This question paper contains 6 printed pages]

Roll No. | | | | | | | | | | | |

S. No. of Question Paper : **1229**

Unique Paper Code : **2353010009**

Name of the Paper : **Mathematical Statstics**

Type of the Paper : **DSE**

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

Semester : **V**

Duration : **3 Hours** Maximum Marks : **90**

(Write your Roll No. on the top immediately on receipt of this question paper.)

All questions are compulsory.

Attempt any *two* parts from each question.

All questions carry equal marks.

Use of non-programmable scientific calculators and statistical tables is permitted.

Part A

1. (*a*) A large but sparsely populated country has two small hospitals, one at the south end of the country and the other at the north end. The south hospital's emergency room has 4 beds, whereas the north hospital's emergency room has only 3 beds. Let X denote the number of south beds occupied at a particular time on a given day, and let Y denote the number of north beds occupied at the same time on the same day. Suppose that these two rvs are independent, that the pmf of X puts probability masses. .1, .2, .3, *.2 and .2 on the x* values 0, 1, 2, 3, and 4, respectively, and that pmf of Y distributes probabilities .1, .3, .4, and .2 on the *y* values 0, 1, 2, and 3, respectively.

P.T.O.

(i) Display the joint pmf of X and Y in a joint probability table.

(ii) Compute $P(X \leq 1 \text{ and } Y \leq 1)$ and verify that this equals the product of $P(X \leq 1)$ and $P(Y \leq 1)$.

(iii) Express the event that the total number of beds occupied at the two hospitals combined is at most 1 in terms of X and Y, and then calculate this probability.

(iv) What is the probability that at least one of the two hospitals has no beds occupied ?

(b) An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X = the number of points earned on the first part and Y = the number of points earned on the second part. Suppose that the joint pmf of X and Y is given in the accompanying table.

		y			
	***p*(*x*, *y*)**	**0**	**5**	**10**	**15**
	0	.02	.06	.02	.10
x	**5**	.04	.15	.20	.10
	10	.01	.15	.14	.01

(i) Compute the covariance for X and Y.

(ii) Compute ρ for X and Y.

(c) A concert has three pieces of music to be played before intermission. The time taken to play each piece has a normal distribution. Assume that the three times are independent of each other. The mean times are 15, 30, and 20 min, respectively, and the standard deviations are 1, 2, and 1.5 min, respectively. What is the probability that this part of the concert takes at most one hour ? Are there reasons to question the independence assumption ? Explain.

2. (*a*) A system consisting of two components will continue to operate only as long as both components function. Suppose the joint pdf of the lifetimes (months) of the two components in a system is given by $f(x, y) = c\,[10 - (x + y)]$, for $x > 0$, $y > 0$, $x + y < 10$. Find the value of the constant c. If the first component functions for exactly 3 months, what is the probability that the second functions for more than 2 months.

(*b*) Define the bivariate normal distribution. For a few years, the SAT consisted of three components : writing, critical reading, and mathematics. Let W = SAT Writing score and X = SAT Critical Reading score for a randomly selected student. According to the College Board, in 2012 W had mean 488 and standard deviation 114, while X had mean 496 and standard deviation 114. Suppose X and W have a bivariate normal distribution with Corr(X, W) = .5. If an English department plans to use X + W, a student's total score on the nonmath sections of the SAT to help determine admission. Then determine the distribution of X + W.

(*c*) (*i*) Suppose that the lifetimes of two components are independent of each other and that the first lifetime, X_1, has an exponential distribtuion with parameter $\lambda_1 = 1/1000$ whereas the second, X_2, has an exponential distribution with parameter $\lambda_2 = 1/1200$. Compute the probability that the sum of their lifetimes is at most 3000 h.

(*ii*) A surveyor wishes to lay out a square region with each side having length L. However, because of measurement error, he instead lays out a rectangle in which the north-south sides both have length X and the east-west sides both have length Y. Suppose that X and Y are independent and that each one is uniformly distributed on the interval [L – A, L + A] (where 0 < A < L). What is the expected area of the resulting rectangle ?

3. (*a*) A particular brand of dishwasher soap is sold in three sizes : 25, 40, and 65 oz. 20% of all purchasers select a 25-oz box, 50% select a 40-oz box, and the remaining 30% choose a 65-oz box. Let X_1 and X_2 denote the package sizes selected by two independently selected purchasers. Determine the sampling distribution of $\overline{X}$, calculate $E(\overline{X})$, and compare to μ.

(*b*) State the Central Limit Theorem.

