Statistics 252 "Practice Exam" (Solutions) – Winter 2006

1. Let $U = \theta^2 Y$ so that for $u > 0$,

$$P(U \le u) = P(Y \le \theta^{-2}u) = \int_0^{\theta^{-2}u} \theta^2 e^{-\theta^2 y}\, dy = 1 - e^{-u}.$$

Thus, we must find a and b so that

$$\int_0^a e^{-u}\, du = \alpha_1 \quad \text{and} \quad \int_b^\infty e^{-u}\, du = \alpha_2.$$

Computing the integrals we find $a = -\log(1-\alpha_1)$ and $b = -\log(\alpha_2)$. Hence,

$$\begin{aligned} 1 - (\alpha_1 + \alpha_2) = P(a \le U \le b) &= P(-\log(1-\alpha_1) \le \theta^2 Y \le -\log(\alpha_2)) \\ &= P\left(\frac{-\log(1-\alpha_1)}{Y} \le \theta^2 \le \frac{-\log(\alpha_2)}{Y} \right) \\ &= P\left(\sqrt{\frac{-\log(1-\alpha_1)}{Y}} \le \theta \le \sqrt{\frac{-\log(\alpha_2)}{Y}} \right). \end{aligned}$$

In other words,

$$\left(\sqrt{\frac{-\log(1-\alpha_1)}{Y}}, \sqrt{\frac{-\log(\alpha_2)}{Y}} \right)$$

is a confidence interval for θ with coverage probability $1 - (\alpha_1 + \alpha_2)$.

2. (a) If $Y \sim \text{Unif}(0, \theta)$, then since $\mathbb{E}(Y) = \theta/2$ we find that setting $\mathbb{E}(Y) = \overline{Y}$ gives

$$\hat{\theta}_{\text{MOM}} = 2\overline{Y}.$$

Since

$$\mathbb{E}(\hat{\theta}_{\text{MOM}}) = 2\mathbb{E}(\overline{Y}) = 2\mathbb{E}(Y_1) = 2 \cdot \frac{\theta}{2} = \theta$$

we conclude that $\hat{\theta}_{\text{MOM}}$ is an unbiased estimator of θ.

(b) In order to find $\mathbb{E}(Y_{(10)})$ we must find the density function of $Y_{(10)}$. Now,

$$P(Y_{(10)} \le t) = \left[\int_0^t \theta^{-1}\, dy \right]^{10} = \frac{t^{10}}{\theta^{10}}, \quad 0 \le t \le \theta$$

so that $f(t) = 10\,\theta^{-10}\,t^9$, $0 \le t \le \theta$. Thus,

$$\mathbb{E}(Y_{(10)}) = \int_0^\theta 10\,\theta^{-10}\,t^{10}\, dt = \frac{10}{11}\,\theta.$$

An unbiased estimator of θ which is a multiple of $Y_{(10)}$ is therefore given by

$$\hat{\theta}_B = \frac{11}{10} \max(Y_1, \ldots, Y_{10}).$$

Also, note that

$$\mathbb{E}(Y_{(10)}^2) = \int_0^\theta 10\,\theta^{-10}\,t^{11}\, dt = \frac{10}{12}\,\theta^2.$$

(c) From **(a)**, we conclude

$$\mathrm{Var}(\hat{\theta}_{\mathrm{MOM}}) = 4\,\mathrm{Var}(\overline{Y}) = \frac{4}{10}\,\mathrm{Var}(Y_1) = \frac{4\theta^2}{10 \cdot 12} = \frac{\theta^2}{30}.$$

From **(b)**, we conclude

$$\mathrm{Var}(\hat{\theta}_B) = \frac{121}{100}\,\mathrm{Var}(\max(Y_1, \ldots, Y_{10})) = \frac{121}{100}\left(\frac{10}{12} - \frac{100}{121}\right)\theta^2 = \frac{\theta^2}{120}.$$

Thus,

$$\mathrm{eff}(\hat{\theta}_{\mathrm{MOM}}, \hat{\theta}_B) = \frac{\mathrm{Var}(\hat{\theta}_B)}{\mathrm{Var}(\hat{\theta}_{\mathrm{MOM}})} = \frac{1}{4}.$$

Both $\hat{\theta}_{\mathrm{MOM}}$ and $\hat{\theta}_B$ are unbiased so we prefer the one with the smaller variance, namely $\hat{\theta}_B$.

