Indian Statistical Institute
Semester 1 (2006-2007)
B. Stat 1st Year
Final Exam
Probability Theory 1

Friday 1.12.2006, 10:30-1:30 Total Points $6 \times 10 = 60$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. A number a is chosen at random from the set $\{1, 2, \cdots, N\}$. If p_N is the probability that $a^2 - 1$ is divisible by 10, then find $\lim_{N \to \infty} p_N$.

2. A pack of cards contains s identical series, each containing n identical cards numbered $1, 2, \cdots, n$. A random sample of $r \geq n$ cards is drawn from the pack *without replacement.* (a) Calculate the probability u_r that each number is represented in the sample. (b) Determine the limit of u_r as $s \to \infty$. $6 + 4 = 10$ pts.

3. Let Q_n denote the probability that in n tosses of a fair coin no run of 3 consecutive heads appears. Notice that $Q_0 = Q_1 = Q_2 = 1$. Find a recursive formula for Q_n in terms of $Q_{n-1}, Q_{n-2}, Q_{n-3}$, by conditioning on the first time tail appears. Evaluate Q_8.

4. Suppose X and Y are independent geometric random variables with parameter $p, 0 < p < 1$. Set $Z = X - Y$ and $M = \min(X, Y)$. Are Z and M independent? Give reasons for your answer.

5. Suppose we have two decks of n cards each numbered $1, 2, \cdots, n$. The two decks are shuffled and the cards are matched against each other. We say that a match occurs at position i if the ith card drawn from each deck has the same number. Let S_n denote the number of matches. Find ES_n and $Var(S_n)$.

P.T.O

6. (a) Suppose $X_1, X_2, \cdots, X_n$ are independent positive random variables. Define

$$Y_k = \frac{X_k}{X_1 + X_2 + \cdots + X_n}, k = 1, 2, \cdots, n.$$

Find the correlation coefficient between Y_k and $Y_l, k \neq l = 1, 2, \cdots, n$. You may use the fact that the Y_k's add to give 1.

(b) Suppose $X_1, X_2, \cdots X_{k+1}$ are i.i.d. random variables. Defining $V_{j+1}^2 = X_1^2 + X_2^2 + \cdots + X_{j+1}^2$, show that for $j < k$ we have

$$E\left(\frac{X_{j+1}^2}{V_{j+1}^2}\frac{X_{k+1}^2}{V_{k+1}^2}\right) = \frac{1}{(j+1)(k+1)}.$$

(Hint: Write $V_{k+1}^2 = V_{j+1}^2 + \tilde{V}_{k-j}^2$ where $\tilde{V}_{k-j}^2 = \sum_{i=j+2}^{k+1} X_i^2$. Now carefully use the identity

$$1 = \left(\frac{V_{j+1}^2(V_{j+1}^2 + \tilde{V}_{k-j}^2)}{V_{j+1}^2(V_{j+1}^2 + \tilde{V}_{k-j}^2)}\right).)$$

$5 + 5 = 10$ pts.

First Semester Examinations (2006 – 2007)
B Stat – 1 year
Remedial English
100 marks
1 ½ hours

Date: 6.12.06

1) Write an essay on any one of the following topics. Five paragraphs are expected.
   a) Mobile phones – a blessing or a curse
   b) A rainy day
   c) My family background

(60 marks)

2) Fill in the blanks with the appropriate prepositions :
   a) The stories in the book are full _____ interest.
   b) Do not indulge _____ strong language.
   c) The cat jumped _____ the chair.
   d) There is a cow _____ the field.
   e) He is fond _____ tea.
   f) He died fighting _____ his country.
   g) Will you come _____ me?
   h) He passed _____ the door.
   i) He was born _____ humble parents in Deulipur.
   j) He is very different _____ his brother.
   k) He did not profit _____ the experience.
   l) He is true _____ his word.
   m) I prefer tea _____ coffee.
   n) It will be done _____ then.
   o) He is devoid _____ common sense.
   p) I am much obliged _______ you.
   q) He sat _____ a stool.
   r) The path leads _______ the woods.

(20 marks)

3) Fill in the blanks with appropriate words:

I stood_________out at the sea. The gray-blue_______came hurtling onto the wet sand and I felt the cold water gather around my bare________. The _______was near the western horizon, a dull red glare almost hidden behind the white _______. Seagulls soared through the darkening blue _______, struggling hard against the strong ________that blew in from the ________. Their shrill ________rented the air.

I________someone call my name and ______around to face the beach. It was a little boy dressed in black, his face somewhat familiar. He stood a little away _______me, a tattered cloth bag hanging from his _______ and a large sea shell held up high in his right ________. It was the latter that he was waving at me, a lopsided grin on his________.

P.T.O

"Very good shell, sir, very good," he________ in broken English. "I ________ it up today morning from this beach. Will you ________ it, sir? Not expensive, not expensive, only forty ____."

I _______ shook my head, urging him to go away.

(20 marks)

INDIAN STATISTICAL INSTITUTE
Final-semester Examination
B.STAT (FIRST YEAR)
Computing Techniques And Programming I
8th Dec. 2006. Full Marks: 80. Duration: Three hours.

Note: Answer all the questions

1. (a) A computer with three 8-bit registers A, B and C is used to run the following program :
   (i) A = (0F) in Hex.
   (ii) B = (03) in Hex.
   (iii) C = 0.
   (iv) C = C + A.
   (v) B = B – 1.
   (vi) If B NOT = 0 jump to (iv).
   (vii) Stop.

   What would be the content of the three registers after the program stops ? Briefly explain your answer.

   (b) RAM chips of 256word x 8 bits capacity are used to design a memory segment of size 1kilo word x 16 bits.
   (i) How many such RAM chips are required for the segment?
   (ii) How many address bus lines are needed?
   (iii) What is the size of the data bus?

   (c) What is a half subtractor? Starting from truth table derive a full subtractor circuit in terms of two half subtractors and an OR gate.

   [2+3+(2+3)]

2. (a) Starting from the truth table definition, derive the Boolean expressions of all the 3-variable linear Boolean functions. (You should try to simplify all the expressions in terms of only EX-OR logic.).

   (b) What is an ODD/EVEN function? Explain their significance.

   [8 + 2]

3. Design the SEC/DED (Single Error Correcting/ Double Error Detecting) code for 8-bit message. (You should illustrate your design with only four code-words instead of describing all the 256 code-words and your illustration should indicate the rectification of at least one single error and the detection of the occurrence of any double

P.T.O

error). [8 +1+1]

4. (a) A disk pack has 19 surfaces. Storage area on each surface has an inner diameter of 22cm. And outer diameter of 33cm. Maximum storage density on any track is 2000bits/cm and spacing between tracks is equal to .25cm. Assuming data storage capacity to be the same on all tracks,
   (i) Find the storage capacity of the disk pack.
   (ii) Find the data transfer rate in bytes/second at a rotational speed of 3600 RPM.

   (b) Considering the message 1101011011 and the generator polynomial as represented by the binary string 10011, calculate the CRC code which has to be appended with the message

   [(3 + 3) + 4]

5. Illustrate the following programming techniques to swap two integer variables with appropriate C code :
   (i) Using only two variables.
   (ii) Using three variables.
   (iii) Using 'Swap' function call
   (iv) Using Exclusive-OR Boolean Logic. [2 +2 +3 + 3]

6. (a) " Most often bubble sort can be speeded up by having successive passes in opposite directions"- Explain the above statement.
   (b) Give the C code for both the simple bubble sort and speeded up bubble sort with full documentations.

   [2 +4 +4]

7. A matrix is called a square matrix if it has the same number of rows and columns. Further a square matrix is a diagonal matrix if its only nonzero elements are on the diagonal from upper left to lower right. It is called upper triangular if all elements below diagonal are zero and lower triangular if all elements above the diagonal are zero. Write a C program that reads a square matrix and determines if it is a diagonal or upper triangular or lower triangular. (You should illustrate all the three cases with full documentation.)

   [2 + 4 +4]

8. Write the C program for creating a simple linked list of several integers and hence extend the C program (with full documentation) for reversing the direction of the pointers in the already created linked list. Your program should print both the linked lists. [5 + 5]

Indian Statistical Institute, Kolkata

B.Stat.I year (2006-07)

Analysis I : Mid-semester Examination

September 13, 2006.

Maximum Marks 100 Maximum Time 3 hrs.

[Note: Answer all questions. Figures on the margin indicate marks alloted to each question.]

1. (a) Consider the subset A of real numbers defined by

$$A = \{x \in \mathbb{Q} : x^2 \leq 3\}.$$

Show that the least upper bound u of A satisfies the relation $u^2 = 3$.

