Statistics 252–Mathematical Statistics
Winter 2005 (200510)
Final Exam Solutions

Instructor: Michael Kozdron

1. (a) To find the method of moments estimator of θ we must solve the equation $\mathbb{E}(Y) = \overline{Y}$ for θ. Since $\mathbb{E}(\overline{Y}) = \theta$, we conclude $\hat{\theta}_{\text{MOM}} = \overline{Y}$.

1. (b) By definition,

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

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

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

We find that

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

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

$$\theta = \frac{2n}{\sum_{i=1}^{n} \frac{1}{y_i}}.$$

Since

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

for all θ, the second derivative test implies

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

1. (d) Since

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

we see that if we let

$$u = \sum_{i=1}^{n} \frac{1}{y_i}, \quad g(u,\theta) = \theta^{2n} e^{-\theta u}, \quad \text{and} \quad h(y_1, \ldots, y_n) = \left(\prod_{i=1}^{n} y_i \right)^{-3},$$

then $L(\theta) = g(u,\theta) \cdot h(y_1, \ldots, y_n)$ so from the Factorization Theorem we conclude that

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

is sufficient for the estimation of θ.

1. (e) If $T(U) = \frac{2n}{U}$, then 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

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

is also sufficient.

1. (f) Since

$$\log f(y|\theta) = 2\log\theta - 3\log y - \frac{\theta}{y},$$

we find

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

Thus,

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

1. (g) An approximate 90% confidence interval for θ is given by

$$\left[\hat{\theta}_{\text{MLE}} - 1.645\,\frac{1}{\sqrt{n\,I(\hat{\theta}_{\text{MLE}})}}\,,\ \hat{\theta}_{\text{MLE}} + 1.645\,\frac{1}{\sqrt{n\,I(\hat{\theta}_{\text{MLE}})}}\right]$$

or

$$\left[\frac{2n}{\sum_{i=1}^{n}\frac{1}{Y_i}} - 1.645\frac{\sqrt{2n}}{\sum_{i=1}^{n}\frac{1}{Y_i}}\,,\ \frac{2n}{\sum_{i=1}^{n}\frac{1}{Y_i}} + 1.645\frac{\sqrt{2n}}{\sum_{i=1}^{n}\frac{1}{Y_i}}\right].$$

1. (h) The generalized likelihood ratio is

$$\begin{aligned}\Lambda = \frac{L(\theta_0)}{L(\hat{\theta}_{\text{MLE}})} = \frac{\theta_0^{2n}\left(\prod_{i=1}^{n} y_i^{-3}\right)\exp\left\{-\theta_0\sum_{i=1}^{n}\frac{1}{y_i}\right\}}{\hat{\theta}_{\text{MLE}}^{2n}\left(\prod_{i=1}^{n} y_i^{-3}\right)\exp\left\{-\hat{\theta}_{\text{MLE}}\sum_{i=1}^{n}\frac{1}{y_i}\right\}} &= \left(\frac{\theta_0}{\hat{\theta}_{\text{MLE}}}\right)^{2n}\exp\left\{\hat{\theta}_{\text{MLE}}\sum_{i=1}^{n}\frac{1}{y_i} - \theta_0\sum_{i=1}^{n}\frac{1}{y_i}\right\} \\ &= \left(\frac{\theta_0}{2n}\sum_{i=1}^{n}\frac{1}{y_i}\right)^{2n}\exp\left\{2n - \theta_0\sum_{i=1}^{n}\frac{1}{y_i}\right\} \\ &= \left(\frac{\theta_0 e}{2n}\right)^{2n}\left(\sum_{i=1}^{n}\frac{1}{y_i}\right)^{2n}\exp\left\{-\theta_0\sum_{i=1}^{n}\frac{1}{y_i}\right\}\end{aligned}$$

and so the rejection region is

$$\begin{aligned}\{\Lambda \le c\} &= \left\{\left(\frac{\theta_0 e}{2n}\right)^{2n}\left(\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{2n}\exp\left\{-\theta_0\sum_{i=1}^{n}\frac{1}{Y_i}\right\} \le c\right\} \\ &= \left\{\left(\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{2n}\exp\left\{-\theta_0\sum_{i=1}^{n}\frac{1}{Y_i}\right\} \le C\right\}\end{aligned}$$

where c and C are constants. In fact, $C = c\left(\frac{\theta_0 e}{2n}\right)^{-2n}$.

