Test Code: STA/STB (Short Answer Type) 2013

Junior Research Fellowship for Research Course in Statistics

The candidates for the research course in Statistics will have to take two short-answer type tests – STA and STB. Each test is of two-hour duration. Test STA will have about 10 questions of equal value, set from selected topics in Mathematics and Statistics at the undergraduate level. Test STB will have roughly 8 questions of equal value, on topics in Statistics at Master's level.

Syllabus for STA and STB

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) *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 : STA

1. Let f be a real-valued, bounded, twice differentiable function defined on $(0, \infty)$ with $f' \geq 0$ and $f'' \leq 0$. Show that

$$\lim_{x \to \infty} f'(x) = 0.$$

2. Let $a_1, ..., a_n$ be any n positive real numbers. Show that

$$\lim_{t \to 0+} \left[\frac{1}{n} \sum_{i=1}^{n} a_i^t \right]^{1/t}$$

is the geometric mean of $a_1, ..., a_n$.

3. Find the set of all positive integers n for which there is a real matrix A of dimension $n \times n$ such that $A^{-1} = -A$.

4. If $E(X^2) = 1,\ E(X^4) = 3,\ E(X^6) = 9$, then find all possible distributions of X.

5. Suppose that $X_1, ..., X_n$ are jointly distributed random variables with joint cdf $H(x_1, ..., x_n)$ and marginal cdf's $F_i(x_i)$, for $i = 1, ..., n$. Let $a_1, ..., a_n$ and $b_1, ..., b_n$ be $2n$ real numbers such that for all i we have $a_i \leq b_i$. Then prove that

$$H(b_1, ..., b_n) - H(a_1, ..., a_n) \leq \sum_{i=1}^{n} \big(F_i(b_i) - F_i(a_i)\big).$$

6. Let $Z_1, ..., Z_5$ be five independent $N(\mu, 1)$ observations. Further, let X be the sum of the first four of them and Y be the sum of last three of them. Then find the best linear unbiased estimator of μ based on X, Y.

7. Two units are selected without replacement by probability proportional to size from the following population of 3 units.

Unit	1	2	3
size	10	20	30

   (a) List the possible samples and their respective selection probabilities.

   (b) If only one of the three is selected with probability proportional to size of the unit for NOT inclusion in the sample, would the new list of selection probabilities of the samples match the corresponding probabilities in the first part? Justify your answer by computing relevant probabilities.

8. Let $X_1, ..., X_n$ be iid with both mean and variance being equal to 1. Find the asymptotic distribution of
$$\sqrt{n}(e^{\bar{X}} - e),$$
where $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.

9. Let $X_1, ..., X_n$ be iid with pdf $f(x|\theta)$. Also, let
$$f(x|0) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise,} \end{cases}$$
and
$$f(x|1) = \begin{cases} \frac{1}{2\sqrt{x}}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$
Consider a prior on θ that takes values 0 and 1 with equal probabilities. Find the Bayes estimator (posterior mean) of θ.

10. Suppose that we have a (possibly biased) coin and an urn having $w(> 0)$ white balls and $b(> 0)$ black balls. A trial is defined as follows:

    Trial: Toss the coin once. Add a white ball to the urn if the outcome is 'Head'; otherwise add a black ball. Then draw a ball at random from the urn having $w + b + 1$ balls, and record the colour of the ball drawn.

    Let X be the number of times a white ball is recorded in n independent trials, as described above. Find the maximum likelihood estimator of the probability of 'Head'.

Sample Questions : STB

1. Suppose that X is a nonnegative random variable with a decreasing density function. Let $E(X) = \mu$ and t both be positive real numbers. Further, let U be a $Unif(0, 2t)$ random variable. For $x \geq 0$, define $\overline{F}(x) = P(X > x)$. Then show that

   (a) $E(\overline{F}(U)) \geq \overline{F}(E(U))$, and

   (b) $\overline{F}(t) \leq \frac{\mu}{2t}$.

2. Let $\{X_n\}$ be a sequence of real-valued random variables (not necessarily independent) with $E(X_n) = 0$ and $Var(X_n) = \sigma_n^2$. Let $\sum_{n=1}^{\infty} \sigma_n < \infty$, and $Y_k = X_1 + \cdots + X_k$. Whatever be the joint distribution of the X_n's, show that

   (a) $\sum_{n=1}^{\infty} E|X_n| < \infty$, and

   (b) there exists a random variable U such that $Y_k \to U$ with probability 1.

3. Let X_1 and X_2 be two independent observations from a distribution F with $Var(X_i) = \sigma^2 > 0$ and $0 < Var(X_i^2) < \infty$. Prove that $T_1 = X_1^2 - X_1 X_2$ is unbiased for σ^2. Find another unbiased estimator for σ^2 having variance *strictly* less than that of T_1.

