INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2013–2014) (Back Paper)

M. Stat First Year

Regression Techniques

Date: 28.7.14 Marks: 100 Duration: 3 hours

Attempt all questions

1. (a) Consider fitting a straight line $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i;\ i = 1, \ldots, n$. Suppose that $2\beta_0 - \beta_1 = 4$. Show the unrestricted and the restricted estimation spaces diagrammatically.

(b) Provide vectors that span the restricted estimation space and its complement.

(c) Consider the regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (assume that it contains an intercept), where $\boldsymbol{\beta}$ is a $p \times 1$ vector and $\mathbf{X}$ is $n \times p$ and has full rank. Consider testing $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{A}$ is $q \times p$ $(q < p)$, has full rank and does not involve the intercept β_0. Show geometrically that the F-statistic for testing this hypothesis can be written in terms of the difference between the R^2-statistics, where R^2 stands for the coefficient of determination.

Marks: 8+7+10=25

2. (a) Using the ridge regression set up establish the existence of biased estimators which have smaller risk than unbiased ones.

(b) Assume that there are infinitely many non-zero β_j's in the model $y_i = \sum_{j=1}^{\infty} \beta_j x_{ij} + \epsilon_i,\ i = 1, \ldots, n$, where $\epsilon_i; i = 1, \ldots, n$, are normal random errors, $\sum_{j=1}^{\infty} \beta_j^2 < \infty$ and, for each n, for $j, k = 1, \ldots, n$, $\sum_{i=1}^{n} x_{ij} x_{ik} = n\delta_{jk}$, where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. Let $\hat{\lambda}$ be the minimizer of $\hat{P}(\lambda) = n^{-1} RSS(\lambda) + 2n^{-1}\sigma^2 tr(S_\lambda)$ over $\Lambda_n = \{1, \ldots, a_n\}$ with $a_n \to \infty$ as $n \to \infty$. Show that $P(\hat{\lambda} \leq x) \to 0$ as $n \to \infty$ for any finite x.

Marks: 10+15=25

3. (a) Describe the Iteratively Weighted Least Squares (IWLS) method for fitting generalized linear models.

(b) Show that for canonical links IWLS reduces to the Newton-Raphson method.

(c) Under what condition is the IWLS method also equivalent to the method of moments approach?

Marks: 15+5+5=25

4. (a) Describe leverage values. What do they measure?

(b) Show that leverage values lie between 0 and 1.

(c) Discuss the role of principal components in detecting multicollinearity.

Marks: 5+10+10=25

MULTIVARIATE ANALYSIS

M Stat 1st Year (2013-14)

Backpaper Examination

Date: 11.08.14 Time: 2 hours 30 minutes

Total Marks: 100

1. (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 $A_j \sim W_p(n_j,\Sigma)$, $j = 1,2,\ldots,k$ and that if 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.

   (c) 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. (7+8+5+5=25)

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 distribution of $\frac{1-\Lambda_{p,1,n}}{\Lambda_{p,1,n}} \cdot \frac{n-p+1}{p}$ and $\frac{1-\sqrt{\Lambda_{2,m,n}}}{\sqrt{\Lambda_{2,m,n}}} \cdot \frac{n-1}{m}$ (8+7+5+5=25)

3. Derive likelihood ratio test for testing $H_0 : \underset{\sim}{\mu_1} = \underset{\sim}{\mu_2}$ against $H_1 : \underset{\sim}{\mu_1} \neq \underset{\sim}{\mu_2}$ based on two random samples of size n and m from $N_p(\underset{\sim}{\mu_1},\Sigma_1)$ and $N_p(\underset{\sim}{\mu_2},\Sigma_2)$ respectively, under the assumption of $\Sigma_1 = \Sigma_2 = \Sigma$, where Σ is unknown. Is there any condition that should be imposed on the sample sizes so that you can carry out your test? If so, explain it. Also derive union-intersection test for the same null hypothesis and check whether the two critical regions are same. Show that the test based on union-intersection principle is unbiased. (7+2+9+7=25)

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. (25)

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination – Semester I : 2014-2015
M.Stat. I Year
Measure Theoretic Probability

Date : 01.09.14 | **Maximum Score : 40** | **Time : 2 Hours**

Note : This paper carries questions worth a **total** of **50 MARKS**. Answer as many as you can. The **MAXIMUM** you can score is 40.

1. Suppose that $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ are three fields on a non-empty set Ω. Show that the class $\mathcal{S}$ consisting of all sets of the form $S = F_1 \cap F_2 \cap F_3$ where $F_i \in \mathcal{F}_i$ for $i = 1, 2, 3$, forms a semifield. Hence describe the smallest field on Ω containing $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$.

   (8+2)=**[10]**

2. Let $\mathcal{C}$ be a non-empty class of subsets of a non-empty set Ω. Show that $\sigma(\mathcal{C}) = \{A \subset \Omega : A \in \sigma(\mathcal{D}) \text{ for some countable } \mathcal{D} \subset \mathcal{C}\}$.

   [8]

3. Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing right-continuous function and let μ denote the Radon measure on $\mathcal{B}$ with distribution function F. Define G on $\mathbb{R}$ by $G(x) = F((x \vee (-2)) \wedge 7)$.
   (a) Show that G is non-decreasing and right-continuous.
   (b) Show that the Radon measure ν induced by G satisfies $\nu(B) = \mu(B \cap (-2, 7]) \ \forall \ B \in \mathcal{B}$.

   (6+6)=**[12]**

4. Show that the σ-field on $\mathbb{R}$ generated by the function $f(x) = x^5$ is the Borel σ-field.

   [8]

5. (a) Show that $\mathcal{A} = \{A \subset \mathbb{R} : \text{ either } A \text{ or } A^c \text{ is countable}\}$ is a σ-field on $\mathbb{R}$.
   (b) Show that if a function $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{A}$-measurable, then there must be a set $A \subset \mathbb{R}$ with A^c countable such that f is constant on A. [Hint : Try to first see that the set $\{a \in \mathbb{R} : f^{-1}(\{a\}) \text{ is countable}\}$ is non-empty and bounded above.]

   (4+8)=**[12]**

INDIAN STATISTICAL INSTITUTE

Mid-semestral Examination

M. Stat (1st Year): 2014-15

Subject: Applied Stochastic Processes

Date: 02-09-2014

Duration: 2hours　　　　　　　　　　　　　　　　　　　Full Marks: 40

Attempt all questions

1. a) Let $\{X_n\}_{n\geq 0}$ be a branching process with p.g.f. = $\phi(s) = p_0 + p_1 s + p_2 s^2 + \ldots\ldots$ and $p_0 > 0$, $p_0 + p_1 < 1$.

   If ϱ is the extinction probability show that ϱ is the smallest non-negative root of the equation $\phi(s) = s$.

   b) A mature individual produces off-spring according to p.g.f $\phi(s)$. Suppose that we have a population of k immature individuals each of which grows to maturity with probability p and then reproduces independently of other individuals. Show that p.g.f of total number of immature individuals at the beginning of next generation is $(1 - p + p\phi(s))^k$. [6 + 6]

2. Let in a culture there be two types of cells Red and White. At the end of each minute each red cell vanishes and produces 2 Red cells and 1 White cell with probability ⅓ and no cell with probability ⅔. Whereas at the end of each minute each White cell vanishes and produces 2 White cells and 1 Red Cell with probability ⅓ and no cell with probability ⅔. Let at the end of n^{th} minute X_n = number of White cells and Y_n = number of Red cells.

   Find expression for $E(X_n)$ when it is given that $X_0 \equiv 10$ and $Y_0 \equiv 0$. Hence find the limiting value of $E(X_n)$ as $n \to \infty$ and extinction probability of the culture.

   [4+4+4]

3. a) Give a short non-mathematical description of Epidemiology with reference to susceptible, infected, removed persons and latent period.

   b) Write a short note on simple deterministic epidemic. [6+10]

——— o ———

DESIGN OF EXPERIMENTS

M.Stat 1st Year

Total Marks : 50 — Time : $1\frac{1}{2}$ hours

Class notes & books are allowed

1. a) Give a <u>non trival</u> example of a design which is not connected. (5)

   b) Write down 2 mutually orthogonal latin squares of size 8 (15)

2. Consider the following BIBD. Yields are put with in parenthesis

Blocks	Treatments			
1	1 (13)	2 (7)	5 (4)	7 (12)
2	1 (15)	2 (11)	3 (6)	6 (11)
3	1 (13)	3 (3)	4 (4)	5 (3)
4	1 (15)	4 (7)	6 (11)	7 (16)
5	2 (8)	3 (3)	4 (4)	7 (12)
6	2 (10)	4 (6)	5 (7)	6 (10)
7	3 (4)	5 (3)	6 (8)	7 (12)

   a) Write down the incidence matrix of the design (5)

   b)Prepare a partial order of the seven treatments (25)

INDIAN STATISTICAL INSTITUTE
Mid- Semester Examination: 2014-15
M. Stat. I Year
Sample Survey

Date: 03/09/2014 **Maximum Marks: 25** **Duration: $1\frac{1}{2}$ Hours**

10:30 – 12:00 hours.

