MEASURE THEORETIC PROBABILITY
M. STAT. IST YEAR SEMESTER 2
INDIAN STATISTICAL INSTITUTE

Backpaper Examination
Time: 3 Hours
Total Marks: 100
Date: 25th August 2018

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

1. State with appropriate justification whether the following statements are true or false:
   (a) Let $\{X_k\}$ be a sequence of independent random variables with $\mathrm{P}[X_k = -k^2] = k^{-2}$, $\mathrm{P}[X_k = -k^3] = k^{-3}$ and $\mathrm{P}[X_k = 2] = 1 - k^{-2} - k^{-3}$. Then $\sum X_k = \infty$ a.e. [8]
   (b) The function $\phi(t) = |\cos t|$ is a characteristic function. [10]
2. On a measure space $(\Omega, \mathcal{A}, \mu)$, consider a sequence real valued measurable functions f_n which converge pointwise to f, another measurable function. Assume that, for each n, there exists a nonnegative measurable function h_n such that $|f_n| \le h_n$ and h_n converges pointwise to a measurable function h with $\int h_n d\mu \to \int h d\mu < \infty$. Show that $\int f_n d\mu \to \int f d\mu$. [12]
3. Define random variables $\{X_t : 0 \le t \le 1\}$ defined on the sample space $[0, 1]$ endowed with Borel σ-field as follows: $X_t(\omega) = \mathbb{1}_{\{t\}}(\omega)$.Show that each X_t is a random variable. Describe the σ-field generated by the entire collection of the random variables $\{X_t : 0 \le t \le 1\}$. [14]
4. Show that
$$\int_{-\infty}^{\infty} \frac{1}{(1+y^2)^2} dy = \frac{\pi}{2}.$$
[12]
5. Let $\{X_n\}$ be an i.i.d. sequence of Unif(0, 1) random variables. Define $M_n = \min(X_1, \dots, X_n)$. Show that nM_n converges weakly to the exponential distribution with mean 1. [8]
6. Prove that
$$\lim_{n\to\infty} \frac{1}{(n-1)!} \int_0^n e^{-x} x^{n-1} dx = \frac{1}{2}.$$
[8]
7. Let $\{X_n\}$ and $\{Y_n\}$ be two sequences such that $\sum \mathrm{P}[X_n \ne Y_n] < \infty$. Then if Y_n converges weakly, so does X_n to the same limit.
   Consider an i.i.d. sequence of random variables with common density $f(x) = |x|^{-3}$ for $|x| > 1$ and 0 otherwise. Define $Y_n = X_n \mathbb{1}_{[|X_n| \le \sqrt{n}]}$ and show that $\sum_1^n Y_i/\sqrt{n \log n}$ converges weakly to standard normal distribution. Hence show that $\sum_1^n X_i/\sqrt{n \log n}$ also converges weakly to standard normal distribution. [8+8=16]
8. Let $\{X_n\}$ be a sequence of nonnegative independent random variables. Show that $\sum X_n < \infty$ a.e. iff for any $c > 0$, both the sums $\sum \mathrm{P}[X_n > c]$ and $\sum \mathrm{E}\left[X_n \mathbb{1}_{[X_n \le c]}\right]$ converge. [12]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination 2018-19

M. Stat. 1st Year
Statistical Inference I

3rd September, 2018

Maximum Marks: 60 | Time: 3 hours

[Note: Notations are as used in the class. Answer as much as you can. Best of Luck!]

1. Under the decision theoretic framework (to be specified by you), define the randomized and behavioral rules and explain their difference through a suitable example. [3+3+4]=10

2. Suppose $X_1, \ldots, X_n$ are independent and identically distributed observations from $N(\mu, \sigma^2)$, where μ is a known real number and consider the loss function $L(\sigma^2, a) = \left(\frac{a}{\sigma^2} - 1\right)^2$ for the purpose of estimating σ^2.

   (a) Find the Bayes estimator of σ^2 with respect to the conjugate prior.

   (b) Find a Minimax estimator of σ^2. [5+5]=10

3. Prove or disprove the following for a standard decision problem as defined in the class.

   (a) Sample mean is admissible for estimation of normal mean, under the absolute error loss, based on n independent identically distributed observations with known variance.

   (b) When the parameter space Θ contains only two elements, the intersection of any two complete risk sets is essentially complete.

   (c) The upper and lower values of a statistical game (or decision problem) are equal for countable parameter spaces.

   (d) When the parameter space is finite, the Bayes rule with respect to a given prior is admissible if the prior has support on the full parameter space.

   (e) When the parameter space is finite and the risk set is closed and bounded from below, a Bayes rule with respect to any proper prior exists. [5×5]=25

4. For a decision problem, assume that the parameter space Θ is finite. Let S and S_0 denote the risk sets of all randomized and non-randomized decision rules, respectively.

(a) Show that S is convex but S_0 may not be convex.

(b) Prove that S is the convex hull of S_0.

[(2+3)+5]=10

5. Suppose we have p independent random samples, each of size n, from $N(\theta_j, \sigma^2)$, $j = 1, \dots, p$, respectively. Assume σ^2 is known and we want to estimate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$.

(a) Derive the MLE, say $\boldsymbol{T}$, of $\boldsymbol{\theta}$.

(b) Define the James-Stein estimator of $\boldsymbol{\theta}$ and prove that it is better than $\boldsymbol{T}$ with respect to the squared error loss (squared L_2-norm).

[3+(1+6)]=10

6. Write down two sets of sufficient conditions for the class of non-randomized decision rules to be essentially complete for the statistical decision problem $(\Theta, \mathcal{D}, \widehat{R})$. Prove your claim.

[(2+2)+(3+3)]=10

Mid - Semestral Examination

M.Stat. - I

(2018 - 2019)

Categorical Data Analysis

Full Marks -70

Time : 2 hours 30 mins.

(1) Define odds ratio and relative risk for 2X2 contingency table with suitable examples. The odds ratio can be used as a measure of association between disease and exposure, irrespective of the study design, either it is follow up or cross-sectional - Justify. Establish the relationship between odds ratio and relative risk. Odds ratio is a good approximation for relative risk for a rare disease - discuss. Find the asymptotic standard error of the log relative risk.

[5 + 3 + 3 + 4 + 3 + 4 = 22]

(2) The yardstick of logistic regression is to convert the combined risk to an individual risk through a sigmoid function - discuss the issue. Discuss elaborately the origin of the form of the logistic function in the context of population ecology. Let $x_i = (x_{i1},\ldots,x_{ip})$ denote setting i of values of p explanatory variables, $i=1\ (1)\ N$. When more than one observation occurs at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of successes. We then let y_i refer to this success count rather than to an individual binary response. Then $Y_1,Y_2,\ldots,Y_N$ are independent binomials with $E(Y_i) = n_i\pi(x_i)$. Using logistic regression set up for the probability $\pi(x_i)$ construct the likelihood equation and show that the Newton-Raphson and Fisher's scoring iterative method for solving likelihood equations are equivalent. Suggest the confidence interval for the true logit function. Proposed an amended estimator for [redacted] logit and also discuss its utility. Interpret the intercept parameter of this logistic regression.

[3 + 5 + 5 + 3 + 3 + 3 = 22]

(3) a) Suppose that cell counts $(n_1,n_2,\ldots,n_N)$ have a multinomial distribution with cell probabilities $\pi = (\pi_1,\pi_2,\ldots,\pi_N)^T$. Let $n = n_1+..+n_N$, and let $p = (p_1,p_2,\ldots,p_N)$ denote the sample proportions, where $p_i = n_i/n$. Denote observations i of the n cross-classified in the contingency table by $Y_i = (Y_{i1},Y_{i2},\ldots,Y_{iN})$, where $Y_{ij} = 1$ if it falls in cell j, and $Y_{ij} = 0$ otherwise, $i = 1(1)N$. Let $g(t_1,t_2,\ldots,t_N)$ be a differentiable function, and let $\phi_i = \partial g/\partial \pi_i$, $i = 1(1)\ N$, denote $\partial g/\partial t_i$ evaluated at $t = \pi$. Then show that $\sqrt{n}\,[g(p) - g(\pi)]$ follows normal distribution with mean 0 and variance $\sum \pi_i \phi_i^2 - (\sum \pi_i \phi_i)^2$ asymptotically.

b) A sample of 156 dairy calves born in a county, Florida, USA, were classified according to whether they caught pneumonia within 60 days of birth. Calves that got a pneumonia infection were also classified according to whether they got a secondary or tertiary infection within 2 and 5 weeks respectively after the first infection cleared up. Let a denote the number of calves that got a primary, secondary and tertiary infection, b the number that received a primary and secondary but not tertiary infection, c denote the number that received a primary but not a secondary infection, and d denote the number that did not receive a primary infection. Let π be the probability of a primary infection. Construct the testing procedure for the hypothesis that the probability of infection at time t, given infection at times $1, \ldots, t-1$, is also π, for $t = 2,3$ and show that $\hat{\pi} = (3a+2b+c)/(3a+3b+2c+d)$.

c) Construct Fisher's exact test of independence for 2×2 contingency table. Discuss the methods of ordinal associations with suitable examples.

[6 + 10 + 5 + 5 = 26]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : Semester I (2018-19)

M. Stat. I Year

Stochastic Processes

Date: 05.09.2018 Maximum marks: 50 Time: 135 minutes

The total mark is 54

1. (a) Is a process with independent increments also a martingale?
   (b) Consider a birth process $X(t)$ with intensity $\lambda(t)$. Identify the following two cases by specific names and the corresponding rate parameters: (a) $\lambda(t) = \lambda X(t)$ and (b) $\lambda(t) = \lambda E[Y(t)]$, for $\lambda > 0$ and another independent non-homogeneous Poisson process $Y(t)$ with intensity $\mu(t)$.
   (c) Prove that the equilibrium probabilities do not exist for a linear birth and death process.
   (d) Give an example of a bivariate Markov process explaining the notation and stating the assumptions clearly.

   [2+(1+2)+2+3=10]

2. (a) Describe a linear birth and death process with unequal rate parameters starting with a single individual and then derive the expression for the corresponding probability generating function (PGF) giving full details.
   (b) Consider a linear birth and death process starting with a single individual under observation during the time interval $[0, \tau]$. Describe the data and then obtain the maximum likelihood estimates (MLEs) of the two rate parameters.

   [12+8=20]

3. Consider a general birth and death process with $\lambda_0 = 0$ so that 0 is an absorbing state and also the rate parameters are such that this absorption from any state is certain. Let w_i be the mean absorption time starting from state i $(i = 0, 1, \cdots)$ with $w_0 = 0$. Write down a recursive relation for the w_i's. Give explanation. [6]

4. Consider the counter model as discussed in the class with the output at time t given by

$$Y(t) = \sum_{n=0}^{N(t)} X_i e^{-\alpha(t-s_i)},$$

   where $N(t)$ counts the number of arrivals by time t, following HPP(λ), X_i is the random amplitude of the ith arriving signal at time s_i with the X_i's being independent and identically distributed with mean μ. Assume also the X_i's to be independent of the signal arriving process given by $N(t)$. Obtain the PGF of $Y(t)$ and hence an expression for its expectation. [8+2=10]

5. Suppose that the growth of normal bacteria is described by a deterministic curve $X(t)$ with $X(0) = n_0$. A mutant type appears from this pool of normal bacteria with mutation rate μ per bacterium. Once a mutant type appears, it grows according to a linear birth and death process with rates α and β, respectively, per bacterium. Write down the differential equation for the PGF of the number of mutant bacteria at time t. Hence, derive an expression for the expected number of mutant bacteria at time t. [6+2=8]

Indian Statistical Institute

First Semestral Midterm Examination 2018-19

M. Stat. I yr

Regression Techniques

Date: September, 06, 2018 (14:30 hrs) Maximum marks: 60 Duration: 2.5 hrs.

Answer all Questions. ***The paper carries*** 70 ***points***

1. Let X, Y be two random variables satisfying $\mathrm{E}(X) = \mathrm{E}(\mathrm{Y}) = 0$, $\mathrm{Var}(X) = \mathrm{Var}(\mathrm{Y}) = 1$ respectively. Find out the minimum value and a solution (a^*, b^*) for the optimization problem

$$\min \mathrm{Var}(\mathrm{aX} + \mathrm{bY}) \quad \text{subject to } |\mathrm{a}| + |\mathrm{b}| = 1.$$

(b) How does the answer to the above optimization problem change if we assume $\mathrm{Var}(X) = 1, \mathrm{Var}(\mathrm{Y}) = \sigma^2$, where $\sigma^2 > 0$ is given but may not be 1.

[12+8 =20]

2 (a) Let $y_i = a + bx_i + \epsilon_i$, $1 \le i \le n$, be a standard simple linear regression moodel. For simplicity we assume $\bar{x}_n = \bar{y}_n = 0$ for this particular data set. Let $\hat{b}_n$ denote the OLS estimated slope for this data.

Suppose a new observation (y, x) is obtained which is assumed to follow the same regression structure described above. Further let $\hat{b}_n(x)$ denote the new OLS slope after adding the data to the previous dataset. Compute $(d/dx)\,\hat{b}_n(x)$ in closed form.