(b) Let $A = \bigcup_{n=1}^{\infty} A_n$ where it is given that each A_n is a subset of real numbers and the union A is a bounded set. For each $n = 1, 2, 3, \cdots$ we denote by u_n the least upper bound of the set A_n and by u the least upper bound of A. Show that u is the least upper bound of the set $B = \{u_i : i = 1, 2, 3, \cdots\}$. [10+10=20]

2. Find:

$$\text{(a)} \lim_{n\to\infty} n^{\frac{1}{n}}, \text{ (b)} \lim_{n\to\infty} \frac{n}{2^n}$$

$$\text{(c)} \lim_{n\to\infty} \frac{n^2 - 3n + 5}{n^3 + n^2 + n}, \text{ (d)} \lim_{n\to\infty} \left(\frac{1}{n^2+1} + \frac{1}{n^2+2} + \cdots + \frac{1}{n^2+n}\right).$$

[5 × 4 = 20]

3. (a) Let (a_n) be a convergent sequence. Show that (a_n) is bounded.
(b) Let (a_n) and(b_n) be convergent sequences. Show that $(a_n b_n)$is convergent.
(c) Let (a_n) be a bounded sequence and $(b_i = a_{n_i})$ a subsequence of (a_n). Prove that

$$\liminf_{n\to\infty} a_n \leq \liminf_{i\to\infty} b_i.$$

[5+7+8 = 20]

4. (a) Let (a_n) be a bounded sequence and a is a given real number with the property that for all $\epsilon \geq 0$, there exists an integer N such that $(a_n) \leq a + \epsilon$ for all $n \geq N$. Show that

$$\limsup_{n\to\infty} (a_n) \leq a.$$

(b) If (a_n) and(b_n) are bounded sequences, show that

$$\limsup_{n\to\infty} (a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n.$$

[10+10 = 20]

5. Test the following series for convergence: (a) $\sum_{n=1}^{\infty} \frac{n+5}{n^2+n+1}$
(b) $\frac{1}{2} + \frac{1}{3} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{2^3} + \frac{1}{3^3} \cdots$. [10+10 = 20]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2006-2007)

B. STAT (FIRST YEAR)

Computing Techniques And Programming I

15th Sept.2006. Full Marks:30. Duration: Two hours.

Note: Answer all the questions.

1. Convert the following with the mathematical proofs for the conversion.
   (a) (101101.1011) in binary to decimal.
   (b) (1998.375) in decimal to binary.
   (c) (5423.76) in Octal to Hexadecimal. [6]

2. The set of three ways of organizing 96 bits of main memory is as follows:
   (a) There are six locations, the content of each location is 16 bits.
   (b) There are eight locations, the content of each location is 12 bits.
   (c) There are twelve locations, the content of each location is 8bits.

   Estimate the possible sizes of Memory Address Register and Memory Data Register in each of the above cases. [3]

3. Derive the novelties of using the 2's complement representation. [4]
4. (a) Distinguish between cross assembler and simulator.
   (b) What do you understand by "Random Access" in connection with memories. [2]
5. (a) What is the value of average after the following program segment is executed:

```
main()
{   int average, sum=0, index=0;
    for ( ;  ;) {   sum=sum+index;
                    ++ index;
                    if (sum>=100) break;
                }
    average =  sum/index;
}
```

   (b) Write a C program to convert a decimal integer number into its binary equivalent. You should illustrate your dry run along with your program. [1+4]
6. Give the flowchart and hence the C program to list all the composite numbers from 11 to 20. You should give the complete dry run for your program. [6]
7. Give the C program to prepare a magic square 3x3 with full documentation. [4]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2006-07
B. Stat. I Year
Analysis I

Date: 27.11.06 **Maximum Marks: 80** **Duration: 3 Hours**

Answer six questions.

1. Find the following limits :

(i) $\limsup_{n\to\infty} a_n$, where $\{a_1, a_2, ...\}$ is an enumeration of the rational numbers in $[0,1]$.

(ii) $\lim_{x\to 0+} \frac{x}{a}\left[\frac{b}{x}\right]$, where $a, b > 0$ and $[y]$ is the integral part of y.

(iii) $\lim_{n\to\infty} n e^{\frac{1}{n}}$

[3+4+3=10]

2. (i) Give an example of a function f such that $\lim_{h\to 0} |f(a+h) - f(a-h)| = 0$, but f is not continuous at the point a. (Merely give f and a, no proof is needed.)

(ii) Show that if f is a continuous function from $[0,1]$ into $[0,1]$, then there exists x such that $f(x) = x$. (Here x may be equal to 0 or 1.)

[4+6=10]

3. For the function $f(x) = \frac{5-3x^2}{1-x^2}$ defined on $\mathbb{R} - \{-1, 1\}$

(i) Find the intervals of increase and decrease, and also locate the local maxima and local minima.

(ii) Is f Lipschitz continuous on $(1, 2]$?

[6+4=10]

4. (a) Prove that $\left(a^{\alpha}\right)^{\frac{1}{2}} = a^{\frac{\alpha}{2}}$ where $a, \alpha > 0$.

(b) Show that the function $f(x) = x^{a}$ is differentiable at all $x > 0$, where $a > 0$.

[4+6=10]

P.T.O

5. (i) If f is twice differentiable on $[0,1]$ show that $f(1) = f(0) + f'(0) + \frac{f''(\theta)}{2!}$ where $\theta \in (0,1)$ (without using the Taylor's theorem, of course).

   (ii) Find $\lim_{x \to 0} \frac{\sin x - x}{x^2}$

[6+4=10]

6. (i) Show that there exists A such that $|\sin^2 x - \sin^2 y| \le A\,|x - y|$ for all x and y.

   (ii) Find the best constant A in (i).

[4+6=10]

7. (i) Let $.\alpha_1\alpha_2\alpha_3 \ldots$ and $.\beta_1\beta_2\beta_3 \ldots$ be distinct decimal expansions representing the same real number $x, 0 \le x \le 1$. Show that there exists i such that from the i th digit onwards, one expansion has all zeros and the other all nines.

   (ii) Let $\sum a_n$ be a convergent series which is not absolutely covergent. Show that there exists a rearrangement $\sum b_n$ of the series such that $\sum_{k=1}^{N} b_k > 0$ for all $N = 1, 2, 3, \ldots$.

[5+5=10]

8. (i) Test the following series for convergence.

   (a) $\sum_{n=1}^{\infty} \sin \frac{\pi}{k}$ (b) $\sum_{n=2}^{\infty} \frac{(-1)^n}{n - \sqrt{n}}$

   (ii) For which values of x does the series

   $$\sum_{k=1}^{\infty} \frac{x^k}{k^k}$$ converge absolutely.

[(3+3) + 4 = 10]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2006-07
B. Stat. I Year
Statistical Methods I

Date: 29.11.06 **Maximum Marks: 100** **Duration: 3 Hours**

Answer all questions.

1. (a) Describe briefly a completely randomized experiment to compare the effectiveness of three diets on the growth of chickens under investigation. What should be your explanatory variable & response variable ? What do you assume about the experimental units ?

   (b) Show how the total sum of squares from (a) may be partitioned into components.

   (c) Comment on the following statement : If there is high within-diet variation in (a), the observed differences among diet means could be due to simply chance.

   (d) An experiment was designed to assess the effects of noise on the educational performance of children. Some hyperactive students were tested as well as a control group of non-hyperactive students. The mean test scores are given below:

Group	High Noise	Low Noise
Control	214	170
Hyperactive	120	140

   Use suitable plots to compare the 2 groups of students and 2 noise levels. Check if any interaction effect is present. Comment on your findings.

[4+3+3+5=15]

2. (a) Consider a 2-way frequency table for bivariate data on X, Y. Suppose the class marks of X and Y (say x_i and y_j, $i = 1, .., p; j = 1, ..., q$) are regarded as the variate values. Let the conditional mean of Y given $X = x_i$ be $\bar{y}_{x_i}$. Fit a least squares regression line to the observed means $\bar{y}_{x_i}$, each weighted by the frequency of x_i, $i = 1 ..., p$. Show that this line is the same as the least squares regression line of y on x. If all the $\bar{y}_{x_i}$ lie on a straight line, what would this imply ?

   (b) Form a bivariate frequency table for the following data on (X,Y) : (2,0), (1,1), (2,1), (1,2), (1,2), (0,3), (0,3), (1,3), (1,3). For this data, is $e^2_{yx} = e^2_{xy}$? (Hint : use (a) above)

   (c) Give an example of a bivariate data set (with 4 data points) for which (i) $e^2_{yx} = e^2_{xy} = r^2 = 1$. (ii) $e^2_{yx} = 1$, $e^2_{xy} = r^2 = 0$.

[6 +4 +5=15]

3. 100 students appear for a test in Mathematics and a test in English. Let $x_1, .. x_{100}$ and $y_1, .. y_{100}$ be their ranks according to the two tests. If there are no ties in the rankings, derive expressions for the means and variances of x and y values and cov (x, y) . Hence, derive an expression for the correlation coefficient (r) between x and y. How would you interpret the cases when r = -1 or r = +1 ?