When a batch of a certain chemical product is prepared, the amount of a particular impurity in the batch is a random variable with mean value 4.0 g and standard deviation 1.5 g. If 50 batches are indepedently prepared, what is the (approximate) probability that the sample average amount of impurity X is between 3.5 and 3.8 g ? Justify why is the Central Limit Theorem applicable here ?

(*c*) (*i*) Define chi-squared distribution with ν degrees of freedom.

(*ii*) Show that if $X_1, X_2, \ldots\ldots\ldots, X_n$ is a random sample from a $N(\mu, \sigma)$ distribution, then $(n - 1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Further deduce that $E(S^2) = \sigma^2$ and $V(S^2) = 2\sigma^4/(n - 1)$.

4. (*a*) (*i*) Show that the sample variance S^2 is an unbiased estimator of σ^2 for any population distribution.

(*ii*) Define minimum variance unbiased estimator (MVUE).

(*b*) (*i*) The traditional criteria for "strong" passwords and the emerging advice is to use longer passphrases by concatenating several everyday words. Suppose that 10 students at a certain university are randomly selected, and it is found that the three of them use passphrases for their email accounts. Let p = P(passphrase); i.e., p is the proportion of all students at the university using a passphrase on their email accounts. Find the maximum likelihood estimate of the parameter p.

(*ii*) Suppose X has the pdf $f(x; \theta) = \theta x^{\theta-1}$ for $0 \leq x \leq 1$. Obtain the Fisher information $I(\theta)$.

(c) State the Neyman factorization theorem. Use it to show that the sufficient statistic for μ is $T = \Sigma X_i$, by considering a random sample $X_1, X_2, \ldots\ldots, X_n$ from a Poissson distribution with parameter μ, describing the numbers of errors in n batches of tax returns where each batch consists of many returns.

5. (a) Assume that the helium porosity (in percentage) of coal samples taken from any of coal samples taken from any particular seam is normally distributed with true standard deviation 75.

(i) Compute a 95% confidence interval (CI) for the true average porosity of certain seam if the average porosity for 20 specimens from the seam was 4.85.

$[Z_{0.05} = 1.96]$

(ii) Compute a 98% confidence interval (CI) for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.

$[Z_{0.01} = 2.33]$

(b) For healthy individuals the level of prothrombin in the blood is approximately normal distributed with mean 20 mg/dL and standard deviation 4 mg/dL. Low levels indicate low clotting ability. In studying the effect of gallstones on prothrombin, the level of each patient in a sample is measured to see if there is a deficiency. Let μ be the true average level of prothrombin for gallstone partients (and assume $\sigma = 4$).

(i) What are the appropriate null and alternative hypotheses ?

(ii) Let $\overline{X}$ denote the sample average level of prothrombin in a sample of $n = 20$ randomly selected gallstone patients. Consider the test procedure with test statistic $\overline{X}$ and rejection region $\overline{x} \leq 17.92$. What is the probability distribution of the test statistic when H_0 is true ? What is the probability of a type I error for the test procedure ?

(*c*) On the label, Pepperidge Farm bagels are said to weigh four ounces each (113 g). A random sample of six bagels resulted in the following weights (in grams) :

117.6, 109.5, 111.6, 109.2, 119.1, 110.8.

Based on this sample, is there any reason to doubt that the population mean is at least 113 *g* at 0.05 level of significance ? [$t_{0.05,5}$ = 2.015]

6. (*a*) A plan for an executive traveler's club has been developed by an airline on the premise that 5% of its current customers would qualify for membership. A random sample of 500 customers yielded 40 who would qualify. Using this data, test at level 0.01 the null hypothesis that the company's premise is correct against the alternative that it is not correct. [$z_{0.01}$ = 2.576]

(*b*) The accompanying data on x = current density (mA/cm^2) and y = rate of deposition (mm/min) appeared in the article "Plating of 60/40 Tin/Lead Solder for Head Termination Metallurgy". Do you agree with the claim by the article's author that a linear relationship was obtained from the tin-lead rate of deposition as a function of current density ? Explain your reasoning.

X	20	40	60	80
Y	0.24	1.20	1.71	2.22

(*c*) A statistics department at a large university maintains a tutoring center for students in its introductory service courses. The center has been staffed with the expectation that 40% of its clients would be from the business statistics course, 30% from engineering statistics, 20% from the statistics course for social science students, and the other 10% from the course for agriculture students. A random sample of n = 120 clients revealed 52, 38, 21, and 9 from the four courses. Does this data suggest that the percentages on which staffing was based are not correct ? State and test the relevant hypotheses using α = 0.05. [$\chi^2_{0.05,3}$ = 7.815]