(b) Find out equation of the limiting regression line after adding the new point (y, x) as $x \longrightarrow \infty$. (note that ϵ_{n+1}, the unobserved error has no dependence on x in this case)

[15+ 10=25]

3 (a) Define the DFFITS and Studentized residual statistics for a standard multiple linear regression model. Can we express DFFITS values in terms of Studentized residuals? If yes, write down the formula.

(b) Consider a simple linear regression model $y_i = a + bx_i + \epsilon_i$, $1 \le i \le n$ where the unobserved errors ϵ_i's are known to satisfy $\epsilon_i = \rho\epsilon_{i-1} + u_i$, $1 \le i \le n$ for some known $0 \le \rho \le 1$. Assume $\epsilon_0 = 0$ and u_i's are iid $\mathrm{N}(0, \sigma^2)$ for some unknown $\sigma^2 > 0$. Describe how would you compute OLS estimate of the slope b. Also derive the distribution of the SSE for your procedure.

[10+ 15=25]

Indian Statistical Institute
Mid-semestral Examination

September 7, 2018

Multivariate Analysis, M1

Total points: 20 Time: 2 hours

1. Principal components analysis [4 + 1]

(a) Let U_1, U_2 be i.i.d. uniform on $(0,1)$. Set $X_1 = U_1, X_2 = U_2, X_3 = U_1 - U_2, X_4 = U_1 + U_2$. Compute the covariance matrix Σ of $X = (X_1, X_2, X_3, X_4)^\top$. How many principal components are relevant? What are these principal components? What is special about them?

(b) Given data $X_{n\times p}$, statistician A first standardizes the data and then does PCA. Statistician B, however, applies the Mahalanobis transform first, and then does PCA. Whose dimension reduction procedure makes sense and why?

2. Multidimensional scaling [5]

Let $C = ((c_{rs}))_{n\times n}$ be a symmetric PSD matrix. Define a matrix D by

$$d_{rs} := (c_{rr} - 2c_{rs} + c_{ss})^{1/2}.$$

Prove that D is an Euclidean distance matrix with the corresponding inner product matrix $B = HCH$, where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is the centering matrix.

3. Multivariate hypothesis testing [3 + 2]

Assume that at time t the locations of the asteroids between Mars and Jupiter are distributed identically and independently as $N_3(\mu_t, \Sigma_t)$ (imagine that the Sun is at the origin of our coordinate system). Distances are measured in AU or Astronomical Unit, i.e. the average distance between the Sun and the Earth. Assume that Σ_t is known to be $0.25I$. Given a measurement $x_t = (2.2, 1.5, 0.76)$ of the location of the asteroid Ceres at time t, derive a test of the hypothesis

$$H_0 : \|\mu_t\| = 2.85.$$

In particular, (a) write down a formula for your test statistic in terms of x_t, and (b) write down an expression for the p-value in terms of the CDF of a distribution that you know.

4. Factor analysis [2 + 2 + 1]

In the context of the factor model, consider the equation

$$S_{p\times p} = \hat{\Lambda}_{p\times k}\hat{\Lambda}^\top + \hat{\Psi},$$

where you have observed that all the sample covariances are non-negative. Show that (a) if $p = 3, k = 1$, then there are exatly two solutions for $(\hat{\Lambda}, \hat{\Psi})$, and, (b) if $p = 2, k = 1$, then there are infinitely many solutions. When are these solutions not statistically meaningful?

INDIAN STATISTICAL INSTITUTE
Semester Examination 2018-19

M. Stat. 1st Year
Statistical Inference I

November 12, 2018

Maximum Marks: 50 | Time: 3 hours

Note: Notations are as used in the class. Best of Luck!

Part A: Answer any TWO among three. Total Marks $5 \times 2 = 10$.

1. Consider a time series sample $\{X_t : t = 1, \ldots, n\}$, where each X_t independently follows a normal distribution with mean ζ_t and variance 1. Derive a set of James-Stein type minimax estimators of $\boldsymbol{\theta} = (\zeta_1, \ldots, \zeta_n)$.

2. Write down, with all the required notations and assumptions, Karlin's theorem on admissibility for the exponential family of distributions. Hence, or otherwise, find all admissible estimators of the mean of a Gamma distribution.

3. For a probability distribution P over a finite sample space $\chi = \{1, \ldots, n\}$, the Renyi entropy is defined as

$$H_\alpha^R(P) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^{n} p_i^\alpha \right), \quad \alpha > 0,$$

where $\boldsymbol{p} = (p_1, \ldots, p_n)$ is the probability mass function of P with respect to the counting measure. Prove that there exists a decision theoretic game which leads to the above Renyi entropy and derive the optimal value (saddle point) of the game.

Part B: Answer as much as you can. Maximum you can score in Part B is 40.

4. Suppose $X \sim Bin(n, \theta)$ for some $\theta \in [0, 1]$ and consider the squared error loss function.

(a) Find the Bayes estimator of θ with respect to the conjugate prior.

(b) Find a Minimax estimator of θ. [3+3]=6

P.T.O

5. Prove or disprove the following for a standard decision problem as defined in the class.

(a) When the parameter space is finite, the Bayes rule with respect to any proper prior is admissible.

(b) For any given convex set $S \subset \mathbf{R}^k$ for a positive integer k, there is a decision problem for which the risk set is S.

(c) If S is a non-empty convex set which is closed from below, then it has at least one lower boundary point.

(d) When the parameter space is finite and the risk set is closed and bounded both from below only, the class of decision rules corresponding to the lower boundary points of the risk set forms a complete class but is not minimal complete.

[5×4]=20

6. Find a minimum risk equivariant estimator of θ in the following cases:

(a) X_1, X_2 are two independent observations from a distribution symmetric about θ and the location-invariant loss function is convex and even at $\theta = 0$.

(b) $X_1, \ldots, X_n$ are n (> 2) independent and identically distributed observations from a Gamma distribution with location parameter μ and scale parameter σ, where $\theta = (\mu, \sigma)$. Take an appropriate invariant loss function.

[4+6]=10

7. Suppose $X \sim p_\theta$, the Cauchy density with location parameter $\theta \in \mathbb{R}$.

(a) Verify if this family has the monotone likelihood ratio property.

(b) Hence, or otherwise, derive a most powerful level-α test for the hypothesis $H_0 : \theta = 0$ against $H_1 : \theta = 1$.

[3+5]=8

8. Suppose $X_1, \ldots, X_n$ are n (> 2) independent and identically distributed observations from $N(\mu, \sigma^2)$, where both parameters are unknown. Derive the UMPU level-α test for the hypothesis $H_0 : \mu \leq 0$ against $H_1 : \mu > 0$. State the result clearly, with all notations and assumptions, that you need to use.

[5+2]=7

Indian Statistical Institute

First Semestral Examination 2018-19

M. Stat. I yr (B-Stream)

Regression Techniques

Date: November 15, 2018 (14:30 hrs) Maximum marks: 100 Duration: 3 hrs.

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

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

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

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

Suppose $\hat{\beta}$ minimizes $||Y - X\beta||^2$ and define $\hat{Y} = X\hat{\beta}$. Compute $E||Y - \hat{Y}||^2$ under the full model.

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

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

Show that $\text{SSE}_1 \geq \text{SSE}$ (of the full model where both X and Z are used).

[15+10 =25]

2 (a) Describe the notion of False Discovery rate (FDR) in a multiple hypothesis testing problem with examples.

(b) Given a standard multiple linear regression model with p explanatory variables. Consider a sequence of hypotheses $\text{H}_i : \beta_i = 0$ for $1 \leq i \leq p$. Describe how you would apply the Benjamini-Hochberg algorithm for subset selection which controls the FDR. Workflow should be precisely described.

(c) Consider a standard simple linear regression model $y_i = a + b\,x_i + \epsilon_i$, $1 \leq i \leq 3$, with $x_1 = -1, x_2 = 0, x_3 = 1$ respectively. Define a cross validation formula for β as

$$CV(\beta) = \sum_{i=1}^{3} \left(y_i - \beta\,x_i - \bar{y}_{(-i)}\right)^2,$$

where $\bar{y}_{(-i)} = (\sum_{j \neq i} y_j)/2$ for $i = 1, 2, 3$ respectively. Simplify $CV(\beta)$ in standard quadratic function form $(c_0 + c_1\,\beta + c_2\,\beta^2)$ for the given model. Is it possible to use $CV(\beta)$ to develop a test statistic for the hypothesis $\beta = 0$? If yes, describe the method in brief.

[5+10+15=30]

PTO

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

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

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

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

[10+ 15=25]

4 (a) Describe the notion of θ-th regression quantiles in a standard multiple linear regression problem for $0 < \theta < 1$ through appropriate minimization problem. Show that solution to the minimization problem indeed produces θ-th quantile of the error distribution if there is only intercept in the model.

(b) Show that the minimization problem can be converted into solving a linear programming problem. Write down the LP problem in its primal form when $\theta = 1/2$.

[10+ 10=20]

5. Data Assignment

[10]

Semestral Examination

M.Stat. - I

(2018 - 2019)

Categorical Data Analysis

DATE : 17.11.2018 Full Marks -100

Time : 3 hours 30 mins.

(Answer as many as you can. The maximum you can score is 100)

1. The following Table contains results of a study by Mendenhall *et al.* (1984) to compare the radiation therapy with surgery in treating cancer of the larynx. The response indicates whether the cancer was controlled for at least two years following treatments.

   (a) Construct an appropriate testing procedure for the hypothesis $H_0 : \theta = 1$ against $H_1 : \theta > 1$, where θ denotes the population odds ratio which can be properly defined from the above retrospective study.

   (b) Explain how you will evaluate the P – value.

   (c) Discuss how the testing procedure can be extended for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, where θ_0 is a specified value of θ.

   (d) Comment on the approximate confidence interval of the odds ratio for the extended case.

   (e) Show that the conditional ML estimate of θ differs from the unconditional ML estimate.

Case	Cancer Controlled	Cancer not Controlled
Surgery	21	4
Radiation Therapy	17	8

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

2. Genotypes AA, Aa, and aa occur with probabilities $[\theta^2, 2\theta(1-\theta), (1-\theta^2)]$. A multinomial sample of size n has frequencies (n_1, n_2, n_3) of these three genotypes.

   (a) Form the log-likelihood and show that $\hat{\theta} = (2n_1 + n_2)/2(n_1 + n_2 + n_3)$.

   (b) Find the asymptotic standard error of θ.

   (c) Explain how to test whether the probabilities truly have this pattern or not.

[3 + 3 + 3 = 9]

3. (a) Derive the asymptotic distribution of the sample log odds ratio.

   (b) Consider the following linear logit model

$$logit(\pi_i) = \alpha + \beta x_i,$$

where y_i is the number of outcomes in the first column (successes) out of n_i trials in row i of an $I \times 2$ table. Let scores $\{x_1, x_2, \ldots, x_I\}$ describe distances between categories of X. When one expect a monotone

P.T.O.

effect of X on Y the above model is suitable. Derive the likelihood equation and suggest a suitable test for $\beta = 0$.

(c) Define the Probit and Log-Log link models.

(d) Evaluate the magnitude of the maximum value of the rate of change of response curve in Probit analysis.

(e) Establish the relationship between Log-Log link and the Cumulative Distribution Function (CDF) of the extreme value distribution.

[4 + 5 + 4 + 3 + 4 = 20]

4. (a) Derive the likelihood equation for a generalized linear model for which random component specifies N independent observations $Y_1, Y_2, \ldots, Y_N$ from the exponential dispersion family.

(b) Suppose that $n_i Y_i$ has a binomial(n_i, π_i) distribution, and Y_i is the sample proportion, rather than number of successes of n_i trials, $i = 1 \ldots N$. Show that this distribution can be expressed in the form of the exponential dispersion family.

(c) Derive the mean, variance and the logit link of the above binomial distribution using the inter-relationship of the exponential dispersion family parameters as expressed in question number (b).

[6 + 3 + 6 = 15]

5. (a) Define loglinear model of independence for $I \times J$ contingency table that cross classifies a multinomial sample of n subjects on two categorical responses.

(b) Using the concept of logit link, interpret the parameters of this model, when $J = 2$.

(c) How this model can be extended to a more complex saturated loglinear models?

(d) Derive the direct relationship between log odds ratio and the association term present in the saturated model.

[3 + 3 + 3 + 3 = 12]

6. Suppose we have 8, 2×2 contingency tables as an output of a specific clinical trial experiment at 8 centres.

(a) Suggest a suitable test under the hypothesis that the conditional odds ratio obtained from the data of 8 centres are equal.

(b) Suggest the Mantel – Haenszel estimate of the common odds ratio under the null.

(c) Propose a non - model based test of conditional independence in $2 \times 2 \times 8$ tables.

(d) Also suggest an alternative test of conditional independence of such tables using the logit model.

[4 + 3 + 4 + 4 = 15]

7. (a) A nasal drainage study involved 30 patients who had colds of recent onset. Two 20-minute steam inhalation treatments, spaced 60-90 minutes apart, were administered at the time of enrolment for each patient. Assessment of subjective response was made on an individual daily score card by the patient from day 1 to-day 4. On each day, the severity of nasal drainage was calibrated into 4 ordered categories with polytomus response such as no symptoms, mild, moderate, and severe. One was interested in examining whether the severity improved following the treatment, and in testing whether the observations over time, for each subject, were likely to be correlated. Derive the suitable test and critical region for this case.

