Stat 252.01 Winter 2005
Assignment #1 Solutions

Printed Lecture Notes

(4.3) It is a simple matter to compute:

- $\mathbb{E}(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p + 0 \cdot (1 - p) = p$;
- $\mathbb{E}(X^2) = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$;
- $\mathbb{E}(e^{-\theta X}) = e^{-\theta \cdot 1} P(X = 1) + e^{-\theta \cdot 0} P(X = 0) = e^{-\theta} \cdot p + 1 \cdot (1 - p) = 1 - p(1 - e^{-\theta})$.

(4.4) In order to solve this problem, we will need to compute several integrals. Since the density function for any random variable integrates to 1, we have

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1.$$

After substituting $u = y^2/2$, and carefully handling the infinite limits of integrations, we find

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy = 0.$$

Finally, using parts with $u = y$, $dv = ye^{-y^2/2} dy$, and carefully handling the infinite limits of integration,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = 1.$$

In fact, it is also straightforward to show that for $n = 1, 2, 3, 4, 5, 6, \ldots$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^n e^{-y^2/2} dy = (n-1) \cdot (n-3) \cdot (n-5) \cdots 3 \cdot 1 \cdot \left(\frac{1 + (-1)^n}{2} \right).$$

As for the expected moments, we apply the *Law of the Unconscious Statistician.*

- By definition,

  $$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

  Substituting $y = \frac{x-\mu}{\sigma}$ so that $x = \sigma y + \mu$, $\sigma dy = dx$ transforms the integral into

  $$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-y^2/2} dy &= \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \sigma \cdot 0 + \mu \cdot 1 = \mu \end{aligned}$$

  using the integrals above.

- By definition,

  $$\mathbb{E}(X^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Substituting $y = \frac{x-\mu}{\sigma}$ so that $x = \sigma y + \mu$, $\sigma dy = dx$ transforms the integral into

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu)^2 e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma^2 y^2 + 2\sigma\mu y + \mu^2) e^{-y^2/2} dy.$$

As in the previous part, splitting up the integral into the three separate pieces, and using the integrals computed above, we find

$$\mathbb{E}(X^2) = \sigma^2 \cdot 1 + 2\sigma\mu \cdot 0 + \mu^2 \cdot 1 = \sigma^2 + \mu^2.$$

- By definition,

$$\mathbb{E}(e^{-\theta X}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\theta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

The first step is to combine and simplify the integrand, namely

$$\begin{aligned} e^{-\theta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} &= \exp\left(-\theta x - \frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{\theta^2\sigma^4 - 2\mu\theta\sigma^2 - (x+\theta\sigma^2-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{\theta\sigma^2}{2} - \mu\theta\right) \exp\left(\frac{-(x+\theta\sigma^2-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

where the last equality was obtained by *completing the square*. Substituting this back into the original integral gives

$$\mathbb{E}(e^{-\theta X}) = \exp\left(\frac{\theta\sigma^2}{2} - \mu\theta\right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x+\theta\sigma^2-\mu)^2}{2\sigma^2}\right) dx.$$

To compute this final integral we make the substitution $y = \frac{x+\theta\sigma^2-\mu}{\sigma}$ so that $\sigma dy = dx$. This gives

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x+\theta\sigma^2-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1,$$

so that

$$\mathbb{E}(e^{-\theta X}) = \exp\left(\frac{\theta\sigma^2}{2} - \mu\theta\right).$$

(5.8)

- By definition, $\mathrm{Cov}(XY) = E((X-\mu_X)(Y-\mu_Y))$. Squaring out the terms and using the linearity of expectation gives $\mathbb{E}(XY - Y\mu_X - X\mu_Y + \mu_X\mu_Y) = \mathbb{E}(XY) - \mathbb{E}(Y\mu_X) - \mathbb{E}(X\mu_Y) + \mathbb{E}(\mu_X\mu_Y)$. Pulling out the constants gives $\mathbb{E}(XY) - \mu_X\mathbb{E}(Y) - \mu_Y\mathbb{E}(X) + \mu_X\mu_Y$ which equals $\mathbb{E}(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ since $\mathbb{E}(X) = \mu_X$ and $\mathbb{E}(Y) = \mu_Y$.

