THE UNIVERSITY OF NEW SOUTH WALES

DEPARTMENT OF STATISTICS

MID SESSION TEST - 2017 -Tuesday, 5th September (Week 7)

MATH5905

Time allowed: 75 minutes

1. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be i.i.d. Bernoulli (θ) random variables (that is,

$$f(x, \theta) = \theta^x (1-\theta)^{1-x}, x = \{0, 1\}; \theta \in (0, 1)).$$

   a) Justify that that $T = \sum_{i=1}^n X_i$ is sufficient and complete for θ.

   b) Derive the UMVUE of $h(\theta) = \theta^2$. Justify each step in your answer.

   c) Calculate the Cramer-Rao bound for the minimal variance of an unbiased estimator of $h(\theta) = \theta^2$. Does the variance of the UMVUE of $h(\theta)$ attain this bound? Give reasons.

   d) Find the MLE $\hat{h}$ of $h(\theta)$. Justify your answer.

   e) When testing $H_0 : \theta \le 0.6$ versus $H_1 : \theta > 0.6$ with a 0-1 loss in Bayesian setting with the prior $\tau(\theta) = 12\theta^2(1-\theta)$, what is your decision when $T = 5$ and $n = 8$. (You may use: $\int_0^{0.6} x^7(1-x)^4 dx = 0.00011$.)
   Note: The continuous random variable X has a beta density f with parameters $\alpha > 0$ and $\beta > 0$ if $f(x; \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in (0, 1)$. It holds: $E(X) = \frac{\alpha}{\alpha+\beta}$. Here

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx, B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \Gamma(\alpha+1) = \alpha\Gamma(\alpha).$$

2. Let $X_1, X_2, \dots, X_n$ be independent random variables, with a density

$$f(x; \theta) = \begin{cases} \frac{\theta}{x^2}, x > \theta, \\ 0 \text{ else} \end{cases}$$

   where $\theta > 0$ is an unknown parameter. If $Z_n = X_{(1)}$, (i.e., the minimal of the n observations) then

   a) Argue that Z_n is a sufficient statistic for θ.

   b) Show that the density of Z_n is

$$f_{Z_n}(z) = \begin{cases} \frac{n\theta^n}{z^{n+1}}, z > \theta, \\ 0 \text{ else} \end{cases}$$

   (**Hint:** You may find the cdf first by using $P(X_{(1)} > x) = P(X_1 > x \cap X_2 > x \dots \cap X_n > x)$.)

   c) Find the MLE of θ. Justify your answer.

   d) (*) Given that $X_{(1)}$ is complete for θ, find the UMVUE of θ.

Solution to Question 1

Part a). 4 marks

Approach 1: Using property of one parameter exponential family (see p25 of lecture notes).

Indeed, we observe that,

$$\begin{aligned} f(x,\theta) &= \theta^x(1-\theta)^{1-x} \\ &= (1-\theta)\exp\left(x\log\left(\frac{\theta}{1-\theta}\right)\right). \end{aligned}$$

Thus, Bernoulli belongs to one parameter exponential family. This then implies $T=\sum_{i=1}^n X_i$ is (minimal) sufficient, and complete.

Marking criteria:

- 2 marks for notifying that binomial belongs to a one parameter exponential family.
- 1 mark for making the conclusion that the test statistic is sufficient.
- 1 mark for notifying that the test statistic is complete.

Approach 2: Lehmann and Scheffe's method (see p23 of lecture notes) and definition of completeness (see p38 of lecture notes).

We calculate the proportion of the joint density

$$\begin{aligned} \frac{L(X,\theta)}{L(Y,\theta)} &= \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{\prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i}} \\ &= \theta^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i}(1-\theta)^{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i}. \end{aligned}$$

This is independent of θ iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Thus, $T=\sum_{i=1}^n X_i$ is (minimal) sufficient.

To show it is also complete, we see that

$$\begin{aligned} \mathbb{E}_\theta(g(T)) &= \sum_{k=0}^n \binom{n}{k}\theta^k(1-\theta)^{n-k}g(k) \\ &= (1-\theta)^n\sum_{k=0}^n \binom{n}{k}\left(\frac{\theta}{1-\theta}\right)^k g(k) \\ &= (1-\theta)^n\sum_{k=0}^n \binom{n}{k}\eta^k g(k) = 0, \end{aligned}$$

where $\eta=\frac{\theta}{1-\theta}$, holds for all $\theta\in(0,1)$, and all $\eta\in(0,\infty)$ only if implies $g(k)=0$ for all $k=0,\dots,n$. Notice that $\binom{n}{k}=\frac{n!}{k!(n-k)!}>0$, $1-\theta>0$, and η^k is a polynomial. As a consequence, we have $\mathbb{P}(g(T)=0)=1$.

