Math 442 – Homework Set 4

p.444[2,5,6,8,9,10,13,14], p.415[4,11]

In this solution packet, we will frequently use the notation

$$\lambda(x|\theta) := \log f_X(x|\theta).$$

1. (p.444 #2) Suppose that X has geometric distribution with parameters p. (See Section 5.5.) Find the Fisher information $I(p)$ in X.

   Solution. It can be seen that

   $$\lambda(x|p) = \ln f_X(x|p) = \ln p(1-p)^x = \ln(p) + x\ln(1-p)\ .$$

   Then

   $$\lambda'(x|p) = \frac{\partial}{\partial p}\ \lambda(x|p) = \frac{1}{p} - \frac{x}{1-p}$$

   and

   $$\lambda''(x|p) = \frac{\partial^2}{\partial p^2}\ \lambda(x|p) = -\frac{1}{p^2} - \frac{x}{(1-p)^2}\ .$$

   Since the Fisher information may be defined as

   $$I(p) = -E[\lambda''(X|p)] \quad \text{Equation (7.8.8) page 436,}$$

   with $E(X) = (1-p)/p$ it follows that

   $$I(p) = -E\left[-\frac{1}{p^2} - \frac{X}{(1-p)^2}\right] = \left[\frac{1}{p^2} + \frac{1-p}{p(1-p)^2}\right] = \frac{1}{p^2} + \frac{1}{p(1-p)}\ .$$

   □

2. (p.444 #5) Suppose that a random variable X has a normal distribution for which the mean is 0 and the variance σ^2 is unknown ($\sigma^2 > 0$). Find the Fisher information $I(\sigma^2)$ in X. Note that in this exercise the variance σ^2 is regarded as the parameter, whereas in Exercise 4 the standard standard deviation σ is regarded as the parameter.

   Solution. Let

   $$\lambda(x|\sigma^2) = \ln f_X(x|0,\sigma^2) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\right) = -\tfrac{1}{2}\ln\sigma^2 - \tfrac{1}{2}\ln(2\pi) - \frac{x^2}{2\sigma^2}\ .$$

   Then

   $$\lambda'(x|\sigma^2) = \frac{\partial}{\partial\sigma^2}\ \lambda(x|\sigma^2) = -\frac{1}{2\sigma^2} + \frac{x^2}{2(\sigma^2)^2}$$

   and

   $$\lambda''(x|\sigma^2) = \frac{\partial^2}{(\partial\sigma^2)^2}\lambda(x|\sigma^2) = \frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}\ .$$

   With $E(X) = 0,\ E(X^2) = \mathrm{Var}(X) + 0 = \sigma^2$ and so

   $$I(\sigma^2) = -E\left[\frac{1}{2\sigma^4} - \frac{X^2}{\sigma^6}\right] = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4}\ .$$

   □

3. (p. 444 #6) Suppose that X is a random variable for which the probability density function or the probability function is $f_X(x|\theta)$, where the value of the parameter θ is unknown but must lie in an open interval Ω. Let $I_0(\theta)$ denote the Fisher information in X. Suppose now that the parameter θ is replaced by a new parameter μ, where $\theta = \psi(\mu)$, and ψ is a differentiable function. Let $I_1(\mu)$ denote the Fisher information in X when the parameter is regarded as μ. Show that

$$I_1(\mu) = [\psi'(\mu)]^2 I_0[\psi(\mu)].$$

Proof. By definition

$$\begin{aligned}
I_1(\mu) &= E_\mu\left\{\left[\lambda'\big(x|\psi(\mu)\big)\right]^2\right\} \\
&= E\left\{\left[\frac{\partial}{\partial\mu}\ \lambda\big(x|\psi(\mu)\big)\right]^2\right\} \\
&= E\left\{\left[\frac{\partial}{\partial\psi(\mu)}\ \lambda\big(x|\psi(\mu)\big)\ \frac{\partial\psi(\mu)}{\partial\mu}\right]^2\right\} && \text{by the chain rule} \\
&= \left(\frac{\partial\psi(\mu)}{\partial\mu}\right)^2\ E\left\{\left[\frac{\partial}{\partial\psi(\mu)}\ \lambda\big(x|\psi(\mu)\big)\right]^2\right\} && \text{properties of expection} \\
&= [\psi'(\mu)]^2\ E_{\psi(\mu)}\left\{\left[\lambda'\big(x|\psi(\mu)\big)\right]^2\right\} \\
&= [\psi'(\mu)]^2\ I_0[\psi(\mu)] && \text{as } \psi(\mu) = \theta\ .
\end{aligned}$$

□

4. (p. 444 #8) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a normal distribution for which the mean μ is unknown and the variance σ^2 is known. Show that $\overline{X}_n$ is an efficient estimator of μ.

