Stats 743 – Foundations of Statistics

Assignment – 2

Date: April 2, 2023 **Instructor: Prof. N. Balakrishnan**

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

---

1. Suppose $\mathbf{C} = ((c_{ij}))_{i,j=1}^{k}$ is a symmetric non-singular product-decomposable matrix with elements c_{ij} of the form $c_{ij} = a_i b_j$. Then, show that $\mathbf{C}^{-1}$ is a symmetric tri-diagonal matrix with its elements as (for $i \le j$)

$$c^{ij} = \begin{cases} \frac{a_2}{a_1(a_2b_1 - a_1b_2)} & \text{for} \quad i = j = 1 \\ \frac{a_{i+1}b_{i-1} - a_{i-1}b_{i+1}}{(a_ib_{i-1} - a_{i-1}b_i)(a_{i+1}b_i - a_ib_{i+1})} & \text{for} \quad 2 \le i = j \le k-1 \\ \frac{b_{k-1}}{b_k(a_kb_{k-1} - a_{k-1}b_k)} & \text{for} \quad i = j = k \\ -\frac{1}{(a_{i+1}b_i - a_ib_{i+1})} & \text{for} \quad j = i+1 \text{ and } 1 \le i \le k-1 \\ 0 & \text{for} \quad j > i+1 \end{cases}.$$

2. Consider a location family of distributions with density function $f_X(x;\mu) = f(x-\mu)$, for $x, \mu \in \mathbf{R}$. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from this location family of distributions.

   (a) Then, by adopting the Lagrangian multiplier method, derive an expression for the Best Linear Unbiased Estimator of μ;

   (b) Obtain an expression for the variance of that estimator.

3. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from Uniform$(\theta, \theta+1)$ distribution.

   (a) By using the results of the last exercise, derive the Best Linear Unbiased Estimator of θ and its variance;

   (b) Derive the MLE of θ in this case;

   (c) Derive the distribution of the MLE, and then find its mean, variance and mean squared error;

   (d) How does the mean squared error of the MLE of θ compare with the variance of the BLUE of θ determined in Part (a)?

4. Let $X_1, \cdots, X_n$ be a random sample of size n from a population with density function

$$f(x;\mu) = \frac{e^{-(x-\mu)}}{(1+e^{-(x-\mu)})^2}, \quad x, \mu \in \mathbf{R}.$$

Further, let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the corresponding order statistics.

(a) What is a sufficient statistic for μ (and comment);

(b) Set up the maximum likelihood estimation method for the parameter μ, and (through Monte Carlo simulations) simulate the Bias and Mean Squared Error of the MLE of μ based on a sample of size $n = 10$ (use 1,000 simulations);

(c) Based on the order statistics, determine the coefficients of the Best Linear Unbiased Estimator of μ and its variance for a sample of size $n = 10$;

(d) Compare the MSE simulated in Part (b) with the variance determined in Part (c), and comment;

(e) Derive the Fisher information about μ in the sample $X_1, \cdots, X_n$, and compute it for a sample of size $n = 10$. How does it compare with the simulated MSE in Part (b)?

(f) Using either of the two estimators discussed above (namely, MLE or BLUE), explain how you would carry out a test for the hypotheses

$$H_0: \mu = a \quad vs. \quad H_1: \mu \neq a,$$

where a is some specified value. Describe the procedure in detail.

5. Consider a scale family of distributions with density function $f_X(x;\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, for $x \in \mathbf{R}$, $\sigma \in \mathbf{R}^+$. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from this scale family of distributions.

(a) Then, by adopting the Lagrangian multiplier method, derive an expression for the Best Linear Unbiased Estimator of σ;

(b) Obtain an expression for the variance of that estimator.

6. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics obtained from a random sample of size n from Uniform$(0, \theta)$ distribution. Then:

(a) Using the results of the last exercise, derive the Generalized Least Squares Estimator (Best Linear Unbiased Estimator) of θ and its variance;

(b) Make some comments about how it compares with the Uniformly Minimum Variance Unbiased Estimator derived from Lehmann-Scheffé theorem.

7. Let $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} \cdots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ be a random sample from the bivariate normal distribution, $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$.

(a) Does the model belong to the exponential family of distributions?

(b) Find the complete sufficient statistic for the model parameter $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$;

(c) Find the natural parameters of the model.

8. Consider the family of bivariate normal distributions, $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$.

   (a) Does the model belong to the exponential family of distributions?

