C

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MAY 2016

Computer Science & Engineering

(Computer Science & Systems Engineering)

04CS6414 Statistical Foundations of Machine Learning

Max. Marks : 60 Duration: 3 Hours

PART A

Answer All Questions

Each question carries 3 marks

1. Explain total probability and Baye's theorem.
2. State Central Limit Theorem.
3. Given that the counting process $\{N(t), t \geq 0\}$ is a poisson process with the mean λt. Then prove the following:
   (a) $P(N(h) = 1) = \lambda h + O(h)$
   (b) $P(N(h) \geq 2) = O(h)$
4. Given that X_1 , X_2 …. X_n a random sample from N (μ, σ^2) population. Derive an expression for the confidence intervals for mean of the population.
5. What are the different steps for testing a hypothesis? Explain with an example.
6. The specifications for a certain kind of ribbon call for a mean breaking strength of 180 pounds. If five pieces of the ribbon have a mean breaking strength of 169.5 pounds with a standard deviation of 5.7 pounds , test the null hypothesis $\mu=180$ pounds against the alternative hypothesis $\mu<180$ pounds at the 0.01 level of significance. Assume that the population distribution is normal.
7. Playing 10 rounds of golf on his home course, a golf professional averaged 71.3 with a standard deviation of 1.32. Test the null hypothesis that the consistency of his game on his home course is actually measured by $\sigma = 1.20$, against the alternative hypothesis that he is less consistent. Use the level of significance 0.05.
8. To check on an ambulance service's claim that at least 40% of its calls are life-threatening emergencies, a random sample was taken from its file, and it was found that only 49 out of 150 calls were life-threatening emergencies. Can the null hypothesis $p \geq 0.40$ be rejected against the alternative hypothesis $p < 0.40$ at 0.01 level of significance.

PART B

Each question carries 6 marks

9. Given that X is a continuous random variable with probability density function (p.d.f) $f_X(x)$. Also given a function of random variable X, as Y, then find the p.d.f for the following values of Y.
   (a) $Y=|X|$ (2)
   (b) $Y=X^n$ (2)
   (c) $Y=g(X)$ (2)

OR

10. For every real number 'x', the cumulative distribution function (c.d.f) of a random variable 'X' is given by $F_X(x) = P(X \leq x)$. Any such c.d.f must satisfy some predefined conditions. Explain the necessary conditions with proof. (6)
11. Given two random variables X and Y and a function g(X,Y) is assigned to a new random variable Z. Given the joint p.d.f (probability density function) of X and Y as $f_{XY}(x,y)$. Determine the p.d.f of Z

($f_Z(z)$)

(a) Z=X-Y (3)

(b) Z=X/Y (3)

OR

12. Explain the following,

(a) Convergence of random variables (3)

(b) Independent or uncorrelated random variables (3)

13. An automobile manufacturing company has a policy of assigning its office employees to the three sections it has. The three sections are production, human resource and sales. There is no set pattern for reassignments. One does not know in which section he/she will be assigned next. The next assignment may depend on the current assignment. Based on the above scenario, explain Markov random process. Also discuss about the transition matrix of the above random process. (6)

OR

14. Explain the following,

(a) Martingales and types of Martingales (2)

(b) Properties of nth order joint pdf (2)

(c) Gaussian random process (2)

15. What you mean by Sampling distribution of mean? If the population variance ‘σ’ is a known parameter, then how we can prove the characteristics of the sampling distribution of mean? (6)

OR

16. Explain the term “Estimation in statistics”. What are the different methods for estimation? Also explain different criterion for estimation? (6)

17. (a) Consider two populations having the means μ_1 and μ_2 and they defined the null hypothesis $\mu_1 - \mu_2 = \delta$, where δ is a specified constant, on the basis of independent random samples of size n1 and n2. Which statistic you will choose for test the concerning difference between two means? Also define the critical regions for testing $\mu_1 - \mu_2 = \delta$. (3)

(b) Analysis of drinking water samples for 100 homes in one part of a city gave the lead levels mean as 34.1 ppm with a S.D of 5.9 ppm, whereas from the other part of city, a sample of 100 homes shows the mean level of lead as 36.0 ppm with a S.D of 6.0 ppm.

(i) Calculate the test statistic and its observed significance level.

(ii) Use 5% level of significance, write your conclusion. (3)

OR

18. The following are the number of sales which a sample of nine salespeople of industrial chemicals in California and a sample of six salespeople of industrial chemicals in Oregon made over a certain fixed period of time.

California:	59	68	44	71	63	46	69	54	48
Oregon:	50	36	62	52	70	41			

Assume that the populations sampled can be approximated closely with normal distributions having the variance, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 \neq 0$ at the 0.01 level of significance. (6)

19. (a)Explain about the estimation of proportions. (2)

(b)In a referendum submitted to students body at university, 850 men and 566 women voted. 530 out of the men and 304 of the women voted “yes”. Does this indicate a significant difference of opinion on the matter at 1% level, between men and women students? (4)

OR

20. Samples of three kinds of materials, subjected to extreme temperature changes, produced the results shown in the following table:

	Material A	Material B	Material C	Total
Crumbled	41	27	22	**90**
Remained intact	79	53	78	**210**
Total	**120**	**80**	**100**	**300**

Use the 0.05 level of significance to test whether, under the stated conditions, the probability of crumbling is the same for the three kinds of materials. (6)