C 3391 (Pages : 10) Name.......................................

Reg. No...................................

Ph.D. ENTRANCE EXAMINATION, APRIL 2021

STATISTICS

Time : Two Hours Maximum : 100 Marks

Part A

All questions *are* Compulsory.
Each question carries 2 marks.

1. The interval $(a, b]$ is a Borel set because (a, b) is a Borel set for all $a, b \in \mathbb{R}$ and :

   (a) $(a,b] = \bigcup_{n=1}^{\infty}\left(a, b-\frac{1}{n}\right)$.

   (b) $(a,b] = \bigcap_{n=1}^{\infty}\left(a, b+\frac{1}{n}\right)$.

   (c) $(a,b] = \bigcup_{n=1}^{\infty}\left(a, b+\frac{1}{n}\right)$.

   (d) $(a,b] = \bigcap_{n=1}^{\infty}\left(a, b-\frac{1}{n}\right)$.

2. Suppose that $\Omega = \{1,2,3,4\}$ and $\mathcal{F} = \mathbb{P}(\Omega)$. For $w \in \Omega$, define δ_w on $\mathcal{F}$ by :

   $$\delta_w(A) = \begin{cases} 1, \text{ if } w \in A; \\ 0, \text{ if } w \notin A. \end{cases}$$

   Let $P = 0.3\delta_1 + 0.7\delta_2$; and $Q = \delta_1 \times \delta_2$. Then,

   (a) Only P is a probability measure.

   (b) Only Q is a probability measure.

   (c) Both P and Q are probability measures.

   (d) Neither P nor Q is a probability measure.

3. Let V denote a vector space of real valued continuous functions on $[-1, 1]$. Let W denote the largest subset of V such that for every function $h \in W$ and $f, g \in V$, $\int_{-1}^{1} f(x)\, g(x)\, h(x)\, dx$ defines the inner product of f and g. Consider the following statements :

   (1) W is a subspace of V.

   (2) $h(x) = e^{-x-1} \in W$.

   (3) $h(x) = x - 1 \in W$.

   The correct statement(s) is (are) :

   (a) 1, 2, 3.

   (b) Only 1 and 2.

   (c) Only 1.

   (d) Only 2.

Turn over

4. Let (χ, d) be a metric space where χ is an infinite set and the metric d is given as :

$$d(x,y)=\begin{cases}0, \text{ if } x=y\\ 1, \text{ if } x\neq y\end{cases}$$

Consider $E \subset X$ given as $E=\left[\frac{1}{n}, n\in\mathbb{N}\right]$, where $\mathbb{N}$ is a set of natural numbers. Identify the INCORRECT statement.

(a) E is a closed subset of χ.

(b) E is a bounded subset of χ.

(c) E is a compact subset of χ.

(d) E is an open subset of χ.

5. Consider a function $f : R \to R$ defined as $f(x)=\frac{1}{1+2x^2}$. Which of the following statements are correct ?

(i) f is uniformly continuous on $\mathbb{R}$.

(ii) f is continuous on $\mathbb{R}$ but is uniformly continuous only on the compact subsets of $\mathbb{R}$.

(iii) f is an unbounded function.

(a) Only (i).

(b) Only (ii).

(c) Only (i) and (iii).

(d) Only (ii) and (iii).

6. Suppose T is a linear transformation on V such that $T^3 - T^2 - T + I = 0$ where I is an identity matrix of appropriate order. Then T^{-1} is :

(a) $I - T - T^2$.

(b) $I + T - T^2$.

(c) $I + T + T^2$.

(d) $I - T + T^2$.

7. Suppose in the interval [0,1] the function $x^4 - 62x^2 + ax + 9$ attains its maximum value at $x = 1$. Then the value of a is :

(a) 120.

(b) 140.

(c) 210.

(d) 180.

8. Which of the following statements are true in association with exchangeability ?

(i) $Y_1, Y_2, \ldots, Y_n$ are exchangeable if $p(y_1, y_2, \ldots, y_n) = p(y_{\pi_1}, y_{\pi_2}, \ldots, y_{\pi_n})$ for all permutations π of $\{1, 2, \ldots, n\}$.

