Statistical Inference

M-Stat I, First Semester 2017-2018

Mid-Semester Examination

Date: 04.09.2017 Maximum marks: 40 Duration: 2 hours

1. Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson(θ). Suppose the loss function is the scaled squared error loss $L(\theta, a) = (\theta - a)^2/\theta^k$. Find the Bayes estimator for θ with the *gamma*(a, b) prior. [The *gamma*(a, b) density is given by $f(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}, x > 0$.] [8]

2. Consider the decision problem where θ is the parameter of interest. The parameter space as well as the action space are both equal to the real line, and the loss function satisfies
$$L(\theta, a) = \begin{cases} k_1|\theta - a| & \text{if } a \leq \theta \\ k_2|\theta - a| & \text{if } a > \theta \end{cases}$$
where $k_1 > 0$ and $k_2 > 0$. Assuming any conditions that may be necessary, show that the Bayes rule is the p-th quantile of the posterior distribution of θ, where p is a suitable function of k_1 and k_2. [10]

3. (a) Define an extended Bayes rule, and give an example.
   (b) Show that if δ is an equalizer rule (a rule which has constant risk) and is an extended Bayes rule, then it is also a minimax rule.
   (c) Show with an example than when a minimax rule is an equalizer rule and is extended Bayes, the corresponding least favourable distribution may not be a proper prior.

   [3+6+3=12]

4. Let $f(x)$ be a convex real valued function defined on a nonempty convex subset S of $\mathbb{R}^k$. Let Z be a k dimension random variable with finite expectation $E(Z)$, for which $P(Z \in S) = 1$. Then, show that
   (a) $E(Z) \in S$.
   (b) $f(E(Z)) \geq E(f(Z))$.

   [5+7=12]

Indian Statistical Institute

M.Stat. First year, Mid Semester Exam: 2017

Topic: Regression Techniques

Maximum Marks: 50, Duration: 2 hours

DATE: 05/09/2017

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

1. Consider the multiple linear regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$. Show that the estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are independently distributed. Find the corresponding distributions. [4+6]

2. (a) Under the standard multiple linear regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$, show that Variance($\hat{\beta}_{ridge}$) $<$ Variance($\hat{\beta}_{ols}$).
(b) Under the multiple regression model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 A)$, A is positive definite, how will you test the hypothesis $H_0 : C\beta = c$, for some $q \times p$ matrix C with rank q. [5+5]

3. (a) Define False Discovery Rate (FDR) in the context of multiple testing problem. State Benjamini-Hochberg algorithm for controlling FDR.
(b) Suppose data are collected on body weight and blood pressure level (BPL) for 100 patients. Assuming that BPL can be modelled as an unknown function of body weight, how will you predict your BPL if you know that your body weight is 65 unit? [5+5]

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

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

   Using a non-parametric approach of modeling the mean trajectories and a parametric approach of modelling the covariance structure, propose a semi-parametric modelling of the above dataset. Write down the joint likelihood function, and explain the parameter estimation method clearly. [10]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : Semester I (2017-2018)

M. Stat 1st Year

Multivariate Analysis

Date: 6. 9. 17 Maximum marks: 60 Time: 2 hours.

Note: Answer all questions. Maximum you can score is 60. If you use any result proved in class, you have to state the result clearly. To get full credit, your arguments have to be complete.

Part - A

1. Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent data matrices having n_1 and n_2 rows respectively from the same $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, $\boldsymbol{\Sigma}$ being positive definite.

   Find the sampling distribution of D^2, where D is the sample Mahalanobis distance between the two samples . [15]

2. Suppose n independent samples are available from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and it is known that $\boldsymbol{\mu} = k\boldsymbol{\mu}_0$, where k and $\boldsymbol{\Sigma}$ are unknown but $\boldsymbol{\mu}_0$ is a known fixed p-dimensional vector. Find the MLE of k, and derive its mean and variance. [7+6]

3. Derive a likelihood ratio test for the hypothesis $\boldsymbol{\mu}'\boldsymbol{\mu} = 1$, based on one observation $\mathbf{x}$ from $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where σ^2 is known. [12]

4. Suppose $\mathbf{M}$ is a $p \times p$ matrix, distributed as $W_p(\Sigma, n)$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

   $\boldsymbol{\Sigma}_{11}$ being a $p_1 \times p_1$ matrix.

   Using the results on partitioned wishart matrices, develop a Union Intersection Test based on F-ratio for the null hypothesis $\boldsymbol{\Sigma}_{12} = 0$. [15]

Part - B

Assignments [10]

INDIAN STATISTICAL INSTITUTE

Mid-semester exam. (Semester I: 2017-2018)

Course Name: M. Stat. 1st year

Subject Name: Categorical Data Analysis

Date: 7, 9 , 2017, Maximum Marks: 40. Duration: 2 hrs.

Note: Answer all questions.

1. Let X and Y be identically distributed categorical random variables. Suppose both X and Y have k categories $C_1, \ldots, C_k$ with probabilities $p_1, \ldots, p_k$, and joint distribution of X and Y is defined by the conditional probabilities

$$P(Y = C_j | X = C_i) = (1 - \alpha)p_j + \alpha I(i = j), \quad i, j = 1, \ldots, k,$$

where $I(i = j)$ is the indicator and $0 \leq \alpha \leq 1$. Find $\tau_{Y|X}$, Goodman and Kruskal's τ, as a function of α. [10]

2. Find the sample values of the measure of association Goodman-Kruskal's τ for the following two tables.

1	3	10	6	7
2	3	10	7	6
1	6	14	12	5
0	1	9	11	3

1	6	14	12	5
0	1	9	11	3
1	3	10	6	7
2	3	10	7	6

[6]

3. (a) Derive the joint asymptotic distribution of log odds ratios in a 2×4 contingency table. [10]

(b) Find the P-value of the test for independence for the following table by Fisher's conditional test procedures against one-sided alternative. Discuss three possible ways to carry out a two-sided test procedure in this context.

	Guess poured first	
Poured first	Milk	Tea
Milk	4	1
Tea	1	4

[5+9]

M.Stat. I / Stochastic Processes
Midsem. Exam. / Semester I 2017-18
Date - September 08, 2017 / Time - 2 hours
Maximum Score - 30

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

1. Passengers arrive in a busy bus station following a Poisson process with intensity parameter α. Each passenger independently of the others wait for a random time, that follows Exp(β) distribution, to get a bus for destination. Let $\{N(t)\}$ be the number of passengers arrives during $(0,t]$, $\{X(t)\}$ be the number of passengers still waiting for a bus at time t and $\{Y(t)\}$ be the number of passengers at time t who have left the bus station after getting a bus,

   (a) (6 marks) Find the conditional probability that a passenger is waiting at time t given that s/he arrived during $(0,t]$. Calculate $P(X(t) = k \mid N(t) = n)$ and consequently find the distribution of $X(t)$.

   (b) (3 marks) Given $X(0) = i$, find the limiting distribution of $X(t)$, as $t \to \infty$, i.e., find $\lim_{t\to\infty} P(X(t) = j \mid X(0) = i)$.

   (c) (2 marks) What is the $\lim_{t\to\infty} P(Y(t) = j \mid Y(0) = k)$?.

2. In a system, each particle splits into two (with probability $0 < p < 1$) or disappear (with probability $q = 1 - p$) after a random amount of time, say T, that follows Exp(α) distribution, independently of others in the system. Assume that there were $k(> 0)$ many particles in the system initially. Let $\{X(t)\}_{t\geq 0}$ be the number of particles in the system at time t.

   (a) (3 marks) Calculate the rate matrix Q for the process.

   (b) (4 marks) Find the probability of extinction.

   (c) (4 marks) Let $\{Y(t)\}_{t\geq 0}$ be the number new particle in the system arriving according to a Poisson process with intensity parameter $\nu(> 0)$. Given that there are i particles in the system initially, find the probability that a new arrival happens before any particle in the system splits or disappear and also find the probability that a particle splits into two, before any new arrival.

3. Arrival of customers in a big supermarket (open for 24 hour) is assumed to follow a Poisson process with intensity parameter λ_1 during 9am to 9pm and $\lambda_2(< \lambda_1)$ during 9pm to 9am.

P.T.O

Assume that after arrival each customer (during 9am to 9pm) decides to buy a product with probability $0 < p_1 < 1$, but the night customers (who come are more of a buyer than window shopper) buy a product with probability $1 > p_2 (> p_1)$ independently of others.

(a) (4 marks) Find the distribution of the customers who buy a product.

(b) (3 marks) In a 24 hour cycle (from 9am to 9am in the following day), calculate the probability that the number of customers who buy a product during 9am to 9pm is larger than the number of customers who buy a product during 9pm to 9am.

