Statistical Inference and Data Analysis

January 11, 2021

1 Closed book part

Question 1

Given is a random variable X with $X \sim \text{Expo}(\theta)$ and a random sample $X_1, ..., X_n$.

1. What is the distribution of $\sum_{i=1}^n X_i$, explain why. Can you make a pivotal quantity for θ by using $\sum_{i=1}^n X_i$?
2. Use part 1) to derive an exact confidence interval of level $(1-\alpha)$ for θ.
3. We are interested in finding the Bayes estimator. As a prior density we use the density of a Gamma(α,β) distribution. Find the posterior distribution.
4. Calculate the Bayes estimator T^B.
5. Give the corresponding credible region for θ and discuss it.

Question 2

1. Proof: $\hat{\beta} \sim N_p(\beta, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$
2. Proof $E(S^2) = \sigma^2$, S^2 was given.
3. Derive the likelihood ratio test for the hypothesis $H_0 : \mathbf{C}\beta = 0$ where $\mathbf{C}$ has r lines. You can use the following as given:
   - $\mathbf{Y}$ is normally distributed
   - (contained the expression of the likelihood function as given on p197)
   - $L(\hat{\beta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2}e^{-n/2}$ (for the complete model) and $L(\hat{\beta}_0, \hat{\sigma_0}^2) = (2\pi\hat{\sigma_0}^2)^{-n/2}e^{-n/2}$ (for the restricted model).
4. Define the F statistic.
5. Give the relation between the F statistic and the likelihood ratio statistic.

2 Open book part

Question 1

Given is the rv. X and a random sample $X_1, ..., X_n$ of X having the following density function:

$$f(x, \theta) = \frac{1}{4n}\theta^{-4}x^7 \exp(-\frac{x^2}{\theta})$$

for $x > 0$ and $\theta > 0$. Also we know that

- $X^2/\theta \sim$ Gamma(4,1)
- $E(X^r) = \frac{1}{6}\Gamma(4 + \frac{r}{2})\theta^{r/2}$
- $\Gamma(m + r/2) = \frac{(2m)!}{m!2^{2m}\sqrt{\pi}}$

1. Show that the MLE for θ is given by:
$$\hat{\theta_n} = \frac{\sum_{i=1}^{n} X_i}{4n}$$
and give its bias, variance and MSE.

2. Discuss the consistency of this MLE.

3. Give the asymptotic normallity result for the MLE.

4. Calculate the Method of Moments estimator (note that it differs from the MLE) and denote it as $\hat{\theta_n}^{MoM}$, give also the asymptotic normality result of $\hat{\theta_n}^{MoM}$ and calculate the ARE between the method of moments estimator and the MLE.

5. Use part 3) to construct an approximate confidence interval for θ.

6. Given the hypothesis: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, find the most powerful test of size α if we say that $\theta_1 < \theta_0$.

Question 2

Given the random vector $X = (X_1, X_2, X_3)^T$ which is normally distributed with mean 0 and variance-covariance matrix:
$$\Sigma_X = \begin{pmatrix} 0.2 & 0 & 0.2 \\ 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 \end{pmatrix}$$

1. Show that the 3 principal components are given by
$$Y_1 = \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3)$$
$$Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$$
$$Y_3 = \frac{1}{\sqrt{3}}(X_1 + X_2 - X_3)$$
Also find their distributions.

2. Given were 3 scree plots and biplots (plot of the first two PCs with arrows for the components of X_1, X_2 and X_3) in , we had to explain which one of the plots belonged to a sample of the random vector X. The same was asked for 3 biplots.