M.Sc. Examination, 2021
Semester-I
Statistics
Course: MSC-11
(Linear Models and Distribution Theory)
Time: 3 Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. (a). Under the assumptions of the Gauss-Markov Model,
   $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{E(\epsilon) = 0}$, $\boldsymbol{Cov(\epsilon) = \sigma^2 I}$
   if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, find the BLUE of $\boldsymbol{\lambda}'\boldsymbol{\beta}$**.**
   (b) Show that the estimator found in (a) is uncorrelated with all unbiased estimators of zero.

   5+5

2. (a). Show that with $\boldsymbol{G}$ a generalized inverse of $\boldsymbol{X}'\boldsymbol{X}$**,** and $\boldsymbol{H} = \boldsymbol{GX}'\boldsymbol{X}$ **,** then $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if $\boldsymbol{\lambda}'\boldsymbol{H} = \boldsymbol{\lambda}'$**.**

   (b) Prove that the BLUE of any linear combination of estimable parametric function is the linear combination of their BLUEs.

   6+4

3. Show that for the linear model $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \boldsymbol{N(0, \sigma^2 I)}$

$$\frac{\frac{(\boldsymbol{\Lambda\hat{\beta}} - \boldsymbol{\Lambda\beta})'(\Lambda S^- \Lambda')^{-1}(\boldsymbol{\Lambda\hat{\beta}} - \boldsymbol{\Lambda\beta})}{m}}{\left(\frac{SSE}{n-r}\right)} \sim F_{m,n-r}$$

   where $\boldsymbol{\Lambda}^{\mathrm{m\times p}}$ is of rank m, $\boldsymbol{S}^-$ is a generalized inverse of $(\boldsymbol{x}'\boldsymbol{x})$. Discuss the case when $m = 1$. (You need to prove all the results to be used) 8+2

4. (a) Let $X_i \sim N(\mu_i, 1), i = 1..n$ and they are independently distributed. Find the distribution of $\sum_{i=1}^{n} X_i^2$. Hence or otherwise find the mgf of the distribution.

   (b) If $\boldsymbol{M} \sim \boldsymbol{W}_p(\boldsymbol{\Sigma}, m)$ and $\boldsymbol{B}$ is a $p \times q$ matrix, then show that

$$\boldsymbol{B}'\boldsymbol{MB} \sim W_q(\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B}, m)$$

   8+2

5. (a) If $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, m), m > p$ then show that the ratio

   $\frac{\boldsymbol{a}'\boldsymbol{\Sigma a}}{\boldsymbol{a}'\boldsymbol{M}^{-1}\boldsymbol{a}}$ has the χ^2_{m-p+1} distribution for any fixed p-vector $\boldsymbol{a}$

   (b). Derive the probability density function of a non-central F distribution. 6+4

6. Consider the following linear model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad i = 1,2,3; j = 1,2$$

Are the following functions estimable?

(i) μ

(ii) τ_1

(iii) $\mu + \tau_1$

(iv) $\tau_1 - \tau_2$

(v) $\tau_1 - \frac{\tau_2+\tau_3}{2}$

10

M.Sc. Examination, 2021
Semester-I
Statistics
Course: MSC-12
(Real Analysis and Measure Theory)
Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin.
Notations have their usual meanings

NOTE: There are total 6 questions. Answer any 4 questions.

1. (a) What is Probability measure? Write the properties of a measure.

(b) Let $\Omega = \Re^2$. Define A_n as the interior of the circle with radius 1 and center $\left(\frac{(-1)^n}{n}, 0\right)$. Find $\limsup_n A_n$ and $\liminf_n A_n$ with suitable explanations.

$[5 + 5 = 10]$

2. (a) Define almost sure convergence, convergence in probability and convergence in distribution.

(b) Define open ball and interior point in $\Re^n$. Prove that, union of finite number of open sets is open.

$[5 + 5 = 10]$

3. (a) State Monotone convergence theorem, Fatou's lemma and Dominated convergence theorem.