(c) (4 marks) Given that the price of products in the supermarket follow a Uniform (α, β) distribution, for some $0 < \alpha < \beta$. Find the expected amount received from the product sale during 9am to 9pm and the same from 9pm to 9am. Find their variances.

All the best.

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

Statistical Inference I, M-Stat I

Date: 20.11.2017 Maximum marks: 100 Duration: 3 hours.

[Answer as many as you can. Unless specifically asked for, any result proved in class may be used here (without proof) with a proper statement of the result.]

1. Suppose that X follows a binomial distribution with parameters n and θ, where $\theta \in \Theta = (0, 1)$ and n is known.

   (a) Consider the $beta(\alpha, \beta)$ prior, and the squared error loss $L(\theta, \delta) = (\delta - \theta)^2$. Show that the maximum likelihood estimate x/n is not a Bayes rule.

   (b) Now consider the loss function $L(\theta, \delta) = (\delta - \theta)^2/[\theta(1 - \theta)]$. Show that the maximum likelihood estimate x/n is a Bayes rule with respect to the uniform prior on (0, 1).

   [9+9=18]

2. If for a given decision problem (Θ, D, R), with finite $\Theta = \{\theta_1, \ldots, \theta_k\}$, the risk set S is bounded below, then prove that

$$\inf_{\delta \in D^*} \sup_{\tau \in \Theta^*} r(\tau, \delta) = \sup_{\tau \in \Theta^*} \inf_{\delta \in D^*} r(\tau, \delta) = V,$$

   and there exists a least favorable distribution τ_0. Moreover, if S is closed from below, prove that there exists an admissible minimax decision rule δ_0, and δ_0 is Bayes with respect to τ_0. [Here Θ^* represents the space of prior distributions and V represents the value of the game. Also $r(\tau, \delta) = \int R(\theta, \delta) d\tau(\theta)$. All other symbols have their usual meanings.]. [18]

3. Suppose that $X_1, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, both parameters unknown. Derive the uniformly most powerful unbiased test for testing the hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. [18]

4. Consider the hypothesis testing problem for the parameter θ under the exponential family model. Suppose we are testing an appropriately defined null hypothesis H_0 against a suitable alternative H_1.

   (a) Define what is meant by a similar test.

   (b) Let ω be the boundary of the null and the alternative hypothesis, and let T be a statistic which is sufficient for $\theta \in \omega$. Define what is meant by a test having Neyman structure with respect to the statistic T.

(c) Prove that every similar test has Neyman structure if and only if T is boundedly complete.

3+3+12=18]

5. For the one parameter exponential family model with parameter θ, consider testing the null hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. Suppose that the family has monotone likelihood ratio in X. Let a_1 and a_0 represent the actions of rejecting the null, and failing to do so, respectively. Consider the one sided tests of the form

$$\phi(x) \begin{cases} 1 & \text{if } x > x_0 \\ \gamma & \text{if } x = x_0 \\ 0 & \text{if } x < x_0. \end{cases} \tag{1}$$

Suppose that the loss function L satisfies

$$\begin{aligned} L(\theta, a_1) - L(\theta, a_0) \geq 0 \quad \text{for } \theta \leq \theta_0 \\ L(\theta, a_1) - L(\theta, a_0) < 0 \quad \text{for } \theta \leq \theta_0. \end{aligned}$$

Show that the one sided tests defined in (1) form an essentially complete class. Show that under additional conditions (to be stated by you), any test of the form (1) is admissible.

[18]

6. Suppose X has a multiparameter exponential family distribution with natural parameters $(\theta_1, \ldots, \theta_k)$. We are interested in testing $H_0 : \theta_1 \leq 0$ versus $H_1 : \theta_1 > 0$, treating the other $(k-1)$ components to be nuisance parameters. Describe how you can develop a UMPU test for the above hypothesis based on a random sample $(X_1, \ldots, X_n)$.

[18]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M. STAT. 1st year

Subject Name: Categorical Data Analysis

Date: 22, 11, 2017, Maximum Marks: 60. Duration: 3 hrs.

Group A: Answer all questions.

1. Geometrically interpret *Odds Ratio* $= 4$ for 2×2 contingency tables. [15]

2. Discuss the latent variable approach to model ordinal categorical variables with possible covariates. How can you model a bivariate response vector using latent variable approach where both the variables are ordinal categorical? Write down the likelihood in this context. How the model and likelihood would change if a third response vector is added which is continuous in nature? [4+3+3+6=16]

3. Each week *Variety* magazine summarizes reviews of new movies by critics in several cities. Each review is categorized as pro, con, or mixed, according to whether the overall evaluation is positive, negative, or a mixture of the two. The following Table summarizes the ratings from April 1995 through September 1996 for Chicago film critics Gene Siskel and Roger Ebert. Discuss kappa and generalized kappa as measures of agreement in this context. Find the values (setting appropriate weights for generalized kappa) in this case, and interpret the results.

	Ebert		
Siskel	Con	Mixed	Pro
Con	24	8	13
Mixed	8	13	11
Pro	10	9	64

[5+5+3=13]

4. Define and interpret *thinning operator* for time series of count data. Suppose Y_1 and Y_2 are two zero-inflated Poisson random variables, where the Poisson part of Y_2 is obtained from the Poisson part of Y_1 by using the *thinning operator*. Find $\text{corr}(Y_1, Y_2)$.

   Discuss how negative binomial margin can be constructed by using the *thinning operator*.

[4+6+6=16]

Indian Statistical Institute

M.Stat. First year, First Semestral Exam: 2017-18

Regression Techniques

Maximum Marks: 50, Duration: 3 hours

24·11·17

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

1. (a) Consider the standard multiple regression model (with p predictors) $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$. Find $\hat{\sigma}^2_M$, the MLE of σ^2. Suppose that $\frac{p}{n} \to c$. Show that $\hat{\sigma}^2_M \xrightarrow{P} (1-c)\sigma^2$. [2+3]

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

2. Consider the setting of toxicology studies where we want to model the probability of a response (death) to different doses of a toxin. Suppose each subject has a tolerance T for a dose x. That is, if $T \leq x$, then the subject dies. Define Y to be the indicator of death ($Y = 1$ corresponds to the subject dying).

   Assume the distribution of tolerances in the population follows an extreme value (or Gumbel) distribution with cdf $F(t) = \exp(-\exp\{-\frac{(t-a)}{b}\})$ with mean $a + .577b$ and standard deviation $\frac{\pi b}{\sqrt{6}}$.

   (a) For a given dose x, derive the probability a randomly selected subject dies, i.e.,$\pi(x) = P(Y = 1|x)$ as a function of the mean and variance of the tolerance distribution.
   (b) Derive the link function g for the regression of the binary response Y on the co-variate dose (x), such that $g(\pi(x))$ has the following form: $g(\pi(x)) = \alpha + \beta x$. How do the parameters α and β relate to a and b?
   (c) Does $\pi(x)$ approach one at the same rate that it approaches zero? Explain.
   (d) We define $LD50$ as the dose x_0 such that $\pi(x_0) = 0.5$. Derive $LD50$ as a function of the regression parameters, α and β. [5+5+2+3]

3. (a) What is the difference between "intermittent missingness" and "monotone missingness" for longitudinal responses?
   (b) Propose a step by step algorithm for imputing the monotone missing values under

the available case missing value (ACMV) restriction. Is this restriction more powerful than the complete-case missing value (CCMV) restriction? Explain. [3+7]

4. Financial regulations require banks to report their daily risk measures called Value at Risk (VAR). Let Y be the financial return of the bank and then for a given θ $(0 < \theta < 1)$, the VAR is the value y^* satisfying $P(Y \leq y^*) = \theta$. The financial return depends on exchange rate (x). Based on a sample of n data points (Y_i, x_i), we need a suitable statistical approach for estimating VAR at θ=0.85. Specify explicitly:
(i) appropriate statistical model, (ii) the estimation method, and (iii) limitations of the model (and method) used. [3+5+2]

5. Define the following terms:
(i) missing completely at random, (ii) missing at random, (iii) missing not at random, (iv) irregular longitudinal data, (v) weighted false discovery rate. [5]

M.Stat. I / Stochastic Processes

Final Exam. / Semester I 2017-18

Date - November 28, 2017 / Time - 3 hours

Maximum Score - 50

NOTE : THE PAPER HAS QUESTIONS WORTH 55 MARKS. SHOW ALL YOUR WORK TO GET THE FULL CREDIT. RESULTS USED MUST BE CLEARLY STATED.

1. Suppose a machine has 4 components that run independently till a random time that follows $Exp(\lambda)$ distribution. The machine fails when all the components fail. When a component is down, the repair can take a random time that follows $Exp(\mu)$ distribution. Let X_t be the no. of component(s) working at time t.