[5+2+3=10]

P.T.O

4. (a) For two quantitative variables X and Y can r_{xy} be high when there is no causal relationship between X and Y and r_{xy} be low even if causal relationship is present ? Explain.

(b) Construct suitable data to illustrate that for two categorical variables, the direction of association may change when the data is aggregated over groups to form a single group.

(c) In the context of linear regression, will an influential observation necessarily be an outlier ?

(d) Consider the line $y = \bar{y} + \frac{s_y}{s_x}(x - \bar{x})$ for a data set (x_i, y_i), $i = 1, \ldots, n$ (with a scatter plot of positive slope), where s_y, s_x are the standard deviations of y and x, respectively. Show that the sum of squares of the vertical distances of the points from the line is $2.(1-r)\,\Sigma(y_i - \bar{y})^2$, where r is the correlation coefficient.

(e) For data grouped into 5 classes with 30 observations in each class, what will be the range of the intraclass coefficient ? How would you interpret the situation when this coefficient is equal to either of the bounds ?

[3x5=15]

5. (a) Consider a population of N individuals. A simple random sample of size = 2 is drawn from the population without replacement, and the sample variance of the heights of the two individuals is calculated. How is the expected value of this sample variance related to the population variance of the heights of all N individuals ?

(b) Consider a population of N individuals where an unknown number m are smokers. A simple random sample of size n was drawn without replacement from the population, and in the sample, s individuals were observed to be smokers. Suggest an estimate of m based on the sample data such that the expected value of the estimate equals m. Justify your answer.

[10+10=20]

6. There are n distinct points in the plane representing the heights and the weights of n individuals. You are given the slopes of each of the $n(n-1)/2$ lines passsing through the two data points corresponding to a pair of individuals. However, you are not given the original data points that contain the information on the heights and the weights of the individuals. You are also given the standard deviations of the heights and the weights and you know n. Using this information, how would you obtain the residual sum of squares when a straight line is fitted by the method of least squares to the data for predicting weight based on height ?

[15]

7. Assignment

[10]

*****

INDIAN STATISTICAL INSTITUTE
Mid-semester Examination: First Semester (2006-2007)
B.Stat. I year
Vectors and Matrices-I

Date: 11 Sept. 2006. Maximum Marks 30. Duration: 3 Hrs.

Note: Class-room notation is used. You may answer any part of any question. All vectors and matrices considered here are over real field unless otherwise stated.

1. Prove the following statements:
   a) If S is a generating subset and T is an independent subset of a finite dimensional vector space V then $|S| \geq |T|$.
   b) $r(AB) = r(B) - d\{C(B) \cap N(A)\}$.

[5+5=10]

2. Prove the equivalence of the following statements:
   a) $r(A) = r(A^2)$.
   b) $C(A) \cap N(A) = \{\phi\}$.
   c) There exists a nonsingular matrix P such that

$$A = P\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}P^{-1}$$ for some nonsingular matrix D.

[8]

3. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix}$.

   a) Find a generalized inverse, a rank factorization and rank of A.
   b) Find bases of column space, row space and null space of A.
   c) For each of the two vectors given by b, find solution set of $Ax = b$ (i) $b = (2\ 3\ 5)'$ and (ii) $b = (2\ 3\ 6)'$.

[5+2+3=10]

4. Prove or disprove the following statements:
   a) The matrix T is triangular and orthogonal => T = I.
   b) The matrix A is idempotent <=> C)A) = N(I – A).
   c) Every symmetric matrix has a symmetric generalized inverse.

------xxx------

[3x2=6]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2006-07

B.STAT I

Vectors and matrices I

BACK PAPER

Date: 6.2.07 Maximum Marks 100 Duration: 3 hours

(a) Define *row rank* and *column rank* of a matrix and show that they are equal.

(b) Define *rank* and *nullity* of a matrix A of order $m \times n$ and show that they add up to n.

(c) Prove that $r(AB) = r(B) - d\{\mathcal{C}(B) \cap \mathcal{N}(A)\}$.

(d) Prove that $r(AB) + r(BC) \leq r(B) + r(ABC)$.

(e) For any two matrices A and B such that $A + B$ is defined, prove that $r(A+B) \leq r(A) + r(B)$, and equality occurs if and only if $\mathcal{C}(A) \cap \mathcal{C}(B) = \{\phi\}$ and $\mathcal{R}(A) \cap \mathcal{R}(B) = \{\phi\}$.

(f) Let $A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}$, where A and B is nonsingular. Obtain the inverse of A in terms of B, C, D and E.

(g) For square matrices A and B of order n show that $|AB| = |A|\ |B|$.

[$7 \times 7 = 49$]

(a) Show that the definiteness of a real quadratic form is unaltered by a nonsingular linear tranformation of the variables.

(b) If A is a nonindefinite matrix then $a_{ii} = 0 \Rightarrow a_{ij} = 0$ for all j.

(c) Show that every nonnegative definite matrix is of the form LL' for some lower triangular matrix L.

[$3 + 3 + 8 = 14$]

Prove or disprove the following:

(a) $ABB^* = CBB^* \Leftrightarrow AB = CB$.

(b) $x^*Ax = 0$ for all real vectors $x \Rightarrow A = 0$.

(c) $ABA = 0 \Rightarrow B$ can be expressed as $C + D$ where $AC = 0$ and $DA = 0$.

(d) $A^2 = A \Rightarrow Ax = x$ for all x in the $\mathcal{C}(A)$.

(e) Rank of a triangular matrix T is equal to the number of nonzero diagonal elements of T.

P.T.O.

(f) A is a nonnegative definite matrix of order n and rank $r \Rightarrow r^{th}$ order leading principal minor of A is greater than zero.

(g) $(A+I)^4 = 0 \Rightarrow A$ is nonsingular.

(h) $A^2 = I \Rightarrow A = \pm I$.

(i) A is a square matrix of order n and $A^2 = 0 \Rightarrow r(A) \leq n/2$.

(j) A is a real square matrix such that $tr(A^2) = tr(AA') \Rightarrow A = A'$.

[$10 \times 4 = 40$]

——— xXx ———

INDIAN STATISTICAL INSTITUTE
First Semester Back Paper Examination: 2006-07
B. Stat. I Year
Statistical Methods I

Date: 08.02.07 Maximum Marks: 100 Duration: 3 Hours

Answer all questions

1.(a) What are the differences between an 'observational study' and a 'controlled experiment'. Give an example of a controlled experiment where the three principles of experimentation have been implemented. Clearly state your assumptions.

(b) Explain the usefulness of a "double blind" experiment.

[7+3=10]

2.(i) The following table summarizes the student applicants to two departments of a university, categorized by sex, department and admission decision.

Sex of Student	Department 1		Department 2	
	Admit	Deny	Admit	Deny
Male	480	120	10	90
Female	180	20	100	200

(a) Form a two-way table of sex by admission decision for the university combining the two departments. Overall, does the university admit a higher percentage of male applicants?

(b) From the table in (a), should you conclude that there is a sex bias in the university admissions in favour of males?

(ii) For bivariate quantitative data on (X,Y) grouped into a 2-way frequency table, prove that $r^2 \leq e_{yx}^2 \leq 1$. When are the equalities attained?

[(3+3)+(5+4)=15]

3.(a) Two persons, a and B, take independent observations on the some variable in an experiment. The medians of the observations of A and B are approximately equal but the variability is larger for A. Both sets of observations are skewed while A has an extremely large observation. Give a rough sketch of side-by-side box plots to represent the two sets of observations.

(b) After fitting a regression line of y on x , the plot of the residuals against the time order in which the data were collected, shows a fan-shaped pattern? What would this indicate?

Contd...(2)

(c) A plot of the residuals against the corresponding values of the explanatory variable x shows a sloping linear pattern. What would this indicate?

(d) The correlation between the weight of an athlete and the amount of weight he can lift was 0.60. Does this imply that the average athlete can lift 60% of his body weight?

(e) In (d) above, if an athlete gains 10 kgs, can he expect to lift an additional 6 kgs?

[3×5=15]

4.(a) The scores of 10 students in an admission test and a subsequent test are given below. Calculate Kendall's rank correlation coefficient for this data. Interpret your answer.

	Student									
	1	2	3	4	5	6	7	8	9	10
Score in Admission test	82	50	70	79	63	45	95	90	64	49
Score in Subsequent test	54	60	51	55	45	75	79	82	48	50

(b) Can the above coefficient be interpreted as the usual correlation coefficient (r) between two suitably chosen variables? Justify your answer with a proof.

[5+5=10]

5.(a) Show that in any data set, the difference between the mean and the median can never exceed the standard deviation.