(b) Suppose $f_n \to f$ pointwise. Can we say that $\int f_n d\mu \to \int f d\mu$? Explain with suitable example.

$[5 + 5 = 10]$

4. (a) State Weak law and Strong law of large numbers for independent and identically distributed (iid) sequences.

(b) Let $X_1, X_2, \cdots$ be independent and uniformly distributed on $(-1, 1)$. Show that

$$\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n} \to \frac{1}{3} \quad \text{in probability as } n \to \infty.$$

$[5 + 5 = 10]$

5. (a) Define Characteristic function. For any characteristic function $\phi(t)$, explain why $|\phi(t)| \leq 1$?

(b) Let $X \sim uniform(-1, 1)$ with pdf $f(x) = 1/2$ for $-1 < x < 1$. Show that, the characteristic function $C_X(t) = \frac{\sin(t)}{t}$.

$[5 + 5 = 10]$

6. (a) Define accumulation point of a set $S \subset \Re^n$. If A is open and B is close, explain why $A - B$ is open.

(b) State Bolzano-Weirstrass theorem and illustrate with a suitable example.

$[5 + 5 = 10]$

M.Sc. Examination, 2021
Semester-I
Statistics
Course: MSC-13
(Statistical Inference-I)
Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin.
Notations have their usual meanings

NOTE: In question 1 answer any four out of six, In question 2 answer any two out of three.

1. Answer any 4 questions. $4 \times 5 = 20$

(a) What is sufficient statistics? Let $X = (X_1, X_2, \cdots, X_n)$ be result of n independent Bernoulli trials with $X_i = 0, 1;\quad \theta = P(X_i = 1)$ for all $i = 1, 2, \cdots, n$. Show that, $T = \sum_{i=1}^{n} X_i$ is sufficient for θ.

(b) Let $X_1, X_2, \cdots, X_n$ are independent and identically distributed (iid) with distribution $uniform(\theta, \theta + 1)$. Find Maximum Likelihood Estimator (MLE) of θ and comment on its uniqueness.

(c) State Rao - Blackwell theorem or state Lehmann-Scheffe theorem. Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed (iid) with distribution $Poisson(\theta)\quad (0 < \theta < \infty)$. Find MVUE of θ.

(d) Define U-statistics. What is symmetric kernel? Illustrate with an example.

(e) What is Loss functions and expected loss? Illustrate square error loss (quadratic loss) and absolute error loss. What is The Bayes estimator? Illustrate with example.

(f) Suppose that Y follows a $Binomial(n, \theta)$, $(0 < \theta < 1)$. A priori let θ follows Beta prior $(Beta(\alpha, \beta))$, where $(\alpha, \beta > 0)$. Find Bayes estimator under square error loss (quadratic loss).

2. Answer any two questions. $2 \times 10 = 20$

(a) (i) State Neyman-Fisher Factorization theorem. Define minimum sufficient statistic.

(ii) Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed (iid) with distribution $N(\mu, \sigma^2)$, where both (μ, σ^2) unknown. Let $\theta = (\mu, \sigma^2)$. Show that, $T = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ sufficient for θ and also minimal sufficient statistic for θ.

$5 + 5 = 10$

(b) (i) Explain The use of U-statistics as an effective way of obtaining unbiased estimators.

(ii) State and illustrate asymptotic distribution of U-statistics.

$5 + 5 = 10$

(c) Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed (iid) with distribution $N(\mu, 1)$ and a priori $\mu \sim N(0, \tau^{-2})$ (τ known). Let $X = (X_1, X_2, \cdots, X_n)$.

(i) Show that, the posterior of μ given X is normally distributed.

(ii) Explain that, in this question the posterior mean and posterior median have same value. (You may use the fact that normal density is symmetric). Hence, or otherwise, find Bayes estimate of μ under both square error loss (quadratic loss) and absolute error loss.

$5 + 5 = 10$

M.Sc. Examination, 2021
Semester-I
Statistics
Course: MSC-14 (Sample Survey)
Full Marks: 40 Time: 3 Hours

(Answer any four questions.)

