INDIAN STATISTICAL INSITUTTE

Mid-Semester Examination: 2016-17

Course Name: M Stat - I

Subject Name: Multivariate Analysis

Date: 05.09.2016 Maximum Marks: 30 Duration: $1\frac{1}{2}$ hours

1. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, \Sigma)$ where $\Sigma = ((\sigma_{ij}))$ is a positive definite matrix. Define $A = XX'$ where $X = (\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n)'$. Based on a random sample of size n, how do you do test,

   (a) $H_0 : \sigma_{11} - 2\sigma_{12} + \sigma_{22} = \sigma_0^2$ against $H_1 : \sigma_{11} - 2\sigma_{12} + \sigma_{22} > \sigma_0^2$? You have to prove the result that you use.
   (b) If the above null hypothesis is accepted, how do you test $H_0 : \sigma_{12} = 0$ against $H_1 : \sigma_{12} > 0$?
   (c) Find out the expectation and the variance of $\underset{\sim}{Z}' A \underset{\sim}{Z}$ where $Z_i (i = 1, \ldots, n)$ are *i.i.d.* $N(0, 1)$ variables. (5+4+5=14)

2. Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be *i.i.d.* random vectors from $N_p(\underset{\sim}{0}, \Sigma)$.

   (a) Derive the likelihood criterion for testing $H_0 : \Sigma = \sigma^2 I$.
   (b) Find the r-th $(r = 1, 2, \cdots)$ order moment of the LR criterion under H_0. (4+7=11)

3. Let Y_{ij} be *i.i.d.* random variables with probability density function

$$f_{Y_{ij}}(y) = \frac{\theta_i}{1-\rho} e^{-\theta_i y/(1-\rho)}, \ y > 0, \ i = 1, 2, \ldots, k; j = 1, 2, \cdots$$

   Also let M have a geometric distribution with probability mass function

$$P(M = m) = (1-\rho)\rho^{m-1}, \ 0 \le \rho < 1, \ m = 1, 2, \ldots$$

   (a) Find the distribution of $X_i = \sum_{j=1}^{M} Y_{ij}$ for $i = 1, 2, \ldots, k$.
   (b) Find the joint density of $\underset{\sim}{X} = (X_1, \ldots, X_k)'$.
   (c) Find $E(X_r X_s)$ explicitly for any $1 \le s < t \le k$. (4+5+4=13)

Statistical Inference

M-Stat I, First Semester 2016-2017

Mid-Semester Examination

Date: 06.09.2016 | Maximum marks: 100 | Duration: 3 hours

1. (a) Suppose that $X \sim N_p(\theta, I)$, where θ is the parameter of interest and I represents the identity matrix of dimension p. Let the prior be multivariate normal with mean zero and covariance matrix $c^2 I$ for some real constant c. Let the loss be squared error loss, i.e., given the estimator $\delta(X)$, the loss is the squared L_2 norm between $\delta(X)$ and θ. Show that $\delta(X) = X$ is minimax.

   (b) In part (a), let $p = 1$. Show that $aX + b$ is not admissible when (i) $a > 1$, (ii) $a < 0$ or (iii) $a = 1, b \neq 0$. [12+3+3+2=20]

2. Define an ϵ-Bayes rule ($\epsilon > 0$), an extended Bayes rule, a generalized Bayes rule and the limit of a Bayes rule. Given an example of each. [3+3+3+3=12]

3. (a) Define a lower boundary point of a convex set $S \in \mathbb{R}^k$.

   (b) Let $\lambda(S)$ denote the set of lower boundary points of S. When is the set S said to be closed from below?

   (c) Consider a finite parameter space $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$. Suppose there exists a rule δ_0 such that $x = (R(\theta_1, \delta_0), \dots, R(\theta_k, \delta_0))$ is in $\lambda(S)$. Show that δ_0 is an admissible rule. [3+3+12=18]

4. Suppose that the parameter space $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ is finite and the risk set S is bounded from below and closed from below. Then the class of decision rules

$$D_0 = \{\delta \in D^* : (R(\theta_1, \delta, \dots, R(\theta_k, \delta) \in \lambda(S)\}$$

   is a minimum complete class. (Here D^* and $\lambda(S)$ have their usual meanings). [15]

5. Suppose that S is a closed, convex subset of $\mathbb{R}^k$ which does not contain the origin. Show that there exists a vector $p \in \mathbb{R}^k$ such that $p^T x > 0$ for all $x \in S$. [15]

6. Suppose that for a given decision problem with a finite parameter space $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ the risk set S is bounded below. Then show that

$$\inf \sup r(\pi, \delta) = \sup \inf r(\pi, \delta) = V,$$

   i.e, the value of the game exists. Also show that there exists a least favourable distribution π_0. (Here "sup" is with respect to the prior, and "inf" is with respect to the decision rule). [20]

MI-2016-17
Midterm Examination
Categorical Data Analysis

Full Marks 30 Date: 7th September, 2017 Time: 10.30-12.30

NOTE: Answer the first four questions. Marks will be deducted for unnecessary writing.

1.A press release about a study stated that the odds of referral for cardiac catheterization for blacks are 57.4 % of the odds for whites. If the probability that a white would be referred for heart catheterization was 0.906, calculate the probability that a black would be referred for heart catheterization. (4)

2a. Using entropy measure of variation, obtain U, the Uncertainty Coefficient, a measure of association between two categorical variables X and Y.

b. Show that $U = 0$ iff X and Y are independent. [3+(1+4)=8]

3. It is required to compare two methods of treating a type of allergy. Method I was used on 15 patients and Method II was used on 14 patients. The results are shown in the following table. Judge, using Fisher's exact test procedure if Method II is better than Method I. (6)

	Cured	Not cured	Total
Method I	6	9	15
Method II	11	3	14
Total	17	12	

Please turn over

4. For a Logistic regression to determine characteristics associated with remission in cancer patients, the most important explanatory variable is LI, that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. Software output gives the estimate of "intercept" and LI as -3.7771 and 0. 1449 respectively with the estimated Covariance matrix as follows: (2+2+3=7)

	Intercept	LI
Intercept	1.900616	-0.07653
LI	-0.07653	0.003521

a) Estimate the success probability at LI=8
b) For a unit increase in LI, by how much the estimated odds will multiply?
c) Obtain 95% Confidence interval for the estimated odd ratio.

5. Class assignments and practical assigned. (5)

STOCHASTIC PROCESSES
M. STAT. IST YEAR SEMESTER 1
INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination
Time: 2 Hours Full Marks: 35
Date: September 8, 2016

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed. Calculators and any other computing devices are strictly not allowed.

1. Find out the extinction probability of a branching process, whose progeny size is distributed as Binomial $(n,\ \frac{1}{2})$, in each of the cases $n = 1, 2, 3$. [1+1+4=6]
2. In a bacteria colony, bacteria are classified as young, reproductive or old. Each young bacteria survive the current time period with probability p and after the current time period moves to the reproductive stage. A bacteria in reproductive stage survives with probability q and produces young bacteria whose count has generating function ψ. After reproduction, the bacteria in reproductive stage becomes old in the next time period. The old bacteria survive each time period with probability r and continue in the old stage till it dies. Pose the problem as a multitype branching problem stating your assumptions clearly. Provide the offspring generating function and the mean matrix. [2+3+3=8]
3. A bus arrives in an empty bus station and leaves after a deterministic time T after picking up all the passengers who arrive while the bus is waiting. If the passengers arrive like a Poisson process with rate λ, what is the expected value of the total waiting time of all the passengers together? [7]
4. The autos plying between Dunlop and Sinthee arrive at Bonhoogly like a Poisson process with intensity 20 per hour. The autos can arrive with 0, 1, 2, 3 or 4 (the maximum capacity after the recent traffic regulation drive) passengers with probability $\frac{1}{20}$, $\frac{1}{20}$, $\frac{1}{5}$, $\frac{1}{5}$, $\frac{1}{2}$. If you are waiting with your best friend to go for dinner, how long, on average, will you have to wait to so that both of you get on an auto? Note that in order to reach the restaurant quickly, you are willing to go separately in the auto. [7]
5. Shounak, Soumya and Tuhin entered the Dean's office simultaneously with the same request, which can be attended to by any of the two office staffs, who start working on the requests of Shounak and Soumya immediately (this is a math problem and the situation has been idealized), while Tuhin waits for one of his friends to finish. Both the office staff work independently and take Exponential(λ) amount of time to get the job done. What is the probability that Shounak will be the last to leave? What is the expected time taken by Tuhin to leave? [4+3=7]

Indian Statistical Institute

First Midsemestral Examination 2016-17

M. Stat. I yr (B-Stream)

Regression Techniques

Date: September 9 , 2016 (14:30 hrs) Maximum marks: 60 Duration: 2 hrs.

Answer all Questions. ***Paper carries 65 points.***

1 (a) Consider a standard linear model with standard notations which can be expressed as

$$Y = X\beta + Z\gamma + \varepsilon,$$

where the design matrix is partitioned into X and Z based on two sets of variables. There errors, ε are iid normal with mean zero and variance σ^2. It is desired to minimize

$$L(\beta, \gamma \,|\, \lambda) = ||Y - X\beta||^2 + ||Y - Z\gamma||^2 + \lambda\, ||X\beta - Z\gamma||^2,$$

simultaneously over β and γ. The Lagrangian penalty parameter $\lambda > 0$ is known. Obtain a minimizer $(\beta^*(\lambda), \gamma^*(\lambda))$ of $L(\beta, \gamma \,|\, \lambda)$.

(b) In the previous example assume $n = 2k + 1$ for some positive integer $k > 5$. Let $X = [\mathbf{1},\ \mathrm{x}\,]$ and $Z = [\mathbf{1},\ \mathrm{z}\,]$ respectively where $\mathbf{1} = (\underbrace{1, 1, \ldots, 1}_{n \text{ times}})^T$,
$\mathrm{x} = (-k, -(k-1), -(k-2), \ldots, 0, 1, 2, \ldots, k)^T$ and $\mathrm{z} = (k^2, (k-1)^2, \ldots, 0, 1, 2^2, 3^2,$ respectively. Find the limit of $L\,(\beta^*(\lambda), \gamma^*(\lambda)\,|\lambda)$ as $\lambda \to \infty$.

[(10+5 = 15]

2. Cosider a standard regression data

$$Y = X\beta + \varepsilon,$$

with n observations and $n \times p$ design matrix of full rank. Let $\hat{\beta}_n$ denote the OLS estimate of β from the given data. Suppose a new observation (y_{n+1}, x^T_{n+1}) is added to the data set so that the new OLS estimate of β is now $\hat{\beta}_{n+1}$. Express the difference $\left(\hat{\beta}_{n+1} - \hat{\beta}_n\right)$ in terms of other observations in a closed form (as a function of Y, X, y_{n+1} and x^T_{n+1}).

(b) Find the variance of $(y_{n+1} - x^T_{n+1}\hat{\beta}_n)$.

[15+5=20]

Please turn over...

3. Describe notions of the following with examples. [Answer must be short. But it should be adequate description as in a good wikipedia page]
   (a) Multicolinearity and variance inflation factor
   (b) Ridge regression and its bias and variance issues.

[10+10 = 20]

4. Take home projects.

[2+2+2+2+2 =10]

Indian Statistical Institute

First Midsemestral Suplementary Examination 2016-17

M. Stat. I yr (B-Stream)

Regression Techniques

Date: October 25, 2016 (16:00 hrs) Maximum marks: 60 Duration: 2 hrs.