3. (a) Suppose that $\hat{\theta}$ is an estimator of θ. The random interval $[L(\hat{\theta}), U(\hat{\theta})]$ is a 93% confidence interval for θ if

$$P(L(\hat{\theta}) \le \theta \le U(\hat{\theta})) = 0.93.$$

Hence, we interpret a 93% confidence interval to mean that before the data have been observed, there is a 93% chance that the parameter will lie in the random interval. However, once the data have been observed, no such probability statement is true. Either the given interval does or does not contain θ. Alternatively, if many, many intervals are observed, each constructed using the same formula, then the long-run average that will contain θ is 0.93.

(b) It is desirable to find unbiased estimators because by having an unbiased estimator we know $\mathbb{E}(\hat{\theta}) = \theta$; that is, the most likely "value" of $\hat{\theta}$ is θ. If we have the unbiased estimator with the smallest variance, then the distribution of $\hat{\theta}$ is clustered as tightly as possible about θ. Thus, the MVUE is the "most likely" of all unbiased estimators to be "closest" to θ.

4. (a) Since

$$\log f(y|\theta) = \log(y) - 2\log(\theta) - \frac{y^2}{2\theta^2},$$

we find

$$\frac{\partial^2}{\partial\theta^2}\log f(y|\theta) = \frac{2}{\theta^2} - \frac{3y^2}{\theta^4}.$$

Thus,

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right) = \frac{3\mathbb{E}(Y^2)}{\theta^4} - \frac{2}{\theta^2} = \frac{4}{\theta^2}.$$

(b) To find $\hat{\theta}_{\mathrm{MOM}}$ we solve the equation $\mathbb{E}(Y) = \overline{Y}$ for θ. This implies

$$\hat{\theta}_{\mathrm{MOM}} = \sqrt{\frac{2}{\pi}}\,\overline{Y}.$$

(c)

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_{\mathrm{MOM}}) = \frac{2}{\pi}\,\mathrm{Var}\,\overline{Y} = \frac{2}{n\pi}\,\mathrm{Var}\,Y_1 = \frac{2}{n\pi}\left(\mathbb{E}(Y_1^2) - [\mathbb{E}(Y_1)]^2\right) &= \frac{2}{n\pi}\left(2 - \frac{\pi}{2}\right)\theta^2 \\ &= \left(\frac{4-\pi}{n\pi}\right)\theta^2\end{aligned}$$

5. (a) We find that

$$\mathbb{E}(\overline{Y}) = \mathbb{E}(Y_1) = 252\,\theta.$$

Thus, if

$$\hat{\theta}_A = \frac{\overline{Y}}{252} = \frac{1}{252n}\sum_{i=1}^{n} Y_i$$

then $\hat{\theta}_A$ is an unbiased estimator of θ.

(b) Since

$$\log f(y|\theta) = -252\log(\theta) - \log(251!) + 251\log(y) - \frac{y}{\theta},$$

we find

$$\frac{\partial^2}{\partial\theta^2}\log f(y|\theta) = \frac{252}{\theta^2} - \frac{2y}{\theta^3}.$$

Thus,

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right) = -\frac{252}{\theta^2} + \frac{2\mathbb{E}(Y)}{\theta^3} = \frac{252}{\theta^2}.$$

(c) The Cramer-Rao inequality tells us that an unbiased estimator $\hat{\theta}$ of θ must satisfy

$$\operatorname{Var}(\hat{\theta}) \geq \frac{1}{nI(\theta)} = \frac{\theta^2}{252n}.$$

Since

$$\operatorname{Var}(\hat{\theta}_A) = \frac{1}{252^2 n}\operatorname{Var} Y_1 = \frac{1}{252^2}\cdot(252\,\theta^2) = \frac{\theta^2}{252n},$$

we have found an unbiased estimator whose variance attains the lower bound of the Cramer-Rao inequality. Hence, $\hat{\theta}_A$ must be the MVUE of θ.