1. (a) Explain the problem you will encounter while asking people directly about their belongings to a dichotomous sensitive characteristic.

   (b) Discuss how will you estimate the proportion of individuals belonging to different political parties of West Bengal using unrelated questionnaire method, where the unrelated questions have 'Yes/No' type response.

4+6=10

2. (a) Compare the relative merits and demerits of cluster sampling and two-stage sampling.

   (b) Suggest an unbiased estimator of population total under two-stage sampling. Find its variance and an unbiased estimator of the variance,

2+(1+7)=10

3. (a) Suggest an unbiased estimator of the population total under PPSWR sampling scheme. Find its variance.

   (b) Estimate the gain in precision if we use PPSWR instead of SRSWR for estimating the population total. You should prove the necessary results.

(2+2)+6=10

4. Find the expressions of the expectation and variance of the effective size of the sample obtained using (i) a SRSWR(n,N) design and (ii) a PPSWR(n,N) design. You need to prove all necessary results.

10

5. (a) State and prove a necessary and sufficient condition for the existence of an unbiased estimator of the population total.

   (b) Show that among the class of HLUEs, no one exists with uniformly minimum variance.

5+5=10

6. Given an unbiased estimator t of the population total Y based on a sample s drawn according to a design p, describe a procedure to construct an estimator having variance smaller than that of t.

10

MSC Semester I Examination, 2021

Statistics

Paper: MSC-15

Full Marks: 40 Time: 4 hours

1. There are four objects w_1, w_2, w_3, w_4 whose weights are to be determined.

Left pan	Right pan	Weight needed for equilibrium
w_1, w_2, w_3, w_4	-------------	20
w_1, w_2	w_3, w_4	10
w_1, w_3	w_2, w_4	5
w_1, w_4	w_2, w_3	1

(a). Obtain best estimate of all weights

(b). Find their dispersion matrix of the estimators and estimate of the variance of each measurement. 5+3

2. Suppose that two observations on each of three treatments are as follows:

Treatments

τ_1	τ_2	τ_3
8	5	12
6	3	14

Check whether the following linear parameter functions are estimable or not. Also find the BLUEs incase they are estimable.

i. $\tau_1 - \tau_2$

ii. $\tau_1 + \tau_2$

iii. $2\mu + \tau_1 + \tau_2$

iv. $\frac{(\tau_1+\tau_2)}{2} - \tau_3$

8

3. A random sample of size 20 is drawn from a population with the probability density function

$$f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}; x, \theta > 0$$

and the sample mean comes out to be 12.6. Find MLE of θ. How do you modify the estimate if 2 sample observations are known to exceed value 60 only? Also how do you modify in drawing the sample observation exceeding 60 is rejected. 10

4. Consider the problem of point estimation of θ in $N(\theta,1)$. Given that θ belongs to [-1, 1]. On the basis of a sample of size n, the following estimator has been defined.

$$
\begin{aligned}
T &= -1 \text{ if } \bar{X} < -1 \\
&= \bar{X} \text{ if } -1 \le \bar{X} \le 1 \\
&= 1 \text{ if } \bar{X} > 1
\end{aligned}
$$

$\bar{X}$ being sample mean. Assuming (i) squared error loss and (ii) absolute error loss draw the risk curve of $\bar{X}$ and T over the range $\theta \in [-1,1]$ on the same graph paper and comment. Take n=10.

10

5. For a linear model, the normal equations are:

$$
\begin{aligned}
10\,\beta_1 - 2\beta_2 - 8\beta_3 &= 12 \\
-2\beta_1 + 5\beta_2 - 3\beta_3 &= 16 \\
-8\beta_1 - 3\beta_2 + 11\beta_3 &= -28
\end{aligned}
$$

(a). Obtain any solution of the normal equations.

(b). Find the maximum number of linearly independent estimable parametric functions. 2+2

M.Sc. Examination, 2021
Semester-I
Statistics (Practical)
Course: MSC-16 (Practical on Sample Survey)
Full Marks: 40 Time: 4 Hours

(Answer all questions.)

