Test Code MS (Short answer type) 2004

Syllabus for Mathematics

Permutations and Combinations. Binomials and multinomial theorem. Theory of equations. Inequalities.

Determinants, matrices, solution of linear equations and vector spaces.

Trigonometry, Coordinate geometry of two and three dimensions.

Geometry of complex numbers and De Moivre's theorem. Elements of set theory.

Convergence of sequences and series. Power series. Functions, limits and continuity of functions of one or more variables.

Differentiation, Leibnitz formula, maxima and minima, Taylor's theorem. Differentiation of functions of several variables. Applications of differential Calculus.

Indefinite integral, Fundamental theorem of Calculus, Riemann integration and properties. Improper integrals. Differentiation under the integral sign. Double and multiple integrals and Applications.

Syllabus for Statistics

Probability and Sampling Distributions

Notions of sample space and probability, combinatorial probability, conditional probability and independence, random variable and expectations, moments, standard discrete and continuous distributions, sampling distributions of statistics based on normal samples, central limit theorem, approximation of Binomial to Normal or Poisson law. Bivariate normal and multivariate normal distributions.

Descriptive Statistics

Descriptive statistical measures, graduation of frequency curves, product-moment, partial and multiple correlation, Regression (bivariate and multivariate).

Inference

Elementary theory and methods of estimation (unbiasedness, minimum variance, sufficiency, maximum likelihood method, method of moments). Tests of hypotheses (basic concepts and simple applications of Neyman-Pearson Lemma). Confidence intervals. Tests of regression. Elements of non-parametric inference.

Design of Experiments and Sample Surveys

Basic designs (CRD/RBD/LSD) and their analyses. Elements of factorial designs. Conventional sampling techniques (SRSWR/SRSWOR) including stratification; ratio and regression methods of estimation.

Sample Questions

Group A

1. Suppose two teams play a series of games, each producing a winner and a loser, until one team has won two more games than the other. Let G be the total number of games played. Assume each team has a chance of 0.5 to win each game, independent of the results of the previous games.

   (a) Find the probability distribution of G.

   (b) Find the expected value of G.

2. Is the following system of equations always consistent for real k? Justify your answer.

$$\begin{array}{rcrcrcl} x & + & y & + & kz & = & 2, \\ 3x & + & 4y & + & 2z & = & k, \\ 2x & + & 3y & - & z & = & 1. \end{array}$$

Find the value of k for which this system admits more than one solution? Express the general solution for the system of equations for this value of k.

Group B

3. Suppose a random vector (X, Y) has joint probability density function

$$f(x, y) = 3y$$

on the triangle bounded by the lines $y = 0$, $y = 1 - x$, and $y = 1 + x$. Compute $E(Y|X \leq \frac{1}{2})$.

4. Two policemen are sent to watch a road that is 1 km long. Each of the two policemen is assigned a position on the road which is chosen according to a uniform distribution along the length of the road and independent of the other's position. Find the probability that the policemen will be less than 1/4 kilometer apart when they reach their assigned posts.

5. Here is a partial key-block of a 2^4 factorial experiment (with factors A, B, C, D) conducted in two blocks of size 8 each:

Partial key-block: $ad \quad bd \quad c \quad \cdots$

Search out the other five treatment combinations for the key-block and also the confounded interaction. Also, give the treatment combination of the second block.

6. Let Y_1, Y_2, Y_3 and Y_4 be four uncorrelated random variables with

$$E(Y_i) = i\theta, \quad \text{Var}\,(Y_i) = i^2\sigma^2, \quad i = 1, 2, 3, 4,$$

where θ and σ (> 0) are unknown parameters. Find the values of c_1, c_2, c_3 and c_4 for which $\sum_{i=1}^{4} c_i Y_i$ is unbiased for θ and has least variance.

7. Suppose X has a normal distribution with mean 0 and variance 25. Let Y be an independent random variable taking values -1 and 1 with equal probability. Define $S = XY + \frac{X}{Y}$ and $T = XY - \frac{X}{Y}$.

   (a) Find the probability distribution of S.

   (b) Find the probability distribution of $(\frac{S+T}{10})^2$.

8. Let $X_1, X_2, \ldots X_n$ be i. i. d. with common density $f(x;\theta)$ given by

$$f(x;\theta) = \frac{1}{2\theta}\exp(-|x|/\theta), \quad -\infty < x < \infty, \quad \theta \in (0, \infty).$$

In case of each of the statistics S and T defined below, decide (a) if it is an unbiased estimator of θ, (b) if it is an MLE for θ and (c) if it is sufficient for θ. Give reasons.

$$S = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad T = \frac{1}{n}\sum_{i=1}^{n} |X_i|.$$

9. Let Y_1, Y_2, Y_3 and Y_4 be a random sample from a population with

9. Let Y_1, Y_2, Y_3 and Y_4 be a r.s from a popl^n with

probability density function

$$f(y,\theta) = \begin{cases} \left(\frac{1}{2\theta^3}\right) y^2 \exp(-y/\theta) & \text{if } y > 0 \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Find the most powerful test for testing $H_0 : \theta = \theta_0$ against the hypothesis $H_1 : \theta = \theta_1$, where $\theta_1 > \theta_0$. Is the test uniformly most powerful for $\theta > \theta_0$?

10. On a particular day let X_1, X_2 and X_3 be the number of boys born before the first girl is born in hospitals 1, 2 and 3 respectively. If the observations are $X_1 = 0$, $X_2 = 3$ and $X_3 = 2$, find the most powerful test to test the null hypothesis that a girl and a boy are equally likely to be born against the alternative that a girl is less likely to be born than a boy.

11. For $n \geq 1$ let $x_n = \dfrac{1}{n\alpha_n}$ where α_n is such that $2^{\alpha_n - 2} < n \leq 2^{\alpha_n - 1}$. Is the series $\sum_{n=1}^{\infty} x_n$ convergent or divergent? Justify your answer.

12. Let f and g be continuous functions such that $f(x) \leq g(x)$ for all $x \in [0,1]$. Determine which of the following statements are true and which are false:

   (a) $\int_0^x |f(t)|dt \leq \int_0^x |g(t)|dt$ for all $x \in [0,1]$,

   (b) $\int_0^x (|f(t)| + f(t))dt \leq \int_0^x (|g(t)| + g(t))dt$ for all $x \in [0,1]$,

   (c) $\int_0^x (|f(t)| - f(t))dt \leq \int_0^x (|g(t)| - g(t))dt$ for all $x \in [0,1]$.

   For any statement which you believe to be true, you need to give a proof and for any statement which you believe to be false, you need to give a counter example.

Test Code MS (Short answer type) 2005
Syllabus for Mathematics

Permutations and Combinations. Binomials and multinomial theorem. Theory of equations. Inequalities.

Determinants, matrices, solution of linear equations and vector spaces.

Trigonometry, Coordinate geometry of two and three dimensions.

Geometry of complex numbers and De Moivre's theorem. Elements of set theory.

Convergence of sequences and series. Power series. Functions, limits and continuity of functions of one or more variables.

Differentiation, Leibnitz formula, maxima and minima, Taylor's theorem. Differentiation of functions of several variables. Applications of differential Calculus.

Indefinite integral, Fundamental theorem of Calculus, Riemann integration and properties. Improper integrals. Differentiation under the integral sign. Double and multiple integrals and Applications.

Syllabus for Statistics

Probability and Sampling Distributions

Notions of sample space and probability, combinatorial probability, conditional probability and independence, random variable and expectations, moments, standard discrete and continuous distributions, sampling distributions of statistics based on normal samples, central limit theorem, approximation of Binomial to Normal or Poisson law. Bivariate normal and multivariate normal distributions.

Descriptive Statistics

Descriptive statistical measures, graduation of frequency curves, product-moment, partial and multiple correlation, Regression (bivariate and multivariate).

Inference

Elementary theory and methods of estimation (unbiasedness, minimum variance, sufficiency, maximum likelihood method, method of moments). Tests of hypotheses (basic concepts and simple applications of Neyman-Pearson Lemma). Confidence intervals. Tests of regression. Elements of non-parametric inference.

Design of Experiments and Sample Surveys

Basic designs (CRD/RBD/LSD) and their analyses. Elements of factorial designs. Conventional sampling techniques (SRSWR/SRSWOR) including stratification; ratio and regression methods of estimation.

Sample Questions

1. Let A be a $n \times n$ upper triangular matrix such that $AA^T = A^T A$. Show that A is a diagonal matrix.

2. Let X and Y be independent random variables with X having a binomial distribution with parameters 5 and 1/2 and Y having a binomial distribution with parameters 7 and 1/2. Find the probability that $|X - Y|$ is even.

3. Let A be a $n \times n$ orthogonal matrix, where n is even and suppose $|A| = -1$, where $|A|$ denotes the determinant of A. Show that $|I - A| = 0$, where I denotes the $n \times n$ identity matrix.

4. Let f be a non-decreasing, integrable function defined on $[0, 1]$. Show that

$$\left(\int_0^1 f(x)dx\right)^2 \leq 2\int_0^1 x(f(x))^2 dx.$$

5. Suppose X and U are independent random variables with

$$P(X = k) = \frac{1}{N+1}, \qquad k = 0, 1, 2, \ldots, N,$$

and U having a uniform distribution on $[0, 1]$. Let $Y = X + U$.

a) For $y \in \mathbb{R}$, find $P(Y \leq y)$.

b) Find the correlation coefficient between X and Y.

6. Consider a randomized block design with v treatments, each replicated r times. Let t_i be the treatment effect of the i-th treatment. Find $Cov(\sum l_i\hat{t}_i, \sum m_i\hat{t}_i)$ where $\sum l_i\hat{t}_i$ and $\sum m_i\hat{t}_i$ are the best linear unbiased estimators of $\sum l_i t_i$ and $\sum m_i t_i$ respectively and $\sum l_i = \sum m_i = \sum l_i m_i = 0$.

7. Suppose that $X_1, \ldots, X_n$ is a random sample of size $n \geq 1$ from a Poisson distribution with parameter λ. Find the minimum variance unbiased estimator of $e^{-\lambda}$.

8. Suppose $X_1, \ldots, X_n$ constitute a random sample from a population with density

$$f(x, \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right), \quad x > 0, \theta > 0.$$

Find the Cramer-Rao lower bound to the variance of an unbiased estimator of θ^2.

9. Let X_1, X_2, X_3 be independent random variables such that X_i is uniformly distributed in $(0, i\theta)$ for $i = 1, 2, 3$. Find the maximum likelihood estimator of θ and examine whether it is unbiased for θ.

10. Suppose that in 10 tosses of a coin we get 7 heads and 3 tails. Find a test at level $\alpha = 0.05$ to test that the coin is fair against the alternative that the coin is more likely to show up heads. Find the power function of this test.

11. Let X and Y be two random variables with joint probability density function

$$f(x, y) = \begin{cases} 1 & \text{if } -y < x < y, \ 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the regression equation of Y on X and that of X on Y.

2006

Sample Questions

1. Let A and B be two invertible $n \times n$ real matrices. Assume that $A+B$ is invertible. Show that $A^{-1} + B^{-1}$ is also invertible.