Solution. It is said that an estimator, $T = T(X_1, X_2, \ldots, X_n)$, is an *efficient estimator of its expectation* $g(\theta)$ if there is equality in the information inequality for every value of $\theta \in \Omega$. The *information inequality* also known as the *Cramér-Rao inequality* may be stated as follows:

$$\mathrm{Var}_\theta(T) \geq \frac{[g'(\theta)]^2}{E\left[\left\{\sum_{i=1}^n \frac{\partial}{\partial\theta}\ \ln f_X(x_i|\theta)\right\}^2\right]}\ .$$

In the given problem, $T = \overline{X}_n$ and so $g(\mu) = E(\overline{X}_n) = \mu$ hence $g'(\mu) = 1$.

$$\begin{aligned}
\sum_{i=1}^n \frac{\partial}{\partial\mu}\ \ln f_X(x_i|\mu,\sigma^2) &= \sum_{i=1}^n \frac{\partial}{\partial\mu}\ \ln\left[\frac{1}{\sqrt{2\pi}\ \sigma}\ \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}\right] \\
&= \sum_{i=1}^n \frac{\partial}{\partial\mu}\left[-\frac{1}{2}\ln(2\pi) - \ln\sigma - \frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\
&= \sum_{i=1}^n -\frac{1}{2\sigma^2}\ \frac{\partial}{\partial\mu}(x_i{}^2 - 2\mu x_i + \mu^2) \\
&= \sum_{i=1}^n -\frac{1}{2\sigma^2}\ (-2x_i + 2\mu) \\
&= \frac{n}{\sigma^2}(\mu - \overline{X}_n).
\end{aligned}$$

Then

$$E\left[\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\mu}\ \ln f_X(x_i|\mu,\sigma^2)\right\}^2\right]=\frac{n^2}{\sigma^4}\ E(\mu-\overline{X}_n)^2=\frac{n^2}{\sigma^4}\mathrm{Var}(\overline{X}_n)=\frac{n}{\sigma^2}\ .$$

Here we have used the fact that $\overline{X}_n \sim N(\mu,\sigma^2/n)$; that is, $\mathrm{Var}(\overline{X}_n)=\sigma^2/n$. Now it can be observed that there is equality in the information inequality,

$$\mathrm{Var}_\mu(T)=\frac{\sigma^2}{n}=\frac{1}{n/\sigma^2}=\frac{[g'(\theta)]^2}{E\left[\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ \ln f_X(x_i|\theta)\right\}^2\right]},$$

for every value of $\mu\in\Omega$. Therefore $\overline{X}_n$ is an efficient estimator of μ. □

5. (p. 444 #9) Suppose that a single observation X is taken from a normal distribution for which the mean is 0 and the standard deviation σ is unknown. Find an unbiased estimator of σ, determine its variance, and show that this variance is greater than $1/I(\sigma)$ for every value of $\sigma>0$. Note that the value of $I(\sigma)$ was found in Exercise 4.

Solution. Since we have not completed Exercise 4,

$$\begin{aligned}\lambda(x|\sigma)=\ln f_X(x|\sigma)&=\ln\left[\frac{1}{\sqrt{2\pi}\ \sigma}\ \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}\right]\\&=-\frac{1}{2}\ \ln(2\pi)-\ln\sigma-\frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)\end{aligned}$$

Then

$$\lambda''(x|\sigma)=\frac{\partial^2}{(\partial\sigma)^2}\lambda(x|\sigma)=\frac{1}{\sigma^2}-\frac{3x^2}{\sigma^4}$$

and so

$$I(\sigma)=-E\left[\frac{1}{\sigma^2}-\frac{3x^2}{\sigma^4}\right]=-\frac{1}{\sigma^2}+\frac{3E(X^2)}{\sigma^4}=\frac{2}{\sigma^2}\ .$$

Next we claim that $\sqrt{(\pi/2)}|X|$ is an unbiased estimator of σ,

$$\begin{aligned}E(\sqrt{(\pi/2)}|X|)&=\int_{-\infty}^{\infty}\sqrt{(\pi/2)}|x|f_X(x|0,\sigma^2)dx\\&=\sqrt{(\pi/2)}\int_0^{\infty}2x\frac{1}{\sqrt{2\pi}\ \sigma}\ \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}dx\\&=\frac{1}{\sigma}\int_0^{\infty}x\exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\}dx\end{aligned}$$

use substitution let $u=x^2/2\Rightarrow du=xdx$, then

$$E(\sqrt{(\pi/2)}|X|)=\frac{1}{\sigma}\int_0^{\infty}e^{-u/\sigma^2}\ du=\frac{1}{\sigma}\left[-\sigma^2e^{-u/\sigma^2}\right]_0^{\infty}=\frac{\sigma^2}{\sigma}=\sigma\ .$$

