36-705 Intermediate Statistics Homework #5 Solutions

October 20, 2016

Problem 1 [20 pts.]

We have that $\mathbb{E}[X_1] = \alpha\beta$ and $E\left[X_1^2\right] = \alpha\beta^2 + \alpha^2\beta^2$. So the system defined by the method of moments is:

$$M_1 = \frac{1}{n}\sum_{i=1}^n X_i = \alpha\beta \qquad \text{and} \qquad M_2 = \frac{1}{n}\sum_{i=1}^n X_i^2 = \alpha\beta^2 + \alpha^2\beta^2$$

Solving for α, β we get:

$$\hat{\alpha} = \frac{M_1^2}{M_2 - M_1^2} \qquad \text{and} \qquad \hat{\beta} = \frac{M_2 - M_1^2}{M_1}$$

Problem 2 [25 pts.]

Since there is only one parameter to estimate and $\mathbb{E}[X_1] = \lambda$, the method of moments estimator $\hat{\lambda}_m$ is given by:

$$\hat{\lambda}_m = \frac{1}{n}\sum_{i=1}^n X_i$$

To find the maximum likelihood estimator $\hat{\lambda}_{mle}$ calculate the log likelihood function:

$$l_n(\lambda) = -n\lambda + \sum_{i=1}^n X_i \log\lambda - \sum_{i=1}^n \log X_i!$$

Its first derivative equals to:

$$\frac{\partial}{\partial\lambda} l_n(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^n X_i$$

Setting it equal to 0 we get the unique solution $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$. This is a global maximum because the second derivative is negative for every λ

$$\frac{\partial^2}{\partial\lambda^2} l_n(\lambda) = -\frac{1}{\lambda^2}\sum_{i=1}^n X_i < 0$$

To get the Fisher information we use the second derivative to get:

$$\mathcal{I}_n(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial\lambda^2} l_n(\lambda)\right] = \frac{1}{\lambda^2}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{n}{\lambda}$$

Problem 3

(a) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

First, let us write ψ in terms of θ. We see that

$$\begin{aligned}
\psi &= \mathbb{P}(Y_1 = 1) \\
&= \mathbb{P}(X_1 > 0) \\
&= 1 - \mathbb{P}(X_1 \le 0) \\
&= 1 - \mathbb{P}(X_1 - \theta \le 0 - \theta) \\
&= 1 - \Phi(-\theta) && \text{(because } X_1 - \theta \sim N(0,1)\text{)} \\
&= \Phi(\theta).
\end{aligned}$$

Now we will find the MLE $\hat{\theta}$ of θ.

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_i - \theta)^2\right) \\
\ell(\theta) &= \sum_{i=1}^{n} \left[-\log(\sqrt{2\pi}) - \frac{1}{2}(X_i - \theta)^2 \right] \\
\frac{\partial \ell(\theta)}{\partial \theta} &= \sum_{i=1}^{n} [-(X_i - \theta)(-1)].
\end{aligned}$$

To solve for $\hat{\theta}$, we solve

$$\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta} &\overset{set}{=} 0 \\
\sum_{i=1}^{n} (X_i - \theta) &= 0 \\
\sum_{i=1}^{n} X_i &= n\theta \\
\hat{\theta} &= \bar{X}_n.
\end{aligned}$$

By the Continuous Mapping Theorem, we conclude

$$\hat{\psi} = \Phi(\hat{\theta}) = \Phi(\bar{X}_n).$$

(c) Define $\tilde{\psi} = (1/n)\sum_i Y_i$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

We determine that

$$\begin{aligned}
\tilde{\psi} &= \frac{1}{n}\sum_{i=1}^{n} Y_i \\
&\xrightarrow{p} \mathbb{E}[Y_i] && \text{(by WLLN)} \\
&= 1 \cdot \mathbb{P}(Y_i = 1) + 0 \cdot \mathbb{P}(Y_i = 0) \\
&= \mathbb{P}(Y_i = 1) \\
&= \mathbb{P}(Y_1 = 1) \\
&= \psi.
\end{aligned}$$

(d) Compute the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$. Hint: Use the delta method to get the standard error of the MLE. Then compute the standard error (i.e. the standard deviation) of $\tilde{\psi}$.