2. Maximize $x + y$ subject to the condition that $2x^2 + 3y^2 \leq 1$.

3. Let $X_1, X_2 \ldots$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{4}$, let $Y_1, Y_2, \ldots$ be another sequence of i.i.d. Bernoulli random variables with parameter $\frac{3}{4}$ and Let N be a geometric random variable with parameter $\frac{1}{2}$ (i. e. $P(N = k) = \frac{1}{2^k}$ for $k = 1, 2 \ldots$). Assume the X_i's, Y_j's and N are all independent.

   Compute $\mathrm{Cov}(\sum_{i=1}^{N} X_i, \sum_{i=1}^{N} Y_i)$.

4. Let U be uniformly distributed on the interval $(0, 2)$ and let V be an independent random variable which has a discrete uniform distribution on $\{0, 1, \ldots, n\}$. *i.e.*

$$P\{V = i\} = \frac{1}{n+1} \qquad \text{for } i = 0, 1, \ldots, n.$$

   Find the cumulative distribution function of $X = U + V$.

5. Suppose X is the number of heads in 10 tossses of a fair coin. Given $X = 5$, what is the probability that the first head occured in the third toss?

6. Let Y_1, Y_2, Y_3 be i.i.d. continuous random variables. For $i = 1, 2$, define U_i as

$$U_i = \begin{cases} 1 & \text{if } Y_{i+1} > Y_i, \\ 0 & \text{otherwise.} \end{cases}$$

   Find the mean and variance of $U_1 + U_2$.

7. Let Y be a random variable with probability density function

$$f_Y(y|\theta) = \begin{cases} \frac{1}{\theta} e^{-y/\theta} & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

   with $\theta > 0$.

   Suppose that the conditional distribution of X given $Y = y$ is $N(y, \sigma^2)$, with $\sigma^2 > 0$. Both θ and σ^2 are unknown parameters. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the joint distribution of X and Y.

   Find a nontrivial joint sufficient statistic for (θ, σ^2).

8. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the discrete distribution with joint probability mass function

$$f_{X,Y}(x,y) = \begin{cases} \frac{\theta}{4} & (x,y) = (0,0) \text{ and } (1,1), \\ \frac{2-\theta}{4} & (x,y) = (0,1) \text{ and } (1,0), \end{cases}$$

with $0 \leq \theta \leq 2$. Find the maximum likelihood estimator of θ.

9. Let $X_1, X_2, \ldots$ be i.i.d. random variables with density $f_\theta(x)$, $x \in \mathbb{R}$, $\theta \in (0,1)$ being the unknown parameter. Suppose that there exists an unbiased estimator T of θ based on sample size 1, i. e. $E_\theta(T(X_1)) = \theta$. Assume that $\mathrm{Var}(T(X_1)) < \infty$.

   (a) Find an estimator V_n for θ based on $X_1, \ldots X_n$ such that V_n is consistent for θ.

   (b) Let S_n be the MVUE (minimum variance unbiased estimator) of θ based on $X_1, \ldots, X_n$. Show that $\lim_{n\to\infty} \mathrm{Var}(S_n) = 0$.

10. For the data collected via a randomized block design with v treatments and b blocks, the following model is postulated:

$$E(y_{ij}) = \mu + \tau_i + \beta_j, \quad 1 \leq i \leq v, 1 \leq j \leq b,$$

where t_i and β_j are the effects of the ith treatment and the jth block respectively, and μ is a general mean. For $1 \leq i \leq v$, define $Q_i = T_i - \frac{G}{v}$, where T_i is the total of observations under the ith treatment and $G = \sum_{i=1}^{v} T_i$. Show that

$$E(Q_i) = (b - \frac{b}{v})\tau_i, \quad \mathrm{Var}(Q_i) = \sigma^2(b - \frac{b}{v}),$$
$$\mathrm{Cov}(Q_i, Q_j) = -(\frac{b}{v})\sigma^2 \text{ for } i \neq j,$$

where σ^2 is the per observation variance.

11. A straight line regression $E(y) = \alpha + \beta x$ is to be fitted using four observations. Assume $\mathrm{Var}(y|x) = \sigma^2$ for all x. The values of x at which observations are to be made lie in the closed interval $[-1, 1]$. The following choices of the values of x where observations are to be made are available:

   (a) two observations each at $x = -1$ and $x = 1$,

   (b) one observation each at $x = -1$ and $x = 1$ and two observations at $x = 0$,

(c) one observation each at $x = -1, -\frac{1}{2}, \frac{1}{2}, 1$.

If the interest is to estimate the slope with least variance, which of the above strategies would you choose and why?

12. Consider a possibly unbalanced coin with probability of heads in each toss being p, where p is unknown. Let X be the number of tails before the first head occurs. Find the uniformly most powerful test of level α for testing $H_0 : p = \frac{1}{6}$ against $H_1 : p > \frac{1}{6}$.

2007

Sample Questions :
Section A

1. Let A be a 2×2 matrix with real entries such that $A^2 = 0$. Find the determinant of $I + A$ where I denotes the identity matrix.

2. Let A and B be $n \times n$ real matrices such that $A^2 = A$ and $B^2 = B$. Suppose that $I - (A + B)$ is invertible. Show that $\text{rank}(A) = \text{rank}(B)$.

3. Let f be a function such that $f(0) = 0$ and f has derivatives of all order. Show that
$$\lim_{h \to 0} \frac{f(h) + f(-h)}{h^2} = f''(0)$$
where $f''(0)$ is the second derivative of f at 0.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. Define $g : [0, \infty) \to \mathbb{R}$ by,
$$g(x) = \int_{-x}^{x} (2xt + 1) f(t) dt.$$
Show that g is differentiable on $(0, \infty)$ and find the derivative of g.

5. Let X and Y be i.i.d. random variables, with $P(X = k) = 2^{-k}$ for $k = 1, 2, 3, \ldots$. Find $P(X > Y)$ and $P(X > 2Y)$.

6. 18 boys and 2 girls are made to stand in a line in a random order. Let X be the number of boys standing in between the girls. Find

   (a) $P(X = 5)$,

   (b) $E(X)$.

Section B

7. Let X and Y be i.i.d. exponentially distributed random variables with mean $\lambda > 0$. Define Z by:
$$Z = \begin{cases} 1 & \text{if } X < Y \\ 0 & \text{otherwise.} \end{cases}$$

Find the conditional mean $E(X \mid Z = 1)$.

8. Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform $(0, 1)$ random variables and suppose

$$X = \max(U_1, U_2, \ldots, U_n) \text{ and } Y = \min(U_1, U_2, \ldots, U_n).$$

Find the distribution of $Z = X - Y$.

9. Let X_1 and X_2 be i.i.d. random variables from Bernoulli(θ) distribution. Verify if the statistic $X_1 + 2X_2$ is sufficient for θ.

10. Suppose X takes three values $1, 2$ and 3 with

$$P(X = k) = \begin{cases} (1-\theta)/2 & \text{if } k = 1 \\ 1/2 & \text{if } k = 2 \\ \theta/2 & \text{if } k = 3 \end{cases}$$

where $0 < \theta < 1$. Suppose that the following random sample of size 10 was drawn from the above distribution :

$$1, 3, 1, 2, 3, 1, 2, 2, 1, 1.$$

Find the m.l.e. of θ based on the above sample.

11. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from the exponential distribution with mean $\theta > 0$. Find the most powerful test for testing $H_0 : \theta = 2$ against $H_1 : \theta = 1$. Find the power of the test.

12. Let Y_1, Y_2 and Y_3 be uncorrelated random variables with common variance $\sigma^2 > 0$ such that

$$E(Y_1) = \beta_1 + \beta_2, E(Y_2) = 2\beta_1 \text{ and } E(Y_3) = \beta_1 - \beta_2$$

where β_1 and β_2 are unknown parameters. Find the residual (error) sum of squares under the above linear model.

2008

Sample Questions:

1. Let

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

Which of the following statements are false. In each case, justify your answer.

(a) A has only one real eigenvalue.

(b) $\text{Rank}(A) = \text{Trace}(A)$.

(c) Determinant of A equals the determinant of A^n for each integer $n > 1$.

2. For $k \geq 1$, let

$$a_k = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{kn} \exp\left(-\frac{1}{2}\frac{m^2}{n^2}\right).$$

Find $\lim_{k\to\infty} a_k$.

3. Let g be a continuous function with $g(1) = 1$ such that

$$g(x+y) = 5g(x)g(y)$$

for all x, y. Find $g(x)$.

[**Hint:** You may use the following result.

If f is a continuous function that satisfies $f(x+y) = f(x) + f(y)$ for all x, y, then $f(x) = xf(1)$.]

4. The unit interval $(0, 1)$ is divided into two sub-intervals by picking a point at random from inside the interval. Denoting by Y and Z the lengths of the longer and the shorter sub-intervals respectively, show that Y/Z does not have finite expectation.

5. Consider the i.i.d. sequence

$$X_1, X_2, X_3, X_4, X_5, X_6$$

where each X_i is one of the four symbols $\{a, t, g, c\}$. Further suppose that

$$P(X_1 = a) = 0.1 \qquad P(X_1 = t) = 0.2$$
$$P(X_1 = g) = 0.3 \qquad P(X_1 = c) = 0.4.$$

Let Z denote the random variable that counts the number of times that the *subsequence* **cat** occurs (*i.e.* the letters c, a and t occur consecutively and in the correct order) in the above sequence. Find $E(Z)$.

6. Let F and G be (one dimensional) distribution functions. Decide which of the following are distribution functions.

   (a) F^2,

   (b) H where $H(t) = \max\{F(t), G(t)\}$.

   Justify your answer.

7. Let X and Y be exponential random variables with parameters 1 and 2 respectively. Another random variable Z is defined as follows.

   A coin, with probability p of Heads (and probability $1 - p$ of Tails) is tossed. Define Z by

$$Z = \begin{cases} X & \text{if the coin turns Heads} \\ Y & \text{if the coin turns Tails} \end{cases}$$

   Find $P(1 \leq Z \leq 2)$.

8. Let $\underline{Y} = (Y_1, Y_2)'$ have the bivariate normal distribution $N_2(\underline{0}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Obtain the mean and variance of $U = \underline{Y}'\Sigma^{-1}\underline{Y} - Y_1^2/\sigma_1^2$.

9. Let $X_1, \ldots, X_m$ be a random sample from a uniform distribution on $\{1, 2, \ldots, N\}$ where N is an unknown positive integer. Find the MLE $\widehat{N}$ of N and find its distribution function.

10. Consider a population with three kinds of individuals labelled 1, 2 and 3. Suppose the proportion of individuals of the three types are given by $f(k, \theta), k = 1, 2, 3$ where $0 < \theta < 1$ and

$$f(k, \theta) = \begin{cases} \theta^2 & \text{if } k = 1 \\ 2\theta(1-\theta) & \text{if } k = 2 \\ (1-\theta)^2 & \text{if } k = 3. \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from this population. Find the most powerful test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ $(\theta_0 < \theta_1 < 1)$.

11. Let r be the number of successes in n Bernoulli trials with unknown probability p of success. Obtain the minimum variance unbiased estimator of $p - p^2$.