(b) Give an example of a data set with positive variance where the difference between the mean and the median equals the standard deviation.

[8+10=18]

6.(a) Consider the data points (4.5, 10.5), (2.0, 8.5), (3.5, 7.0) and (4.0, 9.5). Fit a linear equation $y = \alpha + \beta x$ to the data by the method of least absolute deviations.

(b) Show that for a set of distinct points in the plane, the two least squares lines for predicting y based on x and for predicting x based on y will be identical if and only if all the data points lie on a straight line.

[9+8=17]

7. State and prove Chebyshev's inequality. Give an example of a data set where exactly 75% of the data points lie within one std. dev. of the mean, and all the data points, lie within two std. dev. of the mean.

[8+7=15]

***

Indian Statistical Institute
Semester 1 (2006-2007)
B. Stat 1st Year
Backpaper
Probability Theory 1

Date: 9.2.07 Time: 3 hours Total Points: $7 \times 15 = 105$

Maximum you can score is 100. Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. From an urn containing $(2n+1)$ tickets numbered serially, three tickets are drawn at random without replacement. Prove that the probability that the numbers on them are in arithmetical progression is $3n/(4n^2-1)$.

2. Ten manuscripts are arranged in 30 files (3 files for each manuscript). Find the probability that 6 files selected at random do not contain an entire manuscript.

3. Prove geometrically that there are exactly as many paths ending at $(2n+2, 0)$ and having all interior vertices strictly above the axis as there are paths ending at $(2n, 0)$ and having all vertices above or on the axis. Hence compute the value of

$$P(S_1 \geq 0, \cdots, S_{2n-1} \geq 0, S_{2n} = 0).$$

4. Suppose that the probability that the weather (fine or wet) of one day will be the same as that of the preceding day is a constant p (this model works near the coasts where season changes have little effect). If, on the basis of past records, it is assumed that the probability of the first of July, 2006, being fine is θ, a constant, then determine θ_n, the probability of it being fine on the following nth day. Find the limiting value of θ_n as $n \to \infty$.

5. Of a finite population of N animals in a region, W are caught, marked and released. Members are then caught one by one until w (preassigned) marked animals are obtained, the total number of the animals

P.T.O

in the sample being a random variable X. Find the probability distribution of X and show that

$$EX = \frac{w(N+1)}{W+1}.$$

6. The sum of two independent one-digit numbers ξ and η uniformly distributed on $\{0,1,2,\cdots 9\}$ can be written in the form $\xi + \eta = 10\xi_2 + \xi_1, 0 \le \xi_i \le 9$. Find the probability distributions of ξ_1 and ξ_2. Are they dependent?

7. The random variables $X_1, X_2, \cdots, X_n$ have the same expectation 0 and $Var(X_i) = 1, \forall i$, $Cov(X_i, X_j) = \rho, \forall i \neq j = 1, 2, \cdots, n$. Prove that $\rho \geq -\frac{1}{n-1}$.

Indian Statistical Institute
Semester 2 (2006-2007)
B. Stat 1st Year
Mid-semestral Exam
Probability Theory 2

Monday 19.2.2007, 10:30-1:00 Total Points $5 \times (3+5) = 40$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. A point X is taken at random on rectilinear segment AB of length l whose midpoint is O. What is the probability that AX, BX and AO can form a triangle if

   (a) the distribution of X is assumed to be uniform?

   (b) the distribution of X has density $f(x) = \frac{3}{l^3}x^2, 0 < x < l$?

2. For a continuous random variable X, defined in the range $0 \le x < \infty$, the probability distribution is such that

$$P(X \le x) = 1 - e^{-\beta x^2},$$

   where β is a fixed positive real number.

   (a) Find the median of the distribution.

   (b) If μ, m_0 and σ denote the mean, mode and standard deviation respectively, prove that $2m_0^2 - \mu^2 = \sigma^2$.

3. Suppose X is a continuous random variable with density $f(x)$ on a finite interval $[a,b]$ of length L and zero elsewhere.

   (a) Explicitly write down the formula for $Var(X)$ in terms of a, b and f and show that for any real number β we have the inequality $Var(X) \le E(X-\beta)^2$.

   (b) Show that for X as described above we always have $Var(X) \le \frac{L^2}{4}$ without any assumtion on the form of the density f.

P.T.O

4. Suppose X is a nonnegative continuous random variable with distribu-
tion function F (and density f).

(a) If X has a finite fourth moment, can we say $\lim_{x\to\infty} x^4(1-F(x))$ = 0? Can we say $\lim_{x\to\infty} x^3(1-F(x)) = 0$? Justify your answer.

(b) Prove that for positive r, if the $(2r+2)$th moment of X exists an $\mu_t = EX^t, t > 0$, then $(\mu_{2r+1})^2 \leq \mu_{2r}\mu_{2r+2}$.

5. Consider the following two densities defined over $(0,\infty)$

$$p_1(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-(\log x)^2/2},$$
$$p_2(x) = p_1(x)[1 + a\sin(2\pi\log x)], -1 \leq a \leq 1.$$

(a) Show that p_2 is a density and that p_1 has finite moments of a orders.

(b) If a described in the range $[-1,1]$ is nonzero, are the moments the same (integral) order different for the two densities? Try the secon moment for example.

Indian Statistical Institute, Kolkata

B.Stat.I year (2006-07)

Analysis II : Mid-semester Examination

February 21, 2007.

Maximum Marks 40 Time 2 hrs. 30 mins.

[Note: Answer all questions. Figures on the margin indicate marks alloted to each question.]

1. (a) Show that if f is a bounded function on the interval $[a,b]$ such that for all $\eta > 0$, f is integrable on $[a+\eta, b-\eta]$, then f is integrable on $[a,b]$.
(b) Show that a bounded monotone function on a closed bounded interval $[a,b]$ is integrable.
[5+5= 10]

2. Let f be a nonnegative continuous function on [a,b] with $\sup\{f(x) : x \in [a,b]\} = M$. Show that $\lim_{n\to\infty}(\int_{[a,\ b]} f^n(x))^{\frac{1}{n}} = M$ [8]

3. (a) Assuming the Weiestrass Approximation theorem, show that it is possible to find a sequence of differentiable functions (f_n) converging uniformly to a function f on $(0,1)$which is not differentiable at the point $\frac{1}{2}$.
(b) If f is a continuously differentiable function on $[0,1]$, is it possible to obtain a sequence of polynomials p_n such that $p_n \to f$ uniformly as also $p_n^{'} \to f^{'}$ uniformly.
(c) Give an example to show that the composition of two Riemann integrable functions need not be Riemann integrable.

[2+4+4 = 10]

4. Let $C = \{U_\alpha, \alpha \in A\}$ be a family of open intervals, $U_\alpha = (a_\alpha, b_\alpha)$ such that the closed and bounded interval $[a,b]$ is covered by C.
Let S be the set $S = \{x \in [a,b]$: the interval $[a,x]$ can be covered by a finite subfamily of $C\}$.
Show that if $\sigma = \sup S$, then $\sigma \in S$. [4]

5. Let ϕ be the function

$$\phi(t) = \int_0^1 \sin tx^2, \quad t \in \mathbb{R}.$$

Show that ϕ is continuous on $\mathbb{R}$. [8]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2006-2007)

B. STAT. I year

Vectors and Matrices II

Date: 27 Feb. 2007. Maximum Marks: 30 Duration: 3 Hrs.

(i) All the matrices and vectors considered are over complex field unless otherwise
1. (ii) Class room notation is used.

Show that every square matrix is unitarily similar to a triangular matrix.

Define algebraic multiplicity $am(\lambda, A)$ and geometric multiplicity $gm(\lambda, A)$ of an eigenvalue λ of a matrix A and show that $am(\lambda, A) \geq gm(\lambda, A)$.

Let $\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$ be any set of distinct eigenvalues of A and the set of vectors $S = \{x_1, x_2, \cdots, x_k\}$ be a set of corresponding eigenvectors, that is, $Ax_i = \lambda_i x_i$, for $i = 1, 2, \cdots, k$. Then show that the set S is independent.

Given a square matrix A and a nonnull vector x, show that there exists an eigenvector of A in the space generated by $\{x, Ax, A^2x, \cdots\cdots\}$.

Show that, A and B are nonnegative definite matrices implies there exists a nonsingular matrix Z such that both Z^*AZ and Z^*BZ are diagonal.

Show that if A and B are two matrices such that $A^*A = B^*B$, then there exists a unitary matrix U such that $A = UB$.

[$6 \times 4 = 24$]

ove or disprove the following:

$r(A) = r(A^2) \Leftrightarrow am(0, A) = gm(0, A)$

Real part of every eigenvalue of a real skew symmetric matrix is zero.

$am(\lambda, A) = am(\lambda^k, A)$ for every positive integers k.

$gm(\lambda, A) = gm(\lambda^k, A)$ for every positive integers k.

