θ, if samples $X_1, X_2, ..., X_n$ are drawn from uniform distribution on interval $[1, \theta]$? No justifications required.

   (d) Suppose we sample from a normal distribution with parameters μ, σ^2 (usual convention). Explain why the sample mean is a consistent estimator for the parameter μ. The answer is a single line.

   (e) A news article states that "there is a 95% chance that the average Indian household earns between Rs.48,000 and Rs.55,000 per month." The claim is based on a 95% confidence interval computed from survey data. What is misleading about this interpretation? State the objection precisely.

   (f) What is the conjugate distribution for the success probability of a Bernoulli random variable? Explain why in a single line. You can quote results discussed in class.

   (g) In the Netherlands, nurse Lucia de Berk was convicted after several patient deaths occurred during her hospital shifts. Prosecutors argued that the probability of so many deaths happening "by chance" was 1 in 342 million, and used this to claim near certainty of guilt. Later, statisticians showed this reasoning was flawed. Explain precisely what the logical or statistical mistake was in treating such a small "chance" as strong evidence of guilt.

2. (4 marks) The financial crisis of 2008 highlighted concerns about excessive CEO compensation. In a Gallup poll conducted in 2009, a random sample of 998 adults was asked: *"Do you favor or oppose the federal government taking steps to limit the pay of executives at major companies?"* Of those surveyed, 59% responded in favor.

   (a) Verify the reported margin of error of approximately ± 3 percentage points for a 95% confidence interval.

(b) Construct the 95% confidence interval for the proportion of adults who favor pay limits.

3. (a) (3 marks) State the Central Limit Theorem **precisely and rigorously** for independent and identically distributed random variables with finite mean μ and variance σ^2.

   (b) (3 marks) Prove that, for each fixed $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \int_0^{n+\sqrt{n}\,z} \frac{x^{n-1}e^{-x}}{(n-1)!}\,dx \;=\; \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

   [Careful; Tricky]

4. (4 marks) Five numbers $X_1, \ldots, X_5$ are independently drawn from the Uniform$(0,1)$ distribution. Let Y denote the second largest of these five observations.

   Find the value of y at which the pdf of Y attains its maximum.

5. (6 marks) Let $X_1, X_2, \ldots, X_n$ be i.i.d. Poisson(λ). Let $\bar{X}$ denote the sample mean.

   (a) Show that the statistic $\hat{\lambda} = \bar{X}$ is unbiased and sufficient for λ.

   (b) Show that $\hat{\lambda}$ attains the Cramér–Rao lower bound and hence is the minimum variance unbiased estimator.

6. (5 marks) Let X be a single random sample from a normal distribution

$$X \sim N(\mu, 1),$$

   where μ is unknown.

   (a) Show that the normal family is conjugate for μ under this likelihood.

   (b) Assume the prior on μ is $\mu \sim N(0,1)$. Show that the posterior distribution of μ given the sample is normal again.

   (c) Compute the posterior mean.

7. (6 marks) Each evening, the media report various averages and indices that are presented as portraying the state of the stock market. But do they really convey useful information? Some financial analysts argue that speculative markets tend to rise and fall randomly, as though some hidden roulette wheel were spinning out the figures.

   One way to test this theory is to model the up-and-down behavior of the markets as a geometric random variable. If this model were correct, we could argue that the market does not use yesterday's history to decide whether to rise or fall today, and that this history does not alter the probability p of a rise or $1-p$ of a fall.

   Suppose that on a given Day 0 the market rose, and on the following Day 1 it fell. Let the random variable X represent the number of consecutive days the market falls before it rises again (a "success"). Then X follows a geometric distribution with pmf

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots$$

   A sample of market runs produced the observed frequencies shown below.

k	Observed Frequency	Expected Frequency
1	72	74.14
2	35	31.20
3	11	13.13
4	6	5.52
5	2	2.32
6	2	1.69

For example, the entry "k = 3" with an observed frequency of 11 means that in 11 instances, the market declined for three consecutive days before rising again on the fourth day.

(a) Derive the maximum likelihood estimator $\hat{p}$ for the geometric distribution based on a random sample $X_1, X_2, \ldots, X_n$. Compute $\hat{p}$ using the observed frequency.

(b) Using $\hat{p}$, explain how the expected frequencies would be computed for each k. Explain the idea and show it for $k = 2$ only.

(c) Interpret the meaning of $\hat{p}$ in the context of stock market movements. What does this model suggest about whether markets remember their past?

8. (3 marks) Consider a random variable s that is uniformly distributed on $[0, 1]$. Define

$$X_n(s) = \begin{cases} 1, & 0 \leq s < \dfrac{n+1}{2n}, \\ 0, & \text{otherwise}, \end{cases} \qquad X(s) = \begin{cases} 1, & 0 \leq s < \dfrac{1}{2}, \\ 0, & \text{otherwise}. \end{cases}$$

Show that $X_n \to X$ almost surely. [Be careful at $s = 1/2$.]