   (a) (3 marks) For this continuous parameter Markov chain, give its state space and find its Q matrix.

   (b) (4 marks) Find the long run probability that the machine is operating (i.e., the proportion of time it is operating) in terms of μ and λ.

   (c) (3 marks) Determine the expected time between two consequtive breakdowns of the machine.

   (d) (4 marks) If there are n components determine the expected time of consequtive breakdowns.

2. (a) An individual with a highly contagious disease enters a population. During each subsequent period, the carrier will either infect $k(\geq 1)$ new persons with probability $f_k = dp^{k-1}$, or, infect no individual with probability $f_0 = 1 - d/(1-p)$, before being removed (at the end of that period), where $0 < p,\ d/(1-p) < 1$. An infected individual spreads the infection during subsequent time unit according to the same distribution (independently of other infected individual in the population) and is always removed at the end of the period. Let X_n be number of infected individual at time n.

   i. (3 marks) Thinking this as a Branching process, write down the generating function, $\phi(s)$, of the progeny distribution. Find the relation between d and p, when $m \equiv E(\text{Number of progeny}) = 1$.

   ii. (4 marks) For $m = 1$, determine the probability that at the nth generation there would be no infectious people in the population.

   (b) Consider another scenario, where the infected individual, in each subsequent period, either infect a new person with probability 1/2, or be discovered and removed by public health officials. An infected individual is discovered and removed with probability 1/2 at each unit

P.T.O

of time independent of others. Thus the progeny distribution, in this case, is, $f_2 = 1/2$ and $f_0 = 1/2$. An unremoved infected individual again spread the infection following the same distribution as before, in each time unit (independently of other remaining infected individual in the population).

i. (3 marks) Show that the generating function of the infected patients (from the first individual) in part(a) for $d = 1/4$ and $p = 1/2$ is bigger than or equal to the generating function of the infected patients (from the first individual) in part(b) over $[0, 1]$.

ii. (4 marks) Would the expected time for containing the infection be finite or infinite for the scenario in part (b)? Justfy your answer.

3. Arrival of customers in a big supermarket (open for 24 hour) is assumed to follow a Poisson process with intensity parameter λ_1 (per hour) during 9am to 9pm and $\lambda_2(< \lambda_1)$ (per hour) during 9pm to 9am. Assume that after arrival each customer (during 9am to 9pm) decides to buy a product with probability $0 < p_1 < 1$, but the night customers (who come are more of a buyer than window shopper) buy a product with probability $1 > p_2(> p_1)$ independently of others.

   (a) (4 marks) Find the distribution of the customers who buy a product.

   (b) (5 marks) In a 24 hour cycle (from 9am to 9am in the following day), calculate the probability $P(T_i^D + 12 < T_j^N)$, where T_i^D = time of the ith customer who buy a product during 9am to 9pm, and T_j^N = time of the jth customer who buy a product during 9pm to 9am.

   (c) (4 marks) Given that the price of products in the supermarket follow a Gamma(α, β) distribution (for some $\alpha, \beta > 0$), independently of the arrival process, find the expected amount received from the product sale during 9am to 9pm and the same from 9pm to 9am. Find their variances.

4. A restaurant has a budget $c_1(> 0)$ unit of money per day for maintenance. The owner of the restaurant learnt that there would be check up by the health safety officials at a random interval, whose mean is 30 days, and they would fine a random amount F, that follows $Uniform(\alpha_0, \alpha_1)$, $0 < \alpha_0 < \alpha_1$, if they find a fault in the maintenance. The probability of finding a fault during health safety check up is $p \in (0, 1)$, independent of any other visit. The owner decided to reduce the daily maintenance cost to $c_2(< c_1)$ to cope up with this extra cost. However, s/he is aware if c_2 is too low then the s/he would have to pay the fine more frequently, i.e., the probability p of finding a fault during a health safety check up would increase.

   (a) (4 marks) Find the expected time (in terms of p) until the first violation (or fault) detected by the health officials.

   (b) (5 marks) Find the long run cost of the maintenance for the owner (in terms of the given parameters.

   (c) (5 marks) Let $p = \exp\{-kc_2\}$ for some $k > 0$. Find the optimal value (or range) of c_2 (that minimises the overall maintenance cost of the owner) in terms of the given parameters.

All the best.

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : (2017-2018)

M. Stat 1st Year

Multivariate Analysis

Date: 1. 12. 17 Maximum marks: 100 Time: 3 hours.

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

Part - A

1. Suppose $\mathbf{X}$ and $\mathbf{Y}$ are i.i.d. $N_n(0, \mathbf{I})$ random vectors. Find the distribution of $\frac{\mathbf{X}'\mathbf{Y}}{\mathbf{X}'\mathbf{X}}$. [10]

2. Consider the IRIS data for the Iris Versicolour population only. There are four measurements: sepal length, sepal width, petal length and petal width. Assume that the measurements follow $N_4(\boldsymbol{\mu}, \Sigma)$ distribution. Suppose we want to test whether the mean of sepal length is equal to the mean of petal length for this population.

   (a) Find the m.l.e. of $\boldsymbol{\mu}$, the population mean vector for these four measurements under the null hypothesis. [10]

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

3. In a two-class classification problem with two features, the density of the observations from the first class is $N_2(\mathbf{0}, \mathbf{I})$ while that for the second class is $N_2(\mathbf{0}, 4\,\mathbf{I})$.

   (a) Assuming equal prior probabilities and costs of misclassification, find the Bayes optimal error for this problem. [10]

   (b) If the prior probabilities are unknown, give a rough sketch of the boundary produced by any admissible rule. [4]

   (c) What will be the form of the minimax classification rule? [6]

4. Suppose $\mathbf{X}$ is a random vector having $N_3(0, \Sigma)$ distribution where

$$\Sigma = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

   Find the principal components, and the proportion of total variance explained by the first two of them. [15]

[P. T. O.]

5. (a) Describe the Principal Factor Analysis method. [10]

(b) Suppose in this method, all the initial estimates of the specific variances are given by the same constant c. Derive the estimated factor loadings in this case. [10]

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

Part - B

Assignments [10]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : (2017-2018)

M. Stat 1st Year: First Semester

Multivariate Analysis

Date: 29. 12. 2017 Maximum marks: 100 Time: 3 hours.

1. If $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the mean and covariance of a sample of size $n(> p)$ from $N_p(\boldsymbol{\mu}, \Sigma)$ where Σ is p.d., find the distribution of $(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$. [15]

2. Consider a random vector X which has the $N_p(\boldsymbol{\mu}, \Sigma)$ distribution where $\boldsymbol{\mu}$, Σ (p.d.) are unknown. The problem is to test the hypothesis that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ (known) against $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$ on the basis of a random sample.

   (a) Derive the Likelihood ratio test statistic for this test. [10]

   (b) Derive the Union-Intersection test statistic for the same test, and compare it with the LRT statistic found in part (a). [15]

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

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

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

4. (a) Describe the k-factor model and how the related parameters are estimated in Principal Factor Analysis. [12]

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

5. Describe the three stages of the Profile Analysis in Multivariate Analysis of Variance, and derive the related test statistics. [15]

M.Stat. I / Stochastic Processes

Back Paper Exam. / Semester I 2017-18

Time - 3 hours

Maximum Score - 45

04.01.18

NOTE : THE PAPER HAS QUESTIONS WORTH 100 MARKS. SHOW ALL YOUR WORK TO GET THE FULL CREDIT. RESULTS USED MUST BE CLEARLY STATED.

1. Let $\{X(t)\}$ be a continuous parameter Markov chain on the set of nonnegative integers, say S.

   (a) (6 marks) Assume $X(0) = i$. Show that T_i follows Exponential distribution with some nonnegative parameter, say q_i, where $T_i = \inf\{t > 0 : X(t) \neq X(0)\}$. Note, $q_i = 0$ if and only if i is an absorbing state.

   (b) For this Markov chain, the transition probability, $(p_{ij}(t))$ always has limit at infinity for each $i, j \in S$, i.e., $\lim_{t\to\infty} p_{ij}(t)$ exists, and say it is π_{ij} for each $i, j \in S$.

      i. (4 marks) Show that, for any $i \in S$, $0 \leq \sum_{j\in S} \pi_{ij} \leq 1$.

      ii. (8 marks) Further show that, for any $t > 0$, $\sum_{j\in S} \pi_{ij} p_{jk}(t) = \pi_{ik} = \sum_{j\in S} p_{ij}(t)\pi_{jk}$ for any $i, k \in S$.

   (c) (7 marks) If $\{X(t)\}$ is a linear birth process, show that it is always conservative (i.e., non-explosive), which means, $P(X(t) < \infty)$, for any finite $t \geq 0$.

