Chapter 7

# Point Estimation

7.1 For each value of $x$, the MLE $\hat{\theta}$ is the value of $\theta$ that maximizes $f(x|\theta)$. These values are in the following table.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\hat{\theta}$ | 1 | 1 | 2 or 3 | 3 | 3 |

At $x = 2$, $f(x|2) = f(x|3) = 1/4$ are both maxima, so both $\hat{\theta} = 2$ or $\hat{\theta} = 3$ are MLEs.

7.2 a.

$$\begin{aligned} L(\beta|x) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_i^{\alpha-1} e^{-x_i/\beta} = \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/\beta} \\ \log L(\beta|x) &= -\log \Gamma(\alpha)^n - n\alpha \log \beta + (\alpha-1) \log \left[\prod_{i=1}^{n} x_i\right] - \frac{\sum_i x_i}{\beta} \\ \frac{\partial \log L}{\partial \beta} &= -\frac{n\alpha}{\beta} + \frac{\sum_i x_i}{\beta^2} \end{aligned}$$

Set the partial derivative equal to 0 and solve for $\beta$ to obtain $\hat{\beta} = \sum_i x_i/(n\alpha)$. To check that this is a maximum, calculate

$$\left.\frac{\partial^2 \log L}{\partial \beta^2}\right|_{\beta=\hat{\beta}} = \left.\frac{n\alpha}{\beta^2} - \frac{2\sum_i x_i}{\beta^3}\right|_{\beta=\hat{\beta}} = \frac{(n\alpha)^3}{\left(\sum_i x_i\right)^2} - \frac{2(n\alpha)^3}{\left(\sum_i x_i\right)^2} = -\frac{(n\alpha)^3}{\left(\sum_i x_i\right)^2} < 0.$$

Because $\hat{\beta}$ is the unique point where the derivative is 0 and it is a local maximum, it is a global maximum. That is, $\hat{\beta}$ is the MLE.

b. Now the likelihood function is

$$L(\alpha, \beta|x) = \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/\beta},$$

the same as in part (a) except $\alpha$ and $\beta$ are both variables. There is no analytic form for the MLEs, The values $\hat{\alpha}$ and $\hat{\beta}$ that maximize $L$. One approach to finding $\hat{\alpha}$ and $\hat{\beta}$ would be to numerically maximize the function of two arguments. But it is usually best to do as much as possible analytically, first, and perhaps reduce the complexity of the numerical problem. From part (a), for each fixed value of $\alpha$, the value of $\beta$ that maximizes $L$ is $\sum_i x_i/(n\alpha)$. Substitute this into $L$. Then we just need to maximize the function of the one variable $\alpha$ given by

$$\begin{aligned} & \frac{1}{\Gamma(\alpha)^n \left(\sum_i x_i/(n\alpha)\right)^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/(\Sigma_i x_i/(n\alpha))} \\ =\ & \frac{1}{\Gamma(\alpha)^n \left(\sum_i x_i/(n\alpha)\right)^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-n\alpha}. \end{aligned}$$

For the given data, $n = 14$ and $\sum_i x_i = 323.6$. Many computer programs can be used to maximize this function. From PROC NLIN in SAS we obtain $\hat{\alpha} = 514.219$ and, hence, $\hat{\beta} = \frac{323.6}{14(514.219)} = .0450$.

7.3 The log function is a strictly monotone increasing function. Therefore, $L(\theta|\mathbf{x}) > L(\theta'|\mathbf{x})$ if and only if $\log L(\theta|\mathbf{x}) > \log L(\theta'|\mathbf{x})$. So the value $\hat{\theta}$ that maximizes $\log L(\theta|\mathbf{x})$ is the same as the value that maximizes $L(\theta|\mathbf{x})$.

7.5 a. The value $\hat{z}$ solves the equation

$$(1-p)^n = \prod_i (1 - x_i z),$$

where $0 \le z \le (\max_i x_i)^{-1}$. Let $\hat{k}$ = greatest integer less than or equal to $1/\hat{z}$. Then from Example 7.2.9, $\hat{k}$ must satisfy

$$[k(1-p)]^n \ge \prod_i (k - x_i) \qquad \text{and} \qquad [(k+1)(1-p)]^n < \prod_i (k+1-x_i).$$

Because the right-hand side of the first equation is decreasing in $\hat{z}$, and because $\hat{k} \le 1/\hat{z}$ (so $\hat{z} \le 1/\hat{k}$) and $\hat{k}+1 > 1/\hat{z}$, $\hat{k}$ must satisfy the two inequalities. Thus $\hat{k}$ is the MLE.

b. For $p = 1/2$, we must solve $\left(\frac{1}{2}\right)^4 = (1-20z)(1-z)(1-19z)$, which can be reduced to the cubic equation $-380z^3 + 419z^2 - 40z + 15/16 = 0$. The roots are .9998, .0646, and .0381, leading to candidates of 1, 15, and 26 for $\hat{k}$. The first two are less than $\max_i x_i$. Thus $\hat{k} = 26$.

7.6 a. $f(\mathbf{x}|\theta) = \prod_i \theta x_i^{-2} I_{[\theta,\infty)}(x_i) = \left(\prod_i x_i^{-2}\right) \theta^n I_{[\theta,\infty)}(x_{(1)})$. Thus, $X_{(1)}$ is a sufficient statistic for $\theta$ by the Factorization Theorem.

b. $L(\theta|\mathbf{x}) = \theta^n \left(\prod_i x_i^{-2}\right) I_{[\theta,\infty)}(x_{(1)})$. $\theta^n$ is increasing in $\theta$. The second term does not involve $\theta$. So to maximize $L(\theta|\mathbf{x})$, we want to make $\theta$ as large as possible. But because of the indicator function, $L(\theta|\mathbf{x}) = 0$ if $\theta > x_{(1)}$. Thus, $\hat{\theta} = x_{(1)}$.

c. $\mathrm{E}\,X = \int_\theta^\infty \theta x^{-1}\,dx = \theta \log x|_\theta^\infty = \infty$. Thus the method of moments estimator of $\theta$ does not exist. (This is the Pareto distribution with $\alpha = \theta$, $\beta = 1$.)

7.7 $L(0|\mathbf{x}) = 1$, $0 < x_i < 1$, and $L(1|\mathbf{x}) = \prod_i 1/(2\sqrt{x_i})$, $0 < x_i < 1$. Thus, the MLE is 0 if $1 \ge \prod_i 1/(2\sqrt{x_i})$, and the MLE is 1 if $1 < \prod_i 1/(2\sqrt{x_i})$.

7.8 a. $\mathrm{E}\,X^2 = \mathrm{Var}\,X + \mu^2 = \sigma^2$. Therefore $X^2$ is an unbiased estimator of $\sigma^2$.

b.

$$\begin{aligned}
L(\sigma|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}. \quad \log L(\sigma|\mathbf{x}) = \log(2\pi)^{-1/2} - \log\sigma - x^2/(2\sigma^2).\\
\frac{\partial \log L}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{x^2}{\sigma^3} \stackrel{\text{set}}{=} 0 \Rightarrow \hat{\sigma}X^2 = \hat{\sigma}^3 \ \Rightarrow \hat{\sigma} = \sqrt{X^2} = |X|.\\
\frac{\partial^2 \log L}{\partial \sigma^2} &= \frac{-3x^2\sigma^2}{\sigma^6} + \frac{1}{\sigma^2}, \text{ which is negative at } \hat{\sigma} = |x|.
\end{aligned}$$

Thus, $\hat{\sigma} = |x|$ is a local maximum. Because it is the only place where the first derivative is zero, it is also a global maximum.

c. Because $\mathrm{E}\,X = 0$ is known, just equate $\mathrm{E}\,X^2 = \sigma^2 = \frac{1}{n}\sum_{i=1}^1 X_i^2 = X^2 \Rightarrow \hat{\sigma} = |X|$.