(ii) Independent and identically distributed sequence is exchangeable.

(iii) If $Y_1, Y_2, \ldots, Y_n$ are conditionally i.i.d., given θ and if $\theta \sim P(\theta)$, then $Y_1, Y_2, \ldots, Y_n$ are exchangeable.

(iv) If $Y_1, Y_2, \ldots, Y_n$ are exchangeable for all n, then, $Y_1, Y_2, \ldots, Y_n$ given θ need not be i.i.d.

(a) (i), (ii) and (iii). (b) (ii), (iii) and (iv).

(c) (i), (ii) and (iv). (d) (i), (iii) and (iv).

9. Let $P(Y = y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$, $y \in \{0, 1, 2, \ldots, n\}$. If θ has a *Uniform* (0,1) prior, then the posterior distribution $P(\theta \mid Y = y)$ will be :

(a) *Uniform* (0,1). (b) *Beta* (y + 1, $n - y$ + 1).

(c) *Beta* (y, $n - y$). (d) *Uniform* (0, y).

10. Let $Y_1, Y_2, \ldots, Y_n$ be n independent observations from $P(Y = y \mid \theta) = \theta^y (1-\theta)^{n-y}$, $0 < \theta < 1$, $y \in \{0,1\}$. Let $Y = \sum_{i=1}^{n} Y_i$? If θ has a Beta (1, 1) prior, then which of the following statements are true about the predictive distribution of $\tilde{Y}$, where $\tilde{Y} \in \{0, 1\}$ } is an additional outcome from the same distribution yet to be observed ?

(i) $P(\tilde{Y} = 1 \mid Y = 1) = 2/(n+2)$. (ii) $P(\tilde{Y} = 1 \mid Y = 0) = 1/(n+2)$.

(iii) $P(\tilde{Y} = 0 \mid Y = 1) = 1 - 1/n$. (iv) $P(\tilde{Y} = 0 \mid Y = 0) = (n+1)/(n+2)$.

(a) (i), (ii) and (iii). (b) (ii), (iii) and (iv).

(c) (i), (iii) and (iv). (d) (i), (ii) and (iv).

11. Let Y be Poisson (λ) and a single realization resulted in Y = 3. If λ has a prior distribution given by $p(\lambda) = 9\lambda e^{-3\lambda}$, $\lambda > 0$, then the mean of the posterior distribution is :

(a) 5/4. (b) 4/5.

(c) 2. (d) 3/5.

12. For estimating the population proportion P in a class of a population having N units, the variance of the estimator p of P based on sample of size n is :

(a) $\frac{N}{N-1}\frac{PQ}{n}$.

(b) $\frac{N}{N-1}\frac{PQ}{N}$.

(c) $\frac{N-n}{N-1}\frac{PQ}{n}$.

(d) $\frac{N-1}{N-n}\frac{PQ}{n}$.

13. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta,1)$. Let $\hat{\theta}_1 = \overline{X}$ and $\hat{\theta}_2 = w\overline{X} + (1-w)\mu_0$, where $\hat{\theta}_2$ is the Baye's estimator obtained from posterior mean by considering $N(\mu_0, \tau_0^2)$ as the prior for θ and $\omega = (1/\tau_0^2)\left((1/\tau_0^2) + (n/\sigma^2)\right)^{-1}$. Then, which of the following are not true ?

(i) $E\theta_0\left(\hat{\theta}_1\right) = E\theta_0\left(\hat{\theta}_2\right) = \theta$.

(ii) $V\theta_0\left(\hat{\theta}_1\right) = V\theta_0\left(\hat{\theta}_2\right) = 1/n$.

(iii) $V\theta_0\left(\hat{\theta}_2\right) = w^2/n < 1/n$.

(iv) $MSE_{\theta_0}\left(\hat{\theta}_1\right) < MSE_{\theta_0}\left(\hat{\theta}_2\right)$ always.

