Solutions Resit Exam - Statistics 2020/2021

SOLUTION 1:

(a) Under H_0 we have $T(X) \sim \mathcal{N}(0,1)$.

So a test to the level $\alpha = 0.05$ rejects the null hypothesis if either $T(X) > q_{0.975} = 2$ or $T(X) < q_{0.025} = -q_{0.975} = -2$.

That is, reject H_0 if $T(X) \notin [-2, 2]$.

(b) For $\mu = -1$ we have that the statistic $2 + 2\bar{X}_{16}$ is $\mathcal{N}(0,1)$ distributed.

We reject H_0

$$\textbf{if } \; T(X) > 2 \Leftrightarrow 2 + 2\bar{X}_{16} > 0 \quad \textbf{or if} \quad T(X) < -2 \Leftrightarrow 2 + 2\bar{X}_{16} < -4$$

So the power for the true parameter $\mu = -1$ is:

$$P(2 + 2\bar{X}_{16} > 0) + P(2 + 2\bar{X}_{16} < -4) = 0.5 + (1 - 0.99997) \approx 0.5$$

(c) We have

$$\bar{X}_{16} = -2.65 \Leftrightarrow T(X) = -1.3$$

and so we get for the p-value $\mathbf{p}$:

$$\mathbf{p} = P(T(X) > 1.3) + P(T(X) < -1.3) = 2 \cdot P(T(X) > 1.3) = 2 \cdot (1 - 0.9) = 0.2$$

(d) We have:

$$\sqrt{16} \cdot \frac{\bar{X}_{16} - \mu}{\sqrt{4}} \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad 2\bar{X}_{16} - 2\mu \sim \mathcal{N}(0,1)$$

and so

$$P(-q_{0.9} \leq 2\bar{X}_{16} - 2\mu \leq q_{0.9}) = 0.8$$

Solving for μ yields:

$$P\left(\bar{X}_{16} - \frac{1}{2} \cdot q_{0.9} \leq \mu \leq \bar{X}_{16} + \frac{1}{2} \cdot q_{0.9}\right) = 0.8$$

$\bar{X}_{16} = -2.65$ and $q_{0.9} = 1.3$ we get the 80% confidence interval:

$$[-3.3, -2.0]$$

SOLUTION 2:

(a) We have the likelihood

$$L(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = (1-\theta)^{\sum\limits_{i=1}^{n} X_i - n} \cdot \theta^n$$

and the log likelihood:

$$l(\theta) = (\sum_{i=1}^{n} X_i - n) \cdot \log(1-\theta) + n \cdot \log(\theta)$$

Taking the 1st and the 2nd derivative of $l(\theta)$ yields

$$\begin{aligned} l'(\theta) &= -\frac{\sum\limits_{i=1}^{n} X_i - n}{1-\theta} + \frac{n}{\theta} \\ l''(\theta) &= -\frac{\sum\limits_{i=1}^{n} X_i - n}{(1-\theta)^2} - \frac{n}{\theta^2} \end{aligned}$$

We have:

$$l'(\theta) = 0 \Leftrightarrow \ldots \Leftrightarrow \theta = \frac{n}{\sum\limits_{i=1}^{n} X_i}$$

and

$$l''(\theta) < 0 \text{ for all } \theta \in (0,1)$$

It follows

$$\hat{\theta}_{ML} = \frac{n}{\sum\limits_{i=1}^{n} X_i}$$

(b) Use $l''(\theta)$ with $n = 1$ and the relationship

$$I(\theta) = -E[l''(\theta)] = E\left[\frac{X_1 - 1}{(1-\theta)^2} + \frac{1}{\theta^2}\right] = \frac{\frac{1}{\theta} - 1}{(1-\theta)^2} + \frac{1}{\theta^2} = \frac{1}{\theta(1-\theta)} + \frac{1}{\theta^2} = \frac{1}{\theta^2(1-\theta)}$$

