Stats 743 – Foundations of Statistics

Assignment – 1

Date: February 17, 2023 **Instructor: Prof. N. Balakrishnan**

Solve all the problems, and all of them will be graded equally!

---

1. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics from a random sample of size n from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$. Then, show that the sequence of order statistics $\{(X_{r:n}, r = 1, \cdots, n)\}_{n=1,2,\ldots}$ forms a Markov chain.

2. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics from a random sample of size n from a two-point distribution taking on values a and b $(a < b)$ with probabilities π and $1 - \pi$, respectively. Then, prove that the sequence of order statistics $\{(X_{r:n}, r = 1, \cdots, n)\}_{n=1,2,\ldots}$ **forms a Markov chain**.

3. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics from a random sample of size n from a discrete distribution taking on at least 3 distinct values. Then, prove that the sequence of order statistics $\{(X_{r:n}, r = 1, \cdots, n)\}_{n=1,2,\ldots}$ **does not form a Markov chain**.

4. Let $U_{1:n} < U_{2:n} < \cdots < U_{n:n}$ be the order statistics from a random sample of size n from *Uniform*$(0, 1)$ distribution.
   (a) Show that, for $1 \leq r < s \leq n$, the ratio $\dfrac{U_{r:n}}{U_{s:n}}$ and $U_{s:n}$ are statistically independent;
   (b) Identify the distributions of these two variables;
   (c) Using the results in Parts (a) and (b), derive an expression for $\mathrm{Cov}(X_{r:n}, X_{s:n})$ (Hint: You don't need to do two-dimensional integration!)

5. Let $U_{1:n} < U_{2:n} < \cdots < U_{n:n}$ be the order statistics from a random sample of size n from *Uniform*$(0, 1)$ distribution. Then, extend the result in Part (a) of Problem 4 by proving that

$$\frac{U_{1:n}}{U_{2:n}},\ \frac{U_{2:n}}{U_{3:n}},\ \cdots,\ \frac{U_{n-1:n}}{U_{n:n}},\ U_{n:n}$$

   are statistically independent. Identify their distributions.

6. Let $X_1, X_2, \cdots, X_{n+1}$ be independent *Standard Exponential* $(\theta = 1)$ random variables. Further, let $S_{n+1} = \sum_{i=1}^{n+1} X_i$.

(a) Using the Jacobian method, obtain the joint density function of

$$\left(\frac{X_1}{S_{n+1}}, \ \frac{X_1 + X_2}{S_{n+1}}, \ \cdots, \ \frac{X_1 + X_2 + \cdots + X_n}{S_{n+1}} \right);$$

(b) Can you identify this joint density function, and then comment on it?

7. Let $X_1, X_2, \cdots, X_n$ be a random sample from a *power function distribution* with density function $f(x) = \frac{2x}{\theta^2}$, $0 < x < \theta$. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the corresponding order statistics.

   (a) Write down the joint density function of $(X_{1:n}, X_{2:n}, \cdots, X_{n:n})$;

   (b) Show that $X_{n:n}$ is a sufficient statistic for θ;

   (c) Then, find an unbiased estimator of θ in terms of $X_{n:n}$;

   (d) Discuss whether this family is complete, and if so, what can you say about the estimator of θ found in Part (c).

8. In continuation of Problem 7:

   (a) Explain whether $P_n = \frac{X_{n:n}}{\theta}$ could serve as a pivot or not;

   (b) If it could serve as a pivot, explain how you would use it to construct a $100(1-\alpha)\%$ confidence interval for θ;

   (c) Using the mean and variance of $X_{n:n}$, scale the pivot P_n suitably, and then derive its asymptotic (non-degenerate) distribution.

9. Suppose 2 independent random values from $Uniform(0, 1)$ distribution are drawn and are used to form an interval, say (a, b). Further, suppose another 2 independent random values are drawn (independently of the first two drawn) from $Uniform(0, 1)$ distribution and are used to form yet another interval, say (c, d).

   (a) What is the probability that the two intervals intersect?

   (b) What is the probability that the two intervals do not intersect?

10. Let $X_1, \cdots, X_n$ be a random sample from $Uniform(-\theta, +\theta)$ distribution, for $\theta > 0$, and $X_{1:n} < \cdots < X_{n:n}$ be the corresponding order statistics.

    (a) Show that $(X_{1:n}, X_{n:n})$ is jointly sufficient for θ;

    (b) Derive the MLE, $\hat{\theta}$, of θ;

    (c) Derive the exact distribution of $\hat{\theta}$.

11. In continuation of Problem 10:

    (a) From the exact distribution derived in Part (c) of Problem 10, find the expected value and variance of $\hat{\theta}$. What do you observe from these?

(b) Is $(X_{1:n}, X_{n:n})$ complete sufficient for θ? Justify your answer.

12. Let $X_1, \cdots, X_n$ be a random sample from $Poisson(\lambda)$ distribution, with $n > 3$.
    (a) Find the uniformly minimum variance unbiased estimator of $e^{-\lambda}(1 - e^{-\lambda})^2$;
    (b) Derive its variance and comment on it.

