INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

Second semester

M. Stat - First year 2015

Stochastic Processes

Date: 7th September, 2015

Maximum Marks: 40

Duration: 2 hours 30mins

Anybody caught using unfair means will immediately get 0. Please try to explain every step. Only class notes are allowed in the exam.

(1) Consider the Poisson process $N = \sum_{k\geq 1} \varepsilon_{j_k}$ on $(-\infty, \infty]$ with points $\{j_k\}_{k\geq 1}$ and mean measure $\mu(dx) = e^{-x}dx$. Let $Y_1 = \sup_{k\geq 1} j_k$ be the largest point and Y_2 be the second largest point.

(a) Compute the distribution of Y_1.

(b) Compute the joint distribution of Y_1, Y_2.

[4+4 points]

(2) Vehicles pass a crossing at the instants of a Poisson process of intensity λ; you need a gap length of at least a in order to cross. Let T be the first time at which you could succeed in crossing to the other side.

(a) Show that $E[T] = \frac{e^{a\lambda}-1}{\lambda}$;

(b) Find $E[e^{\theta T}]$ for $\theta > 0$;

(c) Suppose there are 2 lanes to cross, carrying independent Poissonian traffic with respective rates λ and μ. Find the expected time to cross in two cases when (i) there is an island between 2 lanes and (ii) you must cross both in one go. Which one is higher?

[4+2+2 points]

(3) Let Z_n be the size of the n-th generation of a branching process with $Z_0 = 1$ and $P(Z_1 = k) = 2^{-k}$ for $k \geq 0$. Show directly that as $n \to \infty$,

$$P(Z_n \leq 2yn | Z_n > 0) \to 1 - e^{-2y} \text{ for } y > 0.$$

[8 points]

(4) Consider a branching process with generation sizes Z_n satisfying $Z_0 = 1$, and $P(Z_1 = 0) = 0$. Pick two individuals at random (with replacement) from the n-th generation and let L be the index of the generation which contains their most recent ancestor.

(a) Show that $P(L = r) = E[Z_r^{-1}] - E[Z_{r+1}^{-1}]$ for $0 \leq r \leq n$.

(b) What can be said if $P(Z_1 = 0) > 0$?

[6+2 points]

(5) We know that extinction probability $q = \lim_{n \to \infty} P(Z_n = 0) = \lim_{n \to \infty} \varphi_n(0)$. If $p_0 > 0$, then show that for every $s \in [0, 1)$, $\varphi_n(s) \to q$ as $n \to \infty$.

[8 points]

(6) Let N be a Poisson process in $\mathbb{R}^2$ with constant intensity λ, and let $R_{(1)} < R_{(2)} < \cdots$ be the ordered distances from the origin of the points of the process.

(a) Show that $R_{(1)}^2, R_{(2)}^2, \cdots$ are points of a Poisson process on $[0, \infty)$ with intensity $\lambda\pi$.

(b) Show that $R_{(k)}$ has a density function

$$f(r) = \frac{2\pi\lambda r(\lambda\pi r^2)^{k-1}e^{-\lambda\pi r^2}}{(k-1)!}, \; r > 0.$$

[4+4 points]

INDIAN STATISTICAL INSTITUTE

MID-SEMESTRAL EXAMINATION

M.Stat (1st Year) (B-Stream) 2015-16

Subject: Statistical Inference

Full Marks: 40

Date: 09.09.2015 Duration: 1 hr. 40 mins.

Attempt all Questions

1.

a) Define admissible decision rules and complete class of decision rules.

b) Show that if minimal complete class of decision rules exists, then it is the set of all admissible rules.

c) Show by an example that the set of admissible rules may be non-empty, but minimal complete class may not exist.

d) By an example show that intersection of two complete classes may not be complete.

e) Let $X_1, X_2, \ldots, X_n$ $(n>1)$ be a sample of independent and identically distributed observations from $N(0, \sigma^2)$. Show that MLE for σ^2 may not be admissible under square error loss. [2+5+3+5+5]

2.

a) Let for $\Theta = \mathcal{A} = R$ we have the following loss function:

$$L(\theta, a) = \tfrac{1}{3}|\theta - a| \quad \text{if} \quad \theta \le a$$
$$= \tfrac{2}{3}|\theta - a| \quad \text{if} \quad \theta \ge a$$

Let $p(\theta/x)$ be the posterior density of θ when x is observed.

Derive the Bayes decision rule in terms of some t^{th} quantile of $p(\theta/x)$. Find t.

b) Let $\Theta = \{\theta_1, \theta_2\}$ and the risk set (i.e. the set of risk functions as subset of R^2) be given by

$\{(x_1, x_2) : (x_1 - 3)^2 + (x_2 - 5)^2 \le 3\}$

Derive the minimax rule i.e the point denoting minimax rule.

c) For a statistical decision problem the following are given:

$\Theta = \mathcal{A} = [0, \infty)$, $L(\theta, a) = |\theta - a|$ for $\theta \in \Theta$ and $a \in \mathcal{A}$

$f(x/\theta) = 1/\theta$, $0 < x < \theta$

Prior $\Pi(\theta) = \theta e^{-\theta}$, $0 < \theta < \infty$

Find the Bayes decision rule with respect to Π.

[8+6+6]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2015-16

M. Stat. I Year
Categorical Data Analysis

Date: 11. 09. 15 **Full Marks: 60** **Time: 3 hours**

The paper carries 70 marks. Answer as many as you can.

1(a) Consider a multinomial distribution with parameters $(n, \pi_1, \pi_2, \ldots, \pi_k)$. For testing $H_0 : \pi_j = \pi_{j0}, j = 1, 2, \ldots, k$, using observed multinomial proportions $\{\hat{\pi}_j = (n_j/n)\}$, show that the likelihood ratio statistic is

$$G^2 = -2n \sum_{j:\hat{\pi}_j > 0} \hat{\pi}_j \log\left(\frac{\pi_{j0}}{\hat{\pi}_j}\right).$$

Show that G^2 is well defined under the null and $G^2 \geq 0$ with equality if and only if $\hat{\pi}_j = \pi_{j0}$ for all j.

(b) For observed $\{n_j\}$ define the divergence measure

$$D(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_{j:\hat{\pi}_j > 0} n_j \left[(n_j/n\pi_{j0})^{\lambda} - 1\right]$$

Show that $\lim_{\lambda \to 0} D(\lambda) = G^2$.

[12+8=20]

2(a) For $I \times J$ contingency tables, explain why the variables are independent when the $(I-1) \times (J-1)$ differences $\pi_{j|i} - \pi_{j|I} = 0$, $i = 1, 2, \ldots, (I-1)$, and $j = 1, 2, \ldots, (J-1)$.

(b) For a diagnostic test of a certain disease, let π_1 denote the probability that the diagnosis is positive given that the subject has the disease, and let π_2 denote the probability that the diagnosis is positive given that the subject does not have the disease. Let ρ denote the probability that a subject has the disease. Find the probability that the subject has the disease with positive diagnosis.

[15+10 = 25]

P.T.O.....

3 (a) Describe the Fisher's exact test to test the hypothesis of no association in a 2
contingency table. When a test statistic has continuous distribution, the P-v
has an uniform distribution under the null. For Fisher's exact test explain
$\Pr(P - \text{value} \leq \alpha) \leq \alpha$.

(b) Suggest a suitable experimental design in a 2×2 table which produces an unbi
estimate of the odds ratio.

(c) Consider a 2×2 contiongency table of joint distribution of two binary random v
able X and Y. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent and identic
distributed replications of (X, Y). Find the conditional distribution of $\sum_j X$
given $\sum_j X_i = r$ and $\sum_j Y_j = t$ for suitable for different choices of nonnega
integers (r, t).

[10 + 5+ 10 =

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : Semester I (2015-2016)

M. Stat 1st Year

Multivariate Analysis

Date: 14. 9. 15 | Maximum marks: 60 | Time: 2 hours.

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

1. Suppose $\mathbf{X}$ is an $n \times p$ data matrix from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$, $\mathbf{J}$ is a $p \times p$ matrix with all entries equal to one. It is known that $\sigma > 0$ and $-\frac{1}{(p-1)} < \rho < 1$.

   Find the Maximum likelihood estimates of μ, σ^2 and ρ. [16]

   [Hint: You may use the eigen values and eigen vectors of $\boldsymbol{\Sigma}$ and its spectral decomposition.]

2. Suppose based on n_i i.i.d. samples available from two independent multivariate populations $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for $i = 1, 2$, we want to test the hypothesis that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

   (a) Derive the LRT statistic and also derive its distribution under the null hypothesis. [8+8]

   (b) Derive also the UIT statistic for this problem, and compare it with the LRT statistic derived in part (a). [8]

3. For a two-class problem with equal prior probabilities and misclassification costs, let the class densities be $p_i(\mathbf{x}) = N_p(\boldsymbol{\mu}_i, \sigma^2 I)$, for $i = 1, 2$.