Therefore, our claim that $\sqrt{(\pi/2)}|X|$ is an unbiased estimator of σ is true.
The variance of the estimator is

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

Clearly,

$$\pi > 3 \quad \Leftrightarrow \quad \frac{\pi}{2} - 1 > \frac{3}{2} - 1 \quad \Leftrightarrow \quad \left(\frac{\pi}{2} - 1\right)\sigma^2 > \frac{\sigma^2}{2}$$

and so

$$\text{Var}\left(\sqrt{(\pi/2)}|X|\right) = \left(\frac{\pi}{2} - 1\right)\sigma^2 > \frac{\sigma^2}{2} = \frac{1}{I(\sigma)}$$

for every value of $\sigma > 0$, as was to be shown. □

6. (p. 444 #10) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a normal distribution for which the mean is 0 and the standard deviation σ is unknown ($\sigma > 0$). Find the lower bound specified by the information inequality for the variance of any unbiased estimator of $\log\sigma$.

Solution. Let T be an unbiased estimator of $\log\sigma$, then $E(T) = \log\sigma = g(\sigma)$ hence $g'(\sigma) = 1/\sigma$.

$$\begin{aligned}\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \ln f_X(x_i|0,\sigma^2) &= \sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \ln\left[\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right\}\right] \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial\sigma}\left[-\frac{1}{2}\ln(2\pi) - \ln\sigma - \frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right] \\ &= \sum_{i=1}^{n} -\frac{1}{\sigma} + \frac{{x_i}^2}{\sigma^3} \\ &= -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{x_i}{\sigma}\right)^2\end{aligned}$$

Then

$$\begin{aligned}E\left[\left\{\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \ln f_X(x_i|0,\sigma^2)\right\}^2\right] &= E\left[\frac{n^2}{\sigma^2} - \frac{2n}{\sigma^2}\sum_{i=1}^{n}\left(\frac{x_i}{\sigma}\right)^2 + \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sigma}\right)^2\right)^2\right] \\ &= \frac{n^2}{\sigma^2} - \frac{2n(n)}{\sigma^2} + \frac{1}{\sigma^2}(2n + n^2) = \frac{2n}{\sigma^2}\end{aligned}$$

as $\sum_{i=1}^{n}(x_i/\sigma)^2 \sim {\chi^2}_n$. Note this also confirms that the Fisher information of a random sample is n times the Fisher information of a single random variable ($= nI(\sigma)$).
The lower bound specified by the information inequality for the variance of T is then

$$\frac{[g'(\sigma)]^2}{E\left[\left\{\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \ln f_X(x_i|0,\sigma^2)\right\}^2\right]} = \frac{(1/\sigma)^2}{2n/\sigma^2} = \frac{1}{2n}\ .$$

Free of σ, interesting ... □

7. (p. 444 #13) Determine what is wrong with the following argument:
Suppose that the random variable X has a uniform distribution on the interval $[0, \theta]$, where the value of θ is unknown ($\theta > 0$). Then $f_X(x|\theta) = 1/\theta$, $\lambda(x|\theta) = -\log\theta$ and $\lambda'(x|\theta) = -(1/\theta)$. Therefore,

$$I(\theta) = E\{[\lambda'(X|\theta)]^2\} = \frac{1}{\theta^2}\ .$$

Since $2X$ is an unbiased estimator of θ, the information inequality states that

$$\text{Var}(2X) \geq \frac{1}{I(\theta)} = \theta^2\ .$$

But

$$\text{Var}(2X) = 4\text{Var}(X) = 4 \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3} < \theta^2 \ .$$

Hence, the information inequality is not correct.

Solution. The information inequality does not apply to a uniform distribution on the interval $[0, \theta]$, because the Fisher information of distribution cannot be computed. Specifically, when computing the Fisher information it is assumed that the order in which the probability density function or probability function is integrated with respect to x and differentiated with respect to θ may be reversed. In this case, the parameter θ appears as an endpoint in the interval where $f_X(x|\theta) > 0$ which makes this reversal impossible. Hence, a false assumption has been made in the proof which invalids the result. □

8. (p. 444 #14) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a gamma distribution for which the value of the parameter α is unknown and the value of β is known. Show that if n is large the distribution of the maximum likelihood estimator of α will be approximately a normal distribution with mean α and variance

$$\frac{[\Gamma(\alpha)]^2}{n\{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2\}} \ .$$