λ is a nonzero eigenvalue of $A \Leftrightarrow (1/\lambda)$ is an eigen value of every A_r^-.

$gm(1, A) = r(A) \Leftrightarrow A = A^2$.

[$6 \times 1 = 6$]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2006-07

B.Stat. I Year
Statistical Methods II

Date: 23.02.07 **Maximum Marks: 40** **Duration: 2 Hours**

Answer all questions

1. According to a genetic model , the chances that the colour of the flower of a certain type of tree will be red, white and pink are 0.25, 0.25 and 0.5 respectively. A biologist has observed 41 trees with red flowers, 44 trees with white flowers and 115 trees with pink flowers in an experiment producing 200 trees. Is the model supported by the observed data ? Answer this after computing the χ^2 - deviance and the P-value.

[5]

2. Suppose thate Y has χ^2 distribution with 5 degrees of freedom. Compute $P\{Y \geq 7\}$.

[5]

3. A uniform (on $[0,1]$) random number generator on computer has generated the following four values : 0.2487, 0.1638, 0.7923, 0.6410. Compute the Kolmogorov-Smirnov deviance statistic to measure the deviation of the data from uniform distribution on [0, 1]

[5]

4. Suppose that X has p.d.f.

$$f(x) = \begin{cases} 2x^2 & 0 \leq x \leq 1 \\ \frac{4}{21} x^3 & 1 < x \leq 2 \end{cases}$$

Describe a function g (u) such that if u has uniform density on [0, 1], g (u) will have the same distribution as that of X.

[5]

P.T.O

5. (a) Prove that by introducing an additional explanatory variable in a multiple regression equation, one can never decrease its usefulness as a prediction formula.

(b) Given 3 variables X_1, X_2, X_3, you are to find a linear regression equation for X_1. You need to decide whether to use (I) a simple regression with either X_2 or X_3 as regressor or (II) a multiple regression. Which would you choose in the following situations ?

(i) $r_{23} = 1$, (ii) $r_{23} = 0$, (iii) $r_{23} = 0$, $r_{12} = r_{13}$.

(iv) $r_{12} = 0.8$, $r_{13} = 0.4$, $r_{23} = 0.5$.

[3+8=11]

6. (a) From a data set with 3 variables, a student computed $r_{12} = 0.6$, $r_{13} = -0.4$, $r_{23} = 0.7$. Do you think these computations are all correct ?

(b) For 3 variables, let $|r_{ij}| = 1 \quad \forall\ i, j = 1, 2, 3$. How many of these r_{ij} values can be negative ?

(c) If $aX_1 + bX_2 + cX_3 = 0$ holds for all sets of values of X_1, X_2, X_3 , (a, b, c of same sign) what will be the values of the partial correlation coefficients $r_{12.3}$, $r_{13.2}$, $r_{23.1}$?

[3 x 3 = 9]

*****

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination (2006-2007)

BSTAT (HONS) I Year

Computer Techniques and Programming II

Date 2.3.07 Maximum marks 60 Duration 2 Hrs

Note: Answer any five Questions. Give all the necessary steps.

QUESTIONS

1. (a) What is the correct algorithm to calculate the smaller root of the equation $x^2-20x+1=0$. Find the smaller root numerically using three digit arithmetic.

(b) Let x_0 be an approximate cube root of Q and let $s=(Q/x_0^3)-1$. Show that the exact root is given by

$$Q^{1/3} = x_0\,[1+s/3-s^2/9+ 5s^3/81-.......]$$

Apply this formula to obtain the cube root of 2 correct up to 3 places of decimal.

(5+7)

2. Derive the equation

$$x_n = x_{n-1} -(x_{n-1}^p-Q)/(px_{n-1}^{p-1})$$

for finding a pth root of Q. Use this formula to find a fifth root of 29 correct up to 4 places of decimal.

(4+8)

3. (a) Prove that

$$(\Delta^2/E)e^x.E(e^x)/(\Delta^2e^x) = e^x$$

where $\Delta f(x)= f(x+h)- f(x)$ and $Ef(x)=f(x+h)$.,h being a constant.

(b) Find the cubic polynomial of degree 3 which takes the values given below.

P.T.O

x	y
0	1
1	1
2	2
4	5

(6+6)

4. Find a minimum value of y(x) from the data given below

x	y
.60	.6221
.65	.6155
.70	.6138
.75	.6170

(12)

5. (a) Find the truncation error for Simpon's rule for numerical integration.
(b) Apply Trapezoidal rule to evaluate $\int_1^{1.6}\sqrt{x}dx$ correct up to three places of decimal.

(6+6)

6. A Quadrature formula for integration is given by

$\int_{-1}^{1} f(x)dx = af(t_1)+bf(t_2)$

This formula is exact for any polynomial of degree $n \le 3$.

Find a,b,t_1 and t_2. Use this formula to evaluate $\int_0^{\pi/2} \sin x dx$

(8+4)

7. Use the method of Regula falsi to find a real root of the equation

$2x-\log_{10}(x) = 7$ correct up to two places of decimal. (12)

Indian Statistical Institute
Semester 2 (2006-2007)
B. Stat 1st Year
Semestral Examination
Probability Theory 2

Date and Time: 8.5.07, 10:30-1:30 Total Points: $5 \times (6+6) = 60$

Answers must be justified with clear and precise arguments. Each question must be answered on a separate page and more than one answers to a question will not be accepted. If there are more than one answers to a question, only the first answer will be examined. Some of the answers may be in terms of the univariate standard normal cdf Φ.

1. (a) Suppose (U_1, U_2) are the coordinates of a point which is uniformly distributed on the unit square. Consider the events $A_1 = \{U_1 < 1/2\}, A_2 = \{U_2 < 1/2\}, A_3 = \{(U_1 - 1/2)(U_2 - 1/2) < 0\}$. Are they pairwise independent events? Are they mutually independent events?

   (b) Let T_1 and T_2 denote two independent exponential random variables with parameters λ_1 and λ_2 respectively, which represent the lifetimes of two components. Denote $T_{\min} = \min(T_1, T_2)$ and let $X_{\min}$ denote which component failed first, i.e. $X_{\min} = 1$ if $T_1 < T_2$ and $X_{\min} = 2$ if $T_1 > T_2$. Are $T_{\min}$ and $X_{\min}$ independent?

2. (a) Suppose (X, Y) is bivariate normal with means 0, variances 1, and correlation coefficient $\rho, \rho \in (-1, 1)$. Find the joint probability

$$P(X \le x, \rho X > Y - y\sqrt{1-\rho^2}).$$

   (b) Suppose X_1 and X_2 are independent standard Cauchy random variables. Find the distribution of $X_1X_2/(X_1 + X_2)$.

3. Suppose X_1 and X_2 are independent $N(0, 1)$. Consider the following pair of random variables obtained from (X_1, X_2):

$$Y_1 = \frac{X_1 + X_2}{\sqrt{2}}, Y_2 = \frac{X_1 - X_2}{\sqrt{2}}.$$

   (a) Identify

$$\overline{X} = \frac{X_1 + X_2}{2} \text{ and } S^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2}{2-1} = \frac{X_1^2 + X_2^2 - 2\overline{X}^2}{2-1},$$

   in terms of Y_1 and Y_2?

   (b) Are $\overline{X}$ and S^2 independent?

P.T.O

4. (a) Two points are independently and uniformly distributed on $(0,1)$. Suppose d is a fixed positive number, $0 < d < 1/2$. We say that the points are 'isolated' if one point is not within a distance d of the other point. Compute the probability that the points will be isolated.

(b) Suppose (X,Y) has joint density $f_{X,Y}(x,y)$ and X and Y have finite second moments. Define $Var(X|Y=y) = E\{(X - E(X|Y=y))^2|Y=y\}$. Now show that

$$Var(X) = E(Var(X|Y)) + Var(E(X|Y)).$$

5. Let X and Y be independent $N(0,1)$ random variables, and let $Z = X+Y$.

(a) Find the conditional distribution function and the conditional density of Z given that $X > 0$ and $Y > 0$.

(b) Show that $E(Z|X>0, Y>0) = 2\sqrt{2/\pi}$.

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2006-07
B. Stat. - I Year
Analysis II

Date: 11.05.07 Maximum Marks: 60 Duration: 3 Hours

Answer four questions.

1.(a) Let f be a Riemann integrable function on the interval $[a,b]$ and g Lipschits continuous function on an interval J such that $J \supseteq f([a,b])$. Show that gof is integrable on $[a,b]$.

(b) Let f be a bounded non-negative function on $[a,b]$. For $\in > 0$, define the function g by $g(x) = f(x)$ if $x > \in$

$= E$ if $0 \le x \le \in$

Prove that for a partition P of $[a,b]$, $U(g,P) - L(g,P) \le K[U(f^3.P) - L(f^3.P) + \in]$, Where K is a constant independent of P.