Marking criteria:

- 1 mark for calculating the proportion.
- 1 mark for proving the sufficiency.
- 1 mark for using the definition of completeness.

- 1 mark for proving the test statistic is complete.

Approach 3: First Investigation (see p21 of lecture notes)

By definition of sufficient statistics, it is enough to show that

$$\mathbb{P}\Big(X_1 = x_1, ..., X_n = x_n \Big| \sum_{i=1}^{n} X_i = k\Big) = \frac{\theta^k(1-\theta)^{n-k}}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

is independent of θ, which is obvious. For completeness, see Approach 2.

Marking criteria:

- 1 mark for using the definition of sufficiency.
- 1 mark for proving the test statistic is sufficiency.
- 1 mark for using the definition of completeness.
- 1 mark for proving the test statistic is complete.

Part b). 4 marks

Since T is complete and sufficient, we need to find an unbiased estimator of τ and apply Theorem of Lehmann-Scheffe (see p38 lecture notes). Let

$$\tau = X_1 X_2, \tag{1}$$

we see that

$$\mathbb{E}(\tau) = \Big(\mathbb{E}(X_1)\Big)^2 = \theta^2, \tag{2}$$

which is unbiased. Now, we apply Theorem of Lehmann-Scheffe, this then yields

$$\begin{aligned}
a(k) &= \mathbb{E}\Big(X_1X_2\Big|\sum_{i=1}^{n} X_i = k\Big) \\
&= \mathbb{P}\Big(X_1 = 1, X_2 = 1\Big|\sum_{i=1}^{n} X_i = k\Big) \\
&= \frac{\mathbb{P}\Big(X_1 = 1, X_2 = 1, \sum_{i=1}^{n} X_i = k\Big)}{\mathbb{P}\Big(\sum_{i=1}^{n} X_i = k\Big)} \\
&= \frac{\mathbb{P}\Big(\sum_{i=3}^{n} X_i = k-2\Big)\theta^2}{\mathbb{P}\Big(\sum_{i=1}^{n} X_i = k\Big)} \\
&= \frac{\binom{n-2}{k-2}\theta^k(1-\theta)^{n-k}}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} \\
&= \frac{k(k-1)}{n(n-1)} \\
&= \bar{X}(\bar{X} - \frac{1}{n})\frac{n}{n-1}.
\end{aligned}$$

Marking criteria:

- 1 mark for selecting an estimator.

- 1 mark for proving the estimator is unbiased.
- 1 mark for applying Theorem of Lehmann-Scheffe.
- 1 mark for getting the final UMVUE.

Part c). 4 marks

we can calculate the Cramer-Rao bound as:

$$\begin{aligned} Var_\theta(a(k)) &\geq \frac{\left(\frac{\partial}{\partial\theta}\theta^2\right)^2}{-\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log L(X,\theta)\right)} \\ &\geq \frac{4\theta^2}{-\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}\right)} \\ &\geq \frac{4\theta^2}{-\mathbb{E}_\theta\left(\frac{\sum_{i=1}^n x_i}{\theta^2} + \frac{n-\sum_{i=1}^n x_i}{(1-\theta)^2}\right)} \\ &\geq 4\theta^2\frac{\theta(1-\theta)}{n}. \end{aligned}$$

Next, we see that

$$V(X,\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n-\sum_{i=1}^n x_i}{(1-\theta)} = \frac{n}{\theta^2(1-\theta)}\left(\bar{X}\theta - \theta^2\right).$$

and, $\bar{X}\theta$ is not a statistics so the bound is not attainable.

Marking criteria:

- 2 marks for calculating the Cramer-Rao bound.
- 2 marks for make the right conclusion.

Part d). 2 marks

We first calculate the log-likelihood:

$$\begin{aligned} \log\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} &= \log\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= \sum_{i=1}^n x_i\log\theta + (n-\sum_{i=1}^n x_i)\log(1-\theta). \end{aligned}$$

Differentiating with respect to θ and set this to zero yields:

$$\sum_{i=1}^n x_i\frac{1}{\theta} - (n-\sum_{i=1}^n x_i)\frac{1}{(1-\theta)} = 0.$$

Thus, we have the MLE $\hat{\theta}_{MLE} = \bar{X}$. By the invariance property,

$$\hat{h} = (\frac{1}{n}\sum_{i=1}^n x_i)^2 = (\bar{X})^2.$$

Marking criteria:

- 1 mark for getting the MLE for θ.
- 1 mark for using the invariance property.