2. In a system, each particle splits into two (with probability $0 < p < 1$) or disappear (with probability $q = 1 - p$) after a random amount of time, say T, that follows Exp(α) distribution, independently of others in the system. Assume that there were $k(> 0)$ many particles in the system initially. Let $\{X(t)\}_{t\geq 0}$ be the number of particles in the system at time t.

   (a) (5 marks) Thinking this as a Branching process calculate the generating function for of the Progeny distribution. Also find the Q matrix for this Branching process.

   (b) (8 marks) Find the probability of extinction. Find the expected value of $X(t)$. If $p = 1/3$, would the expected time to extinction be finite? Give a brief argument, but no proof needed.

   (c) Let $\{Y(t)\}_{t\geq 0}$ be the number of new particle in the system arriving according to a Poisson process with intensity parameter $\nu(> 0)$. After the arrival the new particle behaves exactly the same way as that of the original particles described above.

P.T.O

i. (5 marks) Find the rate matrix Q for the process with immigration $\{Y(t)\}$. Can this process extinct? Justify your answer.

ii. (7 marks) Give the criteria for positive and null recurrence and transience of the process, in terms of the given parameters, as explicity as possible.

3. An electrical component of a machine assumed to fail following $Gamma(2, 2)$ distribution. Failed component is replaced immediately. Let $N(t)$ be the number of times the equipment is replaced up to time t.

   (a) (8 marks) Assuming the first replacement occurs at time 0, find $E(N(t))$.

   (b) (8 marks) Assuming the first replacement occurs after a random time that has the density $(1 - F)/\mu$ find $E(N(t))$, where F is the distribution function corresponding to $Gamma(2, 2)$ distribution and μ is the expectation of a $Gamma(2, 2)$ random variable.

   (c) (9 marks) Now, let us assume that the first replacement occurs after a finite random whose distribution is unknown. Using the CLT, give an estimate of $P(ta < N(t) \leq tb)$, for large values of t, where $0 < a < b$ are two reals.

4. There are two servers in a supermarket checkout point who works independently. Let the serving times are $Exp(\mu_1)$ and $Exp(\mu_2)$ respectively. Let the arrivals of customer (who are buying products independently of others) follow a Poisson process with intensity λ, but directed to the first queue if the customer has certain kind of products (say, type I) and directed to the second queue if the customer has bought any other kind of products (say, type II for simplicity). Assume that the customers are buying type I products with probability $p_1 > 0$ and buying type II products with probability $p_2 = 1 - p_1 > 0$. Let $X_i(t)$ be the number of people in the queue i, for $i = 1, 2$, at time t.

   (a) (6 marks) Find the rate matrix Q for both the systems.

   (b) (6 marks) Find the conditions for the transience and null recurrence or positive recurrence of both the systems, in terms of the given parameters.

   (c) (5 marks) Find the probability that ith customer in the first queue arrives before the jth customer of the second queue.

   (d) (8 marks) What is probability that the first server will remain idle, in the long run? Find the probability both the servers will remain idle, in the long run. To answer both of these questions assume positive recurrence condition holds.

All the best.

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18 [Back Paper]

M. STAT. 1st year

Subject Name: Categorical Data Analysis

Date: 05.01, 2018, Maximum Marks: 100. Duration: 3 hrs.

Group A: Answer all questions.

1. Geometrically interpret *Risk Ratio = 1* for 2×2 contingency tables. [20]

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

3. Interpret *thinning operator* for time series of count data. Obtain correlation coefficient at lag h for first order autoregressive Poisson process using the *thinning operator*. [5+10=15]

4. Five groups of animals were exposed to a dangerous substance in varying concentration. Let n_i be the number of animals and y_i be the number that died in the ith concentration.

Concentration	n_i	y_i
1×10^{-5}	6	0
1×10^{-4}	6	1
1×10^{-3}	6	4
1×10^{-2}	6	6
1×10^{-1}	6	6

Describe the fit of an appropriate logistic model for π_i (probability of death) as a function of $\log_{10}$ (concentration). [No numerical calculation is needed. Discuss the applicability of Newton-Raphson procedure in this context.] How can you test for LD_{50} in this context? [5+5=10]

5. Let X and Y be two nominal categorical random variables with I and J categories. Define $\pi_{ij} = P(X = C_i, Y = D_j)$, $i = 1, \ldots, I$; $j = 1, \ldots, J$. Define the measure of variability $V(\cdot)$ as the probability that two independent guesses are wrong.

(a) Find $V(Y)$ [8]

P.T.O

(b) Find the measure of association $A_{Y|X}$ as the proportional reduction of the variability of Y with the knowledge of X. [8]

6. (a) Derive the joint asymptotic distribution of log odds ratios in a 2×4 contingency table. [8]

   (b) Find the P-value of the test for independence for the following table by Fisher's conditional test procedures against one-sided alternative.

	Guess poured first	
Poured first	Milk	Tea
Milk	6	2
Tea	2	6

[8]

Measure Theoretic Probability.

~~Introduction to Measure Theory~~

MS. STAT. 1st YEAR SEMESTER 2

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

Time: 2 Hours Full Marks: 40

Date: February 19, 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 D be a dense subset of $\mathbb{R}$ and f be a real valued function defined on $\mathbb{R}$, such that $\{x : f(x) > \alpha\}$ is measurable for each $\alpha \in D$. Then f is measurable. [5]
   (b) Let $f_n \downarrow f$ hold pointwise for a sequence of nonnegative measurable functions $\{f_n\}$. Then $\int f_n d\mu \downarrow \int f d\mu$. [6]
2. Show that $\liminf_{n\to\infty} A_n = \{\omega : \lim_{n\to\infty} \mathbb{1}_{A_n}(\omega) = 1\}$. [3]
3. On the measurable space $(\Omega, \mathcal{F})$, let μ_n be a sequence of measures, such that $\mu_n(E)$ increases, to $\mu(E)$ say, for each $E \in \mathcal{F}$. Show that μ is also a measure on $(\Omega, \mathcal{F})$. [6]
4. (a) If E is a Borel subset of $[0, 1)$, show that, for every $y \in [0, 1)$, the set $E \oplus y$ is also a Borel set, where $\oplus$ denotes addition modulo 1. [5]
   (b) If E is as above, show that E and $E \oplus y$ have same Lebesgue measure for all $y \in [0, 1)$. (You may assume invariance of Lebesgue measure over $\mathbb{R}$ under translation.) [3]
   (c) Let A be the non-Lebesgue measurable subset constructed in the class. Let E be a Borel measurable subset of A. By considering the translates of E by the rational numbers or otherwise, show that Lebesgue measure of E must be 0. [6]
5. Let F be a continuous distribution function. Show that it is also uniformly continuous. [6]

INDIAN STATISTICAL INSTITUTE

Mid-Semeseter of 2nd Semester Examination : 2017-18

Course Name : M.Stat. 1st Year
Subject Name : Sample Surveys and Design of Experiments
Date : Feb 20, 2018

Total duration for two groups: 1 hr + 1 hr = 2 hrs.

Note: Use separate answer sheets for two groups.

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

Answer any two questions.

1. (a) State and prove Godambe's (1965) theorem regarding the existence of uniformly minimum variance estimator for Y within the class of all unbiased estimators.

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

   (4+6=10)

2. (a) Define admissibility of an estimator within a class of estimators.

   (b) State and prove the theorem on admissibility of the Horvitz and Thompson's estimator (HTE) for Y within the class of all homogeneous linear unbiased estimators.

   (3+7=10)

3. (a) Consider a PPS sampling scheme of size n to be drawn from a population of size N having a known size-measure auxiliary variable x well related with the main variable of interest y. Here the first unit is chosen with a probability proportional to its size-measure and follow it up with an SRSWOR of size $(n-1)$ from the remaining $(N-1)$ units in the population. Obtain the inclusion-probabilities of (i) the units and (ii) the distinct paired units.

   (b) State with proof about the nature of the Yates and Grundy's form of the variance estimator of HT estimator of Y for this sampling scheme.

   (5+5=10)

INDIAN STATISTICAL INSTITUTE
KOLKATA
M. Stat. I : Semester 2
Mid-Semestral Examination
Group B : Design of Experiments

Full Marks : 20 Time : 1 hour

Answer any TWO of the following three questions. Marks allotted to a question are indicated in brackets [ ] at the end.

1. Give the rank and the structural definitions of connectedness of a block design and prove their equivalence. Give an example of a disconnected block design.