Assignments to be submitted on 03.09.2014, carry 5 marks.

N.B. Answer any 2 questions each carrying 10 marks.

1 Define Godambe's class of homogeneous linear unbiased estimators for a finite population total. How does a ratio estimator belong to this class ? Can you find a uniformly minimum variance estimator in this class? Proofs in details are needed.

2 Giving detailed proofs show how Yates & Grundy's estimator may be unbiased for the variance of Horvitz and Thompson's estimator for a finite population total. Find a sampling scheme for which this is uniformly non-negative.

3. Show how you may find it difficult to find the best member in the class of all linear unbiased estimators for a finite population total.

*****

INDIAN STATISTICAL INSTITUTE

M. Stat. 1st Year. Semester I. 2014—2015

Mid-term Examination

Large Sample Methods in Statistics

September 04, 2014 | Maximum Marks: 65 | Time: 2 and 1/2 hours

---

- Answer all the questions.
- You must state clearly any result stated (and proved) in class in order to answer a particular question.

---

1. Suppose $\{X, X_n : n \geq 1\}$ are defined over the probability space (Ω, A, P). Let $X_n(\omega) = \frac{1}{n}\ \forall\, \omega \in \Omega, X(\omega) = 0\ \forall\, \omega \in \Omega$.. Let F_n denote the cdf of X_n and F denote that of X.
   a) Verify if $\lim_{n\to\infty} F_n(t) = F(t)\ \forall\, t \in \mathbb{R}$.
   b) Comment on your finding. [4 + 3 = 7]

2. Suppose $X_n = O_p(1)$ and $Y_n = o_p(1)$. Let $Z_n := X_n Y_n$. Show that $Z_n = o_p(1)$. [7]

3. State a theorem that enables one to find the asymptotic distribution of a statistic, suitably centred and scaled, after decomposing the statistic into two parts, one being mean of i.i.d. random variables and the other being asymptotically negligible, in an appropriate sense. [You must state your assumptions clearly.] [4]

4. Suppose $\{X_n : n \geq 1\}$ is a sequence of i.i.d. observations from an $Exp(\theta, 1)$ distribution, $\theta \in \Theta := \mathbb{R}$. Denote by $\hat{\theta}_n$, the MLE of θ based on $X_1, \ldots, X_n$. Argue, for or against, with reasons, the following – without using a closed-form expression for MSE and bias in b) and c).
   a) The estimator $\hat{\theta}_n$ is strongly consistent for θ.
   b) For every $\theta \in \Theta$, the MSE of $\hat{\theta}_n$, as an estimator of θ, converges to 0 as as $n \to \infty$.
   c) For every $\theta \in \Theta$, the bias of $\hat{\theta}_n$, as an estimator of θ, converges to 0 as as $n \to \infty$. [9 + 5 + 4 = 18]

5. Suppose i.i.d. observations are being taken from a Multinomial$(1; \pi_1, \pi_2, \ldots, \pi_k)$ population. We wish to test the hypothesis $H_0: \pi_i = \pi_{i,0}\ \forall\, i$ against $H_1 : H_0$ is false, where $\pi_{i,0} > 0\ \forall\, i$. Suggest, with adequate reasons, a suitable test statistic for this problem. Find the asymptotic null distribution of your test statistic as the sample size goes to infinity. [4 + 10 = 14]

6. Consider the autoregressive scheme

$$X_n = \beta X_{n-1} + \varepsilon_n \text{ for } n = 1, 2, 3, \ldots,$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. with $E(\varepsilon_1) = \mu$ and $Var(\varepsilon_1) = \sigma^2$, $X_0 \equiv 0$, and $|\beta| < 1$. Find suitable real sequences $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$ with $b_n > 0\ \forall\, n$, such that $(\bar{X}_n - a_n)/b_n \to N(0,1)$ in distribution, where $\bar{X}_n := (X_1 + \cdots + X_n)/n$. [15]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2014-15

M. Stat. I Year
Statistical Inference I

Date: 05. 09. 14 **Full Marks: 60** **Time: 3 hours**

Answer all Questions

1. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from two Normal populations with means and variances (θ, σ^2) and (θ, γ^2) respectively. Find out the Bayes decision rule under squared error loss where $(\theta, \sigma^2, \gamma^2)$ are independent, θ having a $\mathrm{N}(0, \tau^2)$ distribution while (σ^2, γ^2) are iid with σ^{-2} having a Gamma (α, λ) distribution where α is the shape and λ is the scale parameters respectively.

   [20]

2. Consider a two person zero sum game where each player has two possible strategies 0 and 1 respectively. The set of strategies are given by mixed strategies on $\{0, 1\}$ fo rboth the players. The loss matrix on pure choices is given by $L(0,0) = 1, L(0,1) = 5, L(1,0) = 2$ and $L(1,1) = 0$. Find out an optimal pair of strategies for the game. What is the value of the matrix game?

   [20]

3. Consider two populations (assume disctrete if convenient) with $f(x, \theta)$ and $g(y, \eta)$ as p.m.f respectively. Assume that both θ and η are real valued parameters. Furthermore we are given two sets of training samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ respectively from the two populations (to be indicated by 1 and 2).

   Next let k iid test samples $(Z_1, I_1), (Z_2, I_2), \ldots, (Z_k, I_k)$ are given where I_j denotes *unobserved* the indicator of the population (1 or 2). Conditional on $I_j = 1$, Z_j has pmf $f(\cdot, \theta)$; while conditional on $I_j = 2$, Z_j has pmf $g(\cdot, \eta)$ for every $j = 1, 2, \ldots, k$. The set of possible pure actions are given by $\{\mathbf{a} = (a_1, a_2, \ldots, a_k) : a_i = 1, 2 \text{ for } i = 1, 2, \ldots, k\}$.

   Note that $\epsilon = (\theta, \eta, \Pr\{I_1 = 1\})$ define unknown consequences as ϵ varies over suitable range defined above . Finally we define the loss function to be total absolute error

$$L(\epsilon, \mathbf{a}) = \sum_{j=1}^{k} E\,|I_j - a_j|.$$

please turn over ..

(a) Discuss how would you use the information in observed test samples $\mathbf{Z}$ =
$(Z_1, Z_2, \ldots, Z_k)$ to find Bayes decision rule $d(\mathbf{Z}) = (d_1(\mathbf{Z}), d_2(\mathbf{Z}), \ldots, d_k(\mathbf{Z})$
?

(b) What would be right frequesntists' risk function to consider to evaluate $d(\mathbf{Z})$

N. B. Qn. 3 is an open ended question. Answers may be different. The natur of reasoning and argument will be evaluated. You may assume specific forms of and g if it helps in getting a specific answer. But, the forms should be standard **Answers must be brief and to the point.**

[10 + 10 =2C

INDIAN STATISTICAL INSTITUTE

End-Semester Examination – Semester I : 2014-2015

M.Stat. I Year

Measure Theoretic Probability

Date : 10.11.14 **Maximum Score : 60** $3\frac{1}{2}$ Hours

Note : This paper carries questions worth a total of **75 MARKS**. Answer as much as you can. The **MAXIMUM** you can score is **60**.

1. For a real-valued measurable function f on $(\Omega, \mathcal{A}, \mu)$, let $I(f) = \{p > 0 : \int |f|^p \, du < \infty\}$.
(a) Show that $I(f)$ is an interval (possibly empty) by showing that if $p_1 < p < p_2$ and $p_1, p_2 \in I(f)$, then $p \in I(f)$.
(b) Show that if $I(f)$ is non-empty, then the function ϕ defined on $I(f)$ by the formula $\phi(p) = \log\left(\int |f|^p \, d\mu\right)$, is a convex function.
(c) Give example of a measurable f on $((0,\infty), \mathcal{B}, \lambda)$ with $I(f) = (0,1)$. (5+5+5)=[15]

2. Show that for any real random variable X on a probability space $(\Omega, \mathcal{A}, P)$, one has $E[e^X] = \int_{\mathbb{R}} e^y \, P(X > y) \, dy$. [8]

3. Suppose $\{f_n\}$ is a sequence of non-negative measurable functions on a measure space $(\Omega, \mathcal{A}, \mu)$ such that $f_n \to f$ μ-a.e.
(a) Show that if $\int f \, d\mu < \infty$, then $\int (f_n - f)^- \, d\mu \to 0$.
(b) Conclude that if $\int f_n \, d\mu \to \int f \, d\mu < \infty$, then $f_n \xrightarrow{L_1} f$. (5+5)=[10]

4. Let X, Y be independent real random variables on a probability space $(\Omega, \mathcal{A}, P)$ and let $Z = X - Y$.
(a) Show that if H is the distribution function of X, then for any real z, $H(z + Y)$ is a bounded random variable and the function $F(z) = E(H(z+Y))$ is the distribution function of Z.
(b) Show that if Y is (absolutely) continuous with density function g, then so also is Z with density function $f(z) = E(g(X - z))$. (5+5)=[10]

