Statistics 252–Mathematical Statistics
Winter 2007 (200710)
Final Exam Solutions

Instructor: Michael Kozdron

1. (a) By definition, the likelihood function $L(\theta)$ is given by

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

1. (b) In order to maximize $L(\theta)$, we attempt to maximize the log-likelihood function

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

We find that

$$\ell'(\theta) = \frac{d}{d\theta}\ell(\theta) = \frac{2n}{\theta} - 2\theta \sum_{i=1}^{n} y_i^2$$

and setting $\ell'(\theta) = 0$ implies that

$$\theta^2 = \frac{n}{\displaystyle\sum_{i=1}^{n} y_i^2} \quad \text{and so} \quad \theta = \sqrt{\frac{n}{\displaystyle\sum_{i=1}^{n} y_i^2}}.$$

Since

$$\ell''(\theta) = \frac{-2n}{\theta^2} - 2\sum_{i=1}^{n} y_i^2 < 0$$

for all θ, the second derivative test implies

$$\hat{\theta}_{\text{MLE}} = \sqrt{\frac{n}{\displaystyle\sum_{i=1}^{n} Y_i^2}}.$$

1. (c) If we let

$$u = \sum_{i=1}^{n} y_i^2, \quad h(y_1, \ldots, y_n) = 2^n \left(\prod_{i=1}^{n} y_i\right), \quad \text{and} \quad g(u, \theta) = \theta^{2n} \exp\left\{-\theta^2 u\right\},$$

then since $L(\theta) = h(y_1, \ldots, y_n) \cdot g(u, \theta)$ we conclude from the Factorization Theorem that $U = \sum_{i=1}^{n} Y_i^2$ is sufficient for the estimation of θ. We now recall that any one-to-one function of a sufficient statistic is also sufficient. Suppose that

$$T(U) = \sqrt{\frac{n}{U}}.$$

Since T is a one-to-one function, we conclude that

$$T(U) = \sqrt{\frac{n}{\sum_{i=1}^{n} Y_i^2}} = \hat{\theta}_{\text{MLE}}$$

is also a sufficient statistic for the estimation of θ.

1. (d) We find

$$\log f_Y(y|\theta) = \log 2 + 2\log\theta + \log y - \theta^2 y^2$$

so that

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

The Fisher information is given by

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

and so we see that we must now compute $\mathbb{E}(Y^2)$. Therefore,

$$\mathbb{E}(Y^2) = \int_{-\infty}^{\infty} y^2\, f_Y(y|\theta)\, dy = \int_0^{\infty} 2\theta^2\, y^3\, e^{-\theta^2 y^2}\, dy = \frac{1}{\theta^2}\int_0^{\infty} ue^{-u}\, du = \frac{1}{\theta^2}\Gamma(2) = \frac{1}{\theta^2}$$

making the substitution $u = \theta^2 y^2$, $du = 2\theta^2 y\, dy$. We therefore conclude that

$$I(\theta) = \frac{2}{\theta^2} + 2\mathbb{E}(Y^2) = \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{4}{\theta^2}.$$

1. (e) Recall that a $(1-\alpha)$ confidence interval based on the Fisher information and the maximum likelihood estimator is given by

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

Using the results of **(b)** and **(d)** we conclude that the required confidence interval is

$$\left[\sqrt{\frac{n}{\sum_{i=1}^{n} Y_i^2}} - \frac{2.58}{2\sqrt{\sum_{i=1}^{n} Y_i^2}}\,,\ \sqrt{\frac{n}{\sum_{i=1}^{n} Y_i^2}} + \frac{2.58}{2\sqrt{\sum_{i=1}^{n} Y_i^2}}\right] \quad \text{or} \quad \left[\frac{\sqrt{n} - 1.29}{\sqrt{\sum_{i=1}^{n} Y_i^2}}\,,\ \frac{\sqrt{n} + 1.29}{\sqrt{\sum_{i=1}^{n} Y_i^2}}\right]$$

since $z_{0.005} = 2.58$ from Table 2.

