Mathematical Statistics

Assignment 5

Hanbin Liu 11912410

5.1 Solution

$\mathbb{C} = \{1, 7, 3, 8, 4\}$, then we have

$$\begin{aligned}\alpha(0) &= \Pr(X \in \mathbb{C} \mid 0)\\ &= \Pr(X = 1 \mid 0) + \Pr(X = 7 \mid 0) + \Pr(X = 3 \mid 0) + \Pr(X = 8 \mid 0) + \Pr(X = 4 \mid 0)\\ &= 0 + 0.01 + 0.02 + 0.07 + 0.05\\ &= 0.15,\end{aligned}$$

and

$$\begin{aligned}\beta(1) &= \Pr(X \in \mathbb{C}' \mid 0)\\ &= \Pr(X = 5 \mid 1) + \Pr(X = 9 \mid 1) + \Pr(X = 10 \mid 1) + \Pr(X = 6 \mid 1) + \Pr(X = 2 \mid 1)\\ &= 0.03 + 0.02 + 0.04 + 0.01 + 0\\ &= 0.1.\end{aligned}$$

$\square$

5.2 Solution

$$0.1 = p(0.5) = \Pr(Y \geq c \mid 0.5) = \Pr\left(\frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \geq \frac{c - n\theta}{\sqrt{n\theta(1-\theta)}}\Big| 0.5\right) = \Pr\left(Z \geq \frac{2c - n}{\sqrt{n}}\right),$$

where $Z \sim N(0, 1)$. Thus,

$$\frac{2c - n}{\sqrt{n}} = z_{0.01}. \tag{1}$$

$$0.95 = p(\frac{2}{3}) = \Pr(Y \geq c \mid \frac{2}{3}) = \Pr\left(\frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \geq \frac{c - n\theta}{\sqrt{n\theta(1-\theta)}}\Big| \frac{2}{3}\right) = \Pr\left(Z \geq \frac{3c - 2n}{\sqrt{2n}}\right),$$

which implies that

$$\frac{3c - 2n}{\sqrt{2n}} = z_{0.95}. \tag{2}$$

By (1) and (2), we have

$$\begin{cases} n = (3z_{0.1} + 2\sqrt{2}z_{0.05})^2 = 72.2086 \\ c = \frac{1}{2}(3z_{0.1} + 2\sqrt{2}z_{0.05})^2 + \frac{1}{2}z_{0.1}(3z_{0.1} + 2\sqrt{2}z_{0.05}) = 41.5495 \end{cases}$$

Approximately, $n = 72$ and $c = 42$. $\square$

5.3 Solution

(a)

$Y = \sum_{i=1}^{n} X_i \sim \Gamma(2n, \theta)$. The pdf of Y is given by

$$f_Y(y) = \begin{cases} \frac{\theta^{2n}}{\Gamma(2n)} e^{-\theta y} y^{2n-1}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

(b)

$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \theta^{2n} (\prod_{i=1}^{n} x_i) e^{-\theta \sum_{i=1}^{n} x_i}$. Since $\theta_1 > 1$, we have

$$\frac{L(1)}{L(\theta_1)} = \theta_1^{-2n} e^{(\theta_1 - 1)\sum_{i=1}^{n} x_i} \leq k,$$

which is equivalent to

$$\sum_{i=1}^{n} x_i \leq \frac{\log k - 2n \log \frac{1}{\theta_1}}{\theta_1 - 1} = c.$$

Thus, a test φ of size α with critical region

$$\mathbb{C} = \{x : \sum_{i=1}^{n} x_i \leq c\}$$

is the most powerful test for testing H_0 against H_1.
Next, we find c. Under H_0, $Y = \sum_{i=1}^{n} X_i \sim \Gamma(2n, 1)$. Then

$$2Y \sim \Gamma(\frac{4n}{2}, \frac{1}{2}) = \chi^2(4n).$$

Hence,

$$\alpha = \Pr(X \in \mathbb{C} \mid H_0) = \Pr(Y \leq c \mid H_0) = \Pr(2Y \leq 2c \mid H_0).$$

i.e.,$1 - \alpha = \Pr(2Y > 2c \mid H_0)$.
Then, $2c = \chi^2(1-\alpha, 4n)$, $c = \frac{1}{2}\chi^2(1-\alpha, 4n)$.

