Stat 252 Winter 2007
Assignment #4 Solutions

2. Textbook

(8.30) If $\hat{\lambda} = \overline{Y}$, then $\mathbb{E}(\hat{\lambda}) = \mathbb{E}(\overline{Y}) = \lambda$ so that $\overline{Y}$ is an unbiased estimator of λ. Since the standard error of $\hat{\lambda}$ is

$$\sigma_{\hat{\lambda}} = \sqrt{\mathrm{Var}(\overline{Y})} = \sqrt{\frac{\lambda}{n}},$$

a natural guess for the estimated standard error is

$$\hat{\sigma}_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{n}}.$$

(8.36) (a) If $Z \sim \mathcal{N}(0,1)$, then using Table 4 gives

$$P(-1.96 \le Z \le 1.96) \approx 0.95.$$

That is, the normal distribution with mean 0 and variance 1 is a *parameter-free distribution*. Thus, if $Y \sim \mathcal{N}(\mu, 1)$, then

$$\frac{Y-\mu}{1} \sim \mathcal{N}(0,1).$$

Substituting for Z gives

$$P(-1.96 \le Y - \mu \le 1.96) \approx 0.95$$

so that solving for μ in the probability statement gives

$$P(Y - 1.96 \le \mu \le Y + 1.96) \approx 0.95.$$

In other words, a 95% confidence interval for μ is

$$[Y - 1.96, Y + 1.96].$$

(b) To find a 95% upper confidence limit for a normal distribution means to find z_α such that if $Z \sim \mathcal{N}(0,1)$, then

$$P(Z \le z_\alpha) = 0.95.$$

Using Table 4, we find that $z_\alpha \approx 1.645$. Similar to **(a)**, we find that

$$P(Y - \mu \le 1.645) \approx 0.95$$

so that solving for μ in the probability statement gives

$$P(\mu \ge Y - 1.645) \approx 0.95.$$

In other words, $Y - 1.645$ is a 95% *lower* confidence limit for μ. You should notice that because $Y - \mu \sim \mathcal{N}(0,1)$, the inequality switched.

(c) To find a 95% lower confidence limit for a normal distribution means to find z_α such that if $Z \sim \mathcal{N}(0,1)$, then

$$P(Z \ge z_\alpha) = 0.95.$$

Again using Table 4, we find that $z_\alpha \approx -1.645$. Similar to **(a)**, we find that

$$P(Y - \mu \geq -1.645) \approx 0.95$$

so that solving for μ in the probability statement gives

$$P(\mu \leq Y + 1.645) \approx 0.95.$$

In other words, $Y + 1.645$ is a 95% *upper* confidence limit for μ. You should notice that because $Y - \mu \sim \mathcal{N}(0, 1)$, the inequality switched.

(Remark: Technically, the answers to ***(b)*** *and* ***(c)*** *should be switched, but because I am most concerned that you intuitively understand what is going on, that is a minor concern.)*

(8.37) (a) If $X \sim \chi^2(1)$, then using Table 6 gives

$$P(0.0009821 \leq X \leq 5.02389) \approx 0.95.$$

Since the pivotal quantity Y^2/σ^2 has a $\chi^2(1)$ distribution, substituting in for X in the probability statement gives

$$P\left(0.0009821 \leq \frac{Y^2}{\sigma^2} \leq 5.02389\right) \approx 0.95$$

so that

$$P\left(\frac{Y^2}{5.02389} \leq \sigma^2 \leq \frac{Y^2}{0.0009821}\right) \approx 0.95.$$

In other words, a 95% confidence interval for σ^2 is

$$\left[\frac{Y^2}{5.02389}, \frac{Y^2}{0.0009821}\right].$$

(b) To find a 95% upper confidence limit for a chi-squared distribution with *df* $= 1$ means to find $\chi^2_{\alpha,1}$ such that if $X \sim \chi^2(1)$, then

$$P(X \leq \chi^2_{\alpha,1}) = 0.95.$$

Using Table 6 gives $\chi^2_{\alpha,1} \approx 3.84146$ so that

$$P(X \leq 3.84146) \approx 0.95.$$

Substituting in for X and solving for σ^2 in the probability statement gives

$$P\left(\sigma^2 \geq \frac{Y^2}{3.84146}\right) = 0.95.$$

In other words, $Y^2/3.84146$ is a 95% *lower* confidence limit for σ^2. You should notice that because $Y^2/\sigma^2 \sim \chi^2(1)$, the inequality switched.

