Q.1) The joint distribution of Y_1, the number of contracts awarded to firm A, and Y_2, the number of contracts awarded to firm B, is given by the entries in the following table. Find $\text{Cov}(Y_1 - Y_2, Y_1 + Y_2)$. [10 marks]

		y_1		
		1	2	3
	0	1/9	2/9	1/9
y_2	1	2/9	2/9	0
	2	1/9	0	0

Table 1: The joint probability of Y_1 and Y_2

Solution: We have

$$E(Y_1) = 1(4/9) + 2(4/9) + 3(1/9) = 15/9$$

$$E(Y_1^2) = 1^2(4/9) + 2^2(4/9) + 3^2(1/9) = 29/9$$

$$E(Y_2) = 0(4/9) + 1(4/9) + 2(1/9) = 6/9$$

$$E(Y_2^2) = 0^2(4/9) + 1^2(4/9) + 2^2(1/9) = 8/9$$

Also,

$$\begin{aligned}
\text{Cov}(Y_1 - Y_2, Y_1 + Y_2) &= \text{Cov}(Y_1, Y_1) + \text{Cov}(Y_1, Y_2) + \text{Cov}(-Y_2, Y_1) + \text{Cov}(-Y_2, Y_2) \\
&= V(Y_1) + \text{Cov}(Y_1, Y_2) - \text{Cov}(Y_2, Y_1) - V(Y_2) \\
&= V(Y_1) - V(Y_2) = (29/9 - 15^2/9^2) - (8/9 - 6^2/9^2) \\
&= \frac{29(9) - 15^2 - 8(9) + 36}{81} = \frac{0}{81} = 0
\end{aligned}$$

Q.2) Let the random variable Y possess a uniform distribution on the interval $(0, 1)$. Derive the distribution of the random variable

$$W = Y^2.$$

[10 marks]

Solution: We use the method of CDF.

$$F_W(w) = P(W \leq w) = P(0 < Y^2 \leq w) = P(0 < Y \leq \sqrt{w}) = \int_0^{\sqrt{w}} dy = \sqrt{w}$$

Therefore the pdf is given by

$$f_W(w) = \frac{d}{dw}F_W(w) = \frac{1}{2\sqrt{w}} \quad \text{for} \quad 0 < w < 1.$$

Q.3) Let $Y_1, Y_2, \ldots, Y_n$ be independent, uniformly distributed random variables on the interval $[0, \theta]$. Find the pdf of

$$Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n).$$

[10 marks]

Solution: We have

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) = P(Y_1 \leq y, \ldots, Y_n \leq y) = \prod_{i=1}^{n} P(Y_i \leq y) = \prod_{i=1}^{n} \int_0^y \frac{1}{\theta} dt \\ &= \prod_{i=1}^{n} \frac{y}{\theta} = \frac{y^n}{\theta^n} \end{aligned}$$

The pdf is then given by

$$f_{Y_{(n)}}(y) = \frac{d}{dy}F_{Y_{(n)}}(y) = n\frac{y^{n-1}}{\theta^n} \quad \text{for} \quad 0 \leq y \leq \theta$$

Q.4) An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed 0.5 inch. [10 marks]

Solution:

$$\begin{aligned} P(|\overline{Y}_n - \mu| \leq 0.5) &= P(|Z| \leq \frac{0.5}{2.5/\sqrt{100}}) = P(|Z| \leq 2) = P(Z \leq 2) - P(Z \leq -2) \\ &= P(Z \leq 2) - P(Z \leq -2) = P(Z \leq 2) - 1 + P(Z \geq -2) \\ &= P(Z \leq 2) - 1 + P(Z \leq 2) = 2P(Z \leq 2) - 1 = 2(0.9772) - 1 \\ &= 0.9544 \end{aligned}$$

Q.5) In a Gallup Poll of $n = 800$ randomly chosen adults, 45% indicated that movies were getting better whereas 43% indicated that movies were getting worse. Find a 98% confidence interval for p, the overall proportion of adults who say that movies are getting better. [10 marks]

Solution: The confidence interval is given as follows:

$$\widehat{p} \mp t_{\alpha/2,n-1}\sqrt{\frac{p(1-p)}{n}}$$

Since the sample size is large, we use the following confidence interval.

$$\widehat{p} \mp z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Therefore,

$$0.45 \mp 2.325\sqrt{\frac{0.45(1-0.45)}{800}} \Rightarrow 0.45 \mp 0.04089 \Rightarrow (0.40911, 0.49089)$$

Q.6) Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y|\theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1, \theta > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that

$$\sum_{i=1}^{n} -\ln(Y_i)$$

is sufficient for θ. [10 marks]

Solution: The likelihood is given as

$$L(\theta; y_1, \ldots, y_n) = \prod_{i=1}^{n} \theta y_i^{\theta-1} = \theta^n e^{(\theta-1)\sum_{i=1}^{n} \ln y_i} = \theta^n e^{\theta \sum_{i=1}^{n} \ln y_i} e^{-\sum_{i=1}^{n} \ln y_i}$$

Therefore, $\sum_{i=1}^{n} \ln Y_i$ is a sufficient statistic. Also $-\sum_{i=1}^{n} \ln Y_i$ is a sufficient statistic.