12. Consider a randomized block experiment with 4 treatments and 3 replicates (blocks) and let τ_i be the effect of the ith treatment $(1 \leq i \leq 4)$. Find all possible covariances between the least squares estimators of the following treatment contrasts:

   (a) $\tau_1 - \tau_2$,

   (b) $\tau_1 + \tau_2 - 2\tau_3$,

   (c) $\tau_1 + \tau_2 + \tau_3 - 3\tau_4$.

* For more smaple questions, visit http://www.isical.ac.in/~deanweb/MSTATSQ.html

2009

Sample Questions:

1. (a) Let A be an $n \times n$ real matrix such that $(A+I)^4 = 0$ where I denotes the identity matrix. Show that A is nonsingular.

   (b) Give an example of a nonzero 2×2 real matrix A such that $x'Ax = 0$ for all real vectors x.

2. Let $\{x_n : n \geq 0\}$ be a sequence of real numbers such that $x_{n+1} = \lambda x_n + (1-\lambda)x_{n-1}, n \geq 1$, for some $0 < \lambda < 1$.

   (a) Show that $x_n = x_0 + (x_1 - x_0)\sum_{k=0}^{n-1}(\lambda - 1)^k$.

   (b) Hence, or otherwise, show that x_n converges and find the limit.

3. Using an appropriate probability distribution or otherwise show that

$$\lim_{n\to\infty} \int_0^n \frac{\exp(-x)x^{n-1}}{(n-1)!}dx = \frac{1}{2}.$$

4. Let R and θ be independent non-negative random variables such that $R^2 \sim \chi_2^2$ and $\theta \sim \mathrm{U}(0, 2\pi)$. Fix $\theta_0 \in (0, 2\pi)$. Find the distribution of $R\sin(\theta + \theta_0)$.

5. Suppose F and G are continuous and strictly increasing distribution functions. Let X have distribution function F and $Y = G^{-1}F(X)$.

   (a) Find the distribution function of Y.

   (b) Hence, or otherwise, show that the joint distribution function of (X, Y), denoted by $H(x, y)$, is given by $H(x, y) = \min(F(x), G(y))$.

6. Suppose $X_1, \ldots, X_n$ are i.i.d. $\mathrm{N}(\theta, 1)$, $\theta_0 \leq \theta \leq \theta_1$, where $\theta_0 < \theta_1$ are two specified numbers. Find the MLE of θ and show that it is better

than the sample mean $\bar{X}$ in the sense of having smaller mean squared error.

7. Let X be a random variable taking values $-1, 0, 1, 2, 3$ such that

$$P_\theta(X = x) = \begin{cases} 2\theta(1-\theta) & \text{if} \quad x = -1, \\ \theta^x (1-\theta)^{3-x} & \text{if} \quad x = 0, 1, 2, 3. \end{cases}$$

where $\theta \in (0, 1)$. Show that $E_\theta(U(X)) = 0$ for all $\theta \in (0, 1)$ if and only if $U(k) = a\, U^\star(k)$ for some $a \in \mathbb{R}$, where

$$U^\star(k) = \begin{cases} 0 & \text{if} \quad k = 0, 3, \\ 1 & \text{if} \quad k = -1, \\ -2 & \text{if} \quad k = 1, 2. \end{cases}$$

8. Let $X_1, \ldots, X_n$ be i.i.d. observations from the density

$$f(x) = \frac{1}{\mu} \exp(-x/\mu), \quad x > 0,$$

where $\mu > 0$ is an unknown parameter.

Consider the problem of testing the hypothesis $\mathrm{H}_0 : \mu \leq \mu_0$ against $\mathrm{H}_1 : \mu > \mu_0$.

(a) Show that the test with critical region $[\bar{X} \geq \mu_0 \chi^2_{2n,1-\alpha}/2n]$, where $\chi^2_{2n,1-\alpha}$ is the $(1-\alpha)$-th quantile of the χ^2_{2n} distribution, has size α.

(b) Give an expression of the power in terms of the c.d.f. of the χ^2_{2n} distribution.

9. The average height and weight of a group of students turned out to be 5 ft 6 inches and 65 kilograms respectively. The correlation between heights and weights was found to be 0.6. Using the regression equation for predicting weight from height, the estimated weight of a 6 ft tall student was calculated to be 80 kilograms. Predict the height of a student whose weight is 60 kilograms.

10. (a) Suppose that the mean of a normal population is unknown. Assuming the population standard deviation to be known, a $100(1-2\alpha)\%$ confidence interval was calculated for the mean, and its length was found to be L. By what factor the sample size needs to be changed to ensure that a $100(1-\alpha)\%$ confidence interval for the same mean will be of length $L/2$.

    (b) An airlines company experiences that 20% of the customers who reserve their tickets actually do not show up on the day of the journey. Suppose the capacity of an aircraft is 200, and the company allows 225 customers to reserve for a flight. What is the probability that the flight will be at least 95% full? (Use any appropriate approximation and clearly state the relevant result/s.)

11. A simple random sample of size $n = n_1 + n_2$ is drawn without replacement from a finite population of size N. Further a simple random sample of size n_1 is drawn without replacement from the first sample. Let $\bar{y}$ and $\bar{y_1}$ be the respective sample means.

    (a) Find $V(\bar{y_1})$ and $V(\bar{y_2})$, where $\bar{y_2}$ is the mean of the remaining n_2 units in the first sample.

    (b) Show that $\text{Cov}(\bar{y_1}, \bar{y_2}) = -S^2/N$, where S^2 is the population variance.

12. If τ_i denotes the effect due to the ith treatment of a latin square design of order $m \times m$, then derive a test for testing the null hypothesis $H_0 : \tau_i - 2\tau_j + \tau_k = 0$. Also obtain a $100(1-\alpha)\%$ confidence interval for $\tau_i - 2\tau_j + \tau_k$.

2010

Sample Questions

1. Let A be a 4×4 matrix with non-negative entries such that the sum of the entries in each row of A equals 1. Find the sum of all entries in the matrix A^5.

2. Find the total number of isosceles triangles such that the length of each side is a positive integer less than or equal to 40.
(Here equilateral triangles are also counted as isosceles triangles.)

3. Let $f : (-1, 1) \to R$ be a continuous function with the property that $f(x) = f(x^4)$ for all x and $f(1/2) = a$. Show that the only function f satisfying this property is the constant function $f(x) = a$.

4. Let $X_1, X_2, ..., X_n$ be a random sample from a Bernoulli distribution with parameter p. Let $\bar{X}_n$ be the sample average given by $\bar{X}_n = \frac{1}{n}(X_1 + \quad + X_n)$. Find the expected value of $(\bar{X}_n - p)^3$.

5. A rectangle is drawn where the lengths of the sides are chosen randomly from $[0, 10]$ and independently of one another. Find the probability that the length of its diagonal is smaller than or equal to 10.

6. A person throws a fair dice repeatedly, each throw being independent of the other throws. At each trial he observes one of the numbers $1, 2, ..., 6$. What is the expected number of throws to observe

   (a) two distinct numbers for the first time?

   (b) three distinct numbers for the first time?

7. Let $Y_i = \alpha + \beta(\frac{i}{n}) + \epsilon_i$, $i = 1, ..., n$, where $\epsilon_1, ..., \epsilon_n$ are i.i.d. $N(0, \sigma^2)$. Here α, β, σ^2 are unknown parameters; $-\infty < \alpha, \beta < \infty$, $0 < \sigma^2 < \infty$, $n \geq 3$.

   (a) Find the least squares estimators of α and β.

   (b) Are the least squares estimates of α, β based on the observed values of Y_i's, UMVUE? Justify your answer.

8. A population contains 10 units, labelled $U_1, U_2, ..., U_{10}$. The value, of a character Y under study, for U_i is Y_i $(1 \leq i \leq 10)$. In order to estimate the population mean, $\bar{Y}$, a sample of size 4 is drawn in the following manner:

   (i) a simple random sample of size 2 is drawn without replacement from the units $U_2, U_3, ..., U_9$;

   (ii) the sample drawn in step (i) is augmented by the units U_1 and U_{10}. Based on the above sample in (ii), suggest an unbiased estimator of $\bar{Y}$ and obtain its variance.

9. Let $X_1, X_2, ..., X_n$ be i.i.d. $N(\mu, 1)$ and independently $Y_1, Y_2, ..., Y_m$ be i.i.d. $N(\mu, 4)$ random variables.

   (a) Test $H_0 : \mu = 0$ against the alternative $H_1 : \mu > 0$ based on the combined sample.

   (b) Find the power function of the test.

10. (a) Let $X \sim N(\mu, 1)$ and $Y = e^X$. Find the probability density function of Y.

    (b) Find $\mu_Y = E(Y)$; here $E(\cdot)$ stands for expectation.

    (c) Let $Y_1, Y_2, ..., Y_n$ be a random sample from the distribution of Y. Find the maximum likelihood estimator of μ_Y based on this sample.

11. Let X be a random variable uniformly distributed over $(0, 2\theta)$, $\theta > 0$ and $Y = \max(X, 2\theta - X)$.

    (a) Find $\mu = E(Y)$.

    (b) Let $X_1, ..., X_n$ be a random sample from the above distribution with unknown θ. Find two distinct unbiased estimators of μ, as defined in (a), based on the entire sample.

* For more sample questions, visit http://www.isical.ac.in/~deanweb/MSTATSQ.html

2011

Sample Questions:

1. Let $\mathbf{A}_{n\times n} = \begin{bmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{bmatrix}$, where $a \neq b$ and $a + (n-1)b = 0$.

   Suppose $\mathbf{B} = \mathbf{A} + \frac{\mathbf{1}\mathbf{1}'}{n}$, where $\mathbf{1} = (1, 1, \ldots, 1)'$ is an $n \times 1$ vector.

   Show that

   (a) $\mathbf{B}$ is non-singular.

   (b) $\mathbf{A}\mathbf{B}^{-1}\mathbf{A} = \mathbf{A}$.

2. A real valued function f is called *lower semi-continuous at a real number* x_0 if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x \in (x_0 - \delta, x_0 + \delta)$, $f(x) \geq f(x_0) - \epsilon$. The function f is called *lower semi-continuous* if it is lower semi-continuous at every real number x_0.

   (a) Show that every real valued continuous function f is lower semi-continuous.

   (b) If f and $-f$ both are lower semi-continuous, show that f is continuous.

   (c) Define the function f as

   $$f(x) = \begin{cases} 0, & x < 0 \\ a, & x = 0. \\ 1, & x > 0 \end{cases}$$

   Is there any value of a for which f will be lower semi-continuous? Justify your answer.

3. Suppose there are n boxes labelled $1, 2, \ldots, n$ and n balls labelled $1, 2, \ldots, n$. Balls are placed at random in the boxes. Let X be the number of empty boxes. Find $E(X)$ and $Var(X)$.

4. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent and identically distributed d-dimensional normal random vectors. Consider a fixed $\mathbf{x}_0 \in \mathbb{R}^d$ and for $i = 1, 2, \ldots,$ define $D_i = \|\mathbf{X}_i - \mathbf{x}_0\|$, the Euclidean distance between $\mathbf{X}_i$ and $\mathbf{x}_0$. Show that for every $\epsilon > 0$,

$$P\left[\min_{1\leq i\leq n} D_i > \epsilon\right] \to 0 \text{ as } n \to \infty.$$

5. Consider a set of n observations $\{x_1, x_2, \ldots, x_n\}$ with mean $\bar{x}$ and variance s^2. When a new observation x_{n+1} is added to this set, the mean decreases, but the variance remains the same. Express x_{n+1} in terms of $n, \bar{x}$ and s^2.

6. Consider one way ANOVA model

$$E(y_{ij}) = \mu + \tau_i, \quad i = 0, 1, 2, 3, 4; \quad j = 1, 2, \ldots, n_i.$$

It it is known that $n_0 + \cdots + n_4 = 18$ and we want to estimate the pairwise differences $\tau_0 - \tau_i$ for $i = 1, 2, 3, 4$, which of the following strategies will you choose and why?

(a) $n_0 = 14$, $n_1 = n_2 = n_3 = n_4 = 1$.

(b) $n_0 = 10$, $n_1 = n_2 = n_3 = n_4 = 2$.

(c) $n_0 = 6$, $n_1 = n_2 = n_3 = n_4 = 3$.

(d) $n_0 = 2$, $n_1 = n_2 = n_3 = n_4 = 4$.

7. An ecologist collected data on pea plants in different fields. Given a field of area x, the number of pea plants Y in the field is modelled as Poisson with mean βx. Based on n observations $\{(x_i, y_i), 1 \leq i \leq n\}$, compute

(a) the maximum likelihood estimate of β,

(b) the estimate of β by minimizing $\sum_{i=1}^{n}(y_i - \beta x_i)^2$.

Compare the two estimators in (a) and (b) in terms of their bias and mean square error.

8. Consider the problem of testing

$$H_0 : f(x) = \frac{1}{\pi(1+x^2)} \quad \text{vs} \quad H_1 : f(x) = \frac{1}{2}e^{-|x|}$$

based on a single observation X with density $f(x)$. Calculate the power of the most powerful test of size α.

9. Consider observations on three variables X_1, X_2 and X_3. Suppose that X_1 is regressed on X_2. When the residual of the above regression is regressed on X_3, the regression coefficient of X_3 is β_3. When X_1 is regressed on X_2 and X_3 simultaneously, the regression coefficient of X_3 is β_3^*.

   (a) Show that $|\beta_3| \leq |\beta_3^*|$.

   (b) When does the equality hold in (a)?

10. The following describes summary statistics for performance in examination of a random sample of 50 students from a population of 250 students of a certain school.

	Passed in English	Failed in English
Passed in Mathematics	23	17
Failed in Mathematics	3	7

   (a) Find an estimate of the percentage of students failing in exactly one subject, and also find an unbiased estimate of its variance.

   (b) If it is further known that overall 202 students have passed in Mathematics, how will you modify the estimate of the percentage in (a) above?

* For more sample questions, visit http://www.isical.ac.in/~deanweb/MSTATSQ.html

2012

Sample Questions

1. Find a basis for the subspace of $\mathbb{R}^4$ spanned by the four vectors
$$\begin{pmatrix}1\\1\\2\\4\end{pmatrix}, \begin{pmatrix}2\\-1\\-5\\2\end{pmatrix}, \begin{pmatrix}1\\1\\-4\\0\end{pmatrix}, \begin{pmatrix}2\\1\\1\\6\end{pmatrix}.$$

2. Let f be a polynomial. Assume that $f(0) = 1$, $\lim_{x\to\infty} f''(x) = 4$ and $f(x) \geq f(1)$ for all $x \in \mathbb{R}$. Find $f(2)$.

3. Let X_1 and X_2 be i.i.d. exponential random variables with mean $\lambda > 0$. Let $Y_1 = X_1 - X_2$ and $Y_2 = RX_1 - (1-R)X_2$, where R is a Bernoulli random variable with parameter $1/2$ and is independent of X_1 and X_2.

   (a) Show that Y_1 and Y_2 have the same distribution.

   (b) Obtain the common density function.

4. Let $\begin{pmatrix}X\\Y\end{pmatrix}$ be a bivariate normal vector such that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ and $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. Let $S \subset \mathbb{R}^2$ be defined by
$$S = \{(a,b) : aX + bY \text{ is independent of } Y\}.$$

   (a) Show that S is a subspace.

   (b) Find its dimension.

5. Let $X_1, X_2, \ldots, X_j \ldots$ be i.i.d. $N(0,1)$ random variables. Show that for any $a > 0$,
$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n} X_i^2 \leq a\right) = 0.$$

6. There are two biased coins – one which has probability $1/4$ of showing heads and $3/4$ of showing tails, while the other has probability $3/4$ of showing heads and $1/4$ of showing tails when tossed. One of the two coins is chosen at random and is then tossed 8 times.

(a) Given that the first toss shows heads, what is the probability that in the next 7 tosses there will be exactly 6 heads and 1 tail?

(b) Given that the first toss shows heads and the second toss shows tail, what is the probability that the next 6 tosses all show heads?

7. Consider a randomized (complete) block design with 4 treatments and 5 replications and, let t_i be the effect of the i-th treatment ($1 \leq i \leq 4$). Consider the following three treatment contrasts.

$$\frac{1}{\sqrt{2}}(t_1 - t_2), \quad \frac{1}{\sqrt{6}}(t_1 + t_2 - 2t_3) \quad \text{and} \quad \frac{1}{\sqrt{12}}(t_1 + t_2 + t_3 - 3t_4).$$

(a) Find the variances of the best linear unbiased estimators of the above treatment contrasts.

(b) Find all the covariances between them.

8. Let V_1 be the variance of the estimated mean from a stratified random sample of size n with proportional allocation. Assume that the strata sizes are such that the allocations are all integers.

Let V_2 be the variance of the estimated mean from a simple random sample of size n.

Show that the ratio V_1/V_2 is independent of n.

9. Suppose X_1 and X_2 are i.i.d. Bernoulli random variables with parameter p where it is known that $\frac{1}{3} \leq p \leq \frac{2}{3}$. Find the maximum likelihood estimator $\widehat{p}$ of p based on X_1 and X_2.

10. Let $X_1, X_2, \ldots, X_{10}$ be i.i.d. Poisson random variables with unknown parameter $\lambda > 0$. Find the minimum variance unbiased estimator of $\exp\{-2\lambda\}$.

Note: For more sample questions you can visit `http://www.isical.ac.in/~deanweb/MSTATSQ.html`.

2013

Sample Questions

1. Suppose V is the space of all $n \times n$ matrices with real elements. Define $T : V \to V$ by $T(A) = AB - BA$, $A \in V$, where $B \in V$ is a fixed matrix. Show that for any $B \in V$

   (a) T is linear;

   (b) T is not one-one;

   (c) T is not onto.

2. Let f be a real valued function satisfying

$$|f(x) - f(a)| \leq C |x - a|^{\gamma},$$

   for some $\gamma > 0$ and $C > 0$.

   (a) If $\gamma = 1$, show that f is continuous at a;

   (b) If $\gamma > 1$, show that f is differentiable at a.

3. Suppose integers are formed by taking one or more digits from the following

$$2, 2, 3, 3, 4, 5, 5, 5, 6, 7.$$

   For example, 355 is a possible choice while 44 is not. Find the number of distinct integers that can be formed in which

   (a) the digits are non-decreasing;

   (b) the digits are strictly increasing.

4. Consider n independent observations $\{(x_i, y_i) : 1 \le i \le n\}$ from the model

$$Y = \alpha + \beta x + \epsilon,$$

where ϵ is normal with mean 0 and variance σ^2. Let $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ be the maximum likelihood estimators of α, β and σ^2, respectively. Let v_{11}, v_{22} and v_{12} be the estimated values of $\mathrm{Var}(\hat{\alpha})$, $\mathrm{Var}(\hat{\beta})$ and $\mathrm{Cov}(\hat{\alpha}, \hat{\beta})$, respectively.

(a) What is the estimated mean of Y when $x = x_0$? Estimate the mean squared error of this estimator.

(b) What is the predicted value of Y when $x = x_0$? Estimate the mean squared error of this predictor.

5. A box has an unknown number of tickets serially numbered $1, 2, \ldots, N$. Two tickets are drawn using simple random sampling without replacement (SRSWOR) from the box. If X and Y are the numbers on these two tickets and $Z = \max(X, Y)$, show that

(a) Z is not unbiased for N;

(b) $aX + bY + c$ is unbiased for N if and only is $a + b = 2$ and $c = -1$.

6. Suppose X_1, X_2 and X_3 are three independent and identically distributed Bernoulli random variables with parameter p, $0 < p < 1$. Verify if the following statistics are sufficient for p:

(a) $X_1 + 2X_2 + X_3$;

(b) $2X_1 + 3X_2 + 4X_3$.

7. Suppose X_1 and X_2 are two independent and identically distributed random variables with Normal $(\theta, 1)$ distribution. Further, consider a Bernoulli random variable V with $P[V = 1] = 1/4$, which is independent of X_1 and X_2. Define X_3 as

$$X_3 = \begin{cases} X_1 & \text{when } V = 0, \\ X_2 & \text{when } V = 1. \end{cases}$$

For testing $H_0 : \theta = 0$ against $H_1 : \theta = 1$ consider the test:

$$\text{Reject } H_0 \text{ if } \ (X_1 + X_2 + X_3)\,/3 > c.$$

Find c such that the test has size 0.05.

8. Suppose X_1 is a standard normal random variable. Define

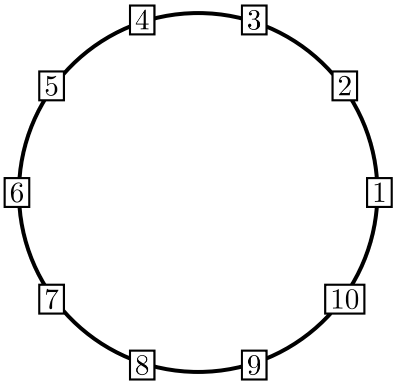
$$X_2 = \begin{cases} -X_1 & \text{if } |X_1| < 1, \\ X_1 & \text{otherwise.} \end{cases}$$

   (a) Show that X_2 is also a standard normal random variable.

   (b) Obtain the cumulative distribution function of X_1+X_2 in terms of the cumulative distribution function of a standard normal random variable.

9. Envelopes are on sale for Rs. 30 each. Each envelope contains exactly one coupon, which can be one of four types with equal probability. Suppose you keep on buying envelopes and stop when you collect all the four types of coupons. What will be your expected expenditure?

10. There are 10 empty boxes numbered $1, 2, \ldots, 10$ placed sequentially on a circle as shown in the figure.



We perform 100 independent trials. At each trial, one box is selected with probability $1/10$ and one ball is placed in each of the two neighbouring boxes of the selected one.

Define X_k to be the number of balls in the k^{th} box at the end of 100 trials.

(a) Find $\mathrm{E}\left[X_k\right]$ for $1 \leq k \leq 10$.