1) A survey on 32 households was conducted and Warner's randomized response technique (Related Question method with q=0.73) was applied among the heads of the households to ask about the habit of underpaying the Income Tax. The actual amount of Income Tax underpaid is given in the following table.

Household		Response	Amount Underpaid
Serial Number	Size	Yes(1) / No(0)	
1	3	1	2300
2	2	1	17000
3	5	0	5568
4	1	1	1304
5	3	0	0
6	2	1	0
7	4	0	711
8	3	0	1203
9	2	1	9874
10	4	1	2200
11	4	1	0
12	7	0	12000
13	2	1	1807
14	3	1	1400
15	4	0	708
16	4	1	1500
17	5	1	0
18	2	0	1100
19	1	0	1825
20	4	1	0
21	5	1	1407
22	3	1	342
23	2	1	645
24	4	1	0
25	3	0	713
26	5	0	1822
27	2	1	0
28	3	1	1623
29	5	1	1108
30	6	0	365
31	2	1	0
32	3	1	1409

(i) Take a sample of 6 households using Rao, Hartley and Cochran's sampling scheme.

(ii) Estimate the total amount of Income Tax underpaid by these 32 households under the above scheme. Also provide an unbiased variance estimate for it.

(iii) Use the sample to estimate the proportion underpaying the Income Tax.

(iv) Now take a sample of 6 distinct households using Lahiri's method. Provide an estimate of the total amount of IT underpaid under this sampling scheme.

5+(2+3)+3+(3+4)=20

2) Following is a sampling frame regarding study of household size:

Village	Area (km^2)	Previous census population	Number of households	Size of households
1	8.7	69	17	7,5,5,4,6,2,3,5,5,6,5,4,4,4,5,3,3
2	10.6	82	18	6,5,4,5,4,5,6,5,3,5,4,4,5,3,3,5,6,4
3	15.0	110	26	6,6,3,5,3,4,5,5,4,4,4,3,7,5,4,6,2,5,5,6,1,5,5,4,6,3
4	6.2	80	18	6,3,6,3,6,3,4,5,4,4,4,5,6,3,5,1,3,5
5	9.6	92	24	5,4,6,5,4,5,6,5,4,4,7,6,6,5,4,4,5,6,3,4,3,3,5,3
6	7.3	65	17	3,4,4,6,5,7,3,5,4,6,4,5,4,5,3,3,6
7	4.5	72	20	6,4,4,5,4,5,6,4,3,5,4,6,5,5,2,2,4,5,4,3
8	10.6	108	24	5,3,3,7,4,4,6,6,4,5,3,7,6,4,5,6,3,5,1,3,5,4,4,6
9	5.4	106	24	5,3,3,7,4,4,6,6,4,5,3,7,5,6,6,3,5,2,7,5,4,3,1,6
10	3.5	80	22	4,5,6,5,4,4,7,6,6,5,4,4,5,6,3,4,3,3,5,3,5,4
11	5.8	72	15	5,4,5,6,5,4,4,7,6,6,5,4,4,5,6

(i) Take a sample of 4 villages using SRSWOR and select 6 households from each villages using SRSWOR.

(ii) Estimate the estimate of average size per household and the estimate of the variance of the estimator.

(iii) How can an estimate obtained from a simple random sample of 24 plots be compared with the estimate obtained in (ii)?

(iv) Select 3 distinct villages by Lahiri's method and find an estimate of the average household size under this sampling scheme.

3+(3+4)+3+(3+4)= 20

M.Sc. Examination, 2022
Semester-III
Statistics
Course: MSC-31
Stochastic Process

Time: 3 hrs Full Marks:40

Answer **all** questions.

