2022

STATISTICS — HONOURS

Paper : CC-9

(Statistical Inference-I and Sampling Distribution)

Full Marks : 50

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer ***any five*** questions from ***question nos.*** **1-8**. 2×5

1. Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 1)$ choose d so that test based on $T_n = (\#\text{ of } |X_i| \leq d)/n$ for testing $H_0 : \theta = 0$ against $H_1 : \theta = 10$ has both types of errors almost zero.

2. Let $X_1, \dots, X_n$ be a random sample from Bernoulli (p) population, $P \in (0,1)$. Define p-value of the test based on $T_n = n - \{\#\text{ of } X_i = 0\}$, for testing $H_0 : p = P(X_1 = 0) = \frac{1}{2}$ against $H_1 : p > \frac{1}{2}$.

3. Let $X_1, \dots X_{n_1}, X_{n_1+1}, \dots, X_n$ be IID $N(\theta, \sigma^2), \theta \in \mathbb{R}, \sigma > 0$. Consider statistics

$$T_{1n} = \frac{\overline{X}_n - \theta}{\left[\frac{1}{n-1}\sum_1^n \left(X_i - \overline{X}_n\right)^2\right]^{\frac{1}{2}}} \text{ and } T_{2n} = \frac{\frac{1}{n-n_1}\sum_{n_1+1}^n X_i - \theta}{\left[\frac{1}{n_1 - 1}\sum_1^{n_1}\left(X_i - \overline{X}_{n_1}\right)^2\right]^{\frac{1}{2}}}, \overline{X}_{n_1} = \frac{1}{n_1}\sum_1^{n_1} X_i, \overline{X}_n = \frac{1}{n}\sum_1^n X_i$$

Are sampling distribution of T_{1n} and T_{2n} same? Justify with derivations.

4. Based on a random sample $\{X_1, \dots, X_{10}\}$ from $N(0, 1)$ construct $T(X_1, \dots, X_{10})$ such that T follows a χ^2 distribution with 3 d.f. Use all the observations for constructing T.

5. Let $X_1, \dots, X_n$ be $N(0, 1)$, show that $E(X_{(1)} + X_{(n)}) = 0$.

6. Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 1)$, $\theta \in \mathbb{R}$. Consider the confidence interval $S = \left(\overline{X}_n - d, \overline{X}_n + d\right)$ for θ where $d > \frac{Z_{\alpha/2}}{\sqrt{n}}$, $Z_{\alpha/2}$ is the upper level $\alpha/2$ point of $N(0, 1)$ distribution and $\alpha \in (0, 1)$. If observed interval $\left(\overline{x}_n - d, \overline{x}_n + d\right)$ contains θ, is it possible to reject alternative at level α for testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ based on critical region S? Justify with necessary derivations.

7. Let $\{X_1, ..., X_{20}\}$ be IID $F_{2,3}$ distribution. Find the value of $E\left(\sum_1^5 X_i^{-1} \Big/ \sum_1^{20} X_i^{-1}\right)$.

8. If $X \sim t_n$ (t-distribution with n d.f.), what is the distribution of $\left(1+\frac{X^2}{n}\right)^{-1}$? Derive in details.

Answer ***any two*** questions from question nos. **9-11**. 5×2

9. Let $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_r}$ be two independent random samples from Poisson (λ_1) and Poisson (λ_2) respectively. Perform a test based on P-value for testing $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 > \lambda_2$.

10. Let $X_1, ..., X_n$ be independent and identically distributed continuous random variables with distribution function $F(x)$. Then show that

$$\lim_{n\to\infty} P_F(n(F(X_{(n)}) - F(X_{(n-1)})) \le t) = e^{-1}\ \forall\, t > 0,$$ when $X_{(i)}$ is the i-th order statistics.

11. Let X and Y be independent variables having common distribution Exponential (λ), $\lambda > 0$. Find
   (a) conditional distribution of w given u and hence
   (b) marginal distribution of w, where $u = X + Y$ and $w = X - Y$.

Answer ***any three*** questions from question nos. **12-16**.