- By definition $\mathrm{Var}(X) = \mathbb{E}((X-\mu_X)^2)$ and $\mathrm{Cov}(X,X) = \mathbb{E}((X-\mu_X)(X-\mu_X))$. Hence, we see that they are both equal to σ^2.

- Using the first identity we find $\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \mathbb{E}(XX) - \mathbb{E}(X)\mathbb{E}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

(5.9) In Exercise 4.4, we computed $\mathbb{E}(X) = \mu$ and $\mathbb{E}(X^2) = \sigma^2 + \mu^2$. Using the computational formula, we find

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \sigma^2 + \mu^2 - (\mu)^2 = \sigma^2.$$

(5.12) If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Hence,

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

so that X and Y are uncorrelated.

(5.13)

- We find the density of Y simply using the *Law of Total Probability*:

  $$\begin{aligned}
  &P(Y = 0) \\
  &\quad = P(Y = 0|X = 1)P(X = 1) + P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = -1)P(X = -1) \\
  &\quad = 0 \cdot 1/4 + 1 \cdot 1/2 + 0 \cdot 1/4 \\
  &\quad = 1/2,
  \end{aligned}$$

  $$\begin{aligned}
  &P(Y = 1) \\
  &\quad = P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0) + P(Y = 1|X = -1)P(X = -1) \\
  &\quad = 1 \cdot 1/4 + 0 \cdot 1/2 + 1 \cdot 1/4 \\
  &\quad = 1/2.
  \end{aligned}$$

- The joint density of (X, Y) is given by

  $$\begin{aligned}
  P(X = 0, Y = 0) &= P(Y = 0|X = 0)P(X = 0) = 1 \cdot 1/2 = 1/2; \\
  P(X = 0, Y = 1) &= P(Y = 1|X = 0)P(X = 0) = 0 \cdot 1/2 = 0; \\
  P(X = 1, Y = 0) &= P(Y = 0|X = 1)P(X = 1) = 0 \cdot 1/4 = 0; \\
  P(X = 1, Y = 1) &= P(Y = 1|X = 1)P(X = 1) = 1 \cdot 1/4 = 1/4; \\
  P(X = -1, Y = 0) &= P(Y = 0|X = -1)P(X = -1) = 0 \cdot 1/4 = 0; \\
  P(X = -1, Y = 1) &= P(Y = 1|X = -1)P(X = -1) = 1 \cdot 1/4 = 1/4.
  \end{aligned}$$

  Since, for example, $P(X = 0, Y = 0) = 1/2$, but $P(X = 0)P(Y = 0) = 1/2 \cdot 1/2 = 1/4$, we see that X and Y cannot be independent.

- The possible values of XY are 0, 1, -1. Hence,

  $$P(XY = 0) = P(X = 0, Y = 0) = 1/2$$

  and

  $$P(XY = 1) = P(X = 1, Y = 1) = 1/4$$

  using the computations above. By the law of total probability,

  $$P(XY = -1) = 1/4.$$

  (Equivalently, $P(XY = -1) = P(X = -1, Y = 1) = 1/4$.) Thus,

  $$\mathbb{E}(XY) = 0 \cdot P(XY = 0) + 1 \cdot P(XY = 1) + (-1) \cdot P(XY = -1) = 0 + 1/4 - 1/4 = 0.$$

Since $\mathbb{E}(X) = 0$ and $\mathbb{E}(Y) = 0$, we see that

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 = 0;$$

whence X and Y are uncorrelated.

(5.15) By definition,

$$\mathbb{E}((g \circ X)(h \circ Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [g(x) \cdot h(y)] f(x, y)\, dx\, dy.$$

Since X and Y are independent, we can write $f(x, y) = f_X(x) \cdot f_Y(y)$. Thus, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [g(x) \cdot h(y)] f(x, y)\, dx\, dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [g(x) \cdot h(y)] f_X(x) \cdot f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \int_{-\infty}^{\infty} h(y) f_Y(y)\, dy \\
&= \mathbb{E}(g \circ X)\mathbb{E}(h \circ Y)
\end{aligned}$$

as required.