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

$$P(U \le u) = P(Y^2 \le u/\theta) = P(Y \le \sqrt{u/\theta}) = \int_0^{\sqrt{u/\theta}} \frac{2\theta y}{(1+\theta y^2)^2}\, dy.$$

Therefore, the density of U is given by

$$f_U(u) = \frac{d}{du} P(U \le u) = \frac{2\theta\sqrt{u/\theta}}{(1+\theta(\sqrt{u/\theta})^2)^2} \cdot \frac{d}{du}\sqrt{u/\theta} = \frac{2\sqrt{\theta u}}{(1+u)^2} \cdot \frac{1}{2\sqrt{\theta u}} = \frac{1}{(1+u)^2}, \quad u > 0.$$

Thus, we must find a and b so that

$$\int_0^a \frac{1}{(1+u)^2}\, du = \frac{\alpha}{2} \quad \text{and} \quad \int_b^\infty \frac{1}{(1+u)^2}\, du = \frac{\alpha}{2}.$$

Computing the integrals we find

$$1 - \frac{1}{1+a} = \frac{\alpha}{2} \quad \text{and} \quad \frac{1}{1+b} = \frac{\alpha}{2} \quad \text{so that} \quad a = \frac{\alpha}{2-\alpha} \quad \text{and} \quad b = \frac{2-\alpha}{\alpha}.$$

Hence,

$$1 - \alpha = P(a \le U \le b) = P\left(\frac{\alpha}{2-\alpha} \le \theta Y^2 \le \frac{2-\alpha}{\alpha}\right) = P\left(\frac{\alpha}{(2-\alpha)\,Y^2} \le \theta \le \frac{2-\alpha}{\alpha\, Y^2}\right).$$

In other words,

$$\left[\frac{\alpha}{(2-\alpha)\,Y^2}, \frac{2-\alpha}{\alpha\, Y^2}\right]$$

is a confidence interval for θ with coverage probability $1 - \alpha$.

2. (b) If $\alpha = 0.10$ and we observe $y = 2$, then the observed confidence interval is

$$\left[\frac{0.10}{(2-0.10)\,2^2}, \frac{2-0.10}{0.10 \cdot 2^2}\right] \quad \text{or} \quad \left[\frac{1}{76}, \frac{19}{4}\right] \quad \text{or, approximately,} \quad [0.013, 4.75].$$

Since $\theta_0 = 3$ lies in this interval, we conclude that there is not sufficient evidence to reject $H_0 : \theta_0 = 3$ in favour of $H_A : \theta \neq 3$ at the $\alpha = 0.10$ significance level.

3. (a) We find that $f_0(y) = 1$ for $0 < y < 1$ and $f_A(y) = 1 - \frac{1}{4}(y - 1/2) = \frac{9}{8} - \frac{y}{4}$ for $0 < y < 1$. Therefore, the likelihood ratio is

$$\Lambda(y) = \frac{f_0(y)}{f_A(y)} = \frac{1}{\frac{9}{8} - \frac{y}{4}} = \frac{8}{9 - 2y}$$

for $0 < y < 1$, and so the rejection region is

$$RR = \{\Lambda(Y) < c\} = \left\{\frac{8}{9-2Y} < c\right\} = \left\{Y < \frac{9c-8}{2c}\right\} = \{Y < c'\}$$

where $c' = \frac{9c-8}{2c}$ is another constant. We now choose c (or, equivalently, c') so that this test has the desired significance level. Since

$$\alpha = P_{H_0}(\text{reject } H_0) = P_{\theta=0}(Y < c') = \int_0^{c'} f_0(y)\,dy = \int_0^{c'} 1\,dy = c'$$

we conclude that $c' = \alpha$ and so $RR = \{Y < \alpha\}$.