(c) Show that the sequence

$$a_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots\cdots + \frac{1}{n} - \log n, \qquad n = 1,2,3,\cdots\cdots$$

is convergent.

[4+6+5]

2.(a) Let $f_n(x) = \frac{1}{nx+1}$ if $0 < x < 1, n = 1,2,.....$

Show that $\{f_n\}$ converges pointwise but not uniformly on $(0,1)$.

(b) Let $\{f_n\}$ be a sequence of real valued continuous functions converging pointwise to a continuous limit function f on $[a,b]$. If $f_n(x) \ge f_{n+1}(x)$ for each $n = 1,2,......$ and each $x \in [a,b]$, then show that f_n converges to f uniformly on $[a,b]$.

(c) Show that if $a_1 \ge a_2 \ge \cdots\cdots$ and $\lim_{n\to\infty} a_n = 0$ the series $\sum_{n=0}^{\infty} a_n \cos nx$ converges uniformly on every interval $(\delta, 2\pi - \delta)$ where $\delta > 0$.

[4+5+6]

P.T.O.

3.(a) Find the radius of convergence of the power series

$$1+2x+3^2\frac{x^2}{2!}+4^3\frac{x^3}{3!}+\cdots\cdots$$

(b) Show that for any $a\in\mathbb{R}$,

$$(1+x)^a=1+ax+\frac{a(a-1)}{2!}x^2+\cdots+\frac{a(a-1)\cdots(a-n+1)}{n!}x^n+\cdots\cdots$$

For all $x\in(-1,1)$.

(c) Show that $\lim_{x\to1}\sum_{n=1}^{\infty}\frac{x^n}{n^4}=\frac{\pi^4}{90}$ (by whatever method you may like).

[5+5+5]

4.(a) Test the following improper integrals for convergence:

(i) $\int_0^{\infty}\frac{|\cos x|}{\sqrt{x}}dx$ (ii) $\int_0^1\frac{x^p-1}{\log x}dx\,(p>-1)$

(b) Show that

$$\int_0^{\infty}\frac{\sin^3 x}{x}dx=\frac{\pi}{4}$$

(c) Show that $\Gamma(p)=\int_0^{\infty}e^{-x}x^{p-1}dx$, $p>0$ is a continuous function on $(0,\infty)$.

[5+5+5]

5. Let f be a continuously differentiable 2π periodic function on $\mathbb{R}$.

(a) Show that the Fourier coefficients satisfy the relations

$$a_n(f)=-\frac{1}{n}b_n(f')$$
$$b_n(f)=\frac{1}{n}a_n(f')$$

where f' is the derivative of f.

Contd....(3)

(b) Show that $\sum|a_n(f')|^2+|b_n(f')|^2<\infty$.

(c) Prove that $\sum_{n=0}^{\infty}|a_n(f)|<\infty, \sum_{n=1}^{\infty}|b_n(f)|<\infty$.

(d) Show that if $s_n(f)$ is the nth partial sum of the Fourier Series of f, then $s_n(f)$ converges and $\lim_{n\to\infty}s_n(f)=f$.

[4+4+3+4]

***

INDIAN STATISTICAL INSTITUTE

Seond Semester Examination: (2006-2007)

B. STAT. I year

Vectors and Matrices II

Date: 14 May 2007. Maximum Marks: 60 Duration: 3 Hrs.

Notes: *(i) All the matrices and vectors considered are over complex field unless otherwise stated. (ii) Class room notation is used. (iii) State clearly the results used.*

1 On eigenvalues of real symmetric matrices, state and prove the following theorems:

(a) Courant-Fischer min-max theorem.

(b) Sturmian separation theorem (Interlacing theorem).

(c) Poincare separation theorem.

[$3 \times 5 = 15$]

2 (a) Given a pair of *nnd* matrices X and Y of same order, show that there exists a nonsingular matrix U such that both U^*XU and U^*YU are diagonal.

(b) Given any pair of matrices $A_{m\times n}$ and $B_{k\times n}$, show that there exist a pair of unitary matrices $P_{m\times m}$ and $Q_{k\times k}$ and a nonsingular matrix $Z_{n\times n}$ such that

$$A = P\begin{pmatrix} \Phi & 0 \\ 0 & 0 \end{pmatrix} Z \quad \text{and} \quad B = Q\begin{pmatrix} \Psi & 0 \\ 0 & 0 \end{pmatrix} Z,$$

where Φ and Ψ are diagonal matrices of order r with nonnegative diagonal elements satisfying $\Phi^2 + \Psi^2 = I_r$, r being the rank of the matrix $(A^* : B^*)$. (Assume that $r \leq min\{m, k\}$).

[$5 + 10 = 15$]

3 (a) Define rank and signature of a quadratic form and prove that a quadratic form $x'Ax$ can be written as the product of two linearly independent linear forms in x if and only if A has rank 2 and signature 0.

(b) Let A, B and $A - B$ be *nnd*. Show that there exist *nnd* g-inverses A^- and B^- of A and B respectively such that $B^- - A^-$ is *nnd*.

[$5 + 5 = 10$]

4 (a) For the matrix A given below, find a lower triangular matrix L such that $A = LL'$ using square root method. Also find $|A|$.

$$A = \begin{pmatrix} 4 & 2 & -2 & 0 \\ 2 & 2 & 1 & 3 \\ -2 & 1 & 6 & 7 \\ 0 & 3 & 7 & 14 \end{pmatrix},$$

P. T. O

(b) For the above given matrix A, reduce the quadratic form $x'Ax$ to diagonal form using a nonsingular linear transformation of the variables. Give also the matrix of the transformation.

[5 + 5 = 10]

5 Prove or disprove the following:

(a) A is *nnd* and $x'Ax = 0$ for some vector x implies $y'Ax = 0$ for all y.

(b) Every 2×2 real normal matrix is either symmetric or skew symmetric.

(c) Let A and B be real symmetric matrices where A is *pd*. Then each eigenvalue of $A^{-1}B$ is 1 implies $A = B$.

(d) $x'J_n x \leq n\, x'x$.

(e) The modulus of any eigenvalue of a square matrix A cannot be greater than the largest singular value of A.

[5 × 2 = 10]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2006-07
B. Stat. I Year
Statistical Methods II

Date: 16.05.07 Maximum Marks: 60 Duration: 3:00 Hours

Answer all questions 1-6.

1. (X_1, X_2, X_3) has a joint normal distribution with mean vector $(0.3, 1.2, 2.6)$ and dispersion matrix $\begin{pmatrix} 4 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix}$. You are given three independent uniform (on $[0,1]$) random numbers 0.2461, 0.8023, 0.6148. From these three random numbers, generate a set of values for X_1, X_2 and X_3.

[8]

2. Suppose that we have two bivariate observations $X_1 = 0.3125, Y_1 = 0.6271$ and $X_2 = 0.4923, Y_2 = 0.8821$. Compute the Kolmogorov-Smirnov distance between the empirical distribution based on these observations and the joint distribution of (X, Y), where X and Y are independent uniform random variables each distributed on $[0,1]$.

[10]

3. Consider the bivariate density function

$$f(x,y) = \begin{cases} \frac{1}{2} & \text{if } |x| + |y| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Describe (with justification) how to generate (X, Y), with p.d.f $f(x, y)$ using (u, v), where u and v are independent random variables with a common uniform distribution on $[0,1]$.

[12]

4.(a) Consider p variables $X_1, X_2, \ldots, X_p$. Show that the multiple correlation coefficient $r_{1.23\cdots p}$ must be numerically at least as high as any of the partial correlation coefficients (of different orders) of X_1 with the other variables.

P.T.O.

(b) Suppose the correlation coefficients r_{ij} $(i, j = 1, \cdots\cdots, p,\ i \neq j)$ are all equal. If this equal value is r, show that $r_{12.34\cdots\cdots p} = \dfrac{r}{1+pr}$. Hence, show that $r \geq -\dfrac{1}{p-1}$.

(c) With $p = 3$, suppose $r_{23} = 0$ and r_{12}, r_{13} are approximately equal. For predicting X_1, should one use a simple regression or a multiple regression?

[6+5+4=15]

5. Consider the model $Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i, i = 1, \cdots\cdots, n$; where the ϵ_i's are independent and identically distributed random variables with mean 0. Suppose $n = 6$ and $x_1 = x_2 = -1$, $x_3 = x_4 = 0$, $x_5 = x_6 = 1$. Check if $\alpha + \beta_2$ is estimable. Is β_1 estimable?

[5]

6. A factory manufactures bolts of cloth and the number of defects on any bolt has mean 5. Suggest, with reasons, a suitable distribution for the number of defects per bolt. In a sample of 100 bolts, find the expected number of bolts with fewer than 2 defects.

