Test Code: RSI/RSII (Short Answer Type) 2008

Junior Research Fellowship for Research Course in Statistics

The candidates for research course in Statistics will have to take two short-answer type tests – RSI and RSII. Each test is of two-hour duration. Test RSI will have about 10 questions of equal value, set from different topics in Mathematics at B. Sc (Pass) level and Statistics at B.Sc. (Honours) level. Test RSII will have roughly 8 questions on topics in Statistics at Master's level.

Syllabus for RSI and RSII

Mathematics

Functions and relations. Matrices - determinants, eigenvalues and eigenvectors, solution of linear equations, and quadratic forms.
Calculus and Analysis - sequences, series and their convergence and divergence; limits, continuity of functions of one or more variables, differentiation, applications, maxima and minima. Integration, definite integrals, areas using integrals, ordinary linear differential equations.

Statistics

(a)*Probability*: Basic concepts, elementary set theory and sample space, conditional probability and Bayes theorem. Standard univariate and multivariate distributions. Transformations of variables. Moment generating functions, characteristic functions, weak and strong laws of large numbers, convergence in distribution and central limit theorem. Markov chains.
(b) *Inference*: Sufficiency, minimum variance unbiased estimation, Bayes estimates, maximum likelihood and other common methods of estimation. Optimum tests for simple and composite hypotheses. Elements of sequential and non-parametric tests. Analysis of discrete data - contingency chi-square.
(c) *Multivariate Analysis*: Standard sampling distributions. Order statistics with applications. Regression, partial and multiple correlations. Basic properties of multivariate normal distribution, Wishart distribution, Hotelling's T-square and related tests.
(d) *Linear Models and Design of Experiments*: Inference in linear models. Standard orthogonal and non-orthogonal designs. Inter and intra-block analysis of general block designs. Factorial experiments. Response surface designs. Variance components estimation in one and two-way ANOVA.
(e) *Sample Surveys*: Simple random sampling, Systematic sampling, PPS sampling, Stratified sampling. Ratio and regression methods of estimation. Non-sampling errors, Non-response.

Sample Questions : RSI

1. Let A and B be two given subsets of Ω. For each subset C of Ω define

$$I_c(\omega) = \begin{cases} 1 & \text{if} \quad \omega \ \in C \\ 0 & \text{if} \quad \omega \ \notin C. \end{cases}$$

   Does there exists a subset C of Ω such that $\mid I_A(\omega) - I_B(\omega) \mid = I_c(\omega)$ for each $\omega \in \Omega$? Justify your answer.

2. (a) Evaluate the determinant

$$\begin{vmatrix} 1+x_1y_1 & 1+x_1y_2 & \dots & 1+x_1y_n \\ 1+x_2y_1 & 1+x_2y_2 & \dots & 1+x_2y_n \\ \vdots & \vdots & & \vdots \\ 1+x_ny_1 & 1+x_ny_2 & \dots & 1+x_ny_n \end{vmatrix}.$$

   (b) Find the inverse of the matrix $\boldsymbol{A} + \boldsymbol{\alpha\alpha}'$, where $\boldsymbol{A}$ is a diagonal matrix $\text{diag}(a_1, a_2 \cdots, a_k)$ and $\boldsymbol{\alpha}' = (\frac{1}{\alpha_1}, \cdots, \frac{1}{\alpha_k})$, $\alpha_i > 0$ for $i = 1, \cdots, k$.

3. Consider a symmetric matrix $\mathbf{A}$ such that neither $\mathbf{A}$ nor $-\mathbf{A}$ is positive definite. Show that there exists a vector $x \neq 0$ such that $x'\mathbf{A}x = 0$.

4. Let f, g and h be defined on $[0, 1]$ as follows:

$$\begin{aligned} f(x) &= g(x) = h(x) = 0 \text{ whenever } x \text{ is irrational;} \\ f(x) &= 1 \text{ and } g(x) = x \text{ whenever } x \text{ is rational;} \\ h(x) &= \frac{1}{n} \text{ if } x \text{ is the rational number } m/n \text{ (in lowest terms);} \\ h(0) &= 1. \end{aligned}$$

   Prove that

   (a) f is not continuous anywhere in $[0, 1]$,

   (b) g is continuous only at $x = 0$, and

   (c) h is continuous only at the irrational points in $[0, 1]$.

5. (a) For any two real numbers a and b, show that:

$$|a^3 \ + b^3| \geq \frac{1}{4}|a + b|^3.$$

   (b) For a sequence of bivariate random variables (X_n, Y_n), suppose ${X_n}^3 + {Y_n}^3 \rightarrow 0$ in distribution. Show that $X_n \ + \ Y_n \rightarrow 0$ in distribution.