(5.20) If $\mathrm{Var}(X) = \mathrm{Var}(Y) = 0$, then the third part of the Cauchy-Schwarz inequality implies $(\mathrm{Cov}(X, Y))^2 \leq 0$. But, for *any* number a, it must be the case $a^2 \geq 0$. Thus, we must have $(\mathrm{Cov}(X, Y))^2 \geq 0$. But the only way for $0 \leq (\mathrm{Cov}(X, Y))^2 \leq 0$ is if $(\mathrm{Cov}(X, Y))^2 = 0$. Thus, since the only number whose square is 0 is 0, we have $\mathrm{Cov}(X, Y) = 0$.

Textbook

(1.1(g)) Briefly: The parameter of interest is the *lifetime of a certain type of transistor*. The population, obviously, consists of these transistors. The inferential objective of the electrical engineer is to determine whether or not the average lifetime is greater than 500 hours. This can be done by collecting a random sample, and either constructing a confidence interval or conducting a hypothesis test. It *should* be a simple matter to collect a sample of transistors since, presumably, they come from an assembly line of some sort.

(1.9) By definition,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 \quad \text{where} \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

Notice that $(y_i - \overline{y})^2 = y_i^2 - 2\overline{y}y_i + \overline{y}^2$. Thus,

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \overline{y})^2 &= \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} (-2)\overline{y}y_i + \sum_{i=1}^{n} \overline{y}^2 \text{ (using c)} \\
&= \sum_{i=1}^{n} y_i^2 - 2\overline{y} \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \overline{y}^2 \text{ (using a)} \\
&= \sum_{i=1}^{n} y_i^2 - 2\overline{y} \sum_{i=1}^{n} y_i + n\overline{y}^2 \text{ (using b).}
\end{aligned}$$

But,

$$\sum_{i=1}^{n} y_i = n\overline{y}$$

so we can substitute that into the above to conclude

$$\sum_{i=1}^{n} y_i^2 - 2\overline{y}\sum_{i=1}^{n} y_i + n\overline{y}^2 = \sum_{i=1}^{n} y_i^2 - 2n\overline{y}^2 + n\overline{y}^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2.$$

Substituting back into this for $\overline{y}$ gives

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right].$$

(1.10) If our data consists of $\{1, 4, 2, 1, 3, 3\}$, then we trivially compute

$$\sum_{i=1}^{6} y_i = 1 + 4 + 2 + 1 + 3 + 3 = 14$$

and

$$\sum_{i=1}^{6} y_i^2 = 1^2 + 4^2 + 2^2 + 1^2 + 3^2 + 3^2 = 40.$$

Thus,

$$s^2 = \frac{1}{6-1}\left[\sum_{i=1}^{6} y_i^2 - \frac{1}{6}\left(\sum_{i=1}^{6} y_i\right)^2\right] = \frac{1}{5}\left[40 - \frac{1}{6}\cdot 14^2\right] = \frac{22}{15}.$$

Note that writing garbage with decimals is unacceptable here!

(1.33) Briefly: Lead content readings must be non-negative. Since 0 is only 0.33 standard deviations below the mean, the population can only extend 0.33 standard deviations below the mean. This radically skews the distribution so that it *cannot* be normal. (If this is unclear, draw a picture.)

(8.2) **(a)** If $\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$, then

$$\mathbb{E}(\hat{\theta}_3) = \mathbb{E}(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = \mathbb{E}(a\hat{\theta}_1) + \mathbb{E}((1-a)\hat{\theta}_2) = a\mathbb{E}(\hat{\theta}_1) + (1-a)\mathbb{E}(\hat{\theta}_2) = a\theta + (1-a)\theta = \theta.$$

(b) If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, then

$$\begin{aligned}\operatorname{Var}(\hat{\theta}_3) = \operatorname{Var}(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = \operatorname{Var}(a\hat{\theta}_1) + \operatorname{Var}((1-a)\hat{\theta}_2) &= a^2\operatorname{Var}(\hat{\theta}_1) + (1-a)^2\operatorname{Var}(\hat{\theta}_2) \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2.\end{aligned}$$

In order to minimize $\operatorname{Var}(\hat{\theta}_3)$ as a function of a, we use the methods of elementary calculus. Suppose that $f(a) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$. Then, $f'(a) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2$. The critical points of this function occur when $f'(a) = 0$. Thus, the only critical point occurs at

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Technically, you should use the second derivative test to verify that a_1 is a minimum. Since $f''(a) = 2\sigma_1^2 + 2\sigma_2^2 > 0$ for all a, it is, in fact, a minimum.

