Zhe Wang
University of Edinburgh
School of Mathematics
Course: Incomplete Data Analysis

Assignment - 2
ID: S2055676
Term: Fall 2020
Due Date: 22^{nd} Nov, 2020

Problem 1
(As I have been struggling with typing align* (so many formulas) in R markdown, so my question 1 is here typed in LaTeX, and the remaining questions are typed in rmd.)

Solution:

1.

(a) Cumulative distribution function is

$$F(y;\theta) = 1 - e^{-\frac{y^2}{2\theta}},$$

Therefore, p.d.f. is given by $f(y;\theta) = F'(y;\theta) = \frac{1}{\theta}e^{-\frac{y^2}{2\theta}}$
Survival function $S(y) = P(Y > y) = \int_y^\infty f(y;\theta) = 1 - F(y;\theta) = e^{-\frac{y^2}{2\theta}}$
Therefore, the likelihood function is

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \left\{ [f(y_i;\theta)]^{r_i} [S(C;\theta)]^{1-r_i} \right\} \\
&= \prod_{i=1}^{n} \left\{ [\frac{1}{\theta}e^{-\frac{y_i^2}{2\theta}}]^{r_i} [e^{-\frac{C^2}{2\theta}}]^{1-r_i} \right\} \\
&= \frac{1}{\theta^{\sum_{i=1}^n r_i}} e^{-\frac{1}{2\theta}\sum_{i=1}^n y_i^2 r_i} e^{-\frac{1}{2\theta}\sum_{i=1}^n c^2(1-r_i)} \\
&= \frac{1}{\theta^{\sum_{i=1}^n r_i}} e^{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2},
\end{aligned}$$

The corresponding log likelihood is

$$l(\theta) = \log L(\theta) = -\textstyle\sum_{i=1}^n R_i \log\theta - \frac{\sum_{i=1}^n X_i^2}{2\theta}$$

We thus have

$$U = \frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta) = -\frac{\sum_{i=1}^n R_i}{\theta} + \frac{\sum_{i=1}^n X_i}{2\theta^2},$$

let $U = 0$, we have

$$\widehat{\theta}_{\mathrm{MLE}} = \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n R_i}.$$

(b) We have

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta) = \frac{\sum_{i=1}^n r_i}{\theta^2} - \frac{\sum_{i=1}^n X_i^2}{\theta^3}.$$

The fisher information is

$$\begin{aligned}
I(\theta) &= -E\left(\frac{\sum_{i=1}^n r_i}{\theta^2} - \frac{\sum_{i=1}^n X_i^2}{\theta^3}\right) \\
&= -\frac{1}{\theta^2}nE(R) + \frac{1}{\theta^3}nE(X^2)
\end{aligned}$$

Note that R is a binary random variable and so

$$\begin{aligned}
E(R) &= 1 \times \Pr(R=1) + 0 \times \Pr(R=0) \\
&= \Pr(R=1) \\
&= \Pr(Y \leq C) \\
&= F(C;\theta) \\
&= 1 - e^{-\frac{y^2}{2\theta}}.
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \int_0^\infty X_i^2 f(x;\theta)\mathrm{d}x \\
&= \int_0^C X_i^2 f(x;\theta)\mathrm{d}x + \int_C^\infty X_i^2 f(x;\theta)\mathrm{d}x \\
&= \int_0^C y^2 f(y;\theta)\mathrm{d}y + \int_C^\infty C^2 f(x;\theta)\mathrm{d}x \\
&= -C^2 e^{-\frac{C^2}{(2\theta)}} + 2\theta\left(1 - e^{-\frac{C^2}{(2\theta)}}\right) + C^2 S(C) \\
&= 2\theta(1 - e^{-\frac{C^2}{(2\theta)}})
\end{aligned}$$

Therefore,

$$\begin{aligned}
I(\theta) &= -\frac{1}{\theta^2} nE(R) + \frac{1}{\theta^3} nE(X^2) \\
&= -\frac{n}{\theta^2}(1 - e^{-\frac{C^2}{2\theta}}) + \frac{2n}{\theta^2}(1 - e^{-\frac{C^2}{(2\theta)}}) \\
&= \frac{n}{\theta^2}(1 - e^{-\frac{C^2}{2\theta}})
\end{aligned}$$

(c) As mle is asymptotic normal distributed, we need to find the mean and variance of mle. From 1(a), we know that $\widehat{\theta}_{\text{MLE}} = \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n R_i}$,
Therefore,

$$\begin{aligned}
Var(\theta) &= \frac{1}{\sqrt{J(\theta)}} \\
&= \frac{1}{\sqrt{-\frac{\sum_{i=1}^n R_i}{\theta^2} + \frac{\sum_{i=1}^n X_i^2}{\theta^3}}} \\
&= \frac{\theta}{\sqrt{-\sum_{i=1}^n R_i + \frac{\sum_{i=1}^n X_i^2}{\theta}}} \\
&= \frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n R_i)^{\frac{3}{2}}}
\end{aligned}$$

therefore, the 95% confidence interval for θ is $[\hat{\theta} - 1.96\frac{1}{\sqrt{J(\theta)}}, \hat{\theta} + 1.96\frac{1}{\sqrt{J(\theta)}}]$,
i.e., $[\frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n R_i} - 0.98\frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n R_i)^{\frac{3}{2}}}, \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n R_i} + 0.98\frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n R_i)^{\frac{3}{2}}}]$