(b) Define concordance and discordance pairs with suitable example in the context of measures of ordinal association.

[10 + 4 = 14]

8. Let us assume, $(n_1, \ldots, n_N)'$ has multinomial distribution with probabilitites $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)'$. Here $n_i, i = 1 \ldots N$ are counts in N cells of a contingency table. The model is $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, where $\boldsymbol{\pi}(\boldsymbol{\theta})$, denotes a function that relates π to a smaller number of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$. Derive the asymptotic distribution of the model based ML estimator $\hat{\pi} = \pi(\hat{\theta})$ of π.

[11]

INDIAN STATISTICAL INSTITUTE

Semestral Examination : Semester I (2018-19)

M. STAT. I Year

Stochastic Processes

Date: 20.11.2018 Maximum marks: 100 Time: $3\frac{1}{2}$ hours

Answer as many as you can. Total mark is 108.

1. Answer the following questions briefly. You may assume the results done in the class by specifically mentioning them.

   (a) Write the $M/M/s$ queue as a birth and death process by specifying the birth and death rates. What is the distribution of Idle Period for this queue?

   (b) What is the specific name of the self-exciting point process when the intensity at time t depends only on the time since the last event occurrence? For a counter model with fixed lock period L and NHPP arrival of signals, what is the intensity of the arrival of the registered signals?

   (c) For a homogeneous Poisson process $X(t)$, derive an expression for $Cov[X(s), X(t)]$ for $s < t$. What is this expression for an NHPP?

   (d) For a delayed renewal process with the first renewal time having density $\frac{x}{50}$, $0 \leq x \leq 10$, and the subsequent renewal times having density $\frac{1}{10}$, $0 \leq x \leq 10$, what is the expected number of renewals $M_D(t)$, by time t (≤ 10)?

   (e) For a discrete time branching process $\{X_n\}$ with $X_0 = i(> 1)$ and m being the mean of the offspring distribution, find an expression for $E[X_n]$ in terms of m.

[(2+3)+(2+3)+(2+1)+3+3=19]

2. Consider a system with M machines, R repairmen ($R < M$) and S spare machines. The machines operate independently, each with *exponential*(λ) failure time. The repairmen act independently, each taking *exponential*(μ) repair time. The repair times and the failure times of the machines are independent. After failure, a machine goes for repair and the repair work starts immediately if a repairman is free, otherwise waits for repair. A spare, if available, replaces the failed machine immediately. After completion of repair, a machine joins operation immediately, if needed, otherwise becomes a spare. Let $X(t)$ denote the number of non-operative machines, under or awaiting repair, at time t. Write this as a birth and death process by specifying the birth and death rates. [6]

3. (a) Explain why, in the context of a renewal process, the total life time is not the same as the renewal time.

   (b) Find the limiting distribution of the total life at time t for a renewal process with renewal distribution F.

   (c) In the context of a renewal process, what is the mean of the limiting distribution of $S_{N(t)+1} - S_{N(t)}$ in terms of the mean and variance of the renewal distribution?

[3+10+4=17]

4. Let $\{N(A) : A \subset \mathcal{R}\}$ be a Poisson process with intensity $\lambda(x) = \Lambda/2$, for $-1 \leq x \leq 1$, and 0, otherwise, where Λ is a constant. A second point process $\{M(B) : B \subset \mathcal{R}\}$ is created by randomly and independently translating each point x of N to a point y of M as $y = x + \epsilon$, where ϵ is a standard normal variate. Obtain the distribution for the number of points of M in a set B. Out of n input points, what is the probability that there are m points of M in the set B? [6+4=10]

5. (a) Describe a self exciting point process by writing down the general form of the intensity process clearly. Derive explicit expression for $P[N(s,t) = 0|N(s) = 0]$, where $N(s,t)$ is the number of point events in $(s,t]$ and $N(s) = N(0,s)$.

   (b) Consider a computer system during peak load when a large unlimited number of jobs are waiting to be processed. Assume that the required processing time of a job is exponentially distributed with mean $1/\mu$. The processing times for different jobs are independent. Further assume that an adjustment time per job is required (before the job starts) by the system to perform overhead functions to be ready for the processing. The adjustment times for different jobs are independent (and also independent of the respective job processing time) and are exponentially distributed with mean $1/\lambda$. At time 0, an adjustment time starts to take up the first job-processing. Consider the point process defined by the point events of successive jobs being over. Identify this process by a renewal process by indicating the renewal distribution and derive an explicit expression for $P[N(s,t) = 0|N(s) = 0]$. For fixed and unknown adjustment time τ per job and based on an observed path until the time of completion of the nth job, find maximum likelihood estimates of μ and τ.

[(2+5)+(2+5+6)=19]

6. (a) Suppose that the primary photoelectrons are generated as a Poisson process with intensity λ. By a cascade process, each primary electron, at its generation time, triggers generation of secondary electrons according to another Poisson process with intensity γ. Let $Y(t)$ denote the total number of secondary electrons produced at time t. Identify the process $Y(t)$ and then give an expression for its characteristic function. Hence, obtain $E[Y(t)]$.

   (b) A rare mutation takes place in one normal cell with probability $\nu(t)\Delta t + o(\Delta t)$ in the time interval $(t, t+\Delta t)$ during division (that is, during mutation a normal cell divides into one normal cell and one mutant cell). Probability of two or more mutations in $(t, t+\Delta t)$ is $o(\Delta t)$. Normal cells do not divide otherwise and they act independently. Start with a body of S normal cells at time 0. Each mutant cell born in this way grows according to a linear birth and death process with rates λ and μ, respectively, per cell, thereby forming a mutant clone. Let $Y(t)$ denote the number of non-extinct mutant clones at time t. Derive the distribution of $Y(t)$.

[(3+4+3)+10=20]

7. (a) Give an argument for the Little's formula $E(W) = E(Q)/\mu$, where μ is the mean service time for a customer, W is the waiting time for an arriving customer before entering into service and Q is the system size in equilibrium.

   (b) Consider an $M/M/1$ queue with arrival rate 10/hour and the service time with mean 4 minutes. Answer the following questions assuming the system to be in equilibrium.

      i. What is the probability of having a queue?

      ii. What is the average queue length?

      iii. What is the average time a customer spends in the system?

   (c) In a tandem queue system with K servers in series, Poisson arrivals and exponential service times with different means, give an expression for the expected system size in equilibrium.

[3+(2+3+5)+4=17]

Indian Statistical Institute
Semestral Examination

November 22, 2018

Multivariate Analysis, M1

Total points: 30 | Time: 3 hours

Note: This is a **closed notes/closed book** examination. Do any **six** of the following problems. Notations, if not explicitly explained, are to be interpreted as defined in class.

1. Fréchet-Hoeffding bounds [2 + (2 + 1)]

Recall the bivariate min (counter-monotonicity), and max (co-monotonicity) copulae, defined respectively as

$$m(u, v) = \max\{u + v - 1, 0\}, \text{and } M(u, v) = \min\{u, v\}.$$

(a) Prove the Fréchet-Hoeffding bounds: for any bivariate copula C,

$$m(u, v) \leq C(u, v) \leq M(u, v), \forall (u, v) \in [0, 1]^2.$$

(b) Hence show that, for any random 2-vector (X, Y), one has

$$\max\{F_X(t), F_Y(t)\} \leq \mathbb{P}(\min\{X, Y\} \leq t) \leq \min\{F_X(t) + F_Y(t), 1\}.$$

Give examples of random vectors for which equalities hold in the above bounds.

2. Multivariate hypothesis testing [4 + 1]

Consider a data matrix $X_{n\times p} \sim N_p(\mu, \Sigma)$. Let $\mathcal{R}$ be the population correlation matrix. Derive the likelihood ratio (LR) statistic (you should simplify its expression as much as possible) for the hypothesis

$$H_0: \text{ all the eigenvalues of } \mathcal{R} \text{ are equal.}$$

What is the limit distribution of $-2 \log \lambda$?

3. Discriminant analysis [2 + (2 + 1)]

(a) Recall that, in the two class case, Fisher's linear discriminant rule allocates an observation to class 1 if

$$h_{\text{LDA}}(x) = (\bar{x}_1 - \bar{x}_2)^\top S^{-1}_{\text{pooled}}\big(x - \frac{\bar{x}_1 + \bar{x}_2}{2}\big) \geq 0.$$

Show that Fisher's rule is equivalent to the following classification rule:

$$\text{assign } x \text{ to class } a \text{ if } a = \arg\min_{1\leq a'\leq 2} (x - \bar{x}_{a'})^\top S^{-1}_{\text{pooled}}(x - \bar{x}_{a'}),$$

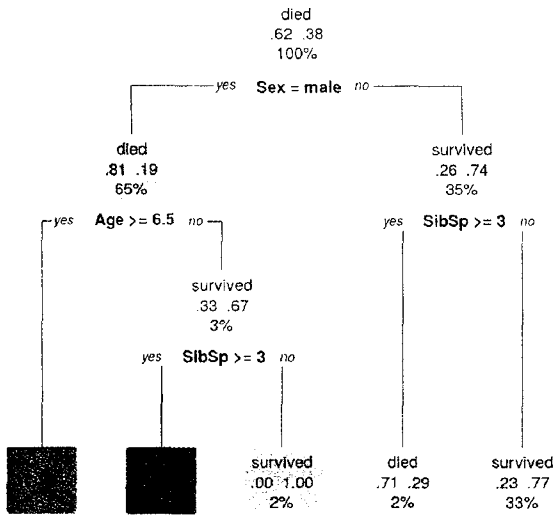
with ties broken in favour of class 1.

(b) In a two-class problem, suppose observations in class 1 come from $\mathcal{N}_2(\mu, \Sigma)$, and from $\mathcal{N}_2(-\mu, \Sigma)$ in class 2. Assuming $\mu = (\mu_1, \mu_2)^\top \neq (0, 0)^\top$ and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ to be known ($|\rho| < 1$), compute the Bayes risk $R_\pi(\mu, \rho) = \pi_1 p_{21} + \pi_2 p_{12}$ of the Bayes discriminant rule with respect to class-prior $\pi = (\pi_1, \pi_2) = (\frac{1}{2}, \frac{1}{2})$. Then find conditions on μ, ρ, under which correlation improves discrimination, i.e. $R_\pi(\mu, \rho) < R_\pi(\mu, 0)$.

4. CART: pruning a particular classification tree [5]

Suppose that we have grown the tree shown in Figure 1 to predict the survival status of a Titanic passenger based on age, sex, and the number of siblings onboard ("SibSp"), call it T. For each $\alpha \geq 0$, find the smallest optimally pruned subtree $T(\alpha)$ (according to the cost-complexity criterion $R_\alpha(T') = R(T') + \alpha|\tilde{T}'|$, where $R(T')$ is the fraction of misclassifications by the tree T').

Figure 1: Classification tree for Titanic survival data. Each node contains the following information: proportion of dead and surviving passengers, and the fraction of data points in that node.



5. CART: the smallest optimally pruned subtrees with respect to different complexity penalties [5]

Let T be a classification tree. Let $\alpha < \beta$. Show, from first principles, that $T(\beta)$ is the smallest optimal β-pruned subtree of $T(\alpha)$, i.e. show that, if $T(\alpha) = T'$, then $T'(\beta) = T(\beta)$.

6. SVM: the linearly separable case [3 + 2]

Consider a two-class dataset of size n where the two classes are linearly separable in the covariate space. Show that, in this case, the hard margin and the soft-margin SVMs are equivalent. Argue why

$$\text{LOOCV} \leq \frac{\#\text{ support vectors}}{n},$$

where LOOCV is the leave-one-out cross-validation error.

7. Spectral clustering [3 + (1 + 1)]

(a) Let A be the adjacency matrix of the ϵ-neighbourhood graph G_ϵ (with respect to a given metric d and a threshold $\epsilon > 0$) of a dataset with n points. Show that v is an eigenvector of the standard graph Laplacian $L = A - D$ corresponding to the eigenvalue 0 if and only if v is constant on the connected components of G_ϵ (i.e. $v_i = v_j$ if i and j are connected by a path in G_ϵ).

(b) How are the eigenvalues and eigenvectors of L related to those of the normalized graph Laplacians $L_{\text{sym}} = I - D^{-1/2}AD^{-1/2}$ and $L_{\text{rw}} = I - D^{-1}A$? Are L_{sym} and L_{rw} positive semi-definite?

8. Hierarchical clustering [2 + 2 + 1]

Apply the agglomerative single and complete linkage hierarchical clustering algorithms on the following distance matrix:

$$D = \begin{pmatrix} 0 & & & & & & \\ 2 & 0 & & & & & \\ 4 & 5 & 0 & & & & \\ 7 & 8 & 4 & 0 & & & \\ 2 & 3 & 2 & 6 & 0 & & \\ 2 & 3 & 2 & 6 & 1 & 0 & \\ 4 & 5 & 1 & 4 & 3 & 3 & 0 \end{pmatrix}.$$

Explain your steps in full detail and draw complete dendograms. Given the ground truth partition $\{\{1,2,5,6\},\{3,4,7\}\}$, compare these dendograms.