1. Write TRUE/FALSE by the following statements.
   (a) A state i is recurrent if $lim_{n\to\infty}{p_{ii}}^{(n)} = 0$.
   (b) For a finite irreducible Marov chain all states are positive recurrent.
   (c) If the counting process $\{N(t); t > 0\}$ has the stationary increment property then for every $t' > t > 0$, $N(t') - N(t)$ has the same distribution function as $N(t' - t)$.
   (d) A i.i.d. sequence of Cauchy variables is covariance stationary.
   (e) $X_t = X_{t-1} + \epsilon_t$ where $\epsilon \sim N(0, 1)$. This model is covariance stationary.

5

2. Answer in short for the following.
   (a) Write down the state space and index set of a Brownian stochastic process.
   (b) For a discrete time Markov process with state space $S = \{0, 1\}$ with $p_{00} = 1$, $p_{01} = 0$, $p_{10} = .5$, $p_{11} = .5$. Does there exist any unique steady state probability? Justify your answer.
   (c) Consider a Markov chain with state space $\{0, 1, 2\}$ and transition matrix $\begin{pmatrix} \frac{1}{4} & \frac{5}{8} & \frac{1}{8} \\ \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}$. Find $lim_{n\to\infty}p_{12}^{(n)}$.
   (d) Let $X_1(t)$ and $X_2(t)$ be two independent Poisson processes with rates $\lambda_1 = 1$ and $\lambda_2 = 2$ respectively. Let $X(t) = X_1(t) + X_2(t)$. Given that $X(1) = 2$ find the probability that $X_1(1) = 1$.
   (e) Customers arrive at a store with the Poisson flow having rate 10/hr. Each is either male or female with probability 1/2. Compute the probability that at least 5 men entered within 10 a.m. to 10.30 a.m.

$5 \times 2 = 10$

3. Choose the most suitable word for the following multiple choice questions.
   (a) For a Wiener process
      i. measure of drift is double to measure of spread.
      ii. measure of drift is a function of time so is measure of spread.
      iii. measure of drift is function of time but measure of spread is function of time square.
      iv. measure of drift is equal to measure of spread.
   (b) Let π_j denote the long run proportion of time that the chain spends in state j where $\pi_j = 0$. Which of the following is false?
      i. No stationary distribution exists for j.
      ii. state j is null recurrent.
      iii. state j is positive recurrent.
      iv. If $i \longleftrightarrow j$ then i is null recurrent.

(c) For a transition density matrix

  i. all off diagonal elements are positive
  ii. all diagonal elements are zero.
  iii. sum of elements of each row is greater than 0.
  iv. sum of the off diagonal elements is opposite in sign to the diagonal elements.

(d) Suppose John is in gambling zone. On each successive gambling either he wins Re.1 or he loses Re.1 with probability of winning .6. John starts playing with Rs.2. What is the probability that John obtains a fortune of Rs. 4 without going broke?

  i. .91
  ii. .5
  iii. .36
  iv. 1

(e) Two persons are catching fish independently with a Poisson flow at rate 2/hr. What is the expected amount of time that all of them will catch at least one fish?

  i. 45 min
  ii. 1 hr
  iii. 30 min
  iv. none of the above

$5 \times 2 = 10$

4. Show that for an one dimensional symmetric random walk process on integer valued countable state space states are all null persistent. (5)

5. Deduce the difference equation on generalized birth and death process clearly stating the necessary assumptions. (5)

6. Define renewal function. Establish the relationship of renewal function upto time t in terms of renewal function upto the time point t and $x < t$. (5)

M.Sc. Examination, 2021
Semester-III
Statistics
Course: MSC-32 (Categorical Data Analysis and Advanced Data Analysis Techniques)
Full Marks: 40 Time: 3 Hours

(Answer any four questions.)