1. (i) If $n = 8$ observations produce $\sum_{i=1}^{8} \frac{1}{Y_i} = 10$, then based on this data, an approximate 90% confidence interval for θ is

$$[1.6 - 0.66, 1.6 + 0.66] \quad \text{or} \quad [0.94, 2.26].$$

Since $\theta_0 = 1$ falls in this interval, we conclude from the confidence interval–hypothesis test duality that we do not reject $H_0 : \theta = 1$ in favour of $H_A : \theta \neq 1$ at significance level $\alpha = 0.10$.

1. (j) If $\theta_0 = 1$ and $\sum_{i=1}^{8} \frac{1}{Y_i} = 10$, then the observed generalized likelihood ratio is

$$\Lambda = \left(\frac{e}{16}\right)^{16} (10)^{16} \exp\{-10\} = e^6 5^{16} 8^{-16}$$

and so

$$-2 \log \Lambda = -2(6 + 16 \log 5 - 16 \log 8) \approx 3.04011.$$

Since $-2 \log \Lambda \overset{\text{approx}}{\sim} \chi^2(1)$, we find from Table 6 that the critical value corresponding to $\alpha = 0.10$ is $\chi^2_{0.10,1} = 2.70554$. Therefore, since $3.04011 > 2.70554$, based on this data we can reject H_0 in favour of H_A at the $\alpha = 0.10$ level.

2. Let $U = Y - \theta$ so that for $-\infty < u < \infty$,

$$P(U \le u) = P(Y \le \theta + u) = \int_{-\infty}^{\theta+u} \frac{e^{(y-\theta)}}{\left[1 + e^{(y-\theta)}\right]^2} \, dy = -\frac{1}{1 + e^{(y-\theta)}} \Bigg|_{-\infty}^{\theta+u} = 1 - \frac{1}{1 + e^u}.$$

The density function of U is therefore $f_U(u) = \frac{e^u}{(1+e^u)^2}$ for $-\infty < u < \infty$. Thus, we must find a and b so that

$$\alpha_1 = P(U < a) = \int_{-\infty}^{a} \frac{e^u}{(1 + e^u)^2} \, du \quad \text{and} \quad \alpha_2 = P(U > b) = \int_{b}^{\infty} \frac{e^u}{(1 + e^u)^2} \, du.$$

Computing the integrals we find

$$\alpha_1 = 1 - \frac{1}{1 + e^a} \quad \text{and} \quad \alpha_2 = \frac{1}{1 + e^b}$$

and so solving for a and b we find

$$a = \log\left(\frac{\alpha_1}{1 - \alpha_1}\right) \quad \text{and} \quad b = \log\left(\frac{1 - \alpha_2}{\alpha_2}\right).$$

Hence,

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

In other words,

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

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

3. (a) If X_1, X_2, X_3 are i.i.d. $\text{Exp}(\lambda)$ random variables and we define $Y = \min\{X_1, X_2, X_3\}$, then for $y > 0$,

$$P(Y > y) = [P(X_1 > y)]^3 = [1 - P(X_1 \le y)]^3 = [1 - (1 - e^{-y/\lambda})]^3 = e^{-3y/\lambda}.$$

That is,

$$F_Y(y) = 1 - e^{-3y/\lambda} \quad \text{and} \quad f_Y(y) = \frac{3}{\lambda} e^{-3y/\lambda}, \ y > 0,$$

implying that $Y \sim \text{Exp}(\lambda/3)$.

3. (b) The likelihood function is $L(\lambda) = f_Y(y|\lambda) = \frac{3}{\lambda} e^{-3y/\lambda}$ (since there is $n = 1$ random variable, namely Y). In order to maximize the likelihood function, we attempt to maximize the log-likelihood function

$$\ell(\lambda) = \log 3 - \log \lambda - \frac{3y}{\lambda}.$$

Since

$$\ell'(\lambda) = -\frac{1}{\lambda} + \frac{3y}{\lambda^2}$$

so that $\ell'(\lambda) = 0$ implies $\lambda = 3y$, and since

$$\ell''(\lambda) = \frac{1}{\lambda^2} - \frac{6y}{\lambda^3} \quad \text{so that} \quad \ell''(3y) = -\frac{1}{9y^2} < 0,$$

we conclude from the second derivative test that $\hat{\lambda}_{\text{MLE}} = 3Y$.