13. Let $X_1, \cdots, X_n$ be the lifetimes of n components from $Exponential(\theta)$ distribution, with $n > 2$. Further, let S denote the lifetime of a series system of 2 such components and P denote the lifetime of a parallel system of 2 such components.
    (a) Find the uniformly minimum variance unbiased estimator for the probability that S is larger than a pre-specified time t;
    (b) Find the uniformly minimum variance unbiased estimator for the probability that P is at most a pre-specified time t^*;
    (c) Derive the distributions of these estimators.

14. Let X_1, X_2, X_3 be iid random variables from $Normal(\mu, 1)$ distribution.
    (a) What is the Fisher information in X_i about μ, for $i = 1, 2, 3$?
    (b) What is the Fisher information in (X_i, X_j) about μ, for $i, j = 1, 2, 3$ and $i \neq j$?
    (c) What is the Fisher information in (X_1, X_2, X_3) about μ? Comment.

15. In continuation of Problem 14, let $X_{1:3} < X_{2:3} < X_{3:3}$ be the order statistics obtained from X_1, X_2, X_3.
    (a) Obtain an expression for the Fisher information in $X_{i:n}$ about μ, for $i = 1, 2, 3$;
    (b) Compute those Fisher information measures (you may use any method that you think is suitable for their computation), and comment;
    (c) Obtain expressions for the Fisher information in $(X_{i:n}, X_{j:n})$, for all choices of $1 \leq i < j \leq 3$;
    (d) Computer those Fisher information measures (here again, you may use any method that you think is suitable for their computation), and comment;
    (e) What is the Fisher information in $(X_{1:3}, X_{2:3}, X_{3:3})$? Explain.

16. Suppose $X_1, \cdots, X_n$ is a random sample from $Normal(\mu, \sigma^2)$ distribution. In this case, it is well-known that $\bar{X}$ and S^2 are unbiased estimators of μ and σ^2, resoectively. Now, let z_p denote the upper p-th quantile of the standard normal distribution and ξ_p denote the upper p-th quantile of $Normal(\mu, \sigma^2)$ distribution.
    (a) Then, using $\bar{X}$ and S^2, construct a pivot for ξ_p;
    (b) Using that pivot, explain how you would construct a $100(1 - \alpha)\%$ confidence interval for ξ_p;
    (c) Explain the process of determining the required percentage points of the involved pivot;

(d) For the case of sample size $n = 20$, determine the percentage points corresponding to 90% and 95% confidence intervals for upper 10% and 25% quantiles for $Normal(\mu, \sigma^2)$ distribution. (Hint: You may use any method that you think is suitable for this process!)

17. Let $X_1, \cdots, X_n$ be a random sample from an absolutely continuous distribution with cdf $F(x)$ and pdf $f(x)$, and let ξ_p denote the upper p-th quantile of this distribution. Further, let $X_{1:n} < \cdots < X_{n:n}$ denote the corresponding order statistics. Then, suppose the random interval $(X_{r:n}, X_{s:n})$, for $1 \leq r < s \leq n$, is provided as a $100(1-\alpha)\%$ confidence interval for ξ_p.

    (a) How should (r, s), for a given sample size n, be chosen so that it is a valid $100(1-\alpha)\%$ confidence interval for ξ_p?

    (b) For the case of sample size $n = 20$, determine **all** the choices of (r, s) corresponding to 90% and 95% confidence intervals for upper 10% and 25% quantiles;

    (c) Of those confidence intervals, which ones you would use as best 90% and 95% confidence intervals for upper 10% and 25% quantiles? (Hint: You may use minimum width as a criterion for choosing the best confidence interval!)

18. It will be of interest to compare the parametric confidence interval in Problem 16 with the non-parametric conference interval in Problem 17. For this purpose, you may carry out a Monte Carlo simulation study as follows.

    Simulate a sample of size $n = 20$ from standard normal distribution, without loss of generality. Then, construct 95% confidence intervals for the upper 10% quantile by both approaches. Repeating this process $N = 1000$ times, find the average width as well as the coverage probability for both confidence intervals, and make some comments about their relative performance.

19. Carry out the same process of comparison as in Problem 18, but by simulating the samples from

    (a) *Logistic distribution* with pdf $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \ -\infty < x < \infty$;

    (b) *Laplace distribution* with pdf $f(x) = \frac{1}{2}e^{-|x|}, \ -\infty < x < \infty$;

    (c) *Normal outlier-model* in which 18 observations are from $Normal(0, 1)$ and 2 observations are from $Normal(0, 3^2)$;

    (d) Make comparative comments about the relative performance of the two approaches for confidence intervals based on the results obtained in Parts (a)–(c).

20. Consider the following Darwin's well-known data (Fisher, 1966), which represent the differences in heights between cross- and self-fertilized plants of the same pair grown together in one pot:

    $$49, \ -67, \ 8, \ 16, \ 6, \ 23, \ 28, \ 41, \ 14, \ 29, \ 56, \ 24, \ 75, \ 60, \ -48$$

    Suppose we wish to assume *Normal*(μ, σ^2) distribution for analyzing these data. Then:

    (a) Obtain estimates of μ and σ^2 by assuming normal distribution for the data;

(b) Construct $P - P$ and $Q - Q$ plots for these data, and pass some comments;

(c) Find an alternate estimate of μ by using a 10% trimmed mean (by removing the 2 smallest and 2 largest observations in the data). Compare this estimate with the sample mean you obtained in Part (a), and make some comments (gleaning on your observations from Part (b)).

GOOD LUCK AND HAVE FUN!!
