BSc and MSci EXAMINATIONS (MATHEMATICS)

May–June 2016

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science.

M3S1/M4S1

Statistical Theory I

Date: Tuesday, 10th May 2016 Time: 9:30 – 12:00

Solutions

1. (a) *Neyman Factorization Criterion:* Suppose that $X = (X_1, ..., X_n)$ has a joint distribution $f_\theta(x)$. Then $T = T(X)$ is a sufficient statistic for θ if and only if $f_\theta(x) = g\,(T(x), \theta)\,h(x)$. $\boxed{\text{seen} \Downarrow}$ $\boxed{2}$

Proof: First suppose that T is sufficient. Then

$$\begin{aligned} f_\theta(x) &= P_\theta(X = x) = \sum_t P_\theta(X = x, T = t) = P_\theta(X = x, T = T(x)) \\ &= P_\theta(T = T(x))P_\theta(X = x | T = T(x)) = g(T(x), \theta)h(x) \end{aligned}$$

Now, suppose that $f_\theta(x) = g\,(T(x), \theta)\,h(x)$. Then, if $T(x) = t$,

$$\begin{aligned} P_\theta(X = x | T = t) &= \frac{P_\theta(X = x)}{P_\theta(T = t)} = \frac{P_\theta(X = x)}{\sum\limits_{T(y)=t} P_\theta(X = y)} \\ &= \frac{g(T(x), \theta)h(x)}{\sum\limits_{T(y)=t} g(T(y), \theta)h(y)} = \frac{h(x)}{\sum\limits_{T(y)=t} h(y)} \end{aligned}$$

which does not depend on θ. If $T(x) \neq t$, then $P_\theta(X = x | T = t) = 0$. In both cases, $P_\theta(X = x | T = t)$ is independent of θ and so T is sufficient. $\boxed{3}$

(b) (i) According to Neyman Factorization Criterion, since

$$f_\theta(x) = \prod_{i=1}^n \theta^{x_i}(1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1 - \theta)^{n - \sum_{i=1}^n x_i},$$

$T = \sum_{i=1}^n X_i$ is a sufficient statistic for θ. $\boxed{2}$

Yes, T is complete because Bernoulli distribution is a member of "full rank" exponential family of distributions since $f_\theta(x) = \exp\left(\ln\left(\frac{\theta}{1-\theta}\right) \sum_{i=1}^n x_i + n \ln(1 - \theta)\right)$. $\boxed{2}$

Also, T is minimal sufficient because it is a complete and sufficient statistic. $\boxed{1}$

(ii) An unbiased estimator of $\theta(1 - \theta)$ is $I(X_1 = 0, X_2 = 1)$. By the Lehmann- $\boxed{\text{sim. seen} \Downarrow}$

Scheffe Theorem, the UMVUE of $\theta(1-\theta)$ is

$$\begin{aligned} E(I(X_1=0,X_2=1)|T=t) &= P(X_1=0,X_2=1|\sum_{i=1}^{n}X_i=t) \\ &= \frac{P(X_1=0,X_2=1,\sum_{i=1}^{n}X_i=t)}{P(\sum_{i=1}^{n}X_i=t)} \\ &= \frac{P(X_1=0,X_2=1,\sum_{i=3}^{n}X_i=t-1)}{P(\sum_{i=1}^{n}X_i=t)} \\ &= \frac{\theta(1-\theta)\binom{n-2}{t-1}\theta^{t-1}(1-\theta)^{n-2-t+1}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{\binom{n-2}{t-1}}{\binom{n}{t}} = \frac{t(n-t)}{n(n-1)}. \end{aligned}$$

Therefore, $\frac{\sum_{i=1}^{n}X_i\left(n-\sum_{i=1}^{n}X_i\right)}{n(n-1)}$ is the UMVUE of $\theta(1-\theta)$. $\boxed{5}$

(iii) The Cramer-Rao lower bound here is

$$\frac{\left(\frac{d}{d\theta}\left(\theta(1-\theta)\right)\right)^2}{I(\theta)} = \frac{(1-2\theta)^2}{nI_{X_1}(\theta)} = \frac{(1-2\theta)^2}{\frac{n}{\theta(1-\theta)}} = \frac{\theta(1-\theta)\,(1-2\theta)^2}{n}.$$