(a) (i), (ii) and (iii).

(b) (ii), (iii) and (iv).

(c) (i), (iii) and (iv).

(d) (i), (ii) and (iv).

14. Which of the following statements are true ?

(i) If T is a boundedly complete sufficient statistic for θ, and S is an ancillary statistic for θ, then, T and S are independent.

(ii) An ancillary statistic may or may not be independent of a sufficient statistic.

(iii) It is possible to have variance smaller than that of the Cramer-Rao lower bound at some points of the parameter space.

(iv) The parametric function need not be differentiable for the Cramer-Rao lower bound to hold.

(a) (i), (ii) and (iii) only.

(b) (ii), (iii) and (iv) only.

(c) (i), (iii) and (iv) only.

(d) (i), (ii) and (iv) only.

15. Suppose you are given random variables X and Y such that

$$X \sim N\left(\mu_x, \sigma_x^2\right),\ Y \mid X = x \sim N\left(\beta_0 + \beta_1 x, \sigma^2\right).$$

Suppose the joint distribution of (X, Y) is bivariate normal. Which of the following statements are true ?

(i) $Var(Y) = \sigma^2\left(1 - \sigma_x^2\right)$. (ii) $Cov(X, Y) = \beta_1\sigma_x^2$.

(iii) $E(Y) = \beta_0 + \beta_1\mu_x$. (iv) $Var(Y) = \sigma^2 + \beta_1^2\sigma_x^2$.

(a) (ii), (iii) and (iv) only. (b) (i), (ii) and (iii) only.

(c) (i), (iii) and (iv) only. (d) (i), (ii) and (iv) only.

16. Suppose that we want to test the equality of mean vectors of five independent trivariate normal populations having the same variance-covariance matrix using Wilk's lambda (Λ_w) on the basis of a sample of size 10 coming from each of them. Which of the following provides the correct distribution of the test statistic under H_0 ?

(a) $-5\ln(\Lambda_w) \sim \chi^2_{12}$. (b) $-45\ln(\Lambda_w) \sim \chi^2_{12}$.

(c) $-4.5\ln(\Lambda_w) \sim \chi^2_{15}$. (d) $-44.5\ln(\Lambda_w) \sim \chi^2_{15}$.

17. Let X be a random variable with probability mass function f_0 under H_0 and f_1 under H_1 defined as follows :

X	1	2	3	4	5	6
$f_0(x)$	.01	.01	.01	.01	.01	.95
$f_1(x)$	.05	.04	.03	.02	.01	.85

Then, which of the following test is a most powerful test of size 0.03 ?

(a) $\phi(x) = \begin{cases} 1, \text{ if } f_1(x)/f_0(x) \geq 2 \\ 0, \quad \text{otherwise.} \end{cases}$ (b) $\phi(x) = \begin{cases} 1, \text{ if } f_1(x)/f_0(x) \leq 4 \\ 0, \quad \text{otherwise.} \end{cases}$

(c) $\phi(x) = \begin{cases} 1, \text{ if } f_1(x)/f_0(x) \geq 3 \\ 0, \quad \text{otherwise.} \end{cases}$ (d) $\phi(x) = \begin{cases} 1, \text{ if } f_1(x)/f_0(x) \leq 2 \\ 0, \quad \text{otherwise.} \end{cases}$

18. In a standard multiple linear regression model $y = X\beta + \epsilon$, which of the following are not true ? (H is the hat matrix defined by $H = X(X'X)^{-1}X'$).

(i) $\text{Var}(\hat{y}) = \sigma^2 H$.

(ii) The least squares estimator $\hat{\beta} = \beta + R\epsilon$ where $R = (X'X)^{-1}X'$.

(iii) $SS_R = y'(I - H)\, y$.

(iv) R^2 is the square of the correlation coefficient between y and $\hat{y}$.

(v) Residual $e = H \in$.

(a) (iii) and (v) only. (b) (ii) and (iv) only.

(c) (i) and (iii) only. (d) (i) and (v) only.