(c)

$$p(\theta) = \Pr(X \in \mathbb{C} \mid \theta) = \Pr(Y \leq c \mid \theta) = \int_0^c \frac{\theta^{2n}}{\Gamma(2n)} y^{2n-1} e^{-\theta y}\, dy.$$

□

5.4 Solution

(a)

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \theta^n \prod_{i=1}^{n} (1 - x_i)^{\theta - 1},\ 0 < x_i < 1.$$

Then,

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{1}{\theta_1^n \prod_{i=1}^{n} (1 - x_i)^{\theta_1 - 1}} \leq k,$$

which is equivalent to

$$\log \prod_{i=1}^{n} (1-x_i) \geq c.$$

Therefore, a test φ for size α with critical region

$$\mathbb{C} = \{x : \log \prod_{i=1}^{n} (1-x_i) \geq c\}$$

is the most powerful test for testing H_0 against H_1. To determine c, we note that

$$Y_i = -\log(1-X_i) \sim \text{Exponential}(\theta) = \Gamma(1,\theta).$$

Then we have

$$-\log \prod_{i=1}^{n} (1-X_i) = -\sum_{i=1}^{n} \log(1-X_i) = \sum_{i=1}^{n} Y_i \sim \Gamma(n,\theta).$$

Under H_0, $\sum_{i=1}^{n} Y_i \sim \Gamma(n,1)$ and

$$2\sum_{i=1}^{n} Y_i \sim \Gamma(\frac{2n}{2}, \frac{1}{2}) = \chi^2(2n).$$

Therefore,

$$\begin{aligned}\alpha &= \Pr(X \in \mathbb{C} \mid H_0) = \Pr(\log \prod_{i=1}^{n} (1-X_i) \geq c \mid H_0) \\ &= \Pr(-2\log \prod_{i=1}^{n} (1-X_i) \leq -2c \mid H_0) = \Pr(\sum_{i=1}^{n} Y_i \leq -2c \mid H_0).\end{aligned}$$

Equivalently,

$$1-\alpha = \Pr(\chi^2(2n) > -2c).$$

Thus,

$$-2c = \chi^2(1-\alpha, 2n) \Longrightarrow c = -\frac{1}{2}\chi^2(1-\alpha, 2n).$$

(b)

$\Theta_0 = \{1\}$, $\Theta_1 = (0,1) \cup (1,\infty)$ and $\Theta_0 \cup \Theta_1 = \Theta = \Theta^* = (0,\infty)$.

$$L(\theta) = \theta^n \prod_{i=1}^{n} (1-x_i)^{\theta-1},\ l(\theta) = n\log\theta + (\theta-1)\log \prod_{i=1}^{n} (1-x_i),\ l'(\theta) = \frac{n}{\theta} + \log \prod_{i=1}^{n} (1-x_i).$$

The MLE of θ is given by

$$\hat{\theta} = \frac{-n}{\log \prod_{i=1}^{n} (1-x_i)}.$$

Then we have

$$\lambda(x) = \frac{L(1)}{L(\hat{\theta})} = \frac{1}{\hat{\theta}^n (\prod_{i=1}^{n} (1-x_i))^{\hat{\theta}-1}} = \frac{1}{(\frac{-n}{\log \prod_{i=1}^{n} (1-x_i)})^n (\prod_{i=1}^{n} (1-x_i))^{\frac{-n}{\log \prod_{i=1}^{n} (1-x_i)}-1}}.$$

Therefore, the critical region that rejecting H_0 is

$$\mathbb{C} = \{x : \lambda(x) \leq \lambda_\alpha\} = \Big\{x : (\frac{-n}{\log \prod_{i=1}^{n} (1-x_i)})^n (\prod_{i=1}^{n} (1-x_i))^{\frac{-n}{\log \prod_{i=1}^{n} (1-x_i)}-1} \geq c\Big\}.$$