(c) To find a 95% lower confidence limit for a chi-squared distribution with *df* $= 1$ means to find $\chi^2_{\alpha,1}$ such that if $X \sim \chi^2(1)$, then

$$P(X \geq \chi^2_{\alpha,1}) = 0.95.$$

Using Table 6 gives $\chi^2_{\alpha,1} \approx 0.0039321$ so that

$$P(X \geq 0.0039321) \approx 0.95.$$

Substituting in for X and solving for σ^2 in the probability statement gives

$$P\left(\sigma^2 \leq \frac{Y^2}{0.0039321}\right) \approx 0.95.$$

In other words, $Y^2/0.0039321$ is a 95% *upper* confidence limit for σ^2. You should notice that because $Y^2/\sigma^2 \sim \chi^2(1)$, the inequality switched.

*(Remark: Technically, the answers to **(b)** and **(c)** should be switched, but because I am most concerned that you intuitively understand what is going on, that is a minor concern.)*

(8.38) (a) From **(8.37)**, we find that

$$P\left(\frac{Y^2}{5.02389} \leq \sigma^2 \leq \frac{Y^2}{0.0009821}\right) \approx 0.95.$$

Since the square root function is monotonic, we can conclude

$$P\left(\sqrt{\frac{Y^2}{5.02389}} \leq \sqrt{\sigma^2} \leq \sqrt{\frac{Y^2}{0.0009821}}\right) \approx 0.95.$$

Since Y is a *non-negative* random variable, and since $\sigma > 0$, we conclude

$$P\left(\frac{Y}{\sqrt{5.02389}} \leq \sigma \leq \frac{Y}{\sqrt{0.0009821}}\right) \approx 0.95,$$

or, in other words, a 95% confidence interval for σ is

$$\left[\frac{Y}{\sqrt{5.02389}}, \frac{Y}{\sqrt{0.0009821}}\right].$$

(b) Similarly, $Y/\sqrt{0.0039321}$ is a 95% upper confidence limit for σ.

(c) Similarly, $Y/\sqrt{3.84146}$ is a 95% lower confidence limit for σ.

(8.25) (a) Let p_1 denote the proportion of Americans who ate the recommended amount of fibrous foods in 1983, and let p_2 denote the proportion who ate the recommended amount in 1992. The data then yield $\hat{p}_1 = 0.59$ and $\hat{p}_2 = 0.53$. The estimated standard errors are easily calculated as:

$$\hat{\sigma}_{\hat{p}_1} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} = \sqrt{\frac{0.59 \cdot 0.41}{1250}} \quad \text{and} \quad \hat{\sigma}_{\hat{p}_2} = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n}} = \sqrt{\frac{0.53 \cdot 0.47}{1251}}.$$

Thus, a point estimate for the difference is given by $\hat{p}_1 - \hat{p}_2 = 0.59 - 0.53 = 0.06$. This indicates that there was a 6% *decrease* in the proportion of Americans who were eating the recommended amount of fibrous foods in 1993 compared with 1982. A bound on the error of estimation is

$$2\sqrt{\hat{\sigma}^2_{\hat{p}_1} + \hat{\sigma}^2_{\hat{p}_2}} \approx 0.04.$$

(b) Note that the answer in **(a)** yields an approximate 95% confidence interval of

$$0.06 \pm 0.04 = [0.02, 0.10].$$

Since this interval *does not cover* 0, there is statistically significant evidence to indicate that there has been a demonstrable decrease in the proportion of Americans who ate the recommended amount of fibrous foods in 1993 compared with 1982.

(8.50) (a) This problem is similar to **(8.25)**. Using the answers to **(8.25)**, and the fact that $t_{0.01,1250} = 2.3293$, we conclude that an approximate 98% confidence interval for the difference is

$$\hat{p}_1 - \hat{p}_2 \pm 2.3293\sqrt{\hat{\sigma}^2_{\hat{p}_1} + \hat{\sigma}^2_{\hat{p}_2}} \quad \text{or} \quad 0.06 \pm 0.046.$$

(b) As before, since this interval *does not cover* 0, there is statistically significant evidence to indicate that there has been a demonstrable decrease in the proportion of Americans who ate the recommended amount of fibrous foods in 1993 compared with 1982.