1. (a) Let a denote the number of calves that got a primary, secondary and tertiary infection, b denote the number that received a primary and secondary but not a tertiary infection, c the number that received a primary but not a secondary infection and d the number that did not receive a primary infection. Let π be the probability of a primary infection. Consider the hypothesis that the probability of infection at time t, given infection at times $1, 2, \ldots t-1$, is also π, for $t = 2, 3$. Find the MLE of π and develop a test procedure for the above hypothesis.
   (b) For testing $H_0 : \pi_j = \pi_{j0}, j = 1, 2, \ldots c$ using sample multinomial proportions $\widehat{\pi_j}$, the likelihood ratio test statistic is defined as $G^2 = -2n \sum_j \widehat{\pi_j} \ln \left(\frac{\pi_{j0}}{\widehat{\pi_j}} \right)$. Show that $G^2 \geq 0$ with equality if and only if $\widehat{\pi_j} = \pi_{j0}$, for all j.

6+4=10

2. (a) Define Yule's Q-statistic for a 2×2 contingency table. Show that for multinomial sampling, the asymptotic variance of $\sqrt{n}\left(\widehat{Q} - Q\right)$ is $\sum_i \sum_j \pi_{ij}{}^{-1} \left(1 - Q^2\right)^2 / 4$, where $\widehat{Q}$ is the sample analogue of Q-statistic. (You need to prove the necessary results.)
   (b) For counts $\{n_i\}$, define the power divergence statistic for testing goodness of fit and find its limits as the power parameter λ tends to 0 and -1 respectively.

(2+5)+3=10

3. (a) What do you mean by a generalized linear model? Describe its components.
   (b) Elaborate explain the fitting procedure of a logistic regression model.
   (c) Define Pearson's residual and Anscombe's residual.

3+5+2=10

4. (a) Find the bootstrap and jackknife estimate of bias and standard error of sample variance with divisor $n - 1$.
   (b) Describe Cross validation technique in details.

6+4=10

5. (a) What do you mean by 'Gibbs sampling'?
   (b) Suppose $(X \mid \theta) \sim N\left(\theta, \sigma^2\right)$; σ^2 known and $\theta \sim$Cauchy (μ, τ), μ, τ known. Describe how Gibbs sampling technique can be applied to simulate from the posterior density of $\theta \mid X$

4+6=10

6. Explain how EM algorithm can be used to cluster data which can be modeled as a mixture of univariate normal populations.

10

M.Sc. Examination, 2021
Semester-III
Statistics
Course: MSC-33(MSE-1)
(Operations Research and Optimization Techniques)
Time: Three Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any five** questions

1. What is a transportation problem? Is it considered to be a Linear Programming Problem? Show that a balanced transportation problem always has a feasible solution. 2+3+3

2. Briefly state the role of modeling in Operations Research. Mention different types of models and their solutions. 2+6

3. For the M/M/1 queuing system find the expected number of customers in the system in the steady state and also the expected queue length. Find the cumulative distribution function for the waiting time of a customer who has to wait in an M/M/1 queuing system. 4+4

4. What is two-person zero-sum game? Transform this game to a Linear Programming Problem. Prove that if mixed strategies be allowed, then there always exists a value of the game. 2+2+4

5. (a) Define an inventory. What are the advantages and disadvantages of having inventories?
(b) Suppose that Q^* is the optimal order quantity and K^* is the corresponding minimum annual variable cost.

Show that if a value of $Q = (1+\alpha)Q^*$ is used, $\frac{K}{K^*} = 1 + \frac{\alpha^2}{2(1+\alpha)}$, where K is the annual variable cost corresponding to an order quantity Q. 4+4

6. (a) What is replacement Problem? Give some illustrations. Discuss replacement policy of equipments that deteriorates gradually with change in time value of money.
(b) What is preventive replacement? Find out criterion for optimal replacement time in such situation. 4+4

7. (a) Distinguish between deterministic and probabilistic models of inventory.
(b) For an inventory model, if P(r) denotes the probability of requiring r units, where r is a discrete variable, C_1 is the inventory holding cost per unit of time, C_2 is the shortage cost per unit per unit of time, then show that the stock level which minimizes the total expected cost is that value of S which satisfies the conditions:

$$\sum_{r=0}^{S-1} P(r) < \frac{C_2}{C_1 + C_2} < \sum_{r=0}^{S} P(r).$$ 4+4

8. Write short notes on any two of the following: 4+4

(a) (s, S) inventory policy
(b) Duality problem in LPP
(c) Saddle point in game theory
(d) Congestion factor in Queuing model

M.Sc. Semester III Examination 2021

Subject: Statistics

Paper: MSC34 (Time Series Analysis)

Full Marks: 40 Time: *Three* hours

Each question carries 10 marks. Answer any *four* questions:

(Symbols have usual meanings)

1. a) Define a time series and describe its components with suitable examples.
   b) Explain the concept of stationarity for time series analysis.