   [8 + 2 = 10]

2. State and prove a necessary and sufficient condition for a connected block design to be balanced. Use your result to examine if this block design :

   $B_1 = (1, 2, 3, 4)$, $B_2 = (1, 2, 3, 5)$, $B_3 = (1, 2, 4, 5)$, $B_4 = (1, 3, 4, 5)$, $B_5 = (2, 3, 4, 5)$,

   balanced? Justify your findings.

   [7 + 3 = 10]

3. Prove Fisher's inequality for a B.I.B.D. Give a method of construction of the following

   BIBD : $v = b = s^2 + s + 1$, $r = k = s^2$, $\lambda = s(s-1)$,

   where s is a prime or a prime power, showing clearly that your method leads to the parametric values shown.

   [6 + 4 = 10]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Stat. – I Year, 2017-2018 (Semester – II)

Optimization Techniques

Date: 21.02.2018 Maximum Marks: 60 Duration: 2 hours 30 minutes

---

Note: The question paper is of 75 marks. Answer as much as you can, but the maximum you can score is 60.

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

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

(Q1) Use SIMPLEX method to solve the following LP.

$$\begin{array}{lll} \text{Minimize} & x_0 & \\ \text{subject to} & 8x_1 - 7x_2 - x_0 & \leq 0 \\ & -2x_1 + x_2 - x_0 & \leq 0 \\ & x_1 + x_2 & = 1 \\ & x_1, x_2 & \geq 0 \end{array}$$

[10]

(Q2) 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]

(Q3) Let $P, Q \subseteq \mathbb{R}^n$ be convex sets and let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Suppose that x^* is an optimum solution to $\min\{f(x) \mid x \in P \cap Q\}$ and x^* lies in the interior of Q. Show that x^* is also an optimum solution to $\min\{f(x) \mid x \in P\}$. [10]

(Q4) (a) Deduce the dual of the following LP, where A is an $m \times n$ matrix, $\mathbf{x}$ is an m-dimensional vector, and $\mathbf{y}$ is an n-dimensional vector:

$$\begin{array}{lll} \text{Minimize} & \mathbf{x}^T A\mathbf{y} & \\ \text{subject to} & \sum_{j=1}^{n} y_j & = 1 \\ & \mathbf{y} & \geq \mathbf{0} \end{array}$$

(b) Deduce the conditions on A, $\mathbf{b}$ and $\mathbf{c}$ so that the primal linear program $\mathcal{P}$ and its dual $\mathcal{D}$ are the same linear program.

(c) Prove that $\mathcal{P}$ has an optimal solution if and only if the following set of constraints has a feasible solution.

$$\begin{array}{rcl} A\mathbf{x} & \leq & \mathbf{b} \\ A^T\mathbf{y} & \geq & \mathbf{c} \\ \mathbf{c}^T\mathbf{x} & \geq & \mathbf{b}^T\mathbf{y} \\ \mathbf{x}, \mathbf{y} & \geq & \mathbf{0} \end{array}$$

[7+3+5=15]

(Q5) (a) Describe the steps of a randomized incremental algorithm for solving LPs where the dimension $d = 2$. Clearly mention any assumptions that you make.

(b) Analyze the expected time taken by your algorithm when the dimension $d = 2$.

(c) Generalize the above analysis of the expected time taken by the algorithm when d is a variable parameter and not a constant by forming a recurrence involving n and d. There is no need of solving the recurrence.

[7+3+5=15]

(Q6) 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.

[5+10=15]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2017-18 (Second Semester)

M. STAT. I YEAR
Abstract Algebra

Date : 21.02.18 | Maximum Marks : 40 | Duration : 3 Hours

For any field k, we use $\overline{k}$ to denote its algebraic closure. As customary, the field of rational numbers and real numbers are denoted by $\mathbb{Q}$ and $\mathbb{R}$ respectively.

GROUP A

Attempt any FIVE questions.
Each question carries SIX marks.

1. Let $L|_k$ be a field extension and $\alpha, \beta \in L$. If $\alpha^m, \beta^n \in k$ for two relatively prime positive integers m and n, then show that $k(\alpha, \beta) = k(\alpha\beta)$.

2. Let F be a field and $\sigma \in \text{Aut } F$ be an element of infinite order. If F^σ is the fixed field of σ, i.e., $F^\sigma := \{a \in F \mid \sigma(a) = a\}$, then show that $F|_{F^\sigma}$ cannot be a finite extension.

3. Let k be an infinite field and $L|_k$ be a finite extension which is not normal. If E is the normal closure of L in $\overline{k}$, prove that there exists an element $\alpha \in E$ such that L does not contain any conjugate of α over k.

4. Let R be a unique factorization domain with a field of fractions $F := Q(R)$. If $R \neq F$, show that for every positive integer n, there exists a finite extension of degree n over F. Deduce that $\overline{F}|_F$ is not a finite extension.

5. Let $L|_k$ be a field extension. If $f(X) \in L[X]$ is a polynomial which is not contained in $k[X]$, then show that the set $\{a \in k \mid f(a) \in k\}$ is finite.

6. Let $L|_k$ be an algebraic extension of fields of characteristic zero. If there exists a positive integer n such that every element of L satisfies a polynomial of degree n over k, then prove that $L|_k$ is a finite extension.

7. If $E|_\mathbb{Q}$ is a finite normal extension of odd degree, show that $E \subseteq \mathbb{R}$.

8. Find an infinite family of quadratic extensions over $\mathbb{Q}$ such that no two extensions of the family are isomorphic to each other.

P.T.O

GROUP B

State whether the following statements are TRUE or FALSE with brief justification. An answer without any justification will not fetch any credit. Attempt any FIVE, each one carries THREE marks.

1. A finite field is not algebraically closed.
2. $\mathbb{Q}(\sqrt{5}, \sqrt{7}) = \mathbb{Q}(\sqrt{5} + \sqrt{7})$.
3. Let $k \subseteq L \subseteq E$ be three fields. If $L|_k$ is an algebraic extension and $t \in E$ is a transcendental element over k then t is a transcendental element over L.
4. Let k_1, k_2 be two subfields of a field L such that L is a finite extension over $k_1 \cap k_2$. Then $[k_1k_2 : k_1] = [k_2 : k_1 \cap k_2]$.
5. Let k_1, k_2 be two subfields of a field L such that L is a finite extension over both k_1 and k_2. If $k_1 \cong k_2$ then $L|_{k_1}$ is normal if and only if $L|_{k_2}$ is normal.
6. Let $L|_k$ be an algebraic extension of fields and $\alpha, \beta \in L$. If f, g are the minimal polynomials over k of α and β respectively, then f is irreducible over $k(\beta)$ if and only g is irreducible over $k(\alpha)$.
7. Let f, g be two irreducible polynomials in $k[X]$, where k is a field. If f and g have the same splitting field in $\overline{k}$, then deg $f =$ deg g.
8. Let $L|_k$ be an algebraic extension of fields of characteristic $p > 0$. Then for any $\alpha \in L$, there exists a positive integer n such that α^{p^n} is separable over k.

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination 2017-2018

M. Stat - First year

Metric Topology and Complex Analysis

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

Answer all questions.

For full credit, you have to state the theorems/results you use. Marks for a question is not related to the length of its answer. Brief, precise and to the point answers will be preferred.

(1) Let (X, d) be a connected metric space. Suppose that there is a nonconstant continuous function $f : X \to \mathbb{R}$. Show that X is uncountable. 6

(2) Let (X, d) be a metric space. Define a new metric $\delta(x, y) = \min\{d(x, y), 1\}$ on X. Prove or disprove:

(a) (X, d) is complete implies (X, δ) is complete,

(b) (X, δ) is complete implies (X, d) is complete. 6

(3) Let $(X, d_1), (Y, d_2)$ be two compact metric spaces. Fix a point $x_0 \in X$. Let M be the set of all functions from X to Y. Define a metric ρ on M by

$$\rho(f, g) = \sup_{x \in X}\{d_2(f(x), g(x))\}.$$

Let A be the subset of functions in M which are discontinuous at x_0. Show that A is an open set of M. 10

(4) Let U be a nonempty open subset of $\mathbb{R}^n$ and $U \neq \mathbb{R}^n$. Construct a function $f : \mathbb{R}^n \to \mathbb{R}$ that is discontinuous at every point of U and continuous on U^c. 6

P. T. O.

(5) Let the metric space (X, d) be totally bounded. Fix an $\epsilon > 0$. Show that every infinite subset $Y \subset X$ has an infinite subset $Z \subset Y$, such that diameter$(Z) < \epsilon$. 8

(6) Let A be a subset of a complete metric space X. Suppose A is of first category. Show that A^c is dense. 8

Indian Statistical Institute
Midterm Examination
2017-2018 Academic Year. Second Semester
M.Stat. First Year
Resampling Techniques

Date: 23 February, 2018 Total Marks : 30 Duration: $2\frac{1}{2}$ Hours

Answer all questions

1. (a) Suppose you have iid observations $X_1, \ldots, X_n$ from a distribution F with a finite second moment. Suppose one estimates μ^2 by $\bar{X}_n^2$, where $\bar{X}_n$ is the sample mean and μ is the mean of the distribution. What is the (delete-1) Jacknife estimator of the bias of $\bar{X}_n^2$ in estimating μ^2 ? Prove your assertion.

