Stat 252.01 Winter 2005
Assignment #2 Solutions

1. (a) Since $Y_1 \sim \mathcal{U}(0,\theta)$, we have

$$f_Y(y) = \begin{cases} 1/\theta, & 0 \le y \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

(b) If $f(t)$ denotes the density of $\hat{\theta} = \min(Y_1, \ldots, Y_n)$, then $f(t) = F'(t)$, where

$$F(t) = P(\hat{\theta} \le t) = 1 - P(\hat{\theta} > t) = 1 - P(Y_1 > t, \ldots, Y_n > t) = 1 - [P(Y_1 > t)]^n.$$

Note that in the last step we have used the fact that Y_i are iid. Next, for $0 \le t \le \theta$, we compute

$$P(Y_1 > t) = \int_t^\infty f_Y(y)\, dy = \int_t^\theta \frac{1}{\theta}\, dy = \frac{\theta - t}{\theta}.$$

Thus, we conclude

$$f(t) = \frac{d}{dt}\left(1 - \left[\frac{\theta - t}{\theta}\right]^n\right) = n\theta^{-n}(\theta - t)^{n-1}, \quad 0 \le t \le \theta.$$

(c) By definition,

$$\mathbb{E}(\hat{\theta}) = \int_{-\infty}^{\infty} t f(t)\, dt = \int_0^\theta n\theta^{-n} t(\theta - t)^{n-1}\, dt.$$

This last integral is solved with a simple substitution. Let $u = \theta - t$ so that $du = -dt$. Thus,

$$\begin{aligned} \int_0^\theta n\theta^{-n} t(\theta - t)^{n-1}\, dt = -n\theta^{-n} \int_\theta^0 (\theta - u)u^{n-1}\, du &= n\theta^{-n} \int_0^\theta \theta u^{n-1} - u^n\, du \\ &= n\theta^{-n}\left(\theta \cdot \frac{\theta^n}{n} - \frac{\theta^{n+1}}{n+1}\right) \\ &= \frac{\theta}{n+1}. \end{aligned}$$

(d) From **(c)**, we clearly see that $\hat{\theta}$ is NOT an unbiased estimator of θ. However,

$$\tilde{\theta} = (n+1)\min(Y_1, \ldots, Y_n)$$

IS an unbiased estimator of θ. (You should check that $2\overline{Y}$ is also an unbiased estimator of θ. Why?)

(8.13) (a) If $Y \sim Bin(n,p)$, then $\mathbb{E}(Y) = np$ and $\text{Var}(Y) = np(1-p)$. Thus,

$$\mathbb{E}(\hat{p}_2) = \frac{\mathbb{E}(Y) + 1}{n+2} = \frac{np+1}{n+2}$$

so that

$$B(\hat{p}_2) = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}.$$

(b) To compute $\text{MSE}(\hat{p}_i)$ we use the computational formula:

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

Thus, we find that since $\hat{p}_1$ is unbiased,

$$\mathrm{MSE}(\hat{p}_1) = \mathrm{Var}(\hat{p}_1) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

As for $\hat{p}_2$, we find

$$\mathrm{Var}(\hat{p}_2) = \frac{\mathrm{Var}(Y+1)}{(n+2)^2} = \frac{\mathrm{Var}(Y)}{(n+2)^2} = \frac{np(1-p)}{(n+2)^2},$$

so that

$$\mathrm{MSE}(\hat{p}_2) = \mathrm{Var}(\hat{p}_2) + [B(\hat{p}_2)]^2 = \frac{np(1-p)}{(n+2)^2} + \left(\frac{1-2p}{n+2}\right)^2 = \frac{(4-n)p^2 - (4-n)p + 1}{(n+2)^2}.$$

(c) The problem in the text should ask, "For what values of p is $\mathrm{MSE}(\hat{p}_2) < \mathrm{MSE}(\hat{p}_1)$?" In order for $\mathrm{MSE}(\hat{p}_2) < \mathrm{MSE}(\hat{p}_1)$, we require

$$\frac{p(1-p)}{n} > \frac{(4-n)p^2 - (4-n)p + 1}{(n+2)^2}.$$

Simplifying gives

$$p^2 - p + \frac{n}{8n+4} < 0.$$

This quadratic inequality in p can be easily solved to give

$$\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{n}{8n+4}} < p < \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{n}{8n+4}}.$$

Note that it is insufficient to simply copy the answer in the back of the book. "p near 1/2" is meaningless unless it is accompanied by the explicit solution above. While it is true that asymptotically p is "near" 1/2, it is necessary to *prove* exactly what "near" means in this case.

(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

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

a natural guess for the estimated standard error is

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

(8.32) 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

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

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.

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

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

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

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

so that solving for μ in the probability statement gives

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

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 $z_\alpha = 1.645$. Similar to **(a)**, we find that

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

so that solving for μ in the probability statement gives

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

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 $z_\alpha = -1.645$. Similar to **(a)**, we find that

$$P(Y - \mu \ge -1.645) = 0.95$$

so that solving for μ in the probability statement gives

$$P(\mu \le Y + 1.645) = 0.95.$$

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 $Z \sim \chi_1^2$, then using Table 6 gives

$$P(0.0009821 \leq Z \leq 5.02389) = 0.95.$$

Since the pivotal quantity Y^2/σ^2 has a χ_1^2 distribution, substituting in for Z in the probability statement gives

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

so that

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

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 χ_α such that if $Z \sim \chi_1^2$, then

$$P(Z \leq \chi_\alpha) = 0.95.$$

Using Table 6 gives $\chi_\alpha = 3.84146$ so that

$$P(Z \leq 3.84146) = 0.95.$$

Substituting in for Z 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 $Y^2/\sigma^2 \sim \chi_1^2$, the inequality switched.

(c) To find a 95% lower confidence limit for a chi-squared distribution with $df = 1$ means to find χ_α such that if $Z \sim \chi_1^2$, then

$$P(Z \geq \chi_\alpha) = 0.95.$$

Using Table 6 gives $\chi_\alpha = 0.0039321$ so that

$$P(Z \geq 0.0039321) = 0.95.$$

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

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

In other words, $Y^2/0.0039321$ is a 95% *upper* confidence limit for σ^2. You should notice that because $Y^2/\sigma^2 \sim \chi_1^2$, 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

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

Since the square root function is monotonic, we can conclude

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

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

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

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.

(8.43) 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

$$\hat{p} \pm z_a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad \frac{1912}{2374} \pm 2.575 \cdot \sqrt{\frac{1912/2374 \cdot 462/2374}{2374}}.$$

In other words, at the 99% confidence level, the proportion of adults in the continental US registered to vote is between 0.784 and 0.826.

(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 the critical value for a 98% normal confidence interval is 2.33, we conclude that an approximate 98% confidence interval for the difference is

$$\hat{p}_1 - \hat{p}_2 \pm 2.33\sqrt{\hat{\sigma}^2_{\hat{p}_1} + \hat{\sigma}^2_{\hat{p}_2}} \quad \text{or} \quad 0.06 \pm 0.046.$$

(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.