THE UNIVERSITY OF MELBOURNE

Semester 2 Final Exam Solution, 2016

School of Mathematics and Statistics

MAST20005 Statistics

Writing time:

Reading time:

This is NOT an open book exam

This paper consists of 4 pages (including this page)

Question 1 (12 marks)

(a) Let $x_{(1)} = \min_i x_i$. The likelihood is

$$L(\theta, \lambda) = \prod_i \theta e^{-\theta(x_i-\lambda)} I(x_i \geq \lambda) = \theta^n e^{-\theta \sum_i (x_i - \lambda)} I(x_{(1)} \geq \lambda) \quad (1)$$

(b) First consider maximising $L(\theta, \lambda)$ in λ for given θ. Since $L(\theta, \lambda)$ is nondecreasing in λ for $\lambda \leq x_{(1)}$, the MLE for lambda is $\hat{\lambda} = x_{(1)}$. Second set $\lambda = \hat{\lambda}$ and solve

$$0 = \frac{\partial}{\partial \theta} L(\theta, \hat{\lambda}) = \frac{n}{\theta} - \sum_i (x_i - x_{(1)}). \quad (2)$$

Therefore the MLE of θ is $\hat{\theta} = \sum_i (x_i - x_{(i)})/n$.

(c) The likelihood function can be written as

$$L(\theta, \lambda) = \theta^n e^{-\theta n \overline{x} + \theta \lambda} I(x_{(1)} \geq \lambda) = \phi(\overline{x}, x_{(i)}, \theta, \lambda) h(x_1, \ldots, x_n) \quad (3)$$

with $h(x_1, \ldots, x_n) = 1$. Thus the factorisation theorem implies that $(\overline{X}, x_{(i)})$ is sufficent for (θ, λ).

(d) Note that $\partial^2 \log L(\theta, \lambda)/\partial \theta^2 = -1/\theta^2$. Thus the CR-LB is θ^2/n.

(e) The maximum likelihood estimate is $\hat{\theta} = 4.94$ and an approximate 99% confidence interval is $\hat{\pm} \theta 2.57 \hat{\theta}/\sqrt{n}$ or $(0.92, 8.96)$.

Question 2 (10 marks)

(a) Note that

$$EY = \int_0^\theta y n y^{n-1} \theta^{-n} dy = \frac{n}{n+1} \left. y^{n+1} \right|_0^\theta = \frac{n}{n+1} \theta$$

Thus, $(n+1)Y/n$ is an unbiased estimator for θ.

(b) Note that for $c < \theta$ we have

$$P(Y < c) = \int_0^c \theta n y^{n-1} \theta^{-n} dy = \left. \frac{y^n}{\theta^n} \right|_0^c = (c/\theta)^n.$$

This implies

$$P(\theta \alpha^{1/n} < Y < \theta) = P(Y < \theta) - P(Y < \theta \alpha^{1/n}) = 1 - \alpha.$$

The aboe probability statement is equivalent to $1 - \alpha = P(Y < \theta < Y/\alpha^{1/n})$. Hence, a $100(1-\alpha)\%$ confidence interval for θ is $(y, y/\alpha^{1/n})$.

(c) Since $y = 9.4$ a 95% confidence interval is $(y, y/0.05^{1/5}) = (9.4, 17.1)$.

Question 3 (12 marks) The following data are the lead concentration (μg/l) in eight samples:

```
17.0   21.4   30.6 5.0   12.2   11.8   17.3   18.8
```

corresponding to sample mean $\overline{x} = 16.76$ and sample standard deviation $s = 7.57$.

(a) $\hat{m} = \hat{\pi}_{0.5} = (17.3 + 17)/2 = 17.15$.

(b) There are 3 negative differences $x_i - 15$:

```
> x-15
[1]    2.0    6.4   15.6 -10.0   -2.8   -3.2    2.3    3.8
```

meaning that 5 out of 8 observations provide evidence against H_0. Let $Y \sim \text{Binomial}(n = 8, p = 0.5)$ pvalue is then computed as $P(Y \geq 5) = 1 - P(Y \leq 4) = 1 - 0.64 = 0.36 > 0.05$. Therefore, we cannot reject H_0 at the 0.05 level of significance.

(c) Note that

$$P(Y_2 < m < Y_7) = \sum_{k=2}^{6} \binom{8}{k} (1/2)^k (1/2)^{8-k} = 0.96 - 0.04 = 0.92$$

Therefore a 92% confidence interval for m is $(11.8, 21.4)$.

(d) A 95% CI is $\overline{x} \pm 1.96 s/\sqrt{n}$, i.e. $(10.43, 23.09)$, which is quite close but larger than the nonparametric confidence interval. This additional uncertainty may be due to the slight skewness of the empirical distribution of the observations.