7.9 This is a uniform$(0,\theta)$ model. So $\mathrm{E}\,X = (0+\theta)/2 = \theta/2$. The method of moments estimator is the solution to the equation $\tilde{\theta}/2 = \bar{X}$, that is, $\tilde{\theta} = 2\bar{X}$. Because $\tilde{\theta}$ is a simple function of the sample mean, its mean and variance are easy to calculate. We have

$$\mathrm{E}\,\tilde{\theta} = 2\mathrm{E}\,\bar{X} = 2\mathrm{E}\,X = 2\frac{\theta}{2} = \theta, \quad \text{and} \quad \mathrm{Var}\,\tilde{\theta} = 4\mathrm{Var}\,\bar{X} = 4\frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

The likelihood function is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}),$$

where $x_{(1)}$ and $x_{(n)}$ are the smallest and largest order statistics. For $\theta \geq x_{(n)}$, $L = 1/\theta^n$, a decreasing function. So for $\theta \geq x_{(n)}$, $L$ is maximized at $\hat{\theta} = x_{(n)}$. $L = 0$ for $\theta < x_{(n)}$. So the overall maximum, the MLE, is $\hat{\theta} = X_{(n)}$. The pdf of $\hat{\theta} = X_{(n)}$ is $nx^{n-1}/\theta^n$, $0 \leq x \leq \theta$. This can be used to calculate

$$\mathrm{E}\,\hat{\theta} = \frac{n}{n+1}\theta, \quad \mathrm{E}\,\hat{\theta}^2 = \frac{n}{n+2}\theta^2 \quad \text{and} \quad \mathrm{Var}\,\hat{\theta} = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

$\tilde{\theta}$ is an unbiased estimator of $\theta$; $\hat{\theta}$ is a biased estimator. If $n$ is large, the bias is not large because $n/(n+1)$ is close to one. But if $n$ is small, the bias is quite large. On the other hand, $\mathrm{Var}\,\hat{\theta} < \mathrm{Var}\,\tilde{\theta}$ for all $\theta$. So, if $n$ is large, $\hat{\theta}$ is probably preferable to $\tilde{\theta}$.

7.10 a. $f(\mathbf{x}|\theta) = \prod_i \frac{\alpha}{\beta^\alpha} x_i^{\alpha-1} I_{[0,\beta]}(x_i) = \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_i x_i\right)^{\alpha-1} I_{(-\infty,\beta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}) = L(\alpha,\beta|\mathbf{x})$. By the Factorization Theorem, $(\prod_i X_i, X_{(n)})$ are sufficient.

b. For any fixed $\alpha$, $L(\alpha,\beta|\mathbf{x}) = 0$ if $\beta < x_{(n)}$, and $L(\alpha,\beta|\mathbf{x})$ a decreasing function of $\beta$ if $\beta \geq x_{(n)}$. Thus, $X_{(n)}$ is the MLE of $\beta$. For the MLE of $\alpha$ calculate

$$\frac{\partial}{\partial\alpha}\log L = \frac{\partial}{\partial\alpha}\left[n\log\alpha - n\alpha\log\beta + (\alpha-1)\log\prod_i x_i\right] = \frac{n}{\alpha} - n\log\beta + \log\prod_i x_i.$$

Set the derivative equal to zero and use $\hat{\beta} = X_{(n)}$ to obtain

$$\hat{\alpha} = \frac{n}{n\log X_{(n)} - \log\prod_i X_i} = \left[\frac{1}{n}\sum_i \left(\log X_{(n)} - \log X_i\right)\right]^{-1}.$$

The second derivative is $-n/\alpha^2 < 0$, so this is the MLE.

c. $X_{(n)} = 25.0$, $\log\prod_i X_i = \sum_i \log X_i = 43.95 \Rightarrow \hat{\beta} = 25.0$, $\hat{\alpha} = 12.59$.

7.11 a.

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_i \theta x_i^{\theta-1} = \theta^n \left(\prod_i x_i\right)^{\theta-1} = L(\theta|\mathbf{x}) \\ \frac{d}{d\theta}\log L &= \frac{d}{d\theta}\left[n\log\theta + (\theta-1)\log\prod_i x_i\right] = \frac{n}{\theta} + \sum_i \log x_i. \end{aligned}$$

Set the derivative equal to zero and solve for $\theta$ to obtain $\hat{\theta} = (-\frac{1}{n}\sum_i \log x_i)^{-1}$. The second derivative is $-n/\theta^2 < 0$, so this is the MLE. To calculate the variance of $\hat{\theta}$, note that $Y_i = -\log X_i \sim \text{exponential}(1/\theta)$, so $-\sum_i \log X_i \sim \text{gamma}(n, 1/\theta)$. Thus $\hat{\theta} = n/T$, where $T \sim \text{gamma}(n, 1/\theta)$. We can either calculate the first and second moments directly, or use the fact that $\hat{\theta}$ is inverted gamma (page 51). We have

$$\begin{aligned} \mathrm{E}\frac{1}{T} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}. \\ \mathrm{E}\frac{1}{T^2} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t^2} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}, \end{aligned}$$

and thus

$$\mathrm{E}\,\hat\theta = \frac{n}{n-1}\theta \quad \text{and} \quad \mathrm{Var}\,\hat\theta = \frac{n^2}{(n-1)^2(n-2)}\theta^2 \to 0 \text{ as } n\to\infty.$$

b. Because $X \sim \text{beta}(\theta,1)$, $\mathrm{E}\,X = \theta/(\theta+1)$ and the method of moments estimator is the solution to

$$\frac{1}{n}\sum_i X_i = \frac{\theta}{\theta{+}1} \Rightarrow \tilde\theta = \frac{\sum_i X_i}{n{-}\sum_i X_i}.$$

7.12 $X_i \sim$ iid Bernoulli$(\theta)$, $0\le\theta\le 1/2$.

a. method of moments:

$$\mathrm{E}\,X = \theta = \frac{1}{n}\sum_i X_i = \bar X \quad\Rightarrow\quad \tilde\theta = \bar X.$$

MLE: In Example 7.2.7, we showed that $L(\theta|\mathbf{x})$ is increasing for $\theta\le\bar x$ and is decreasing for $\theta\ge\bar x$. Remember that $0\le\theta\le 1/2$ in this exercise. Therefore, when $\bar X\le 1/2$, $\bar X$ is the MLE of $\theta$, because $\bar X$ is the overall maximum of $L(\theta|\mathbf{x})$. When $\bar X > 1/2$, $L(\theta|\mathbf{x})$ is an increasing function of $\theta$ on $[0,1/2]$ and obtains its maximum at the upper bound of $\theta$ which is $1/2$. So the MLE is $\hat\theta = \min\left\{\bar X, 1/2\right\}$.

b. The MSE of $\tilde\theta$ is $\mathrm{MSE}(\tilde\theta) = \mathrm{Var}\,\tilde\theta + \mathrm{bias}(\tilde\theta)^2 = (\theta(1-\theta)/n) + 0^2 = \theta(1-\theta)/n$. There is no simple formula for $\mathrm{MSE}(\hat\theta)$, but an expression is

$$\begin{aligned}\mathrm{MSE}(\hat\theta) &= \mathrm{E}(\hat\theta-\theta)^2 = \sum_{y=0}^{n}(\hat\theta-\theta)^2\binom{n}{y}\theta^y(1-\theta)^{n-y}\\ &= \sum_{y=0}^{[n/2]}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n}\left(\frac{1}{2}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y},\end{aligned}$$

where $Y=\sum_i X_i \sim \text{binomial}(n,\theta)$ and $[n/2]=n/2$, if $n$ is even, and $[n/2]=(n-1)/2$, if $n$ is odd.

