Problem Set 9

8.2 **a.** $E\left[\hat{\theta}_3\right] = E\left[a\hat{\theta}_1 + (1-a)\hat{\theta}_2\right] = aE\left[\hat{\theta}_1\right] + (1-a)E\left[\hat{\theta}_2\right] = a\theta + (1-a)\theta = \theta$

b. It is given that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ and $V(\hat{\theta}_1) = \sigma_1^2$, $V(\hat{\theta}_2) = \sigma_2^2$. Assuming that $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, the variance of the new estimator, $\hat{\theta}_3$, will be

$$V\left(\hat{\theta}_3\right) = V\left[a\hat{\theta}_1 + (1-a)\hat{\theta}_2\right] = a^2V\left(\hat{\theta}_1\right) + (1-a)^2V\left(\hat{\theta}_2\right) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$$

In order to choose a value of a such that $V(\hat{\theta}_3)$ is minimized, look at

$$\tfrac{d}{da}V\left(\hat{\theta}_3\right) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2.$$

Setting the derivative equal to 0, we obtain

$$a\sigma_1^2 - (1-a)\sigma_2^2 = 0$$

or

$$a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

Notice that $\frac{d^2}{da^2}V\left(\hat{\theta}_3\right) = 2\sigma_1^2 + 2\sigma_2^2 > 0$, so that the value is in fact a minimum.

8.4 Recall that if Y_i is Exponential(θ) then $E(Y_i) = \theta$ and $V(Y_i) = \theta^2$. Hence we can use Theorem 5.12 to obtain

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = E(\hat{\theta}_3) = E(\hat{\theta}_5) = \theta$$
$$V(\hat{\theta}_1) = \theta^2$$
$$V(\hat{\theta}_2) = \tfrac{1}{4}\left(2\theta^2\right) = \tfrac{\theta^2}{2}$$
$$V(\hat{\theta}_3) = \tfrac{1}{9}\left(\theta^2 + 4\theta^2\right) = \tfrac{5\theta^2}{9}$$
$$V(\hat{\theta}_5) = \tfrac{1}{9}\left(3\theta^2\right) = \tfrac{\theta^2}{3}$$

The distribution of $\hat{\theta}_4$ can be obtained by using the methods of Section 6.6 in the text, with $F(y) = 1 - e^{-y/\theta}$. Then

$$g_1(y) = \tfrac{3}{\theta}e^{-y/\theta}\left(e^{-y/\theta}\right)^2 = \tfrac{3}{\theta}e^{-3y/\theta}$$

which is an exponential distribution with mean $\frac{\theta}{3}$.

$$E(\hat{\theta}_4) = \tfrac{\theta}{3} \qquad V(\hat{\theta}_4) = \tfrac{\theta^2}{9}$$

a. The unbiased estimators are $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_5$.

b. Among these four estimators, $\hat{\theta}_5 = \overline{Y}$ has the smallest variance.

8.5 Since Y has an exponential distribution with mean $\theta + 1$, $E(Y) = \theta + 1$ and $E(\overline{Y}) = \theta + 1$. Hence if we use $\hat{\theta} = \overline{Y} - 1$, $E(\hat{\theta}) = \theta$ and we have constructed an unbiased estimator.

8.8 **a.** For the uniform distribution given here, $E(Y_i) = \theta + \frac{1}{2}$. Hence $E(\overline{Y}) = \theta + \frac{1}{2}$ and the bias is $B = E(\overline{Y}) - \theta = \frac{1}{2}$.

b. An unbiased estimator of θ can be constructed by using $\hat{\theta} = \overline{Y} - \frac{1}{2}$, which has

$$E(\hat{\theta}) = \theta.$$

c. If $\overline{Y}$ is used as an estimator, then

$$V(\overline{Y}) = \frac{V(Y)}{n} = \frac{1}{12n} \qquad \text{and} \qquad \text{MSE} = V(\overline{Y}) + B^2 = \frac{1}{12n} + \frac{1}{4}.$$

8.9 **a.** For a binomial random variable Y, $E(Y) = np$ and $E(Y^2) = V(Y) + n^2p^2$ $= npq + n^2p^2$. Hence

$$E\left\{n\left(\frac{Y}{n}\right)\left[1-\left(\frac{Y}{n}\right)\right]\right\} = E(Y) - \frac{1}{n}E(Y^2) = np - pq - np^2 = np(1-p) - pq$$
$$= (n-1)pq$$

b. An unbiased estimator $\hat{\theta}$ has expected value npq. Hence we can use

$$\left(\frac{n}{n-1}\right) n \times \frac{Y}{n}\left(1-\frac{Y}{n}\right) = \frac{n^2}{n-1}\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$$

