Statistics 252 Winter 2007 Midterm #2 – Solutions

1. If $Y \sim \text{Uniform}(\theta, 2\theta)$, then $f_Y(y|\theta) = \theta^{-1}$, $\theta \le y \le 2\theta$. Let $U = Y/\theta$ so that if $1 \le u \le 2$, then

$$P(U \le u) = P(Y \le \theta u) = \int_\theta^{\theta u} \frac{1}{\theta}\, dy = u - 1,$$

and so $f_U(u) = 1$, $1 \le u \le 2$. We must now find a and b such that $\int_1^a du = \frac{\alpha}{2}$ and $\int_b^2 du = \frac{\alpha}{2}$. Solving gives $a = 1 + \frac{\alpha}{2}$ and $b = 2 - \frac{\alpha}{2}$, and so $1 - \alpha = P(a \le U \le b)$ or, in other words,

$$1 - \alpha = P\left(1 + \frac{\alpha}{2} \le \frac{Y}{\theta} \le 2 - \frac{\alpha}{2}\right) = P\left(\frac{2+\alpha}{2} \le \frac{Y}{\theta} \le \frac{4-\alpha}{2}\right) = P\left(\frac{2Y}{4-\alpha} \le \theta \le \frac{2Y}{2+\alpha}\right).$$

The required confidence interval for θ with coverage probability $1 - \alpha$ is therefore

$$\left[\frac{2Y}{4-\alpha}, \frac{2Y}{2+\alpha}\right].$$

2. (a) If we write $a(\theta) = 3\theta$, $b(y) = y^2$, $c(\theta) = \theta$, $d(y) = y^3$, $\alpha = 0$, and $\beta = \infty$, then we see that $f_Y(y|\theta)$ does, in fact, belong to an exponential family.

2. (b) The likelihood function is given by

$$L(\theta) = \prod_{i=1}^n f_Y(y_i|\theta) = 3^n \theta^n \left(\prod_{i=1}^n y_i^2\right) \exp\left\{-\theta \sum_{i=1}^n y_i^3\right\}.$$

2. (c) In order to maximize $L(\theta)$ we will try to maximize $\ell(\theta)$ instead. Therefore,

$$\ell(\theta) = n \log 3 + n \log \theta + 2\sum_{i=1}^n \log y_i - \theta \sum_{i=1}^n y_i^3$$

and so

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n y_i^3.$$

Setting $\ell'(\theta) = 0$ implies

$$\theta = \frac{n}{\sum\limits_{i=1}^n y_i^3}.$$

Since $\ell''(\theta) = -\frac{n}{\theta^2} < 0$ we conclude from the second derivative test that

$$\hat{\theta}_{\text{MLE}} = \frac{n}{\sum\limits_{i=1}^n Y_i^3}.$$

2. (d) If we let $u = \sum_{i=1}^n y_i^3$, $g(u, \theta) = \theta^n \exp\{-\theta u\}$, and $h(y_1, \ldots, y_n) = 3^n \prod_{i=1}^n y_i^2$, then $L(\theta) = g(u, \theta) \cdot h(y_1, \ldots, y_n)$ so from the Factorization Theorem we conclude that

$$U = \sum_{i=1}^n Y_i^3$$

is sufficient for the estimation of θ.

2. (e) Let $T(U) = \frac{n}{U}$. Since T is a one-to-one function, and since any one-to-one function of a sufficient statistic is also sufficient, we conclude that

$$T\left(\sum_{i=1}^{n} Y_i^3\right) = \frac{n}{\sum_{i=1}^{n} Y_i^3} = \hat{\theta}_{\text{MLE}}$$

is sufficient for the estimation of θ.

2. (f) Since $\log f_Y(y|\theta) = \log 3 + \log\theta + 2\log y - \theta y^3$ so that

$$\frac{\partial}{\partial\theta}\log f_Y(y|\theta) = \frac{1}{\theta} - y^3 \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2}\log f_Y(y|\theta) = -\frac{1}{\theta^2},$$

we find

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f_Y(Y|\theta)\right) = -E\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}.$$

2. (g) Since an approximate $1-\alpha$ confidence interval for θ is given by

$$\left[\hat{\theta}_{\text{MLE}} - z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta}_{\text{MLE}})}}\,,\ \hat{\theta}_{\text{MLE}} + z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta}_{\text{MLE}})}}\right],$$

we conclude that

$$\left[\frac{n}{\sum_{i=1}^{n} Y_i^3} - 1.96\frac{\sqrt{n}}{\sum_{i=1}^{n} Y_i^3}\,,\ \frac{n}{\sum_{i=1}^{n} Y_i^3} + 1.96\frac{\sqrt{n}}{\sum_{i=1}^{n} Y_i^3}\right]$$

is the required 95% confidence interval.

3. In order to determine the method of moments estimator of θ we equate the first population moment and the first sample moment, $E(Y) = \overline{Y}$, and solve for θ. Since

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y|\theta)\, dy = \int_0^1 y\,\theta y^{\theta-1}\, dy = \theta\int_0^1 y^\theta\, dy = \frac{\theta}{\theta+1}$$

we conclude $\frac{\theta}{\theta+1} = \overline{Y}$ and so solving for θ gives $\hat{\theta}_{\text{MOM}} = \frac{\overline{Y}}{1-\overline{Y}}$.

4. The significance level of this hypothesis test is

$$\alpha = P_{H_0}(\text{reject } H_0) = P_{\mu=0}(\overline{Y} > 7.84/\sqrt{n}) = P\left(\frac{\overline{Y}-0}{4/\sqrt{n}} > \frac{7.84/\sqrt{n}-0}{4/\sqrt{n}}\right) = P(Z > 1.96) \approx 0.025$$

where $Z \sim \mathcal{N}(0,1)$ and the last step follows from Table 4.

5. (a) The statement of the Cramer-Rao inequality is as follows. Suppose that $Y_1, \ldots, Y_n$ are i.i.d. with $Y_i \sim f_Y(y|\theta)$. Suppose further that f is a "smooth" function (continuous and differentiable). Let $\hat{\theta}$ be an unbiased estimator of θ based on $Y_1, \ldots, Y_n$. Then $\text{Var}(\hat{\theta}) \geq \frac{1}{nI(\theta)}$.

5. (b) Suppose that the variance of an unbiased estimator $\hat{\theta}$ of θ is $\frac{1}{nI(\theta)}$ and that the other assumptions of the statement in **(a)** have been met. Since the lower bound of the Cramer-Rao inequality has been attained, we know that no other unbiased estimator can have smaller variance than $\hat{\theta}$. Hence, $\hat{\theta}$ must be the MVUE of θ.