   (a) Express the Bayes optimal error rate P^* in terms of the c.d.f. of the standard normal density and $\frac{||\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1||}{2\sigma}$. [10]

   (b) Let, $\boldsymbol{\mu}_1 = 0$ and $\boldsymbol{\mu}_2 = \mu\mathbf{1}_p$, where $\mathbf{1}_p$ is the p-dimensional vector with all entries equal to one. Show that in this case, P^* goes to 0 as p goes to infinity. Give an interpretation of this result. [10]

Part - B

Assignments [10]

Indian Statistical Institute

M.Stat. First Semester, Mid-Semester Exam: 2015

Topic: Regression Techniques

Maximum Marks: 60, Duration: 2 hours

16.09.2015

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 $Y = X\beta + \epsilon$, with p predictors.
   (a) What are the model assumptions for GLS method of estimation?
   (b) Find the GLS estimate of β and give an expression for Variance($\hat{\beta}_{GLS}$). [2+4+4]

2. (a) Let $\beta_1, \beta_2, \beta_3$ be the interior angles of a triangle. Suppose that we have available estimates Y_1, Y_2, Y_3 of $\beta_1, \beta_2, \beta_3$ respectively. We assume $Y_i \sim N(\beta_i, \sigma^2)$, for $i = 1, 2, 3$, σ unknown and that Y_i's are independent. What is the F test for testing the null hypothesis that the triangle is equilateral?
   (b) Discuss the importance of PRESS R^2 in regression diagnostics. [6+4]

3. (a) What is the LASSO estimate of β under the standard multiple regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$? Show that the LASSO estimate of β is indeed the posterior mode with respect to a Laplace prior.
   (b) Let W and V be two orthogonal subspaces with projection matrices P_w and P_v respectively. Show that $||P_w y||^2$ and $||P_v y||^2$ are independently distributed, where $y \sim N(\mu, \sigma^2 I_n)$. [2+4+4]

4. (a) While testing the set of hypotheses $H_0 : \beta_j = 0$, vs. $H_1 : \beta_j \neq 0$, for $j = 1, 2, \ldots, 10$, how can you control the false discovery rate (FDR) at level 0.05?
   (b) Suppose data are collected on body mass index (BMI), blood sugar level (BSL) and hypertension level (HL) for 100 patients. Assuming that HL can be considered as an unknown function of BMI and BSL, how will you predict the HL of a patient if you know that her BMI is 25 and BSL is 88? [5+5]

5. (a) Suppose we have measurements on n subjects for some response variable Y and three predictors X_1, X_2, X_3. A linear regression model is proposed as: $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$, where ϵ_i's are iid with some unknown distribution. Propose a suitable non-parametric approach to test the importance of the predictor X_3 on the response.
   (b) What is the use of variance inflation factor (VIF) in regression analysis? [6+4]

6. The HIV Epidemiology Research Study (HERS) was a longitudinal cohort study of the natural history of HIV in women. Between 1993 and 1996, HERS enrolled 1310 women who were either HIV positive or at high risk of infection. Every 6 months for

up to 5 years, outcomes like CD4 count, body mass index, depression status, drug use behavior were recorded. Our focus is to model CD4 count on the other measures and to investigate the effect of time on the response as well.

(a) Write down a regression model for the above data.

(b) In a fully parametric framework, write down the likelihood function to be maximized for the parameter estimation. [5+5]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

Second semester

M. Stat - First year 2015

Stochastic Processes

Date: 16th November, 2015

Maximum Marks: 60

Duration: 3 hours 30 mins

Anybody caught using unfair means will immediately get 0. Please try to explain every step. Only class notes are allowed in the exam.

1. The excess or residual lifetime of a renewal process $N(t)$ be denoted by $E(t)$. Suppose $\{X_n\}$ denote the interarrival times. Suppose X_1 is non-arithmetic with finite mean μ.
   (a) Show that $E(t)$ converges in distribution as $t \to \infty$, the limit distribution being
   $$H(x) = \int_0^x \frac{1}{\mu}(1 - F(y))dy.$$
   (b) Show that the r-th moment of this limit distribution is given by $\mathsf{E}[X_1^{r+1}]/\mu(r+1)$.
   (c) Show that
   $$\lim_{t\to\infty} \mathsf{E}[E(t)^r] \to \frac{\mathsf{E}[X_1^{r+1}]}{\mu(r+1)}.$$
   [4+2+4=10]
2. Consider the birth-death process X with $\lambda_n = n\lambda$ and $\mu_n = n\mu$ for all $n \geq 0$. Suppose $X(0) = 1$ and let $\eta(t) = P(X(t) = 0)$. Show that η satisfies the differential equation
   $$\eta'(t) + (\lambda + \mu)\eta(t) = \mu + \lambda\eta(t)^2.$$
   Hence find $\eta(t)$ and calculate $P(X(t) = 0|X(u) = 0)$ for $0 < t < u$. [4+2+4=10]
3. Let $\{\chi_i\}$ be i.i.d. integer valued independent of a Poisson process $N(t)$. Define $X(t) = \sum_{i=1}^{N(t)} \chi_i$.

(a) Show that $X(t)$ is a continuous time Markov chain.

(b) Obtain the backward differential equations.

(c) Find the infinitesimal matrix.

[3+3+4=10]

4. (a) Show that $\mathsf{E}[X|X > 0] \le \mathsf{E}[X^2]/\mathsf{E}[X]$ for any random variable X taking non-negative values.

(b) Let Z_n be the size of the n-th generation of a branching process with $Z_0 = 1$ and $P[Z_1 = k] = qp^k$ for $k \ge 0$ where $p > 1/2$. Use (a) to show that $\mathsf{E}[Z_n/\mu^n|Z_n > 0] \le 2p/(p-q)$ where $\mu = p/q$. Also show that $\mathsf{E}[\frac{Z_n}{\mu^n}|Z_n > 0] \to \frac{p}{p-q}$ as $n \to \infty$.

[3+7=10]

5. Consider a pure renewal process $N(t)$. Denote $U(t) = \mathsf{E}[N(t)]$.

(a) Show $V(t) := \mathsf{E}[N(t)^2] = \sum_{n=1}^{\infty}(2n-1)F^{(n-1)*}(t)$.

(b) Find the Laplace transform of $\sum_{n=1}^{\infty} nF^{(n-1)*}(t)$. Use this to write $V(t)$ in terms of $U(t)$ and derive the renewal equation.

(c) Find an expression for V only in terms of U.

(d) What is $V(t)$ if N is a Poisson process?

[3+3+2+2=10]

6. Consider a renewal process N and suppose that arrival is 'overlooked' with probability q, independently of all other arrivals. Let $M(t)$ be the number of arrivals which are deleted upto time t/p where $p = 1 - q$.

(a) Show that M is a renewal process whose inter-arrival time distribution function G_p is given by

$$G_p(x) = \sum_{r=1}^{\infty} pq^{r-1}F_r(x/p)$$

where F_n is the distribution function of the time of the n-th arrival in the original process N.

(b) Find the characteristic function of G_p in terms of F and show that $\lim_{p\to 0} G_p(s) = 1 - e^{-s/\mu}$ for $s > 0$, so long as inter-arrival times in the original process have finite mean μ. [5+5=10]

Indian Statistical Institute

M.Stat. First year, First Semestral Exam: 2015-16

Regression Techniques

Dt: 19.11.15 Maximum Marks: 60, Duration: 3 hours

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

1. (a) 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?
(b) In a biomedical study, data are collected on total medical expenditure and related covariates for a particular season of the year from n individuals. In the data, $n_1 (< n)$ responses are found to be exactly zero and the non-zero (continuous) responses are denoted by $Y_1, Y_2 \ldots, Y_{n_2}$, where $n_1 + n_2 = n$. There are p regressor variables $X_1, X_2, \ldots, X_p$ in the data. Suggest a suitable regression model for predicting the medical expenditure for the $(n+1)$-th individual given the relevant information on the predictors. [5+5]

2. (a) Discuss the importance of distance correlation in the context of variable selection problem. State three important properties of distance correlation.
(b) Consider a time varying coefficient (TVC) linear regression model for network modeling. How will you test the hypothesis that the effect of the "nearest neighbors" is NOT time-varying? [5+5]

3. (a) Define the "systematic component" of a generalized linear model (GLM). What is meant by "canonical link"? For a Binomial GLM, find the canonical link function.
(b) Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables from general exponential family with linear predictor $\eta_i = x_i^T \beta$ and the link function g. Considering p predictors in the model, how will you use Fisher scoring method to obtain $\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ denotes the MLE of the vector of regression coefficients $\boldsymbol{\beta}$. [4+6]

4. (a) What is the difference between "intermittent missingness" and "monotone missingness" for longitudinal responses? Suggest a suitable GLM based approach for imputing the intermittent missing values.
(b) Propose a step by step algorithm for imputing the monotone missing values under the available case missing value (ACMV) restriction. Is this restriction more powerful than the complete-case missing value (CCMV) restriction? Explain. [5+5]

5. (a) In Bayesian framework model parameters are assumed to be random variables with some prior distributions. The regression model with p predictors, response variable (Y) and design matrix X can be written as the following: $Y | X, \boldsymbol{\beta}, \sigma^2 \sim N(X\boldsymbol{\beta}, \sigma^2 I)$.