5. Let $X_n \xrightarrow{P} X$ and $f : \mathbb{R} \to \mathbb{R}$ be a continuous function.
(a) Assuming that X is a bounded random variable, show that $f(X_n) \xrightarrow{P} f(X)$.
(b) Prove or disprove that the conclusion of (a) holds without the boundedness assumption in (a). (5+3)=[8]

6. (a) Let $\{X_n\}$ be a sequence of random variables on a probability space. Define the "tail" σ-field.
Denoting $\{S_n\}$ to be the sequence of partial sums, decide whether the following events are "tail" events or not:
(i) $\{X_n \geq X_{n+1}$ for infinitely many $n\}$, (ii) $\{S_{2n} \geq S_n$ for all but finitely many $n\}$,
(iii) $\{S_n/n$ converges $\}$, (iv) $\{\sqrt{n}\, S_n$ converges to $0\}$.
(b) State and prove Kolmogorov's zero-one law. ((2+4×2)+6)=[16]

7. Let $\{X_n\}$ be a sequence of i.i.d. random variables with common density $f(x) = e^{-x}$ if $x \in (0, \infty)$, and $f(x) = 0$ otherwise. Show that $\limsup_n (X_n / \log n) = 1$, a.s. [8]

MSTAT I - Measure Theoretic Probability

Compensatory Exam. / Semester I 2014-15

Time - 3 hours/ Maximum Score - 100

12·11·14

NOTE : RESULTS USED MUST BE CLEARLY STATED.

1. (11+14=25 marks)

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

(b) Let $\nu = \mu_1 \times \mu_2$ be a σ-finite measure on $(\mathcal{R}^2, \mathcal{B}(\mathcal{R}^2))$. For any set A in $\mathcal{B}(\mathcal{R}^2)$, define $A^x = \{y \in \mathcal{R} : (x,y) \in A\}$. Show that, the function $x \mapsto \mu_2(A^x)$ is Borel measurable, for any set A.

Also, show that, for any integrable function f on $(\mathcal{R}^2, \mathcal{B}(\mathcal{R}^2), \mu_1 \times \mu_2)$,

$$x \mapsto \begin{cases} \int_{\mathcal{R}} f(x,y)\mu_2(dy), & \text{whenever integral exists} \\ 0, & \text{otherwise} \end{cases}$$

is a Borel-measurable function.

2. (14+11=25 marks)

(a) Let $\{X_n\}$ be a sequence of independent random variables with finite variance. Let $S_n = X_1 + \cdots + X_n$. Does S_n converge almost surely, as $n \to \infty$ for the following sequences of independent random variables? Justify your answer.

(i) X_n has the probability density function

$$f_n(x) = \begin{cases} \frac{1}{2a_n} & \text{if } x \in (-a_n, a_n), \\ 0 & \text{otherwise,} \end{cases}$$

where a_n's are nonzero constants with $\sum_{n\geq 1} a_n^2 < \infty$.

(ii) $P(X_n = 2^{n+2}) = 2^{-n-4} = P(X_n = -2^{n+2})$ and $P(X_n = 0) = 1 - 2^{-n-3}$.

(b) (i) 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 $\sum_{k=1}^n a_k B_k = A_n B_n - \sum_{k=1}^n A_{k-1} b_k$.

(ii) Assume further in (i) 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 $n \to \infty$.

3. (10+10+5=25 marks)

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

   (a) Suppose the following condition (A) hold:

   $$(A) \qquad \lim_{k\to\infty} \sup_n \int_{|h_n|\geq k} h_n d\mu = 0.$$

   Show that $h_n \to h$, in L_1, as $n \to \infty$.

   (b) If $\sup_n \int_{\mathcal{R}} |h_n|^2 d\mu < \infty$ 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 condition (A).

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 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.

   (c) Let $\{X_n\}$ be an i.i.d. sequence of random variable, distributed as student-t distribution with 1 degrees of freedom. Let $S_n = X_1 + \cdots + X_n$. Then S_n/n converges with probability one, as $n \to \infty$.

   (d) Let $\{X_n\}$ be a sequence of independent random variables with $\sup_n E|X_n|^{2+c} < \infty$, for some constant $c > 0$. Then, for all $\epsilon > 0$, $\max_{1\leq k\leq n} P(|X_k| \geq \epsilon) \to 0$ as $n \to \infty$.

   (e) Let $\{f_n\}$ be a sequence of Borel-measurable function on the line and f be another Borel-measurable function on the same space. 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, as $n \to \infty$.

All the best.

Date : 12 November 2014

Design of Experiments

M.Stat I
Semestral I Examination

This question carries 60 marks. The maximum you can score is 50
Time allotted : 2 hrs.

Notes and Books are allowed

1.(a) Write down 3 mutually orthogonal latin squares of order 4, showing all necessary steps for constructing GF (4) and then justifying how the MOLS are obtained.

(b) Design an agricultural experiment in 16 plots with factors : type of soil, type of seed, water level and type of manure, each at 4 different levels, using your answer in (a).

(c) Put yield values in the 16 plots : 1,2,3, ..., 16.
Analyze the data. [6 + 6 + 12 = 24]

2. Consider following two block designs with 3 treatments: a, b, c.

Design 1

Block 1	:	a, b
Block 2	:	a, c
Block 3	:	b, c

Design 2

Block 1	:	a, b
Block 2	:	c, c
Block 3	:	b, c

Which design is "better" ? Justify (do <u>not</u> use any theorem). [8]

3. For the following block design with 3 treatments : a, b, c. Write down the C-matrix.

Block 1	:	a	b	c
Block 2	:	a	c	d
Block 3	:	b	c	d

[8]

4. Does there exist an SBIBD (40, 13, 4) ? Justify
If so, write down only two blocks. [20]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2014-15
M. Stat. I Year
Sample Surveys

Date: 12.11.2014 **Maximum Marks: 25** **Duration: $1\frac{1}{2}$ Hours**

(i) Answer any 2 questions each carrying 10 marks
(ii) Submit assignment Records to me on or before 22.11.2014. carrying 5 marks.

1 Obtain explicitly an unbiased estimator for the variance of symmetrized Des Raj estimator for a finite population total based on a PPSWOR sample chosen in $n(> 2)$ draws.

2 Derive an unbiased estimator for the variance of Horvitz and Thompson's estimator for a finite population total based on a PPSWR sample taken in $n\,(> 2)$ draws.

3 Illustrate a situation when and how you may employ an appropriate unbiased estimator for a finite population proportion of persons bearing a stigmatizing characteristic, the complement to it being also sensitive. Find an unbiased variance estimator

*****

INDIAN STATISTICAL INSTITUTE

SEMESTRAL EXAMINATION

M. Stat (1st Year) – B Stream, 2014 – 2015

Subject: Applied Stochastic processes

Date: 14/11/2014 **Full Marks: 60**

Time: 3 hours.

Attempt all questions

1. State and prove 1st Hardy – Weinberg law in connection to ~~unisex~~ the relevant population with two alleles A, a. In this context also show that proportions obtained in generations by random mating and by random pairing of gametes from the gametic pool are same.

[(2 + 6) + 4]

2. Write down Kermack-Mckendrick Equations. Describe how those equations are derived from the model of General Deterministic Epidemics. Using those equations show that limiting number of susceptible with time $t \to \infty$ is positive.

[4 + 6 + 6]

3. a) Consider Neyman – Scott model in the context of spread of Epidemic in R^2. Show that if there are no deserts and there is full mobility then the epidemic goes for extinction at u_0 implies that it goes for extinction at all other points, for all $u_0 \in R^2$.

b) Let the mobility function be given by

$$f_u(x) = 1/\pi \text{ if } d(u, x) < 1$$
$$= 0 \text{ otherwise,}$$

where $d(u, x)$ is distance between u and $x \in R^2$. Probability generating function of the number of infected persons from one infectious at $u = (u_1, u_2) \in R^2$ given by

$$g(t/u) = 1/3 + 1/3t + 1/3\, t^2 \qquad \text{if } u_1 u_2 = 0$$
$$= 1/3t + 1/3t^2 + 1/3t^3 \qquad \text{if } u_1 u_2 \neq 0$$

Discuss the probabilities of the process to become extinct at different points.

[12 + 6]

4. a) Let $\{X_n\}_{n \geq 0}$ be a branching process with $X_0 = 1$, $E(X_1) = m > 1$, $Var(X_1) = \sigma^2 > 0$. Let $W_n = X_n / m^n$. Show that there is a random variable W such that $W_n \to W$ almost surely.

b) Consider a branching chain with initial size N and probability generating function $g(s) = q + ps$, where $0 < p < 1$, $q = 1 - p$. If T = time when the process becomes extinct, show that $P(T = n) = (1 - p^n)^N - (1 - p^{n-1})^N$.