3. (b) By definition, power $= P_{H_A}(\text{reject } H_0)$. From part **(a)** we know that $RR = \{Y < \alpha\}$ and so we compute

$$\text{power} = P_{\theta=1/2}(Y < \alpha) = \int_0^{\alpha} f_A(y)\,dy = \int_0^{\alpha} \frac{9}{8} - \frac{y}{4}\,dy = \frac{9}{8}\alpha - \frac{1}{8}\alpha^2 = \frac{9\alpha(1-\alpha)}{8}.$$

3. (c) If $\alpha = 0.10$, then $RR = \{Y < 0.10\}$ is the rejection region of test constructed in **(a)**. Hence, if a single observation produces $y = 0.25$, we conclude that since 0.25 is not smaller than 0.10 there is not sufficient evidence to reject H_0 in favour of H_A at the $\alpha = 0.10$ significance level.

4. (a) The generalized likelihood ratio test for the simple null hypothesis $H_0 : \theta = \theta_0$ against the composite alternative hypothesis $H_A : \theta \neq \theta_0$ has rejection region $\{\Lambda < c\}$ where

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta}_{\text{MLE}})}$$

is the generalized likelihood ratio and $L(\theta)$ is the likelihood function. In this case,

$$L(\theta) = \theta^n \left(\prod_{i=1}^{n} y_i\right)^{\theta-1}$$

and $\theta_0 = 1$ so that

$$\begin{aligned}
\Lambda = \frac{1^n \left(\prod y_i\right)^{1-1}}{\hat{\theta}_{\text{MLE}}^n \left(\prod y_i\right)^{\hat{\theta}_{\text{MLE}}-1}} = \hat{\theta}_{\text{MLE}}^{-n} \left(\prod y_i\right)^{1-\hat{\theta}_{\text{MLE}}} &= \left(\frac{\sum \log y_i}{-n}\right)^n \left(\prod y_i\right)^{1+\frac{n}{\sum \log y_i}} \\
&= \left(\frac{\sum \log y_i}{-n}\right)^n \left(e^{\sum \log y_i}\right)^{1+\frac{n}{\sum \log y_i}} \\
&= n^{-n} \left(-\sum_{i=1}^{n} \log y_i\right)^n \exp\left\{\sum_{i=1}^{n} \log y_i + n\right\}
\end{aligned}$$

4. (b) We know that

$$-2\log\Lambda = -2\left[n \log\left(\frac{-\sum \log Y_i}{n}\right) + \sum \log Y_i + n\right] \overset{\text{approx}}{\sim} \chi^2(1).$$

Since the observed data imply

$$-2\log\Lambda = -2\left[8\log\left(\frac{4}{8}\right) - 4 + 8\right] \approx 3.09,$$

and since $\chi^2_{1,0.10} = 2.70554$ from Table 6, we conclude that there is sufficient evidence to reject H_0 in favour of H_A at the $\alpha = 0.10$ significance level.

5. Since Y_1, Y_2, Y_3 are independent normal random variables, we know that $Y_1+Y_2+Y_3$ also has a normal distribution with mean $\mathbb{E}(Y_1)+\mathbb{E}(Y_2)+\mathbb{E}(Y_3) = 4+4+4 = 12$ and variance $\mathrm{Var}(Y_1)+\mathrm{Var}(Y_2)+\mathrm{Var}(Y_3) = 2+2+2 = 6$. Therefore,

$$P(Y_1+Y_2+Y_3 > 13) = P\left(\frac{Y_1+Y_2+Y_3-12}{\sqrt{6}} > \frac{13-12}{\sqrt{6}}\right) \approx P(Z > 0.4082) \approx 0.3416$$