Let

$$Q(x) = \prod_{i=1}^{n}(1-x_i), \quad h(Q) = (-n/\log Q)^n Q^{-n/\log Q - 1}.$$

To determine c, we have

$$\log h(Q) = n\log n - n\log(-\log Q) - n - \log Q,$$

and

$$\frac{d\log h(Q)}{dQ} = \frac{n}{Q(-\log Q)} - \frac{1}{Q} = \frac{n+\log Q}{Q(-\log Q)}.$$

Solving $\frac{d\log h(Q)}{dQ} = 0$ yields that $Q = e^{-n}$. Since $Q = \prod_{i=1}^{n}(1-x_i) \in (0,1)$, it then follows that

$$\begin{cases} \frac{d\log h(Q)}{dQ} = \frac{n+\log Q}{Q(-\log Q)} < 0, & \text{if } Q < e^{-n}, \\ \frac{d\log h(Q)}{dQ} = \frac{n+\log Q}{Q(-\log Q)} > 0, & \text{if } Q > e^{-n}. \end{cases}$$

Therefore, $Q = e^{-n}$ is the minimum of $h(Q)$, and $h(Q)$ is decreasing when $Q < e^{-n}$ and increasing when $Q > e^{-n}$. Thus,

$$\mathbb{C} = \{x : (-n/\log Q)^n Q^{-n/\log Q - 1} \geq c\} = \{x : Q \leq c_1 \text{ or } Q \geq c_2\}.$$

Hence, we need to determine c_1 and c_2. Note that

$$\begin{aligned} \alpha &= \Pr(Q \leq c_1 \text{ or } Q \geq c_2 \mid H_0) \\ &= \Pr(Q(x) \leq c_1 \mid H_0) + \Pr(Q(x) \geq c_2 \mid H_0). \end{aligned}$$

Using equal-tail approach, it then follows that

$$\frac{\alpha}{2} = \Pr(Q(x) \leq c_1 \mid H_0) = \Pr(Q(x) \geq c_2 \mid H_0).$$

Then,

$$\begin{aligned} \frac{\alpha}{2} &= \Pr(Q(x) \leq c_1 \mid H_0) \\ &= \Pr(\log Q(x) \leq \log c_1 \mid H_0) \\ &= \Pr(-2\log Q(x) \geq -2\log c_1 \mid H_0), \end{aligned}$$

which implies that $-2\log c_1 = \chi^2(\frac{\alpha}{2}, 2n)$. i.e.,$c_1 = \exp\{-\frac{1}{2}\chi^2(\frac{\alpha}{2}, 2n)\}$.
Similarly, we can obtain that $c_2 = \exp\{-\frac{1}{2}\chi^2(1-\frac{\alpha}{2}, 2n)\}$. Therefore, a test φ with critical region

$$\mathbb{C} = \left\{x : \prod_{i=1}^{n}(1-x_i) \leq \exp\{-\frac{1}{2}\chi^2(\frac{\alpha}{2}, 2n)\} \text{ or } \prod_{i=1}^{n}(1-x_i) \geq \exp\{-\frac{1}{2}\chi^2(1-\frac{\alpha}{2}, 2n)\}\right\}$$

is the LRT for testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$. □

5.5 Solution

Step 1 We consider to test

$$H_{0s} : \theta = \theta_0 \text{ against } H_{1s} : \theta = \theta_1 (< \theta_0).$$

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = (2\pi)^{\frac{-n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}.$$

The ratio is given by

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{(2\pi)^{\frac{-n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2}}{(2\pi)^{\frac{-n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_1)^2}} = e^{\frac{1}{2}\sum_{i=1}^{n}(\theta_0-\theta_1)(2x_i-\theta_0-\theta_1)}.$$