Solution. It suffices to show that

$$I(\alpha) = \frac{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2}{[\Gamma(\alpha)]^2} \ ,$$

for the asymptotic distribution of the maximum likelihood estimator of α shall be normal with mean α and variance $1/[nI(\alpha)]$. Using the book's definition of the gamma distribution,

$$\begin{aligned}
\lambda''(x|\alpha, \beta) &= \frac{\partial^2}{(\partial\alpha)^2} \ln f_X(x|\alpha, \beta) \\
&= \frac{\partial^2}{(\partial\alpha)^2} \ln \left[\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \right] \\
&= \frac{\partial^2}{(\partial\alpha)^2} [\alpha \ln \beta - \ln \Gamma(\alpha) + \alpha \ln x - \ln x - \beta x] \\
&= \frac{\partial}{\partial\alpha} \left[\ln \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \ln x \right] \\
&= \frac{\partial}{\partial\alpha} \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma'(\alpha)\Gamma'(\alpha)}{[\Gamma(\alpha)]^2} \\
&= \frac{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2}{[\Gamma(\alpha)]^2} \ .
\end{aligned}$$

Hence, the variance of the maximum likelihood estimator is

$$1/[nI(\alpha)] = \frac{[\Gamma(\alpha)]^2}{n\{\Gamma(\alpha)\Gamma''(\alpha) - [\Gamma'(\alpha)]^2\}}$$

as was to be shown. □

9. (p. 415 #4) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a normal distribution for which the mean μ is unknown and the variance σ^2 is known. How large a random sample must be taken in order that there will be a confidence interval for μ with confidence coefficient 0.95 and length less than 0.01σ?

Solution. We know that the maximum likelihood estimator of μ, $\overline{X}$, has normal distribution with mean μ and variance σ^2/n. Let the random variable Z be defined as

$$Z = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \quad \Rightarrow \quad Z \sim N(0,1) \ ,$$

it follows that

$$P(-1.96 \le Z \le 1.96) = \Phi(1.96) - \Phi(-1.96) = 0.95 \ .$$

Hence,

$$P\left(\overline{X} - \frac{\sigma}{\sqrt{n}}(1.96) \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}}(1.96)\right) = 0.95$$

and so a 95% confidence interval for μ is $(\overline{x} - \sigma(1.96)/\sqrt{n},\ \overline{x} + \sigma(1.96)/\sqrt{n})$. As we have shown (in-class assignment 04/06/09) this is the shortest confidence interval, having length $2(\sigma(1.96)/\sqrt{n})$. It follows that

$$2\left(\frac{\sigma(1.96)}{\sqrt{n}}\right) < 0.01\sigma \quad \Leftrightarrow \quad \left(\frac{3.92}{0.01}\right)^2 < n \quad \Leftrightarrow \quad 153664 < n \ ,$$

that is, a random sample must be larger than 153,664 in order to ensure there will be a confidence interval for μ with confidence coefficient 0.95 and length less than 0.01σ. □

10. (p. 416 #11) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a Bernoulli distribution with parameter p. Let $\overline{X}_n$ be the sample average. Use the variance stabilizing transformation found in Exercise 29 of Section 5.13 to construct an approximate coefficient γ confidence interval for p.

Solution. Since $\bar{X}_n$ is the MLE and we know (from central limit theorem) that $\sqrt{n}(\bar{X}_n - p) \dot{\sim} N(0, p(1-p))$, the variance stablizing transformation can be found as:

$$h(p) = \int_0^p \frac{1}{\sqrt{x(x-1)}} dx = \arcsin\sqrt{p}.$$

Then we have,

$$Q = \sqrt{n}(\arcsin\sqrt{\bar{X}_n} - \arcsin\sqrt{p}) \ \dot{\sim}\ N(0,1).$$

Now using Q as a pivotal quantity,

$$\begin{aligned} &\gamma = P(Z_{\frac{1-\gamma}{2}} < Q < -Z_{\frac{1-\gamma}{2}}) \\ &\Leftrightarrow \gamma = P\left(\left[\sin\left(\arcsin\sqrt{\bar{X}_n} + \frac{1}{\sqrt{n}}Z_{\frac{1-\gamma}{2}}\right)\right]^2 < p < \left[\sin\left(\arcsin\sqrt{\bar{X}_n} - \frac{1}{\sqrt{n}}Z_{\frac{1-\gamma}{2}}\right)\right]^2\right) \end{aligned}$$

where $Z_{\frac{1-\gamma}{2}} = \Phi^{-1}(\frac{1-\gamma}{2})$ and $\Phi(\cdot)$ is the CDF of $N(0,1)$. So a $100\gamma\%$ CI estimator for p is:

$$\left(\left[\sin\left(\arcsin\sqrt{\bar{X}_n} + \frac{1}{\sqrt{n}}Z_{\frac{1-\gamma}{2}}\right)\right]^2, \left[\sin\left(\arcsin\sqrt{\bar{X}_n} - \frac{1}{\sqrt{n}}Z_{\frac{1-\gamma}{2}}\right)\right]^2\right).$$

□