8.10 The following information is required to answer the question.

$$E(Y) = \int_0^\theta \left[\frac{\alpha y^\alpha}{\theta^\alpha}\right] dy = \left[\frac{\alpha y^{\alpha+1}}{(\alpha+1)\theta^\alpha}\right]_0^\theta = \frac{\alpha\theta}{\alpha+1}$$
$$E(Y^2) = \int_0^\theta \left[\frac{\alpha y^{\alpha+1}}{\theta^\alpha}\right] dy = \left[\frac{\alpha y^{\alpha+2}}{(\alpha+2)\theta^\alpha}\right]_0^\theta = \frac{\alpha\theta^2}{\alpha+2}$$
$$f(y) = \frac{\alpha y^{\alpha-1}}{\theta^\alpha}$$
$$F(y) = \int_0^y \frac{\alpha t^{\alpha-1}}{\theta^\alpha} dt = \left(\frac{y}{\theta}\right)^\alpha$$
$$F_{Y(n)}(y) = \left(\frac{y}{\theta}\right)^{n\alpha}, 0 \le y \le \theta$$
$$f_{Y(n)}(y) = \frac{n\alpha y^{n\alpha-1}}{\theta^{n\alpha}} \quad 0 \le y \le \theta$$

So that $Y_{(n)}$ is also distributed as the power family with parameters $n\alpha$ and θ.

a. $E(Y_{(n)}) = \frac{n\alpha\theta}{n\alpha+1} \neq \theta$.

b. $\left(\frac{n\alpha+1}{n\alpha}\right) Y_{(n)}$ would be unbiased.

c. $\text{MSE}(Y_{(n)}) = E[(Y_{(n)} - \theta)^2] = E\left(Y_{(n)}^2\right) - 2\theta E(Y_{(n)}) + \theta^2$

$$= \frac{n\alpha\theta^2}{n\alpha+2} - 2\theta\left(\frac{n\alpha\theta}{n\alpha+1}\right) + \theta^2$$
$$= \frac{2\theta^2}{(n\alpha+1)(n\alpha+2)}.$$

8.13 Note that

$$E(\hat{p}_1) = E\left(\frac{Y}{n}\right) = \left(\frac{1}{n}\right)(np) = p$$
$$E(\hat{p}_2) = E\left(\frac{Y+1}{n+2}\right) = \frac{1}{(n+2)}(np+1) = \frac{np+1}{n+2}$$

a. $\text{Bias} = \frac{np+1}{n+2} - p = \frac{np+1-np-2p}{n+2} = \frac{1-2p}{n+2}$.

b. $\text{MSE}(\hat{p}_1) = V(\hat{p}_1) + B^2 = V\left(\frac{Y}{n}\right) + 0 = \left(\frac{1}{n^2}\right) np(1-p) = \frac{p(1-p)}{n}$.

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

Solution for 8.13 (c).

To get $MSE(\hat{P}_1) < MSE(\hat{P}_2)$, we need.

$$\frac{p-p^2}{n} < \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$$

That is

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

It can be rewritten as:

$$(n^2+4n+4)(p-p^2) < n\left[(4-n)p^2+(n-4)p+1\right]$$

which can be simplified to

$$(8n+4)p^2-(8n+4)p+n > 0$$

By the quadratic formula, the roots for

$$(8n+4)p^2-(8n+4)p+n=0$$

are

$$p = \frac{8n+4 \pm \sqrt{(8n+4)^2-4n(8n+4)}}{2(8n+4)} = \frac{1}{2} \pm \sqrt{\frac{n+1}{8n+4}}$$

so for $0 < p < \frac{1}{2} - \sqrt{\frac{n+1}{8n+4}}$ or $\frac{1}{2} + \sqrt{\frac{n+1}{8n+4}} < p < 1$

we have

$$MSE(\hat{P}_1) < MSE(\hat{P}_2)$$

That is, for p closes to 0 or 1, we will ge $MSE(\hat{P}_1) < MSE(\hat{P}_2)$.

