THE UNIVERSITY OF NEW SOUTH WALES

DEPARTMENT OF STATISTICS

MID SESSION TEST - 2016 -Tuesday, 6th September (Week 7)

MATH5905

Time allowed: 75 minutes

1. A preliminary test of a possible carcinogenic compound can be performed by measuring the mutation rate of microorganisms exposed to the compound. An experimenter places the compound in $n = 24$ petri dishes and records the following number of mutant colonies $(X_1, X_2, ..., X_{24})$:

   6, 5, 5, 3, 1, 1, 2, 3, 5, 2, 5, 3, 3, 1, 3, 3, 1, 3, 3, 1, 3, 5, 3, 1

   The number $X_i, i = 1, 2, , \dots, 24$ of mutant colonies is believed to follow a Poisson distribution $(P_\lambda(X_i = x) = f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \dots)$, with $\lambda > 0$ being an unknown parameter.

   a) Use any argument to show that $T = \sum_{i=1}^n X_i$ is complete and sufficient for λ.

   b) Suggest a simple unbiased estimator W of the parameter $\tau(\lambda) = \lambda e^{-\lambda} = P(X_1 = 1)$ (i.e., the probability that exactly one colony will emerge). Hence (or otherwise) derive the UMVUE of $\tau(\lambda)$.

   c) What is the MLE $\widehat{\tau(\lambda)}$ of $\tau(\lambda)$? Explain.

   d) For the given data, find the point estimates of the probability $\tau(\lambda)$ by applying the methods in (b) and (c). Compare the two numerical values.

   e) Is the variance of the UMVUE of $\tau(\lambda)$ equal to the variance given by the Cramer-Rao lower bound for an unbiased estimator of $\tau(\lambda)$? Explain.

   f) Using the prior $\tau(\lambda) = \frac{8}{15}\lambda^5 e^{-2\lambda}, \lambda > 0$ test the hypothesis $H_0 : \lambda \leq 3$ versus the alternative $H_1 : \lambda > 3$ using Bayesian hypothesis testing with a zero-one loss. Formulate your conclusion.

   Hint: You may use the following numerical values: $76!/26^{77} = 210.125, \int_0^3 e^{-26x}x^{76}dx = 117.708$. The Gamma density with parameters $\alpha > 0$ and $\beta > 0$ is

   $$g(x) = \begin{cases} \frac{e^{-\frac{x}{\beta}} x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{else} \end{cases}$$

   The mean and the variance are $\alpha\beta$ and $\alpha\beta^2$, respectively.

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

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

   where $\theta > 0$ is an unknown parameter. If $Z_n = X_{(n)}$, then

   a) Argue that Z_n is a sufficient statistic for θ.
   b) Find the density of Z_n (**Hint:** find the cumulative distribution function of Z_n first).
   c) Assuming that Z_n is also complete, find the UMVUE of the parameter θ as a function of Z_n.
   d) Find the MLE of θ.

Q1.) a) $f(x,\lambda)=e^{-\lambda}\frac{1}{x!}e^{\ln\lambda \cdot x}$ is an one-parameter exponential family with $a(\lambda)=e^{-\lambda}$, $b(x)=\frac{1}{x!}$, $c(\lambda)=\ln\lambda$, $d(x)=x$. Hence $T=\sum_{i=1}^n d(X_i)=\sum_{i=1}^n X_i$ is sufficient and complete for λ. Completeness was also shown in the lectures from first principles: we know that $T\sim Po(n\lambda)$. Assuming that for certain $g(T)$ we have

$E_\lambda g(T)=0\ \forall\lambda>0$, implies

$$\underbrace{e^{-n\lambda}}_{\neq 0}\cdot\sum_{t=0}^{\infty}g\frac{(t)n^t\lambda^t}{t!}=0\quad\forall\lambda>0.$$

This implies that coefficients in front of each power of λ are 0: $g\frac{(t)n^t}{t!}=0\ \forall\ t=0,1,2,\dots$. Hence (since $n^t\neq 0$, $t!\neq 0$) we have $g(t)=0\ \forall t=0,1,2,\dots$, that is $P(g(T)=0)=1$.

b) We take $W=I_{\{X_1=1\}}(\cancel{X})$ as an initial unbiased estimator. Indeed: $EW=1*P(X_1=1)=\lambda e^{-\lambda}=\tau(\lambda)$

Lehmann-Scheffe Theorem tells us that $E(W|T=t)$ is the UMVUE. Hence we have:

$$E(W|T=t)=1*P(W=1|T=t)=\frac{P(W=1\cap T=t)}{P(T=t)}=\frac{P(X_1=1\cap\sum_{i=1}^n X_i=t)}{P(\sum_{i=1}^n X_i=t)}=$$

$$=\frac{P(X_1=1\cap\sum_{i=2}^n X_i=t-1)}{P(\sum_{i=1}^n X_i=t)}=\frac{\lambda\cancel{e^{-\lambda}}\cancel{e^{-(n-1)\lambda}}(n-1)^{t-1}\lambda^{t-1}}{\cancel{e^{-n\lambda}}n^t\lambda^t(t-1)!/t!}=$$

$$=\frac{t}{n}\left(1-\frac{1}{n}\right)^{t-1}=\boxed{\bar{X}\left(1-\frac{1}{n}\right)^{n\bar{x}-1}}\quad\text{being the UMVUE}$$

c) For Poisson data we have: $L(X,\lambda)=\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}$

$\frac{\partial}{\partial\lambda}\ln L(X,\lambda)=-n+\frac{\sum_{i=1}^n X_i}{\lambda}=V(X,\lambda)\Rightarrow V(X,\lambda)=0$ implies that $\hat{\lambda}=\bar{X}$ is MLE of λ.