[5]

7. Assignments.

[5]

***

INDIAN STATISTICAL INSTITUTE

Semestral Examination : (2007-2008)

BSTAT (HONS) 1st Year

Computer Techniques and Programming II

Date : 18.05.2007 Maximum Marks : 80 Duration : 3 Hours

Note : Answer any five questions. Give all the necessary steps.

1. A predictor formula is given by

$$y_{k+1} = a_0 y_k + h\left[a_k y'_k + a_2 y'_{k-1} + a_3 y'_{k-2} + a_4 y'_{k-3}\right]$$

where $y_k = y(x_k)$ and $y'_k = \left.\dfrac{dy}{dx}\right|_{x=x_k}$ x_k being given by $x_k = x_0 + hk$. Using Newton's back ward formula of degree 3 or otherwise find a_0, a_1, a_2, a_3 and a_4. Find the local truncation error. (8+8)

2. (a) Use Runge Kutta method (fourth order) to solve $\dfrac{dy}{dx} = xy^{\frac{1}{2}}$ with $y(1) = 1$ and find $y(1.1)$ and $y(1.2)$ (take $h = .1$) (5+5)

   (b) Use Euler's modified method to the problem $\dfrac{dy}{dx} = 1+xy$ with $y(0) = 2$ to find $y(.1)$, correct up to 2 decimal places. (take $h = .1$) (6)

3. Use Jacobis' method to find all the eigen value and eigen vectors of the matrix

$$\begin{bmatrix} 1 & \sqrt{2} & 2 \\ \sqrt{2} & 3 & \sqrt{2} \\ 2 & \sqrt{2} & 1 \end{bmatrix}$$ (10 + 6)

4. Write the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$ in the form $A = LU$ where L is a lower triangular matrix and U is an upper triangular matrix with all diagonal elements equal to 1. Hence solve the system of equations given by

P.T.O

$$\begin{aligned} 2x + y + z &= 6 \\ 3x + 2y + 3z &= 14.5 \\ x + 4y + 9z &= 35.5 \end{aligned}$$ (8 + 8)

5. Fit a cubic spline to the data

x	y
0	0
1	−1
2	2
3	15
4	44

(take the second derivatives at $x = 0$ and $x = 4$ to be zero i.e. $S_0'' = S_4'' = 0$).
Using the spline find $y(1.3)$ (12+4)

6. (a) Use Gauss-Seidel iteration method to solve the system of equations given by

$$\begin{aligned} 27x + 6y + z &= 85 \\ x + y + 54z &= 110 \\ 6x + 15y + 2z &= 72 \end{aligned}$$

correct up to one place of decimal (10)

(b) Use Given's method to reduce the matrix

$$A = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 4 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

to tridiagonal form. (6)

7. For a matrix A, it is given that λ_1 is the dominant eigen value and X_1 is the corresponding eigenvector. Let λ_2 be the next dominant eigenvalue and X_2 be the corresponding eigenvector. Define the matrix

$$B = A - X_1 a^T$$

when a^T is the first row vector of A. Show that $X_1 - X_2$ is an eigenvector of B with eigenvalue λ_2. Use this result to find out all the eigenvalues and eigenvectors of the matrix A given by

$$A = \begin{bmatrix} 1 & -3 & 2 \\ 4 & 4 & -1 \\ 6 & 3 & 5 \end{bmatrix}$$

given that 7 is the dominant eigenvalue of A. (4+12)

INDIAN STATISTICAL INSTITUTE
Seond Semester Examination: (2006-2007)
B. STAT. I year
BACK PAPER
Vectors and Matrices II

Date: 31.7.2007. Maximum Marks: 100 Duration: 3 Hrs.

Notes: *(i) All the matrices and vectors considered are over complex field unless otherwise stated. (ii) Class room notation is used. (iii) State clearly the results used.*

1 (a) If A and B are matrices of same order such that $AA^* = BB^*$, then show that there exists a unitary matrix U such that $A = BU$.
(b) Derive spectral decomposition of a real symmetric matrix.
(c) Derive singular value decomposition of a matrix A of order $m \times n$.
(d) Let $A = \delta_1 u_1 v_1' + \delta_2 u_2 v_2' + \cdots + \delta_r u_r v_r'$ be a singular value decomposition of A of rank r. Define $A_k = \delta_1 u_1 v_1' + \delta_2 u_2 v_2' + \cdots + \delta_k u_k v_k'$, for $k = 1, 2, \cdots, r$. Then show that $trace(A - B)'(A - B) \geq trace(A - A_k)'(A - A_k)$ for all matrices B of rank less than or equal to k.

[5 + 5 + 5 + 10 = 25]

2 (a) If A is a *pd* matrix, then show that any principal submatrix of A and its Schur's complement are *pd*.
(b) Let A be a *pd* matrix of order n. Then Show that $|A| \leq a_{11}a_{22}\cdots a_{nn}$ with equality if and only if A is diagonal.
(c) Let $A_{n\times n}$ be a real matrix such that $|a_{ij}| \leq 1$ for all i, j. The prove that $|A| \leq n^{n/2}$ and equality holds if and only if $a_{ij} = 1$ or -1 for all i, j and A is orthoganal.

[3 × 5 = 15]

3 Prove the following:
(a) Every independent subset of the vector space $\Re^n$ is orthogonal under some inner product on $\Re^n$.
(b) If A is real idempotent matrix, then there exists a *pd* matrix P such that PA is symmetric.
(c) If $A = B + C$ where B is *pd* and C is real skew symmetric matrix, then $|A| \geq |B|$.
(d) If A and B are *pd* matrices of same order and if $A - B$ is *nnd*, then $A^{-1} - B^{-1}$ is *npd*.

[4 × 5 = 20]

P.T.O

4 (a) Find $|A|$ for the matrix

$$A = \begin{pmatrix} 4 & 2 & -2 & 0 \\ 2 & 2 & 1 & 3 \\ -2 & 1 & 5 & 6 \\ 0 & 3 & 6 & 13 \end{pmatrix}.$$

(b) Find the singular values of the matrix

$$\begin{pmatrix} 2 & 1 & 0 & 1 \\ -1 & 1 & 1 & -2 \\ 1 & 2 & 1 & -1 \end{pmatrix}.$$

[10 + 10 = 20]

5 Prove or disprove the following:

(a) A real symmetric matrix A is *nnd* if and only if it has an *nnd* g-inverse.

(b) AB and BA have the same set of nonzero eigenvalues.

(c) $r(A)$ = the number of nonzero eigenvalues of A implies $r(A) = r(A^2)$.

(d) Every singular value of an idempotent matrix is greater than or equal to 1.

(e) For any real square matrix $trace(AA') = trace(A^2)$ implies A is symmetric.

[5 × 4 = 20]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE

Second Semester Backpaper Examination: 2006-07

B. Stat. I Year

Statistical Methods II

Date: 01.08.07 **Maximum Marks: 100** **Duration: 3 Hours**

Answer all questions.

1. A coin is tossed 400 times and HEAD is obtained in 250 tosses. Compute the χ^2-statistic that measures the deviation of this data from Binomial $\left(400, \frac{1}{2}\right)$ model. What is the P-value in this case ?

[12]

2. Let Y be a random variable with χ^2- distribution having 5 degrees of freedom. Compute $Pr(Y \geq 7)$.

[13]

3. Consider the bivariate density function

$$f(x,y) = \begin{cases} \dfrac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Describe (with justification) how to generate (X,Y) with p.d.f $f(x,y)$ using (u,v), where u and v are independent random variables with a common uniform distribution on [0.1].

[12]

4. You have a fair coin that you can toss as many times as you want. How would you use the coin the generate random numbers that are approximately uniformly distributed over the interval [0,1]? Justify your answer.

[13]

5.a. Observations on p variables $X_1, X_2, \ldots, X_p$ are obtained for each of n individuals. Using the least squares method, derive the form of the multiple regression equation of X_1 on $X_2, \ldots, X_p$ expressing the regression coefficients in terms of the elements of the correlation matrix of $X_1, \ldots, X_p$

P.T.O

b. If you now want to include a new variable X_{p+1} in the regression equation, show how the regression coefficients in the new regression equation of X_1 on $X_2,\ldots,X_p,X_{p+1}$ may be obtained in terms of the coefficients of the lower order.

[8+7=15]

6. a. For p variables $X_1,\ldots,X_p$ define the multiple correlation coefficient. What is the range of this coefficient ? Justify your answer.

b. From a data set with 3 variables, the following were computed :

$r_{12} = 0.6,\ r_{13} = -0.4,\ r_{23} = 0.7.$

Can these computations be all correct ? Justify your answer.

c. If $aX_1 + bX_2\, cX_3 = 0$ holds for all sets of values of X_1, X_2, X_3 (a,b,c are of same sign), what will be the values of the partial correlation coefficients $r_{12.3}$, $r_{13.2}$, $r_{23.1}$?