[10 + 4]

-------------------o--------------------

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2014–2015
M.Stat. 1st Year. 1st Semester
Large Sample Statistical Methods

Date: November 18, 2014 **Maximum Marks: 80** **Duration: 3 and 1/2 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. Suppose i.i.d. observations X_i's are being drawn from a distribution with cdf F such that $\mathrm{E}_F(X_1^4) < \infty$. Suppose, moreover, that $\mathrm{E}_F\left((X_1 - \mu)^3\right) = 0$, where $\mu := \mathrm{E}_F(X_1)$. Show that the sample mean and sample variance are asymptotically uncorrelated. [12]

2. Suppose i.i.d. observations X_i's are being drawn from a distribution having a pdf belonging to a parametric family $\mathcal{F} := \{f(x, \theta) : \theta \in \mathbb{R}\}$. We wish to test the hypothesis $\mathrm{H}_0 : \theta = \theta_0$ against $\mathrm{H}_1 : \mathrm{H}_0$ *is false*. Find the asymptotic distribution of the likelihood ratio test statistic under appropriate conditions to be stated by you. [10]

3. Consider the simple linear regression model $Y_i = \alpha + \beta x_i + \epsilon_i,\ i \geq 1$, where ϵ_i's are i.i.d. with $\mathrm{E}(\epsilon_1) = 0$, $\mathrm{E}(\epsilon_1^2) = \sigma^2$. Denote by $(\hat{\alpha}_n, \hat{\beta}_n)$, the least squares estimator of (α, β) based on (Y_i, x_i), $i = 1, \ldots, n$. Find, under appropriate conditions to be stated by you, the asymptotic distribution of $\hat{\beta}_n$. [10]

4. (a) Suppose i.i.d. observations X_i's are being drawn from a distribution with cdf F such that $\mathrm{Var}_F(X_1^2) < \infty$. Suppose, moreover, that F is differentiable at μ with a positive derivative, where $\mu := \mathrm{E}_F(X_1)$. Find the asymptotic distribution of the sample proportion of observations exceeding the sample mean.

(b) Suppose a sample of $N = 75$ i.i.d. observations are drawn from Laplace(0, 1) distribution. The sample mean and sample median are computed from these N observations. This process is repeated $M = 10000$ times so that we have a set of M vectors of sample means and sample medians, denoted by (m_i, M_i), $i = 1, \ldots, M$. The correlation coefficient of the means and medians, computed from the m_i's and M_i's turns out to be 0.50. Is this believable? Give reasons. [10+9 = 19]

[P.T.O.]

5. (a) Suppose i.i.d. observations X_i's are being drawn from a distribution having a pdf f satisfying $f(x) = f(-x)$ for all $x \in \mathbb{R}$. Denote the U-statistic, based on the kernel $h(x_1, x_2) := I_{x_1+x_2>0}$ and on X_i, $i = 1, \ldots, n$, by U_n. Find the asymptotic distribution of U_n.

(b) Suppose i.i.d. observations X_i's are being drawn from a distribution having a cdf F. Denote the U-statistic, based on a symmetric kernel h of order 3 and on X_i, $i = 1, \ldots, n$, by U_n. Let $\sigma^2 := \text{Var}_F\,(h(X_1, X_2, X_3))$. Also, let $0 < \sigma < \infty$. Show that $\text{Var}_F\,(U_n) \leq 3\sigma^2/n$.

(c) Argue for or against the following. *The asymptotic normality theorem of a U-statistic understates its variance.* [8+9+5 = 22]

6. Suppose i.i.d. observations are being drawn from a Multinomial$(1; \pi_1, \pi_2, \ldots, \pi_k)$ population. We wish to test the hypothesis $\text{H}_0 : \pi_i = \pi_{i,0}$ $i = 1, \ldots, k$, against $\text{H}_1 : \pi_i = \pi_i^*$, $i = 1, \ldots, k$, where $\pi_{i,0} > 0 \;\forall\, i$. Denote by T_n, the Pearson Chi-Square test statistic, where n is the sample size. Explain how you can obtain an approximation to the power of T_n at $(\pi_1^*, \ldots, \pi_k^*)$. [7]

****** *Best of Luck!* ******

Indian Statistical Institute

First Semestral Examination 2014-15

M. Stat. I yr

Statistical Inference I

Date: November 21 , 2014   Maximum marks: 100   Duration: 3 hrs.

Answer all Questions. *Paper carries 115 points. (Open self prepared notes)*

1 (a) Let S denote the risk set of randomized rules of a decision problem on a finite parameter space which is bounded from below. Define the lower boundary $\lambda(S)$ of S. Show that when $\lambda(S) \subset S$ then $\lambda(S)$ forms a minimal complete class.

(b) Let $S \subset \mathbb{R}^n$ be a convex subset and let $\bar{S}$ denote its closure. Show that $\text{int}(S) = \text{int}(\bar{S})$ where $\text{int}(A)$ denotes the interior of a set A.

[15 + 10 = 25]

2. Consider point estimation under squared error loss (notations used in class)

(a) Show that if an estimator T is both unbiased and Bayes under some prior π then $r(\pi) = 0$.

(b) Give an example of a model where unbiased Bayes estimators exist.

(c) For estimation of θ in a $\text{Bin}(n, \theta)$ model find out a minimax estimator. Is it admissible?

[10 + 10 + 15 = 35]

3. State and prove Karlin's theorem on admissibility in a one parameter exponential family model.

[20]

4. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be multivariate normal with mean vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ and covariance matrix I_n. Consider simultaneous estimation of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ under the total squared error loss $L(\boldsymbol{\theta}, \mathbf{a}) = \sum_1^n (a_i - \theta_i)^2$ for an action $\mathbf{a} = (a_1, a_2, \ldots, a_n)$.

(a) Consider the estimators $T_0(\mathbf{X}) = \mathbf{X}$ and $T_1(\mathbf{X}) = \mathbf{X} + g(\mathbf{X})$ respectively, where $g : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. Find suitable $\Delta(\mathbf{X})$ such that

$$R(\boldsymbol{\theta}, T_1) - R(\boldsymbol{\theta}, T_0) = E_{\boldsymbol{\theta}} \Delta(\mathbf{X}).$$

[*please turn over* ...]

(b) Hence or otherwise show that $T_0(\mathbf{X})$ in 4(a) is inadmissible under the loss function L for $n \geq 3$.

[10+ 15 =25]

5. Consider the problem of testing the hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ in the multivariate normal model described in 4 above. Describe a group of transformations on $\mathbb{R}^n$ which keeps the testing problem invariant. Derive the maximal invariant statistic and its distribution under a general $\boldsymbol{\theta}$.

[10]

INDIAN STATISTICAL INSTITUTE

Back-Paper Examination – Semester I : 2014-2015

M.Stat.(B-Stream) I Year

Measure Theoretic Probability

Date: 16/01/2015 **Total Marks : 100** Time : 3 Hours

Note : Answer any five of the six questions each of which has 20 marks. In case you attempt more than five and do not strike off the one that you do not want to be graded, the first five attempted questions will be considered for grading. The **maximum** you can score is **45**.

1. Let $\mathcal{A}$ denote the class of all those sets $A \subset \mathbb{R}$ such that either A or A^c is countable.
   (a) Show that $\mathcal{A}$ is the smallest σ-field on $\mathbb{R}$ containing all the singleton sets.
   (b) Show that a function $f : (\mathbb{R}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ is measurable if and only if there exists a set $A \subset \mathbb{R}$ with A^c countable such that f is constant on A. (10+10)=**[20]**

2. For any function $f : \mathbb{R} \to \mathbb{R}$ and any $\alpha > 0$, let $f_\alpha : \mathbb{R} \to \mathbb{R}$ be defined as $f_\alpha(x) = f(x/\alpha)$.
   (a) Show that if f is measurable, then so is f_α for any $\alpha > 0$.
   (b) Show that the integral $\int f_\alpha(x)dx$ exists if and only if $\int f(x)dx$ exists and, in that case, $\int f_\alpha(x)dx = \alpha \int f(x)dx$.
   (c) Show that if $\alpha_n > 0$, $n \geq 1$ is a sequence with $\alpha_n \to 1$, then for any $f \in L_1$, the sequence f_{α_n} converges to f in L_1. (5+5+10)=**[20]**

3. Let $(\Omega, \mathcal{A}, P)$ be a probability space.
   (a) Suppose $\mathcal{S}_1 \subset \mathcal{A}$ and $\mathcal{S}_2 \subset \mathcal{A}$ are two semifields. Show that, if $P(A \cap B) = P(A)P(B)$ for all $A \in \mathcal{S}_1, B \in \mathcal{S}_2$, then the σ-fields $\sigma(\mathcal{S}_1)$ and $\sigma(\mathcal{S}_2)$ are also independent.
   (b) Show that if $X_1, \ldots, X_n$ are independent random variables and $1 < k < n$, then for any borel functions $h : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^{n-k} \to \mathbb{R}$, the random variables $Y = h(X_1, \ldots, X_k)$ and $Z = g(X_{k+1}, \ldots, X_n)$ are independent. (10+10)=**[20]**