3. Textbook

(8.10) (a) If $\hat{\theta} = \max\{Y_1, \ldots, Y_n\}$, then its distribution function is

$$F(t) = \theta^{-n\alpha}\, t^{n\alpha}, \quad 0 \le t \le \theta$$

so that

$$f(t) = n\,\alpha\,\theta^{-n\alpha}\, t^{n\alpha - 1}, \quad 0 \le t \le \theta.$$

We easily calculate that

$$\mathbb{E}(\hat{\theta}) = \int_0^\theta n\,\alpha\,\theta^{-n\alpha}\, t^{n\alpha}\, dt = \frac{n\,\alpha\,\theta^{-n\alpha}\,\theta^{n\alpha+1}}{n\alpha + 1} = \frac{n\,\alpha}{n\,\alpha + 1}\,\theta.$$

Thus, we conclude $\hat{\theta}$ is a biased estimator of θ.

(b) Clearly, the estimator

$$\frac{n\,\alpha + 1}{n\,\alpha}\,\hat{\theta} = \frac{n\,\alpha + 1}{n\,\alpha}\,\max\{Y_1, \ldots, Y_n\}$$

is an unbiased estimator of θ.

(c) In order to calculate $\text{MSE}(\hat{\theta})$ we must find $\text{Var}(\hat{\theta})$. We find

$$\mathbb{E}(\hat{\theta}^2) = \int_0^\theta n\,\alpha\,\theta^{-n\alpha}\, t^{n\alpha+1}\, dt = \frac{n\,\alpha\,\theta^{-n\alpha}\,\theta^{n\alpha+2}}{n\alpha + 2} = \frac{n\,\alpha}{n\,\alpha + 2}\,\theta^2.$$

Thus,

$$\text{Var}(\hat{\theta}) = \mathbb{E}(\hat{\theta}^2) - [\mathbb{E}(\hat{\theta})]^2 = \frac{n\,\alpha}{n\,\alpha + 2}\,\theta^2 - \left[\frac{n\,\alpha}{n\,\alpha + 1}\,\theta\right]^2 = \frac{n\alpha}{(n\alpha + 1)^2(n\alpha + 2)}\,\theta^2.$$

Finally,

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [B(\hat{\theta})]^2 = \left[\frac{n\alpha}{(n\alpha + 1)^2(n\alpha + 2)} + \frac{1}{(n\alpha + 1)^2}\right]\theta^2 = \frac{2\theta^2}{(n\alpha + 1)(n\alpha + 2)}.$$

(8.15) If $Y_{(1)} = \min\{Y_1, \ldots, Y_n\}$, then its distribution function is

$$F(t) = 1 - e^{-tn/\theta}, \quad t > 0$$

so that

$$f(t) = \frac{n}{\theta} e^{-tn/\theta}, \quad t > 0.$$

We easily calculate (use integration by parts) that

$$\mathbb{E}(Y_{(1)}) = \int_0^\infty \frac{n}{\theta} t e^{-tn/\theta} dt = \frac{\theta}{n}$$

so that if $\hat{\theta} = nY_{(1)}$, then $\mathbb{E}(\hat{\theta}) = n\mathbb{E}(Y_{(1)}) = \theta$ so that $\hat{\theta}$ is an unbiased estimator of θ.

In order to calculate $\text{MSE}(\hat{\theta})$ we must find $\text{Var}(\hat{\theta})$. Notice, however, that $Y_{(1)}$ is an exponential random variable with parameter θ/n. Thus,

$$\text{Var}(\hat{\theta}) = \text{Var}(nY_{(1)}) = n^2 \,\text{Var}(Y_{(1)}) = n^2 \frac{\theta^2}{n^2} = \theta^2.$$

This gives

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [B(\hat{\theta})]^2 = \theta^2 + 0 = \theta^2.$$