Answer all Questions. ***Paper carries 65 points.***

1. Consider standard linear model with standard notations which can be expressed as

$$Y = X\beta + Z\gamma + \varepsilon,$$

where the design matrix is partitioned into X $(n \times p)$ and Z $(n \times q)$ based on two sets of variables. The errors, ε are iid normal with mean zero and variance σ^2. It is desired to estimate γ using OLS. For that purpose we first carry out OLS estimation of $Y, Z_1, Z_2, \ldots, Z_q$ on X. Let $e_y, e_z^1, e_z^2, \ldots, e_z^q$ denote respective residual vectors, each $(n \times 1)$. Next we regress e_y by OLS on a matrix whose columns are given by $e_z^1, e_z^2, \ldots, e_z^q$ respectively. Prove or disprove the the assertion that the estimated regression coefficients of the second stage regression, namely $\hat{\gamma}_2$, equals the OLS estimate $\hat{\gamma}$ from the original model stated above.

[15]

2 (a) Define Cook's distance measure in a standard linear model. Demonstrate its usefulness using an example.

(b) In a regression study three types of banks were involved, commercial, mutual savings and savings and loan respectibely. Suppose a type classifier is of these three types of banks are described by a pair of variables (X_2, X_3) taking values $(1,0)$, $(0,1)$ and $(-1,-1)$ for three types of banks respectively. Also we have another regressor X_1 denoting the size of a bank in a data set (assume all three types banks occur with sufficient requency in a data of size n). The response variable is Y (which is previous years profit).

(i) Develop a linear regression model for relating last years profit to size and category of the banks and write down estimates of three regression parameters in terms of Y and X_1 (you may asume there are n_i observations from category i in the data).

(ii) Interpret each of the following quantities (i) β_2, (ii) β_3 and (iii) $-\beta_2 - \beta_3$ in real terms

[10+10=20]

Please turn over...

3 (a) Suppose the true regression equation is $E(Y|X) = 15 + 20X + 3X^2$ (with common error variance σ^2) where a linear regression is fitted by mistake. From a data of size n it turns out $\hat{Y} = 13 + 40X$ (say, $\hat{\alpha}_n + \hat{\beta}_n X$). Also it is given (verified separately) that $E(\hat{\alpha}_n) = 10$ and $E(\hat{\beta}_n) = 45$ respectively under the true model. Find the true MSE of $\hat{\alpha}_n + \hat{\beta}_n\, x$ when $x = 10$ and $x = 20$ respectively.

(b) Find the true value of $\sum_{i=1}^{n} \text{Var}\hat{Y}_i$.

[10+10=20]

4. Take home projects.

[2+2+2+2+2 =10]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2016

Course Name: M. Stat – I

Subject Name: Multivariate Analysis

Date: 15.11.2016 Maximum marks: 70 Duration: 3 hrs

1. (a) If $A \sim W_p(n, I_p)$, then show that

   (i) $E(A^k) = c(k,n,p)I_p$ and

   (ii) $E(A^{-k}) = d(k,n,p)I_p$

   where $c(k,n,p)$ and $d(k,n,p)$ are constants depending on *k, n,* and *p*. (5+5=10)

2. Let $A \sim W_p(n,\Sigma)$ and $B \sim W_p(m,\Sigma)$ and *A* and *B* are independently distributed.

   (a) Show that $\frac{|A|}{|A+B|} \sim \Lambda_{p,m,n}$, where $\Lambda_{p,m,n}$ has the same distribution as that of $\prod_{i=1}^{p} U_i$ where $U_i \sim B\left(\frac{n-i+1}{2}, \frac{m}{2}\right)$ $i = 1,\ldots,p$ and U_i's are independently distributed.

   (b) Also show that the distributions of $\Lambda_{p,m,n}$ and $\Lambda_{m,p,m+n-p}$ are same for any choice of *m, n,* and *p*.

   (c) Find the distributions of $\frac{1-\Lambda_{p,1,n}}{\Lambda_{p,1,n}} . \frac{n-p+1}{p}$ and $\frac{1-\sqrt{\Lambda_{2,m,n}}}{\sqrt{\Lambda_{2,m,n}}} . \frac{n-1}{m}$ (8+5+3+3=19)

3. Let $X_1,\ldots,X_n$ be a random sample from $N_p(\underset{\sim}{\mu_1},\Sigma_1)$ and $Y_1,\ldots,Y_m$ be another random sample from $N_p(\underset{\sim}{\mu_2},\Sigma_2)$.

   (a) Derive likelihood ratio test for testing $H_0 : \underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$ against $H_1 : \underset{\sim}{\mu_1} \neq \underset{\sim}{\mu_2}$, under the assumption of $\Sigma_1 = \Sigma_2 = \Sigma$, where Σ is unknown.

   (b) Also derive union-intersection test for the same null hypothesis and check whether the two critical regions are same.

   (c) Show that the test based on union-intersection principle is unbiased.

   (5+6+3=14)

4. Let $X_1,\ldots,X_n$ be a random sample of *p*-dimensional random variables with distribution function *F*. Consider the problem of testing $H_0 : F = F_0$ against $H_1 : F \neq F_0$, where F_0 is some specified distribution function and define a test statistic $D_n^* = \sup_{x \in \Re^p} |F_n(x_1,\ldots,x_p) - F(x_1,\ldots,x_p)|$. However, D_n^* is not distribution free.

   (a) Suggest a transformation $y = T(x)$ such that $D_n = \sup_{y \in \Re^p} |G_n(y_1,\ldots,y_p) - (y_1 \ldots y_p)|$ would be distribution free under H_0, where G_n is the empirical distribution function based on *y* values.

(b) For $p=2$, define $D_n^+(u)=(G_n(u)-G(u))$, $D_n^-(u)=(G(u)-G_n(u))$, $D_n^+=\sup_u D_n^+(u)$, and $D_n^-=\sup_u D_n^-(u)$, where G is the distribution function of two independent uniform random variables on (0,1) and G_n is the empirical distribution function. Then prove the following:

(i) If $x_0=y_0=0$, then $D_n^+=\max_{v\in I} D_n^+(v)$, where $I=\{(x_j,y_i)\mid x_i\le x_j, y_i\ge y_j;\ i,j=0,1,\ldots,n\}$.

(ii) If $x_0=0, y_0=1, x_{n+1}=1$, and $y_{n+1}=0$, then $D_n^-=\max_{v\in P}(G(v)-G_n(v-))$, where $P=\{(x_j,y_i)\mid x_j>x_i, y_j<y_i;\ i,j=0,1,\ldots,n+1\}$.

(iii) Using (i) and (ii) above or otherwise, show that for $p=2$,
$D_n=\sup_{u\in I, v\in P}\{G_n(u)-G(u), G(v)-G_n(v-)\}$.

(iv) Describe an algorithm step by step to calculate D_n based on sample observations $\{(x_i,y_i), i=1,2,\ldots,n\}$. Explain clearly how you can calculate the p-value using any computational technique. (2+5+5+3+4+4=23)

4. Consider an examination data whose covariance matrix is given by

$$S=\begin{pmatrix} 302.3 & 125.8 & 100.4 & 105.1 & 116.1 \\ & 170.9 & 84.2 & 93.6 & 97.9 \\ & & 111.6 & 110.8 & 120.5 \\ & & & 217.9 & 153.8 \\ & & & & 294.4 \end{pmatrix}$$

and mean vector $\bar{x}'=(39.0,\ 50.6,\ 46.7,\ 42.3)$. A spectral decomposition of S yields principal components

$$y_1=0.51x_1+0.37x_2+0.35x_3+0.45x_4+0.53x_5-99.7$$
$$y_2=0.75x_1+0.21x_2-0.08x_3-0.30x_4-0.55x_5+1.5$$
$$y_3=-0.30x_1+0.42x_2+0.15x_3+0.60x_4-0.60x_5-19.8$$
$$y_4=0.30x_1-0.78x_2-0.10x_3+0.52x_4-0.518x_5+11.1$$
$$y_5=0.08x_1+0.19x_2-0.92x_3+0.29x_4+0.15x_5+13.9$$

with variances 679.2, 199.8, 102.6, 83.7, and 31.8 respectively. It is seen that the marks of 5 students in five subjects are as follows:

Mechanics	Vectors	Algebra	Analysis	Statistics
77	82	67	67	81
63	78	80	70	81
75	73	71	66	81
55	72	63	70	68
63	63	65	64	63

(a) Check whether the objective of dimension reduction is fulfilled in this analysis.
(b) Plot the marks profile of 5 students with respect to first two principal components and comment.
(c) Write a report on the analysis given and what more you can do with this analysis. Write your interpretation of the results thus obtained clearly stating any theory and/or concept that you want to use. (3+5+4=12)

STATISTICAL INFERENCE I

M-Stat I, First Semester 2016-2017

Semestral Examination

Date: 18.11.2016 Duration: 3 hours

[Answer as many as you can. Total points 115. Maximum you can score is 100]

1. Suppose that X follows a binomial distribution with parameters n and θ, where $\theta \in \Theta = (0, 1)$.
   (a) Consider the *beta*(α, β) prior, and the squared error loss $L(\theta, \delta) = (\delta - \theta)^2$. Show that the maximum likelihood estimate x/n is not a Bayes rule.
   (b) Now consider the loss function $L(\theta, \delta) = (\delta - \theta)^2/[\theta(1-\theta)]$. Show that the maximum likelihood estimate x/n is a Bayes rule with respect to the uniform prior on (0, 1). [10+10]

2. Show that the convex hull of a subset S_0 of the k-dimensional Euclidean space $\mathbb{R}^k$ is the set of all convex linear combinations of at most $k+1$ points of S_0. [20]

3. For the parametric model $\{f_\theta, \theta \in \Theta\}$, consider testing the hypothesis $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$ based on a sample of size n. Suppose that the family of distributions has monotone likelihood ratio in the statistic T. Consider the tests of the form

$$\phi(X) = \begin{cases} 1 & \text{if } T > c, \\ \xi & \text{if } T = c, \\ 0 & \text{if } T < c. \end{cases} \tag{1}$$

   (a) Show that for every test ϕ^* and every $\theta_0 \in \Theta$, there exists a test ϕ' of the form (1) such that

$$E_\theta(\phi'(X)) \le E_\theta(\phi^*(X)), \text{ for } \theta < \theta_0,$$

   and

$$E_\theta(\phi'(X)) \ge E_\theta(\phi^*(X)), \text{ for } \theta > \theta_0.$$

   (b) Let a_1 and a_0 represent the actions of rejecting the null hypothesis and failing to do so, respectively. If the loss function satisfies the inequalities

$$L(\theta, a_1) - L(\theta, a_0) \ge 0 \text{ for } \theta < \theta_0,$$

   and

$$L(\theta, a_1) - L(\theta, a_0) \le 0 \text{ for } \theta > \theta_0,$$

   and the family of distributions has monotone likelihood ratio in T, show that the class of one-sided tests (1) is essentially complete. If the set of points $\{x | f_\theta(x) > 0\}$

is independent of θ and if there exist $\theta_1 \in \Theta$, $\theta_2 \in \Theta$, with $\theta_1 \leq \theta_0 \leq \theta_2$ such that $L(\theta_1, a_1) - L(\theta_1, a_0) > 0$ and $L(\theta_2, a_1) - L(\theta_2, a_0) < 0$, then any test of the form (1) is admissible. [10+10]

4. (a) Consider a hypothesis testing problem involving a random variable X belonging to a parametric family of models indexed by $\theta \in \Theta$. Given a subset ω of Θ and a test function ϕ, explain what is meant by saying that the test is similar with respect to ω.