$\boxed{3}$

Only estimators of the form $\left\{a\sum_{i=1}^{n}X_i+b\right\}$ achieve the Cramer-Rao lower bound. So the variance of the UMVUE of $\theta(1-\theta)$ does not attain the lower bound. $\boxed{2}$

$\boxed{\text{meth seen} \Downarrow}$

2. (a) The log-likelihood function is

$$l(\mu_1,\ldots,\mu_n,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(x_i-\mu_i)^2+\sum_{i=1}^{n}(y_i-\mu_i)^2\right\}$$

and the MLEs of the parameters are

$$\widehat{\mu}_i = \frac{X_i+Y_i}{2} \qquad \text{and} \qquad \widehat{\sigma}^2 = \frac{1}{n}\left\{\sum_{i=1}^{n}\left(X_i-\frac{X_i+Y_i}{2}\right)^2+\sum_{i=1}^{n}\left(Y_i-\frac{X_i+Y_i}{2}\right)^2\right\}.$$

$\boxed{6}$

(b) (i) Since $Z_i \overset{\text{i.i.d.}}{\sim} N(0, 2\sigma^2)$, the log-likelihood function based on $Z_1, \ldots, Z_n$ is

$$l(\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{4\sigma^2}\sum_{i=1}^{n} z_i^2$$

and then the MLE of σ^2 is $\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} Z_i^2$. $\boxed{3}$

MLEs are consistent under the regularity conditions. Because normal distribution satisfies the regularity conditions, the MLE $\widehat{\sigma}^2$ of σ^2 is consistent. $\boxed{2}$

(ii) Since $E(Z^2) = 2\sigma^2$, method of moments (MM) estimator of σ^2 is

$$2\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 \qquad \Rightarrow \qquad \widehat{\sigma}^2_{MM} = \frac{1}{2n}\sum_{i=1}^{n} Z_i^2.$$

$\boxed{2}$

(iii) The family $\{N(0, 2\sigma^2) : \sigma^2 > 0\}$ has monotone likelihood ratio in $\sum_{i=1}^{n} z_i^2$. Using the Karlin-Rubin Theorem, the UMP test at level α is

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} z_i^2 \geq k \\ 0 & \text{if } \sum_{i=1}^{n} z_i^2 < k \end{cases}$$

where k can be chosen so that

$$\alpha = P_{\sigma_0^2}\left(\sum_{i=1}^{n} Z_i^2 \geq k\right) = P\left(\chi^2(n) \geq \frac{k}{2\sigma_0^2}\right).$$

We then get $k = 2\sigma_0^2\chi^2_\alpha(n)$. $\boxed{5}$

(iv) The power of the UMP test obtained in (iii) is non-decreasing in θ because of the monotone likelihood ratio property. So the UMP test here is an unbiased test since its power is not less than α. $\boxed{2}$

meth seen $\Downarrow$

3. (a) Applying the Bayes theorem, we can write the posterior distribution as follows

$$\pi(\theta|x) = \frac{\lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^n (x_i-2)}}{\int_0^\infty \lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^n (x_i-2)}d\theta} = \frac{\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{\int_0^\infty \theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}d\theta}$$
$$= \frac{\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{c}$$

where c is a constant which does not depend on θ. Because the posterior distribution is proportional to $\theta^n e^{(-K\theta)}$, where K is $\sum_{i=1}^n (x_i - 2) + \lambda$, the posterior is $Gamma(n+1, K)$. In fact $\theta|x \sim Gamma\left(n+1, \sum_{i=1}^n (x_i-2)+\lambda\right)$.
Alternatively, one can obtain the above posterior distribution straightforwardly by calculation of the integral in the denominator (i.e., the constant c) as follows:

$$\pi(\theta|x) = \frac{\lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^n (x_i-2)}}{\int_0^\infty \lambda e^{-\lambda\theta}\theta^n e^{-\theta\sum_{i=1}^n (x_i-2)}d\theta} = \frac{\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{\int_0^\infty \theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}d\theta}$$
$$= \frac{\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{\frac{1}{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)} E\left(\left(\text{Exponential}\left(\sum_{i=1}^n (x_i-2)+\lambda\right)\right)^n\right)}$$
$$= \frac{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{\frac{n!}{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)^n}}$$
$$= \frac{\left(\sum_{i=1}^n (x_i-2)+\lambda\right)^{n+1}\theta^n e^{-\theta\left(\sum_{i=1}^n (x_i-2)+\lambda\right)}}{n!}$$

which is again $Gamma\left(n+1, \sum_{i=1}^n (x_i-2)+\lambda\right)$. $\boxed{8}$

(b) Yes, because both the prior and the posterior are Gamma distributions. Note that exponential distribution is a special case of Gamma distribution. $\boxed{3}$

(c) Under the squared error loss, the Bayes estimator is the posterior mean. Because the posterior is Gamma distribution, we can easily obtain

$$\widehat{\theta}_{\text{Bayes}} = \frac{n+1}{\sum_{i=1}^n (x_i-2)+\lambda}.$$

The above Bayes estimator can alternatively be obtained as follows:

$$\begin{aligned}
\widehat{\theta}_{\text{Bayes}} &= E(\theta|x) = \int_0^\infty \theta \left(\frac{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)^{n+1} \theta^n e^{-\theta\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)}}{n!} \right) d\theta \\
&= \frac{\left(\sum_{i=1}^{n}(x_i-2)+\lambda\right)^{n}}{n!} E\left(\left(Exponential(\sum_{i=1}^{n}(x_i-2)+\lambda)\right)^{n+1}\right) \\
&= \frac{n+1}{\sum_{i=1}^{n}(x_i-2)+\lambda}.
\end{aligned}$$

6

(d) Because the Bayes estimator obtained in (c) is unique, therefore it is admissible. 3

meth seen ⇓

4. (a) Under the whole parameter space, the MLEs of θ_1 and θ_2 are $\widehat{\theta}_{1MLE} = \frac{1}{\bar{X}}$ and $\widehat{\theta}_{2MLE} = \frac{1}{\bar{Y}}$, respectively. And under H_0, the MLEs of θ_1 and θ_2 are

$$\left(\widehat{\theta}_1\right)_0 = \left(\widehat{\theta}_2\right)_0 = \frac{m+n}{m\bar{X}+n\bar{Y}}.$$

Hence, the likelihood ratio statistic is as follows

$$\begin{aligned}
\lambda(x,y) &= \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \frac{L\left(\left(\widehat{\theta}_1\right)_0, \left(\widehat{\theta}_2\right)_0\right)}{L\left(\widehat{\theta}_{1MLE}, \widehat{\theta}_{2MLE}\right)} \\
&= \frac{\left(\left(\widehat{\theta}_1\right)_0\right)^m e^{-\left(\widehat{\theta}_1\right)_0 \sum_{i=1}^{m} x_i} \left(\left(\widehat{\theta}_1\right)_0\right)^n e^{-\left(\widehat{\theta}_1\right)_0 \sum_{i=1}^{n} y_i}}{\left(\widehat{\theta}_{1MLE}\right)^m e^{-\widehat{\theta}_{1MLE} \sum_{i=1}^{m} x_i} \left(\widehat{\theta}_{2MLE}\right)^n e^{-\widehat{\theta}_{2MLE} \sum_{i=1}^{n} y_i}} \\
&= \left(\frac{m}{m+n} + \frac{n}{m+n}\frac{\bar{Y}}{\bar{X}}\right)^{-m} \left(\frac{n}{m+n} + \frac{m}{m+n}\frac{\bar{X}}{\bar{Y}}\right)^{-n}.
\end{aligned}$$