Suggest suitable priors for $\boldsymbol{\beta}$ and σ^2 so that under squared error loss function, the posterior estimates of $\boldsymbol{\beta}, \sigma^2$ become identical with the respective OLS estimates in the classical approach. Show your work explicitly.

(b) How will you get a 95% Bayesian confidence interval for $\boldsymbol{\beta}$ from the posterior distribution. [6+4]

6. In a state-space model, observations are collected from n subjects at T evenly spaced time points. The observation at time t is modeled as the following: $y_t = x_t\beta_t + e_t$, where e_t's are iid Normal $(0,\sigma_e^2)$. Note that x_t denotes the predictor value at time t. The unknown states β_t's are modeled using a first order auto-regressive model as the following: $\beta_t = \beta_{t-1} + \epsilon_t$, where ϵ_t's are iid Normal $(0,\sigma_\epsilon^2)$.

   (a) Write down the likelihood function for the above model.

   (b) Assuming suitable prior distributions for the model parameters, write down the joint posterior distribution.

   (c) Give a suitable MCMC algorithm for estimating the model parameters. [3+2+5]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : (2015-2016)

M. Stat 1st Year

Multivariate Analysis

Date: 23. 11. 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. Consider a two-class classification problem in which each component $X_j, j = 1, \ldots, M$ of the measurement vector $\mathbf{X}$ is either 0 or 1 with conditional probabilities

$$p_j = P\{X_j = 1 \mid \text{Class } 1\} \quad \text{and} \quad q_j = P\{X_j = 1 \mid \text{Class } 2\}.$$

Assume that the components of $\mathbf{X}$ are independent and the prior class probabilities are equal.

(a) Find the Bayes classification rule for the problem. [6]

(b) If, further $p_j = p > \frac{1}{2}$ and $q_j = 1 - p$, $j = 1, \ldots, M$, show that for M odd, the Bayes error probability is given by

$$P_e(M,p) = \sum_{l=0}^{(M-1)/2} \binom{M}{l} p^l (1-p)^{M-l}.$$ [9]

(c) Show that $P_e(M,p)$ approaches zero as $M \to \infty$. [5]

2. (a) Let X be a p-dimensional random vector. Show that no standardized linear combination (SLC) of X has a variance larger than that of the first principal component of X. [7]

(b) Show that the variance of any SLC of X which is uncorrelated with the first k principal components of X can not be larger than the variance of the $(k+1)$-th principal component of X. [8]

(c) Suppose that $\mathbf{X} = (X_1, X_2)$ has a bivariate multinomial distribution with $n = 1$, so that $X_1 = 1$ with probability p and $X_2 = 1 - X_1$. Find the principal components of X and their variances. [10]

3. Describe the Principal Factor Analysis method. [15]

[P. T. O.]

4. (a) Describe the naive density estimator and express it as a kernel density estimator. 6

(b) Calculate the bias of the naive density estimator $\hat{f}(x)$ for all real values of x where $f(x)$ is the uniform(0,1) density and the bandwidth $h \in (0, 1/3)$. 14

5. Consider the problem of One-way classification in Multivariate Analysis of Variance. Derive the likelihood ratio test statistic for testing the significance of a fixed contrast. 15

Part - B

Assignments 15

INDIAN STATISTICAL INSTIUTE

Semestral Examination
M Stat. (Ist Year) (B – Stream) 2015 – 16

Subject: Statistical Inference

Date: 25.11.2015 Full Marks: 60 Duration: 3 hours

Attempt all questions

1. (a) Let $x_1, x_2, \ldots\ldots, x_n$ are i. i. d. Normal (μ, σ^2). For fixed $\sigma_0 > 0$, derive likelihood ratio procedure to decide between ω_0: $\sigma < \sigma_0$ and ω_1: $\sigma \geq \sigma_0$.
[Here μ and σ are unknown.]

(b) Construct an example to exhibit possible inadmissibility of likelihood ratio test.

[10 + 10]

2. Consider a testing problem where null and alternate hypotheses are composite and have a common boundary. In that context define similar test and test to have Neyman structure with respect to some statistic. Use this to derive UMPU test for one sided hypothesis in context of two population Poisson problem.

[2 + 2 + 11]

3. (a) Let set of States of Nature be finite and risk set S is a subset of R^k. If S is closed and bounded show that for every prior distribution, Bayes rule exists.

(b) If in Problem 3a), S is only closed and bounded below, show by an example that for some prior, Bayes rule may not exist.

[10 + 5]

4. Let in a statistical game, value of the game exists. Show that any minimax rule for the statistician is an extended Bayes rule.

[10]

---------- O ----------

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2015-16

M. Stat. I Year

Categorical Data Analysis

Date: 27. 11. 15 **Full Marks: 100** **Time: 3 hours**

The paper carries 110 points. Answer as many as you can.

1(a) Let Y have a multinomial distribution ($n = 1$), taking values in k cells with probabilities $\pi_1, \pi_2, \ldots, \pi_k$ repectively. Cosider the measure of diversity defined by

$$\mathrm{V}(Y) = \sum_{i=1}^{k} \pi_i (1 - \pi_i).$$

Show that $\mathrm{V}(Y) = \Pr\{Y_1 \neq Y_2\}$ where Y_1 and Y_2 are two independent samples from the same multinomial.

(b) For the proportional reduction in variation in a $I \times J$ table (cell probabilities π_{ij}'s etc. in standard notations) with possibly dependent rows (X) and columns (Y), show that $E\left[V(Y|X)\right]$ where $V(Y|X)$ is the measure of diversity defined above conditional of the rows (X) is equal to $\left(1 - \sum_i \sum_j \pi_{ij}^2/\pi_{i+}\right)$.

[10+15 =25]

2(a) Define generalized linear model with exponential family of distribution for the dependent variable with examples. What is canonical link?
For binary data, define a GLM using the log link. Show that effects refer to the relative risk. Give an example when this link is not advisable.

(b) Let T denote the time to some event, with pdf f and cdf F. For subject i, let t_i denote the time to the event and let $y_i = 1$ if the event occurs and 0 for censoring (that is, experiment terminated before t_i). Let $T = \sum_i t_i$ and $Y = \sum_i y_i$. Assuming $f(t) = \lambda \exp(-\lambda t)$ obtain a simplaified expression of the (censored) log likelihood equation and the MLE of λ.

(c) Tha hazard function represents the instanteneous rate of the occurance of the event for subjects who survided upto time t (that is, $f(t)/[1 - F(t)]$). In the previous example assume there is a real valued regressor X and subjects are independent with T_i having an exponential distribution with mean $[\lambda \exp(\beta x_i)]^{-1}$. Show that the (full) log likelihood (assuming all t_i's are observed).

P. T. O...

$$L(\lambda, \beta) = \sum_i y_i \log \mu_i - \sum_i \mu_i - \sum_i y_i \log t_i,$$

where $\mu_i = t_i \lambda \exp(\beta x_i)$.

[10+15+20 = 45]

3 (a) Consider logistic regression modeling of a binary Y on real valued regressor X. Suppose for the population $Y = j$, X has a $\mathrm{N}(\mu_j, \sigma^2)$ distribution, $j = 0, 1$. Show that $\Pr\{Y = 1|x\}$ satisfies the logistic regression model with $\beta = (\mu_1 - \mu_0)^2/\sigma^2$.

(b) If $[X|Y = j]$, $j = 0,1$ in the above example, are given by $\mathrm{N}(\mu_j, \sigma_j^2)$ for $j = 0,1$ respectively, with $\sigma_0 \neq \sigma_1$ show that the logistic regression for $\Pr\{Y = 1|x\}$ still holds (need not be simple linear).