(8.32) If $\hat{\theta} = \overline{Y}$, then $\mathbb{E}(\hat{\theta}) = \mathbb{E}(\overline{Y}) = \theta$ so that $\overline{Y}$ is an unbiased estimator of θ. The standard error of $\hat{\theta}$ is

$$\sigma_{\hat{\theta}} = \sqrt{\text{Var}(\overline{Y})} = \frac{\theta}{\sqrt{n}}.$$

Thus, if the estimated standard error is

$$\hat{\sigma}_{\hat{\theta}} = \frac{\hat{\theta}}{\sqrt{n}},$$

then

$$\mathbb{E}(\hat{\sigma}_{\hat{\theta}}) = \frac{\mathbb{E}(\hat{\theta})}{\sqrt{n}} = \frac{\theta}{\sqrt{n}} = \sigma_{\hat{\theta}}$$

so that $\hat{\sigma}_{\hat{\theta}}$ is an unbiased estimator of the standard error.

(8.43) From the data presented, we find that $n = 2374$ adults in the continental US were interviewed, of which 1912 were registered voters. Thus, if p denotes the true proportion of registered voters in the continental US, then from this we conclude

$$\hat{p} = \frac{1912}{2374}.$$

Thus, an approximate 99% confidence interval for p is given by

$$\hat{p} \pm t_{0.005,2373} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad \frac{1912}{2374} \pm 2.5779 \cdot \sqrt{\frac{1912/2374 \cdot 462/2374}{2374}}\ .$$

In other words, at the 99% confidence level, the proportion of adults in the continental US registered to vote is between 0.7844 and 0.8263.

4. We begin by finding the distribution of $\hat{\theta}$, namely

$$P(\hat{\theta} \leq t) = P(Y_1 \leq t) \cdots P(Y_n \leq t) = \frac{t^n}{\theta^n}, \quad 0 \leq t \leq \theta.$$

It therefore follows that the density of $\hat{\theta}$ is

$$f_{\hat{\theta}}(t) = \frac{nt^{n-1}}{\theta^n}, \quad 0 \leq t \leq \theta.$$

We must now find a pivotal quantity. Let

$$U = \frac{\hat{\theta}}{\theta}$$

so that

$$P(U \leq u) = P(\hat{\theta} \leq \theta u) = u^n, \quad 0 \leq u \leq 1$$

from which we conclude

$$f_U(u) = nu^{n-1}, \quad 0 \leq u \leq 1.$$

We now find a and b such that

$$\int_a^b f_U(u)\, du = 0.90, \quad \int_0^a f_U(u)\, du = \int_b^1 f_U(u)\, du = 0.05.$$

That is,

$$\int_0^a nu^{n-1}\, du = 0.05$$

implies $a^n = 0.05$ and so $a = \sqrt[n]{0.05}$. Furthermore,

$$\int_b^1 nu^{n-1}\, du = 0.05$$

implies $1 - b^n = 0.05$ and so $b = \sqrt[n]{0.95}$. We then conclude that

$$P\left(\sqrt[n]{0.05} \leq U \leq \sqrt[n]{0.95} \right) = 0.90$$

and so substituting $U = \hat{\theta}/\theta$ gives

$$P\left(\frac{\hat{\theta}}{\sqrt[n]{0.05}} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.95}} \right) = 0.90.$$

The required 90% confidence interval for θ is therefore

$$\left[\frac{\hat{\theta}}{\sqrt[n]{0.05}}, \frac{\hat{\theta}}{\sqrt[n]{0.95}} \right].$$

5. Let $U = Y/\theta$ so that for $0 \leq u \leq 1$,

$$P(U \leq u) = P(Y \leq \theta u) = \int_0^{\theta u} 2\theta^{-2} y\, dy = \frac{y^2}{\theta^2}\bigg|_0^{\theta u} = u^2.$$

The density function of U is therefore $f_U(u) = 2u$ for $0 \le u \le 1$. Thus, we must find a and b so that

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

Computing the integrals we find $a^2 = \alpha/2$ and $1 - b^2 = \alpha/2$. Hence,

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

In other words,

$$\left[\frac{Y}{\sqrt{1-\alpha/2}}, \frac{Y}{\sqrt{\alpha/2}}\right]$$

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

6. 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 $1-(\alpha_1+\alpha_2)$.

7. 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 $1-(\alpha_1+\alpha_2)$.