   (b) When do we say that a test function ϕ has Neyman structure with respect to a statistic T?

   (c) Let X be a random variable with distribution $P \in \mathcal{P}$, and let T be a sufficient statistic for $\mathcal{P}$. Show that a necessary and sufficient condition for all similar tests to have Neyman structure with respect to T is that the family of distributions $\mathcal{P}^T$ of T is boundedly complete. [3+5+12]

5. Consider the two parameter regular exponential family given by the density

$$f_{\theta,\xi}(x) = C(\theta, \xi) \exp[\theta U(x) + \xi T(x)] h(x). \quad (2)$$

Construct the UMP unbiased level α test for the hypothesis

$$H_0 : \theta \leq \theta_0 \text{ versus } H_1 : \theta > \theta_0,$$

where ξ is unspecified. [15]

6. (a) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$. Suppose that μ is fixed at μ_0. Show that

$$\frac{\bar{X} - \mu_0}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \quad \text{and} \quad \frac{\bar{X} - \mu_0}{\sqrt{\sum_{i=1}^{n} (X_i - \mu_0)^2}}$$

are independent of $\sum_{i=1}^{n} (X_i - \mu_0)^2$.

[You may use the following corollary of Basu's Theorem in establishing the above: *Let $\mathcal{P}$ be the exponential family in Equation (2). Based on a random sample $X_1, X_2, \ldots, X_n$, the statistics $U = \sum U(X_i)$ and $T = \sum T(X_i)$ are jointly sufficient. Consider the family of distributions obtained by letting θ have a fixed value. Then a statistic V is independent of T for all ξ provided the distribution of V does not depend on ξ.*]

   (b) Using part (a) or otherwise, find the UMP unbiased test of level α for testing the hypothesis $H_0 : \mu = \mu_0$, σ unspecified, against the two sided alternative. [10+10]

INDIAN STATISTICAL INSTITUTE

Semester exam. (Semester I: 2016-2017)

Course Name: M. Stat. 1st year

Subject Name: Categorical Data Analysis

Date: 21.11 , 2016, Maximum Marks: 70. Duration: 3 hrs.

Note: Answer all questions.
Use separate answer booklets to answer Group A and Group B.

Group A

1. Geometrically interpret *Risk Ratio = 1* for 2×2 contingency tables. [10]

2. Discuss the latent variable approach to model ordinal categorical variables with possible covariates. Write down the likelihood assuming normality of the latent variable. How can you model a bivariate response vector using latent variable approach where one variable is ordinal categorical and the other is continuous? Write down the likelihood in this context. [4+2+4+4=14]

3. Suppose there are three urns, labelled as urn A, urn B and urn C, each having 4 yellow, 5 blue and 6 orange balls initially. Let Y_1, Y_2, Y_3 be three discrete random variables. Draw a ball from the urn A, Y_1 represents the indicator variable corresponding to the colour of the drawn ball. Consequently we add an additional ball of the same colour of the drawn ball to the urn B. Now draw one ball from this urn B, and Y_2 represents the indicator corresponding to the colour of the drawn ball. Consequently add one ball of the same colour as the drawn ball (from urn B) to urn C. Now draw one ball from urn C, and Y_3 corresponds to the colour of this drawn ball. Find the marginal distributions of Y_1, Y_2 and Y_3. Find $\tau_{12}, \tau_{13}, \tau_{23}$, where τ_{ij} is the Goodman and Kruskal's τ for Y_j on Y_i. Find τ_{13} as a function of τ_{12} and τ_{23}. [3+10+2=15]

4. Discuss how negative binomial margin can be constructed by using the *binomial thinning operator*. Discuss the model with covariates. In this context write down the likelihood for two observations Y_1 and Y_2 with possible covariates, where marginally each of Y_1 and Y_2 follows a negative binomial distribution. [5+2+4=11]

p.T.O

Group B

5. Consider the following table showing the Presidential votes in 2004 and in 2008 for Males sampled in 2010 by the general social survey in USA.

	2008 Election	
2004 Election	Democrat	Republican
Democrat	175	16
Republican	54	188

Is there any evidence of a shift of votes in the Democrat direction? [6]
Justify carrying out proper statistical testing of hypothesis.

6. For a multiple logistic regression model, obtain the likelihood equations for estimating the model parameters and the variance –covariance matrix of the MLE of the parameters. Suggest an estimator of this variance –covariance matrix.

[6+6+2=14]

Stochastic Processes
M. Stat. Ist Year Semester 1 (2016-2017)
Indian Statistical Institute

Semestral Examination
Time: 3 Hours   Full Marks: 50
Date: 25th November, 2016

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed. Calculators and any other computing devices are strictly not allowed.

1. For a discrete time branching process where each individual produces either no offspring or two offsprings with equal probability, find out the probability generating function of the total number of individuals ever born in the system. Hence or otherwise, find out the mass function of the random variable. [4+4=8]

2. The restaurants are placed randomly like a two dimensional homogeneous Poisson process with unit intensity. A customer goes to the nearest restaurant. Your restaurant is located independently of the other restaurants. What is the probability that your restaurant will get all the customers within unit distance from you? [5]

3. Consider a pure renewal process $\{N(t)\}$ corresponding to the i.i.d. inter-renewal sequence $\{X_n\}$ with common distribution F. For any $c > 0$ with $F(c) < 1$, define a new binary inter-renewal sequence $\{Y_n\}$ given by:
$$Y_n = \begin{cases} 0, & \text{if } X_n \le c, \\ c, & \text{otherwise.} \end{cases}$$
Let $\{M(t)\}$ be the renewal process corresponding to $\{Y_n\}$. For $t < c$, show that $M(t)$ has a geometric distribution and write down its mass function. What will be the mass function of $M(t)$ for $t \ge c$. (You may want to consider $M(t) - [t/c]$ first.) Hence, show that $N(t)$ has all moments finite. [4+4+4=12]

4. Identify which of the following random variables have distribution supported on the sets of the form $\{nb : n = 0, 1, 2, \ldots\}$, for some $b > 0$. If the distribution is of the required form, find the largest possible value of b. [3×3=9]
   (a) X takes values $(2k-1)\pi$ with probability 2^{-k}, for $k = 1, 2, \ldots$.
   (b) X takes values $\sqrt{2}$ and $\sqrt{3}$ with equal probability.
   (c) $X = 1/(Y+1)$ where Y follows Poisson distribution.

5. Let $X(t)$ be the size of the population at epoch t, where each individual in the population survives for an Exponential(μ) amount time. Until death, each individual gives birth to new individuals according to a Poisson process at rate λ. By making suitable further assumptions, formulate the problem as a Birth and Death process. Write down the infinitesimal generating matrix with justification. Find the stationary distribution when $\lambda = \mu$. (You may use the formula for the stationary distribution for the general Birth and Death process without proof.) [2+4+3=9]

6. Consider an $M/M/\infty$ queue with arrival rate a and service rate b. Obtain the infinitesimal generator matrix for $\{C(t)\}$, the number of customers in the system at time t. Find the stationary distribution. (You may use the formula for the stationary distribution for the general Birth and Death process without proof.) [4+3=7]

Indian Statistical Institute

First Semestral Examination 2016-17

M. Stat. I yr (B-Stream)

Regression Techniques

Date: November 28, 2016 (14:30 hrs) Maximum marks: 100 Duration: 3 hrs.

Answer all Questions. ***Paper carries 110 points.***

1. Consider standard linear model with standard notations which can be expressed as

$$Y = X\beta + Z\gamma + \varepsilon,$$

where the design matrix is partitioned into X $(n \times p)$ and Z $(n \times q)$ based on two sets of variables. The errors, ε components are iid normal with mean zero and variance σ^2.

(a) Consider the SSE for the full model, namely,

$$SSE = \min_{\beta,\gamma} ||Y - X\beta - Z\gamma||^2.$$

Next suppose the regression is carried out in two steps. In the first step Y is regressed on X to obtain residual vector $e_1 = Y - X\hat{\beta}$ where $\hat{\beta}$ is the OLS estimate. Next we obtain SSE_1 where

$$\text{SSE}_1 = \min_{\gamma} ||e_1 - Z\gamma||^2.$$

Show that $\text{SSE}_1 \geq \text{SSE}$.

(b) Describe conditions under which equality is achieved in (a).

[12+8 =20]

2 (a) Consider a simple linear regression model with time series data

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

for $t = 1, 2, \ldots, T$ with $\epsilon_t = \rho\epsilon_{t-1} + u_t$ where $|\rho| < 1$ (known) and $\{u_t\}$ are iid normal with mean 0 and variances $\sigma^2 > 0$. (Assume ϵ_0 is normally distributed with mean 0 and variance $\sigma^2/(1-\rho^2)$). Find approximate distribution of $\hat{\beta}_{\text{OLS}}$ when T is assumed to be large.

(b) Describe the Durbin-Watson test statistic D for time series regression model (as in (a)) clearly stating which hypothesis it tests. Show that

$$D \approx 2(1 - r)$$

where $r = \left(\sum_2^T e_{t-1}e_t\right) / \left(\sum_2^T e_{t-1}^2\right)$, the usual estimate of the slope in error correlation model with unobserved ϵ_t's replaced by the OLS estimated error $\{e_t\}$ from the regression model.

[10+ 15=25]

P.T.O

3 (a) Discuss the differences in statistical characteristics between intrinsically linear and nonlinear regression models with suitable examples.

(b) In enzyme kinetics study the velocity of reaction (Y) is expected to be related to the concentration (X) as follows.

$$Y_i = \frac{\gamma_0 X_i}{\gamma_1 + X_i} + \epsilon_i.$$

In order to apply Gauss Newton method describe how would you choose (i) initial values and (ii) iteration steps (equation) and (iii) termination of iteration in the algorithm. (precise description necessary)

[10+ 15=25]

4 (a) Let (X, Y) have a bivariate normal distribution with standard notations. Obtain the likelihood ratio test statististic for testing $\rho_{XY} = 0$ based on a sample of size n.

(b) Can the LR statistic in (a) be derived using regression of Y on X? Justify.