[10+15=25]

4. Consider a scenario with $J = 3$ outcome categories, suppose that

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta_j x)}{[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)]},$$

j=1,2. Show that $\pi_3(x)$ is

(i) decreasing in x if $\beta_1 > 0$ and $\beta_2 > 0$,

(ii) increasing in x if $\beta_1 < 0$ and $\beta_2 < 0$,

(iii) nonmonotonic when β_1 and β_2 have different signs.

[5+5+5 =15]

Midterm Examination
Large Sample Statistical Methods
Second Semester
2015-2016 Academic Year
M.Stat. First Year (B-Stream Only)

Date : 22.02.16 Maximum Marks: 40 Duration :- $2\frac{1}{2}$ hours

Answer as many quyestions as you can. The maximum you can score 40.

1. Suppose $X_n \sim N(0, 1+\frac{1}{n})$ and $Y_n = X_n I(|X_n| \le n)$ for all $n \ge 1$. Does $e^{X_n} - e^{Y_n}$ converge in probability ? Justify your answer. [3]

2. Suppose $X_n \sim \text{Beta}(\frac{1}{n}, \frac{1}{n})$ for all $n \ge 1$. Does $\{X_n\}$ converge in distribution ? Prove your assertion. [5]

3. Let $A_1, A_2, \ldots$ be independent events. Prove that a necessary and sufficient condition for $P(\cup_{j=1}^{\infty} A_{i_j})$ to equal 1 for every subsequence $1 \le i_1 < i_2 < \ldots$ of integers is : $\liminf_{n\to\infty} P(A_n) > 0$. [4]

4. Let $X_1, \ldots, X_n$ be iid with a Poisson (λ) distribution with $\lambda > 0$. Are there constants a_n and b_n such that $a_n\left((1-\frac{1}{n})^{n\bar{X}_n} - b_n\right)$ converges in distribution to a non-degenerate random variable, where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$? You are allowed to consider constants a_n and/or b_n which are dependent on λ. Justify your answer. [5]

5. Let $X_1, \ldots, X_n$ be iid from some distribution F with mean 0 and unknown variance σ^2. Assume that $E(|X|^4) < \infty$ for all $n \ge 1$. A common test for $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 > 1$ rejects H_0 when $\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 > \chi^2_{\alpha,n-1}$, where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ and $\chi^2_{\alpha,n-1}$ is the upper α point of a central chi-square distribution with $(n-1)$ degrees of freedom. Does $P_{H_0}(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 > \chi^2_{\alpha,n-1})$ converge to α as $n \to \infty$ if the kurtosis κ of the distribution F is non-zero ? Justify your answer. Recall that $\kappa = \frac{\mu_4}{\mu_2^2} - 3$. [6]

6. (a) Let $X_1, X_2, \ldots X_n$ be iid observations from a $N(\mu, 1)$ distribution with $\mu \in \mathcal{R}$ unknown. Find the joint asymptotic distribution of (properly centred and scaled) sample mean and sample median. Suppose now a random sample of size 100 has been drawn from a $N(1,1)$ distribution and you are only told that the sample median is 1.1. Based on this information, give a reasonable estimate of the sample mean. Justify your answer. [4+2=6]

   (b) Suppose $X_1, \ldots, X_n$ are iid from a distribution F given by $F(x) = xI_{(0\le x<\frac{1}{2})} + (2x - \frac{1}{2})I_{(\frac{1}{2}\le x\le\frac{3}{4})}$. Let $\hat{\zeta}_{\frac{1}{2},n}$ denote the smallest sample median based on $X_1, \ldots, X_n$. For $n = 10000$, how will you approximate the probability that the random variable $\hat{\zeta}_{\frac{1}{2},n}$ is larger than 0.51 ? Justify your answer. [2]

7. Let $X_1, X_2, \ldots$ be independent random variables with

$$\begin{aligned} X_k &= -ke^k \text{ with probability } e^{-2k}, \\ &= +ke^k \text{ with probability } e^{-2k}, \\ &= -k \quad \text{with probability } \frac{1}{2} - e^{-2k}, \\ &= +k \quad \text{with probability } \frac{1}{2} - e^{-2k}. \end{aligned}$$

Let $S_n = X_1 + \ldots + X_n$. Can you find constants $a_n > 0$ such that $a_n S_n$ converges to a non-degenerate distribution ? Prove your assertion. [5]

8. Let P and Q be two probability measures on $(\Omega, \mathcal{A})$. Then show that statement (a) below implies statement (b).

   (a) For any $A \in \mathcal{A}$, $\{P(A) = 0\}$ implies $\{Q(A) = 0\}$.

   (b) For each $\epsilon > 0$, there exists $h_\epsilon > 0$ such that $\{A \in \mathcal{A}$ and $P(A) < h_\epsilon\}$ implies $\{Q(A) < \epsilon\}$. [5]

9. Suppose you have a sequence of i.i.d observations $(X_1, Y_1), (X_2, Y_2), \ldots$ from a bivariate distribution G supported on the unit disc $\{(x, y) : 0 \leq x^2 + y^2 \leq 1\}$. Suppose further that the distribution has a continuous density $g(x, y)$ with $g(0, 0) > 0$. Let $D_i = \sqrt{X_i^2 + Y_i^2}$, for $i = 1, 2, \ldots$. Let $D_{(1,n)} = \min\{D_1, \ldots, D_n\}$. Find real constants a_n and $b_n > 0$ such that $b_n(D_{(1,n)} - a_n)$ converges to a non-degenerate distribution. Prove your answer and write the explicit form of the limiting distribution function. [5]

Indian Statistical Institute, Kolkata

Midterm examination: Second semester 2015 - '16

M.Stat I year

Subject	:	**Metric Topology and Complex Analysis**
Time	:	2 hours 30 minutes
Maximum score	:	45

Attempt all the problems. Please use a new page to answer each problem and make sure that the problem number in the margin can be read, even after stapling. If you attempt the same problem several times, please strike out all the attempts except the correct one before submitting your answer script. Justify every step in order to get full credit of your answers. All arguments should be clearly mentioned on the answer script. Points will be deducted for missing or incomplete arguments.

(1) Prove that, if E is an infinite subset of a compact set K, then E has a limit point in K.

[9 marks]

(2) Given a subspace A of a metric space (X, d), prove that the function $x \mapsto d(x, A)$ is uniformly continuous.

[8 marks]

(3) Suppose (X, d) is a compact metric space and $f : X \to X$ is a function satisfying $d(f(x), f(y)) < d(x, y) \; \forall \, x, y \in X, x \neq y$. Prove that there exists $x_0 \in X$ such that $f(x_0) = x_0$.

[14 marks]

(4) Let (X, d_1) and (Y, d_2) be two totally bounded metric spaces. Prove that the product space $X \times Y$ equipped with the metric $d((x_1, y_1), (x_2, y_2)) = \sqrt{d_1(x_1, x_2)^2 + d_2(y_1, y_2)^2}$ is again totally bounded.

[7 marks]

(5) Decide whether the following pairs of sets are homeomorphic. Justify your answer in each case.

(i) $\mathbb{R}$ and $\mathbb{R}^2$

(ii) $[a, b]$ and $[a, b)$ where $a, b \in \mathbb{R}$

(iii) $(0, 1)$ and (a, b) where $a, b \in \mathbb{R}$

[4+4+4=12 marks]

INDIAN STATISTICAL INSTITUTE

Mid-Semeseter of 2nd Semester Examination : 2015-16

Course Name : **M.Stat. 1st Year**

Subject Name : **Sample Survey and Design of Experiments**

Date : Feb 24, 2016 **Total Duration : 45 mins + 45 mins = $1\frac{1}{2}$ hrs**

Note: Use separate answer sheets for two groups.

Group – Sample Survey. (Total Marks = 15)

Answer any three 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. 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)

3. 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)

4. Let $P_i(0 < P_i < 1, \sum_{i=1}^{N} P_i = 1)$ be known numbers associated with the units i of a population U. Suppose on the first draw a unit i is chosen from U with probability P_i and on the second draw a unit $j(\neq i)$ is chosen with probability $\frac{P_j}{1-P_i}$.

For a sample of size 2 drawn under this sampling scheme, write down Des Raj's (1956) unbiased estimator for Y. Improve that estimator through Rao-Blackwellization.

(5)

INDIAN STATISTICAL INSTITUTE
MID-SEMESTRAL EXAMINATION 2016

M.STAT 1st year. Design of Experiments

February 24, 2016, Total marks 15 Duration: 45 minutes

Answer all questions.

Keep your answers brief and to the point.