8.18 The point estimate of μ is $\bar{y} = 7.2\%$, and the bound on the error of estimation is $2\sigma_{\bar{y}}$. With $n = 200$ and $s = 5.6\%$, we have

$$2\sigma_{\bar{y}} = 2\,\frac{\sigma}{\sqrt{n}} \approx 2\,\frac{s}{\sqrt{n}} = \frac{2(5.6)}{\sqrt{200}} = .79$$

8.19 **a.** The point estimate of μ is $\bar{y} = 11.3$, and the bound on the error of estimation is $2\sigma_{\bar{y}}$. With $n = 467$ and $s = 16.6$, this bound is

$$2\sigma_{\bar{y}} = 2\,\frac{\sigma}{\sqrt{n}} = 2\,\frac{s}{\sqrt{n}} = \frac{2(16.6)}{\sqrt{467}} = 1.54$$

b. The point estimate of $\mu_R - \mu_C$ is $\bar{y}_R - \bar{y}_C = 46.4 - 45.1 = 1.3$. The bound on the error of estimation is

$$2\sqrt{\frac{\sigma_R^2}{n_R} + \frac{\sigma_C^2}{n_C}} = 2\sqrt{\frac{s_R^2}{n_R} + \frac{s_C^2}{n_C}} = 2\sqrt{\frac{(9.8)^2}{191} + \frac{(10.2)^2}{467}} = 1.7$$

c. The point estimate of $p_C - p_R$ is $\hat{p}_C - \hat{p}_R = .78 - .61 = .17$. The bound on the error of estimation is

$$2\sqrt{\frac{\hat{p}_C\hat{q}_C}{n_C} + \frac{\hat{p}_R\hat{q}_R}{n_R}} = 2\sqrt{\frac{(.78)(.23)}{467} + \frac{(.61)(.39)}{191}} = .08$$

8.23 **a.** The bound is $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2\sqrt{\frac{(.67)(.33)}{308{,}007}} = .0017$.

b. The bound is $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2\sqrt{\frac{(.71)(.29)}{308{,}007}} = .0016$.

c. No, ± 2 percentage points is too large for the margin of error. The bound on the margin of error is closer to $\pm .2$ percentage points.

8.25 **a.** Let p_1 = proportion of Americans who ate the recommended amount of fibrous foods in 1983 and p_2 = proportion of Americans who ate the recommended amount of fibrous foods in 1992. Then $n_1 = 1250$, $n_2 = 1251$, $\hat{p}_1 = .59$, and $\hat{p}_2 = .53$. The point estimator for the difference in proportions is

$$\hat{p}_1 - \hat{p}_2 = .59 - .53 = .06.$$

The bound on the error of estimation is

$$2\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 2\sqrt{\frac{(.59)(.41)}{1250} + \frac{(.53)(.47)}{1251}} = .04$$

b. Since $.06 - .04 > 0$, we can conclude that there has been a demonstrable decrease in the proportion of Americans who eat the recommended amount of fibrous foods.

8.28 The point estimate of the total accounts receivable is $500\overline{y} = 500(197.1) = 98{,}550$. To find a bound on the error of estimation, we need to find s^2:

$$s^2 = \frac{\sum_{1}^{n} y_i^2 - n(\overline{y})^2}{n-1} = \frac{933{,}814 - 20(197.1)^2}{19} = 8255.04$$

The variance of $500(\overline{y})$ is $500^2\,\sigma_{\overline{y}}^2 = \frac{500^2\sigma^2}{20}$, which we estimate as $\frac{500^2 s^2}{20}$. A bound on the error of estimation is

$$2\sqrt{\frac{500^2 s^2}{20}} = 20{,}316.3$$

The point estimate of the average accounts receivable, μ, is point $\overline{y} = 197.1$. A bound on the error of estimation is

$$2\left(\frac{\sigma}{\sqrt{n}}\right)$$

which may be estimated by

$$2\left(\frac{s}{\sqrt{n}}\right) = \frac{2(90.857)}{\sqrt{20}} = 40.63.$$

The value 250 is beyond the point estimate plus the bound on the error of estimation. Thus, it is unlikely that the average account receivable exceeds \$250.

8.29 The point estimate is $\hat{p} = .3$. A bound on the error of estimation is

$$2\sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(.3)(.7)}{20}} = .205$$

It is fairly likely that the proportion in compliance exceeds .80 since the value .20 is well within the margin of error from the estimation of the proportion not in compliance.