c. Using the notation used in (b), we have

$$\mathrm{MSE}(\tilde\theta) = \mathrm{E}(\bar X-\theta)^2 = \sum_{y=0}^{n}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

Therefore,

$$\begin{aligned}\mathrm{MSE}(\tilde\theta)-\mathrm{MSE}(\hat\theta) &= \sum_{y=[n/2]+1}^{n}\left[\left(\frac{y}{n}-\theta\right)^2-\left(\frac{1}{2}-\theta\right)^2\right]\binom{n}{y}\theta^y(1-\theta)^{n-y}\\ &= \sum_{y=[n/2]+1}^{n}\left(\frac{y}{n}+\frac{1}{2}-2\theta\right)\left(\frac{y}{n}-\frac{1}{2}\right)\binom{n}{y}\theta^y(1-\theta)^{n-y}.\end{aligned}$$

The facts that $y/n>1/2$ in the sum and $\theta\le 1/2$ imply that every term in the sum is positive. Therefore $\mathrm{MSE}(\hat\theta)<\mathrm{MSE}(\tilde\theta)$ for every $\theta$ in $0<\theta\le 1/2$. (Note: $\mathrm{MSE}(\hat\theta)=\mathrm{MSE}(\tilde\theta)=0$ at $\theta=0$.)

7.13 $L(\theta|\mathbf{x}) = \prod_i \frac{1}{2}e^{-\frac{1}{2}|x_i-\theta|} = \frac{1}{2^n}e^{-\frac{1}{2}\Sigma_i|x_i-\theta|}$, so the MLE minimizes $\sum_i|x_i-\theta| = \sum_i|x_{(i)}-\theta|$, where $x_{(1)},\ldots,x_{(n)}$ are the order statistics. For $x_{(j)}\le\theta\le x_{(j+1)}$,

$$\sum_{i=1}^{n}|x_{(i)}-\theta| = \sum_{i=1}^{j}(\theta-x_{(i)}) + \sum_{i=j+1}^{n}(x_{(i)}-\theta) = (2j-n)\theta - \sum_{i=1}^{j}x_{(i)} + \sum_{i=j+1}^{n}x_{(i)}.$$

This is a linear function of $\theta$ that decreases for $j < n/2$ and increases for $j > n/2$. If $n$ is even, $2j - n = 0$ if $j = n/2$. So the likelihood is constant between $x_{(n/2)}$ and $x_{((n/2)+1)}$, and any value in this interval is the MLE. Usually the midpoint of this interval is taken as the MLE. If $n$ is odd, the likelihood is minimized at $\hat{\theta} = x_{((n+1)/2)}$.

7.15 a. The likelihood is

$$L(\mu, \lambda|\mathbf{x}) = \frac{\lambda^{n/2}}{(2\pi)^n \prod_i x_i} \exp\left\{-\frac{\lambda}{2}\sum_i \frac{(x_i-\mu)^2}{\mu^2 x_i}\right\}.$$

For fixed $\lambda$, maximizing with respect to $\mu$ is equivalent to minimizing the sum in the exponential.

$$\frac{d}{d\mu}\sum_i \frac{(x_i-\mu)^2}{\mu^2 x_i} = \frac{d}{d\mu}\sum_i \frac{\left((x_i/\mu)-1\right)^2}{x_i} = -\sum_i \frac{2\left((x_i/\mu)-1\right)}{x_i}\frac{x_i}{\mu^2}.$$

Setting this equal to zero is equivalent to setting

$$\sum_i \left(\frac{x_i}{\mu} - 1\right) = 0,$$

and solving for $\mu$ yields $\hat{\mu}_n = \bar{x}$. Plugging in this $\hat{\mu}_n$ and maximizing with respect to $\lambda$ amounts to maximizing an expression of the form $\lambda^{n/2}e^{-\lambda b}$. Simple calculus yields

$$\hat{\lambda}_n = \frac{n}{2b} \quad \text{where} \quad b = \sum_i \frac{(x_i-\bar{x})^2}{2\bar{x}^2 x_i}.$$

Finally,

$$2b = \sum_i \frac{x_i}{\bar{x}^2} - 2\sum_i \frac{1}{\bar{x}} + \sum_i \frac{1}{x_i} = -\frac{n}{\bar{x}} + \sum_i \frac{1}{x_i} = \sum_i\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right).$$

b. This is the same as Exercise 6.27b.

c. This involved algebra can be found in Schwarz and Samanta (1991).

7.17 a. This is a special case of the computation in Exercise 7.2a.

b. Make the transformation

$$z = (x_2 - 1)/x_1, w = x_1 \quad \Rightarrow \quad x_1 = w, x_2 = wz + 1.$$

The Jacobean is $|w|$, and

$$f_Z(z) = \int f_{X_1}(w) f_{X_2}(wz+1) w dw = \frac{1}{\theta^2}e^{-1/\theta}\int w e^{-w(1+z)/\theta} dw,$$

where the range of integration is $0 < w < -1/z$ if $z < 0$, $0 < w < \infty$ if $z > 0$. Thus,

$$f_Z(z) = \frac{1}{\theta^2}e^{-1/\theta}\begin{cases} \int_0^{-1/z} w e^{-w(1+z)/\theta} dw & \text{if } z < 0 \\ \int_0^{\infty} w e^{-w(1+z)/\theta} dw & \text{if } z \geq 0 \end{cases}$$

Using the fact that $\int w e^{-w/a} dw = -e^{-w/a}(aw + a^2)$, we have

$$f_Z(z) = e^{-1/\theta}\begin{cases} \frac{z\theta + e^{(1+z)/z\theta}(1+z-z\theta)}{\theta z(1+z)^2} & \text{if } z < 0 \\ \frac{1}{(1+z)^2} & \text{if } z \geq 0 \end{cases}$$

c. From part $(a)$ we get $\hat{\theta} = 1$. From part $(b)$, $X_2 = 1$ implies $Z = 0$ which, if we use the second density, gives us $\hat{\theta} = \infty$.

d. The posterior distributions are just the normalized likelihood times prior, so of course they are different.

7.18 a. The usual first two moment equations for $X$ and $Y$ are

$$\begin{aligned} \bar{x} &= \mathrm{E}\,X = \mu_X, & \frac{1}{n}\sum_i x_i^2 &= \mathrm{E}\,X^2 = \sigma_X^2 + \mu_X^2, \\ \bar{y} &= \mathrm{E}\,Y = \mu_Y, & \frac{1}{n}\sum_i y_i^2 &= \mathrm{E}\,Y^2 = \sigma_Y^2 + \mu_Y^2. \end{aligned}$$

We also need an equation involving $\rho$.

$$\frac{1}{n}\sum_i x_i y_i = \mathrm{E}\,XY = \mathrm{Cov}(X,Y) + (\mathrm{E}\,X)(\mathrm{E}\,Y) = \rho\sigma_X\sigma_Y + \mu_X\mu_Y.$$

Solving these five equations yields the estimators given. Facts such as

$$\frac{1}{n}\sum_i x_i^2 - \bar{x}^2 = \frac{\sum_i x_i^2 - \left(\sum_i x_i\right)^2/n}{n} = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

are used.

b. Two answers are provided. First, use the Miscellanea: For

$$L(\boldsymbol{\theta}|\mathbf{x}) = h(\mathbf{x})c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(\mathbf{x})\right),$$

the solutions to the $k$ equations $\sum_{j=1}^n t_i(x_j) = \mathrm{E}_\theta\left(\sum_{j=1}^n t_i(X_j)\right) = n\mathrm{E}_\theta t_i(X_1)$, $i = 1,\ldots,k$, provide the unique MLE for $\boldsymbol{\theta}$. Multiplying out the exponent in the bivariate normal pdf shows it has this exponential family form with $k = 5$ and $t_1(x,y) = x$, $t_2(x,y) = y$, $t_3(x,y) = x^2$, $t_4(x,y) = y^2$ and $t_5(x,y) = xy$. Setting up the method of moment equations, we have

$$\begin{aligned} \sum_i x_i &= n\mu_X, & \sum_i x_i^2 &= n(\mu_X^2 + \sigma_X^2), \\ \sum_i y_i &= n\mu_Y, & \sum_i y_i^2 &= n(\mu_Y^2 + \sigma_Y^2), \\ \sum_i x_i y_i &= \sum_i \left[\mathrm{Cov}(X,Y) + \mu_X\mu_Y\right] &&= n(\rho\sigma_X\sigma_Y + \mu_X\mu_Y). \end{aligned}$$

These are the same equations as in part (a) if you divide each one by $n$. So the MLEs are the same as the method of moment estimators in part (a).