Then using the invariance property of the MLE we have $\widehat{\lambda e^{-\lambda}}_{MLE} = \boxed{\bar{X} e^{-\bar{X}}}$ being the MLE of $\tau(\lambda)$.

d) Here $\sum_{i=1}^{24} X_i = 71 \to \bar{X} = \frac{71}{24} = 2.958$

The MLE : $\widehat{\lambda e^{-\lambda}}_{MLE} = 2.958 \exp(-2.958) = \underline{0.1536}$

The UMVUE : $2.958\left(\frac{23}{24}\right)^{70} = \underline{0.1504}$ so both estimators are very close.

e) Since $V = -n + \frac{\sum_{i=1}^n X_i}{\lambda} = \frac{ne^{\lambda}}{\lambda}\left(\bar{X}e^{-\lambda} - \lambda e^{-\lambda}\right)$

and $\bar{X}e^{-\lambda}$ is not an estimator (involves the unknown parameter) $\longrightarrow$ CR Bound is not attainable.

f) We need to calculate

$$P(\lambda \le 3/\underline{X}) = \int_0^3 h(\lambda/\underline{X})\, d\lambda$$ and compare it to the threshold of 1/2.

Now $h(\lambda/\underline{X}) \propto \lambda^{71+5} e^{-(24+2)\lambda} = \lambda^{76} e^{-26\lambda}$

identifies $h(\lambda/\underline{X})$ as Gamma $(77, \frac{1}{26})$, and $\Gamma(77) = 76!$

Hence we need to calculate $\frac{26^{77}}{76!}\int_0^3 e^{-26x} x^{76} dx = \frac{117.708}{210.125} = .56$

Since this is $> \frac{1}{2}$, we accept H_0

Q21 a) We can write $f(x;\theta) = \frac{2x}{\theta^2} I_{(0,\theta)}(x)$

Then $L(\underline{X};\theta) = \frac{2^n \prod_{i=1}^n X_i}{\theta^{2n}} \prod_{i=1}^n I_{(0,\theta)}(X_i) = \frac{2^n}{\theta^{2n}} \prod_{i=1}^n X_i \cdot I_{(0,\theta)}(X_{(n)})$

Hence we can factorize $L(\underline{X};\theta) = g(\theta, X_{(n)}) \cdot h(\underline{X})$ with, e.g.,

$g(\theta, X_{(n)}) = \frac{1}{\theta^{2n}} I_{(0,\theta)}(X_{(n)})$ and $h(\underline{X}) = 2^n \prod_{i=1}^n X_i$

Hence $X_{(n)}$ is sufficient

b) First, we find the cdf of a single observation via integration of the density:

$$F_{X_1}(x) = \begin{cases} 0 & \text{if } x < \theta \\ \left(\frac{x}{\theta}\right)^2 & \text{if } 0 < x < \theta \\ 1 & \text{if } x > \theta \end{cases}$$

Then $F_{Z_n}(x) = P(X_{(n)} \le x) = P(X_1 \le x \cap X_2 \le x \cap \dots \cap X_n \le x)$

$$= \begin{cases} 0 & \text{if } x < 0 \\ (x/\theta)^{2n} & \text{if } 0 < x < \theta \rightarrow \text{(we multiply because of indep.)} \\ 1 & \text{if } x > \theta \end{cases}$$

Then the density is the derivative of the cdf $\Rightarrow$

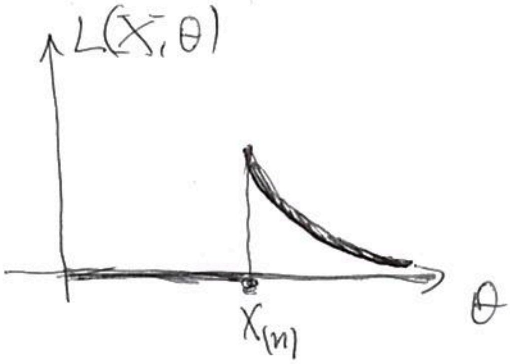
Hence $f_{Z_n}(x) = \begin{cases} \frac{2n x^{2n-1}}{\theta^{2n}} & \text{if } 0 < x < \theta \\ 0 & \text{else} \end{cases}$

c) We first calculate the expected value of Z_n:

$$E Z_n = \int_0^\theta x \frac{2n x^{2n-1}}{\theta^{2n}} dx = \frac{2n}{2n+1} \frac{\theta^{2n+1}}{\theta^{2n}} = \frac{2n}{2n+1}\theta$$

This shows that Z_n is biased but can be easily bias-corrected. We get

$\boxed{\frac{2n+1}{2n} Z_n}$ being unbiased and a function of complete and sufficient statistic → hence Lehmann-Scheffe $\Rightarrow$ it is the UMVUE of θ.

d) 

$$L(X, \theta) = \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} I_{(0,\theta)}(X_{(n)})$$

is 0 before $X_{(n)}$ and is monotonically decreasing after $X_{(n)}$ as a function of θ. Hence it is maximized at $X_{(n)} = Z_n$, that is, Z_n is the MLE.