4. Suppose $X_n \xrightarrow{P} X$, where X_n, $n \geq 1$ and X are all real random variables.
   (a) Show that, if α_n, $n \geq 1$ are such that $P(X_n < \alpha_n) \leq p$, for each $n \geq 1$, then $P(X < \limsup \alpha_n) \leq p$.
   (b) Recall that for $p \in (0, 1)$, a p-th quantile of a r.v. Z is a real number α (not unique in general) such that $P(Z < \alpha) \leq p \leq P(Z \leq \alpha)$. Show that, if α_n is a p-th quantile of X_n and if X has a unique p-th quantile α, then $\alpha_n \to \alpha$. (10+10)=**[20]**

5. (a) Show, directly from definitions, that $X_n \xrightarrow{\text{a.s.}} X$ if and only if $\sup_{k \geq n} |X_k - X| \xrightarrow{P} 0$.
   (b) Show that for random variables X_n, $n \geq 1$ and X defined on a discrete probability space $(\Omega, \mathcal{A}, P)$, one has $X_n \xrightarrow{P} X$ if and only if $X_n \xrightarrow{\text{a.s.}} X$. (10+10)=**[20]**

6. Let $\{X_n\}$ be a sequence of i.i.d. random variables with $E[\sqrt{|X_1|}] < \infty$.
   (a) Denoting $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n^2\}}$, $n \geq 1$, show that the series $\sum_{n=1}^{\infty} V[Y_n/n^2]$ and $\sum_{n=1}^{\infty} E[Y_n/n^2]$ both converge. [You may use the inequality $\sum_{n=j}^{\infty} n^{-\alpha} \leq \frac{\alpha}{\alpha-1} j^{-\alpha+1}$, which holds for all $\alpha > 1$ and for any integer $j \geq 1$.]
   (b) Use (a) to prove that $(X_1 + \cdots + X_n)/n^2 \to 0$ almost surely. (10+10)=**[20]**

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination – Semester II : 2014-2015

M. Stat. I Year

Metric Topology and Complex Analysis

Date : 23.02.15 | **Maximum Score : 40** | Time : $2\frac{1}{2}$ Hours

Note : This paper carries questions worth a **total of 50 MARKS**. Answer as much as you can. The **MAXIMUM** you can score is 40.

1. Let X and Y denote normed linear spaces, equipped with the induced metric.
(a) Show that if $V \subset X$ is open, then for any $A \subset X$, the set $A+V = \{x+y : x \in A, y \in V\}$ is open.
$T : X \to Y$ is said to be a linear map if $T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2)$ for all $x_1, x_2 \in X$ and $a_1, a_2 \in \mathbb{R}$.
(b) Show that a linear map $T : X \to Y$ is continuous if and only if there exists $C \in (0, \infty)$ such that $\|T(x)\| \leq C\|x\|$ for all $x \in X$.
(c) Show that any linear map $T : \mathbb{R}^n \to X$ is continuous. (3+3+3)=**9**]

2. Let X denote the set of real-valued continuous functions on $\mathbb{R}$. For $x \in X$ and integer $k \geq 1$, denote $\|x\|_k = \sup\{|x(t)| : t \in [-k, k]\}$.
(a) Show that $d(x, y) = \sum_{k=1}^{\infty} 2^{-k}(\|x - y\|_k \wedge 1)$, for $x, y \in X$, defines a metric on X.
(b) Show that (X, d) is complete and separable.
(c) Show that $\{x_n\}$ converges to x in (X, d) if and only if $x_n(\cdot) \to x(\cdot)$ uniformly on each compact subset of $\mathbb{R}$. (3+3+3)=**[9]**

3. Let (X, d) be a metric space. Assume usual notations for closures, interiors and boundaries of subsets of X.
(a) Show that $\overline{A \cup B} = \overline{A} \cup \overline{B}$.
(b) Show that $\partial(A \cap B) \subset (\partial A) \cup (\partial B)$.
(c) Denoting $\pi(A) = (\,\overline{A}\,)^0$, show that $\pi(\pi(A)) = \pi(A)$, for any $A \subset X$. (3+3+3)=**[9]**

4. Let (X, d) be a metric space. For $x \in X$ and non-empty $A \subset X$, let $d(x, A) = \inf\{d(x, y) : y \in A\}$.
(a) Show that the map $x \mapsto d(x, A)$ is continuous.
(b) Show that $d(x, A) = 0$ if and only if $x \in \overline{A}$.
(c) Let $V(A, \epsilon) = \bigcup\{B(x, \epsilon) : x \in A\}$. Show that $V(A, \epsilon) = \{x \in X : d(x, A) < \epsilon\}$, called the "$\epsilon$-neighbourhood of A".
(d) Show that if $A \subset X$ is closed, $B \subset X$ is compact and A and B are disjoint, then for some $\epsilon > 0$, A and B have disjoint ϵ-neighbourhoods. (3+3+3+4)=**[13]**

5. Let (X, d) be a metric space. A set $K \subset X$ is called "countably compact" if every countable open cover of K admits a finite subcover. A set $A \subset X$ is said to have the "BW property" if every sequence in A has a convergent subsequence.
(a) Show that a set $A \subset X$ has the BW property if and only if $\overline{A}$ has the BW property.
(b) Show that any countably compact set $K \subset X$ is closed and has the BW property.
(c) Deduce that K is compact if and only if K is countably compact. (4+4+2)=**[10]**

Indian Statistical Institute

M.Stat I

Discrete Mathematics

Mid Semester Examination

Maximum Marks: 70

Date: February 25, 2015
Time 2.5 hours

The question paper contains 7 questions. Total marks is 70. Maximum you can score is 60. Unless otherwise mentioned, all notations are the same as presented in class.

1. (a) Prove that M is a maximum matching if and only if M is contained in a minimum edge cover.

   (b) Prove that L is a minimum edge cover if and only if L contains a maximum matching. (7+3=10)

2. Let T be a minimum spanning tree in G and T' is another spanning tree in G. Prove that T' can be transformed into T by a sequence of steps that exchange one edge of T' for one edge of T, such that the edge set if always a spanning tree and the total weight never increases. (10)

3. Prove that the following is a necessary condition for the existance of k pairwise edge-disjoint spanning trees in a graph G: For any partition of the vertices of G into r parts, there are at least $k(r-1)$ edges of G whose endpoints are in different parts of the partition. (10)

4. Let $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ be a collection of subsets of a set X. A system of distinct representatives (SDR) for $\mathcal{A}$ is a set of distinct elements $a_1, a_2, \ldots, a_m$ in X such that $a_i \in A_i$. Prove that $\mathcal{A}$ has an SDR if and only if $|\cup_{i \in S} A_i| \geq |S|$, for every $S \subseteq \{1, 2, \ldots, m\}$. (10)

5. (a) Prove that every planar graph has a vertex whose degree is less than 5.

   (b) Let G be a graph on n vertices. Prove that if $\omega(G) \leq r$, then $e(G) \leq (1 - 1/r)n^2/2$. Hence prove that, $\chi(G) \geq n^2/(n^2 - 2e(G))$. (5+(3+2) = 10)

6. Prove that every strongly connected tournament has a Hamiltonian circuit. (10)

7. Let G be a connected graph with atmost two odd vertices. Write an algorithm which constructs an Eulerian trail. Prove the correctness of the algorithm and deduce the time complexity. (5 + 3 +2 = 10)

Indian Statistical Institute

M.Stat. First year, Second Semester Exam: 2014-15

Topic: Regression Analysis

Maximum Marks: 60, Duration: 2 hours

Answer all questions. Show your works to get full credit. Marks will be deducted for untidiness and bad handwriting.

1. Consider the multiple regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$.
(a) For a fixed $1 \times p$ vector q, define $\mu(q) = q\beta$. Give 100(1-α)% confidence interval for $\mu(q)$.
(b) Now consider q to be arbitrary. Give 100(1-α)% confidence band for $\mu(q)$. [4+6]

2. (a) Consider the multiple linear regression model:
$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$, where Y is $n \times 1$, X_1 is $n \times p_1$, β_1 is $p_1 \times 1$. X_2 is $n \times p_2$, β_2 is $p_2 \times 1$, and ϵ is $n \times 1$.

Suppose that in fact $\beta_2 = 0$, in other words, the model used by the experimenter is an overfitted model and the true model is $Y = X_1\beta_1 + \epsilon$. Let $\sigma^2_{overfit}$ denote the usual estimate of variance based on the overfitted model, i.e.,$\sigma^2_{overfit} = \frac{Y'(I-P)Y}{n-p_1-p_2}$, where P is the projection onto the space spanned by the columns of X_1 and the columns of X_2. Show that $\sigma^2_{overfit}$ is an unbiased estimate of σ^2 even if the smaller model is true.

