Problem Set 6 - Solutions

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due March 29)

[1] a)

$$L(\theta, X) = \begin{cases} \theta^n \left(\Pi_{i=1}^n x_i\right)^{\theta-1}, & 0 < x_i < 1, \forall i, 0 < \theta < \infty \\ 0, & \text{otherwise} \end{cases}$$

Now take log to find $\ell(\theta)$ and then compute the mle as

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^n \log x_i}$$

b)

$$L(\theta, X) = \begin{cases} e^{-\sum(x_i-\theta)}, & \theta \le x_i, \forall i \\ 0, & \text{otherwise} \end{cases}$$

Which gives $\log L(\theta, X) = -\sum(x_i - \theta)$ and $\partial \log L/\partial\theta = n > 0$. That is $\log L$ is an increasing function of θ provided $x_i \le \theta$, $i = 1, \dots, n$. Thus, $\hat{\theta} = \min\{X_1, \dots, X_n\}$.

[2] a)

$$L(\theta, X) = \begin{cases} \frac{2^n}{\theta^{2n}} \Pi x_i, & 0 < x_i \le \theta, \forall i \\ 0, & \text{otherwise} \end{cases}$$

Note that $x_i \le \theta$ for all i if and only if $\max\{X_1, \dots, X_n\} \le \theta$ hence the likelihood can be written as

$$L(\theta, X) = \begin{cases} \frac{2^n}{\theta^{2n}} \Pi x_i, & \max\{X_1, \dots, X_n\} \le \theta \\ 0, & \text{otherwise} \end{cases}$$

Since the likelihood function is monotonically decreasing in θ, the maximum of $L(\theta)$ occurs at the smallest value in the range of θ, so $\hat{\theta} = \max\{X_1, \dots, X_n\}$.

b) The CDF of X_i is $F(x) = x^2/\theta^2$ and therefore we can compute the CDF and pdf of $Y \equiv \hat{\theta}$ as

$$F_Y(y) = \frac{y^{2n}}{\theta^{2n}}, \qquad 0 < y \le \theta$$

$$f_Y(y) = \frac{2ny^{2n-1}}{\theta^{2n}}, \qquad 0 < y \le \theta$$

$$\implies E(Y) = \int_0^\theta y \frac{2ny^{2n-1}}{\theta^{2n}} \, dy = \frac{2n}{2n+1}\theta$$

so $c = (2n+1)/(2n)$.

c) Note that the median is the value that solves $x^2/\theta^2 = 1/2$, which is $\theta/\sqrt{2}$. The mle of the median is therefore $Y/2$. Note that an unbiased estimate of the median is $[(2n+1)Y]/[2n\sqrt{2}]$.

d) $P(X \leq 2) = 4/\theta^2$, therefore the mle of it is $4/Y^2$.

[3] For $X \sim \text{Poisson}(\theta)$

$$P(X = x) = \frac{\theta^x e^{-\theta}}{x!}$$

as usual, we first write the loglikelihood function and then can find the mle of θ as $\hat{\theta} = \bar{x}$. For the given data we compute $\hat{\theta} = \bar{x} = 2.11$. Then we compute the mle of $P(X = 2)$ as

$$P(X = 2) = \frac{\hat{\theta}^2 e^{-\hat{\theta}}}{2!} = \frac{\bar{x}^2 e^{-\bar{x}}}{2!} = 0.27$$

[4] a)

$$\begin{aligned}
\log f(x;\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2\theta} \\
\frac{\partial \log f(x;\theta)}{\partial \theta} &= -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} \\
\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} &= \frac{1}{2\theta^2} - \frac{x^2}{\theta^3} \\
nI(\theta) &= -nE\left[\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right] \\
&= -\frac{n}{2\theta^2} + \frac{n}{\theta^2} \\
&= \frac{n}{2\theta^2}
\end{aligned}$$

b) Here the mle is $\hat{\theta} = \sum X_i^2/n$. Since $\sum X_i^2/n \sim \chi^2(n)$, we have

$$\text{Var}(\hat{\theta}) = \frac{\theta^2}{n^2}\text{Var}\left(\frac{\sum X_i^2}{\theta}\right) = \frac{n}{2\theta^2} = \frac{1}{nI(\theta)}$$

Therefore, it achieves the minimum variance, so it's an efficient estimator.

c)

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right) \\
\Longrightarrow\ &\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, 2\theta^2\right)
\end{aligned}$$

[5] Note that

$$\begin{aligned}
E[|X_1|] &= 2\int_0^\infty \frac{x}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{x^2}{\theta}\right\}dx \\
&= 2\sqrt{\theta}\int_0^\infty \frac{1}{\sqrt{2\pi}}\exp\{-z\}\,dz \\
&= \sqrt{\theta}\sqrt{\frac{2}{\pi}}
\end{aligned}$$

So $E[Y] = n\sqrt{\theta}\sqrt{\frac{2}{\pi}}\, c$ so for $c = \frac{1}{n}\sqrt{\frac{2}{\pi}}$, Y is an unbiased estimator of $\sqrt{\theta}$.