6. Let $f : \Re \to \Re$ be differentiable at α. Further, $f(\alpha) = 0$ and let $g(x) = |f(x)|$. Show that g is differentiable at α if and only if $f'(\alpha) = 0$.

7. Find the maximum of $\int\int_{\triangle} dxdy$ as a function of m for $0 < m < 1$ where

$$\triangle = \{(x, y) : \frac{x^2}{m} + \frac{y^2}{1-m} \leq 1\}$$

8. (a) Find $\int_{-2}^{4} [x]dx$, where $[x]$ is the largest integer less than or equal to x.
   (b) If $f(x), x \geq 0$ is a differentiable function such that $f'(x) \to \infty$ as $x \to \infty$, then show that $f(x) \to \infty$ as $x \to \infty$.

9. (a) Find the limit

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{n}{n^2 + k^2}$$

   (b) Let $f(t)$ be a continuous function on $(-\infty, \infty)$ and let

$$F(x) = \int_0^x (x - t) f(t) dt. \quad \text{Find } F'(x).$$

10. (a) Find $\max(xyz)$ subject to $x^2 + 2y^2 + 9z^2 = 6$.
   (b) Let $f(x) = \sum_{k=0}^{n} a_k x^k$, where a_k's satisfy $\sum_{k=0}^{n} \frac{a_k}{k+1} = 0$.
   Show that there exists a root of $f(x) = 0$ in the interval $(0, 1)$.

11. One of the sequences of letters $AAAA, BBBB, CCCC$ is transmitted over a communication channel with respective probabilities p_1, p_2, p_3, where $(p_1 + p_2 + p_3 = 1)$. The probability that each transmitted letter will be correctly understood is α and the probabilities that the letter will be confused with two other letters are $\frac{1}{2}(1 - \alpha)$ and $\frac{1}{2}(1 - \alpha)$. It is assumed that the letters are distorted independently. Find the probability that $AAAA$ was transmitted if $ABCA$ was received.

12. In a sequence of independent tosses of a fair coin, a person is to receive Rs.2 for a head and Re.1 for a tail.
   (a) For a positive integer n, find the probability that the person receives exactly Rs. n at some stage.
   (b) What is the limit of this probability as $n \to \infty$?

13. Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with $P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$. Let $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Let $N = \min\{n \geq 1, \text{ such that } S_n = 0\}$.

(a) Show that $P(N = 2k + 1) = 0$ for every integer $k \geq 0$.

(b) Find $P(N = 2k,\ S_k = k)$ for every integer $k \geq 1$.

14. (a) Show that if X_1 and X_2 are identically distributed (not necessarily independent) random variables then, for $t > 0$,

$$P[|X_1 - X_2| > t] \leq 2P[|X_1| > \frac{t}{2}].$$

(b) If $X_1, X_2, \ldots$ are independent and identically distributed random variables and $V_j = (X_1^2 + X_2^2 + \ldots + X_j^2)^{\frac{1}{2}}$, then show that

$$E(\frac{X_k^2}{V_j^2}) = \frac{1}{j},\ \forall\ 1 \leq k \leq j.$$

15. Let $X_1, \cdots, X_n, \cdots,$ be a sequence of independent random variables with $E(X_r) = 0$, $\mathrm{Var}(X_r) = \sqrt{r}$ for $r = 1, \cdots, n, \cdots$. Prove that $\frac{1}{n}(X_1 + \cdots + X_n)$ converges almost surely and find the limit.

16. Suppose die A has 4 red faces and 2 green faces while die B has 2 red faces and 4 green faces. Assume that both the dice are unbiased. An experiment is started with the toss of an unbiased coin. If the toss results in a Head, then die A is rolled repeatedly while if the toss of the coin results in a Tail, then die B is rolled repeatedly. For $k = 1, 2, 3, \cdots$, define

$$X_k = \begin{cases} 1 & \text{if the } k\text{th roll of the die results in a red face} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the probability mass function of X_k.

(b) Calculate $\rho(X_1, X_7)$.

17. A model that is often used for the waiting time X to failure of an item is given by the probability mass function

$$p_X(k|\theta) = \theta^{k-1}(1-\theta),\ k = 1, 2, \ldots,\ 0 < \theta < 1.$$

Suppose that we only record the time to failure if the failure occurs on or before r, and otherwise just note that the item has lived at least $(r+1)$ periods. Let Y denote this censored waiting time.