19. Let X be a non-negative non-degenerate random variable (i.e., not identically equal to a constant). Assume that E (X) and $E(X^{-1})$ exists. Then, which of the following statements is true always ?

(a) $E(X) = E(X^{-1})$. (b) $[E(X)]^{-1} < E(X^{-1})$.

(c) $[E(X)]^{-1} > E(X^{-1})$. (d) $[E(X)]^{-1} = E(X^{-1})$.

20. Which of the following statements are correct ?

(i) Cook's distance D_i always lies between 0 and 1.

(ii) In polynomial regression, the matrix X'X becomes ill-conditioned, if there are small number of knots.

(iii) In a linear regression $y = X\beta + \in$ with usual assumptions having $p = 4$ and $n = 20$, if the third diagonal element of the hat matrix is 0.845, then it can be considered as a leverage point.

(iv) When the regressors are orthogonal, the multicollinearity problem doesn't exist.

(a) (i), (ii) and (iii) only. (b) (ii), (iii) and (iv) only.

(c) (i), (iii) and (iv) only. (d) (i), (ii) and (iv) only.

21. Let Y_1 and Y_2 be independent random variables with means β and 2β respectively with a finite variance σ^2. If the realized values of Y_1 and Y_2 are 1 and 2 respectively, then the least squares estimate of β and the residual sum of squares are respectively.

(a) 1, 0. (b) 1, 1.

(c) 0, 1. (d) 1, 2.

22. Let $Y_t = Z_t - \theta Z_{t-1}^2$, where Z_t's are iidNormal $\left(0, \sigma_z^2\right)$. Then which of the following statements are true ?

(i) $Cov(Y_t, Y_{t-k}) = 0$ for all $k \neq 0$.

(ii) $\{Y_t\}$ is a stationary series.

(iii) $\{Y_t\}$ is a non-Gaussian white noise.

(iv) $Var(Y_t) = = 2\theta^2\sigma_z^4$.

(a) (i), (ii) and (iii). (b) (ii), (iii) and (iv).

(c) (i), (iii) and (iv). (d) (i), (ii) and (iv).

23. Suppose $\{X_t\}$ is a stationary process with $E(X_t) = 0$ and $Var(X_t) = \sigma_x^2$. Then, which of the following statements is not true with respect to the process $\{Y_t\}$, where $Y_t = (-1)^t X_t$.

(i) $E(Y_t) = 0$.

(ii) $Var(Y_t) = Var(X_t)$.

(iii) $\{Y_t\}$ is a stationary process.

(iv) $\{Y_t\}$ is not a stationary process.

(a) Only (ii). (b) Only (iii).

(c) (i) and (ii). (d) Only (iv).

24. Let A be a matrix given as follows. What are the rank and nullity of A ?

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 4 & 1 & 2 \\ -1 & 2 & 0 & 4 \end{bmatrix}$$

(a) Rank = 2 and nullity = 2. (b) Rank = 3 and nullity = 2.

(c) Rank = 1 and nullity = 3. (d) Rank = 3 and nullity = 1.

25. Let X_1 and X_2 be i.i.d. exponential random variables. Which of the following statements is false ?

(a) $X_1 + X_2$ has a gamma distribution.

(b) [max (X_1, X_2) – *min* (X_1, X_2)] has exponential distribution.

(c) min (X_1, X_2) has exponential distribution.

(d) max (X_1, X_2) has exponential distribution.

(25 × 2 = 50 marks)

Part B

Answer any **five** *questions.*
Each question carries 10 *marks.*

26. (a) Let $f : \mathcal{R} \to \mathcal{R}$ be a function such that $f(0) = 0$ and $|f'(x)| \le 5$ for all x. Find the set of all possible values of $f(1)$.

(b) Let $\{a_n, n \ge 1\}$ be a sequence of real numbers with $a_1 = 1$ and for $n \ge 1$.

$$a_{n+1} = \frac{(-1)^n}{2}\left(|a_n| + \frac{2}{|a_n|}\right).$$

Find lim sup a_n and lim in f a_n.