Note that

$$\begin{aligned}\mathrm{Var}\left[\sqrt{\frac{2}{\pi}}|X_1|\right] &= \frac{\pi}{2}\left[E(X_1^2) - [E(|X_1|)]^2\right] \\ &= \frac{\pi}{2}\left[\theta\left(1 - \frac{2}{\pi}\right)\right] = \theta\left[\frac{\pi}{2} - 1\right].\end{aligned}$$

By independence,

$$\mathrm{Var}\,[Y] = \theta\left[\frac{\pi}{2} - 1\right]\frac{1}{n}$$

To finish, we need the efficiency of the parameter $\sqrt{\theta}$. For convenience, let $\beta = \sqrt{\theta}$. Then

$$\log f(x;\beta) = -\log\sqrt{2\pi} - \log\beta - \frac{x^2}{2\beta^2}$$

The second partial derivative of this expression is

$$\frac{\partial^2 \log f(x;\theta)}{\partial\beta^2} = \frac{1}{\beta^2} - 3\frac{x^2}{\beta^4}$$

Hence using $\beta = \sqrt{\theta}$ we can compute information of the sample

$$nI(\sqrt{\theta}) = -nE\left[\frac{1}{\theta} - 3\frac{X^2}{\theta^2}\right] = \frac{2n}{\theta}$$

To compare the efficiency of the estimator let us look at the ratio of $nI(\theta)$ to $\mathrm{Var}(Y)$, which is $1/(\pi - 2)$. Therefore, this is not an efficient estimator.

[6] a) Recall the Δ-method:

$$\sqrt{n}\,(X_n - \theta) \xrightarrow{d} N\left(0, \sigma^2\right) \implies \sqrt{n}\,(g(X_n) - g(\theta)) \xrightarrow{d} N\left(0, \sigma^2[g'(\theta)]^2\right)$$

On the other hand, from Theorem 6 we know that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right)$$

Now just choose $X_n = \hat{\theta}_n$, $\theta = \theta_0$ and $\sigma^2 = 1/I(\theta_0)$, then the result is immediate from Δ-method.

b) This result follows immediately from the equations we used in proving Theorem 6.

c) Recall the equivalent expression given in Theorem 6:

$$\hat{\theta}_n \xrightarrow{d} N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$$

Therefore, the asymptotic standard deviation of the mle $\hat{\theta}$ is $\left[\frac{1}{nI(\theta_0)}\right]^{1/2}$. Because $I(\theta)$ is a continuous function of θ, it follows that $I(\hat{\theta}_n) \xrightarrow{p} I(\theta_0)$. Thus we have a consistent estimate of the asymptotic standard deviation of the mle. Based on this result and the general idea of constructing confidence intervals we obtain the stated result (for an analogy just think of how we construct C.I. for the population mean μ using the sample mean $\bar{X}$).

[7] a) The loglikelihood contribution is given by

$$\log L_i(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2}$$

Because the error terms are independent across observations, the loglikelihood function is simply the sum of all N loglikelihood contributions.

b) Differentiating with respect to β_1, β_2, respectively, gives

$$\frac{\partial \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_1} = \frac{(y_i - \beta_1 - \beta_2 x_i)}{\sigma^2} = \frac{\varepsilon_i}{\sigma^2}$$
$$\frac{\partial \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_2} = \frac{(y_i - \beta_1 - \beta_2 x_i)x_i}{\sigma^2} = \frac{\varepsilon_i x_i}{\sigma^2}$$

c)

$$\begin{aligned}\frac{\partial \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{1}{2}\frac{2\pi}{2\pi\sigma^2} - \frac{1}{2}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^4} \times (-1) \\ &= \frac{\left[(y_i - \beta_1 - \beta_2 x_i)^2 - \sigma^2\right]}{2\sigma^4} \\ &= \frac{\left[\varepsilon_i^2 - \sigma^2\right]}{2\sigma^4}\end{aligned}$$

where we used $\varepsilon_i = y_i - \beta_1 - \beta_2 x_i$ for the last equality. Using $E[\varepsilon_i^2] = \sigma^2$ shows that this term also has expectation zero (for the true parameter values). (You might find it easier to define $\gamma = \sigma^2$ and substitute this in $\log L_i$, then $\partial \log L_i(\boldsymbol{\beta}, \sigma^2)/\partial\sigma^2 = \partial \log L_i(\boldsymbol{\beta}, \gamma)/\partial\gamma$)

d) The first order conditions from part (b) are given by

$$\sum_{i=1}^{N}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$
$$\sum_{i=1}^{N}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)x_i = 0$$

which we can write as

$$\bar{y} - \hat{\beta}_1 - \frac{N_1}{N}\hat{\beta}_2 = 0$$
$$\frac{1}{N_1}\sum_{i=1}^{N_1}(y_i - \hat{\beta}_1 - \hat{\beta}_2) = 0$$

which - after some rewriting - result in the required expressions. In this setting, $\hat{\beta}_1$ corresponds to the sample average for females and $\hat{\beta}_2$ is the difference in sample averages between males and females. The true parameter values correspond to the expected value for females and the expected differential in y_i between males and females.