2. a) Comment on the stationarity and invertibility of the following model

$$(1-L)X_t = (1-1.5L)\varepsilon_t$$

where ε_t's are iid$(0,\sigma^2)$

   b) Show that if AR(2) process is stationary, then $\rho_1^2 < (\rho_2+1)/2$.

3. a) Find an invertible process which has the following ACF:

$$\rho_0 = 1,\ \rho_1 = 0.25 \text{ and } \rho_k = 0 \text{ for } k \geq 2$$

   b) For the process

$$(1-0.6L)X_t = (1-0.8L)\varepsilon_t$$

where ε_t is white noise process with mean zero and variance σ^2
Find the ACF ρ_k and PACF φ_{kk} for $k = 1,2$

4. a) Explain with example why frequency domain analysis of time series is necessary. Express the autocovariance function in terms of power spectrum of a stationary time series. Also show that the variance of the time series can also be broken down by frequency.

   b) Deduce the power spectrum of white noise and justify its name.

5. a) State and verify the properties of ACF of a stationary process.

   b) If exists, find the MA presentation of the series

$$(1-0.6L)X_t = (1-1.2L+0.2L^2)\varepsilon_t$$

ε_t is white noise process with mean zero and variance σ^2.
Comment on its stationarity.

6. a) Describe in detail the single exponential smoothing technique for forecasting time series analysis. Also justify the name. Describe an intuitive method for choosing the best smoothing constant.
   b) Explain how one can determine if a single or double exponential smoothing is best for a particular dataset?

Visva Bharati University
M.Sc. Semester III Examination 2019
Subject: Statistics (Practical)
Paper: MSC-35

Full Marks: 40 Time: 4 Hrs.

1. Consider the following data

t	Y_t
1	0.08368
2	0.07456
3	0.11617
4	0.19309
5	0.21007
6	0.15519
7	0.10032
8	0.03708
9	0.02195
10	0.04116
11	0.02517

   (a) Fit a non-linear regression curve of the form $Y_t = ae^{-bt}t^c + \epsilon_t$.

   (b) Find the bootstrap estimate of the standard error of the parameters. Take the number of bootstrap replications (B) to be 50.

   (c) Draw the histogram of the bootstrap replications of each of the parameters. Also check whether they can be regraded to come from a Normal distribution. 3+5+4=12

2. Consider the following sample from a Poisson distribution.
   7, 5, 6, 7, 3, 5, 7, 0, 9, 4
   Based on the sample, find the bootstrap and jackknife estimate of the standard error of (i) sample mode, (ii) sample median, (iii) sample quartile deviation, (iv) sample range , (v) sample Ginis mean difference and (vi) sample standard deviation. Draw histograms of the replications in each case. 8

3. Suppose the random variable X follows Normal distribution with mean θ and variance 1. We assume the prior density to be Cauchy with location parameter 5 and scale parameter 1. Using Gibbs sampling technique, simulate 50 observations from the posterior distribution of θ given X. Also find the Bayes estimate of the parameter θ under squared error loss and absolute error loss. 10

4. Consider the following data set reporting the number of beetles killed after 5 hours exposure to gaseous carbon disulphide at various concentrations.

Log Dose	No. of Beetles	No. of killed
1.691	59	6
1.724	60	13
1.755	62	18
1.784	56	28
1.811	63	52
1.837	59	53
1.861	62	61
1.884	60	60

Fit logit and probit models to this data and comment on your fit. 5

5. The following table gives the result of a study to compare radiation therapy with surgery in treating cancer in larynx. Use R to perform a suitable test for checking whether the population odds ratio equals unity. Discuss your findings.