(c) Build the joint density (or likelihood) ratio:

$$\lambda(X) := \frac{L_X(0.6)}{L_X(0.4)} = \frac{(1-0.6)^{Y-n} \cdot 0.6^n}{(1-0.4)^{Y-n} \cdot 0.4^n} = \left(\frac{2}{3}\right)^{Y-n} \cdot \left(\frac{3}{2}\right)^n \quad \text{where} \quad Y := \sum_{i=1}^{n} X_i$$

The NP lemma states that the UMP test is of the form that it rejects H_0 if $\lambda(X) < k$. As $\lambda(X)$ is monotone decreasing in Y, we have the equivalence relationship:

$$\lambda(X) < k \quad \Leftrightarrow \quad Y = \sum_{i=1}^{n} X_i > k_0$$

(d) For $n = 1$ and $k_0 = 3$ we reject H_0 if $X_1 > 3$. Thus the UMP test is to the level:

$$\alpha = P(X_1 > 3) = 1 - P(X_1 \leq 3) = 1 - (0.6 + 0.4 \cdot 0.6 + 0.4^2 \cdot 0.6) = 0.064$$

(e) For large n we have:

$$\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0, I(\theta)^{-1}) \Leftrightarrow \sqrt{I(\theta)}\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \sim \mathcal{N}(0, 1)$$

It follows:

$$P(-q_{0.9} \leq \sqrt{I(\theta)}\sqrt{n} \cdot (\hat{\theta}_{ML} - \theta) \leq q_{0.9}) = 0.8$$

$$\Leftrightarrow P(\hat{\theta}_{ML} - q_{0.9}\frac{1}{\sqrt{I(\theta)}\sqrt{n}} \leq \theta \leq \hat{\theta}_{ML} + q_{0.9}\frac{1}{\sqrt{I(\theta)}\sqrt{n}}) = 0.8$$

Here we have: $\hat{\theta}_{ML} = \frac{5}{10} = 0.5$ and the observed Fisher information

$$I(\hat{\theta}_{ML}) = \frac{1}{\hat{\theta}_{ML}^2(1 - \hat{\theta}_{ML})} = \frac{1}{0.125} = 8$$

It follows that an approximate two-sided 80% CI for θ is given by

$$\left[0.5 - 1.3 \cdot \frac{1}{\sqrt{40}}, 0.5 + 1.3 \cdot \frac{1}{\sqrt{40}}\right] = [0.2945, 0.7055]$$

SOLUTION 3: We note the following:

For $x < 0$ we have: $f_{\theta_1,\theta_2}(x) = \frac{1-\theta_1}{\theta_2} \cdot \exp\left\{\frac{x}{\theta_2}\right\}$
For $x \geq 0$ we have: $f_{\theta_1,\theta_2}(x) = \frac{\theta_1}{\theta_2} \cdot \exp\left\{-\frac{x}{\theta_2}\right\}$

(a) We assume that

$$X_{(1)} \leq \ldots \leq X_{(n-K)} < 0 \leq X_{(n-K+1)} \leq \ldots < X_{(n)}$$

The likelihood is then:

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f_{\theta_1,\theta_2}(X_i) = \prod_{i=1}^{n} f_{\theta_1,\theta_2}(X_{(i)})$$

$$= \left(\frac{1-\theta_1}{\theta_2}\right)^{n-K} \cdot \exp\left\{\frac{\sum_{i=1}^{n-K} x_{(i)}}{\theta_2}\right\} \cdot \left(\frac{\theta_1}{\theta_2}\right)^{K} \cdot \exp\left\{\frac{-\sum_{i=n-K+1}^{n} x_{(i)}}{\theta_2}\right\}$$