(7+5+3=15)

7. The following table gives the frequencies of eggs laid by flies on flower-heads. The count of flower-heads with no eggs is not available.

No. of eggs	1	2	3	4	5	6	7	8	9	≥ 10
No. of flower heads	22	18	18	11	9	6	3	0	1	0

Assuming that the number of eggs laid follows a Poisson distribution, fit a suitable distribution to the data and calculate the expected frequencies for the classes.

How will you check if the distribution fits the data ?

(15+5=20)

*****

INDIAN STATISTICAL INSTITUTE
Second Semester Backpaper Examination: 2006-07

B. Stat. I Year
Analysis II

Date: 03.08.07 **Maximum Marks : 100** **Duration: 3 Hours**

Answer all questions.

1. (a) Let f be a non-negative continuous function on [a,b]. Given that $\int_a^b f = 0$, show that $f(x) = 0$ for all $x \in [a, b]$.

(b) Let f and φ be defined on [a,b] such that f is monotone on [a,b] and f' and φ' are continuous on [a,b]. Show that

$$\int_a^b f(x)\,\varphi(x)\,dx = f(a)\int_a^\xi \varphi(x)\,dx + f(b)\int_\xi^b \varphi(x)\,dx \text{ for some } \xi \in [a,b].$$

(c) Show that if $q > p > 0$, $\left| \int_p^q \frac{\sin x}{x}\,dx \right| < \frac{2}{p}$.

[5 + 8 + 7]

2. (a) Show that the improper integrals

$$\int_0^1 \log x\,dx \qquad \int_0^\pi e^x \log|\sin x|\,dx$$

exist.

(b) Prove that a twice differentiable function φ on an open interval I is convex if and only if $\varphi''(x) \geq 0$ for all $x \in I$.

[5 + 5 + 10]

3. (a) Show that if a sequence of continuous functions f_n converge uniformly a function f on the interval I, then f is continuous on I.

(b) Show that if the power series $\sum_{n=0}^{\infty} a_n x^n$ has the radius of convergence R , then the function $f(x) = \sum_{n=0}^{\infty} a_n x^n$, $|x| < R$, is infinitely differentiable on $(-R, R)$.

P.T.O

(c) Decide if the sequence of functions

$$f_n(x) = n\,x\,e^{-n^2x^2} \qquad x \in \mathbb{R},\quad n = 1, 2, \ldots$$

Converges uniformly on $\mathbb{R}$.

[7+7+6]

4. (a) Let $s_n(x)$ is the nth partial sum of the Fourier series of the integrable and absolutely integrable function f on $[-\Pi, \Pi]$,

$$s_n(x) = \sum_{k=0}^{n} a_k \cos kx + \sum_{k=1}^{n} b_k \sin kx .$$

If $\tilde{f}$ is the periodic extension of f on $\mathbb{R}$, show that

$$s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{f}(x-y) \frac{\sin\left(n+\frac{1}{2}\right)y}{\sin y/2}\, dy.$$

(b) Show that for a function f as in (a), $a_n(f) \to 0$ as $n \to \infty$.

(c) Show that if a continuous periodic function f has $a_n(f) = 0$, $n = 0, 1, 2, \ldots$ and $b_n(f) = 0$, $n = 1, 2, \ldots$, then $f(x) = 0$ for all x.

[7+8+5]

5. (a) Obtain the power series expansion of the function $f(x) = (\tan^{-1} x)^2$ where $\tan^{-1} 0 = 0$, indicating the range of validity of the expansion.

(b) Find the radius of convergence of the power series $\sum a_n x^n$ where

$$a_n = \frac{c(c+1)(c+2)\ldots(c+n-1)}{n!}$$

(c) Find the solution, in the form of a power series, of the the differential equation $y'' = xy$ that satisfies the conditions $y = 0$, $y' = 0$ when $x = 0$.

[7+7+6]

*****

Indian Statistical Institute
Semester 1 (2006-2007)
B. Stat 1st Year
Mid-semestral Exam
Probability Theory 1

Monday 4.9.2006, 10:30-1:00 Total Points $5 \times 8 = 40$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. An urn contains a white and b black balls. Balls are drawn one by one until those of the same color are left. Show that the probability that they are all white is $a/(a+b)$.

2. An urn contains 5 identical balls of each of five colors V, I, B, G, Y. You will be drawing one ball after another at random without replacement and someone else watching the contents will stop you as soon as balls of only two colors are left. Find the probability that when you are stopped the color combinations left are (V, I) or (I, B) or (G, Y) ignoring order.

3. A publishing company is bringing out 5 different books in each of 4 different areas, mathematics, statistics, operations research and computer science. A library will buy 10 different books from the publisher by choosing 10 books at random from the above collection. Find the probability that the library will get at least two books from each of the four different areas.

4. A person tosses an unbiased coin repeatedly and moves on the upper right quadrant as follows: if the result of a toss is H he moves 1 unit to the right, if it is T he moves 1 unit up. Thus after the kth toss his coordinates are the number of heads and the number of tails upto that point. $4 + 4 = 8$ pts.

(a) If he tosses $2n$ rimes, show that the probability that he does not touch or cross the diagonal is

$$\frac{1 \cdot 2 \cdots (2n-1)}{2 \cdot 4 \cdots 2n}.$$

(b) If after $2n$ tosses he has reached (n, n), find the probability that he did not touch or cross the diagonal before reaching (n, n).

5. Two players each possessing Rs. 2 agree to play a series of games. The probability of winning a single game is 1/2 (which brings the problem into an equally likely set up) and the loser pays Re. 1 to his adversary. Let p_m denote the probability that after playing $2m$ games neither of them is ruined.

(a) Show that $p_{m+1} = \frac{1}{2} p_m$.

(b) Show that P(one of them is ruined at or before the nth game) is $\frac{1}{2} - \frac{1}{2^{m+1}}$ if $n = 2m$ or $n = 2m+1$. $6 + 2 = 8$ pts.

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination

B.Stat/ I year, First Semester, 2006-2007

Statistical Methods

Date: 7.9. 2006 Maximum Marks: 100 Duration: 3 hours.

1. State whether the following statements are True or False. Give brief justifications to your answers. [$4 \times 7 = 28$]

   (a) A stem and leaf plot cannot be used for observations on a continuous variable.

   (b) A box plot of a set of observations can indicate the center and the variability of the distribution of the observations, but cannot give information about the skewness of the distribution.

   (c) For two variables X and Y, when the numerical value of the correlation coefficient r_{xy} increases, the least squares line of Y on X and that of X on Y approach each other.

   (d) A high positive value of the correlation coefficient between X and Y always imply a strong causal relationship between X and Y.

   (e) The correlation coefficient between monthly income X and monthly expenditure (Y) of several individuals over a period of 10 years is found to be r. If the monthly figures are totaled to yearly income Z and yearly expenditure W, then the correlation coefficient between Z and W will be higher than r.

   (f) For a data set on variables X and Y, it is reported that the linear model for Y on X has a negative slope and 36 percent of the total Y variability is not accounted for by X for this model. This implies that $r_{xy} = 0.60$.

   (g) There are 2 data sets, each of size 100, giving the heights (X) and weights(Y) of a sample of 100 students from two different colleges. The least squares(l.s) regression line is computed for each data set separately and they are found to be identical. Now, if the data sets are pooled to form a set of size 200, the l.s. line from this pooled data will be identical with the one obtained earlier.

2. Suppose that in a bivariate data set consisting of 20 data points the means of the variables X (the independent variable) and Y (the dependent variable) are 5.0 and 3.5, and their standard deviations are 2.5 and 1.5 respectively. The correlation coefficient between the two variables is 0.8. Obtain the equation of the straight line fitted to the data by the method of least squares and also the residual sum of squares after that line is fitted to the

data. Suppose next that one data point was wrongly recorded as $(6.2, 2.9)$ while the correct values should be $(5.2, 4.9)$. Obtain the correct equation of the least squares line and the corresponding correct value of the residual sum of squares. $[10 + 15 = 25]$

3. State and prove Chebyshev's inequality. Give an example of a data set where no value differs from the mean by more than one standard deviation. Give an example of a data set where exactly 75% of the data points differ from the mean by one standard deviation or less, and no data point deviates from the mean by more than two standard deviations. $[8 + 7 + 10 = 25]$

4. Consider the data set consisting of the four data points (1.5, 2.5), (3.0, 4.5), (4.0, 5.5) and (5.0, 8.5). Compute the equation of the least absolute deviations line that will fit these ~~three~~ data points. [14]

5. There are two classes, one consisting of 10 students and another consisting of 15 students. You are given the absolute difference in the mathematics scores for all 300 pairs of students. Describe clearly how you would use this data to compute the within class and the between class variations for the mathematics scores for these 25 students. [8]