First, let us compute the asymptotic variance of $\tilde{\psi}$. By definition, $\tilde{\psi} = \bar{Y}_n$. By the Central Limit Theorem,

$$\frac{\sqrt{n}(\bar{Y}_n - \mathbb{E}[Y_i])}{\sqrt{\mathbb{V}(Y_i)}} \rightsquigarrow N(0,1).$$

So

$$\frac{\sqrt{n}(\tilde{\psi} - \mathbb{E}[Y_i])}{\sqrt{\mathbb{V}(Y_i)}} \rightsquigarrow N(0,1).$$

In part (c), we showed that $\mathbb{E}[Y_i] = \psi$. So

$$\frac{\sqrt{n}(\tilde{\psi} - \psi)}{\sqrt{\mathbb{V}(Y_i)}} \rightsquigarrow N(0,1).$$

Additionally,

$$\mathbb{V}(Y_i) = \mathbb{E}[Y_i^2] - (\mathbb{E}[Y_i])^2.$$

We determine

$$\mathbb{E}[Y_i^2] = 1^2 \cdot \mathbb{P}(Y_i = 1) + 0^2 \cdot \mathbb{P}(Y_i = 0) = \mathbb{P}(Y_i = 1) = \mathbb{P}(Y_1 = 1) = \psi.$$

So

$$\mathbb{V}(Y_i) = \psi - \psi^2 = \psi(1 - \psi).$$

In part (a), we showed that $\psi = \Phi(\theta)$. So

$$\mathbb{V}(Y_i) = \Phi(\theta)[1 - \Phi(\theta)].$$

Now we see that

$$\frac{\sqrt{n}(\tilde{\psi} - \psi)}{\sqrt{\Phi(\theta)[1 - \Phi(\theta)]}} \rightsquigarrow N(0,1).$$

That means that

$$\sqrt{n}(\tilde{\psi} - \psi) \rightsquigarrow N(0, \Phi(\theta)[1 - \Phi(\theta)]).$$

Thus, the asymptotic variance of $\tilde{\psi}$ is $\sigma^2_{\tilde{\psi}} = \Phi(\theta)[1 - \Phi(\theta)]$.

Next, let us compute the asymptotic variance of $\hat{\psi}$. To begin, by the Central Limit Theorem,

$$\sqrt{n}(\bar{X}_n - \theta) \rightsquigarrow N(0,1).$$

$\Phi(\theta)$ (the standard normal CDF) is a differentiable function such that $\Phi'(\theta) \neq 0$. Then by the Delta Method,

$$\frac{\sqrt{n}\left(\Phi(\bar{X}_n) - \Phi(\theta)\right)}{|\Phi'(\theta)|} \rightsquigarrow N(0,1).$$

In part (a), we showed that $\hat{\psi} = \Phi(\bar{X}_n)$ and $\psi = \Phi(\theta)$. This means that

$$\frac{\sqrt{n}(\hat{\psi} - \psi)}{|\Phi'(\theta)|} \rightsquigarrow N(0,1),$$

so

$$\sqrt{n}(\hat{\psi} - \psi) \rightsquigarrow N(0, (\Phi'(\theta))^2).$$

Since Φ' is the normal pdf, we know that

$$\begin{aligned}(\Phi'(\theta))^2 &= \left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\theta^2\right)\right)^2 \\ &= \frac{1}{2\pi}\exp(-\theta^2).\end{aligned}$$

That means that

$$\sqrt{n}(\hat{\psi}-\psi) \rightsquigarrow N\left(0, \frac{1}{2\pi}\exp(-\theta^2)\right),$$

so the asymptotic variance of $\hat{\psi}$ is $\sigma^2_{\hat{\psi}} = \frac{1}{2\pi}\exp(-\theta^2)$.