(b) We obtain the log-likelihood:

$$\begin{aligned} l(\theta_1,\theta_2) &= (n-K)\log\left(\frac{1-\theta_1}{\theta_2}\right) + \frac{\sum_{i=1}^{n-K} x_{(i)}}{\theta_2} + K\log\left(\frac{\theta_1}{\theta_2}\right) + \frac{\sum_{i=n-K+1}^{n} -x_{(i)}}{\theta_2} \\ &= (n-K)\log\left(\frac{1-\theta_1}{\theta_2}\right) + K\log\left(\frac{\theta_1}{\theta_2}\right) - \frac{S}{\theta_2} \qquad \text{(see HINT)} \\ &= K\cdot\log(\theta_1) + (n-K)\cdot\log(1-\theta_1) - n\cdot\log(\theta_2) - \frac{S}{\theta_2} \end{aligned}$$

(c) In θ_1 we have the likelihood:

$$L(\theta_1) = \left(\frac{1-\theta_1}{\theta_2}\right)^{n-K} \cdot \exp\left\{\frac{\sum_{i=1}^{n-K} x_{(i)}}{\theta_2}\right\} \cdot \left(\frac{\theta_1}{\theta_2}\right)^{K} \cdot \exp\left\{\frac{-\sum_{i=n-K+1}^{n} x_{(i)}}{\theta_2}\right\}$$

The factorization theorem with

$$\begin{aligned} h(x) &= \exp\left\{\frac{\sum_{i=1}^{n-K} x_{(i)}}{\theta_2}\right\} \cdot \exp\left\{\frac{-\sum_{i=n-K+1}^{n} x_{(i)}}{\theta_2}\right\} = \exp\left\{\frac{-S}{\theta_2}\right\} \\ g(K,\theta_1) &= \left(\frac{1-\theta_1}{\theta_2}\right)^{n-K} \cdot \left(\frac{\theta_1}{\theta_2}\right)^{K} \end{aligned}$$

implies that K is a sufficient statistic for θ_1

(d)

$$P(X_i > 0) = \int_0^\infty \frac{\theta_1}{\theta_2}\cdot\exp\left\{-\frac{x}{\theta_2}\right\}\,dx = \frac{\theta_1}{\theta_2}\cdot\left(-\theta_2\exp\left\{-\frac{\infty}{\theta_2}\right\} + \theta_2\exp\left\{-\frac{0}{\theta_2}\right\}\right) = \theta_1$$

(e) As we have a random sample of size n, it follows from part (d):

$$E[K] = \sum_{i=1}^{n}\theta_1 = n\cdot\theta_1$$

(f) Take the derivatives w.r.t. θ_1:

$$\begin{aligned} l'(\theta_1) &= \frac{K}{\theta_1} - (n-K)\cdot\frac{1}{1-\theta_1} \\ l''(\theta_1) &= -\frac{K}{\theta_1^2} - (n-K)\cdot\frac{1}{(1-\theta_1)^2} \end{aligned}$$

Setting the first derivative equal to zero, yields:

$$l'(\theta_1) = 0 \Leftrightarrow K(1-\theta_1) - (n-K)\cdot\theta_1 = 0 \Leftrightarrow \theta_1 = \frac{K}{n}$$

And as $l'(\theta_1) < 0$ for all θ_1, we indeed have a maximum, so that $\hat{\theta}_{1,ML} = \frac{K}{n}$.

To check whether the ML estimator is unbiased, we compute:

$$E\left[\hat{\theta}_{1,ML}\right] = E\left[\frac{K}{n}\right] = \frac{E[K]}{n} = \frac{n \cdot \theta_1}{n} = \theta_1$$

where we used the result from exercise part (e).

(g) We re-use the 2nd derivative of the log-likelihood from part (f) and we recall from part (e) that $E[K] = n \cdot \theta_1$. For $n = 1$ we get:

$$I(\theta_1) = -E\left[-\frac{K}{\theta_1^2} - (1-K)\frac{1}{(1-\theta_1)^2}\right] = \frac{E[K]}{\theta_1^2} + \frac{1-E[K]}{(1-\theta_1)^2} = \frac{1}{\theta_1} + \frac{1}{1-\theta_1} = \frac{1}{\theta_1(1-\theta_1)}$$

END OF SOLUTIONS