   (b) Recall the Jackknife estimator of variance (v_{jack}) of an estimator $\hat{\theta}$. Consider a large sample $X_1, \ldots, X_n$ from a distribution and let $\hat{\theta} = g(\bar{X}_n)$, where g is a smooth real valued function. Show that $v_{\text{jack}} \approx [g'(\bar{X}_n)]^2 \frac{s^2}{n}$, where $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ and g' is the derivative of g. [4+5=9]

2. (a) Suppose you have obtained an iid sample $X_1, \ldots, X_n$ from a certain distribution F on R and let $T_n = \sqrt{n}(g(\bar{X}_n) - g(\mu))$, where $\mu = E_F(X_1)$ and g is a real valued function. Stating appropriate assumptions, prove consistency (in appropriate sense) of bootstrap in approximating the distribution of T_n.

   (b) Give one example where the bootstrap fails to be consistent in approximating the distribution of a functional of interest. Prove your assertion.

   (c) Explain the main steps in proving the consistency of the bootstrap in approximating the distribution function (at a given point) of (appropriately centred and scaled) sample quantile. You may assume the necessary regularity conditions without stating them explicitly. [9+4+5=18]

3. Explain with an example why it is important to standardize a random variable for achieving higher order accruracy in approximating distributions through bootstrap than the Central Limit Theorem. [3]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2017-2018

M.Stat. First Year

LARGE SAMPLE STATISTICAL METHODS

Date: 26 February, 2018 Max. Marks: 100 Duration: 2 Hours

Note: $X, X_n, Y_n (n \geq 1)$ denote random variables. For any limiting statement we assume $n \to \infty$.

1. Let $F_n, n \geq 1$, and F be distribution functions on R and F is continuous. If for all $x \in R$, $F_n(x) \to F(x)$, show that $\sup_x |F_n(x) - F(x)| \to 0$. [15]

2. If $g(x) = x^2/(1+x^2)$, show that $X_n \xrightarrow{p} 0$ if and only if $E[g(X_n)] \to 0$. [11]

3. If $X_n \xrightarrow{p} 0$, show that for any median M_n of X_n, $M_n \to 0$. [11]

4. If $X_n - Y_n \xrightarrow{p} 0$, $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ are stochastically bounded and g is a continuous function, show that $g(X_n) - g(Y_n) \xrightarrow{p} 0$. [12]

5. Suppose that $(X_i, Y_i), i \geq 1$ are i.i.d. bivariate random vectors with $E(X_1) = \mu_x, E(Y_1) = \mu_y, Var(X_1) = \sigma_x^2, Var(Y_1) = \sigma_y^2$ and $Corr(X_1, Y_1) = \rho$. If X_i and Y_i are positive random variables, using univariate Central Limit Theorem (CLT) show that

$$\sqrt{n}\left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i} - \frac{\mu_x}{\mu_y}\right)$$

converges in distribution to a normal variable with mean 0 and variance $= (\mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2 - 2\rho\mu_x\mu_y\sigma_x\sigma_y)/\mu_y^4$ (Note that one can use multivariate CLT and delta method to find the asymptotic distribution. You have been asked to use univariate CLT). [12]

6. What is variance stabilizing transformation? Illustrate with an example. [12]

7. Let $X_1, \ldots, X_n$ be i.i.d. with mean μ, variance σ^2 and finite 4th central moment μ_4.

(a) Find the joint asymptotic distribution of $\sqrt{n}(\bar{X}_n - \mu)$ and $\sqrt{n}(S_n^2 - \sigma^2)$ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. (Hint: You may use the Cramer-Wold Theorem and univariate CLT.)

(b) Assume that $\mu > 0$ and $\mu_3 = 0$. Let $T_n = S_n/\bar{X}_n$ and $\gamma = \sigma/\mu$. Find the asymptotic distribution of $\sqrt{n}(T_n - \gamma)$. [13+14]

Indian Statistical Institute
Second Semestral Examination
2017-2018 Academic Year
M. Stat. First year
Resampling Techniques

Date: April 20 , 2018 | Maximum marks: 50 | Duration: $3\frac{1}{2}$ hrs.

Answer all questions

(1) Suppose that $X_1, X_2, \ldots, X_n$ are iid from a certain distribution F and let $T_n = \sqrt{n}(\bar{X}_n - \mu)$, where $\mu = E_F(X_1)$. Stating appropriate assumptions, prove strong consistency (with respect to Kolmogorov metric) of the Bootstrap estimator in approximating the distribution of T_n as $n \to \infty$. [8]

(2) (a) State Miller's Theorem about consistency of (delete-1) Jackknife variance estimators for smooth functions of sample mean. [1]

(b) Give an example of inconsistency of the (delete-1) Jackknife variance estimator. Prove your answer. [1+5]

(c) What can be done to get rid of this problem of inconsistency ? State a general theoretical result to substantiate your answer and indicate the main steps of the proof. [1+2+4]

(3) (a) Describe the Bootstrap percentile method of obtaining confidence interval for a parameter of interest and indicate a situation where it will give an exact lower confidence bound in a one-sided confidence interval. Prove your answer. [2+3]

(b) Stating appropriate assumptions, derive the formula for the lower confidence bound of a one-sided confidence interval obtained from a bias corrected bootstrap percentile method. [6]

(4) Consider the simple linear regression $Y_i = \beta x_i + \epsilon_i$, $i = 1, \ldots, n$ where the x_i's are non-random and ϵ_i's are iid with mean 0 and variance 1.

(a) Describe the residual bootstrap method of approximating the variance of $\hat{\beta}$, where $\hat{\beta}$ is the least squares estimate. [3]

(b) Stating appropriate assumptions, sketch the proof of consistency (in probability) with respect to the Mallows metric for the residual bootstrap method for approximating the sampling distribution of (appropriately centred and rescaled) $\hat{\beta}$. [5]

(5) (a) Explain with an example why use of the ordinary bootstrap may not be a good idea when the observations are dependent. [2]

(b) Discuss three methods of bootstrapping from a stationary time series data stating the conditions under which these bootstrap methods are consistent for approximating the distribution of appropriately centred and rescaled sample mean. [7]

INDIAN STATISTICAL INSTITUTE

Semestral Examination 2017-2018

M. Stat - First year

Metric Topology and Complex Analysis

Date: April 23, 2018 Maximum Marks: 60 Duration: 3 hours

Answer all questions.

For full credit, you have to state the theorems/results you use. Marks for a question is not related to the length of its answer. Brief, precise and to the point answers will be preferred.

(1) Let f be a meromorphic function on a connected open set $\Omega \subset \mathbb{C}$. Suppose that f has exactly one zero at $z_0 \in \Omega$ of order 2 and exactly one simple pole at $z_1 \in \Omega$. Show that

$$f(z) = (z - z_0)^2 (z - z_1)^{-1} h(z) \qquad \forall z \in \Omega$$

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

(2) Let $\Omega = \mathbb{C} \setminus (-\infty, 0]$. Define a holomorphic function $\log_\Omega z$ on Ω such that $e^{\log_\Omega z} = z$ for $z \in \Omega$. Prove that

$$\log_\Omega z = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots, \text{ if } |z| < 1.$$

[5+3=8]

(3) Suppose f is an entire function which is not a polynomial. Show that f is not injective. [10]

P. T. O.

(4) (a) Show that an automorphism f of the upper-half space $\mathbb{H}$ cannot have two fixed points.

(b) For $w, z \in \mathbb{D}$, let $\psi_w(z) = (w-z)/(1-\overline{w}z)$. Let $f : \mathbb{D} \to \mathbb{D}$ be a holomorphic function. Show that

$$|\psi_{f(w)}(f(z))| \leq |\psi_w(z)|$$

[7+10=17]

(5) Using Cauchy's integral formula, find the value of

$$\int_0^{2\pi} e^{e^{i\theta}-i\theta}\, d\theta.$$

[8]

(6) Let f be a holomorphic function on the open unit disc $\mathbb{D}$. Suppose that f has no zero on $\mathbb{D}$. Show that for any $0 < r < 1$,

$$\log|f(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})| d\theta.$$

[12]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Stat. – I Year, 2017-2018 (Semester – II)

Optimization Techniques

Date: 25.04.2018 Maximum Marks: 100 Duration: 4 hours

Note: The question paper is of 120 marks. Answer as much as you can, but the maximum you can score is 60 from Group-A and 40 from Group-B.