[15+ 10=25]

5. Submitted project reports

[15]

INDIAN STATISTICAL INSTITUTEM Stat 1st Year (2016-17)

Back paper First Semester Examination: 2016

Course Name: M. Stat – I

Subject Name: Date: Multivariate Analysis

Date: 26.12.2016 Maximum marks: 100 Duration: 2 hrs 30 mins

1. Let U_i be the i-th principal component of **X**, $i = 1,2,...,p$. Then show that
(a) U_1 has maximum variance among all normed linear combinations of $X_1, ..., X_p$
(b) U_i has maximum variance among all normed linear combinations of $X_1, ..., X_p$ uncorrelated with $U_1,...,U_{i-1}$; $i = 1,...,p$.
(c) In principal component analysis explain how you decide on the number of principal components. (6+6+6=18)

2. (a) Let $A \sim W_p(n,\Sigma)$ and $B \sim W_p(m,\Sigma)$ and A and B are independently distributed. Then show that $\frac{|A|}{|A+B|} \sim \Lambda_{p,m,n}$, where $\Lambda_{p,m,n}$ has the same distribution as that of $\prod_{i=1}^{p} U_i$ where $U_i \sim B\left(\frac{n-i+1}{2}, \frac{m}{2}\right)$ $i = 1,...,p$ and U_i's are independently distributed.
(b) Assuming $n = r + 2\alpha$ where α is fixed, derive the asymptotic distribution of Wilk's lambda statistic $\Lambda_{p,m,n}$ and determine α suitably. (10+12=22)

3. Let a random sample of size N_i be drawn from $N_p(\mu_i,\Sigma)$, $i = 1,2,...,k$; Σ is an unknown positive definite matrix. Based on these samples, we want to test
$H_0 : \mu_1 = \mu_2 = ... = \mu_k$ against H_1 : *at least one equality in* H_0 *is false*
(a) Derive the likelihood ratio test statistic for the above test and find its distribution.
(b) Also construct a test for H_0 using the union-intersection principle. Check whether the test statistic thus obtained has the same distribution as that in (a). Justify your answer. (8+12=20)

4. (a) Derive explicitly the characteristic function of A where $A \sim W_p(n,\Sigma)$.
Hence find the distribution of $A = \sum_{j=1}^{k} A_j$ where $A_j \sim W_p(n_j,\Sigma)$, $j = 1,2,...,k$ and A_j's are independently distributed.
(b) Let $A \sim W_p(n,\Sigma)$. Also let $B = HAH'$ where H is any $p \times p$ orthogonal matrix, the elements of which are random variables distributed independently of A. Show that the distribution of B is distributed independently of H and also find the distribution of B. (10+10=20)

P.T.O

5. Let a random sample of size N be drawn from $N_p(\mu_i, \Sigma)$, where Σ is positive definite and $N > p$.

   (a) Derive the likelihood ratio test criterion (*T, say)* for testing $H_0 : \Sigma = \sigma^2\Sigma_0$ against $H_1 : \Sigma \neq \sigma^2\Sigma_0$.

   (b) Calculate r^{th} order moments of T under H_0.

   (c) Find the exact distribution of T for $p=2$. (6+7+7=20)

INDIAN STATISTICAL INSTITUTE

M-Stat I, First Semester 2016-2017

Back Paper Examination

STATISTICAL INFERENCE I

Date: 27.12.2016 Maximum marks: 100 Duration: 3 hours.

1. Consider the decision problem where the θ is the parameter of interest. The parameter space as well as the action space are both equal to the real line, and the loss function satisfies

$$L(\theta, a) = \begin{cases} k_1|\theta - a| & \text{if } a \leq \theta \\ k_2|\theta - a| & \text{if } a > \theta \end{cases}$$

where $k_1 > 0$ and $k_2 > 0$. All the other symbols have their usual meaning. Assuming any conditions that may be necessary, show that the Bayes rule is the p-th quantile of the posterior distribution of θ, where p is a suitable function of k_1 and k_2. [20 points]

2. Show that the convex hull of a subset S_0 of the k-dimensional Euclidean space $\mathbb{R}^k$ is the set of all convex linear combinations of at most $k+1$ points of S_0. Show, with a counter example, that this number cannot be strictly smaller than $k+1$. [20 points]

3. Let $X_1, \ldots, X_n$ be a random sample from a $N(\theta, 1)$ distribution. Let $0 < \alpha < 1$, and $\theta_1 < \theta_2 < \theta_3 < \theta_4$. Consider testing the following set of hypotheses:

$$H_0 : \theta \leq \theta_1 \quad \text{or} \quad \theta_2 \leq \theta \leq \theta_3 \quad \text{or} \quad \theta \geq \theta_4$$
$$H_1 : \theta_1 < \theta < \theta_2 \quad \text{or} \quad \theta_3 < \theta < \theta_4.$$

Does there exist a level α UMP test for the above? If yes, derive the test. If not, explain why not. [20 points]

4. Let the parameter space and the action space both be equal to $[0, 1]$. Let the loss be

$$L(\theta, a) = (1 - \theta)a + \theta(1 - a).$$

All the symbols have their usual meaning. Let the random observable X have any distribution depending on θ. Show that the decision rule $d(x) = 1/2$ is a minimax rule. [20 points]

5. . Let (X, Y) be distributed according to the exponential family density

$$f_{\theta_1,\theta_2}(x, y) = c(\theta_1, \theta_2)h(x, y) \exp(\theta_1 x + \theta_2 y).$$

Show that the only unbiased test for testing $H_0 : \theta_1 \leq a, \theta_2 \leq b$ against $H_1 : \theta_1 > a$, or $\theta_2 > b$, or both, is $\phi(x, y) = \alpha$, where α is the level of the test. [20 points]

Stochastic Processes
M. Stat. Ist Year Semester 1 (2016-2017)
Indian Statistical Institute

Backpaper Examination

Time: 3 Hours   Full Marks: 100

Date: 28 ~~January~~ Dec., 2017

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed. Calculators and any other computing devices are strictly not allowed.

1. Consider a business family line modelled as a branching process with offspring generating function $P(s) = 1/(2-s)$. If each individual, including the first person in the family, earns profits over his/her lifetime like i.i.d. unit exponential, find out the probability that no member of the business house earns more than x individually. [12]

2. Consider a branching process with immigration, where the offspring generating function is A and in each generation there are independent number of immigrants having a generating function B. If Z_n is the population size in the generation n and the corresponding transition matrix P, show that the generating function of the i-th row of P is given by $B(s)A(s)^i$. If the chain has the stationary distribution with generating function Π, show that $\Pi(s) = B(s)\Pi(A(s))$. [9+9=18]

3. Consider a nonhomogeneous Poisson process on $[0, \infty)$ with the mean intensity function

$$\lambda(t) = \begin{cases} \lambda, & \text{if } 2m \le t < 2m+1 \text{ for some integer } m, \\ 0, & \text{otherwise.} \end{cases}$$

We sample from this process by first choosing an initial point M, a nonnegative integer, according to a probability mass function $\{p_k\}$, and then counting all the points in the interval $[M, M+1)$. What is the probability of counting n points? What are the mean and variance of the counted points? [6+6+6=18]

4. The lighting department of the corporation uses the following policy to change the bulbs which have unit exponential lifetime. It replaces any failed bulb immediately at a cost x and any other bulb, which has survived time T and not failed, at time T at a cost y. Find the optimal time T so as to minimize the long run cost per unit time. [12]

5. Consider a homogeneous Poisson process on the plane with unit intensity. What sort of process the Euclidean norms of the points will form? [10]

6. Show that the limiting remaining lifetime distribution is same as the original inter-renewal distribution iff the inter-renewal distribution is exponential. (You may use the formula for the remaining lifetime distribution without proof.) [15]

7. Show that the transition matrix of a CTMC is continuous everywhere. [15]

INDIAN STATISTICAL INSTITUTE

Semester exam. (Semester I: 2016-2017)

Course Name: M. Stat. 1st year

Subject Name: Categorical Data Analysis (BACK PAPER)

Date: 30/12/2016, Maximum Marks: 100. Duration: 3 hrs. 30 min.

Note: Answer all questions.
Use separate answer booklets to answer Group A and Group B.

Group A

1. Geometrically interpret *Odds Ratio* $= 1$ for a 2×2 contingency table. [12]

2. Data on pre- and post-operative conditions (classified as bad, moderate, good) of 100 patients are given along with their age, sex and another important covariate related to the initial condition of the disease. Give a latent variable based model and illustrate an approach to test whether there is significant improvement in the operation or not. Discuss any computational problem that might be encountered in the analysis. [6+7+4=17]

3. Consider the $2 \times 2 \times 2$ three-way table with cell frequencies $\{n_{ijk}\}$, where $n_{111} = 12$, $n_{112} = 15$, $n_{121} = 25$, $n_{122} = 22$, $n_{211} = 12$, $n_{212} = 10$, $n_{221} = 10$, $n_{222} = 20$. Illustrate the iterative proportional fitting for the models (XY, YZ) and (XY, XZ, YZ). [10]

4. Interpret *thinning operator* for time series of count data. Obtain correlation coefficient at lag h for first order autoregressive Poisson process using the *thinning operator*. [3+8=11]

P.T.O

Group B

5. Consider the following cross classification of job satisfaction by age of respondent in a social survey where the job satisfaction categories being 1= not satisfied, 2= fairly satisfied and 3= completely satisfied.

	Job Satisfaction		
Age (yrs)	1	2	3
<30	34	53	88
30-50	80	174	304
>50	29	75	172

Is there a tendency for higher age to be associated with higher job satisfaction? Justify your claim. [10]

6. For a multinomial model with N cells, obtain the asymptotic joint distribution of the sample proportions in the cells . Hence stating and using the Delta method, obtain the asymptotic distribution of the sample log odds ratio in a 2x2 contingency table. [6+4+6=16]

7. Obtain the likelihood equations to obtain the estimators for the model parameters of a multiple logistic regression model and discuss the method of obtaining the MLE of the parameters.

[8+8=16]

8. For a simple Logistic regression to determine characteristics associated with remission in cancer patients, the most important explanatory variable is LI, that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. Software output gives the estimate of "intercept" and LI as -3.7771 and 0. 1449 respectively with the Estimated Variance-Covariance matrix as follows:

[3+5=8]

	Intercept	LI
Intercept	1.900616	-0.07653
LI	-0.07653	0.003521

a) Estimate the success probability at LI=8.

b) Obtain a 95% Confidence interval for the logit ($\pi(14)$).

Indian Statistical Institute

Mid-semester Examination 2017
M.Math 1st year/M.Stat 1st year

Course name: **Algebra/Abstract Algebra**
Date: **20 February 2017**
Maximum marks: **40**
Duration: **2 hours**

Note: There are 4 questions with total marks 45. Answer as many questions as you like. The maximum you can score is 40. This is closed notes/books exam.

1. Let H be a proper subgroup of G of index $i(H)$.
(a) Suppose H is a normal subgroup. Show that G is solvable if and only if both H and G/H are solvable. (4 marks)
(b) Suppose $|G| | i(H)!$, then show that G is not simple. (4 marks)
(c) Let

$$D_n = \{a^i b^j : 0 \leq i < n, 0 \leq j < 2, o(a) = n, o(b) = 2, ba = a^{-1}b\}$$

be the dihedral group. Show that D_n is solvable for all n. (5 marks)

2. A complex number is said to be *algebraic* if it is a solution of a polynomial equation with integer coefficients. Let $\mathbb{A}$ be the set of all algebraic numbers.
(a) Show that $\mathbb{A}$ is a field. (4 marks)
(b) Show that $\mathbb{A} = \mathbb{Q}^a$. (5 marks)

3. Let $f(x) = x^6 + x^3 + 1$. Suppose K is the splitting field of f in $\mathbb{C}$.
(a) Is f irreducible over $\mathbb{Q}$? Give reasons. (3 marks)
(b) Calculate the degree $[K : \mathbb{Q}]$. (4 marks)
(c) Give an example of a field on which f is not separable. (3 marks)

4. Let K be a normal extension of k. Let G be the group of automorphisms of K over k. Set

$$K^G = \{x \in K : gx = x \ \forall g \in G\}.$$

Let K_0 be the maximal separable subextension of k in K.
(a) Show that K^G is a purely inseparable extension of k, and K is a separable extension of K^G. (3+3 marks)
(b) Show that $K^G K_0 = K$ and $K^G \cap K_0 = k$. (3 marks)
(c) Show that any finite extension of $\mathbb{F}_{p^r}$ is separable. (4 marks)

INDIAN STATISTICAL INSTITUTE

Mid-Semeseter of 2nd Semester Examination : 2016-17

Course Name : **M.Stat. 1st Year**
Subject Name : **Sample Survey and Design of Experiments**
Date : Feb 21, 2017 **Total Duration: 2 hrs**

Note: Answer questions from both Groups A and B, using separate answer-booklets for the groups

Group A – Sample Survey. (Total Marks = 20)

Answer any four questions.