1. Suppose you are to weigh 4 objects using a weighing balance with two pans; the balance needing zero bias correction. You are allowed to make 4 observations and suppose all observations are independent with constant variance σ^2.

   a) Derive a lower bound to the variance of the best linear unbiased estimators of these weights. [3]

   b) Let the design matrix be $X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$

   Let Y_1, Y_2, Y_3, Y_4 be the 4 observations. Obtain the estimators of the 4 unknown weights and the variance of the estimators. [4]

2. a) Give an example of an experimental situation where you would recommend the use of a block design. [2]

   b) Write down a suitable model for analysing data from general block designs. [1]

   c) Consider the following design with 5 treatments labeled $1, 2, \ldots, 5$ and 4 blocks as follows:

Block No.	Treatments
1	2, 3, 3
2	1, 5
3	1, 2
4	4, 5

   (i) Check if the following pairwise contrasts are estimable: [3]

   a) contrast between 1 and 2.

   b) contrast between 3 and 5

   c) contrast between 4 and 5.

   (ii) Is this design connected? [2]

INDIAN STATISTICAL INSTITUTE

Mid-Semeseter of 2nd Semester Examination : 2015-16

Course Name : **M.Stat. 1st Year**

Subject Name : **Sample Survey and Design of Experiments**

Date : Feb 24, 2016 **Total Duration : 45 mins + 45 mins** $= 1\frac{1}{2}$ **hrs**

Note: Use separate answer sheets for two groups.

Group – Sample Survey. (Total Marks = 15)

Answer any three 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. 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)

3. 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)

4. Let $P_i(0 < P_i < 1, \sum_{i=1}^{N} P_i = 1)$ be known numbers associated with the units i of a population U. Suppose on the first draw a unit i is chosen from U with probability P_i and on the second draw a unit $j(\neq i)$ is chosen with probability $\frac{P_j}{1-P_i}$.

For a sample of size 2 drawn under this sampling scheme, write down Des Raj's (1956) unbiased estimator for Y. Improve that estimator through Rao-Blackwellization.

(5)

INDIAN STATISTICAL INSTITUTE
MID-SEMESTRAL EXAMINATION 2016

M.STAT 1st year. Design of Experiments

February 24, 2016, Total marks 15 Duration: 45 minutes

Answer all questions.

Keep your answers brief and to the point.

1. Suppose you are to weigh 4 objects using a weighing balance with two pans; the balance needing zero bias correction. You are allowed to make 4 observations and suppose all observations are independent with constant variance σ^2.

   a) Derive a lower bound to the variance of the best linear unbiased estimators of these weights. [3]

   b) Let the design matrix be $X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$

   Let Y_1, Y_2, Y_3, Y_4 be the 4 observations. Obtain the estimators of the 4 unknown weights and the variance of the estimators. [4]

2. a) Give an example of an experimental situation where you would recommend the use of a block design. [2]

   b) Write down a suitable model for analysing data from general block designs. [1]

   c)Consider the following design with 5 treatments labeled $1, 2, \ldots, 5$ and 4 blocks as follows:

Block No.	Treatments
1	2, 3, 3
2	1, 5
3	1, 2
4	4, 5

   (i) Check if the following pairwise contrasts are estimable: [3]

   a) contrast between 1 and 2,

   b) contrast between 3 and 5

   c) contrast between 4 and 5.

   (ii) Is this design connected? [2]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2015-16 (Second Semester)

M. STAT. I YEAR

Abstract Algebra

Date: 25.02.16 | Maximum Marks: 60 | Duration: $3\frac{1}{2}$ Hours

Attempt Question 6 and ANY FOUR from the rest.

$\mathbb{Q}$ denotes the field of rational numbers,
$\mathbb{R}$ denotes the field of real numbers and
$\mathbb{C}$ denotes the field of complex numbers.

1. Let $A = \mathbb{R}[X]/(X^4 - X)$.

   (i) Prove that A is isomorphic (as a ring) to the product ring $\mathbb{R} \times \mathbb{R} \times \mathbb{C}$.

   (ii) Which element of $\mathbb{R} \times \mathbb{R} \times \mathbb{C}$ corresponds to the element $\overline{X}$ of A?

   (iii) Describe all prime and maximal ideals of A. [5+2+5=12]

2. Let $B = \mathbb{R}[X, Y, Z]/(X^2 + Y^2 + Z^2 - 1)$.

   (i) Prove that B is a Noetherian integral domain.

   (ii) Show that $\overline{X} - 1$ is a prime element of B.

   (iii) Examine if B is a PID. [5+3+4=12]

3. Let k be a field.

   (i) Prove that there are infinitely many distinct monic irreducible polynomials in $k[X]$.

   (ii) Deduce that $k(X)$ is not finitely generated as a k-algebra. [5+7=12]

4. (i) Let R be a UFD with field of fractions $K (\neq R)$. Prove that K cannot be an algebraically closed field.

   (ii) Let k be a field. If t is an element in a field extension of k such that t is transcendental over k, then show that the polynomial $X^n - t$ is irreducible in $k(t)[X]$. [6+6=12]

5. (i) Prove that $h(X,Y) := X^6 + 3X^2Y^2 + Y^4 + 2X^2Y + 3Y^2 + 5Y$ is irreducible in $\mathbb{Q}[X,Y]$.

   (ii) Let $f(X,Y)(\neq 0) \in \mathbb{Q}[X,Y]$. Show that there exist $a, b \in \mathbb{Q}$ such that $f(a,b) \neq 0$. [6+6=12]

6. State whether the following statements are TRUE or FALSE with brief justification. Attempt ANY FIVE.

   (i) If R is any commutative ring and $f(X)$ a non-constant polynomial in $R[X]$ then $f(X)$ has at most finitely many roots in R.

   (ii) If P is a prime ideal of a commutative ring R and I any ideal of R, then the image of P in R/I is a prime ideal of R/I.

   (iii) If B is a subring of A such that B and A are isomorphic as rings, then $B = A$.

   (iv) If $R \subset A \subset R[X_1, \dots, X_n]$ are commutative rings with A being a Noetherian ring, then R must be a Noetherian ring.

   (v) If $L|_k$ is a field extension such that $[L : k]$ is a prime number, then $L = k(\alpha)$ for every $\alpha \in L \setminus k$.

   (vi) If L is a field containing $\mathbb{C}$ and a is an element of L such that $\mathbb{C}[a] = \mathbb{C}(a)$, then $a \in \mathbb{C}$. [$3 \times 5 = 15$]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination (2015–2016)

M STAT I

Optimisation Techniques

Date : 25.02.2016 Maximum Marks : 60 Time : 2 hrs.

This paper carries 70 marks. Maximum you can score is 60. Precisely justify all your steps. Carefully state all the results you are using.

1. Does there exist $x_1, x_2, x_3 \geq 0$ such that

$$\begin{aligned} 2x_1 + 5x_2 + x_3 &= 3 \\ -3x_1 + 8x_2 + x_3 &= -5? \end{aligned}$$

Justify your answer. [10]

2. Consider the following LP problem P:

Minimise $x_1 + x_2$ subject to

$$\begin{aligned} x_1 + 2x_2 &\geq 3 \\ 2x_1 + x_2 &\geq 5 \\ x_1, x_2 &\geq 0 \end{aligned}$$

(a) Solve the problem graphically. [4]

(b) Write down the dual P^*. [3]

(c) Solve P^* using Duality Theorem. [3]

(d) Consider the LP problem P without the constraint $x_1 \geq 0$. How does that affect the optimal solution? [3]

(e) Write the LP problem P in equational form, call it P_1. [2]

(f) Write down the dual P_1^*. Is this different from P^*? [3+2]

3. Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$ where every entry of both A and b are nonnegative and A is of full rank. Let $c \in \mathbb{R}^n$. Show that the following LP problem has an optimal solution: [10]

$$\text{Minimise } c^T x \text{ subject to } x \geq 0 \text{ and } Ax \leq b.$$

PTO

4. For $i = 1, 2$, let P_i stand for the following LP problem:

$$\text{Maximise } c^T x \text{ subject to } x \geq 0 \text{ and } Ax = b_i,$$

where A is an $m \times n$ matrix, $c \in \mathbb{R}^n$, $b_i \in \mathbb{R}^m$ for $i = 1, 2$.

Let $x_i \in \mathbb{R}^n$ be feasible for P_i such that exactly the same coordinates of both x_1 and x_2 are positive.

Show that x_1 is optimal for P_1 if and only if x_2 is optimal for P_2. [10]

5. Use Simplex method to solve the LP problem and its dual: [20]

Maximise $x_1 + x_2 + 2x_3$ subject to $x_1, x_2, x_3 \geq 0$ and

$$\begin{array}{rcrcrcl} & & x_2 & + & 2x_3 & \leq & 3 \\ -x_1 & & & + & 3x_3 & \leq & 2 \\ 2x_1 & + & x_2 & + & x_3 & \leq & 1 \end{array}$$

Indian Statistical Institute
Semester 2 (2015-2016)
M. Stat. 1st Year
Mid-semester Examination
Measure Theoretic Probability

Date and Time: 29.2.16, 2:30 - 4:30 Total Points: $5 \times 6 = 30$

Answers must be justified with clear and precise arguments. If you use any theorem/proposition proved in class state it explicitly. All the functions are assumed measurable and the integrals are wrt the appropriate measure.