(a) Write down the probability mass function of Y.

(b) If $Y_1, Y_2, \ldots, Y_n$ is a random sample from the censored waiting time, write down the likelihood function and find the MLE of θ.

18. Let X be exponentially distributed with mean 1 and let U be a $U[0, 1]$ random variable, independent of X. Define

$$I = 1 \text{ if } U \leq e^{-X} \quad \text{and } I = 0 \quad \text{otherwise .}$$

Show that the conditional distribution of X given $\{I = 1\}$ is exponential with mean 0.5.

19. Let X be an exponential $r.v.$ with mean $\frac{1}{2}$. Let Y be the largest integer less than or equal to X. Find the probability distribution of Y.

20. Let X_1, X_2, X_3 be independent standard normal variables. Find the distribution of
$$U = \frac{X_1 + X_2 + X_3}{X_2 - 2X_3 + X_1}.$$

21. Let Y have a chi-squared distribution with $k(\geq 3)$ degrees of freedom. Express $E(1/Y)$ in terms of k.

22. Let X follow an $N_p(\theta, \Sigma)$ distribution, where Σ is a known positive definite matrix and θ is either θ_1 or θ_2, θ_i's being known. Show that $(\theta_2 - \theta_1)' \sum^{-1} X$ is a sufficient statistic.

23. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population with probability density
$$p(x|\theta) \;=\; \frac{x}{\theta^2} \exp(-\frac{x^2}{2\theta^2}),\ x > 0,\ \theta > 0.$$
Find the Fisher information number $I(\theta)$ and give a lower bound to the variance of an unbiased estimator of θ.

24. Show that $X_1 + 2X_2$ is not sufficient for μ where X_1 and X_2 are two random observations from $N(\mu, 1)$ distribution.

25. The number of accidents X per year in a manufacturing plant may be modelled as a Poisson random variable with mean λ. Assume that accidents in successive years are independent random variables and suppose that you have n observations.
    (a) How will you find the minimum variance unbiased estimator for the probability that in a year the plant has at most one accident ?
    (b) Suppose that you wish to estimate λ. Suggest two unbiased estimators of λ and hence find the UMVUE of λ.

26. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a negative exponential distribution with mean θ. The experiment is terminated after the first $r(r \leq n)$ smallest observations have been noted.

    Write down the likelihood for θ based on these censored observations. Find the *mle* of θ. Obtain the UMP test for testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$ at level α.

27. Consider the following fixed effects linear model:
$$\begin{aligned}
Y_1 &= \theta_1 + \theta_2 + \theta_4 + \epsilon_1 \\
Y_2 &= \theta_1 + \theta_3 + 2\theta_4 + \epsilon_2 \\
Y_3 &= \theta_1 + \theta_2 + \theta_4 + \epsilon_3 \\
Y_4 &= -\theta_2 + \theta_3 + \theta_4 + \epsilon_4 \\
\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4) &\sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_4)
\end{aligned}$$

(a) Find the full set of error functions.

(b) Hence or otherwise, find an unbiased estimator of σ^2 with maximum possible degrees of freedom.

28. An experimenter wanted to use a Latin Square Design (LSD) but instead, used the following row-column design

$$\begin{array}{ccc} A & B & C \\ B & A & C \\ C & A & B \end{array}$$

Are all treatment contrasts estimable ? Give reasons for your answer.

29. An unbiased coin is to be used to select a Probability Proportional to Size With Replacement (PPSWR) sample in 2 draws from the following population of 3 units, where X is the size measure.

i	1	2	3
X_i	2	4	1

Consider the following procedure.

a) Toss the coin thrice independently.

b) If the outcome is $\{HHH\}$ or $\{HHT\}$, select the first unit.

c) If the outcome is $\{HTH\}$, $\{HTT\}$, $\{THH\}$ or $\{THT\}$, then select the second unit.

d) If the outcome is $\{TTH\}$, then select the third unit.

e) If the outcome is $\{TTT\}$, do not select any of the units, go back to (a) above.

f) Continue till a unit is selected.

Show that the above procedure is a PPS method. If W = number of tosses required to select a unit, find $E(W)$.

30. Show that under PPSWOR of size 2 or 3, inclusion probability of ith unit exceeds that of jth unit, whenever ith unit has a larger size measure than unit jth unit.

