Department of Computer Science and Engineering, IIT Roorkee
End Semester Examination - Autumn, 2024
CSE-373 Probability Theory for Computer Engineering

Time: 3 hrs. (Answer all questions: *the paper is printed on both sides*) Full Marks: 100

1.

a. Let (X, Y) be a random point drawn according to the Uniform distribution on the unit disk $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. Are X and Y independent? Are they correlated?
[2+3]

b. Let $X_1, X_2, \cdots, X_n$ be *i.i.d.* random variables each with Poisson distribution with mean 1. Let $S_n = X_1 + \cdots + X_n$, $n \geq 1$, and $\phi_n(t)$ be the moment generating function of S_n. Find the smallest n such that $\Pr(\phi_n(S_n) > 1) \geq 0.99$. Approximate the smallest n using the central limit theorem?
[4+3]

c. Let $(X_1, X_2, X_3) \in \mathbb{R}^3$ be a random point where X_i's are *i.i.d.* from $N(0,1)$. Let C be the sphere centered at the origin and passing through (X_1, X_2, X_3). What is the probability that the surface area of C is less than π? Find the conditional probability for the same using the information that (X_1, X_2, X_3) is in the first quadrant. What happens to the conditional probability if the information is $|X_1| < 1/5$, instead of (X_1, X_2, X_3) in the first quadrant?
[3+2+3]

2.

a. Suppose that $U_1, U_2, \ldots, U_n$ are *i.i.d.* Unif (0, 1) random variables and $S_n = \sum_{i=1}^{n} U_i$. Define the random variable N by $N = \min\{k \mid s_k > 1\}$.
   i. Show that $\Pr(S_k \leq t) = t^k/k!$, for $0 \leq t \leq 1$.
   ii. Find the probability mass function (pmf) of N using (i). Hence find the expected value of N.
   iii. Using the above results, propose an algorithm to calculate the value of e (the base of the natural log) by simulation. Justify your answer.
[4+4+3]

b. Let X_1 and X_2 be two $i.i.d.$ observations from the discrete distribution that satisfies $\Pr(X = \theta) = \Pr(X = \theta + 1) = \Pr(X = \theta + 2) = 1/3$, where θ is an unknown integer.
   i. Is median of X_1 and X_2 an unbiased estimator for θ? Why?
   ii. Find MLE of θ when $|X_1 - X_2| = 2$. Describe the role of the condition $|X_1 - X_2| = 2$ in connection with the estimator described in (i).
[3+3]

c. Suppose your teacher asked each of you to write a random integer X between 0 to 10 during one of the tutorial sessions. The following was the result.

X	0	1	2	3	4	5	6	7	8	9	10
No. of students who chose X	0	02	03	04	04	01	03	20	03	05	05

(Please turn over)

Suppose that the teacher before checking the above data, tries to guess the integer that is chosen by the maximum number of students by asking questions only of the form: is $X = x$, with the guess being independent of his previous guess. He uses Binomial $(10, 2/5)$ distribution for this process. What is the expected number of guesses for the teacher to guess it correctly? [3]

3. Let $X_1, X_2, ..., X_n$ be n *i.i.d.* random numbers on the real line chosen from Unif (θ_L, θ_R) distribution where $\theta_L < \theta_R$. Let $[X_L, X_R]$ be the closed interval with the smallest distance containing the points $X_1, X_2, ..., X_n$.
   a. Find the MLE of the interval (θ_L, θ_R). What is the bias of the length of the MLE of (θ_L, θ_R)? Is it bigger than the bias of the length of $[X_L, X_R]$? Is the MLE consistent for (θ_L, θ_R)? [5 + 8 + 2 + 2]
   b. Assume that $\theta_L = 0$ and θ_R follows Pareto distribution with parameters (α, β) whose probability density function is given by $p(\theta_R) = \frac{\alpha\beta^\alpha}{\theta_R^{\alpha+1}}$, $\alpha > 0,\ \theta_R > \beta > 0$. Find the Bayes estimator for θ_R. [6]
   c. Compare the MLE and Bayes estimator of θ_R with respect to Mean Squared Error. [5]

4.
   a. Suppose the daily dissolved oxygen concentration for a water stream follows a Normal distribution with unknown mean $\mu \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. It has been recorded over 30 days which sums up to a sample average of 2.5 mg/liter and sample standard deviation of 2.12 mg/liter. Find a 90% confidence interval for variance. [6]

   b. Let $X_1, X_2, ..., X_n$ be n *i.i.d.* random variables drawn from Normal distribution with unknown mean $\mu \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. Using the idea from **Item a.** above, find an unbiased estimator for σ^3. [5]

   c. Suppose a mechanic has claimed that the average time to fill up a cycle tire is at least 98.6 seconds. In a sample of 100 flat tires, we observed that the average time required to fill a tire is 98.74 seconds with a standard deviation of 1.1 seconds. Will you accept the original claim at the 5% level? [6]

5. Let X and Y be two binary random variables with $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$ and $\Pr(Y = 1 \mid X = 1) = \alpha$, $\Pr(Y = 1 \mid X = 0) = \beta$.
   a. Find Shannon entropy and the Fisher information of Y. How do these two measures of information behave for various p? [4+4]
   b. Find the mutual information $I(X, Y)$. Explain how this can be used as a measure of independence between X and Y. How is $I(X, Y)$ related to $Cov(X, Y)$, if any? [3 + 2 + 2]

The End