1. Suppose F is a distribution function with total mass 1 on $\mathbb{R}$ and μ_F is the probability measure on $\mathcal{B}(\mathbb{R})$ obtained through the Caratheodory procedure with sets in the algebra of the form $(a, b]$. Show that the measure of a set A can be approximated from above by open sets and from below by compact sets. (Notice that F may have jumps.)
2. Let $(X, \mathcal{F}, \mu)$ be a finite measure space and f is a real valued function such that $\int_X f^n d\mu = c, n = 1, 2, 3,$ where c is a positive constant. What can you say about f?
3. Let f_n be a sequence of nonnegative measurable and integrable functions on $(-\infty, \infty)$ such that $f_n \to f$ a.e., and suppose that $\int f_n \to \int f < \infty$. Then show that for each measurable set E we have $\int_E f_n \to \int_E f$.
4. Let f_n be a sequence of functions in $L^p, 1 \leq p < \infty$, which converge a.e. to a function $f \in L^p$. Show that f_n converges to f in L^p iff $||f_n||_p \to ||f||_p$.
5. Let f and g be functions in $L^1(\mathbb{R})$ and define $f \star g$ to be the function $h(y) = \int f(y - x)g(x)dx$. For $f \in L^1(\mathbb{R})$ define $\hat{f}$ by $\hat{f}(s) = \int e^{ist} f(t)dt$. If $f, g \in L^1(\mathbb{R})$ then show that $\widehat{f \star g} = \hat{f}\hat{g}$. (In fact, first you have to show $f \star g \in L^1(\mathbb{R})$.)

Indian Statistical Institute
Semestral Examination Second Semester (2015-2016)
M.Stat. First Year
Large Sample Statistical Methods

Maximum Marks: 60 Date : 26.04.2016 Duration :- $3\frac{1}{2}$ hours

Answer all questions

1. Suppose, for each integer $n \geq 1$, X_n denotes a random variable distributed uniformly on the set of points $\{1/n, 2/n, \ldots, 1\}$. Show that X_n converges in distribution to X where X has a Uniform$[0,1]$ distribution. [4]

2. Suppose we have non-negative random variables X_n which converge almost surely to X, and $E(X) < \infty$. Show that X_n converges to X in L_1. [4]

3. Suppose $X_1, \ldots, X_n$ are iid Bernoulli($\frac{1}{2}$) random variables. Does there exist a sequence of constants a_n such that $a_n(s_n^2 - 1/4)$ converges to a non-degenerate limit, where $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$? Prove your answer and identify the limit if you claim that such a limit exists. [5]

4. Suppose you have two sequences of random variables $\{X_n\}$ and $\{Y_n\}$ (defined on the same probability space) at least one of which is stochastically bounded. Suppose for each $t < s$, $t, s \in \mathcal{R}$, $P(X_n < t, Y_n > s) \to 0$ as $n \to \infty$ and $P(Y_n < t, X_n > s) \to 0$ as $n \to \infty$. Show that $X_n - Y_n$ converges to zero in probability. Explain how this fact is useful in deriving the weak Bahadur representation of sample quantiles. [6+4=10]

5. Suppose $X_1, \ldots, X_n$ are iid having a Uniform$[\theta - 1/2, \theta + 1/2]$ distribution where $\theta \in \mathcal{R}$ is unknown. Suppose $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are the sample order statistics and $\bar{X}_n$ is the sample mean. Define Y_n as follows :

$$Y_n = e^{-\bar{X}_n^2}(X_{(n)} - 1/2) + (1 - e^{-\bar{X}_n^2})(X_{(1)} + 1/2).$$

   Is Y_n a maximum likelihood estimator of θ ? Is Y_n consistent for θ ? Justify your answers. [3+3=6]

6. Let $X_1, \ldots, X_n$ be iid with common density $f(x, \theta)$ given by

$$f(x,\theta) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2}, -\infty < x < \infty,$$

   where $\theta \in \mathbf{R}$ is unknown. Argue that the maximum likelihood estimator exists and is consistent in this case. Can you give any explicit sequence of functions $g_n(.)$ (possibly dependent on n) of $\bar{X}_n$ such that $g_n(\bar{X}_n)$ (after suitable centering and scaling) converges in distribution to $N(0, 1/I(\theta))$, where $I(\theta)$ is the Fisher Information based on one observation? Justify your answer. [3+3=6]

7. Suppose $X_1, \ldots, X_n$ are iid with common density $f(x, \theta)$, where $\theta \in \Theta$, Θ consisting of *only finitely many real numbers*. Assume also that 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 $\theta = \theta_0$. [5]

8. Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$ where $\theta \in \mathbf{R}$. Let T_n denote the Hodges' estimator that estimates θ by 0 if the sample mean lies in $[-n^{-1/4}, n^{-1/4}]$ and by the sample mean otherwise. Derive the asymptotic distribution of Hodges' estimator after appropriate centering and scaling. Can you propose a criterion with respect to which the asymptotic performance of the sample mean is better than that of the Hodges' estimator? Justify your answer. [6+2=8]

9. Suppose you have an iid sample of size n from a multinomial population with k classes. Let $\pi_i (> 0), i = 1, 2, \ldots, k$ denote the probabilities of the classes. Find the asymptotic distribution of $T_n = \sum_{i=1}^{k} \frac{(n_i - n\pi_i)^2}{n_i}$, where $n_i, i = 1, 2, \ldots, k$ denote the number of members in the sample falling in the i-th class. If some of the n_i's are zero, you may redefine the corresponding term in the sum as zero. [6]

10. Stating appropriate assumptions and the null hypothesis, derive the asymptotic null distribution of the Wilcoxon Signed-Rank Test statistics. [6]

INDIAN STATISTICAL INSTITUTE

M. Stat First Year (2015-16)

Second Semester

Resampling Techniques

Date: 29/04/2016 Marks: 100 Duration: 3 hours

Attempt all questions

(1) Let $X_1, \dots, X_n \stackrel{iid}{\sim} F$ and $Y_1, \dots, Y_n \stackrel{iid}{\sim} G$. Let F and G be absolutely continuous with densities f and g respectively.

(i) Develop a bootstrap-based test of the null hypothesis of equality of the medians.

(ii) With the given information can you also devise a permutation test for testing median equality? Justify.

(iii) If zero is the median for both the distributions with $f(0) \neq g(0)$ (both positive), is your bootstrap test asymptotically consistent with the traditional test? Justify.

[7+3+10=20]

(2) Let $X_1, \dots, X_n \stackrel{iid}{\sim} F$ (univariate) with density f. A kernel density estimator of f is

$$\hat{f}_\lambda(x) = \frac{1}{n\lambda} \sum_{i=1}^{n} \kappa\left(\frac{x - X_i}{\lambda}\right),$$

where $\kappa(\cdot)$ is a given kernel function symmetric about zero, $\int \kappa(x)dx = 1$, $\kappa_2 = \int x^2 \kappa(x) dx \neq 0$, and $\lambda > 0$ is the bandwidth that determines the smoothness of the estimated kernel density.

(i) Propose a method based on bootstrap and MISE (mean integrated squared error) to determine λ.

(ii) If $\kappa(\cdot)$ is the standard normal density, then compute the bootstrap-based MISE.

(iii) Show that the bootstrap-based MISE has a bias of the order $n^{-1}\lambda^{-1}$ as an estimator of the actual MISE of $\hat{f}_\lambda$.

[3+7+10=20]

(3) Consider the following degradation model for studying reliability of system components:

$$y_{ij} = z(t_j)'\Theta_i + \epsilon_{ij};\ i = 1, \dots, n;\ j = 1, \dots, m, \tag{1}$$

where y_{ij} is the measurement of the i-th sample component at time t_j; $z(t) = (z_1(t), \ldots, z_q(t))^T$ is a q-vector whose components are known functions of time t; $\Theta_i = (\Theta_{1i}, \ldots, \Theta_{qi})^T$ are unobservable random q-vectors that are *iid* with a q-variate normal distribution $N_q(\theta, \Sigma_\Theta)$, where θ and Σ_Θ are unknown; ϵ_{ij} are *iid* measurement errors with mean zero and variance 1, and also with finite fourth moment; and Θ_i and ϵ_{ij} are mutually independent. Consider the function

$$R(t) = P\left(z_r(t)\Theta_r \in C_r;\ r = 1, \ldots, q\right),$$

for some appropriate sets $C_1, \ldots, C_q$.