(8.4) (a) Recall that if Y has the exponential density as given in the problem, then $\mathbb{E}(Y) = \theta$. This was done in Stat 251. In order to decide which estimators are unbiased, we simply compute $\mathbb{E}(\hat{\theta}_i)$ for each i. Four of these are easy:

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_1) &= \mathbb{E}(Y_1) = \theta; \\
\mathbb{E}(\hat{\theta}_2) &= \mathbb{E}\left(\frac{Y_1 + Y_2}{2}\right) = \frac{\mathbb{E}(Y_1) + \mathbb{E}(Y_2)}{2} = \frac{\theta + \theta}{2} = \theta; \\
\mathbb{E}(\hat{\theta}_3) &= \mathbb{E}\left(\frac{Y_1 + 2Y_2}{3}\right) = \frac{\mathbb{E}(Y_1) + 2\mathbb{E}(Y_2)}{3} = \frac{\theta + 2\theta}{3} = \theta; \\
\mathbb{E}(\hat{\theta}_5) &= \mathbb{E}(\overline{Y}) = \mathbb{E}\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) = \frac{\mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_3)}{3} = \frac{\theta + \theta + \theta}{3} = \theta.
\end{aligned}$$

In order to compute $\mathbb{E}(\hat{\theta}_4) = \mathbb{E}(\min(Y_1, Y_2, Y_3))$ we need to do a bit of work.

$$\begin{aligned}
P(\min(Y_1, Y_2, Y_3) > t) = P(Y_1 > t, Y_2 > t, Y_3 > t) &= P(Y_1 > t) \cdot P(Y_2 > t) \cdot P(Y_3 > t) \\
&= [P(Y_1 > t)]^3 \\
&= e^{-3t/\theta}.
\end{aligned}$$

Thus, $f(t) = (3/\theta)e^{-3t/\theta}$ which, as you will notice, is the density of an Exponential$(\theta/3)$ random variable. (WHY?) Thus,

$$\mathbb{E}(\hat{\theta}_4) = \mathbb{E}(\min(Y_1, Y_2, Y_3)) = \frac{\theta}{3}.$$

Hence, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, and $\hat{\theta}_5$ are unbiased, while $\hat{\theta}_4$ is biased.

(b) To decide which has the smallest variance, we simply compute. Recall that an Exponential(θ) random variable has variance θ^2. Thus,

$$\begin{aligned}
\operatorname{Var}(\hat{\theta}_1) &= \operatorname{Var}(Y_1) = \theta^2; \\
\operatorname{Var}(\hat{\theta}_2) &= \operatorname{Var}\left(\frac{Y_1 + Y_2}{2}\right) = \frac{\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2)}{4} = \frac{\theta^2 + \theta^2}{4} = \frac{\theta^2}{2}; \\
\operatorname{Var}(\hat{\theta}_3) &= \operatorname{Var}\left(\frac{Y_1 + 2Y_2}{3}\right) = \frac{\operatorname{Var}(Y_1) + 4\operatorname{Var}(Y_2)}{9} = \frac{\theta^2 + 4\theta^2}{9} = \frac{5\theta^2}{9}; \\
\operatorname{Var}(\hat{\theta}_5) &= \operatorname{Var}(\overline{Y}) = \operatorname{Var}\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) = \frac{\operatorname{Var}(Y_1) + \operatorname{Var}(Y_2) + \operatorname{Var}(Y_3)}{9} = \frac{\theta^2 + \theta^2 + \theta^2}{9} = \frac{\theta^2}{3}.
\end{aligned}$$

Thus, $\hat{\theta}_5$ has the smallest variance. In fact, we will show later that it is the *minimum variance unbiased estimator*. That is, *no* other unbiased estimator of the mean will have smaller variance than $\overline{Y}$.