	Cancer controlled	Cancer not controlled
Surgery	21	2
Radiation therapy	15	3

5

x——— Best of Luck———x

M.Sc Semester III Examination, 2021

Statistics

MSC-36(Practical)

Time: Four Hours   Full Marks: 40

Group-A

1. Maximize $z=5x_1+8x_2$

Such that $3x_1+2x_2 \geq 3$

$x_1+4x_2 \geq 10$

$x_1+x_2 \leq 5$

$x_1 \geq 0,\ x_2 \geq 0$

Solve the problem graphically. 4

2. Solve the transportation problem

	DI	DII	DIII	Supply
OI	4	3	2	10
OII	1	5	0	13
OIII	3	8	5	12
Demand	8	5	4	

8

3. A self service store employs one cashier at its counter. Nine customers arrive on an average of every five minute while the cashier can serve 10 customers in every five minute. Assuming Poisson distribution for arrival rate and exponential distribution for service rate, find (i) average number of customers in the system, (ii) average number of customers in the queue and (iii) average waiting time a customer spends in the system. 4

4. A fish vendor sells fish at the rate of Rs.50 per kg on the day of the catch. He pays Rs.2 per kg of fish not sold on the day of the catch for cold storage. Fish one day old is sold at the rate of Rs.30 per kg and there is unlimited demand for it. The demand of fresh fish is known to follow a uniform distribution over the range from 30 to 50.

(a) Determine the optimum quantity of fish that should be procured by the vendor.

(b) Calculate the maximum profit. Assume that the cost of procurement is Rs.35 per kg. 4

Group-B

Answer <u>all</u> questions. Notations carry usual meanings.
R software may be used for problems 1 and 2. Provide the detail R-code if used.

5. a) Simulate and plot 100 observations from the model
$(1 - 0.6L)Y_t = (1 - 0.8L)\varepsilon_t$
where ε_t is white noise process with mean 0 and variance σ^2
b) Calculate and plot its sample auto-correlation functions and partial auto-correlation functions for lag (k)= 0, 1, 2, …, 20. 5

6. Sales of printing papers (in thousands) from January 1983 to December 1972 are given below:

562.674	560.727	701.108	795.337	742	835.088
599	602.53	790.079	788.421	847.152	934.595
668.516	626.379	594.621	889.968	731.675	832.5
597.798	605.508	230.716	797.393	898.527	300
579.889	646.783	617.189	751	778.139	791.443
668.233	658.442	691.389	821.255	856.075	900
499.232	712.906	701.067	691.605	938.833	781.729
215.187	687.714	705.777	290.655	813.023	880
555.813	723.916	747.636	727.147	783.417	875.024
586.935	707.183	773.392	868.355	828.11	992.968
546.136	629	813.788	812.39	657.311	976.804
571.111	237.53	766.713	799.556	310.032	968.697
634.712	613.296	728.875	843.038	780	871.675
639.283	730.444	749.197	847	860	1006.852
712.182	734.925	680.954	941.952	780	832.037
621.557	651.812	241.424	804.309	807.993	345.587
621	676.156	680.234	840.307	895.217	849.528
675.989	748.183	708.326	871.528	856.075	913.871
501.322	810.681	694.238	656.33	893.268	868.746
220.286	729.363	772.071	370.508	875	993.733

a) Plot the time series and comment on its stationarity.
b) Fit an appropriate ARIMA model to the above data.
c) Forecast 12 more observations of the series using the fitted ARIMA model
d) Plot the original series along with the forecasted values with suitable error bands. 8

7. The following are observations recorded every minute on a scientific experiment:

200	201	208	204	205	207	207	204
202	199	201	198	200	202	203	208
206	210	205	207	203	202	201	199
198	206	207	206	200	201	201	203
200	196	203	205				

Calculate the periodograms for the periods 36, 18, 12, 9, 36/5, 6 and write down the fitted periodic time series model. 7

****