(b) Find $\mathrm{Cov}\left(X_k, X_5\right)$ for $1 \leq k \leq 10$.

Note: For more sample questions you can visit
`http://www.isical.ac.in/~deanweb/MSTATSQ.html`.

2013

BOOKLET No. TEST CODE: PSB

Afternoon

Questions: 10 Time: 2 hours

Write your Name, Registration number, Test Code, Number of this booklet, etc. in the appropriate places on the answer–booklet.

- All questions carry equal weight.
- Answer at least one question from GROUP A.
- Best five answers subject to the above condition will be considered.

Answer to each question should start on a fresh page.

ALL ROUGH WORK MUST BE DONE ON THIS BOOKLET AND/OR THE ANSWER–BOOKLET. YOU ARE NOT ALLOWED TO USE CALCULATORS.

STOP! WAIT FOR THE SIGNAL TO START.

GROUP A

1. Let $E = \{1, 2, \dots, n\}$, where n is an odd positive integer. Let V be the vector space of all functions from E to $\mathbb{R}^3$, where the vector space operations are given by

$$\begin{aligned}(f+g)(k) &= f(k)+g(k), \quad \text{for } f, g \in V, k \in E,\\ (\lambda f)(k) &= \lambda f(k), \quad \text{for } f \in V, \lambda \in \mathbb{R}, k \in E.\end{aligned}$$

   (a) Find the dimension of V.

   (b) Let $T\colon V \to V$ be the map given by

$$Tf(k) = \frac{1}{2}\left(f(k) + f(n+1-k)\right), \quad k \in E.$$

   Show that T is linear.

   (c) Find the dimension of the null space of T.

2. Let $a_1 < a_2 < \cdots < a_m$ and $b_1 < b_2 < \cdots < b_n$ be real numbers such that

$$\sum_{i=1}^{m} |a_i - x| = \sum_{j=1}^{n} |b_j - x| \ \text{ for all } x \in \mathbb{R}.$$

   Show that $m = n$ and $a_j = b_j$ for $1 \le j \le n$.

GROUP B

3. Let $S = \{1, 2, \dots, n\}$.

   (a) In how many ways can we choose two subsets A and B of S so that $B \neq \emptyset$ and $B \subseteq A \subseteq S$?

   (b) In how many of these cases is B a proper subset of A?

4. Consider a machine with three components whose times to failure are independently distributed as exponential random variables with mean λ. The machine continues to work as long as at least two components work. Find the expected time to failure of the machine.

5. A pin whose centre is fixed on a flat table is randomly and independently spun twice. Each time, the final position is noted by drawing a line segment.

   (a) What is the probability that the smallest angle between the two segments is more than half of the largest angle ?

   (b) What is the probability that at least one of the two segments makes an angle which is less than 45° with the x-axis (when measured in the anti-clockwise direction) ?

6. There are twenty individuals numbered $1, 2, \ldots, 20$. Each individual chooses 10 others from this group in a random fashion, independently of the choices of the others, and makes one phone call to each of the 10.

   (a) Let X be the number of calls handled (incoming as well as outgoing) by Individual 1. Find $E(X)$.

   (b) Let Y be the number of calls between Individual 1 and Individual 2. Find $E(Y)$.

   (c) Find $E(X|Y = 1)$.

7. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $Uniform(\theta, 1)$ population, where $\theta < 1$.

   (a) Find the MLE $\hat{\theta}$ of θ.

   (b) Find c such that $c\hat{\theta}$ is unbiased for θ.

8. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from some distribution with mean μ and variance σ^2. Let

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

   where $\bar{X}$ is the sample mean. Show that s underestimates σ (that is, it has negative bias).

9. Let $X_1 \sim Geo(p_1)$ and $X_2 \sim Geo(p_2)$ be independent random variables, where $Geo(p)$ refers to the Geometric distribution whose p.m.f. f is given by

$$f(k) = p(1-p)^k, \quad k = 0, 1, \ldots .$$

   We are interested in testing the null hypothesis $H_0 : p_1 = p_2$ against the alternative $H_1 : p_1 < p_2$. Intuitively, it is clear that we should reject if X_1 is large, but unfortunately we cannot compute a cutoff because the distribution of X_1 under H_0 depends on the unknown (common) value of p_1 and p_2.

   (a) Let $Y = X_1 + X_2$. Find the conditional distribution of $X_1 | Y = y$ when $p_1 = p_2$.

   (b) Based on the result obtained in (a), derive a level 0.05 test for H_0 against H_1 that rejects H_0 when X_1 is large.

10. Let y_1, y_2, y_3, y_4 be uncorrelated observations with common variance σ^2 and expectations given by

$$E(y_1) = E(y_2) = \beta_1 + \beta_2 + \beta_3$$
$$E(y_3) = E(y_4) = \beta_1 - \beta_2$$

An observational function $\sum_{i=1}^{4} a_i y_i$ is said to be an *error function* if its expectation is zero.

(a) Obtain a maximal set of linearly independent error functions for the above model (that is, the set should be such that adding any other error function to the set would make the set linearly dependent). Justify your answer.

(b) Obtain an unbiased estimator of $3\beta_1 - \beta_2 + \beta_3$ such that it is uncorrelated with each of the error functions obtained in (a).

2014

BOOKLET No. TEST CODE: PSB

Afternoon

Questions: 9	Time: 2 hours

Write your Registration number, Test Code, Number of this booklet, etc. in the appropriate places on the answer–booklet.

- All questions carry equal weight.
- Answer any two questions from GROUP A and any three questions from GROUP B.

Answer to each question should start on a fresh page.

ALL ROUGH WORK MUST BE DONE ON THIS BOOKLET AND/OR THE ANSWER–BOOKLET. YOU ARE NOT ALLOWED TO USE CALCULATORS.

STOP! WAIT FOR THE SIGNAL TO START.

GROUP A

1. Let $f : R \to R$ be a function which is continuous at 0 and $f(0) = 1$. Also assume that f satisfies the following relation for all x:

$$f(x) - f(x/2) = \frac{3x^2}{4} + x.$$

Find $f(3)$.

2. For any $n \times n$ matrix $A = ((a_{ij}))$, consider the following three properties:

   1. a_{ij} is real valued for all i, j and A is upper triangular.
   2. $\sum_{j=1}^{n} a_{ij} = 0$, for all $1 \leq i \leq n$.
   3. $\sum_{i=1}^{n} a_{ij} = 0$, for all $1 \leq j \leq n$.

   Define the following set of matrices:

$$\mathcal{C}_n = \{A : \ A \text{ is } n \times n \text{ and satisfies (1), (2) and (3) above}\}.$$

   (a) Show that $\mathcal{C}_n$ is a vector space for any $n \geq 1$.

   (b) Find the dimension of $\mathcal{C}_n$, when $n = 2$ and $n = 3$.

3. Let A be a real valued and symmetric $n \times n$ matrix with entries such that $A \neq \pm I$ and $A^2 = I$.

   (a) Prove that there exist non-zero column vectors v and w such that $Av = v$ and $Aw = -w$.

   (b) Prove that every vector z has a unique decomposition $z = x + y$ where $Ax = x$ and $Ay = -y$.

GROUP B

4. Suppose that 15 identical balls are placed in 3 boxes labeled A, B and C. What is the number of ways in which Box A can have more balls than Box C?

5. Suppose that X and Y are random variables such that

$$\begin{aligned} E(X+Y) = E(X-Y) &= 0\,, \\ \mathrm{Var}(X+Y) &= 3\,, \\ \mathrm{Var}(X-Y) &= 1\,. \end{aligned}$$

   (a) Evaluate $\mathrm{Cov}(X, Y)$.

   (b) Show that $E\big(|X+Y|\big) \leq \sqrt{3}$.

   (c) If in addition, it is given that (X, Y) has a bivariate normal distribution, calculate $E\big(|X+Y|^k\big)$ for all positive integers k.

6. Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with mean μ and variance 1. Also assume that $Y_1, \ldots, Y_n$ are i.i.d. with probability mass function $P(Y_i = \pm 1) = 1/2$ for all $1 \leq i \leq n$ and independent of $X_1, \ldots, X_n$. Define T_n as follows:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \cdot |X_i|, \quad n \geq 1.$$

   (a) For any fixed $z \in R$, find

$$\lim_{n\to\infty} P\left(\sqrt{n}T_n \leq z\right).$$

   (b) Using the result in part (a) above, find random quantities L_n and U_n, based on T_n, such that

$$\lim_{n\to\infty} P\left(L_n \leq \mu \leq U_n\right) = 0.95\ .$$

7. Suppose that $X_1, \ldots, X_n$ are i.i.d. with probability density function

$$f_\theta(x) = \begin{cases} \dfrac{1}{2\sqrt{\theta x}} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $X_{(n)}$ is sufficient for θ.

(b) Consider a test of size α $(0 < \alpha < 1)$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ $(> \theta_0)$, that rejects H_0 if and only if $X_{(n)} > k$.

i. Determine the value of k.

ii. Find the minimum sample size required such that the power of the test is at least β $(\alpha < \beta < 1)$.

8. Consider the regression model:

$$y_i = bx_i + e_i, \quad 1 \leq i \leq n,$$

where x_i's are fixed non-zero real numbers and e_i's are independent random variables with mean zero and equal variance.

(a) Consider estimators of the form $\sum_{i=1}^{n} a_i y_i$ (where a_i's are non random real numbers) that are unbiased for b. Show that the least squares estimator of b has the minimum variance in this class of estimators.

(b) Suppose that x_i's take values -1 or $+1$ and e_i's have density

$$f(t) = \frac{1}{2}\, e^{-|t|}, \quad t \in R.$$

Find the maximum likelihood estimator of b.

9. Consider a collection of N cards, numbered $1, 2, \ldots, N$, where $N \geq 2$. A card is drawn at random and set aside. Suppose that n cards are selected from the remaining $(N - 1)$ cards using SRSWR and their numbers noted as $Y_1, \ldots, Y_n$. If $S = \sum_{i=1}^{n} Y_i$, find $E(S)$ and $\mathrm{Var}(S)$.

1. Let $f : R \to R$ be a function which is continuous at 0 and $f(0) = 1$. Also assume that f satisfies the following relation for all x:

$$f(x) - f(x/2) = \frac{3x^2}{4} + x.$$

Find $f(3)$.

2. For any $n \times n$ matrix $A = ((a_{ij}))$, consider the following three properties:

   1. a_{ij} is real valued for all i, j and A is upper triangular.
   2. $\sum_{j=1}^{n} a_{ij} = 0$, for all $1 \leq i \leq n$.
   3. $\sum_{i=1}^{n} a_{ij} = 0$, for all $1 \leq j \leq n$.

   Define the following set of matrices:

   $$\mathcal{C}_n = \{A : \ A \text{ is } n \times n \text{ and satisfies (1), (2) and (3) above}\}.$$

   (a) Show that $\mathcal{C}_n$ is a vector space for any $n \geq 1$.

   (b) Find the dimension of $\mathcal{C}_n$, when $n = 2$ and $n = 3$.

3. Let A be a real valued and symmetric $n \times n$ matrix with entries such that $A \neq \pm I$ and $A^2 = I$.

   (a) Prove that there exist non-zero column vectors v and w such that $Av = v$ and $Aw = -w$.