INDIAN STATISTICAL INSTITUTE

Semestral Examination (Backpaper) : Semester I (2018-19)

M. STAT. I Year

Stochastic Processes

Date: 28.12.2018 Maximum marks: 100 Time: 3 hours

Answer as many as you can. Total mark is 105.

1. State if the following statements are true or false giving suitable reasons.

   (a) A renewal process is Markovian.

   (b) For a renewal process with renewal distribution $F(\cdot)$, the limiting probability that there are odd number of renewals up to a time t is 1/2.

   (c) If the intensity process of a point process $X(t)$ is $\lambda E[Y(t)]$, where $Y(t)$ is another independent linear birth and death process and $\lambda > 0$, then $X(t)$ is a doubly stochastic Poisson process.

   (d) A delayed renewal process with the first renewal time having density $\frac{x}{50}$, $0 \leq x \leq 10$, and the subsequent renewal times having density $\frac{1}{10}$, $0 \leq x \leq 10$, has the expected number of renewals $M_D(t)$, by time t (≤ 10), as $\frac{t}{5}$.

   (e) For a discrete time branching process $\{X_n\}$ with $X_0 = i(> 1)$ and m being the mean of the offspring distribution, $E[X_n] = m^i$.

[4 × 5 = 20]

2. Consider a linear death process $X(t)$ with death rate μ per individual, starting with N individuals at time 0.

   (a) Obtain the differential equation for the probability generating function of $X(t)$ and solve for it. Hence, or otherwise, find the distributions of $X(t)$ (for a fixed t) and T, the time of extinction. Also, derive $E[X(t)]$ and $E[T]$.

   (b) Prove or disprove if the times between successive deaths form a renewal process.

   (c) If immigration with rate λ is allowed to this process, then obtain the limiting distribution of the resulting process.

   (d) Suppose this process with immigration is observed from time 0 to τ. Describe the data and obtain maximum likelihood estimates of λ and μ.

[(2+3+3+3+2+2)+4+4+(2+5)=30]

3. Consider a Markov linear growth process with two sexes in which a female gives birth to either a male or a female offspring with rates λ_1 and λ_2, respectively, and dies with rate μ_1; males do not give birth but die with rate μ_2 per individual. Assume the individuals to act independently. Starting with a female at time 0, derive

   (a) the expected number of females at time t and the expected number of males ever born up to time t, and

   (b) the distribution of the time of first birth of a male child by specifying the corresponding hazard rate as explicitly as possible.

[(4+6)+5=15]

4. (a) For a Poisson process with rate λ, derive $P[\beta_t > z]$, where β_t is the total life at time t. Hence, prove that the mean total life time at time t is $\frac{1}{\lambda}(2 - e^{-\lambda t})$.

   (b) Find the limiting distribution of β_t, the total life at time t, for a renewal process with renewal distribution F. Prove that the mean of this limiting distribution is greater than or equal to the mean renewal time.

[(10+5)+((10+5)=30]

5. A rare mutation takes place in one normal cell with probability $\nu(t)\Delta t + o(\Delta t)$ in the time interval $(t, t + \Delta t)$ during division (that is, during mutation a normal cell divides into one normal cell and one mutant cell). Probability of two or more mutations in $(t, t + \Delta t)$ is $o(\Delta t)$. Normal cells do not divide otherwise and they act independently. Start with a body of S normal cells at time 0. Each mutant cell born in this way grows according to a linear birth and death process with rates λ and μ, respectively, per cell, thereby forming a mutant clone. Let $Y(t)$ denote the number of non-extinct mutant clones at time t. Derive the distribution of $Y(t)$. [10]

Indian Statistical Institute
Backpaper Examination

January 28, 2019

Multivariate Analysis, M1

Total points: 100 — Time: 4 hours

Note: This is a **closed notes/closed book** examination. Notations, if not explicitly explained, are to be interpreted as defined in class.

1. Multivariate hypothesis testing [8 + 2]

Consider n i.i.d. data points from $\mathcal{N}(\mu, \Sigma)$. Derive the likelihood ratio test for testing

$$H_0: \text{ the } \sigma_{ij}\text{'s are all equal}, i \neq j.$$

What is the asymptotic null distribution of $-2\log\lambda$?

2. Principal components analysis [(4 + 1) + (3 + 2)]

(a) Consider data from $\mathcal{N}(\mu, \Sigma)$. Let ℓ_i, λ_i be the i-th largest eigenvalues of S and Σ, respectively. Assuming all the eigenvalues of Σ to be distinct, show that

$$\sqrt{n}(\log \ell_i - \log \lambda_i) \xrightarrow{d} \mathcal{N}(0, 2).$$

Use this to construct an asymptotic $(1-\alpha)\%$ confidence interval for λ_i.

(b) How does performing PCA on the covariance matrix of (X_1, X_2) differ from performing PCA on the covariance matrix of (cX_1, X_2), $c > 0$? Imagine X_1 as being measured in metres and X_2 in grams, and c as being a conversion factor changing the units to kilometres. How would you do PCA on such a dataset?

3. Multidimensional scaling [2 + 2 + 5 + 1]

Consider the five neighbouring states of Madhya Pradesh (MP) as shown in Figure 1. Let the distance between state i and state j be $d(i,j)$ = "the minimum number of borders one needs to cross to reach state i from state j". Construct the corresponding distance matrix between these five states. Is this an Euclidean distance matrix? Perform Multidimensional scaling with $d = 2$ and plot the resulting points. How does this plot compare to the actual map? (Hint: to compute eigenvalues and eigenvectors, you may use the fact that the an $n \times n$ circulant matrix C, whose i-th row, $i \in \{0, 1, \ldots, n-1\}$, is a cyclic permutation of its first row $(c_0, \ldots, c_{n-1})^\top$ with offset i, has eigenvalues $\sum_{i=0}^{n-1} c_i\omega^i$ with corresponding eigenvectors $\frac{1}{\sqrt{n}}(1, \omega, \ldots, \omega^{n-1})^\top$, where ω is an n-th root of unity.)

4. Factor analysis [2 + (2 + 2) + 4]

(a) Describe the orthogonal factor model.

(b) Describe the method of principal factors and the method of principal components for fitting an orthogonal factor model.

(c) Fit the orthogonal factor model with a single factor to an equicorrelation matrix of order 3 and interpret the results.

5. Admissibility of the Bayes classification rule [2 + (2 + 6))]

Consider a k-class classification problem. Define the concept of admissibility of a classification rule with respect to a general loss function $c(i\,|\,j)$. Derive the Bayes rule with respect to some prior π and show that it is admissible.

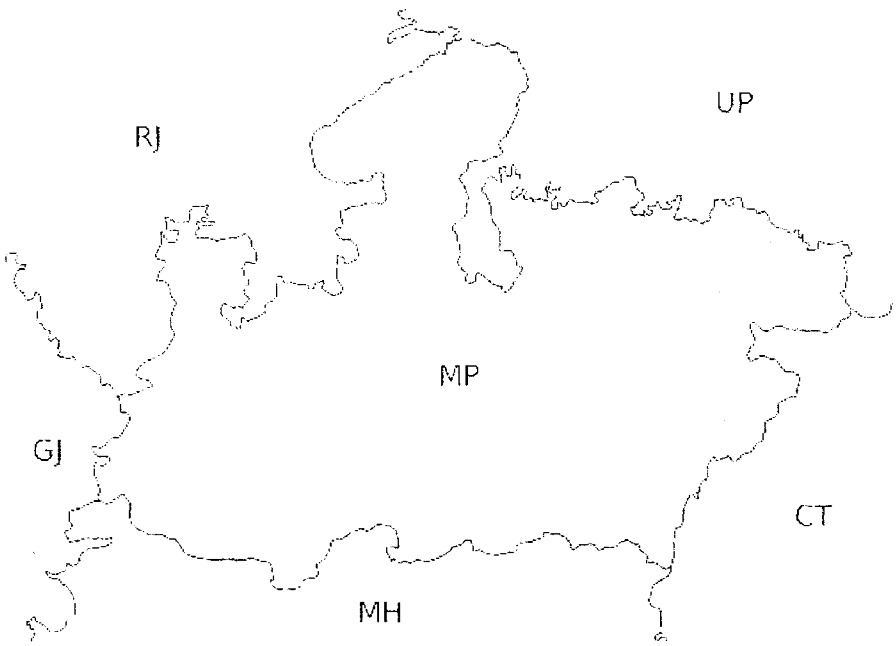


Figure 1: Madhya Pradesh and its five neighbouring states.

6. Cross-validation [(2 + 2) + 6]

What is cross-validation? Why is it used in supervised learning? Show that if one chooses a classifier among several competing ones based on cross-validation, then the cross-validation error of the chosen classifier is an underestimate of its true error.

7. Fisher's LDA [5 + 5]

In a binary classification problem, suppose that we have training data $(X_{n\times p}, Y_{n\times 1})$. Derive Fisher's linear discriminant rule for this problem (from Fisher's original considerations). Suppose now that we change the predictors X to $\hat{Y} = X(X^\top X)^{-1}X^\top Y$ via linear regression (assume X to have full column rank). Show that performing Fisher's LDA on the transformed data $(\hat{Y}, Y)$ is equivalent to LDA on the original data.

8. CART [5 + (1 + 4)]

(a) Describe the CART algorithm for the classification problem in detail.

(b) Define the notion of the *smallest* optimally pruned subtree with respect to a given complexity parameter α and prove that such a subtree always exists.

9. SVM [5 + (2 + 1 + 2)]

(a) Derive the dual of the soft-margin SVM

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(1 - y_i(w^\top x_i + b))_+.$$

(b) Perform support vector classification on the following dataset:

X	1	2	3.5	4	5
Y	−1	−1	−1	1	1

Find optimal values of w and b. Identify the support vectors. Calculate the leave-one-out-cross-validation (LOOCV) error for this dataset.

10. Hierarchical clustering [1 + 1 + 2]

Write down the single and complete linkage hierarchical clustering algorithms. Discuss their advantages and disadvantages. In the dataset depicted in Figure 2, which one would perform better and why?

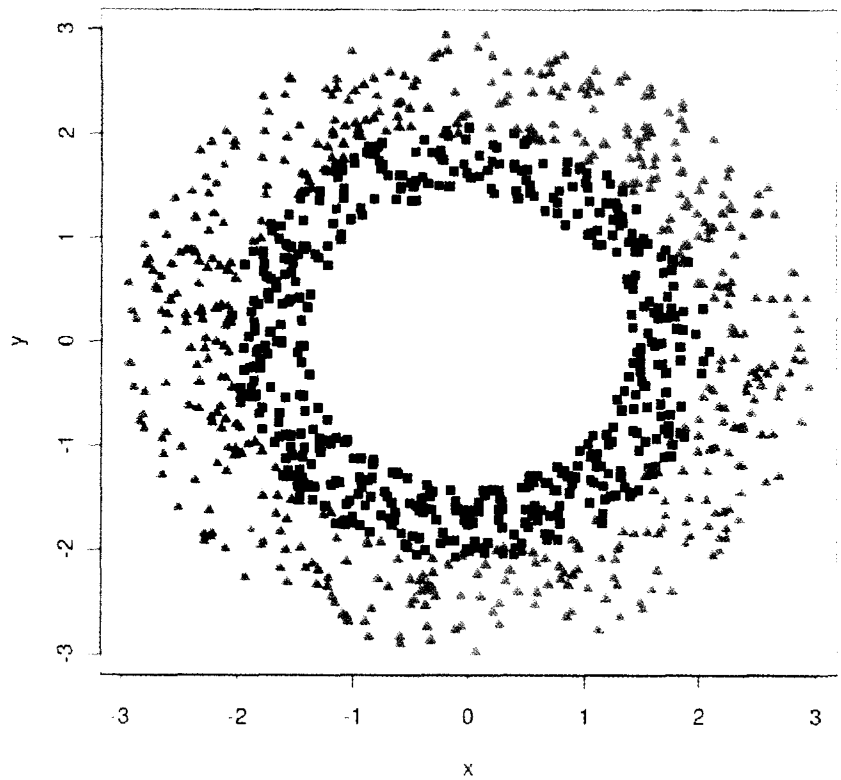


Figure 2: Two annular clusters whose points are plotted respectively as triangles and squares.

INDIAN STATISTICAL INSTITUTE
Back-paper Examination 2018-19

M. Stat. 1st Year
Statistical Inference I

29 January, 2019

Maximum Marks: 100 Time: 3 hours

[Note: Notations are as used in the class. Answer as much as you can. Best of Luck!]

1. $X_1, \ldots, X_n$ are n (> 2) independent and identically distributed observations from a Uniform $U(\theta - \sigma, \theta + \sigma)$ with known $\sigma > 0$. Find out a minimum risk equivariant estimator of θ under squared error loss function, which is independent of the value of σ. [10]

2. Suppose $X_1, \ldots, X_n$ are independent and identically distributed observations from $N(\mu, \sigma^2)$, where σ is a known positive real number and consider the squared error loss function for the estimation of μ.