(b) Let σ^2_{red} be the estimate of σ^2 based on the reduced (and correct) model $Y = X_1\beta_1 + \epsilon$. Show that the expected length of the confidence interval for σ^2 based on the reduced model is smaller than that under the overfitted model. [5+5]

3. (a) Under the standard multiple regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$, show that Variance($\hat{\beta}_{ridge}$) is smaller than Variance($\hat{\beta}_{ols}$).
(b) Under the multiple regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 A)$, A is positive definite, how will you test the hypothesis $H_0 : C\beta = c$, for some $q \times p$ matrix C with rank q. [4+6]

4. (a) While testing the set of hypotheses $H_0 : \beta_j = 0$, vs. $H_1 : \beta_j \neq 0$, for $j = 1, 2, \ldots, p$, how can you control the family-wise error rate at level α=0.025? How can you control the per comparison error rate at the same level?
(b) Suppose data are collected on body mass index (BMI) and hypertension level (HL) for 50 patients. Assuming that HL can be considered as an unknown function of BMI, how will you predict your HL if you know that your BMI is 24.6? [3+7]

5. (a) Given two parallel lines of y on x:
$E(y|x) = \alpha_1 + \beta x$, based on n_1 observations and

$E(y|x) = \alpha_2 + \beta x$, based on n_2 observations.
Find the 95% confidence limit of the horizontal distance between the lines.
(b) What is Cook's D statistic? What is it used for? [7+3]

6. (a) Suppose data are collected on HDL (good) cholesterol, LDL (bad) cholesterol and Triglycerides for n patients longitudinally. Different subjects are measured at different time points and suppose we get T_i measurements from the i-th patient. Thus the response from the i-th patient at the j-th time point on the k-th feature can be denoted by $Y_{ik}(t_{ij})$, where $i = 1, \ldots, n$, $j = 1, \ldots, T_i$ and $k = 1, 2, 3$.

Using a parametric approach of modeling the mean trajectory and the covariance structure, write down the joint likelihood function using the appropriate notations.
(b) Give a brief outline of parameter estimation based on the likelihood function you derived in (a). [6+4]

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination : 2014-15
M. STAT. I YEAR
Optimization Techniques

Date: 26 February 2015 | Maximum Marks: 50 | Duration: 2 hours

---

Notation have usual meaning.

This paper carries 60 marks. However, maximum you can score is 50.

---

1 Let

$$f(x) = \begin{cases} a_1x + b_1 & \text{if } x_0 \le x \le x_1 \\ a_2x + b_2 & \text{if } x_1 < x \le x_2 \\ \vdots & \\ a_mx + b_m & \text{if } x_{m-1} < x \le x_m \end{cases}$$

be a piecewise-linear convex function. Formulate the problem of minimizing $f(x)$ as a linear program. [10]

2 Let $S \subseteq R^n$ be a closed convex set and $\bar{x} \in S$. Suppose that non-zero $d \in R^n$, and that $\bar{x} + \lambda d \in S$ for all $\lambda \ge 0$. Show that d is a direction of S. [10]

3 Consider $S = \{x = (x_1, x_2)^T : -x_1 + 2x_2 \le 4, x_1 - 3x_2 \le 3, x \ge 0\}$. Identify the extreme points and extreme directions of S. Represent $(4,1)^T$ as a convex combination of extreme points plus a non-negative combination of extreme directions. [10]

4 Suppose that at some iteration of simplex method with $\bar{x}$ as basic feasible solution, $z_j - c_j < 0\ \forall j$. Prove that $\bar{x}$ is unique optimal solution. [10]

5 Suppose that Big-M problem has an optimal solution. What can you say about the original problem of interest? Justify your answer. [10]

6 Consider the following linear programming problems:

$$\max \; c^T x \quad \text{subject to } Ax \le b, \qquad \text{and} \qquad \min \; c^T x \quad \text{subject to } Ax \ge b.$$

Suppose that both of these problems are feasible. Prove that

(a) if one of these problems has an optimal solution, then so does the other, and

(b) the first problem is unbounded if and only if the second problem is unbounded.

[10]

——*** xXx ***——

INDIAN STATISTICAL INSTITUTE

Mid-semestral examination

M.stat (1st year) 2014-2015

Subject-Time Series Analysis

Date-27/2/2015 Duration-2 hours Full Marks-40

Attempt All Questions

1. Define a Time Series. Write down main uses of Time Series Analysis. [1+4]

2. Show that filter with coefficients $[a_{-2}, a_{-1}, a_0, a_1, a_2]=1/9\ [-1, 4, 3, 4, -1]$ passes 3rd degree polynomials and eliminates seasonal components with period 3. [5]

3. Consider a moving average process $Y_t = \sum_{i=0}^{\infty} a_i Z_{t-i}$, where $\{Z_t\}\sim WN(0,v^2)$, $\sum_{i=0}^{\infty} |a_i| < \infty$ and $v^2>0$. Show that the sum is almost surely convergent. [6]

4. Consider ARMA (1,1) process $\{X_t\}$ with equation $X_t - 0.5 X_{t-1} = Z_t + Z_{t-1}$, where $\{Z_t\}\sim WN(0,1)$. Show that the process is not invertible, but Z_t can be shown to be Root-Mean-Square limit of random variables which are finite linear combinations of X_t's. [6+6]

5. Let $Y_t = X_t + W_t$, $t= 0,\pm1,\pm2,\ldots$ where $\{X_t\}\sim AR(1)$ process given by the equation $X_t - 0.4X_{t-1} = Z_t$, $\{Z_t\}\sim WN(0,1)$ with $\{W_t\}\sim WN(0,2)$ and $E(W_t Z_s)=0$, for all s and t.
   i) Show that $\{Y_t\}$ is stationary and find its autocovariance function.
   ii) Show that $\{Y_t\}$ is an ARMA(1,1) process.
   iii) Find the values of the parameters of the above ARMA(1,1) process. [4+4+4]

---------x-x-x--------

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : Semester II (2014-2015)

M. Stat 1st Year

Multivariate Analysis

Date: 2. 3. 15 Maximum marks: 60 Time: 2 hours.

Note: Answer all questions. Maximum you an score is 60.

1. Suppose $\mathbf{X}$ follows $N_4(0, \mathbf{I})$. Find the distribution of

$$\frac{\sqrt{2}(X_1X_3 + X_2X_4)}{X_3^2 + X_4^2}$$

[10]

2. Let $\mathbf{X}_{n\times p}$ be a data matrix from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\Gamma_{n\times n}$ be an orthogonal matrix with the last row $\sqrt{n}$ times the unit vector. Suppose $\mathbf{Y} = \Gamma\mathbf{X}$. Denote the rows of $\mathbf{Y}$ by $Y_1, Y_2, \ldots, Y_n$. Show that

   (a) $Y_i, i = 1, 2, \ldots, n$ are independent. [3]

   (b) Y_n follows $N_p(\sqrt{n}\boldsymbol{\mu}, \boldsymbol{\Sigma})$. [3]

   (c) $Y_i, i = 1, 2, \ldots, n-1$ are i.i.d. $N_p(0, \boldsymbol{\Sigma})$. [5]

   (d) Express $\bar{X}$ in terms of Y_n only, and S in terms of $Y_i, i = 1, 2, \ldots, n-1$. [4]

   (e) Find the distribution of $\bar{X}$ and S, and show that they are independent. [5]

3. Suppose $\mathbf{x}$ follows $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = k\boldsymbol{\Sigma}_0$, and $\boldsymbol{\mu}$ is unconstrained. Find the m. l. e. of k. Verify whether this estimate is unbiased or not. [5+3]

4. Consider the IRIS data for the Iris Versicolour population only. There are four measurements: sepal length, sepal width, petal length and petal width. Suppose we want to test whether the mean of sepal length is equal to the mean of petal length for this population.

   (a) Find the m.l.e. of $\boldsymbol{\mu}$, the population mean vector for these four measurements under the null hypothesis. [8]

   (b) Assuming that $\boldsymbol{\Sigma}$ is known, derive the LRT for this test. What is the distribution of the test statistic? How will this test differ from the UIT and why? [10]

   (c) How will your answers in part (b) change, if $\boldsymbol{\Sigma}$ is unknown. [4]

Part - B

Assignments [10]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination : (2014-2015)

M. Stat 1st Year

Multivariate Analysis

Date: 20. 4. 15   Maximum marks: 100   Time: 3 hours.

Note: Answer all questions. Maximum you can score is 100.
You may use calculators. You can use any result that has been proved in class.
However you need to write the results clearly to get credit.

