INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2019-2020

M.Stat. First Year

LARGE SAMPLE STATISTICAL METHODS

Maximum Marks: 60

Answer all questions

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

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

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

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

2. If $X_n \xrightarrow{p} 0$, show that for any median M_n of X_n, $M_n \to 0$.

3. Let $X_1, \ldots, X_n$ be a random sample from a distribution with a density $f(x, \theta)$ where θ is an unknown real parameter. Give an example to show that there may exist a consistent estimator T_n of θ for which $E(T_n) \to \theta + 1$ as $n \to \infty$. [Hint: Consider an example where there is an unbiased consistent estimator $\hat{\theta}_n$ and a set A_n, independent of $\hat{\theta}_n$, with probability tending to one and with some other property so that $\hat{\theta}_n$ can be modified with the help of A_n to construct T_n.]

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

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

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

5. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ, variance σ^2 and finite 4th central moment μ_4. A common test for $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 > 1$ rejects H_0 when $nS_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 > \chi^2_{\alpha,n-1}$ where $\chi^2_{\alpha,n-1}$ is the upper α point of a central chi-square distribution with $(n-1)$ degrees of freedom. Show that $P_{H_0}\left(nS_n^2 > \chi^2_{\alpha,n-1}\right)$ converges to α as $n \to \infty$ only if the value of the kurtosis $\kappa = \frac{\mu_4}{\sigma^4} - 3$ is equal to zero. [Hint: First show (using the CLT and Polya's Theorem) that $\left(\chi^2_{\alpha,n-1} - (n-1)\right)/\sqrt{2n-2}$ converges to the upper α point z_α of the standard normal distribution.]

6. *Simulation experiment.* (a) For $n = 20$ and 30, draw 1000 random samples of size n from a bivariate normal distribution $N_2(0, 0, 1, 1, \rho)$ for $\rho = 0.2$, 0.4, 0.6, 0.8. Tabulate the estimated coverage probabilities of the (approximate) confidence intervals for ρ formed using the $\tanh^{-1}$ transformation of the correlation coefficient.

(b) Plot the histogram of the 1000 sample correlation coefficents r_n's for $n = 30$ and $\rho = 0.4$ and also the histogram of the corresponding $\tanh^{-1} r_n$'s.

7. Let $X_1, \ldots, X_n$ be a random sample from a double exponential distribution with density $f(x, \theta) = \frac{1}{2}\exp(-|x - \theta|)$, $-\infty < x < \infty$, where $\theta \in R$ is an unknown location parameter.

(a) Find the maximum likelihood estimator (MLE) of θ and find the asymptotic (non-degenerate) distribution of properly normalized MLE.

(b) Obtain a confidence interval for the poulation interquartile range (i.e., the difference between the third and first quartiles) with confidence coefficient approximately equal to $1 - \alpha$.

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

9. Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, 1)$ where θ is an unknown integer. Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of θ . Prove that

it is not possible to find a sequence of real constants a_n such that $a_n(\hat{\theta}_n - \theta)$ converges to a non-degenerate limit distribution.

10. Let $X_1, \ldots, X_n$ be i.i.d. with a common density $f(x, \theta)$ where $\theta \in \Theta$, and Θ consists of only finitely many real numbers. Assume also that the set $\{x : f(x, \theta) > 0\}$ is the same for all $\theta \in \Theta$ and that the distributions under different θ's are different. If θ_0 is the true value of θ, prove that with probability tending to one (under θ_0) as $n \to \infty$, the likelihood function will be maximized at the value $\theta = \theta_0$.

11. Let $X_1, \ldots, X_n$ be i.i.d. with a common density $f(x, \theta)$ given by

$$f(x, \theta) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2}, \quad -\infty < x < \infty,$$

where $\theta \in R$ is unknown. Assume that Conditions (A1)–(A7) of the section on MLE hold.

(a) Does there exist a consistent maximum likelihood estimator of θ for this problem? Justify your answer.

(b) Can you suggest any sequence of functions $g_n(\cdot)$ (depending on n) of $\bar{X}_n$ such that $g_n(\bar{X}_n)$ (after suitable centering and scaling) converges in distribution to $N(0, 1/I(\theta))$ where $I(\theta)$ is the Fisher Information? Justify your answer.

12. Let $X_1, \ldots, X_n$ be i.i.d. with a common density $f(x, \theta)$ given by

$$f(x, \theta) = e^{-(x-\theta)}, \ x \geq \theta, \ -\infty < \theta < \infty.$$

(a) Find the MLE $\hat{\theta}_n$ of θ and the limiting non-degenerate distribution of suitably normalized $(\hat{\theta}_n - \theta)$.

(b) Given that $\theta \leq 1$, find the MLE $\widetilde{\theta}_n$ of θ and also find the limiting non-degenerate distribution of suitably normalized $(\widetilde{\theta}_n - \theta)$ under $\theta = \theta_0 < 1$.

13. Draw a random sample of size 25 from a Cauchy distribution with density

$$f(x, \theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \quad -\infty < x < \infty,$$

where θ is chosen by you (you may take $\theta = 1$). Use the method of scoring to estimate θ.

14. Consider the set up for the chi-square test for independence of two attributes (see the document "Large Sample Chi-square test 2" sent to you by email). Note that under independence the $k \times l$ cell probabilities can be expressed as functions of $(k-1)+(l-1)$ independent parameters. Find the maximum likelihood estimates of these parameters under independence.