Part e). 3 marks

We note that, from p17 of lecture notes,

$$\begin{aligned}
P(\theta \leq 0.6) &= \int_0^{0.6} h(\theta|X)d\theta \\
&= \int_0^{0.6} \frac{f(X|\theta)\tau(\theta)}{\int_0^1 f(X|\theta)\tau(\theta)d\theta}d\theta \\
&= \int_0^{0.6} \frac{\theta^7(1-\theta)^4}{B(8,5)}d\theta \\
&= 0.4356 < 0.5
\end{aligned} \tag{3}$$

(note the threshold for 0-1 loss is 0.5). Thus, we reject the H_0.

Marking criteria:

- 1 mark for getting the expression for $P(\theta \leq 0.6)$.
- 1 mark for getting the correct probability.
- 1 mark for making the correct conclusion.

Solution to Question 2

Part a). 3 marks

We first calculate the joint density

$$L(X, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{\theta^n}{\prod_{i=1}^{n} x_i^2} I_{(\theta,\infty)}(X_{(1)}).$$

Thus, Z_n is sufficient by the Neyman Fisher Factorization Criterion (see p22 of lecture notes).

Marking criteria:

- 1 mark for writing the joint.
- 1 mark for making a correct calculation.
- 1 mark for noting the Neyman Fisher Factorization Criterion.

Part b). 3 marks

We first calculating the survival function.

$$\begin{aligned}
1 - F_{Z_n}(z) &= \mathbb{P}(X_{(1)} > z) \\
&= \mathbb{P}(X_1 > z, ..., X_n > z) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i > z) \\
&= \frac{\theta^n}{z^n}.
\end{aligned}$$

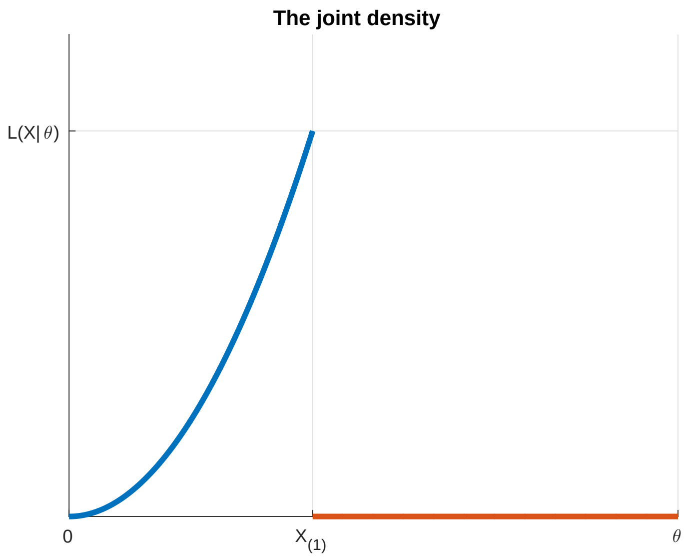

Figure 1: Graph of Joint Density

Then, we obtain the density

$$f_{Z_n}(z) \quad = \quad \frac{n\theta^n}{z^{n+1}} I_{(\theta,\infty)}(z).$$

Marking criteria:

- 2 marks for calculating the survival function (or the cdf).
- 1 mark for obtaining the density via differentiation.

Part c). 3 marks

We first calculate the joint:

$$L(X, \theta) \quad = \quad \frac{\theta^n}{\prod_{i=1}^n x_i^2} I_{(\theta,\infty)}(X_{(1)}),$$

Thus, the likelihood will be maximized at $\theta = \min(x_1, ..., x_n) = X_{(1)}$. See also Figure 1.

Marking criteria:

- 1 mark for calculating the joint.
- 1 mark for giving the correct mle.
- 1 mark for giving the correct reasoning.

Part d). 2 marks

We first check if this is a unbiased estimator.

$$\mathbb{E}(X_{(1)}) = \int_\theta^\infty z \frac{n\theta^n}{z^{n+1}} dz = \frac{n}{n-1}\theta.$$

The UMVUE is then given by

$$\mathbb{E}(\frac{n-1}{n} X_{(1)} | X_{(1)}) = \frac{n-1}{n} X_{(1)}.$$

Marking criteria:

- 1 mark for checking that $X_{(1)}$ is not unbiased and needs a bias-correction.
- 1 mark for finding the UMVUE.