For the second answer, use the hint in the book to write

$$\begin{aligned} L(\theta|\mathbf{x},\mathbf{y}) &= L(\theta|\mathbf{x})L(\theta,\mathbf{x}|\mathbf{y}) \\ &= \underbrace{(2\pi\sigma_X^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_X^2}\sum_i (x_i - \mu_X)^2\right\}}_{A} \\ &\quad \times \underbrace{\left(2\pi\sigma_Y^2(1-\rho^2)\right)^{-\frac{n}{2}}\exp\left[\frac{-1}{2\sigma_Y^2(1-\rho^2)}\sum_i\left\{y_i - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x_i - \mu_X)\right)\right\}^2\right]}_{B} \end{aligned}$$

We know that $\bar{x}$ and $\hat{\sigma}_X^2 = \sum_i (x_i - \bar{x})^2/n$ maximizes $A$; the question is whether given $\sigma_Y$, $\mu_Y$, and $\rho$, does $\bar{x}$, $\hat{\sigma}_X^2$ maximize $B$? Let us first fix $\sigma_X^2$ and look for $\hat{\mu}_X$, that maximizes $B$. We have

$$\begin{aligned}\frac{\partial \log B}{\partial \mu_X} &\propto -2\left(\sum_i \left[(y_i - \mu_Y) - \frac{\rho\sigma_Y}{\sigma_X}(x_i - \mu_X)\right]\right)\frac{\rho\sigma_Y}{\sigma_X} \stackrel{\text{set}}{=} 0 \\ \Rightarrow \sum_i (y_i - \mu_Y) &= \frac{\rho\sigma_Y}{\sigma_X}\Sigma(x_i - \hat{\mu}_X).\end{aligned}$$

Similarly do the same procedure for $L(\theta|\mathbf{y})L(\theta,\mathbf{y}|\mathbf{x})$ This implies $\sum_i (x_i - \mu_X) = \frac{\rho\sigma_X}{\sigma_Y}\sum_i (y_i - \hat{\mu}_Y)$. The solutions $\hat{\mu}_X$ and $\hat{\mu}_Y$ therefore must satisfy both equations. If $\sum_i (y_i - \hat{\mu}_Y) \neq 0$ or $\sum_i (x_i - \hat{\mu}_X) \neq 0$, we will get $\rho = 1/\rho$, so we need $\sum_i (y_i - \hat{\mu}_Y) = 0$ and $\sum_i (x_i - \hat{\mu}_X) = 0$. This implies $\hat{\mu}_X = \bar{x}$ and $\hat{\mu}_Y = \bar{y}$. ($\frac{\partial^2 \log B}{\partial \mu_X^2} < 0$. Therefore it is maximum). To get $\hat{\sigma}_X^2$ take

$$\begin{aligned}\frac{\partial \log B}{\partial \sigma_X^2} &\propto \sum_i \frac{\rho\sigma_Y}{\sigma_X^2}(x_i - \hat{\mu}_X)\left[(y_i - \mu_Y) - \frac{\rho\sigma_Y}{\sigma_X}(x_i - \mu_X)\right] \stackrel{\text{set}}{=} 0. \\ \Rightarrow \sum_i (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) &= \frac{\rho\sigma_Y}{\hat{\sigma}_X}\sum (x_i - \hat{\mu}_X)^2.\end{aligned}$$

Similarly, $\sum_i (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) = \frac{\rho\sigma_X}{\hat{\sigma}_Y}\sum_i (y_i - \hat{\mu}_Y)^2$. Thus $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ must satisfy the above two equations with $\hat{\mu}_X = \bar{X}$, $\hat{\mu}_Y = \bar{Y}$. This implies

$$\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\sum_i (x_i - \bar{x})^2 = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y}\sum_i (y_i - \bar{y})^2 \Rightarrow \frac{\sum_i (x_i - \bar{x})^2}{\hat{\sigma}_X^2} = \frac{\sum_i (y_i - \bar{y})^2}{\hat{\sigma}_Y^2}.$$

Therefore, $\hat{\sigma}_X^2 = a\sum_i (x_i - \bar{x})^2$, $\hat{\sigma}_Y^2 = a\sum_i (y_i - \bar{y})^2$ where $a$ is a constant. Combining the knowledge that $\left(\bar{x}, \frac{1}{n}\sum_i (x_i - \bar{x})^2\right) = (\hat{\mu}_X, \hat{\sigma}_X^2)$ maximizes $A$, we conclude that $a = 1/n$. Lastly, we find $\hat{\rho}$, the MLE of $\rho$. Write

$$\begin{aligned}&\log L(\bar{x}, \bar{y}, \hat{\sigma}_X^2, \hat{\sigma}_Y^2, \rho|\mathbf{x}, \mathbf{y}) \\ &= -\frac{n}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\sum_i \left[\frac{(x_i - \bar{x})^2}{\hat{\sigma}_X^2} - \frac{2\rho(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X, \hat{\sigma}_Y} + \frac{(y_i - \bar{y})^2}{\hat{\sigma}_Y^2}\right] \\ &= -\frac{n}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\left[2n - 2\rho\underbrace{\sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X\hat{\sigma}_Y}}_{A}\right]\end{aligned}$$

because $\hat{\sigma}_X^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$ and $\hat{\sigma}_Y^2 = \frac{1}{n}\sum_i (y_i - \bar{y})^2$. Now

$$\log L = -\frac{n}{2}\log(1-\rho^2) - \frac{n}{1-\rho^2} + \frac{\rho}{1-\rho^2}A$$

and

$$\frac{\partial \log L}{\partial \rho} = \frac{n}{1-\rho^2} - \frac{n\rho}{(1-\rho^2)^2} + \frac{A(1-\rho^2) + 2A\rho^2}{(1-\rho^2)^2} \stackrel{\text{set}}{=} 0.$$

This implies

$$\begin{aligned}\frac{A + A\rho^2 - n\hat{\rho} - n\hat{\rho}^3}{(1-\rho^2)^2} = 0 \quad &\Rightarrow \quad A(1+\hat{\rho}^2) = n\hat{\rho}(1+\hat{\rho}^2) \\ &\Rightarrow \quad \hat{\rho} = \frac{A}{n} = \frac{1}{n}\sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X\hat{\sigma}_Y}.\end{aligned}$$

7.19 a.

$$\begin{aligned} L(\theta|\mathbf{y}) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\beta x_i)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2)\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\beta^2 \sum_i x_i^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i\right). \end{aligned}$$

By Theorem 6.1.2, $(\sum_i Y_i^2, \sum_i x_i Y_i)$ is a sufficient statistic for $(\beta, \sigma^2)$.

b.

$$\log L(\beta,\sigma^2|\mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i - \frac{\beta^2}{2\sigma^2}\sum_i x_i^2.$$