(i) Obtain a consistent (as $n \to \infty$) estimator $\hat{R}(t)$ of $R(t)$.

(ii) Propose, with suitable justifications, the methods of estimating the variance of $\hat{R}(t)$. Which method do you recommend and why?

[10+10=20]

(4) Consider the following linear regression model:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i;\ i = 1, \ldots, n,$$

where $\mathbf{x}_i$ is a $k \times 1$ deterministic vector, $\boldsymbol{\beta}$ is the $k \times 1$ vector of parameters, and ϵ_i are uncorrelated errors with mean zero and unknown variance σ_i^2. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ be the design matrix and assume that $\mathbf{X}^T\mathbf{X}$ is non-singular.

(i) Propose a bootstrap estimator of the covariance matrix of the least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

(ii) Is your estimator unbiased? Justify.

[10+10=20]

(5) (i) In question (4), now propose a jackknife estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$.

(ii) Is it unbiased? Justify.

(iii) Consider the following weighted jackknife estimator

$$v_{J,w} = \sum_{i=1}^{n} (1 - w_{ii}) \left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}\right)^T \left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}\right),$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ is the least squares estimator with the i-th observation deleted, and for $i, j = 1, \ldots, n$, $w_{ij} = \mathbf{x}_i \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{x}_j$. Under what condition on w_{ij} is $v_{J,w}$ unbiased?

[5+5+10=20]

Indian Statistical Institute
Semester 2, Academic Year: 2015-16
Semestral Examination
Course: M. Stat 1st Year
Subject: Measure Theoretic Probability

Total Points: $5 \times 14 = 70$
Date: 2.5.16 **Time: 3 Hours**

Answers must be justified with clear and precise arguments. If you refer to a theorem/result proved in class, state it explicitly. More than one answer to a question will not be entertained and only the first uncrossed answer will be graded.

1. (a) Let g be a real valued integrable function on a finite measure space such that for some constant M and all simple h, $|\int ghd\mu| \leq M||h||_p, 1 < p < \infty$. Show that $g \in L^q$, where $\frac{1}{p} + \frac{1}{q} = 1$.

   (b) Suppose E is a subset of measure zero of $\mathbb{R}$. Show that $A = \{(x, y) : x - y \in E\}$ is a Lebesgue measurable subset of $\mathbb{R}^2$ and also that any subset of A is Lebesgue measurable. (Hint: transform the axis by $u = (x-y)/\sqrt{2}, v = (x+y)/\sqrt{2}$.)

2. Suppose X_i are independent random variables such that $P(X_n = 0) = 1 - p_n$, $P(X_n = 1) = p_n$. If $\sum_1^\infty p_n < \infty$ then

   (a) show that $0 < P(S_n = 1 \text{ i. o.}) < 1$.

   (b) Show that $\{S_n = 1 \text{ inf. often }\}$ is not a tail event but a symmetric event.

3. Suppose, for $n \geq 1$, X_n are independent random variables with the probability distribution

$$\begin{aligned} X_n &= 0 \text{ with probability } 1 - \frac{1}{n} \\ &= n \text{ with probability } \frac{1}{n^2} \\ &= \frac{1}{n} \text{ with probability } \frac{1}{n}(1 - \frac{1}{n}) \end{aligned}$$

   (a) Does the sequence $\{X_n\}$ converge with probability 1 or 0?

   (b) Consider the sum $\sum_{k=1}^\infty X_k$. Does this sum converge with probability 1 or 0?

4. (a) Suppose a_n's are positive reals increasing to infinity and b_n's are reals. Show that $\sum_1^\infty (b_n/a_n)$ converges implies $(\sum_1^n b_i)/a_n$ converges to zero. (Your argument must use the tails $t_n = \sum_n^\infty (b_i/a_i)$, which go to zero.)

(b) Suppose Y_n's are independent zero mean random variables with finite variances $1/n$. What is the almost sure limit of

$$\frac{\sum_1^n Y_i}{\log n}?$$

5. Suppose $X_1, X_2, \ldots$ are independent random variables such that $P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n})$ and $P(X_n = 2^k) = 2^{-k}, k = n+1, n+2, \ldots$. Let $S_n = X_1 + \cdots + X_n$. Does $S_n/\sqrt{n}$ converge weakly to $N(0,1)$?

INDIAN STATISTICAL INSTITUTE

SEMESTRAL EXAMINATION
M.STAT I YEAR

OPTIMIZATION TECHNIQUES

Date:04.05.2016 Maximum marks: 60 Duration: 2 hours

The paper contains 67 marks. Answer as much as you can, the maximum you can score is 60.

1. (a) Prove that a graph is bipartite if and only if its incidence matrix is totally unimodular. [8]
   (b) Derive a polynomial time algorithm for matching in a bipartite graph. [4]
   (c) Prove that, in a bipartite graph G, the cardinality of the maximum coclique is equal to the cardinality of the minimum edge cover. [8]
   (d) Prove that the incidence matrix of a directed graph is totally unimodular. [6]

2. Let $a \in \mathbb{Z}_+^n$, and $b \in \mathbb{N}$. Let

$$\mathcal{K}(a,b) \stackrel{def}{=} \text{conv}\left\{x \in \{0,1\}^n : a^T x \leq b\right\}$$

denote the knapsack polytope. A set $C \subseteq [n]$ is a cover if $\sum_{j \in C} a_j > b$.

   (a) Show that the cover inequality $\sum_{j \in C} x_j \leq |C| - 1$ is a valid inequality for $\mathcal{K}(a,b)$. [4]
   (b) Prove that, if C is a cover for $\mathcal{K}$, then the extended cover inequality

$$\sum_{j \in E(C)} x_j \leq |C| - 1$$

   is valid for $\mathcal{K}$, where $E(C) = C \cup \{j : a_j \geq a_i \; \forall i \in C\}$. [4]
   (c) Consider the knapsack polytope $X = \{x \in \{0,1\}^7 : 11x_1 + 6x_2 + 6x_3 + 5x_4 + 5x_5 + 4x_6 + x_7 \leq 19\}$. Find four minimal cover inequalities for X. Find the corresponding extended cover inequalities. [4+4]

3. Consider the Maximum Alternating Sum Subsequence Problem. Given a sequence $S = \{x_1, x_2, \cdots, x_n\}$ of positive integers, find the subsequence $A = \{x_{i_1}, x_{i_2}, \cdots, x_{i_k}\}$ where $i_1 < i_2 < \cdots < i_k$, that maximizes the alternating sum $y = x_{i_1} - x_{i_2} + \cdots + (-1)^{k-1} x_{i_k}$. You are asked to develop an $\mathcal{O}(n)$ dynamic programming algorithm to find the optimal y (NOT the actual subsequence).

   (a) List the table(s) that your algorithm will use and explain the meaning of each entry. [3].
   (b) Specify the recurrence and the base case(s) of your algorithm. Argue the correctness. [6]

(c) Write a pseudocode for your algorithm. [3].

4. Consider a post office that sell stamps in three different denominations, Rs. 1, Rs. 7, and R 10. Design a dynamic programming algorithm that will find the *minimum* number of stam necessary for a postage value of Rs. N as well as the denominations.

   (a) List the table(s) that your algorithm will use and explain the meaning of each entry.

   (b) Specify the recurrence and the base case(s) of your algorithm. Argue the correctness.

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2015–16

M. STAT. I YEAR
Abstract Algebra

Date: 6.5.16 Maximum Marks: 70 Duration: 4 Hours

Attempt FIVE questions from Group A and THREE from Group B.

$\mathbb{Q}$ denotes the field of rational numbers.
$\mathbb{F}_q$ denotes the finite field with q elements.

GROUP A
Answer ANY FIVE questions.

1. Let L be an algebraic extension of a field k. Show that any k-algebra endomorphism of L is an automorphism. Give an example to show that the result would not be true without the hypothesis that L is algebraic over k. [8]

2. Let $f(X)$ be an irreducible polynomial of degree n over k. Let $g(X) \in k[X]$ and $h(X)$ an irreducible factor of $f(g(X))$ in $k[X]$. Prove that the degree of $h(X)$ is divisible by n. [Hint: consider field extensions.] [8]

3. Let L be a finite normal extension of k such that no element of $L \setminus k$ is purely inseparable over k. Prove that L is a separable extension of k. [8]

4. Let $L = \mathbb{F}_4(t)$, where t is transcendental over $\mathbb{F}_4$. Let $u = t^4 + t$, $K = \mathbb{F}_4(u)$ and $f(X) = X^4 + X + u$. Show that f is irreducible in $K[X]$ and splits completely in $L[X]$. Deduce that L is a Galois extension of K of degree 4. [(3+3)+2=8]

5. (i) If α is a root of $X^3 + X + 1 \in \mathbb{F}_2[X]$, then show that α is a primitive 7th root of unity over $\mathbb{F}_2$.