We conclude that the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$ is

$$\text{ARE}(\tilde{\psi},\hat{\psi}) = \frac{\sigma^2_{\hat{\psi}}}{\sigma^2_{\tilde{\psi}}} = \frac{\frac{1}{2\pi}\exp(-\theta^2)}{\Phi(\theta)[1-\Phi(\theta)]}.$$

(e) Suppose that the data are not really normal. Show that $\hat{\psi}$ is not consistent. What, if anything, does $\hat{\psi}$ converge to?

Suppose we have a random variable X such that $X_i = 0$ for all i. Then for all choices of n,

$$\hat{\psi} = \Phi(\bar{X}_n) = \Phi(0) = \frac{1}{2}.$$

So $\hat{\psi} \xrightarrow{p} \frac{1}{2}$. But $\psi = \mathbb{P}(Y_i = 1) = \mathbb{P}(X_i > 0) = 0$. So $\hat{\psi} \overset{p}{\nrightarrow} \psi$.

Problem 4 [30 pts.]

(a) By assuming X_1, X_2 are independent we have that the likelihood function is

$$\begin{aligned}L(p_1,p_2) &= \binom{n_1}{X_1} p_1^{X_1}(1-p_1)^{n_1-X_1}\binom{n_2}{X_2}p_2^{X_2}(1-p_2)^{n_2-X_2} \\ &= q(p_1;X_2)\times q(p_2;X_2)\end{aligned}$$

so that $(\widehat{p}_1,\widehat{p}_2) = \left(\text{argmax}_{p_1\in[0,1]}q(p_1;X_1),\ \text{argmax}_{p_2\in[0,1]}q(p_2;X_2)\right)$. We have:

$$\log q(p_i;X_i) = \log\binom{n_i}{X_i} + X_i\log p_i + (n_i - X_i)\log(1-p_i).$$

The first order condition is

$$\frac{X_i}{p_i} - \frac{n_i - X_i}{1-p_i} = 0$$

which gives $\widehat{p}_i = X_i/n_i$. This is a maximum point since the second derivative

$$-\frac{X_i}{p_i^2} - \frac{n-X_i}{(1-p_i)^2} < 0$$

is negative for any $p_i \in (0,1)$. Thus, the MLE of p is $\widehat{p} = (\widehat{p}_1, \widehat{p}_2) = (X_1/n_1, X_2/n_2)$, and by the equivariance property of the MLE,

$$\widehat{\psi} = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

(b) We have

$$\frac{\partial^2}{\partial p_i p_i}\ell(p_1,p_2) = -\frac{n_i}{p_i(1-p_i)}, \quad i = 1,2$$

and

$$\frac{\partial^2}{\partial p_1 p_2}\ell(p_1,p_2) = \frac{\partial^2}{\partial p_2 p_1}\ell(p_1,p_2) = 0.$$

Therefore, the Fisher information matrix is

$$I_n(p_1,p_2) = \begin{bmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{bmatrix}.$$

(c) Note that we can write $X_i = \sum_{j=1}^{n_1} X_{ij}$, where $X_{i1}, \dots, X_{in_i} \sim \text{Bernoulli}(p_i)$ so that $\widehat{p}_i$ is a sample mean and the Central Limit Theorem implies

$$\sqrt{n_i}(\widehat{p}_i - p_i) \rightsquigarrow N(0, p_i(1-p_i))$$

for $i = 1,2$. Moreover, since $\widehat{p}_1, \widehat{p}_2$ are independent, we also have, setting $n_1 = n_2 = n$

$$\sqrt{n}(\widehat{p} - p) \rightsquigarrow N(0, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} p_1(1-p_1) & 0 \\ 0 & p_2(1-p_2) \end{bmatrix}.$$

By applying the multivariate Delta method on the transformation $\widehat{\psi} = g(\widehat{p}) = \widehat{p}_1 - \widehat{p}_2$ we obtain:

$$\sqrt{n}(\widehat{\psi} - \psi) \rightsquigarrow N(0, p_1(1-p_1) + p_2(1-p_2)).$$

Problem 5 [25 pts.]