For a fixed value of $\sigma^2$,

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2}\sum_i x_i y_i - \frac{\beta}{\sigma^2}\sum_i x_i^2 \stackrel{\text{set}}{=} 0 \quad\Rightarrow\quad \hat\beta = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

Also,

$$\frac{\partial^2 \log L}{\partial \beta^2} = \frac{1}{\sigma^2}\sum_i x_i^2 < 0,$$

so it is a maximum. Because $\hat\beta$ does not depend on $\sigma^2$, it is the MLE. And $\hat\beta$ is unbiased because

$$\mathrm{E}\,\hat\beta = \frac{\sum_i x_i \mathrm{E}\,Y_i}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot \beta x_i}{\sum_i x_i^2} = \beta.$$

c. $\hat\beta = \sum_i a_i Y_i$, where $a_i = x_i/\sum_j x_j^2$ are constants. By Corollary 4.6.10, $\hat\beta$ is normally distributed with mean $\beta$, and

$$\mathrm{Var}\,\hat\beta = \sum_i a_i^2 \mathrm{Var}\,Y_i = \sum_i \left(\frac{x_i}{\sum_j x_j^2}\right)^2 \sigma^2 = \frac{\sum_i x_i^2}{(\sum_j x_j^2)^2}\sigma^2 = \frac{\sigma^2}{\sum_i x_i^2}.$$

7.20 a.

$$\mathrm{E}\frac{\sum_i Y_i}{\sum_i x_i} = \frac{1}{\sum_i x_i}\sum_i \mathrm{E}\,Y_i = \frac{1}{\sum_i x_i}\sum_i \beta x_i = \beta.$$

b.

$$\mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right) = \frac{1}{(\sum_i x_i)^2}\sum_i \mathrm{Var}\,Y_i = \frac{\sum_i \sigma^2}{(\sum_i x_i)^2} = \frac{n\sigma^2}{n^2\bar{x}^2} = \frac{\sigma^2}{n\bar{x}^2}.$$

Because $\sum_i x_i^2 - n\bar{x}^2 = \sum_i (x_i - \bar{x})^2 \geq 0$, $\sum_i x_i^2 \geq n\bar{x}^2$. Hence,

$$\mathrm{Var}\,\hat\beta = \frac{\sigma^2}{\sum_i x_i^2} \leq \frac{\sigma^2}{n\bar{x}^2} = \mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right).$$

(In fact, $\hat\beta$ is BLUE (Best Linear Unbiased Estimator of $\beta$), as discussed in Section 11.3.2.)

7.21 a.

$$\mathrm{E}\frac{1}{n}\sum_i \frac{Y_i}{x_i} = \frac{1}{n}\sum_i \frac{\mathrm{E}\,Y_i}{x_i} = \frac{1}{n}\sum_i \frac{\beta x_i}{x_i} = \beta.$$

b.

$$\mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i} = \frac{1}{n^2}\sum_i \frac{\mathrm{Var}\,Y_i}{x_i^2} = \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2}.$$

Using Example 4.7.8 with $a_i = 1/x_i^2$ we obtain

$$\frac{1}{n}\sum_i \frac{1}{x_i^2} \geq \frac{n}{\sum_i x_i^2}.$$

Thus,

$$\mathrm{Var}\,\hat{\beta} = \frac{\sigma^2}{\sum_i x_i^2} \leq \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2} = \mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i}.$$

Because $g(u) = 1/u^2$ is convex, using Jensen's Inequality we have

$$\frac{1}{\bar{x}^2} \leq \frac{1}{n}\sum_i \frac{1}{x_i^2}.$$

Thus,

$$\mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right) = \frac{\sigma^2}{n\bar{x}^2} \leq \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2} = \mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i}.$$

7.22 a.

$$f(\bar{x},\theta) = f(\bar{x}|\theta)\pi(\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-n(\bar{x}-\theta)^2/(2\sigma^2)}\frac{1}{\sqrt{2\pi}\tau}e^{-(\theta-\mu)^2/2\tau^2}.$$

b. Factor the exponent in part (a) as

$$\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2 = -\frac{1}{2v^2}(\theta-\delta(\mathbf{x}))^2 - \frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2,$$

where $\delta(\mathbf{x}) = (\tau^2\bar{x}+(\sigma^2/n)\mu)/(\tau^2+\sigma^2/n)$ and $v = (\sigma^2\tau^2/n)\Big/(\tau+\sigma^2/n)$. Let $\mathrm{n}(a,b)$ denote the pdf of a normal distribution with mean $a$ and variance $b$. The above factorization shows that

$$f(\mathbf{x},\theta) = \mathrm{n}(\theta,\sigma^2/n)\times\mathrm{n}(\mu,\tau^2) = \mathrm{n}(\delta(\mathbf{x}),v^2)\times\mathrm{n}(\mu,\tau^2+\sigma^2/n),$$

where the marginal distribution of $\bar{X}$ is $\mathrm{n}(\mu,\tau^2+\sigma^2/n)$ and the posterior distribution of $\theta|\mathbf{x}$ is $\mathrm{n}(\delta(\mathbf{x}),v^2)$. This also completes part (c).

7.23 Let $t = s^2$ and $\theta = \sigma^2$. Because $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, we have

$$f(t|\theta) = \frac{1}{\Gamma\left((n-1)/2\right)2^{(n-1)/2}}\left(\frac{n-1}{\theta}t\right)^{[(n-1)/2]-1}e^{-(n-1)t/2\theta}\frac{n-1}{\theta}.$$

With $\pi(\theta)$ as given, we have (ignoring terms that do not depend on $\theta$)

$$\begin{aligned}\pi(\theta|t) &\propto \left[\left(\frac{1}{\theta}\right)^{((n-1)/2)-1}e^{-(n-1)t/2\theta}\frac{1}{\theta}\right]\left[\frac{1}{\theta^{\alpha+1}}e^{-1/\beta\theta}\right]\\ &\propto \left(\frac{1}{\theta}\right)^{((n-1)/2)+\alpha+1}\exp\left\{-\frac{1}{\theta}\left[\frac{(n-1)t}{2}+\frac{1}{\beta}\right]\right\},\end{aligned}$$

which we recognize as the kernel of an inverted gamma pdf, IG$(a,b)$, with

$$a = \frac{n-1}{2} + \alpha \qquad \text{and} \qquad b = \left[\frac{(n-1)t}{2} + \frac{1}{\beta}\right]^{-1}.$$

Direct calculation shows that the mean of an IG$(a,b)$ is $1/((a-1)b)$, so

$$\mathrm{E}(\theta|t) = \frac{\frac{n-1}{2}t + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1} = \frac{\frac{n-1}{2}s^2 + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1}.$$

This is a Bayes estimator of $\sigma^2$.