Part - A

1. (a) Consider the problem of classification into one of two multivariate normal populations with unequal mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and common positive definite dispersion matrix $\boldsymbol{\Sigma}$. Assuming $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ to be unknown, express the Bayes error i.e., the probability of misclassification for the Bayes rule in terms of the standard normal distribution function, the prior probabilities (π_1 and π_2), the losses due to misclassification $C(1 \mid 2)$ and $C(2 \mid 1)$ and Δ, the Mahalanobis distance between the two populations. [15]

   (b) Suppose, $\pi_1 = 0.6$, $\pi_2 = 0.4$ and

$$\boldsymbol{\mu}_1 = (2,4), \quad \boldsymbol{\mu}_2 = (6,8), \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$$

   Compute the Bayes error for $C(1 \mid 2) = C(2 \mid 1) = c$. [15]

2. Consider a p-dimensional random vector $\mathbf{X}$ with dispersion matrix $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I} + \rho\mathbf{1}$, where $\mathbf{1}$ is a matrix with all elements being unity.

   (a) Find the principal components of $\mathbf{X}$. [10]

   (b) Determine the value of k such that the first k principal components can explain at least 90% of the total variation of $\mathbf{X}$. [5]

3. (a) Describe the k-factor model. Derive the necessary condition for this model to hold. [5]

   (b) Show that the derived condition in part(a) is also sufficient. [7]

   (c) If the k-factor model holds, show that it is scale-invariant but the factor loadings may not be unique. [4 + 4]

   (d) Give an example when a k-factor model may not hold. [5]

[P. T. O.]

4. Describe the three stages of Profile Analysis in Multivariate Analysis of Variance, and derive the related test statistics. [3 + 12]

5. A researcher considered three indices measuring severity of heart attacks. The values of these indices for 40 heart-attack patients arriving at a hospital emergency room produced the following sample covariance matrix

$$S = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

Construct an appropriate statistical test for testing whether the population multiple correlation coefficient between the first index and the others is 0. What will be your decision based on a significance level of 5%. [8 + 7]

Part - B

Assignments [10]

Indian Statistical Institute

M.Stat. First year, Second Semester Final Exam: 2014-15

Topic: Regression Techniques

Maximum Marks: 60, Duration: 3 hours

22·04·2015

Answer all questions. Show your works to get full credit. Marks will be deducted for untidiness and bad handwriting.

1. (a) Define the terms "missing completely at random" (MCAR), "missing at random" (MAR) and "missing not at random" (MNAR) in the context of general missing data problem.
(b) Explain why for MCAR case, list-wise deletion is a bad choice for handling the missing data.
(c) In a multiple regression case with p predictors, consider the situation when the predictor values are given for all n subjects, but the response values are available only for n_1 subjects ($n_1 < n$). Suggest a propensity score based multiple imputation technique for imputing the missing responses. [3+2+5]

2. Consider the standard multiple regression model (with p predictors) $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$.
(a) Derive the UMVU estimator of σ^2. Show that it is a consistent estimator of σ^2 as long as $n - p \to \infty$.
(b) Find $\hat{\sigma}^2_M$, the MLE of σ^2. Suppose that $\frac{p}{n} \to c$. Show that $\hat{\sigma}^2_M \xrightarrow{P} (1-c)\sigma^2$. [5+5]

3. (a) What do we mean by "varying coefficients regression models (VCRM)"? Give a real example where such models are more appropriate than the usual multiple regression models.
(b) Consider a multiple regression problem with two predictors X_1 and X_2. A VCRM is suggested as the following: $Y = \beta_1(u)X_1 + \beta_2(u)X_2 + \epsilon$, where u is some covariate of interest. Suppose a higher degree of smoothness is expected in $\beta_2(u)$ than in $\beta_1(u)$. Provide an appropriate estimation method to achieve this.
(c) To check whether the unknown coefficients really depend on u, give a non-parametric goodness of fit test. Why does the traditional likelihood ratio test (LRT) fail to give appropriate inference here? [3+4+3]

4. (a) Define the "systematic component" of a generalized linear model (GLM). What is meant by "canonical link"? For a Poisson GLM, find the canonical link function.
(b) Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables following a Poisson GLM with linear predictor $\eta_i = x_i^T\beta$ and the link function g. Define $Z_i=1$, if $Y_i > 0$ and $Z_i=0$, otherwise. Show that $Z_1, Z_2, \ldots, Z_n$ follow a binary GLM. Find the link function of this GLM.
If $Z_1, Z_2, \ldots, Z_n$ have to follow a binary GLM with logit link, what will be the link function of the Poisson GLM? [4+6]

5. Consider the standard multiple regression model (with p predictors) $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$.

(a) What is the objective function in LASSO regression? Show that LASSO estimates are indeed the posterior estimates if a double exponential prior is taken for the parameter vector β.

(b) Discuss the relative advantage of Fused LASSO over the ordinary LASSO.

(c) Suppose data are collected on (X_k, Y_k), for $k = 1, 2, \ldots, n$. Define "distance correlation" between X and Y. What are the properties of distance correlation? [5+2+3]

6. Suppose data are collected on HDL (good) cholesterol, LDL (bad) cholesterol and Triglycerides for n patients longitudinally. Subjects are followed at T evenly spaced time points and for the i-th patient, we have r_i available data and $T - r_i$ data points are missing. However, the missingness is monotone in the sense that if patient i is missing at time t, then it is also missing at time t', for $t' > t$.

(a) Write down a selection model for such data assuming the "missing not at random" mechanism. Give the likelihood function and the estimation procedure.

(b) How will you suggest a pattern mixture model assuming ignorable missingness (i.e. missing at random). Write down the likelihood function explicitly. [5+5]

INDIAN STATISTICAL INSTITUTE

Semestral Examination – Semester II : 2014-2015

M. Stat. I Year

Metric Topology and Complex Analysis

Date : 28.04.15 | **Maximum Score : 60** | Time : $3\frac{1}{2}$ Hours

Note : This paper carries questions worth a **total** of **80 MARKS**. Answer as much as you can. The **MAXIMUM** you can score is **60**.

1. (a) State the following:
(i) Liouville's Theorem, (ii) Morera's Theorem, (iii) Cauchy's Residue Theorem, (iv) Casorati-Wierstrass Theorem, (v) Schwarz's Lemma.
(b) Prove any two of the above. ((5×3)+(2×5)=**[25]**

2. (a) State clearly the Cauchy-Riemann equations for an analytic function.
(b) Let f be a non-constatnt analytic function on a connected open set D. Show that the function $h(z) = \overline{f(z)}$, $z \in D$ is not analytic on D, but the function g defined on $D^* = \{\overline{z} : z \in D\}$ by $g(z) = \overline{f(\overline{z})}$, $z \in D^*$ is analytic. (3+7)=**[10]**

3. (a) Let f be an entire function satisfying $|f(z)| \leq 8|z|^5$ for all $z \in \mathbb{C}$ with $|z| > 23$. Show that f must be a polynomial of degree at most 5.
(b) What are all the entire functions f satisfying $f(x) = e^{-2ix}$ for all real x? Justify your answer. (5+5)=**[10]**

4. (a) Let f_n, $n \geq 1$ be a sequence of analytic functions on a connected open set D, converging to a function f, uniformly on all compact subsets of D. Is f necessarily analytic on D? Justify your answer either by a proof or by a counterexample.
(b) Find the value of the contour integral $\int_\gamma \frac{e^z}{(z-1)^4}dz$, where γ is the path defined as $\gamma(t) = 1 + e^{2it}$, $0 \leq t \leq 2\pi$. (5+5)=**[10]**

5. (a) Describe clearly what is meant by the Laurent expansion of a function f around an isolated singularity at $z = a$.
(b) State and prove a result chracterizing the type of the singularity of a function f at $z = a$ in terms of the coefficients in its Laurent expansion around $z = a$. (3+7)=**[10]**

6. Let f be a non-constant analytic function on the open unit ball $D = \{z : |z| < 1\}$ such that $Re(f(z)) \geq 0$ for all $z \in D$.
(a) Show that $Re(f(z)) > 0$ for all $z \in D$.
(b) Denoting $V = \{w \in \mathbb{C} : Re(w) > 0\}$, show that $h(w) = \frac{1-w}{1+w}$ defines a one-one analytic function on V onto D.
(c) Show that, if $f(0) = 1$, then $|f(z)| \leq \frac{1+|z|}{1-|z|}$ for all $z \in D$. [Hint: Use Schwarz's Lemma appropriately.] (3+5+7)=**[15]**

Indian Statistical Institute

M.Stat I

Discrete Mathematics

Semester Examination

Maximum Marks: 100

Date: May 01, 2015

Time 3.5 hours

The question paper contains 10 questions. Total marks is 110. Answer all the questions. Maximum you can score is 100. Unless otherwise mentioned, all notations are the same as presented in class.