   (b) Prove that every vector z has a unique decomposition $z = x + y$ where $Ax = x$ and $Ay = -y$.

4. Suppose 15 identical balls are placed in 3 boxes labeled A, B and C. What is the number of ways in which Box A can have more balls than Box C?

5. Suppose that X and Y are random variables such that

$$\begin{aligned} E(X+Y) = E(X-Y) &= 0\,, \\ \mathrm{Var}(X+Y) &= 3\,, \\ \mathrm{Var}(X-Y) &= 1\,. \end{aligned}$$

(a) Evaluate $\mathrm{Cov}(X,Y)$.

(b) Show that $E|X+Y| \le \sqrt{3}$.

(c) If in addition, it is given that (X,Y) is bivariate normal, calculate $E\big(|X+Y|^3\big)$.

6. Suppose $X_1,\ldots,X_n$ are i.i.d. random variables with mean μ and variance 1. Also assume that $Y_1,\ldots,Y_n$ are i.i.d. with probability mass function $\mathbf{P}(Y_i = \pm 1) = 1/2$ for all $1 \le i \le n$ and independent of $X_1,\ldots,X_n$. Define T_n as follows:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \cdot |X_i|, \quad n \ge 1.$$

(a) For any fixed $z \in R$, find

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n}T_n \le z\right).$$

(b) Using the result in part (a) above, find random quantities L_n and U_n, based on T_n, such that

$$\lim_{n\to\infty} \mathbf{P}\left(L_n \le \mu \le U_n\right) = 0.95\ .$$

7. Suppose that $X_1, \ldots, X_n$ are i.i.d. with probability density function

$$f_\theta(x) = \begin{cases} \frac{1}{2\sqrt{\theta x}} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $X_{(n)}$ is sufficient for θ.

(b) Consider a test of size α ($0 < \alpha < 1$) for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ ($> \theta_0$), that rejects H_0 if and only if $X_{(n)} > k$.

i. Determine the value of k.

ii. Find the minimum sample size required such that the power of the test is at least β ($\alpha < \beta < 1$).

8. Consider the regression model:

$$y_i = bx_i + e_i, \quad 1 \leq i \leq n,$$

where x_i's are fixed non-zero real numbers and e_i's are independent random variables with mean zero and equal variance.

(a) Consider estimators of the form $\sum_{i=1}^n a_i y_i$ (where a_i's are non random real numbers) that are unbiased for b. Show that the least squares estimator of b has the minimum variance in this class of estimators.

(b) Suppose that x_i's take values -1 or $+1$ and e_i's have density

$$f(t) = \frac{1}{2}\, e^{-|t|}, \quad t \in R.$$

Find the maximum likelihood estimator of b.

9. Consider a collection of N cards, numbered $1, 2, \ldots, N$, where $N \geq 2$. A card is drawn at random and set aside. Suppose n cards are selected from the remaining $(N-1)$ cards using SRSWR and their numbers noted as $Y_1, \ldots, Y_n$. If $S = \sum_{i=1}^{n} Y_i$, find $E(S)$ and $Var(S)$.

2016

BOOKLET No. TEST CODE: PSB

Afternoon

Questions: 10	**Time: 2 hours**

Write your Registration number, Test Centre, Test Code, and the Number of this booklet in the appropriate places on the answer-book.

- All questions carry equal weight.
- Answer two questions from GROUP A and four questions from GROUP B.
- Best six answers subject to the above conditions will be considered.

Answer to each question should start on a fresh page.

ALL ROUGH WORK MUST BE DONE ON THIS BOOKLET AND/OR THE ANSWER-BOOK. YOU ARE NOT ALLOWED TO USE CALCULATORS IN ANY FORM.

STOP! WAIT FOR THE SIGNAL TO START.

GROUP A

1. Let x, y be real numbers such that $xy = 10$. Find the minimum value of $|x + y|$ and all also find all the points (x, y) where this minimum value is achieved.

   Justify your answer.

2. Determine the average value of

$$i_1 i_2 + i_2 i_3 + \cdots + i_9 i_{10} + i_{10} i_1$$

   taken over all permutations $i_1, i_2, \ldots, i_{10}$ of $1, 2, \ldots, 10$.

3. For any two events A and B, show that

$$(\mathrm{P}(A \cap B))^2 + (\mathrm{P}(A \cap B^c))^2 + (\mathrm{P}(A^c \cap B))^2 + (\mathrm{P}(A^c \cap B^c))^2 \geq \frac{1}{4}.$$

4. Let X, Y, and Z be three Bernoulli $(\frac{1}{2})$ random variables such that X and Y are independent, Y and Z are independent, and Z and X are independent.

   (a) Show that $\mathrm{P}(XYZ = 0) \geq \frac{3}{4}$.

   (b) Show that if equality holds in (a), then $Z = \begin{cases} 1 & \text{if } X = Y, \\ 0 & \text{if } X \neq Y. \end{cases}$

GROUP B

5. Let $n \geq 2$, and $X_1, X_2, \ldots, X_n$ be independent and identically distributed Poisson (λ) random variables for some $\lambda > 0$. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the corresponding order statistics.

   (a) Show that $\mathrm{P}(X_{(2)} = 0) \geq 1 - n(1 - e^{-\lambda})^{n-1}$.

   (b) Evaluate the limit of $\mathrm{P}(X_{(2)} > 0)$ as the sample size $n \to \infty$.

6. Suppose that random variables X and Y jointly have a bivariate normal distribution with $\mathrm{E}(X) = \mathrm{E}(Y) = 0, \mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, and correlation ρ. Compute the correlation between e^X and e^Y.

7. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with probability mass function

$$f(x;\theta) = \frac{x\theta^x}{h(\theta)} \qquad \text{for } x = 1, 2, 3, \ldots$$

where $0 < \theta < 1$ is an unknown parameter and $h(\theta)$ is a function of θ. Show that the maximum likelihood estimator of θ is also a method of moments estimator.

8. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be independent and identically distributed pairs of random variables with $\mathrm{E}(X_1) = \mathrm{E}(Y_1), \mathrm{Var}(X_1) = \mathrm{Var}(Y_1) = 1$, and $\mathrm{Cov}(X_1, Y_1) = \rho \in (-1, 1)$.

   (a) Show that there exists a function $c(\rho)$ such that

   $$\lim_{n\to\infty} \mathrm{P}\left(\sqrt{n}(\overline{X} - \overline{Y}) \le c(\rho)\right) = \Phi(1)$$

   where Φ denotes the standard normal cumulative distribution function.

   (b) Given $\alpha \in (0, 1)$, obtain a statistic L_n which is a function of $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ such that

   $$\lim_{n\to\infty} \mathrm{P}(L_n < \rho < 1) = \alpha.$$

9. Suppose $X_1, X_2, \ldots, X_N$ are independent exponentially distributed random variables with mean 1, where N is unknown. We only observe the largest X_i value and denote it by T. We want to test $H_0 : N = 5$ against $H_1 : N = 10$. Show that the most powerful test of size 0.05 rejects H_0 when $T > c$ for some c, and determine c.

10. Consider a population with $N > 1$ units having values $y_1, y_2, \ldots, y_N$. A sample of size n_1 is drawn from the population using SRSWOR. From the remaining part of the population, a sample of size n_2 is drawn using SRSWOR. Show that the covariance between the two sample means is

$$\frac{-\sum_{i=1}^{N}(y_i - \bar{y})^2}{N(N-1)},$$

where $\bar{y} = \dfrac{1}{N}\sum_{i=1}^{N} y_i$.

2017

BOOKLET No. TEST CODE: PSB

Afternoon

Questions: 8	**Time: 2 hours**

Write your Registration number, Test Centre, Test Code, and the Number of this booklet in the appropriate places on the answer-book.

- Answer two questions from GROUP A and four questions from GROUP B.
- Best six answers subject to the above conditions will be considered.
- Each question carries 20 marks.
- The maximum possible marks is 120.

Answer to each question should start on a fresh page.

All rough work must be done on this booklet and/or the answer-book. You are not allowed to use calculators in any form.

STOP! WAIT FOR THE SIGNAL TO START.

GROUP A

1. Let a and b be real numbers. Show that there exists a unique 2×2 real symmetric matrix A with $\mathrm{trace}(A) = a$ and $\det(A) = b$ if and only if $a^2 = 4b$.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be an infinitely differentiable function and suppose that for some $n \geq 1$,
$$f(1) = f(0) = f^{(1)}(0) = f^{(2)}(0) = \cdots = f^{(n)}(0) = 0,$$
where $f^{(k)}$ denotes the k-th derivative of f for $k \geq 1$. Prove that there exists $x \in (0,1)$ such that $f^{(n+1)}(x) = 0$.

3. Consider an urn containing 5 red, 5 black, and 10 white balls. If balls are drawn *without* replacement from the urn, calculate the probability that in the first 7 draws, at least one ball of each colour is drawn.

GROUP B

4. Let $X_1, X_2, \ldots, X_n$ be independent random variables, with X_i having probability mass function
$$P(X_i = k) = \left(\frac{i}{i+1}\right)^k \frac{1}{i+1}, \text{ for } k = 0, 1, 2, \ldots$$
and for all $i = 1, \ldots, n$. Let $M = \min\{X_i : 1 \leq i \leq n\}$. Derive the probability mass function of M.

5. The lifetime in hours of each bulb manufactured by a particular company follows an independent exponential distribution with mean λ. To test the null hypothesis $H_0 : \lambda = 1000$ against the alternative $H_1 : \lambda = 500$, a statistician sets up an experiment with 50 bulbs, with 5 bulbs in each of 10 different locations, to examine their lifetimes.

   To get quick preliminary results, the statistician decides to stop the experiment as soon as one bulb fails at each location. Let Y_i denote the lifetime of the first bulb to fail at location i. Obtain the most powerful test of H_0 against H_1 based on $Y_1, Y_2, \ldots, Y_{10}$, and compute its power.

6. Suppose you have a 4-digit combination lock, but you have forgotten the correct combination. Consider the following three strategies to find the correct one:

   (i) Try the combinations consecutively from 0000 to 9999.

   (ii) Try combinations using simple random sampling *with* replacement from the set of all possible combinations.

   (iii) Try combinations using simple random sampling *without* replacement from the set of all possible combinations.

   Assume that the true combination was chosen uniformly at random from all possible combinations. Determine the expected number of attempts needed to find the correct combination in all three cases.

7. Consider independent observations $\{(y_i, x_{1i}, x_{2i}) : 1 \le i \le n\}$ from the regression model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i,\ i = 1, \ldots, n\,,$$

   where x_{1i} and x_{2i} are scalar covariates, β_1 and β_2 are unknown scalar coefficients, and ϵ_i are uncorrelated errors with mean 0 and variance $\sigma^2 > 0$. Instead of using the correct model, we obtain an estimate $\hat{\beta}_1$ of β_1 by minimizing

$$\sum_{i=1}^{n} (y_i - \beta_1 x_{1i})^2.$$

   Find the bias and mean squared error of $\hat{\beta}_1$.