7.24 For $n$ observations, $Y = \sum_i X_i \sim \text{Poisson}(n\lambda)$.

a. The marginal pmf of $Y$ is

$$\begin{aligned} m(y) &= \int_0^\infty \frac{(n\lambda)^y e^{-n\lambda}}{y!} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}\, d\lambda \\ &= \frac{n^y}{y!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{(y+\alpha)-1} e^{-\frac{\lambda}{\beta/(n\beta+1)}}\, d\lambda \;=\; \frac{n^y}{y!\Gamma(\alpha)\beta^\alpha} \Gamma(y+\alpha) \left(\frac{\beta}{n\beta+1}\right)^{y+\alpha}. \end{aligned}$$

Thus,

$$\pi(\lambda|y) = \frac{f(y|\lambda)\pi(\lambda)}{m(y)} = \frac{\lambda^{(y+\alpha)-1} e^{-\frac{\lambda}{\beta/(n\beta+1)}}}{\Gamma(y+\alpha)\left(\frac{\beta}{n\beta+1}\right)^{y+\alpha}} \sim \text{gamma}\left(y+\alpha, \frac{\beta}{n\beta+1}\right).$$

b.

$$\begin{aligned} \mathrm{E}(\lambda|y) &= (y+\alpha)\frac{\beta}{n\beta+1} \;=\; \frac{\beta}{n\beta+1}y + \frac{1}{n\beta+1}(\alpha\beta). \\ \mathrm{Var}(\lambda|y) &= (y+\alpha)\frac{\beta^2}{(n\beta+1)^2}. \end{aligned}$$

7.25 a. We will use the results and notation from part (b) to do this special case. From part (b), the $X_i$s are independent and each $X_i$ has marginal pdf

$$m(x|\mu,\sigma^2,\tau^2) = \int_{-\infty}^{\infty} f(x|\theta,\sigma^2)\pi(\theta|\mu,\tau^2)\, d\theta = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma\tau} e^{-(x-\theta)^2/2\sigma^2} e^{-(\theta-\mu)^2/2\tau^2}\, d\theta.$$

Complete the square in $\theta$ to write the sum of the two exponents as

$$-\frac{\left(\theta - \left[\frac{x\tau^2}{\sigma^2+\tau^2} + \frac{\mu\sigma^2}{\sigma^2+\tau^2}\right]\right)^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} - \frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}.$$

Only the first term involves $\theta$; call it $-A(\theta)$. Also, $e^{-A(\theta)}$ is the kernel of a normal pdf. Thus,

$$\int_{-\infty}^{\infty} e^{-A(\theta)}\, d\theta = \sqrt{2\pi}\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}},$$

and the marginal pdf is

$$\begin{aligned} m(x|\mu,\sigma^2,\tau^2) &= \frac{1}{2\pi\sigma\tau}\sqrt{2\pi}\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}} \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+\tau^2}} \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\right\}, \end{aligned}$$

a n$(\mu,\sigma^2+\tau^2)$ pdf.

b. For one observation of $X$ and $\theta$ the joint pdf is

$$h(x,\theta|\tau) = f(x|\theta)\pi(\theta|\tau),$$

and the marginal pdf of $X$ is

$$m(x|\tau) = \int_{-\infty}^{\infty} h(x,\theta|\tau)\,d\theta.$$

Thus, the joint pdf of $\mathbf{X} = (X_1,\ldots,X_n)$ and $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$ is

$$h(\mathbf{x},\boldsymbol{\theta}|\tau) = \prod_i h(x_i,\theta_i|\tau),$$

and the marginal pdf of $\mathbf{X}$ is

$$\begin{aligned} m(\mathbf{x}|\tau) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_i h(x_i,\theta_i|\tau)\,d\theta_1\ldots d\theta_n \\ &= \int_{-\infty}^{\infty}\cdots\left\{\int_{-\infty}^{\infty} h(x_1,\theta_1|\tau)\,d\theta_1\right\}\prod_{i=2}^{n} h(x_i,\theta_i|\tau)\,d\theta_2\ldots d\theta_n. \end{aligned}$$

The $d\theta_1$ integral is just $m(x_1|\tau)$, and this is not a function of $\theta_2,\ldots,\theta_n$. So, $m(x_1|\tau)$ can be pulled out of the integrals. Doing each integral in turn yields the marginal pdf

$$m(\mathbf{x}|\tau) = \prod_i m(x_i|\tau).$$

Because this marginal pdf factors, this shows that marginally $X_1,\ldots,X_n$ are independent, and they each have the same marginal distribution, $m(x|\tau)$.

7.26 First write

$$f(x_1,\ldots,x_n|\theta)\pi(\theta) \propto e^{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2-|\theta|/a}$$

where the exponent can be written

$$\frac{n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{|\theta|}{a} = \frac{n}{2\sigma^2}\left(\theta-\delta_{\pm}(\mathbf{x})\right) + \frac{n}{2\sigma^2}\left(\bar{x}^2-\delta_{\pm}^2(\mathbf{x})\right)$$

with $\delta_{\pm}(\mathbf{x}) = \bar{x} \pm \frac{\sigma^2}{na}$, where we use the "+" if $\theta > 0$ and the "−" if $\theta < 0$. Thus, the posterior mean is

$$\mathrm{E}(\theta|\mathbf{x}) = \frac{\int_{-\infty}^{\infty}\theta e^{-\frac{n}{2\sigma^2}(\theta-\delta_{\pm}(\mathbf{x}))^2}\,d\theta}{\int_{-\infty}^{\infty} e^{-\frac{n}{2\sigma^2}(\theta-\delta_{\pm}(\mathbf{x}))^2}\,d\theta}.$$

Now use the facts that for constants $a$ and $b$,

$$\begin{aligned} \int_0^{\infty} e^{-\frac{a}{2}(t-b)^2}\,dt &= \int_{-\infty}^{0} e^{-\frac{a}{2}(t-b)^2}\,dt = \sqrt{\frac{\pi}{2a}}, \\ \int_0^{\infty} te^{-\frac{a}{2}(t-b)^2}\,dt &= \int_0^{\infty}(t-b)e^{-\frac{a}{2}(t-b)^2}\,dt + \int_0^{\infty} be^{-\frac{a}{2}(t-b)^2}\,dt = \frac{1}{a}e^{-\frac{a}{2}b^2} + b\sqrt{\frac{\pi}{2a}}, \\ \int_{-\infty}^{0} te^{-\frac{a}{2}(t-b)^2}\,dt &= -\frac{1}{a}e^{-\frac{a}{2}b^2} + b\sqrt{\frac{\pi}{2a}}, \end{aligned}$$

to get

$$\mathrm{E}(\theta|\mathbf{x}) = \frac{\sqrt{\frac{\pi\sigma^2}{2n}}\left(\delta_-(\mathbf{x})+\delta_+(\mathbf{x})\right) + \frac{\sigma^2}{n}\left(e^{-\frac{n}{2\sigma^2}\delta_+^2(\mathbf{x})} - e^{-\frac{n}{2\sigma^2}\delta_-^2(\mathbf{x})}\right)}{2\sqrt{\frac{\pi\sigma^2}{2n}}}.$$

7.27 a. The log likelihood is

$$\log L = \sum_{i=1}^{n} -\beta\tau_i + y_i \log(\beta\tau_i) - \tau_i + x_i \log(\tau_i) - \log y_i! - \log x_i!$$

and differentiation gives

$$\begin{aligned}\frac{\partial}{\partial\beta}\log L &= \sum_{i=1}^{n} -\tau_i + \frac{y_i\tau_i}{\beta\tau_i} \quad\Rightarrow\quad \beta = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \tau_i}\\ \frac{\partial}{\partial\tau_j}\log L &= -\beta + \frac{y_j\beta}{\beta\tau_j} - i + \frac{x_j}{\tau_j} \quad\Rightarrow\quad \tau_j = \frac{x_j + y_j}{1+\beta}\\ &\Rightarrow \sum_{j=1}^{n}\tau_j = \frac{\sum_{j=1}^n x_j + \sum_{j=1}^n y_j}{1+\beta}.\end{aligned}$$

Combining these expressions yields $\hat\beta = \sum_{j=1}^n y_j / \sum_{j=1}^n x_j$ and $\hat\tau_j = \frac{x_j+y_j}{1+\hat\beta}$.

b. The stationary point of the EM algorithm will satisfy

$$\begin{aligned}\hat\beta &= \frac{\sum_{i=1}^n y_i}{\hat\tau_1 + \sum_{i=2}^n x_i}\\ \hat\tau_1 &= \frac{\hat\tau_1 + y_1}{\hat\beta + 1}\\ \hat\tau_j &= \frac{x_j + y_j}{\hat\beta + 1}.\end{aligned}$$

The second equation yields $\tau_1 = y_1/\beta$, and substituting this into the first equation yields $\beta = \sum_{j=2}^n y_j / \sum_{j=2}^n x_j$. Summing over $j$ in the third equation, and substituting $\beta = \sum_{j=2}^n y_j / \sum_{j=2}^n x_j$ shows us that $\sum_{j=2}^n \hat\tau_j = \sum_{j=2}^n x_j$, and plugging this into the first equation gives the desired expression for $\hat\beta$. The other two equations in (7.2.16) are obviously satisfied.

c. The expression for $\hat\beta$ was derived in part (*b*), as were the expressions for $\hat\tau_i$.