27. (a) Let $X_1, X_2, \ldots, X_n$ be a random sample from a double-exponential (Laplace) distribution having density f with median θ. Let M_n be the sample median. Derive the asymptotic distribution of M_n (You may assume a relevant theorem in this context).

(b) Let the random variable X is having a probability density function

$f(x;\theta) = \theta(\theta + x)^{-2}, x > 0, \theta > 0$. Derive the UMP test of size α, for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

28. (a) $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i. i. d.) random variables with $E(X_1) = \mu \in R$ and $Var\,(X_1) = \sigma^2 \in (0, \infty)$ Then, by central limit theorem (CLT),

$$\sqrt{n}\,\frac{(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Z,$$

where $Z \sim N(0,1)$. Show that CLT implies WLLN for $\bar{X}_n$.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with distribution function $F(x)$. Define T_n as $T_n = (1/n)\sum_{i=1}^{n} I_{[X_i \leq x]}$, where $I_{[.]}$ is an indicator function. Show that T_n converges in probability to $F(x)$ for all fixed $x \in R$. Also examine whether T_n satisfy the central limit theorem.

29. Let $\{X_n\}$ be a sequencen of random variables defined as :

$$X_n = \begin{cases} n^c, \text{ with probability } 1/n, \\ 0, \quad \text{with probability } 1 - 2/n, \\ n^{-c}, \text{ with probability } 1/n. \end{cases}$$

where c is a positive constant.

(a) Show that $X_n \xrightarrow{p} 0$.

(b) Examine the convergence of $E|X_n|^r$.

30. (a) Suppose that $X_1, X_2, \ldots, X_{50}$ are independent and identically distributed six-variate normal vectors with mean vector μ. We want to test $H_0 : \mu_1 = \mu_2 = \mu_3$ and $\mu_4 = \mu_5 = \mu_6$, where μ_i is the i^{th} component of μ. Obtain an appropriate Hotelling's T^2 test statistic and discuss the testing procedure.

(b) State the null hypothesis, the alternative hypothesis, the analysis of variance table, the test statistic and the null distribution of the test statistic for multivariate analysis of variance. Explain all the notations involved in the above expressions.

31. (a) Derive the likelihood equations and the solutions of the likelihood equations when you have a random sample $(X_1, X_2, \ldots, X_n)$ of size n from a multivariate normal distribution. You need not show that these solutions actually maximize the likelihood.

(b) Derive the distribution of the sample mean computed from a sample of size n from a p-variate normal distribution.

32. (a) What is a variance stabilizing transformation ? Illustrate with an example.

(b) Consider a randomized block design involving 3 treatments and 3 replicates and let t_i denote the effect of the i – th treatment (i = 1, 2, 3). Derive the variance of the BLUE of $(t_1 - 2t_2 + t_2)/\sqrt{6}$.

33. (a) Discuss when the cluster sampling is more efficient than simple random sampling. Justify your claims.

(b) Suppose we draw a random sample of size n from a population of size N, $1 < n < N$ using a SRSWOR scheme. Let P be the population proportion of units possessing a particular attribute and p be the corresponding sample proportion. Obtain an unbiased estimator of P (1 – P). Justify your claim.

34. (a) Consider the matrix :

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \text{ where } \theta = 2\pi / 31.$$

Find A^{2015}.

(b) Examine the definiteness of the matrix A where

$$A = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}.$$

35. (a) Consider the Markov Chain with state space $S = \{1, 2, ..., n\}$ where $n > 10$. Suppose that the transition probability matrix satisfies $p_{ij} > 0$, if $|i - j|$ is even, and $P_{ij} = 0$, if $|i - j|$ is odd. Then, examine whether the Markov chain is irreducible. Is the chain aperiodic? Are there infinitely many stationary distributions ? Justify your claims.

(b) Consider an M/M/1 queue with arrival rate λ and service rate μ with $\mu > \lambda$, What is the probability that no customer exited the system before time 5 ?

(5 × 10 = 50 marks)