1. Prove that $\chi(\overline{H}) = \omega(\overline{H})$, when H is bipartite and has no isolated vertices. (10)

2. Let $d_1 \leq d_2 \leq \cdots \leq d_n$ be the vertex degrees of a simple graph G. Prove that G is connected if $d_k \geq k$ for every k such that, $k \leq n - 1 - d_n$. (10)

3. A graph is *Hamiltonian-connected* if for every pair of vertices u, v there is a Hamiltonian path from u to v. Prove that a simple graph G is Hamiltonian-connected if, $e(G) \geq \binom{n(G)-1}{2} + 3$. (10)

4. Prove that $R(mK_2, mK_2) = 3m - 1$. (10)

5. Design an algorithm to test if a graph is bipartite. Prove the correctness of the algorithm and calculate the time complexity. (4 + 3 + 3 = 10)

6. Suppose that G is a simple graph with diameter 2 and that $[S, \bar{S}]$ is the minimum edge cut with $|S| \leq |\bar{S}|$.

   (a) Prove that every vertex of S has a neighbor in $\bar{S}$.

   (b) Use (a) to prove that $\kappa'(G) = \delta(G)$. (6+4 = 10)

7. Let J be an $n \times n$ matrix of all 1's and $A(G)$ be the adjacency matrix of a graph G. Prove that G is regular and connected if and only if J is a linear combination of powers of $A(G)$. (15)

8. Consider a graph $G(n, p)$ with $p = 1/n$. Let X be the number of triangles in the graph.

   (a) Show that, $Pr(X \geq 1) \leq 1/6$.

   (b) Use conditional expectation inequality to show that $\lim_{n\to\infty} Pr(X \geq 1) \geq 1/7$.

   (5 + 10 = 15)

9. Using Perron-Frobenius theorem show that the scaled Page Rank computation converges in the limiting condition. (10)

10. Suppose X is a random variable and k is an even integer. Prove that,

$$Pr(|X - E[X]| > t\sqrt[k]{E[(X - E[X])^k]}) \leq 1/t^k.$$

Why is it difficult to derive a similar inequality when k is odd? (8+2 = 10)

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination : 2014-15
M. STAT. I YEAR
Optimization Techniques

Date: 02 May 2015 | Maximum Marks: 50 | Duration: $2\frac{1}{2}$ hours

---

This paper carries 60 marks. However, maximum you can score is 50.

---

1 The starting and current (next) simplex tableaux of a given linear program are shown below. Find the values of the unknowns a through l.

Starting Tableau

z	x_1	x_2	x_3	x_4	x_5	RHS
1	a	1	-3	0	0	0
0	b	c	d	1	0	6
0	-1	2	e	0	1	1

Current Tableau

z	x_1	x_2	x_3	x_4	x_5	RHS
1	0	$-\frac{1}{3}$	j	k	l	-4
0	g	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	0	f
0	h	i	$-\frac{1}{3}$	$\frac{1}{3}$	1	3

[10]

2 Let A be an $m \times n$ matrix, b an m-vector and c an n-vector. Show that the following problem is either infeasible or has an optimal solution having objective value zero.

$$\begin{aligned} \min \quad & c^{\mathrm{T}}x - b^{\mathrm{T}}y \\ \text{subject to} \quad & \\ Ax \quad & \geq b \\ -A^{\mathrm{T}}y & \geq -c \\ x, y & \geq 0, \end{aligned}$$

where x and y are vectors of appropriate order. [10]

3 Prove that the coefficient matrix $A_{(m+n)\times mn}$ of balanced transportation problem is totally unimodular. [10]

[P.T.O.]

4 Consider the Hungarian method for solving assignment problem. Show that the reduced cost matrix $\hat{C}$, at any iteration, has the properties: (i) all the elements are non-negative, and (ii) each row and column contains zero. [10]

5 Solve the following ILP by Balas' enumeration algorithm:

$$\begin{aligned} &\max\ z = x_1 + 2x_2 - 3x_3 \\ &\text{subject to} \\ &20x_1 + 15x_2 - \ x_3 \ \leq\ 10 \\ \text{or.}\quad &12x_1 - \ 3x_2 + 4x_3 \ \leq\ 20 \\ &\qquad x_1, x_2, x_3 \in \{0, 1\}. \end{aligned}$$

[10]

6 For $k > 0$, consider the non-linear programming problem:

$$\begin{aligned} &\max\ z = x_1 x_2^3 x_3 \\ &\text{subject to} \\ &x_1 + x_2 + x_3^3 \ \leq\ k, \\ &\quad x_1, x_2, x_3 \ \geq\ 0. \end{aligned}$$

Solve the above using dynamic programming formulation. [10]

———*** xXx ***———

INDIAN STATISTICAL INSTIUTE

Back paper Examination

M Stat. (Ist Year) (B – Stream) 2014 – 15

Subject: Time Series Analysis

Date: 16.07.2015 Full Marks: 100 Duration: 3 hours

Attempt all questions

1. Under infinite summability of ACVF, define spectral density.
Show that it is non-negative. Using suitable results check if the following functions on set of integers are ACVF's of some time series.

i) $\eta(h)$ = 1 if h = 0
= 0.5 if h = ±1
= 0 o.w.

ii) $\gamma(h)$ = 1 if h = 0
= -0.5 if h = ±1
= -0.25 if h = ±3
= 0 o.w.

[3 + 10 +(6 +6)]

2. Define time series and state main uses of it.
Discuss different components of a time series and state their significance.
How to estimate trend components using (i) Moving Average Method
(ii) Curve fitting Method.

[5 + 5 + (5 + 5)]

3. a) Let P (Y|$\underline{W}$) be the best linear predictor of Y on the basis of random vector $\underline{W}$.
Let Y, V and components of $\underline{W}$ have finite variances –

Show that P(P(Y|V,$\underline{W}$)|$\underline{W}$)
= P (Y|$\underline{W}$)

b) Consider an ARMA(1, 1) process with suitable difference equation with parameters $(\emptyset, \theta, \delta^2)$.
For different value of $\emptyset$, express the process as infinite linear combination of white Noise that defines it.
Hence find the ACVF

[10 + (10 + 5)]

4. a) What is partial Autocorrelation function? Discuss its properties and asymptotic distributions under AR(p) model.
How these can be used to test if a time series is AR(p)?

b) Describe two standard methods to test for a time series to be IID(0, δ^2) for some δ > o.

[(4 + 12 + 4) + 10]

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination : 2014-15
M. STAT. I YEAR
Optimization Techniques
Back Paper

Date: 17.04.2015 | Maximum Marks: 100 | Duration: 3 hours

Notation have usual meaning.

1 (a) Given the observations $x_1, x_2, \ldots, x_n$, we wish to find the median. Give a linear programming formulation of the problem.

(b) Show that the set of optimal solutions of convex programming problem is convex.

[10+10 = 20]

2 (a) Prove that the North-West corner rule produces basic feasible solution for balanced transportation problem.

(b) Solve the transportation problem with the data: $m = 3$, $n = 4$, $s = [15, 25, 10]^T$, $d = [5, 15, 15, 15]^T$, and the cost matrix

$$C = \begin{bmatrix} 10 & 2 & 20 & 11 \\ 12 & 7 & 9 & 20 \\ 4 & 14 & 16 & 8 \end{bmatrix}.$$

[10+20 = 30]

3 (a) Consider the Hungarian method for solving assignment problem. Prove that maximum number of independent zero cells in a reduced cost matrix is equal to the minimum number of lines required to cover all zeros in the matrix.

(b) Solve the assignment problem with the data: $m = 3$ and the cost matrix

$$C = \begin{bmatrix} 2 & 5 & 7 \\ 4 & 2 & 1 \\ 2 & 6 & 5 \end{bmatrix}.$$

[15+10 = 25]

4 (a) Consider a company that manufactures three products (A, B and C), each of which requires two types of resources (Machinery X and Machinery Y).

[P.T.O.]

Technology and resource capacity restrictions are given in the following table. For example, each unit of product A requires 7 units (say, of time) of Machinery X and 2 units of Machinery Y, and yields a profit of p_A.

Resource type (Machinery)	Amount of resources required to produce per unit of product			Current Available Capacity
	A	B	C	
X	7	3	1	≤ 28
Y	2	4	6	≤ 19
Unit Profit	p_A	p_B	p_C	

The company is considering capacity expansion by addition of new equipment. If the management decides to expand capacity of Machinery X, it must choose between adding either 5 or 15 new units of capacity, with an associated investment expense of 50 and 80, respectively. Similarly, if the management expands capacity of Machinery Y, the choices are 12 and 32 additional units of capacity, with investment costs of 30 and 90, respectively. However, the total investment expense must not exceed 150. The company is interested in the determination of production plan that maximizes its profit through consideration of possible capacity expansion.

(*i*) Formulate the above as an optimization problem.

(*ii*) In addition, how do you modify your formulation under the following two different situations:

(I) the company does not want to expand capacity of Machinery Y unless capacity of Machinery X is increased.

(II) the company does not want to expand capacity of Machinery Y unless it decides to increase capacity of Machinery X by 15 units.

(b) Formulate the following as a dynamic programming problem and solve the same.

$$\max \ z = \prod_{j=1}^{N} x_j$$

subject to

$$\sum_{j=1}^{N} x_j \leq q$$

$$x_j \geq 0, \ \forall j.$$

[(10+3)+(5+7) = 25]

——*** xXx ***——