   (a) Find the Bayes estimator of μ with respect to the uniform prior $\pi(\mu) = 1$.

   (b) Find the Bayes estimator of μ with respect to the conjugate prior.

   (b) Find a Minimax estimator of μ. Is it also a Bayes estimator? [3+5+(5+2)]=15

3. Suppose $X \sim U(\theta, \theta + 1)$, the uniform distribution with location parameter $\theta \in \mathbb{R}$.

   (a) Verify if this family has the monotone likelihood ratio property.

   (b) Hence, or otherwise, derive a uniformly most powerful level-α test for the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ for a given θ_0. [5+5]=10

4. Suppose $X_1, \ldots, X_n$ are n (> 2) independent and identically distributed observations from $N(0, \sigma^2)$. Derive the UMPU level-α test for the hypothesis $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$. State the result clearly, with all notations and assumptions, that you need to use for this purpose. [10+5]=15

5. For a decision problem with finite parameter space Θ, write down the Minimax theorem with all notations and assumptions. Also prove it from the first principle. [5+15]=20

6. Prove or disprove the following for a standard decision problem as defined in the class.

(a) Sample mean vector is admissible for estimation of the population mean vector under multivariate normal distribution with known covariance matrix $\sigma^2\mathbf{I}$ and the squared error loss.

(b) A minimal complete class, if it exists, consists of exactly the admissible rules.

(c) The subclass of behavioral decision rules based on a sufficient statistic is an essentially complete class.

(d) When the parameter space is finite and the risk set is closed and bounded from below, a Bayes rule with respect to a proper prior exists if the prior has support on the full parameter space.

(e) The power function of any non-randomized test for a parametric hypothesis is nondecreasing in the parameter θ. [6×5]=30

Mid-Semestral Examination (Supplementary)

M.Stat. - I

(2018 - 2019)

Categorical Data Analysis

Full Marks -50

Time : 2 hours 30 mins.

(Answer as many as you can. The maximum you can score is 50)

1. Muriel Bristol, A colleague of Ronald Fisher's, claims that when drinking tea, she could distinguish whether milk or tea was added to the cup first (she preferred milk first). To test her claim, Fisher asked her test 8 cups of tea, 4 of which had milk added first and 4 of which had tea added first. She knew there were 4 cups of each type and had to predict which 4 had the milk added first. The order of presenting the cups to her was randomized. The result of the experiment are as follows:

	Guess Poured First		
Poured First	Milk	Tea	Total
Milk	3	1	4
Tea	1	3	4
Total	4	4	

Construct the hypothesis, test statistics and evaluate the P – value to judge Dr. Bristol's claim.

[3 + 4 + 3 = 10]

2. A sample of 156 dairy calves born in a county, Florida, USA, were classified according to whether they caught pneumonia within 60 days of birth. Calves that got a pneumonia infection were also classified according to whether they got a secondary infection within 2 weeks after the first infection cleared up. The following table shows the data. Calves that did not gave a primary infection could not get a secondary infection, so no observations can fall in the category for "no" primary infection and "yes" secondary infection.

Test whether the probability of primary infection was the same as the conditional probability of the secondary infection given that the calf already got the primary infection.

	Secondary Infection	
Primary Infection	Yes	No
Yes	30 (38.1)	63 (39.0)
No	0 (-)	63 (78.9)

[10]

3. (a) Define Odds, risk, odds ratio and risk ratio for 2×2 contingency tables with suitable numerical examples.

(b) Deriving necessary results (using delta methods) find the asymptotic distribution of the sample logit.

[6 + 5 + 4 = 15]

4. Let $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})$ denote setting i of values of p explanatory variables, $i = 1(1)N$. When more than one observation occurs at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of successes. Let y_i refer to this success count rather than to an individual binary response. Then $Y_1, Y_2, \ldots, Y_N$ are independent binomials with $E(Y_i) = n_i\pi(x_i)$. Using logistic regression set up for the probability $\pi(x_i)$.

(a) Construct the likelihood equation.

(b) Discuss Fisher's scoring and Newton-Raphson iterative methods for evaluating the estimate of the parameters of the likelihood equation.

[4 + 6 = 10]

5. (a) Discuss the measure of ordinal association with suitable examples, which was proposed by Goodman and Kruskal.

(b) The likelihood-ratio statistic G^2 is asymptotically equivalent to χ^2 as $n \to \infty$ based on the counts $(n_1, n_2, \ldots, n_N)'$ in N cells of a contingency tables (The asymptotic regards N as fixed and $n = \sum n_i \to \infty$).

[5 + 5 = 10]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2018 − 19

M. Stat 1st Year

Measure Theoretic Probability

Date:18/02/2019 Maximum Marks: 40 Duration:2hr 30mins

(1) If μ is a Radon measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then show that the set

$$\{x \in \mathbb{R} : \mu(\{x\}) > 0\}$$

is countable. [10 marks]

(2) Suppose $\mathcal{F}_0$ is a field on Ω and $\mu : \mathcal{F}_0 \to [0, \infty]$ be such that
 (a) $\mu(\emptyset) = 0$;
 (b) μ is finitely additive on $\mathcal{F}_0$;
 (c) whenever $A_n \in \mathcal{F}_0$ is such that $A_n \downarrow \emptyset$, it holds that $\mu(A_n) \downarrow 0$.

Show that μ is a field measure.

[10 marks]

(3) Let μ be a probability measure on $(\Omega, \mathcal{F})$ where $\mathcal{F} = \sigma(\mathcal{A})$ for some field $\mathcal{A}$. Show that for each $A \in \mathcal{F}$, and $\varepsilon > 0$, there exists $B \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$. [8 marks]

(4) Answer **any three** from the following: [$4 \times 3 = 12$ marks]
 (a) Let $\mathbb{Q}$ be the set of rationals and $f : \mathbb{R} \to \mathbb{R}$ be $f = \mathbf{1}_{\mathbb{Q}}$. Find $\int f d\lambda$ where λ is the Lebesgue measure.

 (b) Suppose $\mathcal{A}$ is a collection of subsets of Ω and P is a probability measure on $(\Omega, \sigma(\mathcal{A}))$. Assume that $P(A) = 0$ or 1 for all $A \in \mathcal{A}$. Show that the same is true for all $A \in \sigma(\mathcal{A})$.

(c) Prove or disprove: Suppose that $(\Omega_i, \mathcal{F}_i)$ are measurable spaces for $i = 1, 2$. If $f : \Omega_1 \to \Omega_2$ is a 1-1, onto, measurable function, then f^{-1} is a measurable function from Ω_2 to Ω_1.

(d) Let A be a rectangle in $\mathbb{R}^2$ and λ_2 be the 2-dimensional Lebesgue measure. Show that $\lambda_2(\partial A) = 0$ where ∂A is the boundary of a rectangle.

(e) If f and g are measurable functions from a measurable space $(\Omega, \mathcal{F})$ to $\overline{\mathbb{R}}$ then show that

$$\{\omega \in \Omega : f(\omega) = g(\omega)\} \in \mathcal{F}.$$

INDIAN STATISTICAL INSTITUTE

M. Stat First Year (2018-19)

Second Semester

Resampling Techniques

Date: 20/02/2019 Marks: 30 Duration: 1 hour 30 mins

Attempt all questions

1. (i) Obtain the jackknife estimator corresponding to the unbiased sample variance. Is it also unbiased? [5]

   (ii) Consider the statistic

$$T_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}\left(X_i + X_j \leq 0\right),$$

   where $\mathbb{I}$ stands for the indicator function. Obtain the jackknife estimator corresponding to T_n. [5]

2. Suppose that $X_1, \ldots, X_n$ are *iid* from a distribution with density $\tau^{-1} f_0\left(\frac{x-\mu}{\tau}\right)$, where μ and τ (> 0) are unknown parameters and f_0 is a known density saisfying $\int x f_0(x)dx = 0$ and $\int x^2 f_0(x)dx = 1$. Suppose that we are interested in estimating the quantity

$$H(x) = P\left(\frac{(\bar{X}_n - \mu)}{S_n} \leq x\right),$$

   for $-\infty < x < \infty$, where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ and $S_n^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 / n(n-1)$.

   (i) Using unbiased estimators whenever relevant, propose, with justification, a bootstrap methodology for estimating $H(\cdot)$. [5]

   (ii) Show that your bootsrap estimator is exactly the same as $H(\cdot)$. [5]

3. Suppose that $X_1, \ldots, X_n$ are *iid* with distribution function F having mean zero and finite variance. Consider the V-statistic $V_n = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j)$, with $h(x, y) = xy$.

   (i) Obtain the large sample distribution of nV_n. [4]

   (ii) Obtain the large sample distribution of nV_n^*, where V_n^* is the non-parametric bootstrap estimator of V_n. [6]

   [Hint: You may use bootstrap CLT of the following form: $\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \xrightarrow{\mathcal{L}} N(0, Var(X))$, almost surely.]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2018-19

Course Name : M.Stat. 1st Year
Subject Name : Sample Surveys and Design of Experiments
Date : Feb 21st, 2019 Total Duration : 2.30p.m.-4.30p.m.

Note: Use separate answer sheets for two groups.

Group A : Sample Surveys. (Total Marks = 20)

Answer any four questions.

Notations are as usual.

1. State and prove the theorem regarding the existence of uniformly minimum variance estimator for $Y = \sum_{i=1}^{N} Y_i$ of a variable of interest y within the class of **all unbiased estimators**. (5)

2. 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)

3. 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)

4. Define admissibility of an estimator within a class of estimators. Prove that Horvitz and Thompson's estimator (HTE) for Y is an admissible estimator within the class of **all homogeneous linear unbiased estimators**. (5)

5. Suppose $\nu(s)$ is the effective sample size (number of distinct units) of a sample s obtained by a sampling design p. Then prove that

   (i) $\Sigma_{i=1}^{N} \pi_i = E_p(\nu(s))$ and

   (ii) $\Sigma_{i=1}^{N} \Sigma_{j=1, j\neq i}^{N} \pi_{ij} = V_p(\nu(s)) + E_p(\nu(s)) \left(E_p(\nu(s) - 1\right)$. (5)

Please turn over

Group B: Design of Experiments : Total Marks = 30

Answer all questions

6. Let d_1 be a Randomised Block Design (RBD) with parameters b, v, k_1 and d_2 be a Balanced Incomplete Block Design(BIBD) with parameters b, v, r_2, k_2, λ. Let $d = [d_1 : d_2]$. Then show that d is a connected, nonorthogonal, balanced block design and obtain average variance of the BLUEs of all estimable elementary treatment contrasts(you can assume the formula) under the design d. [2+ 2+8+3=15]

7. Construct a BIBD with parameters $b = 12, v = 9, r = 4, k = 3, \lambda = 1$ (no need to discuss the method of construction). [7]

8. Consider the following row-column design d

$d =$

1	3	4	5
3	4	7	8
8	5	1	3
7	4	3	6

Obtain one unbiased estimator of each of $\tau_5 - \tau_7$ and $\tau_1 - \tau_4$. [4+4=8]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Stat. – I Year, 2018-2019 (Semester – II)

Optimization Techniques

Date: 22.02.2019 Maximum Marks: 60 Duration: 2 hours 30 mins

Note: The question paper is of 80 marks. Answer as much as you can, but the maximum you can score is 30 from Group-A and 30 from Group-B.

Notations: Vectors would be written in small letters with boldface, e.g. $\mathbf{b}$; matrices would be written in capital letters, e.g., A. Transpose of A would be denoted by A^T and transpose of $\mathbf{b}$ would be denoted by $\mathbf{b}^T$. Whenever we say that, $\mathcal{P}$ is a linear program, we mean $\mathcal{P}$ is of the form

$$\begin{array}{lll} \text{Maximize} & \mathbf{c}^T\mathbf{x} & \\ \text{subject to} & A\mathbf{x} & \leq \mathbf{b} \\ & \mathbf{x} & \geq 0 \end{array}$$

and $\mathcal{P}_{eq}$ will denote a linear program of the form

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array}$$

Let $C \subseteq \mathbb{R}^n$ be a non-empty convex set. Then $\mathbf{x}$ is an *extreme point* of C if there are no two points $\mathbf{x}_1, \mathbf{x}_2$ in C different from $\mathbf{x}$ and $\lambda \in (0,1)$ such that $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$.