e)

$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_1 \partial \sigma^2} = -\frac{y_i - \beta_1 - \beta_2 x_i}{\sigma^4} = -\frac{\varepsilon_i}{\sigma^4}$$
$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_2 \partial \sigma^2} = -\frac{(y_i - \beta_1 - \beta_2 x_i) x_i}{\sigma^4} = -\frac{\varepsilon_i x_i}{\sigma^4}$$

On the other hand, starting from $\frac{\partial \log L_i(\boldsymbol{\beta},\sigma^2)}{\partial \sigma^2} = \frac{\left[\varepsilon_i^2 - \sigma^2\right]}{2\sigma^4}$ and differentiating with respect to β_1 and β_2 we obtain the same results.

The expected value of the two terms (under the true parameter values) is zero. This implies that the covariance matrix of the ML estimator is block diagonal (with the blocks corresponding to $\hat{\boldsymbol{\beta}}$ and σ^2, respectively).

f) First compute

$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_1 \partial \beta_1} = -\frac{1}{\sigma^2}$$
$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_1 \partial \beta_2} = -\frac{x_i}{\sigma^2}$$
$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_2 \partial \beta_2} = -\frac{x_i^2}{\sigma^2}$$
$$\frac{\partial^2 \log L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2 \partial \sigma^2} = -\frac{1}{2\sigma^4}$$

Then the information matrix of a sample of size one is

$$I_i(\beta_1, \beta_2, \sigma^2) = -E \begin{bmatrix} -\frac{1}{\sigma^2} & -\frac{x_i}{\sigma^2} & 0 \\ -\frac{x_i}{\sigma^2} & -\frac{x_i^2}{\sigma^2} & 0 \\ 0 & 0 & -\frac{1}{2\sigma^4} \end{bmatrix} = \frac{1}{\sigma^2} E \begin{bmatrix} 1 & x_i & 0 \\ x_i & x_i^2 & 0 \\ 0 & 0 & \frac{1}{2\sigma^2} \end{bmatrix}$$

Then if you want to the information matrix of the whole sample multiply this by N.

[8] a) Straightforward algebra show that the likelihood function is

$$L(\theta) = \theta^{-n} \exp\{-(n/\theta)\bar{x}\}$$

From which we find the mle $\hat{\theta} = \bar{x}$.

b) Then the likelihood ratio test statistic simplifies to

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = e^n \left(\frac{\bar{x}}{\theta_0}\right)^n \exp\{-n\bar{x}/\theta_0\}$$

c) Other than the constant e^n, the test statistic is of the form $g(t) = t^n \exp\{-nt\}$ $(t > 0)$, where $t = \bar{x}/\theta_0$. Using differentiation it is easy to show that $g(t)$ has a unique critical value at 1, i.e. $g'(1) = 0$, and further that $t = 1$ provides a maximum because $g''(1) < 0$. So we conclude that the shape of g is hump shaped and that $g(t) \leq c$ if and only if $t \leq c_1$ or $t \geq c_2$, for some c_1 and c_2. This leads to

$$\Lambda \leq c \iff \frac{\bar{x}}{\theta_0} \leq c_1 \text{ or } \frac{\bar{x}}{\theta_0} \geq c_2$$

d) Note that under the null hypothesis, H_0, the statistic $(2/\theta_0)\sum_{i=1}^{n} X_i$ has a χ^2 distribution with $2n$ degrees of freedom. Based on this, the following decision rule results in a level α test:

$$\text{Reject } H_0 \text{ if } (2/\theta_0)\sum_{i=1}^{n} X_i \leq \chi^2_{1-\alpha/2}(2n) \text{ or } (2/\theta_0)\sum_{i=1}^{n} X_i \geq \chi^2_{\alpha/2}(2n)$$

where $\chi^2_{1-\alpha/2}(2n)$ is the lower $\alpha/2$ quantile of a χ^2 distribution with $2n$ degrees of freedom and $\chi^2_{\alpha/2}(2n)$ is the upper $\alpha/2$ quantile of a χ^2 distribution with $2n$ degrees of freedom.