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

6. Let $X := \min\{Y_1, Y_2\}$ so that

$$P(X > x) = P(Y_1 > x,\, Y_2 > x) = P(Y_1 > x)P(Y_2 > x) = [P(Y_1 > x)]^2$$

since Y_1 and Y_2 are independent and identically distributed. Thus, we calculate (for $x > 4$) that

$$P(Y_1 > x) = \int_x^\infty \frac{1}{2}\exp\left\{-\frac{1}{2}(y-4)\right\} dy = \int_{x-4}^\infty \frac{1}{2}e^{-u/2}\, du = -e^{-u/2}\Big|_{x-4}^{\infty} = \exp\left\{-\frac{1}{2}(x-4)\right\}$$

making the substitution $u = y-4$, $du = dy$, and so

$$P(X \le x) = 1 - P(X > x) = 1 - \left[\exp\left\{-\frac{1}{2}(x-4)\right\}\right]^2 = 1 - e^{4-x}$$

from which we conclude that

$$f_X(x) = \frac{d}{dx}P(X \le x) = e^{4-x}, \quad x > 4.$$

7. To find the method of moments estimators for α and β, we must solve the system of equations

$$\mathbb{E}(Y) = \overline{Y} \quad \text{and} \quad \mathbb{E}(Y^2) = \frac{1}{n}\sum_{i=1}^n Y_i^2.$$

Since $\mathbb{E}(Y) = \frac{\alpha+\beta}{2}$ and $\mathrm{Var}\, Y = \frac{(\beta-\alpha)^2}{12}$, we find

$$\mathbb{E}(Y^2) = \mathrm{Var}\, Y + [\mathbb{E}(Y)]^2 = \frac{(\beta-\alpha)^2}{12} + \frac{(\alpha+\beta)^2}{4} = \frac{\alpha^2+\alpha\beta+\beta^2}{3}.$$

Thus,

$$\alpha+\beta = 2\overline{Y} \quad \text{and} \quad \alpha^2+\alpha\beta+\beta^2 = \frac{3}{n}\sum_{i=1}^n Y_i^2.$$

Since $\alpha = 2\overline{Y} - \beta$, write $\alpha^2 + \alpha\beta + \beta^2 = \alpha(\alpha+\beta) + \beta^2$ to see that

$$\alpha^2+\alpha\beta+\beta^2 = 2\overline{Y}(2\overline{Y}-\beta)+\beta^2 = \beta^2 - 2\overline{Y}\beta + 4\overline{Y}^2 = \beta^2 - 2\overline{Y}\beta + \overline{Y}^2 + 3\overline{Y}^2 = (\beta-\overline{Y})^2 + 3\overline{Y}^2.$$

Hence, we find

$$(\beta - \overline{Y})^2 + 3\overline{Y}^2 = \frac{3}{n}\sum_{i=1}^n Y_i^2 \quad \text{so that} \quad \beta = \overline{Y} + \sqrt{\frac{3}{n}\sum_{i=1}^n Y_i^2 - 3\overline{Y}^2}$$

and

$$\alpha = 2\overline{Y} - \beta = \overline{Y} - \sqrt{\frac{3}{n}\sum_{i=1}^n Y_i^2 - 3\overline{Y}^2}.$$

In other words,

$$\hat{\alpha}_{\text{MOM}} = \overline{Y} - \sqrt{3}\sqrt{\frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2} \quad \text{and} \quad \hat{\beta}_{\text{MOM}} = \overline{Y} + \sqrt{3}\sqrt{\frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2}.$$

8. (a) The method of moments estimator of θ is found by equating the first population moment $\mathbb{E}(Y)$ with the first sample moment $\overline{Y}$ and solving for θ. We find

$$\mathbb{E}(Y) = 0 \cdot P(Y=0) + 1 \cdot P(Y=1) = 0 + \theta^1(1-\theta)^{1-1} = \theta$$

and so we conclude that $\hat{\theta}_{\text{MOM}} = \overline{Y}$.

8. (b) Since $\mathbb{E}(Y_1) = \cdots = \mathbb{E}(Y_n) = \theta$ we conclude that

$$\mathbb{E}(\hat{\theta}_{\text{MOM}}) = \mathbb{E}(\overline{Y}) = \frac{\mathbb{E}(Y_1) + \cdots + \mathbb{E}(Y_n)}{n} = \frac{n\theta}{n} = \theta$$

which shows that $\hat{\theta}_{\text{MOM}}$ is, in fact, an unbiased estimator of θ.