   (b) Find the minimal sufficient statistic for ρ;

   (c) Prove that the model belongs to a *curved 2-parameter exponential family of distributions*;

   (d) Find the natural parameters of the model.

9. Consider the class of all joint distributions of $\mathbf{X} = (X_1, X_2)$, where $X_1 \sim Bin(n, p)$ and $X_2 \sim Bin(m, p^2)$ and are independent, with parameter $\theta = p$. Assume that m, n are positive integers and $p \in (0, 1)$. Then, prove that the model belongs to a *curved 2-parameter exponential family of distributions*.

10. Let $X_1, \cdots, X_n$ be a random sample from the *Inverse Gaussian Distribution* with PDF

$$f(x \mid \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad 0 < x < \infty.$$

   The family of Inverse Gaussian distributions possesses many important and interesting properties and characteristics, and have found key applications; see Chapter 15 of Johnson, Kotz and Balakrishnan (*Continuous Univariate Distributions - Vol. 1*, John Wiley & Sons, New York, 1994) and Seshadri (*The Inverse Gaussian Distribution*, Oxford University Press, England, 1994).

   (a) Show that the statistics

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad T = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i} - \frac{1}{\bar{X}}}$$

   are sufficient and complete;

   (b) For $n = 2$, show that $\bar{X}$ has an Inverse Gaussian Distribution, $\frac{n\lambda}{T}$ has a χ^2_{n-1} distribution, and that they are independent.

11. Let $X_1, X_2, \cdots, X_n$ be a random sample from Normal(μ, σ^2) distribution. Suppose we wish to test the hypothesis that the largest order statistic $X_{n:n}$ is an outlier, and that we wish to use the test statistic

$$T = \frac{X_{n:n} - X_{n-1:n}}{S},$$

   where S is the sample standard deviation.

(a) Under the null hypothesis that $X_{n:n}$ is not an outlier, show that the statistic T is ancillary;

(b) Then, using Basu's theorem, derive an expression for the mean and variance of T, under H_0;

(c) Using the tables of means, variances and covariances of order statistics (see, for example, H.L. Harter and N. Balakrishnan, *CRC Handbook of Tables for the Use of Order Statistics in Estimation*, CRC Press, Florida, 1996), compute the mean and variance of T for sample size $n = 10$, using the expressions derived in Part (b);

(d) Using 1,000 Monte Carlo simulation runs, simulate the values of mean and variance of T under H_0, and compare them with those determined in Part (c) and comment;

(e) Explain what the critical region will be for the test based on T for testing whether $X_{n:n}$ is a large outlier;

(f) Then, determine the upper 5% critical value for T, for sample size $n = 10$, through Monte Carlo simulations (use 1,000 simulation runs);

(g) With the logarithms of the number of trees in orchards being assumed to be normal, Singh et al. (1982) presented the following observed values of logarithms of the number of trees:

1.7918, 2.3026, 2.7726, 3.2581, 3.5264, 3.8067, 3.9703, 4.0943, 4.2905, 4.5747

Then, using the statistic T, test whether the largest observation 4.5747 is an outlier or not, at 5% level of significance.

12. For the data presented in Part (g) of the last exercise, assuming that *the largest observation was not observed*, and that only the first 9 observations of the 10 were observed (i.e., the largest observation was censored), determine the Best Linear Unbiased Predictor of that unobserved largest observation. When comparing it with the observed value of 4.5747 of that largest order statistic, what do you observe?

13. Suppose $X_1, \cdots, X_n$ are independent random variables with X_i having the density function

$$f(x \mid \theta_i, \eta_i) = h(x \mid \eta_i) \exp\left\{\theta_i T(x) - A(\theta_i, \eta_i)\right\}.$$

Then, show that $V = \sum_{i=1}^{n} T(X_i) - nT(U)$ is independent of $U \equiv U(X_1, \cdots, X_n)$ if the density function of U is of the form

$$g(x \mid \theta_1, \cdots, \theta_n, \eta_1, \cdots, \eta_n) = h^*(x \mid \eta_1, \cdots, \eta_n) \exp\left\{\left(\sum_{i=1}^{n} \theta_i\right) T(x) - A^*(\theta_1, \cdots, \theta_n, \eta_1, \cdots, \eta_n)\right\}$$

for some A^* and h^*.

GOOD LUCK AND HAVE FUN!!