Group-A

(AQ1) Let $P = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$. Show that $\mathbf{x}$ is an extreme point of P if and only if $\mathbf{x}$ is a basic feasible solution of P. [10]

(AQ2) Let $C \subseteq \mathbb{R}^n$ be a non-empty closed convex set and $\mathbf{y} \in \mathbb{R}^n \setminus C$. Show that $\mathbf{z} = \text{argmin}_{\mathbf{x}\in C}||\mathbf{y}-\mathbf{x}||$ if and only if, $\forall \mathbf{x} \in C$, we have $(\mathbf{y}-\mathbf{z})^T(\mathbf{x}-\mathbf{z}) \leq 0$. [10]

(AQ3) Show that if the objective function of $\mathcal{P}_{eq}$ is bounded above, then for every feasible solution $\mathbf{x}_0$, there exists a basic feasible solution $\mathbf{x}'$ such that $\mathbf{c}^T\mathbf{x}' \geq \mathbf{c}^T\mathbf{x}_0$. [10]

(AQ4) Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then exactly one of the following sets must be empty:

(i) $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$

(ii) $\{\mathbf{y} : A^T\mathbf{y} \leq \mathbf{0}, \mathbf{b}^T\mathbf{y} > 0\}$

[10]

Group-B

(BQ1) Let $f(x) = \max(\mathbf{c}_1{}^T\mathbf{x} + d_1, \mathbf{c}_2{}^T\mathbf{x} + d_2, \ldots, \mathbf{c}_p{}^T\mathbf{x} + d_p)$. For such a function f, consider the mathematical program

$$\begin{array}{lll} \text{Minimize} & f(x) & \\ \text{subject to} & A\mathbf{x} & \leq \mathbf{b} \\ & \mathbf{x} & \geq \mathbf{0} \end{array}$$

Can you convert this mathematical program to a linear program? Explain with proper arguments. [10]

(BQ2) In the *set cover* problem, we have an universe $\mathcal{U} = \{u_1, \ldots, u_n\}$ of n elements. Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a set of m sets, where each set $S_i \subseteq \mathcal{U}$. Each set S_i has a weight $w_i \geq 0$. The problem in *set cover* is to find a minimum weight collection of subsets of $\mathcal{S}$ that covers all elements of $\mathcal{U}$.

(a) Write an integer linear program (ILP) for the *set cover problem* using decision variables x_i to indicate whether the set S_i is included in the solution or not.

(b) Relax the above ILP and round the optimal solution of the linear program as follows: given the optimal solution $\mathbf{x}^*$ of the linear program, we include the subset S_i in our solution if and only if $x_i^* \geq \frac{1}{f}$, where f is the maximum number of sets in which any element appears and x_i^* is the i-th component of $\mathbf{x}$.

For this rounding scheme, show that the set generated is a set cover and is an f-factor approximation algorithm.

[3+7=10]

(BQ3) Let $\mathcal{C}, \mathcal{D} \subseteq \mathbb{R}^n$ be two non-empty disjoint convex sets such that the set $\mathcal{C} - \mathcal{D}$, where $\mathcal{C} - \mathcal{D} := \{\mathbf{x} - \mathbf{y} : \mathbf{x} \in \mathcal{C} \text{ and } \mathbf{y} \in \mathcal{D}\}$, is closed. Using the result from (AQ2), prove that $\exists\, \mathbf{a} \in \mathbb{R}^n$, with $\mathbf{a} \neq \mathbf{0}$, such that $\sup_{\mathbf{x} \in \mathcal{C}} \mathbf{a}^T\mathbf{x} \leq \inf_{\mathbf{y} \in \mathcal{D}} \mathbf{a}^T\mathbf{y}$. [10]

(BQ4) Consider $\mathcal{P}_{eq}$, where all entries of $A = [a_{ij}]$, $\mathbf{b}$ and $\mathbf{c}$ are integers. Let $\mathbf{x} = \{x_1, \ldots, x_i, \ldots, x_n\}$ be a basic feasible solution. Then, prove that $|x_i| \leq m!\ \alpha^{m-1}\beta$, where $\alpha = \max_{i,j}\{|a_{ij}|\}$ and $\beta = \max_{j=1,\ldots,m}\{|b_j|\}$. [10]

Midterm Examination
Large Sample Statistical Methods
Second Semester
2018-2019 Academic Year
M.Stat. First Year

Date : 25.02.2019 Maximum Marks: 40 Duration : $2\frac{1}{2}$ hours

Answer as many questions as you can. The maximum you can score is 40

1. Let $X_n \sim$ Binomial$(n, \frac{1}{2})$ for $n \geq 1$ and suppose $\Phi(.)$ denotes the distribution function of a $N(0,1)$ distribution. Does $P(X_n \leq \frac{n+n^{3/4}}{2}) - \Phi(n^{1/4})$ converge to a limit as $n \to \infty$? Prove your assertion. [4]

2. Consider a sequence X_n of random variables on a common probability space. Show that one can find a sequence of *strictly positive* real constants b_n such that $b_n X_n$ converges to zero almost surely. [5]

3. Suppose a sequence X_n of random variables is $O_p(1)$. Does this mean that $P(\sup_n |X_n| < \infty) = 1$? Prove or give a counterexample. [6]

4. 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. [6]

5. (a) State (with assumptions) the weak Bahadur representation theorem of a sample quantile. [2]

   (b) Argue how this result implies asymptotic normality of (suitably centred and scaled) sample quantile. [2]

   (c) Let $X_1, X_2, \ldots X_n$ be iid observations from a $N(\mu, 1)$ distribution with μ 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, can you give an estimate of the sample mean ? Justify your answer. [4+3=7]

6. (a) Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$, where $\mu \in \mathcal{R}$ and $\sigma > 0$ are unknown. Let $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Find a variance stabilizing transformation for S_n^2 and hence an approximate 95 percent asymptotic confidence interval for σ^2. Prove your answer. You may assume the asymptotic distribution for S_n^2. [2+1=3]

(b) Consider now the problem of testing $H_0 : \sigma^2 \leq 1$ versus $H_1 : \sigma^2 > 1$. The standard test rejects H_0 is nS_n^2 exceeds the upper α point of the χ^2_{n-1} distribution and this is an exact level α test under iid sampling from a normal distrubution. Consider now iid sampliing from a distribution (not necessarily normal) having finite fourth moments. Suppose one uses the same critical region as above for testing H_0. Will the test be approximately level α if n is large ? Justify your answer. [7]

INDIAN STATISTICAL INSTITUTE

Inference for High Dimensional Data

Time: $2\frac{1}{2}$ Hours | Mid-Semestral Examination | Full Marks: 60

[Answer all questions. The maximum you can score is 60.]

1. Let $\{(\mathbf{x}_i, y_i) : i = 1, 2, \ldots, 50\}$ be a data set consisting of 50 observations, where $\mathbf{x}_i \in R^{100}$ and y_i takes the values 1 and -1, if $\mathbf{x}_i$ comes from F_1 and F_2, respectively. If the observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{50}$ are in general position, show that there exists $\boldsymbol{\alpha} \in R^d$ and $\beta \in R$ such that $(\boldsymbol{\alpha}^\top \mathbf{x}_i + \beta) y_i \geq 0$ for all $i = 1, 2, \ldots, 50$. [6]

2. Suppose that we have a sample $\{\mathbf{z}_1 = \binom{\mathbf{x}_1}{\mathbf{y}_1}, \mathbf{z}_2 = \binom{\mathbf{x}_2}{\mathbf{y}_2}, \ldots, \mathbf{z}_{50} = \binom{\mathbf{x}_{50}}{\mathbf{y}_{50}}\}$ of 50 independent observations from the joint distribution of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ of dimensions 30 and 60, respectively. Show that if the observations $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{50}$ are in general position, we can always find $\boldsymbol{\alpha} \in R^{30}$ and $\boldsymbol{\beta} \in R^{60}$ such that the sample correlation coefficient between $\boldsymbol{\alpha}^\top \mathbf{X}$ and $\boldsymbol{\beta}^\top \mathbf{Y}$ is unity. [6]

3. Consider a data set of the form $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots, (y_n, \mathbf{x}_n)$, where $\mathbf{x}_i \in R^p$ and $y_i \in R$ for $i = 1, 2, \ldots, n$. Also assume that (i) $\sum_{i=1}^n y_i = 0$, (ii) $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$, the zero vector and (iii) $\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top = \mathbf{I}$, the $p \times p$ identity matrix.

   (a) If $\boldsymbol{\beta}_0$ is the minimizer of $\sum_{i=1}^n (y_i - \boldsymbol{\beta}^\top \mathbf{x}_i)^2$ and $\boldsymbol{\beta}_\lambda$ is the minimizer of $\sum_{i=1}^n (y_i - \boldsymbol{\beta}^\top \mathbf{x}_i)^2 + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}$, show that $\boldsymbol{\beta}_\lambda = \boldsymbol{\beta}_0 / (1 + \lambda)$. [4]

   (b) Find the minimizer of the function $\psi(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \boldsymbol{\beta}^\top \mathbf{x}_i)^2 + \lambda \sum_{j=1}^p |\beta_j|$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$. [8]

4. Consider a multiple testing problem involving a family of n null hypotheses. Show that if $\sum_{k=1}^n 1/k > 1/\alpha$, the Benjamini-Hochberg method can have false discovery rate unity. [8]

5. Prove or disprove the following statements:- [6 × 4 =24]

   (a) If $X_1, X_2, \ldots, X_n$ are positively regression dependent statistics, for any increasing set D and $i = 1, 2, \ldots, n$, $P[(X_1, X_2, \ldots, X_n) \in D \mid X_i \leq x]$ is an increasing function of x.

   (b) If $X_1, X_2, \ldots, X_n$ are positively regression dependent statistics and $\psi_1, \psi_2, \ldots, \psi_n$ are monotonically decreasing functions, $\psi_1(X_1), \psi_2(X_2), \ldots, \psi_n(X_n)$ are positively regression dependent.

   (c) Like LASSO, non-negative Garotte can lead to automatic model selection by making some of the regression coefficients equal to zero.

   (d) Let $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed $N_d(\mathbf{0}, \sigma^2 \mathbf{I})$ variates, If $\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$, $\frac{1}{d} \|\mathbf{X}_0 - \bar{\mathbf{X}}_n\|^2$ converges in probability to $(1 + \frac{1}{n}) \sigma^2$ as the dimension d diverges to infinity.

6. Consider a two-class classification problem between two d-dimensional normal distributions $N_d(\mathbf{0}_d, \mathbf{I}_d)$ and $N_d(\boldsymbol{\mu}_d, 4\mathbf{I}_d)$, where $\mathbf{0}_d = (0, 0, \ldots, 0)^\top$, $\boldsymbol{\mu}_d = (\mu, \mu, \ldots, \mu)^\top$ and $\mathbf{I}_d$ is the $d \times d$ identity matrix. Suppose that the training sample is formed by taking 10 observations from each of the two distributions. Consider 1-nearest neighbor (1-NN) classifier based on the Euclidean distance constructed using these observations.

   (a) Calculate the asymptotic misclassification probability of this classifier when $\mu = 1.7$ and d diverges to infinity. [6]

   (b) Will this asymptotic misclassification probability change if instead of $\mu = 1.7$, we use $\mu = 1.75$? Justify your answer. [2]

INDIAN STATISTICAL INSTITUTE

M. Stat First Year (2018-19)

Second Semester

Resampling Techniques

Date: 18/04/2019 Marks: 100 Duration: 2 hrs 30 mins

Attempt all questions

(1) Suppose that we have n *iid* observations $X_1, X_2, \cdots, X_n$ such that $Pr(X_1 = 1) = 1 - Pr(X_1 = 0) = \theta$, where $0 < \theta < 1$ is an unknown parameter. Bootstrap is done by drawing *iid* observations $X_1^*, X_2^*, \cdots, X_n^*$ from the empirical distribution.

(i) Obtain the probability distributions of $\widehat{\theta}_n =$ the sample proportion of 1's and $\widehat{\theta}_n^* =$ the bootstrap proportion of 1's.

(ii) What is the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n)$ as $n \longrightarrow \infty$? Justify.

[4+6+15]

(2) Let $X_1, \ldots, X_n \stackrel{iid}{\sim} F$ (univariate) with density f. Let $T(F) = F^{-1}\left(\frac{1}{2}\right)$. Let $n = 2m$ and assume that $g = F^{-1}$ has a continuous derivative in a neighborhood of $1/2$.

(i) Obtain the jackknife variance estimator $v_{JACK}^{(n)}$ of $T(F_n)$.

(ii) Obtain $\lim_{n\to\infty} nv_{JACK}^{(n)}$. [Hint: Use $X_{(i)} = F^{-1}\left(U_{(i)}\right)$, where $U_i \stackrel{iid}{\sim} Uniform(0,1)$, apply the mean value theorem, and use the result $n\left(U_{(i)} - U_{(i-1)}\right) \xrightarrow{\mathcal{L}} Exponential(1)$, as $n \to \infty$.]

(iii) Is the limit in (2) (ii) consistent with the limit of $nVar_F\left(T(F_n)\right)$? Here $Var_F\left(T(F_n)\right)$ is the variance of $T(F_n)$ under the distribution F.

[4+18+3]

(3) (i) Assume that for $i = 1, \ldots, n$, $X_i \stackrel{iid}{\sim} Bernoulli(\theta)$. With the prior $\theta^{-1}(1-\theta)^{-1}$, show that the posterior of θ is in agreement with the Bayesian bootstrap distribution.

(ii) Discuss with an example, a method to incorporate prior information in Bayesian bootstrap, when the data distribution is unknown.

(iii) Describe the weighted likelihood bootstrap. How can you make it simulation consistent?

[6+10+9]

(4) (i) Consider the functional $T(F) = Var_F(X)$, the variance of the random variable X under the distribution F. Obtain the Gateaux derivative of F at F, and the corresponding influence curve.