1. State and prove Godambe's (1955) theorem regarding the existence of uniformly minimum variance estimator for Y within the class of all homogenious linear unbiased estimators.

(5)

2. Define admissibility of an estimator within a class of estimators. Prove that Horvitz and Thompson's estimator (HTE) for Y is an admissible estimator within the class of all homogeneous linear unbiased estimators.

(5)

3. Prove that for a given sample s, if s^* denotes the reduced set equivalent to s obtained by ignoring the order and multiplicities of the units appearing in s, and if d^* denotes the data corresponding to s^*, then d^* is a minimal sufficient statistic.

(5)

4. Given any design p and an unbiased estimator t for Y depending on order and/or multiplicity of units in sample s, derive an improved estimator for Y through Rao-Blackwellization.

(5)

5. State Rao(1979)'s theorem on the generalized expression for mean squared error of a homogeneous linear estimator for Y. Clearly state the condition essential for this theorem.

   Derive the expression of variance of HT estimator of Y by applying this theorem. Obtain the necessary condition to be satisfied for this.

(5)

Group B: Design of Experiments

Total marks: 24, Maximum you can score is 20.

Answer all questions. Keep your answers brief and to the point.

1. Let d be a connected, equireplicate and proper design with v treatments arranged in b blocks. A new design d^* is constructed from d by augmenting each block of d by all the v treatments applied once each. Thus, d^* also has v treatments and b blocks, but each block of d^* has size $k + v$. (a) If d is balanced, will d^* be balanced? (b) If d is orthogonal, will d^* be orthogonal? Justify your answers. [3+3=6]

2. Given a block design d with v treatments and b blocks, define its dual design d^* as a design with b treatments and v blocks such that if treatment i occurs in block j of design d, then treatment j occur in block i of d^*, for every i, j where $i = 1, \ldots, v$, $j = 1, \ldots, b$. Now, if d is a BIB design, will d^* necessarily be also a BIB design? Justify your answer. [6]

3. A factory has 6 machines and 9 operators, all machines are to be operated each day, employing one worker per machine. Suggest a plan of assignment of these 9 workers to the machines for 12 days, such that all operators work for an equal number of days and every two workers are assigned to work on the same day a constant number of times. [6]

4. Let $\mathcal{D}$ be the class of all block designs with v treatments and b blocks, each of size k.

   Write down the usual additive model for analyzing designs in $\mathcal{D}$.

   Suppose d^* is a BIB design in $\mathcal{D}$. Show that under this model d^* minimizes, over $\mathcal{D}$, the determinant of the variance-covariance matrix of the BLUE of any full set of orthonormal treatment contrasts. [2+4=6]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination : 2016-17
M.STAT. I YEAR, M.S.(QE) I and II YEAR
Optimization Techniques

Date: 22 February 2017 Maximum Marks: 75 Duration: $2\frac{1}{2}$ hours

---

Notation have usual meaning.

This paper carries 85 marks. However, maximum you can score is 75.

---

1 Two cities generate waste and their waste are sent to incinerators (furnaces) for burning. Daily waste production and distances among cities and incinerators are as below:

	Waste produced (ton/day)	Distance to incinerator (in km.)	
		A	B
City 1	500	30	20
City 2	400	36	42

Incineration reduces each ton of waste to 0.2 tons of debris, which must be dumped at one of the two landfills. It costs \$3 per kilometer to transport a ton of material (either waste or debris). Distances (in km) among the incinerators and landfills are given below.

	Capacity (ton/day)	Incineration Cost (\$/ton)	Distance to landfill (in km.)	
			Northern	Southern
Incinerator A	500	40	35	38
Incinerator B	600	30	51	48

Formulate a linear program that can be used to minimize the total cost of disposing waste of the cities. [15]

2 Let $\bar{\mathbf{x}}$ be an extreme point of a convex set $S(\subseteq R^n)$. Then $\bar{\mathbf{x}}$ lies on the boundary of S. [8]

3 Find all the extreme points and extreme directions of the polyhedral set given by:

$$\begin{aligned} x_1 + 2x_2 + x_3 &\leq 10 \\ -x_1 + 3x_2 \quad &= 6 \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

[12]

[P.T.O.]

4 Consider a maximization linear programming problem with extreme points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$, and extreme directions $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_3$, and with an objective function having cost vector $\mathbf{c}$ such that $\mathbf{c}^T\mathbf{x}_1 = 5$, $\mathbf{c}^T\mathbf{x}_2 = 7$, $\mathbf{c}^T\mathbf{x}_3 = 4$, $\mathbf{c}^T\mathbf{x}_4 = 7$, $\mathbf{c}^T\mathbf{d}_1 = 0$, $\mathbf{c}^T\mathbf{d}_2 = -3$ and $\mathbf{c}^T\mathbf{d}_3 = 0$. Is this problem unbounded or has an optimal solution? Hence, depending upon your answer, give either the set of all rays or the set of all optimal solutions. [8]

5 The following is the current simplex tableau (second iteration) of a given linear programming problem in canonical form with the objective to maximize $2x_1 - 3x_2$. The two constraints are of $\leq$ type with non-negative right-hand-sides. In the tableau, x_3 and x_4 are slack variables.

	x_1	x_2	x_3	x_4	RHS
z	b	1	f	g	6
x_3	c	0	1	$\frac{1}{5}$	4
x_1	d	e	0	2	a

Find the unknowns a through g above. [12]

6 Suppose that simplex algorithm terminates declaring Big-M problem as unbounded with all the artificial variables are at zero level in the current solution. Show that the original problem of interest is unbounded. [10]

7 State and prove the complementary slackness theorem. [10]

8 Consider the linear program (P): min $\mathbf{c}^T\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}_1$, $\mathbf{x} \geq 0$. Suppose that $\mathbf{u}_1$ is an optimal solution of P and $\mathbf{v}_1$ is optimal for its dual. Further, let $\mathbf{u}_2$ is an optimal solution of P when $\mathbf{b}_1$ is changed to $\mathbf{b}_2$. Show that

$$\mathbf{v}_1^T(\mathbf{b}_2 - \mathbf{b}_1) \leq \mathbf{c}^T(\mathbf{u}_2 - \mathbf{u}_1).$$

[10]

——*** xXx ***——

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination 2016-2017

M. Stat - First year

Metric Topology and Complex Analysis

Date: February 22, 2017 Maximum Marks: 40 Duration: 2:30 hours

Answer all questions.

For full credit, you have to state the theorems/results you use.

(1) Let (X, d) be a locally connected metric space and $\Omega \subset X$ is an open set.

(a) Show that connected components of Ω are clopen subsets of X.

(b) If $X = \mathbb{R}^n$ with the usual metric then show that Ω has at most countably many components.

(c) If $X = \mathbb{C}$ with the usual metric and Ω is an open subset of $\mathbb{C}$, then using Cauchy-Riemann equation show that the image of a non-constant holomorphic function $f : \Omega \to \mathbb{C}$ cannot be a parabolic arc (of finite or of infinite length).

6+4+6=16

(2) Let $\gamma : [0, 1] \to \mathbb{C}$ be a closed curve on $\mathbb{C}$. Show that, $\mathbb{C} \setminus \gamma([0, 1])$ has exactly one unbounded component. 5

(3) Let (X, d) be a complete metric space.

(a) Show that an open ball B in X is of second category.

(Hint: Consider the boundary ∂B.)

(b) Let $\{A_k\}$ be a sequence of closed sets in X. Suppose that $\cup_k A_k$ contains an open ball B. Then show that at least one of the A_k contains an open ball.

4+3

P. T. O.

(4) For $j \in \mathbb{N}$, define open disc and circle respectively of radius j by

$$D_j = \{z \in \mathbb{C} \mid |z| < j\} \text{ and } C_j = \{z \in \mathbb{C} \mid |z| = j\}.$$

Give brief answers, 1-2 lines of the main points to these questions. Draw diagram if necessary.

(a) Let f be a holomorphic function on $\Omega = D_5 \setminus \overline{D_1}$. Then prove or disprove

$$\int_{C_2} f = \int_{C_3} f.$$

(b) Find $\int_{C_3} \frac{1}{(\xi - z)^2} d\xi$ and $\int_{C_3} \frac{1}{\xi - z} d\xi$ for a point $z \in D_3$ using Cauchy's integral formulas.

(c) Find $\int_{C_3} \frac{1}{(\xi - z)^2} d\xi$ and $\int_{C_3} \frac{1}{\xi - z} d\xi$ for a point $z \notin D_4$.

(d) Solve (b) without using Cauchy's integral formula.

$4 \times 4 = 16$

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2016-2017
M. Stat. 1st Year. 2nd Semester
Large Sample Statistical Methods

Date: February 23, 2017 Maximum Marks: 40 Duration: 2 hours

- Answer all the questions.
- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Consider a sequence of random variables $\{X_n : n \geq 1\}$ such that $X_n \sim$ Poisson$(1/n)$. It is easy to prove that $X_n \xrightarrow{P} 0$. Does $X_n \xrightarrow{a.s} 0$ necessarily? Justify your answer. [6]

2. Consider i.i.d. observations $X_i,\ i \geq 1$, from a distribution with finite second moment. Let $\mu := \mathrm{E}(X_1)$. While testing $\mathrm{H}_0 : \mu = \mu_0$ against $\mathrm{H}_1 : \mu > \mu_0$ one uses the following test: "*Reject H_0 if* $\sqrt{n}(\bar{X}_n - \mu_0)/S_n > t_{n-1,\alpha}$", where $\bar{X}_n := (1/n)\sum_{i=1}^n X_i$, $S_n^2 := (1/n)\sum_{i=1}^n (X_i - \bar{X}_n)^2$, and $t_{n-1,\alpha}$ is the $(1-\alpha)$-th quantile of a t-distribution with $n-1$ degrees of freedom. He claims that this test is asymptotically of level α. Examine, with adequate reasons, if this claim is valid. [8]

3. Suppose $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. random variables all having the same mean μ and variance σ^2. Define the autoregressive scheme $X_n = \beta X_{n-1} + \epsilon_n$, for $n = 1, 2, 3, \ldots$, where $X_0 \equiv 0$ and $-1 < \beta < 1$. Let $\bar{X}_n := (1/n)\sum_{i=1}^n X_i$. Also, let $\theta := \mu/(1-\beta)$, $\tau^2 := \sigma^2/(1-\beta)^2$. Show that $\bar{X}_n$ is A.N.$(\theta, \tau^2/n)$. [9]

4. Suppose $X_i,\ i \geq 1$, are i.i.d. observations from double exponential distribution with pdf $f(x) := (1/2)\exp(-\mid x \mid)$, $x \in \mathbb{R}$. Find the asymptotic correlation between sample mean and sample median (after suitable centering and scaling). [8]