(8.6) Recall that a Poisson(λ) random variable has mean λ and variance λ. This was also done in Stat 251.

(a) Since λ is the mean of a Poisson(λ) random variable, then a natural unbiased estimator for λ is

$$\hat{\lambda} = \overline{Y}.$$

(As you saw in problem **(8.4)**, there is NO unique unbiased estimator, so many other answers are possible.) It is a simple matter to compute that

$$\mathbb{E}(\hat{\lambda}) = \mathbb{E}(\overline{Y}) = \lambda \quad \text{and} \quad \mathrm{Var}(\hat{\lambda}) = \frac{\lambda}{n}.$$

We will need these in **(c)**.

(b) If $C = 3Y + Y^2$, then

$$\mathbb{E}(C) = \mathbb{E}(3Y) + \mathbb{E}(Y^2) = 3\mathbb{E}(Y) + [\mathrm{Var}(Y) + \mathbb{E}(Y)^2] = 3\lambda + [\lambda + \lambda^2] = 4\lambda + \lambda^2.$$

(c) This part is a little tricky. There is NO algorithm to solve it; instead you must THINK. Since $\mathbb{E}(C)$ depends on the *parameter* λ, we do not know its actual value. Therefore, we can *estimate* it. Suppose that $\theta = \mathbb{E}(C)$. Then, a *natural* estimator of $\theta = 4\lambda + \lambda^2$ is

$$\hat{\theta} = 4\hat{\lambda} + \hat{\lambda}^2,$$

where $\hat{\lambda} = \overline{Y}$ as in **(a)**. However, if we compute $\mathbb{E}(\hat{\lambda})$ we find

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}(4\hat{\lambda}) + \mathbb{E}(\hat{\lambda}^2) = 4\mathbb{E}(\hat{\lambda}) + [\mathrm{Var}(\hat{\lambda}) + \mathbb{E}(\hat{\lambda})^2] = 4\lambda + \frac{\lambda}{n} + \lambda^2.$$

This does not equal θ, so that $\hat{\theta}$ is NOT unbiased. However, a little thought shows that if we *define*

$$\tilde{\theta} := 4\hat{\lambda} + \hat{\lambda}^2 - \frac{\hat{\lambda}}{n} = 4\overline{Y} + \overline{Y}^2 - \frac{\overline{Y}}{n}$$

then, $\mathbb{E}(\tilde{\theta}) = 4\hat{\lambda} + \hat{\lambda}^2$ so that $\tilde{\theta}$ IS an unbiased estimator of $\theta = \mathbb{E}(C)$.

(8.8) If Y is a uniform $(\theta, \theta + 1)$ random variable, then its density is

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

It is a simple matter to compute

$$\mathbb{E}(Y) = \frac{2\theta + 1}{2} \quad \text{and} \quad \mathrm{Var}\, Y = \frac{1}{12}.$$

(a) Hence,

$$\mathbb{E}(\overline{Y}) = \mathbb{E}\left(\frac{Y_1 + \cdots + Y_n}{n}\right) = \frac{\mathbb{E}(Y_1) + \cdots + \mathbb{E}(Y_n)}{n} = \frac{\frac{2\theta+1}{2} + \cdots + \frac{2\theta+1}{2}}{n} = \frac{2n\theta + n}{2n} = \theta + \frac{1}{2}.$$

We now find

$$B(\overline{Y}) = \mathbb{E}(\overline{Y}) - \theta = \left(\theta + \frac{1}{2}\right) - \theta = \frac{1}{2}.$$

(b) A little thought shows that our calculation in **(a)** iummediately suggests a natural unbiased estimator of θ, namely

$$\hat{\theta} = \overline{Y} - \frac{1}{2}.$$

(c) We first compute that

$$\operatorname{Var}(\overline{Y}) = \operatorname{Var}\left(\frac{Y_1 + \cdots + Y_n}{n}\right) = \frac{\operatorname{Var}(Y_1) + \cdots + \operatorname{Var}(Y_n)}{n^2} = \frac{1/12 + \cdots + 1/12}{n^2} = \frac{1}{12n}.$$