12. (a) If random variables X_1 and X_2 are independent and each follows X^2-distribution with n d.f., show that $T = \frac{\sqrt{n}\,(X_1 - X_2)}{2\sqrt{X_1 X_2}}$ follows a student's t-distribution with $n-1$ d.f. and distribution of T is independent of $X_1 + X_2$.

   (b) Let $X_1, ..., X_{n_1}$ and $Y_1, Y_2, ..., Y_{n_2}$ be independent random samples. from $N(\mu_1, \sigma_1^2$ and $N(\mu_2, \sigma_2^2)$ respectively and all the parameters are unknown. Find a confidence interval of $\frac{\sigma_1^2}{\sigma_2^2}$ with confidence coefficient $(1-\alpha)$. $\alpha \in (0, 1)$.

6+4

13. (a) Derive the m.g.f. of χ^2 distribution with n d.f. Hence or otherwise show that $\mu_{r+1} = 2r(\mu_r + \mu_{r-1}), r \ge 1$, where $\mu_k = E(\chi_n^2)^k, k \ge 1$.

(b) Let (Z_1, Z_2) have bivariate normal distribution with $E(Z_1) = E(Z_2) = 0$, $V(Z_1) = V(Z_2) = 1$ and correlation $(Z_1, Z_2) = \rho$. Suppose Z_1 and Z_2 are unobservable and the observable random variables are

$$X_i = \begin{cases} 0, & \text{if } Z_i \le 0 \\ 1, & \text{if } Z_i > 0 \end{cases} \quad i = 1, Z$$

Let τ be the correlation coefficient between X_1 and X_2. Prove that $\rho = \sin(\pi\tau/2)$. 5+5

14. (a) For a bivariate sample $\{(Y_i, X_i); i = 1,...,n\}$, consider regression model $Y_i = \beta X_i + \varepsilon_i$, $i = 1,...,n$, where ε_i is independent of x_i and ε_i's are independent and identically distributed as $N(0, \sigma^2)$. Derive a test for $H_0 : \beta = 0$ against $H_1 : \beta \ne 0$.

(b) If $X_1, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$, find the sampling distribution of

$$\sqrt{\frac{n}{n-1}}\left(\overline{X}_n - X_n\right) \Big/ \sqrt{\left\{(n-1)S_n^2 - \frac{n}{n-1}(X_n - \overline{X}_n)^2\right\} \Big/ (n-2)},$$

where $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$ and $S_n^2 = \frac{1}{n-1}\sum_1^n \left(X_i - \overline{X}_n\right)^2$. 6+4

15. (a) Let the random variables X and Y be distributed as $\chi^2_{n_1}$ and F_{n_1,n_2} respectively. For any $\alpha \in (0, 1)$, χ^2_{α, n_1} and F_{α,n_1,n_2} be defined by $P(X \ge \chi^2_{\alpha,n_1}) = p\left(Y \ge F_{\alpha, n_1, n_2}\right) = \alpha$. Then show that for large n_2, $\chi^2_{\alpha,n_1} \approx n_1 F_{\alpha,n_1,n_2}$.

(b) For a bivariate sample $\{(X_{1i}, X_{2i}), i = 1, ..., n\}$ from bivariate normal distribution with unknown parameters $\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$, derive a test for $H_0 : {\sigma_1}^2 = {\sigma_2}^2$ against $H_1 : {\sigma_1}^2 \ne {\sigma_2}^2$. 5+5

16. (a) Let $X_1, ..., X_n$ be i.i.d. random variables with continuous d.f. F and let $X_{(i)} < ... < X_{(n)}$ be the order statistics. If M_0 be the unique population median, then show that

$$P_F\left(X_{(r)} \le M_0 \le X_{(s)}\right) = \left[\sum_{k=r}^{s-1} \binom{n}{k}\right]\left(\frac{1}{2}\right)^n, r < s.$$

Hence or otherwise find a confidence interval of M_0 with coverage probability at least $1 - \alpha$ for some $\alpha \in (0, 1)$.

(b) Find the mean and variance of Student's t-distribution. Show that its density tends to $N(0, 1)$ as degrees of freedom becomes large. 6+4

———————