7

(b) We know the likelihood ratio test rejects H_0 for small values of $\lambda(x,y)$. Now, because $\lambda(x,y)$ depends only on $T = \frac{\bar{X}}{\bar{Y}}$ and we can make $\lambda(x,y)$ small by making T small or T large, so a test based on $T = \frac{\bar{X}}{\bar{Y}}$ would reject H_0 for small or large values of T. In fact, a level α test based on $T = \frac{\bar{X}}{\bar{Y}}$ rejects $H_0 : \theta_1 = \theta_2$ if $T \le c_1$ or $T \ge c_2$, where c_1 and c_2 can be chosen so that $P_{H_0}(T \le c_1) + P_{H_0}(T \ge c_2) = \alpha$, where the distribution of $T = \frac{\bar{X}}{\bar{Y}}$, under H_0, is $F(2m, 2n)$. By considering equal tails of the F distribution, we can reject H_0 if $T \le F_{1-\alpha/2}(2m, 2n)$ or $T \ge F_{\alpha/2}(2m, 2n)$. 5

(c) Under H_0 and under regularity conditions, the asymptotic distribution of $-2\log(\lambda(x,y))$ is $\chi^2(1)$. $\boxed{2}$

The likelihood ratio level α test based on the asymptotic distribution rejects H_0 if $-2\log(\lambda(x,y)) \leq \chi^2_{1-\alpha}(1)$. $\boxed{2}$

(d) From (b), we have

$$P_{H_0:\theta_1=\theta_2}\left(F_{1-\alpha/2}(2m,2n) \leq \frac{\theta_1\bar{X}}{\theta_2\bar{Y}} \leq F_{\alpha/2}(2m,2n)\right) = 1-\alpha,$$

and hence using the connection between confidence intervals and hypothesis tests, a confidence interval for $\frac{\theta_1}{\theta_2}$ with confidence coefficient $1-\alpha$ is $\left(\frac{\bar{Y}}{\bar{X}}F_{1-\alpha/2}(2m,2n), \frac{\bar{Y}}{\bar{X}}F_{\alpha/2}(2m,2n)\right)$. $\boxed{4}$

5. (a) The likelihood equation is $\boxed{\text{part seen} \Downarrow}$

$$S(\widehat{\theta}) = \sum_{i=1}^{n} \frac{2(x_i - \widehat{\theta})}{1 + \left(x_i - \widehat{\theta}\right)^2} = 0.$$

$\boxed{2}$

Because $S(\theta)$ is not monotone in θ, the equation $S(\widehat{\theta}) = 0$ may have more than one solution for given sample $x_1,...,x_n$. $\boxed{2}$

(b) Using the Newton-Raphson method a new estimate is given by

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{S(\widehat{\theta}^{(k)})}{H(\widehat{\theta}^{(k)})}$$

where $S(\theta)$ is given in (a) and

$$H(\theta) = 2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{\left(1+(x_i-\theta)^2\right)^2}.$$

$\boxed{4}$

(c) The Fisher scoring algorithm gives a new estimate as follows

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{S(\widehat{\theta}^{(k)})}{H^*(\widehat{\theta}^{(k)})}$$

where $H^*(\theta) = E(H(\theta))$. Considering the hint, we get

$$\begin{aligned} H^*(\theta) &= E\left(2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{\left(1+(x_i-\theta)^2\right)^2}\right) = 2n\int_{-\infty}^{\infty} \frac{1-(x_i-\theta)^2}{\pi\left(1+(x_i-\theta)^2\right)^3}dx_i \\ &= \frac{4n}{\pi}\int_{0}^{\infty} \frac{1-(x_i-\theta)^2}{\left(1+(x_i-\theta)^2\right)^3}dx_i = \frac{4n}{\pi}\left(\frac{\pi}{8}\right) = \frac{n}{2}, \end{aligned}$$

and hence

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} + \frac{4}{n}\sum_{i=1}^{n} \frac{x_i - \widehat{\theta}^{(k)}}{1+\left(x_i-\widehat{\theta}^{(k)}\right)^2}.$$

$\boxed{5}$

(d) Because $E(X_i)$ is not well-defined, the sample mean may not be a good initial estimate of θ. $\boxed{2}$
Since the density of the X_is is symmetric around θ, one may use the sample median as an initial estimate. $\boxed{2}$

(e) The convergence of the Newton-Raphson algorithm is often faster (when both algorithms converge) because it uses the observed Fisher information rather than the expected Fisher information which needs integral calculation. $\boxed{3}$