8. (c) We begin by calculating

$$\mathbb{E}(Y^2) = 0^2 \cdot P(Y=0) + 1^2 \cdot P(Y=1) = 0 + \theta^1(1-\theta)^{1-1} = \theta$$

so that $\text{Var}(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = \theta - \theta^2 = \theta(1-\theta)$. Therefore,

$$\text{Var}(\hat{\theta}_{\text{MOM}}) = \text{Var}(\overline{Y}) = \frac{\text{Var}(Y_1) + \cdots + \text{Var}(Y_n)}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}.$$

8. (d) By definition, the likelihood function $L(\theta)$ is given by

$$L(\theta) = \prod_{i=1}^n f_Y(y_i|\theta) = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} = \theta^{n\overline{y}}(1-\theta)^{n-n\overline{y}}.$$

In order to maximize $L(\theta)$, we attempt to maximize the log-likelihood function

$$\ell(\theta) = \log L(\theta) = n\overline{y}\log\theta - (n - n\overline{y})\log(1-\theta).$$

We find that

$$\ell'(\theta) = \frac{d}{d\theta}\ell(\theta) = \frac{n\overline{y}}{\theta} - \frac{n - n\overline{y}}{1-\theta}$$

and setting $\ell'(\theta) = 0$ implies that

$$\theta = \overline{y}.$$

Since

$$\ell''(\theta) = \frac{-2n\overline{y}}{\theta^2} - \frac{2(n - n\overline{y})}{(1-\theta)^2} < 0$$

for all θ, the second derivative test implies

$$\hat{\theta}_{\text{MLE}} = \overline{Y}\,.$$

8. (e) We find

$$\log f_Y(y|\theta) = y\log\theta + (1-y)\log(1-\theta)$$

so that

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

We now find the Fisher information is

$$\begin{aligned}
I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f_Y(Y|\theta)\right) = -\mathbb{E}\left(-\frac{Y}{\theta^2} - \frac{1-Y}{(1-\theta)^2}\right) &= \frac{\mathbb{E}(Y)}{\theta^2} + \frac{1-\mathbb{E}(Y)}{(1-\theta)^2} \\
&= \frac{1}{\theta} + \frac{1-\theta}{(1-\theta)^2} \\
&= \frac{1}{\theta} + \frac{1}{1-\theta} \\
&= \frac{1}{\theta(1-\theta)}
\end{aligned}$$

since $\mathbb{E}(Y) = \theta$ as calculated in **(a)**.

8. (f) It is shown in **(b)** that $\hat{\theta}_{\text{MOM}} = \overline{Y}$ is an unbiased estimator of θ. Since

$$\frac{1}{nI(\theta)} = \frac{1}{n \cdot \frac{1}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n} = \text{Var}(\overline{Y}) = \text{Var}(\hat{\theta}_{\text{MOM}})$$

from **(c)**, we see that $\hat{\theta}_{\text{MOM}} = \overline{Y}$ attains the lower bound of the Cramer-Rao inequality. Therefore, we conclude that $\hat{\theta}_{\text{MOM}} = 2\overline{Y}$ is the minimum variance unbiased estimator of θ.

9. (a) This is incorrect because the researcher is claiming that $(1-p\text{-value})$ is the probability that the null hypothesis is false. The p-value is not a probability of a null hypothesis being true or false. It is the probability of observing a value of the sample statistic that is as or more extreme than what was observed, when the null hypothesis is true.

9. (b) With four units, the null hypothesis is unlikely to be rejected because the variability in the test statistic will be large. Hence, there is not enough data to support the researcher's claim that the alternative hypothesis is clearly not right.

10. The correct answers, in order, are: True, False, True, False, False, False.