5. Consider a sequence of i.i.d. random $k \times 1$ vectors $\{\boldsymbol{X}_i : i \geq 1\}$ such that $\boldsymbol{X}_i \sim$ Multinomial$(1; p_1, \ldots, p_k)$. We wish to test the hypothesis $\mathrm{H}_0 : p_i = p_{i,0}$ for $i = 1, \ldots, k$. Let $(n_1, \ldots, n_k) := \sum_{i=1}^n \boldsymbol{X}_i$ and consider the following test:

$$\text{Reject H}_0 \text{ if } T_n \overset{def}{=} 4n\sum_{i=1}^k \left(\sqrt{n_i/n} - \sqrt{p_{i,0}}\right)^2 \text{ is large.}$$

   Find the asymptotic null distribution of T_n. [9]

Indian Statistical Institute
Midterm Examination
Second Semester, 2016-2017 Academic Year
M.Stat. First Year
Resampling Techniques

Date: 24 February, 2017 Total Marks : 35 Duration: 2 Hours

Answer all questions

1. (a) Write down the (delete-1) Jackknife estimator of bias of an estimator and the (delete-1) Jackknife bias-adjusted estimator. Is this estimator expected to reduce bias in moderate and large sample sizes ? Justify your answer. [1+1+3=5]

   (b) What is the (delete-1) Jackknife estimator of variance of an estimator ? What was Tukey's inuition behind proposing this estimator ? Give an example where this estimator is inconsistent and prove your assertion. Is there any way to get rid of this inconsistency ? Justify your answer by stating appropriate results. [1+1+5+3=10]

2. (a) Let $X_1, \ldots, X_n$ be iid following a distribution F and we want to approximate the sampling distribution of $T(X_1, \ldots, X_n, F)$ in large samples, where $T()$ is real valued function depending on the sample and F. Give an example of F and $T()$ such that the usual Bootstrap may not be suitable for the aforesaid purpose. Prove your answer. Can the usual Bootstrap be modified be get rid of this difficulty ? Justify your answer by stating one relevant result. [4+2=6]

   (b) Suppose you have obtained an iid sample $X_1, \ldots, X_n$ from a certain distribution F and let $T_n = \sqrt{n}(\bar{X}_n - \mu)$, where $\mu = E_F(X_1)$. Stating appropriate assumptions, prove strong consistency of the Bootstrap (with respect to Kolmogorov metric) in approximating the distribution of T_n as $n \to \infty$. [10]

   (c) Can the Bootstrap approximation to distributions be more accurate than the Central Limit Theorem in moderate and large sample sizes ? Explain your answer. [4]

MSTAT I - Measure Theoretic Probability

Midsem. Exam. / Semester I 2016-17

Time - 2 hours 15 mins/ Maximum Score - 30

Date : 27.02.17

NOTE : SHOW ALL YOUR WORK TO GET THE FULL CREDIT. RESULTS USED MUST BE CLEARLY STATED.

1. (4+4+4+4=16 marks) Write TRUE or FALSE and justify clearly.

   (a) Set of point of discontinuities of a right continuous function (from $\mathcal{R}$ to $\mathcal{R}$) must be measurable.

   (b) A closed set $D \subset \mathcal{R}$ with positive measure must contain a nonempty open interval I.

   (c) Let $\{f_n\}$ be a sequence of real-valued measurable function on $\mathcal{R}$, such that, $f_n \uparrow f$. Assume, $\int_{\mathcal{R}} f_n d\mu$ exists for all n and also $\int_{\mathcal{R}} f d\mu$ exists, where μ is a probability. Then $\int_{\mathcal{R}} f_n d\mu \to \int_{\mathcal{R}} f d\mu$.

   (d) Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let X be a random variable on this space with finite expectation (i.e., $X : \Omega \mapsto \mathcal{R}$ be measurable function and X integrable w.r.t. P). If for some reals a, b, and for all $A \in \mathcal{A}$, $a \leq \int_A X dP \leq b$ then $a \leq X \leq b$, almost surely $[P]$.

2. (4+4=8 marks)

   (a) Show that for any Borel measurable set $A \subset \mathcal{R}^2$, the map, $x \mapsto \lambda_1(A^x)$ is a Borel-measurable function. where λ_1 is the one-dimensional Lebesgue measure and $A^x = \{y : (x, y) \in A\}$.

   (b) Suppose $A \subset \mathcal{R}^2$, is a Lebesgue measurable would the map $x \mapsto \lambda_1(A^x)$ be a Lebesuge-measurable function? Justify your answer.

3. (4+5=9 marks)

   (a) Let $\{f_n\}$ be a sequence of real-valued Borel measurable functions on $(\Omega, \mathcal{F}, \mu)$ where μ is a σ-finite measure. If $\sum_{n=1}^{\infty} \int_{\Omega} |f_n| d\mu < \infty$ then, show that $\sum_{n=1}^{\infty} f_n$ is absolutely convergent $a.e.$ $[\mu]$ and integrable, and $\int_{\Omega} \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_{\Omega} f_n d\mu$.

   (b) Let $f_n(s) = s^n - 2s^{2n-1}$ for $s \in [0, 1]$ and zero elsewhere. Show that, $\sum_{n=1}^{\infty} f_n$ is absolutely convergent $a.e.$ $[\mu]$ where μ is the Lebesgue measure. Calculate $\int_0^1 \sum_{n=1}^{\infty} f_n d\mu$, $\sum_{n=1}^{\infty} \int_0^1 f_n d\mu$ and compare and comment in the light of part (a).

All the best.

Indian Statistical Institute
Semestral Examination Second Semester (2016-2017)
M.Stat. First Year
Large Sample Statistical Methods

Maximum Marks: 60 Date : 21.04.2017 Duration :- $3\frac{1}{2}$ hours

Answer all questions

1. Suppose $\{X_n : n \geq 1\}$ is a sequence of i.i.d. random variables with X_1 having pdf given by
$$f(x) = K\frac{x^2}{1+x^4},\ x \in \mathbb{R},$$
where $K > 0$ is a suitable constant. Let $Y_n := X_n/n$. Decide, with adequate reasons, if $Y_n \xrightarrow{a.s.} 0$. [4]

2. Suppose $\{X_n : n \geq 1\}$ is a sequence of i.i.d. random variables with X_1 having pdf given by
$$f(x) = \begin{cases} \phi(x) & \text{if } x \leq 0, \\ 2\phi(2x) & \text{if } x > 0, \end{cases}$$
where ϕ is the pdf of standard normal distribution. Find the asymptotic distribution of the sample median. [4]

3. (a) Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with a common distribution function F given by $F(x) = 1 - \frac{1}{x^\theta}$ for $x \geq 1$, where $\theta > 0$. Does the distribution of the sample maximum $X_{(n)}$, suitably scaled and normalized, converge to a non-degenerate disribution asymptotically ? Prove your answer. [5]

   (b) Suppose $X_1, \ldots, X_n$ are i.i.d. having the standard normal distribution. Derive the asymptotic distribution of $X_{(n)}$, the sample maximum, after appropriate centering and scaling. [8]

4. Suppose $X_1, \ldots, X_n$ are i.i.d. with common density $f(x, \theta)$, where $\theta \in \Theta$, Θ consisting of *only finitely many real numbers.* Assume also that the density under each θ has the same support and that the distributions under different θ's are different. If θ_0 is the true value of θ, prove that with probability tending to 1 (under θ_0) as $n \to \infty$, the likelihood will be maximized at the value $\theta = \theta_0$. [4]

5. Suppose $X_1, \ldots, X_n$ are i.i.d. with density $f(x) = \frac{1}{\sigma}e^{-(x-\mu)/\sigma}I_{x\geq\mu}$. Show that the maximum likelihood estimators of μ and σ have different limiting distributions, assuming that the true value of μ is 0 and that of σ is 1. [6]

6. Suppose $X_1, \ldots, X_n$ are i.i.d. with density $f(x, \theta)$ where $\theta \in \Theta$, Θ being an open interval on $\mathbb{R}$. Assume the regularity conditions needed to prove asymptotic normality of a consistent sequence of roots of the likelihood equation. Suppose now that $\hat{\theta}_n$ is a consistent sequence of roots of the likelihood equation and θ_0 is the true value of the parameter θ. Prove that, under θ_0, with probability tending to 1 as $n \to \infty$, $\hat{\theta}_n$ is a local maximum of the likelihood equation. [5]

7. Consider a sequence of i.i.d. random $k \times 1$ vectors $\{\boldsymbol{X}_i : i \geq 1\}$ such that $\boldsymbol{X}_i \sim$ Multinomial$(1; p_1, \ldots, p_k)$, where $p_i > 0$ for each i and $p_1 + \cdots + p_k = 1$. Let $(n_1, \ldots, n_k) := \sum_{i=1}^n \boldsymbol{X}_i$. Also, for $n \geq 1$, let

$$T_n^{(1)} := \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}, \quad T_n^{(2)} := \sum_{i=1}^k \frac{(n_i - np_i)^2}{n_i}.$$

If some of the n_i's are zero, you may redefine the corresponding terms in the sum in $T_n^{(2)}$ as zero. Show that $T_n^{(1)} - T_n^{(2)} \xrightarrow{P} 0$. [7]

8. (a) State and prove Walker's result about the lower bound on the asymptotic variance of CAN estimators. [4]

   (b) Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta. 1)$ where $\theta \in \mathbf{R}$. Let T_n denote the Hodges' estimator that estimates θ by half of the sample mean if the sample mean lies in $[-n^{-1/4}, n^{-1/4}]$ and by the sample mean otherwise. How does this estimator violate the condition(s) assumed for proving Walker's result ? [3]

9. (a) Sketch the main idea of the proof of asymptotic normality of U-Statistics. [4]

   (b) Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) and $s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ where $\bar{X}_n$ is the sample mean. Prove that for any $p \in (0,1)$, s_n^2 has a non-degenerate limit distribution (after appropriate centering and scaling). Identify the limit distribution and the centering and scaling constants corresponding to each p. [6]

INDIAN STATISTICAL INSTITUTE
Semester Examination 2016-17

Course Name: M.STAT 1st year
Subject Name: Sample Surveys and Design of Experiments
Date: April 24, 2017 Maximum Marks: 60 Duration: 3 hours
Note: Answer questions from both Group A and Group B below. Answers from each Group are to be written in separate answer-booklets.

Group A: Sample Surveys Total marks: 30.
Answer all questions. Notations are as usual.

1. State Murthy (1957)'s unbiased estimator of Y based on PPSWOR sampling scheme. Derive its variance and variance estimator. [5]

2. State Ðurbin (1967)'s method of unequal probability sampling scheme for sample size n = 2 and show that it is an IPPS scheme. [5]

3. State how the double sampling approach can be useful in ratio method of estimation of Y. Examine the nature of the estimator, biased or unbiased. Also derive its MSE. [5]

4. Show how Politz and Simmon's 'Not-at-home' technique can be used to estimate $\bar{Y}$ under SRSWR, and estimate its error. [5]

5. Show how the population mean of a sensitive quantitative variable can be estimated by a general sampling design with p(s) being the selection probability of a sample s of respondents. Obtain its variance and variance estimator. [5]

6. Determine how many clusters need to be selected for a sample survey that will be conducted by a one-stage cluster sampling to estimate a population proportion of persons having a particular disease, subject to the following given information.

   1. The population size is 387.
   2. We may tolerate a small risk, say, $\alpha = 5\%$ for facing the absolute deviation in between the estimate and the unknown proportion greater than 0.1.
   3. Clusters are more or less of same size, say, 15.
   4. Clusters will be selected by SRSWOR.
   5. The units within clusters are somewhat similar, though the degree of similarity is low based on some prior knowledge, say, $\rho = 0.07$. [5]

Group B: Design of Experiments

Total marks: 34, Maximum you can score is 30.