As on page 367,

$$MSE(\overline{Y}) = \operatorname{Var}(\overline{Y}) + (B(\overline{Y}))^2$$

so that

$$MSE(\overline{Y}) = \frac{1}{12n} + \left(\frac{1}{2}\right)^2 = \frac{3n+1}{12n}.$$

(8.19) (a) The average calcium concentration in drinking water for kidney stone patients in the Carolinas is 11.3 ppm. A bound on the error of estimation is given by

$$2\mathrm{SE}_{\text{calcium}} \approx 2 \cdot \frac{16.6}{\sqrt{467}} \approx 1.54 \text{ ppm}.$$

It is WRONG if you write an equality at the second or last step. There is no equality there! In other words, an approximate 95% confidence interval is given by 11.3 ± 1.54.

(b) The difference in mean ages for kidney stone patients in the Carolinas and in the Rockies is $46.4 - 45.1 = 1.3$ years. A bound on the error of estimation is given by

$$2\,\mathrm{SE}_{\text{age}_{\text{Rockies}} - \text{age}_{\text{Carolinas}}} = 2\sqrt{\mathrm{SE}^2_{\text{age}_{\text{Rockies}}} + \mathrm{SE}^2_{\text{age}_{\text{Carolinas}}}} \approx 2\sqrt{\left(\frac{10.2}{\sqrt{467}}\right)^2 + \left(\frac{9.8}{\sqrt{191}}\right)^2} \approx 1.7 \text{ ppm}.$$

It is WRONG if you write an equality at the second or last step. There is no equality there! In other words, an approximate 95% confidence interval is given by 1.3 ± 1.7.

(c) The difference in proportions of kidney stone patients from the Carolinas and the Rockies who were smokers at the time of the study is $0.78 - 0.61 = 0.17$. A two standard deviation bound on the difference in proportions is given by

$$\begin{aligned} 2\,\mathrm{SE}_{\text{smoke}_{\text{Rockies}} - \text{smoke}_{\text{Carolinas}}} &= 2\sqrt{\mathrm{SE}^2_{\text{smoke}_{\text{Rockies}}} + \mathrm{SE}^2_{\text{smoke}_{\text{Carolinas}}}} \\ &\approx 2\sqrt{\left(\sqrt{\frac{0.78 \cdot 0.22}{467}}\right)^2 + \left(\sqrt{\frac{0.61 \cdot 0.39}{191}}\right)^2} \\ &\approx 0.08. \end{aligned}$$

It is WRONG if you write an equality at the second or last step. There is no equality there! In other words, an approximate 95% confidence interval is given by 0.17 ± 0.08.

(8.24) Let p_1 denote the unknown proportion of first-born or only child college graduates. Thus, $\hat{p}_1 = 126/180 = 0.7$, and

$$\sigma_{\hat{p}_1} \approx \sqrt{\frac{126/180 \cdot 54/180}{180}} \approx 0.034.$$

Let p_2 denote the unknown proportion of first-born or only child college graduates. Thus, $\hat{p}_2 = 54/100 = 0.54$, and

$$\sigma_{\hat{p}_2} \approx \sqrt{\frac{54/100 \cdot 46/100}{100}} \approx 0.050.$$

Hence, the difference in proportions is given by $\hat{p}_1 - \hat{p}_2 = 126/180 - 54/100 = 4/25 = 0.16$ with standard error

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2}} \approx 0.0604.$$

A bound on the error of estimation is therefore $2\sigma_{\hat{p}_1 - \hat{p}_2} \approx 0.121$. Note that an approximate 95% confidence interval for $p_1 - p_2$ is therefore

$$(\hat{p}_1 - \hat{p}_2) \pm 2\sigma_{\hat{p}_1 - \hat{p}_2} \quad \text{or} \quad 0.16 \pm 0.121.$$

It is equivalent to consider $\hat{p}_2 - \hat{p}_1$ instead. The only difference is the minus sign in the estimate of the difference. The bound on the error of estimation is unchanged. (why?)