Then,

$$\begin{aligned}\frac{L(\theta_0)}{L(\theta_1)} \le k &\iff \frac{1}{2}\sum_{i=1}^{n}(\theta_0-\theta_1)(2x_i-\theta_0-\theta_1) \le \log k \\ &\iff 2\sum_{i=1}^{n} x_i \le \frac{2\log k}{\theta_0-\theta_1} + \pi(\theta_0+\theta_1) \\ &\iff \bar{x} \le c.\end{aligned}$$

The critical region is $\mathbb{C} = \{x : \bar{x} \le c\}$. To find c,

$$\alpha = \Pr(x \in \mathbb{C} \mid \theta_0) = \Pr(\bar{x} \le c \mid \theta_0).$$

Since $\bar{X} \sim N(\theta, \frac{1}{n})$, it then follows that

$$\alpha = \Pr(\frac{\bar{x}-\theta_0}{\sqrt{1/n}} \le \frac{c-\theta_0}{\sqrt{1/n}}) = \Pr(z \le \sqrt{n}(c-\theta_0)),$$

which implies that

$$\sqrt{n}(c-\theta_0) = z_{1-\alpha} \Longrightarrow c = \frac{z_{1-\alpha}}{\sqrt{n}} + \theta_0.$$

Thus, a test φ with critical region $\mathbb{C} = \{x : \bar{x} \le \frac{z_{1-\alpha}}{\sqrt{n}} + \theta_0\}$ is the MPT of size α for testing $H_{0s} : \theta = \theta_0$ against $H_{1s} : \theta = \theta_1 (< \theta_0)$.
Step 2 Note that $c = \frac{z_{1-\alpha}}{\sqrt{n}} + \theta_0$ does not depend on the value of θ_1. It only needs $\theta_1 < \theta_0$. Therefore, the test φ is also the UMPT of size α for testing $H_{0s} : \theta = \theta_0$ against $H_1 : \theta < \theta_0$.
Step 3

$$\begin{aligned}\sup_{\theta \ge \theta_0} p_\varphi(\theta) &= \sup_{\theta \ge \theta_0} \Pr(\bar{X} \le \frac{z_{1-\alpha}}{\sqrt{n}} + \theta_0 \mid \theta) \\ &= \sup_{\theta \ge \theta_0} \Pr\Big(\frac{\bar{X}-\theta}{\sqrt{1/n}} \le \frac{\frac{z_{1-\alpha}}{\sqrt{n}} + \theta_0 - \theta}{\sqrt{1/n}}\Big) \\ &= \sup_{\theta \ge \theta_0} \Pr(Z \le z_{1-\alpha} + \sqrt{n}(\theta_0-\theta)) \\ &= \Pr(Z \le z_{1-\alpha} + \sqrt{n}(\theta_0-\theta_0)) \\ &= \Pr(Z \le z_{1-\alpha}) \\ &= \alpha.\end{aligned}$$

Therefore, the test φ is also the UMPT of size α for testing $H_0 : \theta \ge \theta_0$ against $H_1 : \theta < \theta_0$. □

5.10 Solution

n = 556, m = 4, it then follows that

$$\begin{aligned}Q_{n0} &= \sum_{j=1}^{4} \frac{(N_j - np_{j0})^2}{np_{j0}} \\ &= \frac{(315-312.75)^2}{321.75} + \frac{(108-104.25)^2}{104.25} + \frac{(101-104.25)^2}{104.25} + \frac{(32-34.75)^2}{34.75} \\ &= 0.4700 < \chi^2(0.05, 3) = 7.8147.\end{aligned}$$

We cannot reject H_0, so the data are consistent at the size of 0.05 with the null hypothesis. □

5.11 Solution

n = 300, m = 6, it then follows that

$$
\begin{aligned}
Q_{n0} &= \sum_{j=1}^{6} \frac{(N_j - np_{j0})^2}{np_{j0}} \\
&= \frac{(43-50)^2 + (49-50)^2 + (56-50)^2 + (45-50)^2 + (66-50)^2 + (41-50)^2}{50} \\
&= 8.96 < \chi^2(0.05, 5) = 11.07.
\end{aligned}
$$

We cannot reject H_0, so the data are consistent at the size of 0.05 with the null hypothesis. $\square$