7.29 a. The joint density is the product of the individual densities.

b. The log likelihood is

$$\log L = \sum_{i=1}^{n} -m\beta\tau_i + y_i \log(m\beta\tau_i) + x_i \log(\tau_i) + \log m! - \log y_i! - \log x_i!$$

and

$$\begin{aligned}\frac{\partial}{\partial\beta}\log L &= 0 \quad\Rightarrow\quad \beta = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m\tau_i}\\ \frac{\partial}{\partial\tau_j}\log L &= 0 \quad\Rightarrow\quad \tau_j = \frac{x_j + y_j}{m\beta}.\end{aligned}$$

Since $\sum \tau_j = 1$, $\hat\beta = \sum_{i=1}^n y_i/m = \sum_{i=1}^n y_i / \sum_{i=1}^n x_i$. Also, $\sum_j \tau_j = \sum_j (y_j + x_j) = 1$, which implies that $m\beta = \sum_j (y_j + x_j)$ and $\hat\tau_j = (x_j + y_j)/\sum_i (y_i + x_i)$.

c. In the likelihood function we can ignore the factorial terms, and the expected complete-data likelihood is obtained by on the $r^{th}$ iteration by replacing $x_1$ with $\mathrm{E}(X_1|\hat\tau_1^{(r)}) = m\hat\tau_1^{(r)}$. Substituting this into the MLEs of part (*b*) gives the EM sequence.

The MLEs from the full data set are $\hat{\beta} = 0.0008413892$ and

$$\begin{aligned}\hat{\tau} &= (0.06337310, 0.06374873, 0.06689681, 0.04981487, 0.04604075, 0.04883109,\\ &\quad 0.07072460, 0.01776164, 0.03416388, 0.01695673, 0.02098127, 0.01878119,\\ &\quad 0.05621836, 0.09818091, 0.09945087, 0.05267677, 0.08896918, 0.08642925).\end{aligned}$$

The MLEs for the incomplete data were computed using $R$, where we take $m = \sum x_i$. The $R$ code is

```
#mles on the incomplete data#
xdatam<-c(3560,3739,2784,2571,2729,3952,993,1908,948,1172,
    1047,3138,5485,5554,2943,4969,4828)
ydata<-c(3,4,1,1,3,1,2,0,2,0,1,3,5,4,6,2,5,4)
xdata<-c(mean(xdatam),xdatam); for (j in 1:500) {
xdata<-c(sum(xdata)*tau[1],xdatam) beta<-sum(ydata)/sum(xdata)
tau<-c((xdata+ydata)/(sum(xdata)+sum(ydata))) } beta tau
```

The MLEs from the incomplete data set are $\hat{\beta} = 0.0008415534$ and

$$\begin{aligned}\hat{\tau} &= (0.06319044, 0.06376116, 0.06690986, 0.04982459, 0.04604973, 0.04884062,\\ &\quad 0.07073839, 0.01776510, 0.03417054, 0.01696004, 0.02098536, 0.01878485,\\ &\quad 0.05622933, 0.09820005, 0.09947027, 0.05268704, 0.08898653, 0.08644610).\end{aligned}$$

7.31 a. By direct substitution we can write

$$\log L(\theta|\mathbf{y}) = \mathrm{E}\left[\log L(\theta|\mathbf{y},\mathbf{X})|\,\hat{\theta}^{(r)},\mathbf{y}\right] - \mathrm{E}\left[\log k(\mathbf{X}|\theta,\mathbf{y})|\,\hat{\theta}^{(r)},\mathbf{y}\right].$$

The next iterate, $\hat{\theta}^{(r+1)}$ is obtained by maximizing the expected complete-data log likelihood, so for any $\theta$, $\mathrm{E}\left[\log L(\hat{\theta}^{(r+1)}\mathbf{y},\mathbf{X})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right] \geq \mathrm{E}\left[\log L(\theta|\mathbf{y},\mathbf{X})|\,\hat{\theta}^{(r)},\mathbf{y}\right]$

b. Write

$$\mathrm{E}\left[\log k(\mathbf{X}|\theta,\mathbf{y})|\theta',\mathbf{y}\right] = \int \log k(\mathbf{x}|\theta,\mathbf{y}) \log k(\mathbf{x}|\theta',\mathbf{y})d\mathbf{x} \leq \int \log k(\mathbf{x}|\theta',\mathbf{y}) \log k(\mathbf{x}|\theta',\mathbf{y})d\mathbf{x},$$

from the hint. Hence $\mathrm{E}\left[\log k(\mathbf{X}|\hat{\theta}^{(r+1)},\mathbf{y})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right] \leq \mathrm{E}\left[\log k(\mathbf{X}|\hat{\theta}^{(r)},\mathbf{y})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right]$, and so the entire right hand side in part $(a)$ is decreasing.

7.33 Substitute $\alpha = \beta = \sqrt{n/4}$ into $\mathrm{MSE}(\hat{p}_B) = \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2$ and simplify to obtain

$$\mathrm{MSE}(\hat{p}_B) = \frac{n}{4(\sqrt{n}+n)^2},$$

independent of $p$, as desired.

7.35 a.

$$\begin{aligned}\delta_p\left(g(\mathbf{x})\right) &= \delta_p(x_1+a,\ldots,x_n+a)\\ &= \frac{\int_{-\infty}^{\infty} t\prod_i f(x_i+a-t)\,dt}{\int_{-\infty}^{\infty}\prod_i f(x_i+a-t)\,dt} = \frac{\int_{-\infty}^{\infty}(y+a)\prod_i f(x_i-y)\,dy}{\int_{-\infty}^{\infty}\prod_i f(x_i-y)\,dy} \qquad (y=t-a)\\ &= a+\delta_p(\mathbf{x}) = \bar{g}\left(\delta_p(\mathbf{x})\right).\end{aligned}$$

b.

$$\prod_i f(x_i - t) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\Sigma_i (x_i - t)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}n(\bar{x}-t)^2} e^{-\frac{1}{2}(n-1)s^2},$$

so

$$\delta_p(\mathbf{x}) = \frac{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} t e^{-\frac{1}{2}n(\bar{x}-t)^2} dt}{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} e^{-\frac{1}{2}n(\bar{x}-t)^2} dt} = \frac{\bar{x}}{1} = \bar{x}.$$

c.

$$\prod_i f(x_i - t) = \prod_i I\left(t - \frac{1}{2} \le x_i \le t + \frac{1}{2}\right) = I\left(x_{(n)} - \frac{1}{2} \le t \le x_{(1)} + \frac{1}{2}\right),$$

so

$$\delta_p(\mathbf{x}) = \frac{\int_{x_{(n)}+1/2}^{x_{(1)}+1/2} t\, dt}{\int_{x_{(n)}+1/2}^{x_{(1)}+1/2} 1\, dt} = \frac{x_{(1)} + x_{(n)}}{2}.$$

7.37 To find a best unbiased estimator of $\theta$, first find a complete sufficient statistic. The joint pdf is

$$f(\mathbf{x}|\theta) = \left(\frac{1}{2\theta}\right)^n \prod_i I_{(-\theta,\theta)}(x_i) = \left(\frac{1}{2\theta}\right)^n I_{[0,\theta)}(\max_i |x_i|).$$