(a) Let $\psi = g(\theta)$ where g is a smooth, invertible function. Hence, $\theta = h(\psi)$ where $h = g^{-1}$. Let $I(\theta)$ denote the Fisher information for θ and let $I(\psi)$ denote the Fisher information for ψ. Show that $I(\psi) = I(\theta)(h'(\psi))^2$.

We see that

$$\begin{aligned}
I(\psi) &= \mathbb{E}_\psi\left[\left(\frac{\partial}{\partial\psi}\log p(x;\psi)\right)^2\right] \\
&= \mathbb{E}_\psi\left[\left(\frac{\partial}{\partial\theta}\frac{\partial\theta}{\partial\psi}\log p(x;\psi)\right)^2\right] \\
&= \mathbb{E}_\psi\left[\left(\frac{\partial}{\partial\theta}\frac{\partial h(\psi)}{\partial\psi}\log p(x;\psi)\right)^2\right] \\
&= \mathbb{E}_\psi\left[\left(\frac{\partial}{\partial\theta}h'(\psi)\log p(x;\psi)\right)^2\right] \\
&= \int\left(\frac{\partial}{\partial\theta}h'(\psi)\log p(x;\psi)\right)^2 p(x;\psi)dx
\end{aligned}$$

$h'(\psi)$ does not depend on x and has been evaluated at the ψ on which the expectation was conditioned. So we can take $(h'(\psi))^2$ outside of the integral.

$$\begin{aligned}
&= (h'(\psi))^2\int\left(\frac{\partial}{\partial\theta}\log p(x;\psi)\right)^2 p(x;\psi)dx \\
&= (h'(\psi))^2\int\left(\frac{\partial}{\partial\theta}\log p(x;g(\theta))\right)^2 p(x;g(\theta))dx
\end{aligned}$$

Since g is an invertible function, there is a one-to-one correspondence between values of θ and values of $g(\theta)$. So we can write the density of x parametrized by θ as $f(x;\theta) = p(x;g(\theta))$ for all θ.

$$\begin{aligned}
&= (h'(\psi))^2\int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)dx \\
&= (h'(\psi))^2\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right] \\
&= (h'(\psi))^2 I(\theta).
\end{aligned}$$

(b) Let $X \sim N(\theta, 1)$ and let $\psi = e^\theta$. Find the Fisher information for ψ.

The function h such that $\theta = h(\psi)$ is given by $h(\psi) = \log(\psi)$. Then $h'(\psi) = \frac{1}{\psi}$, so

$$(h'(\psi))^2 = \frac{1}{\psi^2}.$$

Now let us find $I(\theta)$. The likelihood function $L(\theta)$ is given by

$$L(\theta) = f(x;\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\theta)^2\right).$$

So the log likelihood is

$$\ell(\theta) = -\log(\sqrt{2\pi}) - \frac{1}{2}(x-\theta)^2.$$

Then

$$\begin{aligned}
I(\theta) &= -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\left(-\log(\sqrt{2\pi}) - \frac{1}{2}(X-\theta)^2\right)\right] \\
&= -\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\left(-(X-\theta)(-1)\right)\right] \\
&= -\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}(X-\theta)\right] \\
&= -\mathbb{E}_\theta[-1] \\
&= 1.
\end{aligned}$$

We conclude that

$$I(\psi) = I(\theta)(h'(\psi))^2 = 1\left(\frac{1}{\psi^2}\right) = \frac{1}{\psi^2}.$$