   (ii) If ω is another primitive 7th root of unity over $\mathbb{F}_2$, does it necessarily follow that (a) $\mathbb{F}_2(\omega) = \mathbb{F}_2(\alpha)$? (b) ω is a root of $X^3 + X + 1$? [4+(2+2)=8]

6. Suppose that the Galois group of an irreducible and separable polynomial $f(X)$ is Abelian. Let E be a splitting field of $f(X)$ over k and let $\alpha_1, \ldots, \alpha_n$ be the roots of $f(X)$ in E. Show that $E = k(\alpha_i)$ for each i, $1 \le i \le n$, and $[E : k] = \deg f$. [8]

7. Let $\alpha = \cos 20°$. Examine whether

   (i) $\mathbb{Q}(\alpha)$ is a cyclic extension of $\mathbb{Q}$,

   (ii) α is solvable by radicals over $\mathbb{Q}$,

   (iii) α is a constructible number. [4+1+3=8]

8. Prove the following statements:

   (i) For every normal subgroup N of a group G, the commutator subgroup $[N, N]$ is normal in G.

   (ii) Every group of order 70 is solvable. [3+5=8]

GROUP B

Answer ANY THREE questions.

1. Let G be a finite group of automorphisms of a field L and let k be the fixed field of G. Prove that $L|_k$ is a finite Galois extension whose Galois group is G. [12]

2. Let $f(X) = X^6 + 1 \in \mathbb{Q}[X]$.

   (i) Compute the Galois group of $f(X)$.

   (ii) Describe all subfields of the splitting field of $f(X)$. [5+7=12]

3. (i) Let k be a field and $L = k(\alpha)$, where α is algebraic over k. Show that there are only finitely many intermediary fields between k and L.

   (ii) Let $L = \mathbb{F}_2(x, y)$ (where x and y are algebraically independent over $\mathbb{F}_2$) and $k = \mathbb{F}_2(x^2, y^2)$. Show that $[L : k] = 4$ but there are infinitely many intermediary fields between k and L.

   (iii) Give an example of a field extension K of $\mathbb{F}_2$ such that K is a simple extension of $\mathbb{F}_2$ but K contains infinitely many subfields. [4+6+2=12]

4. (i) Let p be a prime number. Prove that if a regular p-gon is constructible, then p is a Fermat prime. Examine whether the regular heptagon (7-gon) is constructible.

   (ii) Show that the polynomial $X^5 - 4X + 2$ cannot be solved by radicals over $\mathbb{Q}$. Clearly state all the results that you use. [5+7=12]

5. State whether the following statements are TRUE or FALSE with brief justification. Attempt ANY FOUR. Each carries 3 marks.

   (i) If $k(\alpha)$ and $k(\beta)$ are finite normal extensions of the field k, then $k(\alpha + \beta)$ is also a finite normal extension of k.

   (ii) The finite field $\mathbb{F}_{81}$ has a subfield isomorphic to the finite field $\mathbb{F}_{27}$.

   (iii) The group of units of the finite field $\mathbb{F}_{31}$ has an element of order 6.

   (iv) The Galois group of the polynomial $X^3 + X + 1$ is S_3.

   (v) Any finite solvable group G always has an Abelian normal subgroup K.

   (vi) $Q(\sqrt[3]{2})$ is not a subfield of any cyclotomic field over $\mathbb{Q}$. [$4 \times 3 = 12$]

INDIAN STATISTICAL INSTITUTE

Final Semeseter Examination of 2nd Semester : 2015-16

Course Name : **M.Stat. 1st Year**

Subject Name : **Sample Survey and Design of Experiments**

Date : 04/05/ 2016

Duration : 3 hrs

Note: Use separate answer sheets for two groups.

Group – Sample Survey. (Total Marks = 35)

Answer all questions.

1. State and prove Rao (1979)'s theorem on MSE of a general homogeneous linear estimator for the population total Y of a variable of interest y.

(10)

2. From a population of N units a sample of n units was drawn under SRSWR. Of these, only n_1 responded. Out of the remaining $n_2(= n - n_1)$ non-responding units, information was later collected on u units, chosen again under SRSWR. Show that an unbiased estimator of $\bar{Y}$ is

$$\hat{\bar{Y}} = (n_1\bar{y}_{n_1} + n_2\bar{y}_u)/n,$$

where $\bar{y}_{n_1}$ is the mean of y for the n_1 units responding initially and $\bar{y}_u$ is the mean of y for the u units responding later. Also show that

$$V(\hat{\bar{Y}}) = \frac{\sigma^2}{n}\left[1 + \frac{n_2(n_2-1)}{nu}\right].$$

(10)

3. Describe how an unrelated question model can be used to estimate a sensitive population proportion by a general sampling design with $p(s)$ being the selection probability of a sample s of respondents. Obtain its variance and variance estimator.

(10)

4. State the situation when double sampling is needed for effective use of stratified random sampling in estimating population mean $\bar{Y}$ of a variable of interest y. Obtain an unbiased estimator and variance of that.

(5)

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2015-16

Course Name: M.Stat. 1st Year **Subject Name:** Sample Surveys and Design of Experiments

Date: 06.05.16

Total marks: 70

Duration: 3 hours

Use separate answer scripts for the two groups.

Group: Design of Experiments: Total marks: 35 **Answer all Questions.**

1. (a) Define a Balanced Incomplete Block (BIB) design. Derive the conditions which must be satisfied by the parameters of a BIB design.

   (b) Does a BIB design exist with parameters: $v = b = 22, r = k = 7, \lambda = 2$? Justify your answer.

   (c) Let N be the incidence matrix of a BIB design. Consider a new design with incidence matrix N', where $'$ denotes the transpose. Verify if the new design will also be a BIB design.

   (d) Construct a BIB design with 11 treatments in 11 blocks, each block of size 5, starting with on the elements of GF(11). [You need to show only first 3 blocks, all blocks need not be shown. You may 2 as a primtive element of GF(11).]

   (e) From the BIB design constructed in (d) above, how will you construct a Hadamard matrix of order 12. Clearly state the result you use for this construction.

   $[(2+2)+2+2+4+(1+1)=14]$

2. (a) What are the advantages of using a factorial experiment? Give your answer with the help of an example of an experimental situation where you would use a factorial experiment.

   (b) A factorial experiment is to be designed with 4 factors, A, B, C, D, each at 2 levels. All main effects and all interactions are to be estimated. Furthermore, the main effects and 2-factor interactions have to be estimated with more precision than the other interactions. **Indicate** how you can construct a design for this in 4 replicates, each replicate consisting of 2 blocks of size 8 each. Show **any one** replicate in full. Give the loss of information for each effect.

   (c) Give a suitable confounding scheme for the above factorial experiment if each of the 4 replicates consisted of 4 blocks of size 4 each. [Design construction not needed]

   (d) What is meant by a resolution(2,3) plan in the context of fractional factorial experiments?

   (e) Obtain a 12-run resolution (1,1) plan for an experiment with 11 factors, each at 2 levels.

   [3+4+2+2+3=14]

3. (a) A shop has 9 salespersons and 3 departments. It needs to prepare a duty roster for assigning the salepersons to its departments, 3 persons to each department each day, for a total of 4 days. The assignment should be such that no two salespersons are assigned together to the same department more than once and each person must be assigned a duty every day. Give a possible assignment for such a roster.

   (b) Construct a Youden square design with 7 treatments arranged in 3 rows and 7 columns.

   [4+3=7]

Indian Statistical Institute
Backpaper
M.Stat. First Year
Second Semester, 2015-16 Academic Year
Large Sample Statistical Methods

Total Marks: 100 Duration :- 4 hours

Answer all questions

1. 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]

2. Show that a sequence of random variables X_n converges in probability to zero if and only if $E(|X_n|/(1+|X_n|)) \to 0$ as $n \to \infty$. [10]

3. Suppose $X_1, \ldots, X_n$ are iid with a $N(\mu, \sigma^2)$ distribution with both parameters being unknown. Find the asymptotic distribution of sample coefficient of variation. [15]

4. State and prove the Weak Bahadur Representation of sample quantiles under iid sampling from a common distribution. [25]

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

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

7. Suppose $X_1, \ldots, X_n$ are iid having density $f(x, \theta)$, where $\theta \in \mathbf{R}$. Invoking appropriate regularity assumptions, derive the asymptotic null distribution of the likelihood ratio test statistic for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. [10]