Answer all questions. Keep your answers brief and to the point.

1. Answer True/False to the following statements and justify your answers.

   [$3 \times 4 = 12$]

   (a) A Balanced Incomplete Block (BIB) design with parameter values $v = b = 22, r = k = 7, \lambda = 2$ exists.

   (b) A BIB design with parameter values $v = b = 31, r = k = 15, \lambda = 7$ does not exist.

   (c) In an equireplicate 2^n factorial, the Best Linear Unbiased Estimators (BLUEs) of all main effects have the same variance.

   (d) A 2^3 experiment (treatment combinations ijk, $i, j, k = 0, 1$) which is designed with the following 2 blocks, does not confound any main effect with the blocks.

Block 1:	000, 101
Block 2:	001, 010, 100, 011, 110, 111

2. (a) Define a Youden Square design.

   (b) An experimenter initially plans an experiment as a block design with $v = b = 20, r = k = 8$, and constructs a design d_1. Subsequently, she decides to instead use a row-column design d_2, with 20 treatments, 8 rows and 20 columns, with the blocks of d_1 written as the columns of d_2. She wonders if it is possible to re-arrange the treatments within each column of d_2 so that each row of d_2 is a block of a Randomized Block Design (RBD).

   Which **one** of the following options will be correct? Justify your choice.

   (i) Irrespective of the choice of d_1, the above rearrangement in d_2 will **always** be possible.

   (ii) Irrespective of the choice of d_1, the above rearrangement in d_2 will **never** be possible.

   (iii) The actual choice of d_1 will determine whether the above rearrangement in d_2 will be possible or not. [2+6=8]

3. Consider a 2^2 factorial experiment with the four treatment combinations written as $00, 01, 10, 11$. The adopted model is:

$$E(y_{00}) = 0,\ E(y_{01}) = \beta,\ E(y_{10}) = \alpha,\ E(y_{11}) = \alpha + \beta,$$

where, $y_{00}, y_{01}, y_{10}, y_{11}$ are typical observations arising from the treatment combinations $00, 01, 10, 11$, respectively; all observations being independent and homoscedastic. With a total of N observations, it is intended to find the replication numbers of the 4 treatment combinations so as to minimize the determinant of the BLUEs of α and β. Find such a choice of the replication numbers when (a) $N = 8$ and (b) $N = 12$.

[Note that the 4 replication numbers must add up to N.] [8]

4. (a) Give an expression for a full set of orthonormal contrasts belonging to the 2-factor interaction effect of the first two factors in a $2 \times 3 \times 2$ factorial experiment. Justify why this set is indeed a full set.

(b) For a 3^4 factorial experiment, obtain an orthogonal main effect plan in 9 runs.

[3+3=6]

Indian Statistical Institute

Second Semestral Examination 2016-17

M. Stat. I yr

Resampling Techniques

Date: April 28 , 2017 | Maximum marks: 50 | Duration: 3 hrs.

Answer all questions. Use separate answerscripts for different groups

Group A

1 (a) State Miller's Theorem about consistency of (delete-1) Jackknife Variance estimators for smooth functions of sample mean. Give a concise proof of this result highlighting the main steps.

(b) Why would one use the Jackknife in this setup instead of more traditional methods of estimation ?

[(2+5)+ 2 = 9]

2. Suppose $X_1, \ldots, X_n$ are iid having distribution F and let $H_n(x) = P_F(\sqrt{n}(\bar{X}_n - \mu) \leq x)$, where $\mu = E_F(X_1)$. Motivate the definition of $H_{\text{Jack}}(x)$, the Jackknife Histogram estimator of $H_n(x)$. State a result about consistency of the Jackknife Histogram with respect to the Kolmogorov metric.

[4+2 =6]

3. Suppose $X_1, \ldots, X_n$ are iid having distribution F with finite second moments. Sketch the main steps in the proof of strong consistency of Bootstrap (with respect to Mallows metric) for estimating the distribution of $\sqrt{n}(\bar{X}_n - \mu)$, where $\mu = E_F(X_1)$.

[6]

Group B

4. Consider a multiple regression model

$$y_i = x_i^t \beta + \epsilon_i,$$

for $i = 1, 2, \ldots, n$ where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are iid from some distribution F with finite fourth moment. Let X_n denote the design matrix and it is assumed that $n^{-1} X_n' X_n \to Q$ as $n \to \infty$ where Q is a positive definite matrix. Describe a suitable bootstrap technique (of your choice) to estimate sampling distribution of the OLS estimator, $\hat{\beta}_{n,ols}$, and provide justifications (each statement should be provable) why your methodology should be consistent in large samples for this regression model.

[4+6=10]

5. Suppose $X_1, X_2, \ldots, X_n$ are iid samples from some distribution F with finite moments. Let $\bar{X}_n^*$ denote naive bootstrap sample mean given the data. Use Efron's jackknife-after-bootstrap technique to provide an estimate of the accuracy $E_F E^* \left(\bar{X}_n^* - \bar{X}_n\right)^2$ where E^* denotes expectation under bootstarp scheme given the data.

[7]

6 (a) Discuss three methods of bootstrapping from a stationary time series data stating the conditions under which these bootstrap methods are consistent in the long term.

(b) For an m-dependent stationary time series data describe suitable bootstrap method to obtain a confidence interval of the autocorrelation at lag one, namely $\hat{\rho}(1)$.

[4+8 =12]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination : 2016-17
M.STAT. I YEAR, M.S.(QE) I and II YEAR
Optimization Techniques

Date: 2. 5. 2017 Maximum Marks: 75 Duration: 3 hours

This paper carries 85 marks. However, maximum you can score is 75.
Notations have usual meaning.

---

1 Consider the following linear program:

$$\begin{aligned} &\max \quad 4x_1 + 5x_2 + 9x_3 + 11x_4 \\ &\text{subject to} \\ &\quad x_1 + x_2 + x_3 + x_4 \le 15, \\ &\quad 7x_1 + 5x_2 + 3x_3 + 2x_4 \le 120, \\ &\quad 3x_1 + 5x_2 + 10x_3 + 15x_4 \le 100, \\ &\quad x_1, x_2, x_3, x_4 \ge 0. \end{aligned}$$

The final (optimal) simplex tableau of this problem is presented below, where x_5, x_6 and x_7 are slack variables.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
z	0	$\frac{3}{7}$	0	$\frac{11}{7}$	$\frac{13}{7}$	0	$\frac{5}{7}$	$\frac{695}{7}$
x_1	1	$\frac{5}{7}$	0	$-\frac{5}{7}$	$\frac{10}{7}$	0	$-\frac{1}{7}$	$\frac{50}{7}$
x_6	0	$-\frac{6}{7}$	0	$\frac{13}{7}$	$-\frac{61}{7}$	1	$\frac{4}{7}$	$\frac{325}{7}$
x_3	0	$\frac{2}{7}$	1	$\frac{12}{7}$	$-\frac{3}{7}$	0	$\frac{1}{7}$	$\frac{55}{7}$

(a) Identify the optimal basis and its inverse from the above.

(b) Find the range of θ in order to ensure that the basis in (a) remains optimal in the following two situations: (i) $c_1 = 4$ is changed to $c_1 = 4+\theta$, and (ii) $b_2 = 120$ is changed to $b_2 = 120 + \theta$.

[4+(6+3) = 13]

[P.T.O.]

2 Consider the linear program: $\min c^{\mathrm{T}}x$ subject to $Ax = b$. Show that it is either infeasible, unbounded or all feasible solutions are optimal.

[10]

3 Solve the following optimization problem:

$$\begin{aligned} \min \quad z = {} & x_1 + 2x_2 + \ldots + nx_n \\ \text{subject to} & \\ & x_1 \geq 1, \\ & x_1 + x_2 \geq 2, \\ & \vdots \\ & x_1 + x_2 + \ldots + x_n \geq n, \\ & x_1, x_2, \ldots, x_n \geq 0. \end{aligned}$$

[10]

4 (a) State the balanced transportation problem using matrix notation. Write down its dual.

(b) Consider the transportation problem with data: $m = 2$, $n = 3$, $s = [3, 5]^{\mathrm{T}}$, $d = [1, 3, 4]^{\mathrm{T}}$, and the cost matrix

$$C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 4 & 4 \end{bmatrix}.$$

Prove by duality theory that $[x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}] = [0, 3, 0, 1, 0, 4]$ is an optimal solution.

[(3+2)+5 = 10]

5 Prove that the coefficient matrix $A_{(m+n)\times mn}$ of transportation problem is totally unimodular.

[10]

6 Let $a_i, b_j \in \mathbb{R}$ (for $i, j = 1, 2, \ldots, m$) be given constants. Solve the assignment problem having cost matrix $C = ((c_{ij}))$, where $c_{ij} = a_i + b_j$ for all i, j.

[10]

7 Consider the Hungarian method for solving assignment problem. State the *labelling algorithm* to draw the minimum number of lines to cover all zeros in a reduced cost matrix, and give a proof for its fitness.

[4+8 = 12]

[P.T.O.]

8 Re-formulate the following optimization problem as an equivalent mixed integer linear programming problem:

$$
\begin{aligned}
&\min\ x_1 - x_2 \\
&\text{subject to} \\
&\ x_1 + x_2 \leq 4, \\
&\ x_1 \geq 1 \ \text{ or } \ x_2 \geq 1 \quad (\text{but not both } \ x_1 > 1,\ \ x_2 > 1), \\
&\ x_1, x_2 \ \geq \ 0.
\end{aligned}
$$

[10]

——*** xXx ***——

INDIAN STATISTICAL INSTITUTE

Semestral Examination 2016-2017

M. Stat - First year

Metric Topology and Complex Analysis

Date: May 2, 2017 | Maximum Marks: 60 | Duration: 3 hours

Answer all questions.

For full credit, state the theorems/results you use.

(1) Let $\Omega \subset \mathbb{C}$ be an open and connected set and $a \in \Omega$. Suppose that f is a holomorphic function on Ω, $f(a) \neq 0$ and $|f(z)| \geq |f(a)|$ for all $z \in \Omega$. Show that f is constant. 8

(2) Suppose that for a function $f : \mathbb{D} \to \mathbb{C}$ both f^2 and f^3 are holomorphic on $\mathbb{D}$. Show that f is holomorphic on $\mathbb{D}$. Here the functions f^2 and f^3 are defined by $f^2(z) = f(z)^2$ and $f^3(z) = f(z)^3$. 8

(3) Let f be a holomorphic function on the upper half plane $\mathbb{H} = \{z \in \mathbb{C} \mid \Im z > 0\}$ such that $|f(z)| < 1$ for all $z \in \mathbb{H}$ and $f(i) = 0$. Show that $|f(2i)| \leq 1/3$.
(Hint: Transfer the problem to $\mathbb{D}$ by composing with the standard conformal map $\mathbb{D} \to \mathbb{H}$ and use Schwarz lemma.) 10

(4) Let f be a holomorphic function on a connected open set $\Omega \subset \mathbb{C}$. Suppose that f has exactly two zeros: a simple zero at z_1 and a zero of order 2 at z_2. Show that

$$f(z) = (z - z_1)(z - z_2)^2 h(z) \qquad \forall z \in \Omega$$

where h is a holomorphic function on Ω which does not vanish on Ω. 10

P. T. O.

(5) Let f be a holomorphic function on an open set G and it has a simple zero at $z_0 \in G$. Then show that $1/f$ has a simple pole at z_0 and the residue of $1/f$ at z_0 is $1/f'(z_0)$. 4+4

(6) Suppose that f is an entire function which is also injective. Show that f does not have essential singularity at ∞. 10

(7) Use contour integration to find:

$$\int_{-\infty}^{\infty} \frac{\cos x}{x^2 + a^2} dx \text{ for } a > 0.$$

10

MSTAT I - Measure Theoretic Probability

Final Exam. / Semester II 2016-17

Time - 3:00 hours/ Maximum Score - 50

<u>NOTE</u> : SHOW ALL YOUR WORK. RESULTS USED MUST BE CLEARLY STATED.

