INDIAN STATISTICAL INSTITUTE

M.Tech (CS) I Year: 2024-25 (Semester II)

Statistical Methods: Class Test 2

Date: 05/04/25 | Time: 60 minutes | Full Marks: 20

Note: Show necessary steps to solve all of the following problems. There is no credit for a solution if the appropriate work is not shown even if the answer is correct.

---

1. Suppose that X has the probability mass function $p(x)$ and we want to test $H_0 : p(x) = p_0(x)$ vs $H_1 : p(x) = p_1(x)$, where

$$p_0(x) = \begin{cases} \frac{1}{20} & \text{if } x = 1, 2, \ldots, 20, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad p_1(x) = \begin{cases} 0.6 & \text{if } x = 1 \\ 0.15 & \text{if } x = 2, 3, \\ 0.1/17 & \text{if } x = 4, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases}$$

For a prefixed $\alpha \in (0, 1)$, state the Neyman-Pearson lemma for finding a most powerful (MP) size-α test for testing H_0 vs H_1. For $\alpha = 0.1$, construct a *non-randomized* MP size-α test, say ϕ_1, for testing H_0 vs H_1. Verify whether the *randomized* test

$$\phi_2(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0.5 & \text{if } x = 2, 3, \\ 0 & \text{otherwise.} \end{cases}$$

is also a most powerful (MP) size-α test for testing H_0 vs H_1. [2+4+3]

2. A machine that automatically controls the length of ribbon on a tape will be judged to be effective if the standard deviation σ of the length of ribbon on a tape is less than or equal to 0.5 cm. It is assumed that length of ribbon on a tape is normally distributed.

   (a) Assume that the true mean length of ribbon is 4 cm. Construct an UMP size-α test to verify whether the machine is ineffective. State appropriate null and alternative hypotheses. [1+4]

   (b) Is the power function for the UMP test increasing or decreasing? [2]

   (c) The sum and sum of squares of the lengths of 40 randomly selected tapes are 197 and 1000 respectively. Based on the UMP test, can we conclude that the machine is ineffective at 5% level of significance? [2]

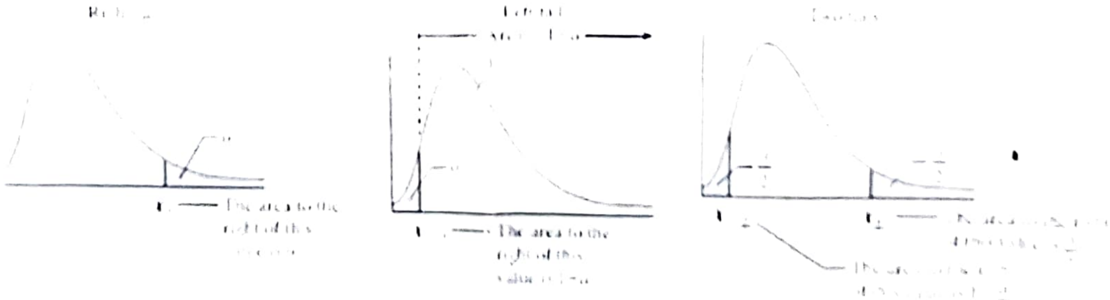
   (d) Compute the p-value of the UMP test and draw conclusion at 15% level of significance. [2]

Table VII

Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169



INDIAN STATISTICAL INSTITUTE

M.Tech (CS) I Year: 2024-25 (Semester II)

Statistical Methods: Mid-Semester Exam

Date: 26/02/2025 | Time: 2:30 PM - 4:30 PM | Full Marks: 30

Note: Show necessary steps to solve all of the following problems. There is no credit for a solution if the appropriate work is not shown even if the answer is correct. The total marks allotted in this question paper is 32. However, if you score, say x marks out of 32 marks, your final score will be counted as $\min\{x, 30\}$ marks. Here, $n > 2$ is a natural number. Use of calculator is allowed in this exam.

---

1. The following data represents interest rates charged by ten credit card companies.

$$8.3,\ 7.9,\ 7.7,\ 8.4,\ 8.3,\ 8.7,\ 8.1,\ 7.8,\ 7.9,\ 9.4$$

   (a) Find the upper, lower fences and outliers (if any) in the dataset. [3]

   (b) Construct a box plot and comment on the skewness of the data. [3+1]

   (c) Suppose the above observations are values of a variable x. Find the Kendall's Tau between x and z where $z = 3 + \frac{1}{x} - 4x^5$. Justify your answer. [2]

2. Suppose $X_1, \ldots, X_n$ are i.i.d. copies of a random variable X with probability density function

$$f(x \mid \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \mathbb{I}_{\{0<x\leq\beta\}}$$

   (a) Find a two-dimensional sufficient statistic for (α, β). [2]

   (b) Derive the maximum likelihood estimator (MLE) of (α, β). [4]

   (c) Find the median m of the distribution of X and compute its MLE. [2]

3. Suppose $X_1, \ldots, X_n$ are i.i.d. observations having common density function

$$f(x \mid \theta) = \frac{4\theta^4}{x^5} \mathbb{I}_{\{x\geq\theta\}}, \text{ where } \theta > 0.$$

   (a) Find the method of moment estimator of θ, say $\widehat{\theta}$. [2]

   (b) If $X_{(1)} = \min\{X_1, \ldots, X_n\}$, then show that $\delta(\mathbf{X}) = \frac{4n-1}{4n} X_{(1)}$ is both unbiased and consistent estimator of θ. [2+2]