(ii) Let ρ be a suitable metric such that

$$\rho(F_n, F) \xrightarrow{a.s.} 0$$

and

$$\sum_{i=1}^{n} [\rho(F_{n-1,i}, F_n)]^2 = O\left(n^{-1}\right), a.s.$$

Here F_n is the empirical distribution function based on *iid* observations $X_1, \ldots, X_n$ and $F_{n-1,i}$ is that based on $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$. If T is continuously ρ-Frechet differentiable at F with $\phi_F \neq 0$, then show that

$$nv_{JACK}^{(n)}/\sigma^2 \xrightarrow{a.s.} 1,$$

where $v_{JACK}^{(n)}$ is the jackknife variance estimator of $T(F_n)$, $\sigma^2 = E\left[\phi(X_1)\right]^2$ and ϕ is the influence function of T.

[10+15]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Stat. – I Year, 2018-2019 (Semester – II)

Optimization Techniques

Date: 22.04.2019 Maximum Marks: 100 Duration: 4 hours

Note: The question paper is of 125 marks. Answer as much as you can, but the maximum you can score is 40 from Group-A and 60 from Group-B.

Notations: Transpose of a matrix A would be denoted by A^T and transpose of a vector b would be denoted by b^T.

Whenever we say that, $\mathcal{P}$ is a linear program, we mean $\mathcal{P}$ is of the form

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}^T x \\ \text{subject to} & Ax \leq b \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

Let A and B be $n \times n$ matrices, then we write $A \succeq B$ if $A - B$ is positive semi-definite matrix.

The *domain* of a function f will be denoted by $dom(f)$.

Gradient and *Hessian* of a function f at x will be denoted by $\nabla f(x)$ and $\nabla^2 f(x)$, respectively.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *strongly convex* if there exists $\alpha > 0$ such that the function $f(x) - \alpha\|x\|^2$ is a convex function.

$\mathcal{MP}$ will denote the following mathematical program in $\mathbb{R}^n$:

$$\begin{array}{ll} \text{Minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \;\; i \in \{1, \ldots, m\} \\ & h_j(x) = 0, \;\; j \in \{1, \ldots, p\} \end{array}$$

Let $\Omega \subseteq \mathbb{R}^n$, and $f : \Omega \to \mathbb{R}$. Then $g \in \mathbb{R}^n$ is a *subgradient* of f at $x \in \Omega$ if for any $y \in \Omega$ one has

$$f(x) - f(y) \leq g^T(x - y).$$

The set of subgradients of f at x is denoted as $\partial f(x)$.

Group-A

(AQ1) Consider the linear program $\mathcal{P}$ defined in the notations and let $S = \{x \in \mathbb{R}^n : Ax \leq b\}$. Define

$$S' = \{d \in \mathbb{R}^n : \forall x \in S, \forall \lambda \geq 0, x + \lambda d \in S\}.$$

(a) Show that $S' = \{d : Ad \leq 0\}$.

(b) Show that S' is a convex set.

(c) Show that $\mathcal{P}$ is unbounded above if and only if there exists a $d \in S'$ such that $c^T d > 0$.

[3 + 2 + 5 = 10]

(AQ2) Let $\mathbf{x}^*$ and (λ^*, μ^*) (where λ^* and μ^* are the Lagrangian multiplier vectors for the inequality and equality constraints, respectively) be any primal and dual optimal solutions to the mathematical program $\mathcal{MP}$ with zero duality gap.

(a) State and prove the necessary KKT optimality conditions in terms of $\mathbf{x}^*$ and (λ^*, μ^*).

(b) State and prove the conditions under which KKT conditions are sufficient for zero duality gap.

[5 + 5 = 10]

(AQ3) State and prove Slater's strong duality condition for $\mathcal{MP}$. [2 + 8 = 10]

(AQ4) Let $C \subseteq \mathbb{R}^n$ be a convex set, and $f : C \to \mathbb{R}$.

(a) Show that if $\forall x \in C, \partial f(x) \neq \emptyset$ then f is a convex function.

(b) Show that if f is a convex function then for any $x \in int(C)$ (interior of C), $\partial f(x) \neq \emptyset$.

(c) Show that if f is differentiable at x then $\nabla f(x) \in \partial f(x)$.

[3 + 5 + 2 = 10]

(AQ5) (a) Consider the following mathematical program in $\mathbb{R}^n$:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x \in \Omega \end{array} \qquad (1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function and Ω is a convex differentiable set. Then $x^* \in \Omega$ is an optimal point for the above mathematical program if and only if for all $x \in \Omega$ we have

$$\nabla f(x^*)^T (x - x^*) \geq 0.$$

(b) Consider the following mathematical program in $\mathbb{R}^n$:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{2}$$

where f is a differentiable convex function and $A \in \mathbb{R}^{m\times n}$. Show that a point $x^* \in \mathbb{R}^n$ is optimal for the above mathematical program if and only if x^* is feasible and there exists $\mu^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) = A^T \mu^*.$$

[7 + 3 = 10]

Group-B

(BQ1) Given a set I of intervals $[a_i, b_i] \subset \mathbb{R}$ for each $1 \leq i \leq n$ and a weight function w on intervals. A subset J of I is a *k-interval packing* if any point in $\mathbb{R}$ is contained in at most k intervals from J. Consider the following combinatorial optimization problem $\mathcal{CP}$:

$$\begin{aligned} \text{Maximize} \quad & w(J) \\ \text{subject to} \quad & J \subseteq I \\ & J \text{ is a } k\text{-interval packing} \end{aligned} \tag{3}$$

where $w(J) = \sum_{\sigma \in J} w(\sigma)$.

(a) Formulate $\mathcal{CP}$ as an integer linear program.

(b) Show that only $n-1$ constraints are needed to be imposed in the above integer linear program apart from the constraints on the variables, i.e., constraints of the form $x_i \in \{0, 1\}$.

(c) Show that the linear programming relaxation of the above integer linear program is integral.

[2 + 5 + 8 = 15]

(BQ2) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strongly convex function, and there exists $x^* \in dom(f)$ such that $f(x^*) = \min_{x \in dom(f)} f(x)$. Also, let $x_0 \in dom(f)$, and define $S = \{x \in dom(f) \mid f(x) \leq f(x_0)\}$.

Assume that the set S and the function f satisfy the following properties: there exist m and M, such that $0 < m < M$ and for all $x \in S$, $m\, I_n \preceq \nabla^2 f(x) \preceq M\, I_n$.

(a) Show that for all $x \in S$, we have

$$f(x) - f(x^*) \leq \frac{\|\nabla f(x)\|^2}{2m}.$$

(b) Show that for all $x \in S$, we have

$$\|x - x^*\| \leq \frac{2\|\nabla f(x)\|}{m}.$$

(c) Define

$$W_{min}(S) := \inf_{\|q\|=1} \left(\sup_{x \in S} q^T x - \inf_{x \in S} q^T x \right)$$

and

$$W_{max}(S) := \sup_{\|q\|=1} \left(\sup_{x \in S} q^T x - \inf_{x \in S} q^T x \right).$$

Show that

$$\frac{W_{max}(S)}{W_{min}(S)} \leq \sqrt{\frac{M}{m}}.$$

[4 + 4 + 7 = 15]

(BQ3) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strongly convex function. Also, let $x_0 \in dom(f)$, and define $S = \{x \in dom(f) \mid f(x) \leq f(x_0)\}$.

The set S and the function f satisfy the following properties:

- There exist m and M, such that $0 < m < M$ and for all $x \in S$, $m\, I_n \preceq \nabla^2 f(x) \preceq M\, I_n$.
- There exists $L > 0$ such that for all $x, y \in S$, $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$.
- There exists $\alpha \in (0, 1/2)$, such that for all $x \in S$, we have

$$\|\nabla f(x)\|_2 < \eta, \quad \text{where } \eta = \min\left\{ \frac{m^2}{L}, 3(1 - 2\alpha)\frac{m^2}{L} \right\}.$$

We are interested in minimizing $f(x)$ with the starting point $x_0 \in dom(f)$.

(a) Describe Newton's Method for minimizing $f(x)$ with the starting point x_0.

(b) Show that Newton's Method is a valid descent method.

(c) Compute the convergence rate of Newton's Method, with starting point x_0, to minimize $f(x)$. Assume that the Backtracking subroutine, inside the Newton's Method, uses parameters (α, β) where $\beta \in (0, 1)$.

[2 + 3 + 10 = 15]

(BQ4) Let $\mathcal{F}$ be a finite family of halfspaces in $\mathbb{R}^n$, $|\mathcal{F}| > n$, such that any $n + 1$ halfspaces in $\mathcal{F}$ have a non-empty intersection. Using Strong Duality Theorem of Linear Programs show that $\cap_{H \in \mathcal{F}} H \neq \emptyset$. [15]

(BQ5) Given a closed convex set $C \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, the *projection problem* $\Pi_C(x)$ is defined as $\Pi_C(x) = \text{argmin}_{z \in C} \|z - x\|$. Also, assume that there exists an oracle $\mathcal{O}_\Pi$ that can solve the projection problem for any compact convex set, given as a system of convex inequalities, in any dimension.

Consider the following mathematical program $\mathcal{MP}1$ in $\mathbb{R}^n$:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega \end{aligned} \tag{4}$$

where $f : \Omega \to [0, B]$ is a convex function and Ω is a compact convex set with diameter R. Note that $x \in \Omega$ if

$$g_i(x) \leq 0, \quad \forall i \in \{1, \ldots, m\}$$

where functions g_i, for all $i \in \{1, \ldots, m\}$, are convex functions. We have access to f via an oracle $\mathcal{O}_f$ that given any $x \in \mathbb{R}^n$ returns the value of f at x.

(a) Design an algorithm for $\mathcal{MP}1$ along the lines of the projected descent method assuming that we are given $x_0 \in \Omega$, the functions g_i, and the value of B as inputs, and we also have access to the oracles $\mathcal{O}_f$ and $\mathcal{O}_\Pi$.

(b) Derive the convergence analysis of the algorithm in terms of number of steps taken, B and R.

[9 + 6 = 15]

Indian Statistical Institute
Semester Examination
M. Stat I Yr. 2018-2019
Sample Survey and Design of Experiments
Full Marks. 80

Date : 24/04/2019 Time : 2.30-6.00 p.m.

Note: Use separate booklet to answer questions of Group A and Group B.

Group A: Sample Surveys. (Total Marks = 30)

Answer any three. Notations are as usual.

1. State Rao, Hartley and Cochran (1962)'s sampling scheme of drawing n units with known positive size measure variable x. Obtain an unbiased estimator of population total Y based on this scheme. Also obtain the variance and variance estimator of that estimator.

   State the condition when that estimator will attain the minimum variance. (10)

2. State and prove Rao (1979)'s general theorem on Mean Squared Error of a homogeneous linear estimator of Y.

   Use this theorem to obtain the Yates and Grundy's variance and variance estimator of the Horvitz and Thompson's estimator of population total Y.

   (10)

3. (a) Define IPPS scheme. State Durbin (1967)'s method of unequal probability sampling scheme for sample size n = 2 and show that it is an IPPS scheme.

   (b) Show how Politz and Simmon's 'At-home-probability' technique can be used to estimate the population mean $\bar{Y}$ under SRSWR, and estimate its error.

   $(5 + 5 = 10)$

4. State how the double sampling approach can be useful in stratified random sampling for the estimation of population mean $\bar{Y}$, when the strata are not pre-formed well. Write down the estimator and examine whether it is biased or unbiased. Also derive its MSE.

   (10)

5. Describe how the population mean of a sensitive continuous quantitative variable can be estimated unbiasedly by a SRSWR of respondents selected in n draws. Obtain the variance of this estimator and variance estimator.

   (10)

Please Turn Over

Group B: Design of Experiments
Total Marks: 50
Answer all questions.

1. Prove or disprove the following statements.

   (a) A balanced orthogonal block design is necessarily a proper block design.

   (b) A balanced incomplete block design (BIBD) always exists with parameters $b = v = 13, r = k = 9$ and $\lambda = 5$ (the notations have their usual meanings).

   (c) A Hadamard matrix of order n exists only if n is a multiple of 4 and such a matrix of order 2^t for some positive integer t can always be constructed. [8+8+(6+4) =26]

2. (a) Let the information matrices of the treatment effects for a row column design d_1 and a block design d_2 obtained from d_1 treating the rows as blocks and ignoring the columns, be denoted by C_{d_1} and C_{d_2} respectively. Show that $C_{d_1} \leq C_{d_2}$. You can start with the usual form of the information matrices, clearly explaining your notations.

   (b) Starting with the following block design with 7 treatments in 7 blocks of size 4 each, construct a suitable row-column design in 7 rows and 4 columns such that the information matrix of the treatment effects for the newly constructed row-column design is the same as the C- matrix of the original block design. Give reasons for the coincidence of these two information matrices.