3. (c) We begin by noting that $\text{MSE}(\hat{\lambda}_{\text{MLE}}) = \text{Var}(\hat{\lambda}_{\text{MLE}}) + [\text{bias}(\hat{\lambda}_{\text{MLE}})]^2$. Since $Y \sim \text{Exp}(\lambda/3)$, we find

$$\mathbb{E}(Y) = \frac{\lambda}{3} \quad \text{and} \quad \text{Var}(Y) = \frac{\lambda^2}{9}.$$

This implies that $\text{Var}(\hat{\lambda}_{\text{MLE}}) = \text{Var}(3Y) = 9\,\text{Var}(Y) = \lambda^2$ and $\mathbb{E}(\hat{\lambda}_{\text{MLE}}) = \mathbb{E}(3Y) = 3\mathbb{E}(Y) = \lambda$ so that $\text{bias}(\hat{\lambda}_{\text{MLE}}) = 0$. Hence, $\text{MSE}(\hat{\lambda}_{\text{MLE}}) = \text{Var}(\hat{\lambda}_{\text{MLE}}) + [\text{bias}(\hat{\lambda}_{\text{MLE}})]^2 = \lambda^2 + 0 = \lambda^2$.

3. (d) We find

$$\log f_Y(y|\lambda) = \log 3 - \log \lambda - \frac{3y}{\lambda} \quad \text{and so} \quad \frac{\partial^2}{\partial \lambda^2} \log f_Y(y|\lambda) = \frac{1}{\lambda^2} - \frac{6y}{\lambda^3}.$$

Thus,

$$I(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial \lambda^2} \log f_Y(Y|\lambda)\right) = -\mathbb{E}\left(\frac{1}{\lambda^2} - \frac{6Y}{\lambda^3}\right) = \frac{6\mathbb{E}(Y)}{\lambda^3} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

The Cramer-Rao inequality tells us that any unbiased estimator $\hat{\lambda}$ of λ must satisfy

$$\text{Var}(\hat{\lambda}) \ge \frac{1}{I(\lambda)} = \lambda^2$$

(since $n = 1$ in this problem). Since $\text{Var}(\hat{\lambda}_{\text{MLE}}) = \lambda^2 = \frac{1}{I(\lambda)}$ we have found an unbiased estimator whose variance attains the lower bound of the Cramer-Rao inequality. Hence, $\hat{\lambda}_{\text{MLE}}$ must be the MVUE of λ.

4. (a) By the confidence interval–hypothesis test duality, we do not reject $H_0 : \theta = 5$ if and only if $5 \in (X-1, X+2)$. In other words, we reject H_0 if $5 \le X-1$ or if $5 \ge X+2$. Hence, the required rejection region is

$$RR = \{X \le 3 \ \text{ or } \ X \ge 6\}.$$

4. (b) By definition, the significance level α is the probability of a Type I error; that is, the probability under H_0 that H_0 is rejected. Hence, we must find c so that

$$\frac{19}{100} = P_{H_0}(\text{reject } H_0) = P_{\theta=1}(\max\{Y_1, Y_2\} > c).$$

Since Y_1, Y_2 are independent Uniform$(0, \theta)$ random variables, we find

$$P_{\theta=1}(\max\{Y_1, Y_2\} \le c) = [P_{\theta=1}(Y_1 \le c)]^2 = \left[\int_0^c 1\, dy\right]^2 = c^2$$

and so $P_{\theta=1}(\max\{Y_1, Y_2\} > c) = 1 - c^2$. Setting

$$\frac{19}{100} = 1 - c^2 \quad \text{implies} \quad c = \frac{9}{10}.$$

By definition, the power of a hypothesis test is the probability under H_A that H_0 is rejected. Hence, we find

$$\begin{aligned}
\text{power} = P_{H_A}(\text{reject } H_0) = P_{\theta>1}\left(\max\{Y_1, Y_2\} > \frac{9}{10}\right) &= 1 - P_{\theta>1}\left(\max\{Y_1, Y_2\} \le \frac{9}{10}\right) \\
&= 1 - \left[P_{\theta>1}\left(Y_1 \le \frac{9}{10}\right)\right]^2 \\
&= 1 - \left[\int_0^{9/10} \frac{1}{\theta}\, dy\right]^2 \\
&= 1 - \frac{81}{100\,\theta^2}.
\end{aligned}$$

5. (a) The samplig distribution of this estimator is vital in order to construct confidence intervals (either exactly by the pivotal method or approximately using the MLE and Fisher information) and to conduct hypothesis tests (either exactly or using the likelihood ratio test approximation). The sampling distribution is also needed so that the accuracy (bias, mean-squared error, etc.) of the estimator can be evaluated.