1. ($5 \times 4 = 20$ marks)

   Write 'True' or 'False' and justify your answer.

   (a) Let A be a subset of $\mathbb{R}^2$ whose every x-crossection, $A_x = \{y : (x, y) \in A\}$, and every y-crossection, $A^y = \{x : (x, y) \in A\}$, are Lebesgue measurable in $\mathbb{R}$ with respect to one-dimensional Lebesgue measure. Then A must be Lebesgue measurable in $\mathbb{R}^2$ with respect to two-dimensional Lebesgue measure.

   (b) Let λ be the Lebesgue measure on $\mathbb{R}$ and let $A \subset \mathbb{R}$ be finite union of disjoint intervals, such that $\lambda(A) < \infty$. Then for all $\epsilon > 0$ there exists a bounded continuous function $g : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} |I_A - g| \, d\lambda < \epsilon$.

   (c) Every compact set in $\mathbb{R}$ without a nonempty interior must have Lebesgue measure zero.

   (d) Let Ω be an uncountable set and μ be a measure on the σ-field that consists of all countable and co-countable subsets of Ω, with $\mu(\Omega) = \infty$. Then μ cannot be σ-finite.

   (e) Let $\{X_n\}$ be a sequence of random variables with mean zero and variance one. Then $X_n \to 0$, a.s., implies $E|X_n| \to 0$.

2. (6+3+3=12 marks)

   Let $\{Y_n\}$ be i.i.d. random variables with $E[|Y_1|^{1/2}] < \infty$. For $n \geq 1$, denote $Z_n = Y_n I_{\{|Y_n| \leq n^2\}}$.

   (a) Prove that $\sum_{n=1}^{\infty} Var(Z_n/n^2)$ converges, and $\sum_{n=1}^{\infty} E(Z_n/n^2)$ converges.

   (b) Show that $(Y_1 + \cdots + Y_n)/n^2 \to 0$ almost surely, as $n \to \infty$.

   (c) Does there always exist a $\delta > 0$ such that $(Y_1 + \cdots + Y_n)/n^{2-\delta} \to 0$ almost surely, as $n \to \infty$? Justify your answer.

P.T.O

3. (6+6=12 marks)

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with finite variance. Let $S_n = X_1 + \cdots + X_n$ and $B_n^2 = Var(S_n)$.

Does $(S_n - E(S_n))/B_n$ converge to $N(0,1)$ in distribution, as $n \to \infty$ for the following $\{X_n\}_{n\geq 1}$? Justify your answer (for all cases).

(a) $P(X_n = n^\alpha) = \frac{1}{2n^2} = P(X_n = -n^\alpha)$, $P(X_n = 0) = 1 - \frac{1}{n^2}$ for some real constant $\alpha > 0$.

(b) X_n is a Uniform$(0, b_n)$ random variable with parameter $b_n > 0$, such that, $\sum_{n\geq 1} b_n = \infty$.

4. (6+6=12 marks)

Let $\{X_n\}$ be a sequence of independent random variables with mean zero. Let $S_n = X_1 + \cdots + X_n$.

(a) Show that, for all $\epsilon > 0$, $P(\max_{1\leq k\leq n} |S_k| \geq 4\epsilon) \leq 4 \max_{1\leq k\leq n} P(|S_k| \geq \epsilon)$.

(b) Assuming $S_n \to 0$ in probability, prove that, $S_n \to 0$, almost surely.

All the best.

MSTAT I - Measure Theoretic Probability

Back Paper Exam. / Semester II 2016-17

Time - 3:00 hours/ Maximum Score - 100

12·07·17

NOTE : SHOW ALL YOUR WORK. RESULTS USED MUST BE CLEARLY STATED.

1. (10+15=25 marks)

   (a) Let μ be a non-negative, finitely additive set-function on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$, such that $\mu(\mathbb{R}) < \infty$. Define $\mu_*(A) = \sup\{\mu(K) : K \subset A,$ compact set $\}$. Then show that μ_* is a measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$.

   (b) Let $\mathcal{C}$ be the class of sets in $\mathbb{R}^2$ of the form $\{(x, y) : y \in A$ for some $A \in \mathbb{B}(\mathbb{R})\}$. Show that $\mathcal{C}$ is a σ-field in $\mathbb{R}^2$ but two-dimensional Lebesgue measure is not σ-finite on $\mathcal{C}$. However, show that one dimensional measure $\mu_0(\mathbb{R} \times B) = \mu(B)$ is σ-finite on $\mathcal{C}$, where μ is the one-dimensional Lebesgue measure.

2. (10+10+5=25 marks)

   Let $\{h_n\}$, h be a sequence of Borel-measurable functions on $(\mathbb{R}, B(\mathbb{R}), \mu)$, such that, $h_n \to h$ in measure (with respect to the finite measure μ), as $n \to \infty$.

   (a) If $\{h_n\}$ is uniformly integrable, show that $h_n \to h$, in L_1, as $n \to \infty$.

   (b) If $\sup_n \int_{\mathbb{R}} |h_n|^2 d\mu < \infty$ then show that $h_n \to h$, in L_1, as $n \to \infty$.

   (c) Give an example in (a) where L_1 convergence fails in absence of uniform integrability.

3. (6+6+6+7=25 marks)

   (a) Let $\{a_n\}_{n\geq 1}$ be a sequence of reals and $\{b_n\}_{n\geq 1}$ be a sequence of nonnegative numbers. Define, $A_0 = 0 = B_0$ and for $n \geq 1$, $A_n = a_1 + \cdots + a_n$ and $B_n = b_1 + \cdots + b_n$. Show that $\sum_{k=1}^n a_k B_k = A_n B_n - \sum_{k=1}^n A_{k-1} b_k$.

   (b) Assume further in (a) that A_n converges to A (real number), and $B_n \to \infty$ as $n \to \infty$. Show that, $(1/B_n) \sum_{k=1}^n a_k B_k \to 0$ as $n \to \infty$.

P.T.O

(c) Let $\{X_k\}_{k\geq 1}$ be a sequence of independent random variables with $E(X_k) = 0$ and $\text{Var}(X_k) < \infty$. Assume further, $\sum_{k=1}^{\infty} \text{Var}(X_k)/k^2 < \infty$. Show that $E\left(\sum_{k=1}^{\infty}(X_k/k)\right)^2 < \infty$ and hence prove that $\sum_{k=1}^{n}(X_k/k) \to \sum_{k=1}^{\infty}(X_k/k)$ (a finite r.v.), almost surely, as $n \to \infty$.

(d) Let $\{X_k\}_{k\geq 1}$ be a sequence of independent random variables, defined as in (c). Define $S_n = X_1 + \cdots + X_n$, with $S_0 = 0$. Use (a), (b), (c) to prove that $S_n/n \to 0$ a.s. as $n \to \infty$. [Direct use of SLLN won't be acceptable!].

4. (5+5+5+5+5=25 marks)

Write TRUE or FALSE and justify by proving or disproving it.

(a) A right continuous function on the real line must have at most countably many discontinuities.

(b) Let $\{X_k\}$ be a sequence of independent random variables with mean zero and finite variance $\{\sigma_k^2\}$, such that

$$\max_{1\leq k\leq n} \frac{\sigma_k^2}{B_n^2} \to 0, \quad \text{as } n \to \infty,$$

where $B_n^2 = \sum_{k=1}^{n} \sigma_k^2$. Then $\frac{\sum_{k=1}^{n} X_k}{B_n} \to N(0,1)$ in distribution, as $n \to \infty$.

(c) Let $\{X_n\}$ be a sequence of random variables such that $X_n \to X$ in probability, as $n \to \infty$. Assume further that, $\sup_n E(X_n^2) < \infty$. Then, as $n \to \infty$, $X_n \to X$ in L_2.

(d) Let $f_n(s) = s^n - 2s^{2n-1}$ for $s \in [0,1]$ and zero elsewhere. Then, $\int_0^1 \sum_{n=1}^{\infty} f_n d\lambda = \sum_{n=1}^{\infty} \int_0^1 f_n d\lambda$ where λ is the Lebesgue measure.

(e) Let $\{f_n\}$, f be a sequence of Borel measurable functions on the line. Let $f_n \to f$ almost everywhere (with respect to the Lebesgue measure μ), as $n \to \infty$. Then $f_n \to f$ in measure (μ), as $n \to \infty$.

All the best.

Indian Statistical Institute
Backpaper
M.Stat. First Year
Second Semester, 2016-17 Academic Year
Large Sample Statistical Methods

14.7.17

Total Marks: 100 Duration :- 4 hours

Answer all questions

1. Suppose $X_n \sim$ Binomial(n, θ), where $0 < \theta < 1$. Let $\psi(\theta) := \theta(1-\theta)$ and $\hat{\psi}_n$ be the UMVUE of ψ. Find the asymptotic distribution of $\hat{\psi}_n$, suitably normalized, when $\theta = 1/2$. [10]

2. Suppose X_n converges in distribution to X where X_n and X are real valued random variables. Show that $g(X_n)$ converges in distribution to $g(X)$ where $g : \mathcal{R} \to \mathcal{R}$ is continuous. [10]

3. State and prove the Glivenko-Cantelli theorem. [15]

4. Consider a sequence of i.i.d. random variables $\{X_i : i \geq 1\}$, with X_1 having cdf F. Let $\xi_p := F^{-1}(p)$, where $0 < p < 1$. Based on $(X_1, \ldots, X_n)$, we construct the sample df F_n, and let $\hat{\xi}_{p,n} := F_n^{-1}(p)$. Suppose $F'(\xi_p)$ exists and is strictly positive. Show that

$$\hat{\xi}_{p,n} = \xi_p + \frac{p - F_n(\xi_p)}{F'(\xi_p)} + R_n,$$

where $\sqrt{n} R_n \xrightarrow{p} 0$. [20]

5. Consider a sequence of i.i.d. random $k \times 1$ vectors $\{\boldsymbol{X}_i : i \geq 1\}$ such that $\boldsymbol{X}_i \sim$ Multinomial$(1; p_1, \ldots, p_k)$. We wish to test the hypothesis $\mathrm{H}_0 : p_i = p_{i,0}$ for $i = 1, \ldots, k$, where $p_{i,0} > 0$ for each i and $p_{1,0} + \cdots + p_{k,0} = 1$. Let $(n_1, \ldots, n_k) := \sum_{i=1}^n \boldsymbol{X}_i$ and consider the following test:

$$\text{Reject } \mathrm{H}_0 \text{ if } T_n := \sum_{i=1}^k \frac{(n_i - np_{i,0})^2}{np_{i,0}} \text{ is large.}$$

Find the asymptotic null distribution of T_n. [15]

6. Give an example where the maximum likelihood estimator is inconsistent. Prove your answer. [10]

7. Stating appropriate assumptions, prove asymptotic normality of sequences of consistent roots of the likelihood equation (after appropriate centering and scaling). [20]