Notations: Vectors would be written in small letters with boldface, e.g. **b**; matrices would be written in capital letters, e.g., A. Transpose of A would be denoted by A^T and transpose of 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 A and B be $n \times n$ matrices, then we write $A \succeq B$ (and $A \succ B$) if $A - B$ is positive semi-definite matrix (and $A - B$ is positive definite matrix).

Group-A

(AQ1) (a) Define a basic feasible solution for $\mathcal{P}_{eq}$.

(b) 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 $\mathbf{c}^T x' \geq \mathbf{c}^T x_0$. [2+8=10]

(AQ2) Describe an interior point method for solving an LP of the form $\mathcal{P}_{eq}$. You can assume that a point inside the feasible polyhedron corresponding to the LP has been given to you. Your description of the algorithm should include the following:

(a) the auxiliary problem generated;

(b) the use of Lagrangian multipliers in solving the auxiliary problem and the system of equations derived as a result;

(c) description of the steps of the algorithm and the number of iterations performed.

[1+4+(3+2)=10]

(AQ3) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable over an open convex domain. Show that the following two statements are equivalent:

(i) f is convex.

(ii) $\forall x \in dom(f)$, we have $\nabla^2 f(x) \succeq 0$.

[10]

(AQ4) Consider the following mathematical program in $\mathbb{R}^n$:

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

Assume that $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable functions.

Let 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 above mathematical program with zero duality gap.

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

(b) State and prove the conditions under which KKT conditions are sufficient.

[5+5=10]

(AQ5) Show that system of inequalities in $\mathbb{R}^n$

$$Ax \leq 0, \ \ c^t x < 0, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$, and the following system of inequalities in $\mathbb{R}^m$

$$A^T y + c = 0, \ \ y \geq 0 \tag{2}$$

are strong alternatives. [10]

(AQ6) 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 and the function f satisfy the following properties:

There exist $0 < m < M$ such that 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

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

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 $\beta \in (0, 1)$. [10]

(AQ7) Consider the following mathematical program in $\mathbb{R}^n$:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \qquad (3)$$

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

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

[10]

Group-B

(BQ1) Use SIMPLEX method to solve the following LP.

$$\begin{aligned} \text{Minimize} \quad & x_1 + 2x_2 - x_3 \\ \text{subject to} \quad & 2x_1 + x_2 + x_3 \leq 14 \\ & 4x_1 + 2x_2 + 3x_3 \leq 28 \\ & 2x_1 + 5x_2 + 5x_3 \leq 30 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

[10]

(BQ2) Consider $\mathcal{P}_{eq}$, where all entries of $A = [a_{ij}]$, b and c are integers. Let 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]

(BQ3) (a) Define the *minimum cut* problem in a directed graph $G = (V, E)$ with two distinguished nodes namely *source* (s) and *sink* (t), where s has only outgoing edges and t has only incoming edges, and positive integral edge weights.

(b) Formulate the minimum cut problem as an integer linear program (ILP).

(c) Show, by using total unimodularity or otherwise, that each extreme point solution of a suitably relaxed linear program of the above ILP is a 0/1 solution, and represents a valid cut.

[2+2+6=10]

(BQ4) Let p^* be the optimal objective value of the following mathematical program in $\mathbb{R}^n$:

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

where $A \in \mathbb{R}^{m \times n}$, and d^* be the optimal objective value of the following mathematical program in $\mathbb{R}^m$:

$$\begin{array}{ll} \text{Maximize} & -b^T y \\ \text{subject to} & A^T y + c \geq 0. \end{array}$$

Prove that $d^* = p^*$. [10]

(BQ5) Let $f : \mathbb{R}^n \to \mathbb{R}$, where $f(x) = \frac{1}{2}x^T Q x + b^T x + c$ and $Q \in \mathbb{R}^{n \times n}$ is symmetric.

(a) Show that if $Q \succ 0$, then f is strongly convex function.

(b) Let $Q \succ 0$, and let s_1 and s_n denote the smallest and largest singular value of Q with $\kappa = \frac{s_1}{s_n}$. Compute the convergence rate of Steepest Descent Method with respect to $\| \cdot \|_2$ norm, starting with $x_0 = (0, \ldots, 0)$, to minimize the function f in terms of κ, b and c. Assume that the Steepest Descent Method is using the Exact Line Search subroutine.

[3+7=10]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2017-2018

M.Stat. First Year

LARGE SAMPLE STATISTICAL METHODS

Date: 27 April, 2018 Max. Marks: 100 Duration: 3 Hours

Answer all questions

1. (a) State the representation theorem for sample quantiles due to J.K. Ghosh.

(b) Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 1)$. Find the joint asymptotic distribution of suitably normalized sample mean and sample median.

[2+13=15]

2. Let $X_1, \dots, X_n$ be a random sample from a distribution with a density $f(x, \theta)$, $\theta \in \Theta$, an open interval in R. Assuming suitable regularity conditions on the densities, to be stated by you, prove the following results.

(a) If $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ is the unique solution of the likelihood equation for all n and all $(X_1, \dots, X_n)$, show that $\hat{\theta}_n$ is consistent for estimation of θ and it also maximizes the likelihood function with probability tending to one.

(b) If $\hat{\theta}_n$ is a consistent solution of the likelihood equation, show that the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ (under $\theta = \theta_0$) is $N(0, I^{-1}(\theta_0))$ where $I(\theta_0)$ is the Fisher information number.

You may use the following auxiliary result: If X_n, Y_n and Z_n are random variables such that $X_n Y_n - Z_n \xrightarrow{p} 0$, $Y_n \xrightarrow{p} c > 0$ and $Z_n \xrightarrow{d} Z$, then $X_n c - Z_n \xrightarrow{p} 0$.

[20+20=40]

3. Suppose a population consists of k mutually exclusive classes, the proportion of members falling in the i-th class being π_i, $i = 1, \dots, k$. Let a random

P.T.O

sample of size n be drawn from the population and let n_i be the number of members of the sample belonging to the i-th class. Show that the asymptotic distribution of Pearson's chi-square statistic is chi-square with $(k-1)$ degrees of freedom.

[18]

4. Consider the setup of Question (2). Assume regularity conditions so that $\sqrt{n}(\hat{\theta}_n - \theta)$ is $AN(0, I^{-1}(\theta))$ where $\hat{\theta}_n$ is a maximum likelihood estimator of θ based on a random sample of size n. Fix $\theta_0 \in \Theta$ and set

$$T_n = \begin{cases} \hat{\theta}_n, & \text{if } |\hat{\theta}_n - \theta_0| > n^{-1/4} \\ \theta_0 & \text{if } |\hat{\theta}_n - \theta_0| \leq n^{-1/4}. \end{cases}$$

Find the asymptotic distribution of $\sqrt{n}(T_n - \theta)$ (under θ).

[15]

5. Let $X, X_n, n \geq 1$, be random variables such that

$$\liminf_{n\to\infty} E[g(X_n)] \geq E[g(X)]$$

for every bounded continuous function $g : (-\infty, \infty) \to [0, \infty)$. Show that $X_n \xrightarrow{d} X$. [It is known that $X_n \xrightarrow{d} X$ if and only if $\lim_{n\to\infty} E[g(X_n)] = E[g(X)]$ for all bounded continuous functions $g : R \to R$.]

[12]

INDIAN STATISTICAL INSTITUTE

Final Semeseter Examination of 2nd Semester : 2017-18

Course Name : **M.Stat. 1st Year**
Subject Name : **Sample Surveys and Design of Experiments**
Date : May,...04.... 2018 **Total duration : 3 hrs 30 mins**

Note: Use separate answer sheets for two groups.

Group – Sample Surveys. (Total Marks = 30)

Answer any three questions. Notations are as usual.

1. State Rao, Hartley and Cochran's unequal probability sampling scheme. Obtain an unbiased estimator of Y based on this scheme. Derive its variance and variance estimator. Also state when this scheme attains its optimal situation. [10]

2. State Midzuno, Sen and Lahiri's method of unequal probability sampling. Obtain the first and second order inclusion probabilities of this scheme. Write down the unbiased estimator of Y under this scheme. Mention with reason the property of the Yates and Grundy's form of variance estimator formula under this scheme.

   Is it an IPPS scheme? If no, modify the scheme to obtain an IPPS scheme. [10]

3. Explain the situation when double sampling is to be utilized in stratified random sampling. Show how to obtain an unbiased estimator for $\bar{Y}$ and variance of this estimator under that situation.