0	3	5	6
1	4	6	0
5	0	1	2
3	6	1	2
5	1	3	4
6	2	4	5
0	4	2	3

[5+7=12]

3. Construct a $\frac{1}{4}(2^5)$ fractional factorial design of Resolution III. Write down the defining relations and the complete alias set. [6+6=12]

INDIAN STATISTICAL INSTITUTE
Final-Semester Examination: 2018 − 19

M. Stat 1st Year

Measure Theoretic Probability

Date: 27/4/19 **Maximum Marks: 60** **Duration:3 hr**

(1) If a random variable X admits a density with respect to the Lebesgue measure, then show that $\lim_{u\to\pm\infty} \phi_X(u) = 0$ where $\phi_X(u)$ is the characteristic function of X. [10]

(2) Let $(X_i)_{i\geq 1}$ be i.i.d Exp(λ) with $\lambda \in (0, \infty)$. Find the radius of convergence of the power series $\sum_{n=0}^{\infty} X_n z^n$, $z \in \mathbb{C}$. [10]

(3) (a) Let $f : [a, b] \to \mathbb{R}$ be a bounded Lebesgue measurable function. Consider $G_f = \{(x, f(x)) : x \in [a, b]\}$ to be the graph of the function f. Show that G_f is measurable with respect to the product Lebesgue σ-algebra and G_f has 2-dimensional Lebesgue measure 0. [6]

(b) Show that for a distribution function F,

$$\int_{\mathbb{R}} (F(x + 1) - F(x))dx = 1.$$

[4]

(4) Suppose that f, g are densities of a measure μ on $\mathbb{R}$, that is,

$$\int_A f(x)dx = \int_A g(x)dx = \mu(A) \text{ for all } A \in \mathcal{B}(\mathbb{R}).$$

If furthermore, f and g are continuous functions, then show that

$$f(x) = g(x) \text{ for all } x.$$

Is the assertion above true without the continuity? give reasons. [10]

(5) Let $(X_j)_{j\geq 1}$ be an i.i.d. sequence of random variables such that $S_n = \sum_{j=1}^{n} X_j$ satisfies $S_n/n \to Y$ almost surely for some real valued random variable Y and suppose $\mathsf{E}[X_1]$ is well defined. Show that $\mathsf{E}[|X_1|] < \infty$ and $Y = E[X_1]$ almost surely. [10]

(6) Let $(X_j)_{j\geq 1}$ be an i.i.d. sequence with mean zero and unit variance. Let $S_n = X_1 + \cdots X_n$ and $T_n = S_1 + \cdots + S_n$. Find a_n and b_n such that

$$\frac{T_n - b_n}{a_n} \xrightarrow{d} N(0,1).$$

[10]

INDIAN STATISTICAL INSTITUTE, KOLKATA

FINAL EXAMINATION: SECOND SEMESTER 2018 - '19

M.STAT I YEAR

Subject : **Metric Topology and Complex Analysis**
Time : 2 hours 30 minutes
Maximum score : 50

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) Show that function $f(z) = |z|^2 - \bar{z}^2$ satisfies the Cauchy-Riemann equations only at the point $z = 0$. Is f analytic at 0?

[4 + 2 = 6 marks]

(2) Evaluate the following two integrals using contour integration.

$$\text{(i)} \int_{-\infty}^{\infty} \frac{\cos x}{x^2 + a^2} dx \text{ (where } a > 0\text{)}, \qquad \text{(ii)} \int_{0}^{\infty} \frac{1}{1 + x^4} dx$$

[10 + 10 = 20 marks]

(3) Let D_r denote the closed disc $|z| \leq r$. Let $f(z)$ be an analytic function on D_1 whose image is contained in D_r for some $r < 1$. Prove that $f(z)$ has a unique fixed point (i.e., a point z_0 such that $f(z_0) = z_0$). (Hint: Apply Rouche's theorem to $f(z)$ and $-z$.)

[8 marks]

(4) Decide whether the following statements are correct. Prove or give counterexamples in each case.

(i) $f(z)$ is an entire function satisfying $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$, then $f(z + 2\pi) = f(z)$ for all $z \in \mathbb{C}$.

(ii) Let $f(z)$ be an entire function satisfying the equation $f(z) = f(1/z)$ for all $z \neq 0$. Then $f(z)$ is a constant function.

(iii) The polynomial $p(z) = 4z^5 - z + 2$ has at least one zero outside the open unit disc.

[5 + 5 + 5 = 15 marks]

(5) Find the singularities of the function $f(z) = 1/(z^2 + 1)^2$, and at every point of singularity determine its nature: if it is a removable singularity, redefine the function at that point so that the function becomes analytic; if it is a pole, find the order and residue.

[6 marks]

Indian Statistical Institute
Second Semestral Examination
2018-2019 Academic Year
M.Stat. First Year
Large Sample Statistical Methods

Date : 03.05.2019 Maximum Marks: 60 Duration :- 3 hours

Answer as many questions as you can. The maximum you can score is 60.

1. Suppose $X_1, \ldots, X_n$ are iid N(0,1). Find the asymptotic distribution of
$$\frac{n(X_1X_2 + X_3X_4 + \ldots + X_{2n-1}X_{2n})^2}{(\sum_{i=1}^{2n} X_i^2)^2}.$$
Prove your assertion. [6]

2. Suppose $X_1, X_2, \ldots$ are iid Poisson(λ), $\lambda > 0$. Find suitable constants a_n and b_n (allowed to be dependent on λ also) such that $a_n(Y_n - b_n)$ converges to a non-degenerate limit where $Y_n = (1 - \frac{1}{n})^{n\bar{X}_n}$ and $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$, for $n \geq 1$. Prove your answer. [8]

3. Suppose you have i.i.d. observations from a distribution having a unique median. Prove that the (smallest) sample median converges to the population median almost surely. [8]

4. (a) Give an example of an inconsistent maximum likelihood estimator and prove that it is indeed inconsistent. [5]

   (b) 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 $\theta = \theta_0$. [7]

5. (a) Listing appropriate assumptions, state a result on asymptotic normality of one-sample U-statistic. [2]

   (b) Suppose $X_1, \ldots, X_n$ are iid having a continuous distribution which is symmetric about its (unknown) median θ. Derive the asymptotic null distribution of the Wilcoxon Signed-Rank test statistic for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. [7]

   (c) Suppose $X_1, \ldots, X_n$ are iid Bernoulli($\frac{1}{2}$) random variables. Do there exist sequences of constants a_n and b_n such that $a_n(s_n^2 - b_n)$ 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]

6. Suppose $X_1, \ldots, X_n$ are iid having density $f(x, \theta)$, where $\theta \in \mathbf{R}$. Stating appropriate regularity assumptions, derive the asymptotic null distribution of the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where θ_0 is a given constant. [8]

7. (a) Let $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. random variables with distribution function F. Define $w(F) = \sup\{x : F(x) < 1\}$. Show that $X_{(n)}$ converges to $w(F)$ almost surely, where $X_{(n)}$ is the maximum of $\{X_1, \ldots, X_n\}$. [5]

   (b) Suppose $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. observation from a distribution given by $F(x) = \{1 - \frac{1}{\log x}\}I_{x\geq e}$. Does the sample minimum $X_{(1)}$ based on observations $\{X_1, \ldots, X_n\}$ converge to a non-degenerate limit after appropriate centering and scaling ? Prove your assertion. [5]

INDIAN STATISTICAL INSTITUTE
Final-Semester Examination: 2018 – 19

M. Stat 1st Year

Measure Theoretic Probability-Back Paper

Date: 15-07-19 **Total Marks: 100** **Duration:3 hr**

(1) Consider the following σ-fields on $\mathbb{N}$:

$$\mathcal{F}_n := \sigma\left(\{\{1\}, \ldots, \{n\}\}\right), \, n \geq 1\,.$$

Define

$$\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n\,.$$

Show that $\mathcal{F}$ is a field but not a σ-field on $\mathbb{N}$.

[10]

(2) A σ-field $\mathcal{F}$ is countably generated if there exists a countable $\mathcal{A}$ such that $\mathcal{F} = \sigma(\mathcal{A})$. Show that the Borel σ-field is countably generated.

[10]

(3) Let $X_1, X_2, \cdots$ be independent $N(0,1)$ random variables. Prove that

$$\limsup_n \left(\frac{X_n}{\sqrt{2\log n}}\right) = 1 \text{ almost surely.}$$

[15]

(4) Suppose that $\int_0^\infty |f(x)|dx < \infty$. Show that for each $\varepsilon > 0$,

$$\lim_{\alpha\to\infty} \lambda\left\{x > \alpha : |f(x)| > \varepsilon\right\} = 0\,.$$

[12]

(5) If X is a random variable whose distribution function F is continuous, show that

$$\int_{\mathbb{R}} F(x)P(X \in dx) = 1/2\,.$$

[10]

(6) If $X_1, \ldots, X_n$ are i.i.d. from the Cauchy distribution, show that $(X_1 + \ldots + X_n)/n$ also has a Cauchy distribution. Argue that $(X_1 + \ldots + X_n)/n$ does not converge almost surely or in probability to any degenerate distribution, as $n \to \infty$. Why does this not contradict the strong law of large numbers?

[8]

(7) For $\alpha \in (0, \infty)$ and $x \in (0, \infty)$, define $f_\alpha(x) = x^{-\alpha}$. Let $p \in [1, \infty)$. Show carefully that $f_\alpha \in L_p((0,1], dx)$ if and only if $\alpha p < 1$. Show also that $f_\alpha \in L_p([1, \infty), dx)$ if and only if $\alpha p > 1$.

[5+5=10]

(8) Let f be an integrable function on a measure space $(E, \mathcal{E}, \mu)$. Suppose that, for some π-system $\mathcal{A}$ containing E and generating $\mathcal{E}$, we have $\int_A f d\mu = 0$ for all $A \in \mathcal{A}$. Show that $f = 0$ almost everywhere. [10]

(9) Show that if $X_n \geq 0$ and $X_n \downarrow X$ almost surely and $\mathsf{E}[X_k] < \infty$ for some k then $\mathsf{E}[X_n] \to \mathsf{E}[X]$ as $n \to \infty$ [15]

INDIAN STATISTICAL INSTITUTE

M. Stat First Year (2018-19)

Second Semester (Back Paper)

Resampling Techniques

Date: 16.7.19 Marks: 100 Duration: 3 hours

Attempt all questions

(1) Let $T_n = F_n^{-1}(\frac{1}{2})$ be the sample median of an *iid* sample $X_1, \ldots, X_n$, where $F_n^{-1}(t) = \inf\{x : F_n(x) \geq t\}$. Assume that $n = 2m - 1$ for an integer m.

(i) Obtain the distribution of the bootstrap estimator of the median, given $X_1, \ldots, X_n$.

(ii) Obtain the jackknife estimator of the variance of the sample median.

(iii) Under appropriate assumptions show that the above jackknife estimator of variance is inconsistent.

[8+7+10=25]

(2) In the study of income shares or wealth distributions, sometimes we need to estimate the poverty line (or low income cut-off) of the population. For the i-th sampled family, let z_i be the expenditure on "necessities", y_i be total income and x_{it}; $t = 1, \ldots, m$, be variables such as urbanization category and family size. Assume the following model:

$$\log z_i = \gamma_1 + \gamma_2\sqrt{|\log y_i|} + \sum_{t=1}^{m} \beta_t x_{it} + \epsilon_i,$$

where γ_1, γ_2, β_t; $t = 1, \ldots, m$, are unknown parameters. Let γ_0 be the overall proportion of income spent on "necessities". Then let us define the poverty line θ as the solution of

$$\log\left[(\gamma_0 + 0.2)\theta\right] = \gamma_1 + \gamma_2\sqrt{|\log \theta|} + \sum_{t=1}^{m} \beta_t x_{0t},$$

for a particular set of $x_{01}, \ldots, x_{0m}$.

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

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

[10+15=25]

(3) Assume that the functional T is continuously Gateaux differentiable at F with non-zero influence function ϕ_F. Let $\sigma^2 = E[\phi_F(X_1)]^2$. Then show that the jackknife estimator v_{JACK} of the variance of T is strongly consistent, that is,

$$\frac{n v_{JACK}}{\sigma^2} \xrightarrow{a.s.} 1,$$

as n, the sample size, tends to infinity. [25]

(4) (i) Compare and contrast bootstrap and permutation tests for two-sample hypothesis testing problems.

(ii) Propose methods of estimating P-values for bootstrap and permutation tests.

(iii) Give an example of a two-sample testing problem where permutation test is inapplicable but bootstrap test is appropriate. Develop the bootstrap testing procedure.

[5+5+15]

Second Semester
2018-2019 Academic Year
M.Stat. First Year
Large Sample Statistical Methods

num Marks: 100 Duration :- 4 hours

Date: 17.07

Answer all questions

Stating appropriate conditions, prove asymptotic normality of a consistent sequence of roots of likelihood equation. [25]

State and prove the Weak Bahadur Representation of sample quantiles (due to J. K. Ghosh) under iid sampling from a common distribution. [25]

Stating appropriate assumptions, derive in the context of iid sampling from a multinomial population with k classes, the asymptotic null distribution of the chi-square test statistic for the composite hypothesis testing problem $H_0 : \pi_i = \pi_i(\theta),\ i = 1, \ldots, k$, where $\pi_i(\theta),\ i = 1, \ldots, k$ are specified functions of θ, an unknown vector of dimension q, where $q < k - 1$. [30]

Stating appropriate assumptions, prove asymptotic normality of one-sample U-statistic based on iid observations from a common distribution. [20]