4. Suppose that $\{\mathbf{X}_{ij} : i = 1, ..., m;\ j = 1, ..., n\}$ are independent p-dimensional random vectors, and for each $i = 1, ..., m$, $\mathbf{X}_{ij}$ follows $N(\mu_i, \Sigma)$ distribution for $j = 1, ..., n$. Let

$$P = \sum_i \sum_j (\mathbf{X}_{ij} - \overline{\mathbf{X}}_{i\bullet})(\mathbf{X}_{ij} - \overline{\mathbf{X}}_{i\bullet})' \quad \text{and}$$
$$Q = \sum_i (\overline{\mathbf{X}}_{i\bullet} - \overline{\mathbf{X}}_{\bullet\bullet})(\overline{\mathbf{X}}_{i\bullet} - \overline{\mathbf{X}}_{\bullet\bullet})',$$

   where

$$\overline{\mathbf{X}}_{i\bullet} = \frac{1}{n} \sum_j \mathbf{X}_{ij}, \quad \text{and} \quad \overline{\mathbf{X}}_{\bullet\bullet} = \frac{1}{mn} \sum_i \sum_j \mathbf{X}_{ij}.$$

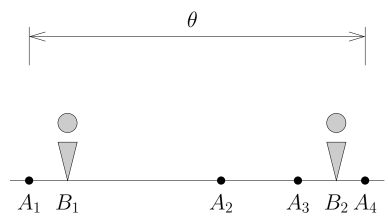
   (a) Show that the eigenvalues of QP^{-1} are all real and nonnegative. (Assume that P is nonsingular.)

   (b) Suppose, to test the null hypothesis

$$H_0 : \mu_1 = \cdots = \mu_m,$$

   we reject H_0, if $H_{0\mathbf{a}} : \mathbf{a}'\mu_1 = \cdots = \mathbf{a}'\mu_m$ is rejected for at least one nonnull vector $\mathbf{a} \in \mathbf{R}^p$ (based on one-way ANOVA F-test). Show that this test rejects H_0 for large values of the maximum eigenvalue of QP^{-1}.

5. There are 4 landmarks $A_1, ..., A_4$ along a straight road to be surveyed. Two surveyors are standing at the points B_1 and B_2, as shown in the diagram below. Each surveyor makes an approximate measurement of his distance to the nearest landmark on either side of himself. The measurements are assumed to be unbiased, homoscedastic and independent.



   (a) Write down the underlying linear model for the data introducing suitable parameters.

   (b) Let θ be the distance between A_1 and A_4. Is it estimable? Justify your answer using your model.

6. We have a finite set $S = \{0, ..., n-1\}$, and a probability distribution $(p_0, ..., p_{n-1})$ on S, where $p_i > 0$ and $\sum p_i = 1$. Suppose that $X_1, X_2, ...$ are iid with this distribution. Let α be any arbitrary element in S. Define

$$T_k = \left\{ \alpha + \sum_{i=1}^{k} X_i \right\} \pmod{n}, \quad \text{for } k = 1, 2, ...$$

   Note that $\{T_k\}$ is a Markov chain.

   (a) Compute the transition matrix for the chain, $\{T_k\}$.

   (b) Show that, for each $j \in S$,

$$P(T_k = j) \to \frac{1}{n} \text{ as } k \to \infty.$$

7. Let $X_1, ..., X_n$ be iid $Unif(0, \theta)$ and $Y_1, ..., Y_n$ be iid $Unif(-\theta, \theta)$ random variables, where θ belongs to the set of natural numbers. Consider the following two hypotheses:

$$H_0 : \theta \text{ is even} \quad vs. \quad H_1 : \theta \text{ is odd}.$$

   (i) On the basis of $X_1, ..., X_n$, give a test, say ϕ_1, for testing H_0 *vs.* H_1 such that both $P(\text{type I error})$ and $P(\text{type II error})$ tend to 0 as $n \to \infty$.

   (ii) On the basis of $Y_1, ..., Y_n$, give another test, say ϕ_2, for testing H_0 *vs.* H_1 whose $P(\text{type I error})$ and $P(\text{type II error})$ are the same as those of ϕ_1.

8. Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$ where the mean μ is unknown and the variance σ^2 is known with $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$ and $n \geq 3$. Let $b > 0$ be a known constant.

   (a) Find the MVUE of b^μ.

   (b) Find $E(b^{X_1+X_2}|S)$, where $S = X_1 + \cdots + X_n$.

9. Consider a finite population of $2N$ units. Let Y be a variable taking values Y_i for $i = 1, ..., 2N$. It is proposed to select $2n$ units according to either design A or design B which are given below.

   Design A:

   > Select $2n$ units from $2N$ units by SRSWR.

   Design B:

   > Divide the $2N$ units into two equal groups $\{1, ..., N\}$ and $\{N + 1, ..., 2N\}$. Select two SRSWR samples of size n from each group separately.

   It is proposed to use the sample mean of the units in the sample to estimate the population mean. Which of the two designs would you use? Justify your answer by comparing the variances of the estimator for the two designs.

***For more sample questions, you are requested to visit the webpage** http://www.isical.ac.in/~deanweb/JRFSTATSQ.html