Sample Questions : RSII

1. Express the following as the probability of an event and evaluate without integrating by parts:
$$\int_{-\infty}^{\infty} \Phi(x) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx,$$
where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

2. Suppose that $X_1, X_2, X_3, \cdots,$ are i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$, $EX_1^4 < \infty$. Show that
$$n^{-1/2}[\sum_1^n (X_i - \overline{X}_n)^2 - n]$$
converges in law to a normal distribution with zero mean, as $n \to \infty$. Here $\overline{X}_n = \sum_1^n X_i/n$.

3. Let $X_1, X_2, \cdots, X_n$ be i.i.d. observations from a continuous distribution and $R_1, \cdots, R_n$ be the ranks of the observations. Find $Cov(R_{n-1}, R_n)$.

4. Let $\{X_n\}$ be a sequence of independent random variables. Show that for each real number $\alpha, P(X_n \to \alpha) = 0$ or 1 by explicitly applying the Borel-Cantelli lemmas.

5. Suppose X and Y are two positive random numbers.
Define $U = max(X, Y)$ and $V = min(X, Y)$. Show that:
   (a) $E(X)E(Y) \geq E(U)E(V)$.
   (b) $Var(X) + Var(Y) \geq Var(U) + Var(V)$.

6. A company desires to operate S identical machines. These machines are subject to failure according to a given probability law. To replace these failed machines the company orders new machines at the beginning of each week to make up the total S. It takes one week for each new order to be delivered. Let X_n be the number of machines in working order at the beginning of the nth week, and let Y_n denote the number of machines that fail during the nth week.
   (a) Establish the recursive formula $X_{n+1} = S - Y_n$, and show that $X_n,\ n \geq 1$ constitutes a Markov Chain.
   (b) Suppose that the failure rate is uniform i.e.,
$$P[Y_n = j \mid X_n = i] = \frac{1}{i+1}, \quad j = 0, 1, \cdots, i.$$
   Find the transition matrix of the Chain, its stationary distribution, and expected number of machines in operation in the steady state.

7. Let N_n be the number of heads observed in the first n tosses of a fair coin, with $N_0 = 0$. Find

$$\lim_{n\to\infty} P[N_n \text{ is a multiple of } 3]$$

by formulating the problem in relation to a Markov Chain with state space $\{0, 1, 2\}$ consisting of the remainders obtained from non-negative integers after division by 3.

8. (a) In a forest there are 50 tigers and an unknown number L of lions. Assume that the $50 + L$ animals randomly move in the forest. A naturalist sights 5 different lions and 15 different tigers in the course of a trip in the forest. Estimate, stating assumptions and with theoretical support, the number of lions in the forest.

   (b) A bank decided to examine a sample of vouchers before conducting a thorough audit. From a very large number of accumulated vouchers, samples were drawn at random. The first defective voucher was obtained in the 23rd draw, and the second defective voucher in the 57th draw. Estimate, giving reasons, the proportion of vouchers that are defective.

9. Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be an independent sequence of bivariate normal random variables such that

$$E(X_i) = E(Y_i) = \mu_i; \quad Var(X_i) = Var(Y_i) = \sigma^2; \quad Cov(X_i, Y_i) = \sigma^2/2.$$

Find the maximum likelihood estimator of σ^2 based on the above data. Is this estimator consistent? Justify your answer.

10. Let $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ be the ordered random variables of a sample of size n from the rectangular $(0, \theta)$ distribution with θ unknown, $0 < \theta < \infty$. By a careless mistake the observations $Y_{(k+1)}, \cdots, Y_{(n)}$ were recorded incorrectly and so they were discarded subsequently (Here $1 \leq k < n$).

   (a) Show that the conditional distribution of $Y_{(1)}, \cdots, Y_{(k-1)}$ given $Y_{(k)}$ is independent of θ.

   (b) Hence, or otherwise, obtain the maximum likelihood estimator of θ and show that it is a function of $Y_{(k)}$.

   (c) If $\frac{k}{n} \to p$ as $n \to \infty$, for some $0 < p < 1$, what can you say about the asymptotic distribution of the maximum likelihood estimator of θ ?

11. Let $X \sim N_6(\mu, \Sigma)$, where $\mu = (\mu_1, \cdots, \mu_6)^T$, and Σ is unknown. Obtain an exact test for testing $H_0 : \mu_l = l\eta$, $l = 1, 2, \cdots, 6$, η being unknown, against H_1 : not H_0. Obtain the cut-off point of this test.

12. The value of Y is estimated from $X = x_0$ and the linear regression of Y on X. Let this estimated value of Y be y_0. Then the value of X corresponding to $Y = y_0$ is estimated from the linear regression of X on Y. Let this estimated value of X be x_0^*. Compare x_0 and x_0^*. Interpret your answer.