8. Let $\theta > 0$ be an unknown parameter, and $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with density

$$f(x) = \begin{cases} 2x/\theta^2 & , 0 \le x \le \theta\,, \\ 0 & , \text{otherwise.} \end{cases}$$

   Find the maximum likelihood estimator of θ and its mean squared error.

2018

GROUP A

1. Find all real solutions (x_1, x_2, x_3, λ) for the system of equations

$$x_2 - 3x_3 - x_1\lambda = 0,$$
$$x_1 - 3x_3 - x_2\lambda = 0,$$
$$x_1 + x_2 + x_3\lambda = 0.$$

2. Let $\{x_n\}_{n\geq 1}$ be a sequence defined by $x_1 = 1$ and

$$x_{n+1} = \left(x_n^3 + \frac{1}{n(n+1)(n+2)}\right)^{1/3}, \quad n \geq 1.$$

Show that $\{x_n\}_{n\geq 1}$ converges and find its limit.

3. Consider all permutations of the integers $1, 2, \ldots, 100$. In how many of these permutations will the 25th number be the minimum of the first 25 numbers and the 50th number be the minimum of the first 50 numbers?

GROUP B

4. An urn contains $r > 0$ red balls and $b > 0$ black balls. A ball is drawn at random from the urn, its colour noted, and returned to the urn. Further, $c > 0$ additional balls of the same colour are added to the urn. This process of drawing a ball and adding c balls of the same colour is continued. Define $X_i = 1$ if at the i-th draw the colour of the ball drawn is red, and 0 otherwise. Compute $\mathrm{E}(\sum_{i=1}^n X_i)$.

5. Suppose X_1 and X_2 are identically distributed random variables, not necessarily independent, taking values in $\{1, 2\}$. If $\mathrm{E}(X_1X_2) = 7/3$ and $\mathrm{E}(X_1) = 3/2$, obtain the joint distribution of (X_1, X_2).

6. A fair 6-sided die is rolled repeatedly until a 6 is obtained. Find the expected number of rolls conditioned on the event that none of the rolls yielded an odd number.

7. Suppose $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is a random sample from a bivariate normal distribution with $\mathrm{E}(X_i) = \mathrm{E}(Y_i) = 0$, $\mathrm{Var}(X_i) = \mathrm{Var}(Y_i) = 1$ and unknown $\mathrm{Corr}(X_i, Y_i) = \rho \in (-1, 1)$, for all $i = 1, \ldots, n$. Define $W_n = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$.

(a) Is W_n an unbiased estimator of ρ? Justify your answer.

(b) For large n, obtain an approximate level $(1 - \alpha)$ two-sided confidence interval for ρ, where $0 < \alpha < 1$.

8. Let $\{X_1, \ldots, X_n\}$ be an i.i.d. sample from $f(x : \theta)$, $\theta \in \{0, 1\}$, with

$$f(x:0) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad f(x:1) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Based on the above sample, obtain the most powerful test for testing $H_0 : \theta = 0$ against $H_1 : \theta = 1$, at level α, with $0 < \alpha < 1$. Find the critical region in terms of the quantiles of a standard distribution.

9. Suppose (y_i, x_i) satisfies the regression model,

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad \text{for } i = 1, \ldots, n,$$

where $\{x_i : 1 \leq i \leq n\}$ are fixed constants and $\{\epsilon_i : 1 \leq i \leq n\}$ are i.i.d. $N(0, \sigma^2)$ errors, where α, β and $\sigma^2 (> 0)$ are unknown parameters.

(a) Let $\widetilde{\alpha}$ denote the least squares estimate of α obtained assuming $\beta = 5$. Find the mean squared error (MSE) of $\widetilde{\alpha}$ in terms of the model parameters.

(b) Obtain the maximum likelihood estimator of this MSE.

PSB
2019

GROUP A

1. Let $f(x) = x^3 - 3x + k$, where k is a real number. For what values of k will $f(x)$ have three distinct real roots?

2. Let A and B be 4×4 matrices. Suppose that A has eigenvalues x_1, x_2, x_3, x_4 and B has eigenvalues $1/x_1, 1/x_2, 1/x_3, 1/x_4$, where each $x_i > 1$.

   (a) Prove that $A + B$ has at least one eigenvalue greater than 2.

   (b) Prove that $A - B$ has at least one eigenvalue greater than 0.

   (c) Give an example of A and B so that 1 is not an eigenvalue of AB.

3. Elections are to be scheduled on any seven days in April and May. In how many ways can the seven days be chosen such that elections are not scheduled on two consecutive days?

GROUP B

4. Let X and Y be independent and identically distributed random variables with mean $\mu > 0$ and taking values in $\{0, 1, 2, \dots\}$. Suppose, for all $m \geq 0$,

$$\mathrm{P}(X = k \mid X + Y = m) = \frac{1}{m+1}, \quad k = 0, 1, \dots, m.$$

   Find the distribution of X in terms of μ.

5. Suppose $X_1, X_2, \ldots, X_n$ are independent random variables such that

$$\mathrm{P}(X_i = 1) = p_i = 1 - \mathrm{P}(X_i = 0),$$

where $p_1, p_2, \ldots, p_n \in (0, 1)$ are all distinct and unknown. Consider $X = \sum_{i=1}^{n} X_i$ and another random variable Y which is distributed as Binomial$(n, \bar{p})$, where $\bar{p} = \frac{1}{n} \sum_{i=1}^{n} p_i$. Between X and Y, which is a better estimator of $\sum_{i=1}^{n} p_i$ in terms of their respective mean squared errors?

6. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from Uniform$(0, \theta)$ for some unknown $\theta > 0$. Let Y_n be the minimum of $X_1, X_2, \ldots, X_n$.

   (a) Suppose F_n is the cumulative distribution function (c.d.f.) of nY_n. Show that for any real x, $F_n(x)$ converges to $F(x)$, where F is the c.d.f. of an exponential distribution with mean θ.

   (b) Find $\lim_{n \to \infty} \mathrm{P}(n[Y_n] = k)$ for $k = 0, 1, 2, \ldots,$ where $[x]$ denotes the largest integer less than or equal to x.

7. Suppose an SRSWOR of size n has been drawn from a population labelled $1, 2, \ldots, N$, where the population size N is unknown.

   (a) Find the maximum likelihood estimator $\widehat{N}$ of N.

   (b) Find the probability mass function of $\widehat{N}$.

   (c) Show that $\frac{n+1}{n}\widehat{N} - 1$ is an unbiased estimator of N.

8. Suppose $\{(x_i, y_i, z_i) : i = 1, 2, \ldots, n\}$ is a set of trivariate observations on three variables: X, Y, and Z, where $z_i = 0$ for $i = 1, 2, \ldots, n-1$ and $z_n = 1$. Suppose the least squares linear regression equation of Y on X based on the first $n-1$ observations is

$$y = \widehat{\alpha}_0 + \widehat{\alpha}_1 x$$

and the least squares linear regression equation of Y on X and Z based on all n observations is

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 z.$$

Show that $\widehat{\alpha}_1 = \widehat{\beta}_1$.

9. Let Z be a random variable with probability density function

$$f(z) = \frac{1}{2} e^{-|z-\mu|}, z \in \mathbb{R}$$

with parameter $\mu \in \mathbb{R}$. Suppose we observe $X = \max(0, Z)$.

(a) Find the constant c such that the test that "rejects when $X > c$" has size 0.05 for the null hypothesis $H_0 : \mu = 0$.

(b) Find the power of this test against the alternative hypothesis $H_1 : \mu = 2$.

2020

GROUP A

1. Let $f(x) = x^2 - 2x + 2$. Let L_1 and L_2 be the tangents to its graph at $x = 0$ and $x = 2$ respectively. Find the area of the region enclosed by the graph of f and the two lines L_1 and L_2.

2. Find the number of 3×3 matrices A such that the entries of A belong to the set $\mathbb{Z}$ of all integers, and such that the trace of $A^t A$ is 6.
(A^t denotes the transpose of the matrix A).

3. Consider n independent and identically distributed positive random variables $X_1, X_2, \ldots, X_n$. Suppose S is a fixed subset of $\{1, 2, \ldots, n\}$ consisting of k distinct elements where $1 \le k < n$.

   (a) Compute
$$\mathbb{E}\left[\frac{\sum_{i \in S} X_i}{\sum_{i=1}^{n} X_i}\right].$$

   (b) Assume that X_i's have mean μ and variance σ^2, $0 < \sigma^2 < \infty$. If $j \notin S$, show that the correlation between $(\sum_{i \in S} X_i)X_j$ and $\sum_{i \in S} X_i$ lies between $-\frac{1}{\sqrt{k+1}}$ and $\frac{1}{\sqrt{k+1}}$.

GROUP B

4. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Let $S_n = X_1 + \cdots + X_n$. For each of the following statements, determine whether they are true or false. Give reasons in each case.

   (a) If $S_n \sim Exp$ with mean n, then each $X_i \sim Exp$ with mean 1.

   (b) If $S_n \sim Bin(nk, p)$, then each $X_i \sim Bin(k, p)$.

5. Let $U_1, U_2, \ldots, U_n$ be independent and identically distributed random variables each having a uniform distribution on $(0,1)$. Let

$$X = \min\{U_1, U_2, \ldots, U_n\}, \qquad Y = \max\{U_1, U_2, \ldots, U_n\}.$$

Evaluate $\mathbb{E}\left[X|Y=y\right]$ and $\mathbb{E}\left[Y|X=x\right]$.

6. Suppose individuals are classified into three categories C_1, C_2 and C_3. Let $p^2, (1-p)^2$ and $2p(1-p)$ be the respective population proportions, where $p \in (0,1)$. A random sample of N individuals is selected from the population and the category of each selected individual recorded. For $i = 1, 2, 3$, let X_i denote the number of individuals in the sample belonging to category C_i. Define $U = X_1 + \frac{X_3}{2}$.

   (a) Is U sufficient for p? Justify your answer.

   (b) Show that the mean squared error of $\frac{U}{N}$ is $\frac{p(1-p)}{2N}$.

7. Consider the following model:

$$y_i = \beta x_i + \varepsilon_i x_i, \ \ i = 1, 2, \ldots, n,$$

where $y_i, i = 1, 2, \ldots, n$ are observed; $x_i, i = 1, 2, \ldots, n$ are known positive constants and β is an unknown parameter. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent and identically distributed random variables having the probability density function

$$f(u) = \frac{1}{2\lambda} \exp\left(-\frac{|u|}{\lambda}\right), \ \ -\infty < u < \infty,$$

and λ is an unknown parameter.

   (a) Find the least squares estimator of β.

   (b) Find the maximum likelihood estimator of β.

8. Assume that $X_1, \ldots, X_n$ is a random sample from $N(\mu, 1)$, with $\mu \in \mathbb{R}$. We want to test $H_0 : \mu = 0$ against $H_1 : \mu = 1$. For a fixed integer $m \in \{1, \ldots, n\}$, the following statistics are defined:

$$\begin{aligned} T_1 &= (X_1 + \ldots + X_m)/m, \\ T_2 &= (X_2 + \ldots + X_{m+1})/m, \\ \vdots &= \qquad \vdots \\ T_{n-m+1} &= (X_{n-m+1} + \ldots + X_n)/m. \end{aligned}$$

Fix $\alpha \in (0, 1)$. Consider the test

$$\text{reject } H_0 \quad \text{if} \qquad \max\{T_i : 1 \leq i \leq n - m + 1\} > c_{m,\alpha}.$$

Find a choice of $c_{m,\alpha} \in \mathbb{R}$ in terms of the standard normal distribution function Φ that ensures that the size of the test is at most α.