Question 6 continues on the next page. . .

(b) What is the MVUE for θ? Provide a detailed statistical argument for your answer. [10 marks]

Solution: For $X = -\ln Y$ we have

$$f_X(x) = e^{-x}\theta(e^{-x})^{(\theta-1)} = \theta e^{-\theta x} \quad \text{for} \quad x > 0$$

Let $W = 2\theta X$ we then have

$$f_W(w) = \frac{1}{2\theta}\theta e^{-w} = \frac{1}{2}e^{-w} = \frac{1}{2^1\Gamma(1)}w^{1-1}e^{-w}$$

Therefore,

$$W = 2\theta X \sim \chi^2_2$$

Hence,

$$V = \sum_{i=1}^{n} W_i = \sum_{i=1}^{n}(-2\theta \ln Y_i) \sim \chi^2_{2n}$$

We also observe that

$$E[V^{-1}] = \int_0^\infty v^{-1}\frac{1}{2^n\Gamma(n)}v^{n-1}e^{-v/2}dv = \frac{1}{2n-2}$$

Therefore,

$$\mathrm{E}\left[\frac{1}{\sum_{i=1}^{n}(-2\theta \ln Y_i)}\right] = \frac{1}{2n-2}$$

Therefore,

$$\mathrm{E}\left[\frac{n-1}{-\sum_{i=1}^{n}\ln Y_i}\right] = \theta$$

A direct result of Rao-Blackwell theorem guarantees that

$$\frac{n-1}{-\sum_{i=1}^{n}\ln Y_i}$$

is a MVUE.

Q.7) For the geometric probability mass function with probability of success equals p, find the MLE for p. [10 marks]

Solution:

$$L(p; y_1, \dots, y_n) = \prod_{i=1}^{n} p(1-p)^{y_i - 1} = p^n (1-p)^{\sum_{i=1}^{n} y_i - n}$$

The logarithm of likelihood is

$$\ell(p; y_1, \dots, y_n) = n \ln p + \sum_{i=1}^{n} y_i \ln(1-p) - n \ln(1-p)$$

The MLE is given by solving

$$\frac{d}{dp} \ell(p; y_1, \dots, y_n) = 0$$

We have

$$\frac{n}{p} - \frac{\sum_{i=1}^{n} y_i}{1-p} + \frac{n}{1-p} = 0 \Rightarrow n - p \sum_{i=1}^{n} y_i = 0 \Rightarrow p = \frac{n}{\sum_{i=1}^{n} y_i}$$

Hence the MLE is

$$\widehat{p} = \frac{n}{\sum_{i=1}^{n} Y_i} = \frac{1}{\overline{X}_n}$$

Q.8) The output voltage for an electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean 128.6 and standard deviation 2.1. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a test with level 0.05. [10 marks]

Solution: The test statistic is given by

$$\begin{cases} H_0: & \mu = 130 \\ H_a: & \mu < 130 \end{cases}$$

The Null is rejected if

$$t < t_{1-0.05, 40-1} \approx -1.684$$

We have

$$t = \frac{128.6 - 130}{2.1/\sqrt{40}} = -4.21637$$

Since -4.21637 is smaller than -1.684 the null is rejected.

Q.9) Use the method of least squares to fit a straight line to the $n = 5$ data points given in Table 2 below. Y is the response variable and X is the independent variable (covariate). [10 marks]

x	y
-2	0
-1	0
0	1
1	1
2	3

Table 2: Observed values of random variables Y and X.

Solution: We have

$$\bar{x} = 0, \bar{y} = 1$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = (4 + 1 + 0 + 1 + 4) - 0 = 10$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = (0 + 0 + 0 + 1 + 6) - 0 = 7$$

Therefore,

$$\widehat{\beta}_1 = \frac{7}{10} \quad \widehat{\beta}_0 = 1$$

The fitted line is

$$\widehat{y} = 1 + \frac{7}{10}x$$

Copyright Notice: The present document and the contents thereof are the property and copyright of the professor(s) who prepared this exam at Concordia University. No part of the present document may be used for any purpose other than research or teaching purposes at Concordia University. Furthermore, no part of the present document may be sold, reproduced, republished or re-disseminated in any manner or form without the prior written permission of its owner and copyright holder.

Question 9 continues on the next page...