5. (b) You might want additional pieces of information such as the p-value of the test, the power (which can be computed exactly since both hypotheses are simple), how the data was collected, the sampling distribution of the test statistic, how the test was conducted (likelihood ratio test, CI-HT duality, Z-test, T-test, χ^2-test, etc.), and whether or not any approximations were made.

6. (a) By definition, the significance level α is the probability of a Type I error; that is, the probability under H_0 that H_0 is rejected. Hence, since $\frac{4S^2}{\sigma^2} \sim \chi^2(4)$,

$$\alpha = P_{H_0}(\text{reject } H_0) = P(S^2 > 1.945 | \sigma^2 = 1) = P\left(\frac{4S^2}{1} > \frac{4 \cdot 1.945}{1}\right) = P(X > 7.78) \approx 0.10$$

where $X \sim \chi^2(4)$. (This last step follows from Table 6.) Hence, we see that the hypothesis test does, in fact, have significance level $\alpha = 0.10$.

6. (b) By definition, the power of an hypothesis test is the probability under H_A that H_0 is rejected. Hence, when $\sigma = 2.7$, we find

$$\begin{aligned} \text{power} = P_{H_A}(\text{reject } H_0) = P(S^2 > 1.945 | \sigma^2 = 2.7^2) &= P\left(\frac{4S^2}{2.7^2} > \frac{4 \cdot 1.945}{2.7^2}\right) \\ &\approx P(X > 1.067) \approx 0.90 \end{aligned}$$

where $X \sim \chi^2(4)$. (The last step follows from Table 6.) Hence, the power of this test when $\sigma = 2.7$ is 0.90.

7. In order to find the minimizing values of $\hat{b}_0$ and $\hat{b}_1$, we begin by computing derivatives:

$$\frac{\partial}{\partial \hat{b}_0} \text{SSE}(\hat{b}_0, \hat{b}_1) = -2 \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_i^2) \quad \text{and} \quad \frac{\partial}{\partial \hat{b}_1} \text{SSE}(\hat{b}_0, \hat{b}_1) = -2 \sum_{i=1}^{n} x_i^2 (y_i - \hat{b}_0 - \hat{b}_1 x_i^2).$$

From the first equation,

$$\frac{\partial}{\partial \hat{b}_0} \text{SSE}(\hat{b}_0, \hat{b}_1) = 0 \quad \text{implies} \quad \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_i^2) = n\bar{y} - n\hat{b}_0 - \hat{b}_1 \sum_{i=1}^{n} x_i^2 = 0$$

so that

$$\hat{b}_0 = \bar{y} - \frac{\hat{b}_1}{n} \sum_{i=1}^{n} x_i^2.$$

From the second equation,

$$\frac{\partial}{\partial \hat{b}_1} \text{SSE}(\hat{b}_0, \hat{b}_1) = 0 \quad \text{implies} \quad \sum_{i=1}^{n} x_i^2 (y_i - \hat{b}_0 - \hat{b}_1 x_i^2) = \sum_{i=1}^{n} x_i^2 y_i - \hat{b}_0 \sum_{i=1}^{n} x_i^2 - \hat{b}_1 \sum_{i=1}^{n} x_i^4 = 0.$$

Substituting in the value for $\hat{b}_0$ found above gives

$$\sum_{i=1}^{n} x_i^2 y_i - \bar{y} \sum_{i=1}^{n} x_i^2 + \frac{\hat{b}_1}{n} \left(\sum_{i=1}^{n} x_i^2 \right)^2 - \hat{b}_1 \sum_{i=1}^{n} x_i^4 = 0$$

which upon solving for $\hat{b}_1$ produces

$$\hat{b}_1 = \frac{\displaystyle\sum_{i=1}^{n} x_i^2 y_i - \bar{y} \sum_{i=1}^{n} x_i^2}{\displaystyle\sum_{i=1}^{n} x_i^4 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i^2 \right)^2}.$$