[This question paper contains 4 printed pages.]

**Your Roll No.....**

**Sr. No. of Question Paper : 5577** **J**

Unique Paper Code : 2222013603

Name of the Paper : Statistical Analysis in Physics

Name of the Course : **B.Sc. (H) Physics**

Semester : VI

Duration : 2 Hours Maximum Marks : 60

**Instructions for Candidates**

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt four questions in all.
3. Question Number 1 is compulsory.
4. Non-programmable calculators are allowed. Use of statistical tables are allowed.

1. Attempt **any five** of the following: (3×5=15)

   (a) What do we mean by Cumulative Distribution Function (CDF)? Mention a few properties of a CDF. How can we get CDF from a given probability density function?

   (b) A particular study showed that 12 % of men will likely develop prostate cancer at some point in their lives. A man with prostate cancer has a 95% chance of a positive test result from a medical screening exam. A man without prostate cancer has a 6% chance of getting a false positive test result. What is the probability that a man has cancer given that he has a positive test result?

(c) State central limit theorem. Give an example of one distribution for which we cannot use this theorem.

(d) Suppose we have a 6-sided white dice and a 4-sided black dice. Each dice is fair. If X and Y are the outcomes of white and black dice respectively. Assuming each dice is fair, find the covariance of X and Y.

(e) What do you mean by joint probability distribution? Consider X and Y be two jointly continuous random variables with joint probability density function given by

$$f_{XY}(x,y) = x + Cy^2 \qquad 0 \le x \le 1, 0 \le y \le 1$$
$$= 0 \qquad \text{otherwise}$$

Find $P[0 \le x \le 1/2,\ 0 \le y \le 1/2]$.

(f) What does conjugate priors mean? If you are using bayesian linear regression to estimate the parameters and the distribution of the likelihood is binomial, what should be the conjugate prior (write expression also) so that the posterior is normalized?

2. (a) State and derive Bayes theorem. Discuss at least two advantages of using Bayes theorem for the parameter estimation. What does the denominator in Bayes theorem represent? (7)

(b) A coin has an unknown bias θ, representing the probability of landing heads. If our prior belief is that θ Beta(2, 2) and we toss the coin 5 times and observe 4 heads and 1 tail.

(a) Determine the posterior distribution of θ given this data.

(b) Compute the posterior mean.

(c) Compute the MAP estimate of θ. (8)

3. (a) Derive posterior mean for Gaussian likelihood with known variance and Gaussian prior. (7)

(b) Normally distributed IQ scores have a mean of 100 and standard deviation of 15. Use the standard z-table to answer the following questions.

(i) What is the probability of randomly selecting someone with an IQ score less than 80?

(ii) What is the probability of randomly selecting someone having IQ between 95 and 100?

(iii) What IQ score corresponds to the 90th percentile?

(iv) The middle 30% of the IQs fall between what two values? (8)

4. (a) What do you mean by point estimation? State the properties that a good point estimate should have. Write two different methods that can be used for the point estimation. (7)

(b) Suppose we have a factory which makes diodes. The diodes have a lifetime which is modelled by an exponential distribution with parameter $\lambda$. Suppose we check 5 diodes and find that their lifetimes are 2,3,1,3 and 4 years. What is the maximum likelihood estimate for $\lambda$.

(Exponential distribution: $f(x|\lambda) = \lambda e^{-\lambda x}$). (8)

5. (a) What do you understand by the term "Bayes Factor"? At the ITO intersection, vehicles pass through at an average rate of 600 per hour.

(i) Find the probability that none passes in a given minute.

(ii) What is the expected number passing in 5 minutes? (7)

(b) Consider a crime scene. The police find that there are two different types of blood on the crime scene left by the criminals- one is type O and other type AB. We know that in the population, type O is found in 60% of the population and type AB in 1% of the population. The police arrest two people R and S. R has a blood group O. Does the data, that is two types O and AB are found on the scene, support the hypothesis that R is one of the criminals? (8)

6. (a) What do we mean by Bayesian linear regression? Discuss (in brief) the significance of using conjugate priors while estimating the posterior density function of the parameter. (7)

(b) In a bayesian linear regression, the prior ($\theta$) follows the normal distribution with mean 0 and standard deviation 1. If the data is as follows: (Assume variance =1)

| x | 2 | 4 | 6 | 8 |
|---|---|---|---|---|
| y | 4 | 16 | 36 | 64 |

Compute the posterior distribution of the $\theta$. (8)