   (c) What is the efficiency of $\widehat{\theta}$ relative to $\delta(\mathbf{X})$? [4]

4. Suppose X_1 and X_2 are i.i.d. *Bernoulli*(θ) with probability mass function $p(x|\theta) = \theta^x(1-\theta)^{1-x}$ for $x \in \{0, 1\}$ and $\theta \in (0, 1)$. If $T = 3X_1 - 2X_2$, then argue why the Fisher information $I_T(\theta)$ in T must be less than or equal to $\frac{2}{\theta(1-\theta)}$. State the result you may need to use. [4+1]

INDIAN STATISTICAL INSTITUTE

M.Tech (CS) I Year: 2024-25 (Semester II)

Statistical Methods: Final Exam

Date: 06/05/2025 Time: 2:30 PM - 5:30 PM Full Marks: 50

Note: Show necessary steps to solve all of the following problems. There is no credit for a solution if the appropriate work is not shown even if the answer is correct. The total marks allotted in this question paper is 52. However, the maximum you can score is 50. Suppose $\mathbb{I}_{\{A\}}$ denotes the indicator variable of an event A, and $n > 2$ is a natural number. Use of a calculator is allowed in this exam.

---

1. Suppose $X_1, X_2, \ldots, X_n$ are independently and identically distributed (i.i.d.) samples having the density

$$f(x|\,\theta) = \frac{1}{24\theta^5} x^4 \exp\left\{-\frac{x}{\theta}\right\} \text{ where } x > 0 \text{ and } \theta > 0.$$

   (a) Find the method of moment estimator δ of θ. [2]

   (b) Derive the Cramer-Rao lower bound for the variances of unbiased estimators of θ. [4]

   (c) Is δ the uniformly minimum variance unbiased estimator of θ? Justify. [2]

2. Suppose $X_1, X_2, \ldots, X_n$ are random samples from an exponential distribution with the density

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma}\right\} \mathbb{I}_{\{x \geq \mu\}} \text{ where } -\infty < x < \infty, -\infty < \mu < \infty,\ \sigma > 0.$$

   (a) What is a two-dimensional sufficient statistic for (μ, σ)? [2]

   (b) Find the maximum likelihood estimator (MLE) $(\widehat{\mu}, \widehat{\sigma})$ of (μ, σ). [4]

   (c) Using the above sample, find an unbiased estimator of $\mu + \sigma$. [2]

   (d) Is $\widehat{\sigma}$ a consistent estimator of σ? Justify. [3]

3. Suppose that X has the probability mass function $p(x)$, and we want to test $H_0 : p(x) = p_0(x)$ against $H_1 : p(x) = p_1(x)$ where

$$p_0(x) = \begin{cases} \left(\frac{1}{2}\right)^{x+1} & \text{if } x = 0, 1, 2 \ldots, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad p_1(x) = \begin{cases} \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^x & \text{if } x = 0, 1, 2 \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Based on the single observation X, derive a most powerful test ϕ of size $\alpha = 0.05$ for testing H_0 vs H_1. What is the power of ϕ? [5+2]

4. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. observations from $U(0, \theta)$ for $\theta > 0$.

   (a) For a prefixed $\alpha \in (0, 1)$, derive a non-randomized uniformly most powerful (UMP) test of size α for testing $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ where θ_0 is a known positive number. [5]

   (b) If $X_{(n)} = \max_{1 \leq i \leq n} X_i$, then show that the randomized test defined as

$$\phi_1(X_1, X_2, \ldots, X_n) = \begin{cases} 1 & \text{if} \quad X_{(n)} > \theta_0 \\ \alpha & \text{if} \quad X_{(n)} \leq \theta_0 \end{cases}$$

   is also a UMP test of size α. Is ϕ_1 a consistent test? [3+1]

5. Suppose $X_1, X_2, \ldots, X_n$ are random samples from a Laplace distribution with the density

$$f(x|\theta) = \frac{1}{2\theta} \exp\left\{-\frac{|x|}{\theta}\right\} \quad \text{where} \quad -\infty < x < \infty \text{ and } \theta > 0.$$

   (a) For a prefixed $\alpha \in (0, 1)$, derive the likelihood ratio test (LRT) of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ for some known $\theta_0 > 0$. The rejection region of your test should be derived based on quantiles of a chi-squared distribution. Note that $\frac{2}{\theta}|X_1| \sim \chi^2_2$. [6]

   (b) Derive the p-value of the LRT in part (a) in terms of the cumulative distribution function of a chi-squared distribution. [2]

6. Two sections of a mathematics course were taught by two professors using the same textbook. Suppose professor 1 and professor 2 taught section 1 and section 2 respectively. After the final exam, professors 1 and 2 selected 14 and 16 students randomly from their respective sections and noted that the sample means of exam scores are 75 and 70 respectively. The sample variances of scores in sections 1 and 2 were 9 and 4 respectively. It is assumed that the scores of exams for sections 1 and 2 are independent of each other and are normally distributed with different population means for different sections but common population variance for both sections.

   (a) Derive a 90% two-sided confidence interval for the difference of the means of exam scores of the two sections. [5]

   (b) Does the confidence interval have shortest expected length? Justify. [2]

   (c) Based on the confidence interval in part (a), is there sufficient evidence, at 10% level of significance, in favour of the hypothesis that the population means corresponding to the two sections are different? Justify. [2]