9. A finite population has N units, with x_i being the value associated with the i^{th} unit, $i = 1, 2, \ldots, N$. Let $\bar{x}_N$ be the population mean. A statistician carries out the following experiment.

- Step 1: Draw a SRSWOR of size n ($< N$) from the population. Call this sample S_1 and denote the sample mean by $\bar{X}_n$.
- Step 2: Draw a SRSWR of size m from S_1. The x-values of the sampled units are denoted by $\{Y_1, \ldots, Y_m\}$.

An estimator of the population mean is defined as,

$$\widehat{T}_m = \frac{1}{m} \sum_{i=1}^{m} Y_i.$$

(a) Show that $\widehat{T}_m$ is an unbiased estimator of the population mean.

(b) Which of the following has lower variance: $\widehat{T}_m$ or $\bar{X}_n$?

2021

Group A

1. Define $f : \mathbb{R} \to \mathbb{R}$ as

$$f(x) = \begin{cases} \cos(2x) & \text{if } x \text{ is rational,} \\ \sin^2(x) & \text{if } x \text{ is irrational.} \end{cases}$$

Find all the real numbers where

   (a) f is continuous,

   (b) f is differentiable.

2. Consider the matrix

$$\mathbf{P} = \begin{pmatrix} 5 & 3 \\ 3 & 2 \\ 8 & 5 \end{pmatrix}_{3\times 2}.$$

Find a matrix $\mathbf{G}$ such that $\mathbf{AGA} = \mathbf{A}$ where $\mathbf{A} = \mathbf{PP}^{\mathsf{T}}$.

3. Consider the set of all five digit integers formed by permuting the digits $1, 2, 4, 6$ and 7. Let X denote a randomly chosen integer from this set and let Y denote the position, from the right, of the digit 4 in the randomly chosen integer. For example, if the integer chosen is 12647, then Y is 2 and for 41276, Y is 5.

   (a) Find $E(X)$ and $E(Y)$.

   (b) Find $E[X|Y = y]$ for all possible values of y.

   (c) Show that X and Y are uncorrelated but not independent.

Group B

4. Let U_1, U_2, U_3 be i.i.d. random variables which are uniformly distributed on $(0,1)$. Let $X = \min(U_1, U_2)$ and $Y = \max(U_2, U_3)$.

   (a) Find $P(X \le x, Y \le y)$ for all $x, y \in \mathbb{R}$.

   (b) Find $P(X = Y)$.

   (c) Find $E[XI_{\{X=Y\}}]$ where I_A is the indicator function of A.

5. Consider a game with six states $1, 2, 3, 4, 5, 6$. Initially a player starts either in state 1 or in state 6. At each step the player jumps from one state to another as per the following rules.

   A perfectly balanced die is tossed at each step.

   (i) When the player is in state 1 or 6: If the roll of the die results in k then the player moves to state k, for $k = 1, \ldots, 6$.

   (ii) When the player is in state 2 or 3: If the roll of the die results in $1, 2$ or 3 then the player moves to state 4. Otherwise the player moves to state 5.

   (iii) When the player is in state 4 or 5: If the roll of the die results in $4, 5$ or 6 then the player moves to state 2. Otherwise the player moves to state 3.

   The player wins when s/he visits 2 more states, besides the starting one.

   (a) Calculate the probability that the player will eventually move out of states 1 and 6.

   (b) Calculate the expected time the player will remain within states 1 and 6.

   (c) Calculate the expected time for a player to win, i.e., to visit 2 more states, besides the starting one.

6. Let $X_1, \ldots, X_n$ be i.i.d. random variables which are uniformly distributed on $(\theta, 2\theta)$, $\theta > 0$.

   (a) Show that $\frac{X_{(n)}}{2}$ is the maximum likelihood estimator (MLE) of θ where $X_{(n)} = \max(X_1, \ldots, X_n)$.

   (b) Find an unbiased estimator for θ based on the MLE.

   (c) Given any $\epsilon > 0$, show that

$$\lim_{n\to\infty} P\left(\left|\frac{X_{(n)}}{2} - \theta\right| > \epsilon\right) = 0.$$

7. There are two urns each contains N balls numbered from 1 to N. From each urn a sample of size n is selected without replacement. Denote the set of numbers appearing in the first and second samples by $s_1 = \{i_1, \ldots, i_n\}$ and $s_2 = \{j_1, \ldots, j_n\}$ respectively. Let

$$X = |s_1 \cap s_2| = \text{number of common elements in } s_1 \text{ and } s_2.$$

   (a) Find the probability distribution of X.

   (b) Suppose $n = 6$ and the observed value of X is 4. Obtain a method of moments estimate of N.

8. Let X_1, X_2, X_3 be i.i.d. random variables from $N(\mu, \sigma^2)$. Let

$$\bar{X} = \frac{1}{3}\sum_{i=1}^{3} X_i, \quad T_1 = \sum_{i=1}^{3} X_i^{\,2}, \quad T_2 = \frac{1}{3}\sum_{i=1}^{3}(X_i - \bar{X})^2.$$

   (a) Compute $E[T_1|\bar{X}]$ and $E[T_1|T_2]$

   (b) Obtain the exact critical region of a level α $(0 < \alpha < 1)$ test for $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ that rejects H_0 if and only if $\dfrac{\bar{X}^2}{T_1}$ is sufficiently large.

9. Consider a linear regression model:

$$y_i = \alpha + \beta x_i + e_i, \;\; i = 1, 2, \ldots, n$$

where x_i's are fixed and e_i's are i.i.d. random errors with mean 0 and variance σ^2.

Define two estimators of β as follows

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \quad \text{and} \quad \widehat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

(a) Obtain an unbiased estimator of β as a linear combination of $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

(b) Find mean squared errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Which, between $\widehat{\beta}_1$ and $\widehat{\beta}_2$, has lower mean squared error?

2022

GROUP A

1. Consider a sequence $\{r_k\}_{k\geq 1}$ of rational numbers lying in the interval $(0,1)$. Further, assume that r_k converges to an irrational number as $k \to \infty$. Suppose $r_k = \frac{m_k}{n_k}$, for all $k \geq 1$, where m_k and n_k are positive integers with no common divisors. Show that the set of integers $\{n_k : k \geq 1\}$ is not bounded.

2. Consider a 4×4 real matrix A which has positive trace and negative determinant.

   (a) Show that A must have at least two real eigenvalues.

   (b) Show that A can have non-real eigenvalues.

3. Let P be a regular polygon with 24 sides. Consider all the triangles whose vertices are also vertices of P. Find the number of such triangles that are neither isosceles nor equilateral.

GROUP B

4. Suppose U and V are independent and identically distributed random variables following a binomial distribution with parameters n and $\frac{1}{2}$. Let the random variable T denote the number of distinct real roots of the quadratic equation

$$x^2 + 2Ux + V^2 = 0.$$

   Find $\mathrm{E}(T)$ and $\mathrm{Var}(T)$.

5. Suppose $r \geq 1$ distinct books are distributed at random among $n \geq 3$ children.

   (a) For each $j \in \{0, 1, 2, \ldots, r\}$, compute the probability that the first child gets exactly j books.

   (b) Let X be the number of children who do not get any book, and Y be the number of children who get exactly one book. Show that

$$\mathrm{Cov}(X, Y) = \frac{r(n-1)(n-2)^{r-1}}{n^{r-1}} - \frac{r(n-1)^{2r-1}}{n^{2r-2}}.$$

6. Consider two random variables (X, Y) distributed as bivariate normal with parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Based on a random sample from this bivariate distribution, the fitted least squares regression line of Y on X and that of X on Y were as follows:

$$Y = 22 - 3X$$

$$X = 5.84 - 0.12Y$$

   (a) Compute the maximum likelihood estimates of the following parametric functions:

   (i) $\min\{\mu_1, \mu_2\}$ (ii) $\dfrac{\sigma_1}{\sigma_2}$ (iii) ρ

   You are not required to derive the expressions for the maximum likelihood estimators of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ.

   (b) If a new observation (5, 5) is included in the above set of observations, determine whether each of the maximum likelihood estimates obtained in (a) will increase, decrease or remain unchanged.

7. Suppose N students arriving at a college are all equally likely to have a particular disease with an unknown probability p. The disease status (affected / not affected) of all students are independent. Blood samples are collected from all N students. In order to estimate p, two strategies are proposed.

Strategy 1 Test all samples separately to obtain the status for all N students.

Strategy 2 Randomly partition the students into m disjoint groups, each comprising $K = N/m$ students (with $K \geq 2$ being an integer). For each group, pool (mix) the blood samples from all K students within the group and test the pooled sample. If the pooled sample tests positive, then at least one student within that group is affected. If the pooled sample tests negative, all students within that group are unaffected.

(a) Based on the data obtained from Strategy 2, find a real-valued sufficient statistic for p and the maximum likelihood estimator of p.

(b) If a group tests positive, then all students within that group are further tested individually. Suppose that each test (for individual sample or pooled sample) has equal cost. Then, which of the two strategies would you prefer to identify all students affected with the disease when the underlying $p = 0.5$, $N = 200$, and $m = 20$?

8. Based on historical data, a positive random variable X arising from an unknown distribution is believed to have the chi-square distribution with 1 degree of freedom. A new theory suggests that it may be better to model $\sqrt{X}$ as exponentially distributed with mean λ, where λ is such that $\mathrm{E}(X)$ for this model is the same as that for the earlier model.

   (a) Compute λ.

   (b) Suppose $X_1, X_2, \ldots, X_n$ are independent observations from this unknown distribution. For $\alpha \in (0, 1)$, consider the most powerful level α test for the null hypothesis that the earlier model is correct against the alternative that the new model is correct. Show that the rejection region of this test is

$$\left\{ (x_1, x_2, \ldots, x_n) \,:\, \sum_i x_i - 2\sqrt{2} \sum_i \sqrt{x_i} \,>\, c \right\}$$

   where c is a constant that depends on α.

9. The following ANOVA table for three factors A, B, and C was obtained (under a suitable model) from some data, but several values were illegible and marked '$*$'. It is known that A had two levels and there were 24 observations in all.

Source	df	Sum of squares	Mean Square	F ratio
A	$*$	$*$	$*$	$*$
B	$*$	$*$	$*$	0.62
C	$*$	$*$	25.4	$*$
AB	2	$*$	5.4	0.54
BC	2	$*$	4.2	$*$
Error	$*$	$*$	$*$	
Total	$*$	257.4		

(a) Fill in the missing values, providing suitable justification.

(b) Suppose that a student has access only to tables for t distributions. Can she test the hypothesis of equality of effects of levels of A and that of equality of effects of levels of B, both at 5% level of significance? Justify your answer.