13. Let X be a random variable having a density $\frac{1}{\theta}e^{-x/\theta}, x, \theta > 0$. Consider $H_0 : \theta = 1$ vs. $H_1 : \theta = 2$. Let ω_1 and ω_2 be two critical regions given by $\omega_1 : \sum_1^n X_i \geq C_1$ and $\omega_2 :$ (number of X_i's $\geq 2) \geq C_2$.

    (a) Determine approximately the values of C_1 and C_2 for large n so that both tests are of size α.

    (b) Show that the powers of both tests tend to 1 as $n \to \infty$.

    (c) Which test would require more sample size to achieve the same power ? Justify your answer.

14. Let $X_1, X_2, \ldots, X_n$ be i.i.d observations with a common exponential distribution with mean μ. Show that there is no uniformly most powerful test for testing $H_0 : \mu = 1$ against $H_A : \mu \neq 1$ at a given level $0 < \alpha < 1$ but there exists a uniformly most powerful unbiased test and derive that test.

15. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ have a multivariate normal distribution with unknown mean vector $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$ and unknown variance covariance matrix Σ which is nonsingular. Let $\mathbf{X_1}, \cdots, \mathbf{X_n}$ be a random sample of size n from the population. Develop a test for testing the hypothesis $H_0 : \mu_1 + 2\mu_2 = \mu_2 + 2\mu_3 = \mu_3 + 2\mu_4$. State the distribution of the test statistic.

16. Suppose that

$$Y_i = \rho x_i + \epsilon_i, \; i = 1, 2, \ldots, n$$

    where ρ is an unknown parameter and $x_i = i/n$. Assume that ϵ_is are independent random variables satisfying $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2 x_i$ with an unknown $\sigma^2 > 0$.

    (a) Obtain the least squares estimator of ρ.

    (b) Obtain the BLUE of ρ.

    (c) Obtain the variances of the two estimators and the relative efficiency of the BLUE with respect to the least squares estimator.

17. Consider the following 2^4 factorial design with factors A, B, C and D in the usual order.

    Block 1 : (0,0,0,0), (0,1,0,1), (1,0,1,0), (1,1,1,1), (1,1,0,1), (0,0,1,0).

    Block 2 : (0,0,1,1), (0,1,1,0), (1,0,0,1), (1,1,0,0), (1,1,1,0), (0,0,0,1).

    Block 3 : (0,1,0,0), (0,1,1,1), (1,0,0,0), (1,0,1,1).

    Examine whether the main effect of A is estimable.

18. Let $(5, 0.023), (3, 0.189), (5, 0.023)$ and $(2, 0.272)$ denote the (i, p_i)-values for the labels i selected in 4 draws with replacement from a population of size N with selection probabilities p_i on each draw. Give (a) an unbiased estimate of N and (b) indicate, without actual evaluation, how a suitable confidence interval for N could be set up.

19. A sample $s^{(1)}$ of n units is selected from a population of N units using SRSWOR and the values of a variable y are ascertained for those n_1 units among the n units of $s^{(1)}$ who responded. Later, a further sub-sample $s^{(2)}$ of m units is selected using SRSWOR out of the $(n - n_1)$ units of $s^{(1)}$ which did not respond. Assuming that the y-values of all the m units of $s^{(2)}$ could be obtained, find the following :

   (a) an unbiased estimator of the population mean $\overline{Y}$ on the basis of available y-values.

   (b) an expression for the variance of the proposed estimator.

20. Let the size of a population be $N = 3$ and the sample size $n = 2$. Let $s_1 = (1, 2)$, $s_2 = (1, 3)$, and $s_3 = (2, 3)$ denote the three possible samples. Under simple random sampling, you have $p(s_i) = 1/3$, $i = 1, 2, 3$. Define the estimator t by

$$t = \begin{cases} t_1 = \frac{y_1+y_2}{2} & \text{if } s_1 \text{ occurs} \\ t_2 = \frac{y_1}{2} + \frac{2y_3}{3} & \text{if } s_2 \text{ occurs} \\ t_3 = \frac{y_2}{2} + \frac{y_3}{3} & \text{if } s_3 \text{ occurs.} \end{cases}$$

   Show that t is unbiased for $\bar{Y}$ and that there exist values (Y_1, Y_2, Y_3) for which $V(t) < V(\bar{y})$, where $\bar{y}$ denotes the conventional sample mean. What does this example imply?