   Also discuss about some variants of this method. [10]

4. Show how Hansen and Hurwitz's technique can be utilized to deal with the non-response cases. Obtain an unbiased estimator of $\bar{Y}$ under SRSWOR in this method and variance of this estimator. [10]

5. (a) Explain how the population proportion (θ) of a sensitive binary variable can be estimated by Warner's randomized response model based on a SRSWR sample of n respondents. Based on this model, obtain an unbiased estimator of θ, its variance and variance estimator.

   (b) In a survey to estimate the prevalence rate (θ) of alcoholism among undergraduate college students in Kolkata, Warner's randomized response technique was applied. An SRSWR sample of 100 students was taken and Warner's RRD with two types of cards having statements 'You drink alcohol' and 'You do not drink alcohol' was used with probability $p = 0.4$. Every selected respondent was asked to draw randomly one card and to report 'match' or 'non-match' of his own characteristic with the statement written on card drawn by him. After completion of survey, 57 'matches' and 43 'non-matches' were found. Based on this RR data, estimate the prevalence rate (θ) of alcoholism and also estimate its standard error.

   [5 + 5 = 10]

INDIAN STATISTICAL INSTITUTE
203 B. T. Road
Kolkata 700108
M. Stat I : Sample Surveys and Design and Analysis of Experiments
Semestral Examination
April, 2018

Group: Design of Experiments

Full Marks : 30

Note : Answer any one of Q1 and Q2 , and any two out of the remaining three questions.

1. (a) State and prove a necessary and sufficient condition for a connected block design to be orthogonal. [4]

   (b) Suppose D ia a BIBD with $\lambda = 1$. Delete one block from D, and call the resulting design D*. Is D* connected? Is it balanced? Explain your answers mentioning the results (without proofs) that you may be using in support of your claim. [6]

2. Show that a Hadamard matrix of order 4t coexists with an SBIBD ($v = 4t - 1, k = 2t - 1, \lambda = t - 1$). Hence show that SBIBD ($v = 2^m - 1, k = 2^{m-1} - 1, \lambda = 2^{m-2} - 1$) exists for all m greater than 2. [10]

3. Give a balanced confounding scheme for a ($2^5, 2^3$) experiment saving all the main effects and two factor interactions. Construct one replication of your balanced confounded design. Also give the appropriate ANOVA for your design, showing clearly how the various sums of squares are to be computed. [3 + 3 + 4 = 10]

4. Construct the key block of a replication of a ($3^4, 3^2$) factorial design confounding the independent interactions ABC^2 and AB^2D^2, and describe how the other 8 non-key blocks are to be generated. Suppose in another replication of this factorial design the independent interactions ABC and and AB^2D are confounded. Give the ANOVA of this ($3^4, 3^2$) factorial design in two replications, showing clearly how the various sums of squares are to be computed. [4 + 6 = 10]

5. (a) Construct a one-fourth fraction of a 2^6 factorial experiment, so that only four - factor or higher order interactions are taken in the identity group of interactions defining the fraction.

   (b) Assuming all the three-factor and higher order interactions negligible, give the ANOVA for your fraction, indicating clearly how the various sums of squares are to be computed. [5 + 5 = 10]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2017-18

M. STAT. I YEAR
Abstract Algebra

Date : 07.05.18 | Maximum Marks : 60 | Duration : 4 Hours

For any field k, we use $\bar{k}$ to denote its algebraic closure. As customary, the field of rational numbers, real numbers and complex numbers are denoted by $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ respectively. If $f(X) \in k[X]$ is a separable polynomial, then its Galois group over k is denoted by $\mathrm{Gal}_k(f)$, or simply by $\mathrm{Gal}(f)$ if the underlying field k is understood from the context.

GROUP A

State whether the following statements are TRUE or FALSE with brief justification. An answer without any justification will not fetch any credit. Attempt any FIVE, each one carries THREE marks.

1. A purely inseparable algebraic extension is normal.
2. Let $L|_k$ be a finite extension and $f(X) \in k[X]$ be an irreducible polynomial. If $f(X)$ remains irreducible in $L[X]$ then $\mathrm{Gal}_k(f) \cong \mathrm{Gal}_L(f)$.
3. If $L|_k$ is finite and $\sigma \in \mathrm{Aut}_k L$ then the order of σ divides $[L : k]$.
4. The polynomial $X^{2^n} + 1 \in \mathbb{Q}[X]$ is irreducible.
5. Any solvable group has an Abelian normal subgroup.
6. Let G be a finite group. If H, K are solvable subgroups of G then HK is solvable.
7. Let $k \subset L_1, L_2 \subset L$ be fields with $L|_k$ being a finite extension. If $[L_1 : k] = 2^m$ and $[L_2 : k] = 2^n$ for two positive integers m and n, then $[L_1 L_2 : k]$ is a power of 2.
8. A regular 21-gon is constructible over $\mathbb{Q}$.

GROUP B

Attempt any FIVE questions.
Each question carries EIGHT marks.

1. Let p, q be distinct primes with $p > q$, and $L|_k$ be a finite extension with ch $k \neq p$. If $\alpha, \beta \in L$ are two elements satisfying $[k(\alpha) : k] = p$ and $[k(\beta) : k] = q$, then prove that $k(\alpha, \beta) = k(\alpha + \beta)$.
(Hint : *If $h(X) \in k[X]$ is the minimal polynomial of $\alpha + \beta$, you may consider the polynomial $h(X + \beta) \in k(\beta)[X]$.*)

2. Let R be a commutative ring with unity. A nonzero element $x \in R$ is called a root of unity if there exists a positive integer $n(x)$ such that $x^{n(x)} = 1$. Show that everey nonzero element of R is a root of unity if and only if R is an algebraic extension of a finite field.

3. If $\overline{k}|_k$ is a finite extension, then show that k is a perfect field.

4. Let $L|_k$ be an algebraic extension. If $L(\alpha)|_L$ is a Galois extension for some $\alpha \in \overline{k}$, then prove that there exists a finite extension $L'|_k$ with $L' \subseteq L$, such that $L'(\alpha)|_{L'}$ is a Galois extension and $\mathrm{Gal}(L'(\alpha)|_{L'}) \cong \mathrm{Gal}(L(\alpha)|_L)$. Moreover, if $L|_k$ is normal, then show that $L'|_k$ can be chosen to be a finite normal extension.

5. Let $L|_{\mathbb{Q}}$ be a simple radical Galois extension over $\mathbb{Q}$. If $L \subset \mathbb{R}$, prove that $[L : \mathbb{Q}] \leqslant 2$.

6. If k is an algebraically closed field of characteristic p and t is an indeterminate over k, then show that the Galois group of the polynomial $f(X) = X^p - X - t \in k(t)[X]$ is solvable, but f is not solvable by radicals.

7. The *Abelian closure* of $\mathbb{Q}$, denoted by $\overline{\mathbb{Q}}^{\mathrm{ab}}$, is defined to be the set-theoretic union of all finite Abelian extensions over $\mathbb{Q}$ in $\overline{\mathbb{Q}}$. Then prove the following assertions.

    (i) $\overline{\mathbb{Q}}^{\mathrm{ab}}$ is a subfield of $\overline{\mathbb{Q}}$

    (ii) $\overline{\mathbb{Q}}^{\mathrm{ab}}|_{\mathbb{Q}}$ is a normal extension.

    (iii) $\mathrm{Gal}_{\overline{\mathbb{Q}}^{\mathrm{ab}}}(X^3 - 2) \simeq \mathbb{Z}/3\mathbb{Z}$.

    Conclude that $\overline{\mathbb{Q}}^{\mathrm{ab}} \neq \overline{\mathbb{Q}}$.

8. Let $f(X) \in \mathbb{Q}[X]$ be an irreducible polynomial. If $\alpha \in \mathbb{C}$ is a root of $f(X)$, show that α is constructible over $\mathbb{Q}$ if and only if $|\mathrm{Gal}(f)| = 2^n$ for some non-negative integer n.

GROUP C

Attempt any ONE question.
Each question carries TWELVE marks.

1. If $L|_k$ is a finite normal extension, prove that $L = k^s k^i$, where k^s and k^i denote the separable closure and the purely inseparable closure of k in L respectively.
Using this, or otherwise, prove that if $L|_k$ is an algebraic extension such that every non-constant polynomial in $k[X]$ has a root in L then $L = \overline{k}$.

2. If $n > 1$ is a positive integer, then prove that the arithmetic progression given by $\{n+1, 2n+1, 3n+1, \ldots\}$ contains infinitely many primes.
Let G be a finite Abelian group. Using the above result, or otherwise, prove that there exist infinitely many distinct Galois extensions over $\mathbb{Q}$, each having Galois group isomorphic to G.
(*You may assume that any finite Abelian group is isomorphic to a finite product of cyclic groups.*)