Question 4 (10 marks)

(a) The posterior distribution of θ is proportional to

$$\frac{\alpha x_0^\alpha}{\theta^{\alpha+1}} I(\theta \geq x_0) \prod_i \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{\alpha x_0^\alpha}{\theta^{\alpha+1}} I(\theta \geq x_0) \prod_i \frac{1}{\theta} I(M \leq \theta) = \frac{\alpha x_0^\alpha}{\theta^{n+\alpha+1}} I(\theta > \max\{x_0, M\})$$

(b) Note that the above expression has the same form of the prior distribution $h(\theta)$ with new parameters $\alpha^* = \alpha + 1$ and $x_0^* = \max\{M, x_0\}$ up to a constant not depending on θ. Thus the posterior pdf must take the form

$$\frac{(\alpha + n) x_0^{*(\alpha+n)}}{\theta^{n+\alpha+1}}, \theta \geq x_0^*.$$

(c) The posterior mean is

$$E[\theta | x_1, \ldots, x_n] = (\alpha + n) \max\{x_0, M\} / (\alpha + n - 1)$$

(d) $E[\theta | x_1, \ldots, x_n] = (5 + 15) \max\{20, 30\} / (5 + 15 - 1) = 20(30)/19 = 31.58$.

Question 5 (5 marks)

(a) $n > 2^2 1.96^2 p(1-p)/(0.1)^2$. So if $p = 1/2$, the lecturer needs at least $n = 384.16 \approx 385$ observations.

(b) If $p < 3/4$, $n > 2^2 1.96^2 (3/4)(1/4)/(0.1)^2 = 288.12$ or $n \geq 289$.

Question 6 (7 marks) Suppose that among males aged 50-59 the Prostate Specific Antigen (PSA) follows a $N(\mu, 1)$ distribution (i.e. $\sigma = 1$ is assumed known). Consider a a test for the null hypothesis $H_0 : \mu = 4$ rejecting H_0 when $|Z| = \sqrt{n}|\overline{X} - 4| > 1.645$.

(a) $\alpha = P(|Z| > 1.645) = P(-1.645 \leq N(0,1) \leq 1.645) = 0.90$.

(b)

$$\begin{aligned}
1 - P(|Z| > 1.645|\mu = 5) &= 1 - P(4 - 1.645/3 < \overline{X} < 4 + 1.645/3) \\
&= 1 - P(-1 - 1.645/3 < \overline{X} - 5 < -1 + 1.645/3) \\
&= 1 - (P(N(0,1) < -1 + 1.645/3) - P(N(0,1) < -1 - 1.645/3)) \\
&= 1 - 0.3257 + 0.06077 = 0.7350
\end{aligned}$$

Question 7 (8 marks)

(a) Recall that $S_X^2/\sigma^2 \sim \chi^2(n-1)$ and $S_Y^2/\sigma^2 \sim \chi^2(m-1)$. The hint implies $E[S_X^2] = \sigma^2$ and $E[S_Y^2] = \sigma^2$. Then

$$E[S^2_{\text{pool}}] = \frac{(n-1)E[S_X^2] + (m-1)E[S_Y^2]}{n+m-2} = \frac{(n-1)\sigma^2 + (m-1)\sigma^2}{n+m-2}$$

(b) Note that the variance function $\phi(\alpha) = Var[\alpha S_X^2 + (1-\alpha)S_Y^2] = \alpha^2 Var(S_X^2) + (1-\alpha)^2 Var(S_Y^2)$ is minimized at

$$\alpha = \frac{1/Var(S_X^2)}{1/Var(S_X^2) + 1/Var(S_Y^2)}$$

By the hint note that $Var[S_X^2 \sigma^2] = 2(n-1)$ or $Var[S_X^2] = 2(n-1)\sigma^4$. Similarly, $Var[S_Y^2] = 2(m-1)\sigma^4$. Thus, $\alpha = (n-1)/(n+m-2)$ and $1 - \alpha = (m-1)/(m+n-2)$.

Question 8 (6 marks) The first moment equation is $\overline{X} = E(X) = \dfrac{r(1-p)}{p}$. The second moment equation is

$$E(X^2) = \frac{r(1-p)}{p^2} - \frac{r^2(1-p)^2}{p^2} = \frac{r(1-p)}{p}\left(\frac{1}{p} - \frac{r(1-p)}{p}\right) = \overline{X}(1/p - \overline{X}).$$

Solving in p gives

$$\tilde{p} = \frac{\overline{X}}{\overline{X^2} - (\overline{X})^2}$$

where $\overline{X^2} = \sum_i X_i^2/n$. Replace $\tilde{p}$ in the first equation and get

$$\tilde{r} = \frac{p}{(1-p)}\overline{X} = \frac{\overline{X}^2}{\overline{X^2} - (\overline{X})^2 - \overline{X}}.$$