By the Factorization Theorem, $\max_i |X_i|$ is a sufficient statistic. To check that it is a complete sufficient statistic, let $Y = \max_i |X_i|$. Note that the pdf of $Y$ is $f_Y(y) = ny^{n-1}/\theta^n$, $0 < y < \theta$. Suppose $g(y)$ is a function such that

$$\mathrm{E}\, g(Y) = \int_0^\theta \frac{ny^{n-1}}{\theta^n} g(y)\, dy = 0, \text{ for all } \theta.$$

Taking derivatives shows that $\theta^{n-1} g(\theta) = 0$, for all $\theta$. So $g(\theta) = 0$, for all $\theta$, and $Y = \max_i |X_i|$ is a complete sufficient statistic. Now

$$\mathrm{E}\, Y = \int_0^\theta y \frac{ny^{n-1}}{\theta^n}\, dy = \frac{n}{n+1}\theta \quad \Rightarrow \quad \mathrm{E}\left(\frac{n+1}{n} Y\right) = \theta.$$

Therefore $\frac{n+1}{n}\max_i |X_i|$ is a best unbiased estimator for $\theta$ because it is a function of a complete sufficient statistic. (Note that $\left(X_{(1)}, X_{(n)}\right)$ is not a minimal sufficient statistic (recall Exercise 5.36). It is for $\theta < X_i < 2\theta$, $-2\theta < X_i < \theta$, $4\theta < X_i < 6\theta$, etc., but not when the range is symmetric about zero. Then $\max_i |X_i|$ is minimal sufficient.)

7.38 Use Corollary 7.3.15.

a.

$$\begin{aligned} \frac{\partial}{\partial\theta} \log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial\theta} \log \prod_i \theta x_i^{\theta-1} = \frac{\partial}{\partial\theta} \sum_i [\log\theta + (\theta-1)\log x_i] \\ &= \sum_i \left[\frac{1}{\theta} + \log x_i\right] = -n\left[-\sum_i \frac{\log x_i}{n} - \frac{1}{\theta}\right]. \end{aligned}$$

Thus, $-\sum_i \log X_i/n$ is the UMVUE of $1/\theta$ and attains the Cramér-Rao bound.

b.

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial\theta}\log\prod_i \frac{\log\theta}{\theta-1}\theta^{x_i} = \frac{\partial}{\partial\theta}\sum_i\left[\log\log\theta - \log(\theta-1) + x_i\log\theta\right] \\ &= \sum_i\left(\frac{1}{\theta\log\theta}-\frac{1}{\theta-1}\right)+\frac{1}{\theta}\sum_i x_i = \frac{n}{\theta\log\theta}-\frac{n}{\theta-1}+\frac{n\bar{x}}{\theta} \\ &= \frac{n}{\theta}\left[\bar{x}-\left(\frac{\theta}{\theta-1}-\frac{1}{\log\theta}\right)\right].\end{aligned}$$

Thus, $\bar{X}$ is the UMVUE of $\frac{\theta}{\theta-1}-\frac{1}{\log\theta}$ and attains the Cramér-Rao lower bound.

Note: We claim that if $\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{X}) = a(\theta)[W(\mathbf{X}) - \tau(\theta)]$, then $\mathrm{E}\,W(\mathbf{X}) = \tau(\theta)$, because under the condition of the Cramér-Rao Theorem, $\mathrm{E}\,\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) = 0$. To be rigorous, we need to check the "interchange differentiation and integration" condition. Both (a) and (b) are exponential families, and this condition is satisfied for all exponential families.

7.39

$$\begin{aligned}\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X}|\theta)\right] &= E_\theta\left[\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)\right] \\ &= \mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right)\right] = \mathrm{E}_\theta\left[\frac{\frac{\partial^2}{\partial\theta^2}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}-\left(\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right)^2\right].\end{aligned}$$

Now consider the first term:

$$\begin{aligned}\mathrm{E}_\theta\left[\frac{\frac{\partial^2}{\partial\theta^2}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right] &= \int\left[\frac{\partial^2}{\partial\theta^2}f(\mathbf{x}|\theta)\right]d\mathbf{x} = \frac{d}{d\theta}\int\frac{\partial}{\partial\theta}f(\mathbf{x}|\theta)\,d\mathbf{x} \qquad \text{(assumption)} \\ &= \frac{d}{d\theta}\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right] = 0, \qquad (7.3.8)\end{aligned}$$

and the identity is proved.

7.40

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial p}\log\prod_i p^{x_i}(1-p)^{1-x_i} = \frac{\partial}{\partial p}\sum_i x_i\log p + (1-x_i)\log(1-p) \\ &= \sum_i\left[\frac{x_i}{p}-\frac{(1-x_i)}{1-p}\right] = \frac{n\bar{x}}{p}-\frac{n-n\bar{x}}{1-p} = \frac{n}{p(1-p)}[\bar{x}-p].\end{aligned}$$

By Corollary 7.3.15, $\bar{X}$ is the UMVUE of $p$ and attains the Cramér-Rao lower bound. Alternatively, we could calculate

$$\begin{aligned}&-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right) \\ &= -n\mathrm{E}\left(\frac{\partial^2}{\partial p^2}\log\left[p^X(1-p)^{1-X}\right]\right) = -n\mathrm{E}\left(\frac{\partial^2}{\partial p^2}\left[X\log p + (1-X)\log(1-p)\right]\right) \\ &= -n\mathrm{E}\left(\frac{\partial}{\partial p}\left[\frac{X}{p}-\frac{(1-X)}{1-p}\right]\right) = -n\mathrm{E}\left(\frac{-X}{p^2}-\frac{1-X}{(1-p)^2}\right) \\ &= -n\left(-\frac{1}{p}-\frac{1}{1-p}\right) = \frac{n}{p(1-p)}.\end{aligned}$$

Then using $\tau(\theta) = p$ and $\tau'(\theta) = 1$,

$$\frac{\tau'(\theta)}{-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} = \frac{1}{n/p(1-p)} = \frac{p(1-p)}{n} = \mathrm{Var}\bar{X}.$$

We know that $\mathrm{E}\bar{X} = p$. Thus, $\bar{X}$ attains the Cramér-Rao bound.

7.41 a. $\mathrm{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathrm{E}\, X_i = \sum_i a_i\mu = \mu\sum_i a_i = \mu$. Hence the estimator is unbiased.

b. $\mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathrm{Var}\, X_i = \sum_i a_i^2\sigma^2 = \sigma^2\sum_i a_i^2$. Therefore, we need to minimize $\sum_i a_i^2$, subject to the constraint $\sum_i a_i = 1$. Add and subtract the mean of the $a_i$, $1/n$, to get

$$\sum_i a_i^2 = \sum_i\left[\left(a_i - \frac{1}{n}\right) + \frac{1}{n}\right]^2 = \sum_i\left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n},$$

because the cross-term is zero. Hence, $\sum_i a_i^2$ is minimized by choosing $a_i = 1/n$ for all $i$. Thus, $\sum_i (1/n)X_i = \bar{X}$ has the minimum variance among all linear unbiased estimators.

7.43 a. This one is real hard - it was taken from an *American Statistician* article, but the proof is not there. A cryptic version of the proof is in Tukey (Approximate Weights, *Ann. Math. Statist.* 1948, 91-92); here is a more detailed version.

Let $q_i = q_i^*(1+\lambda t_i)$ with $0 \le \lambda \le 1$ and $|t_i| \le 1$. Recall that $q_i^* = (1/\sigma_i^2)/\sum_j(1/\sigma_j^2)$ and $\mathrm{Var}W^* = 1/\sum_j(1/\sigma_j^2)$. Then

$$\begin{aligned}\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) &= \frac{1}{(\sum_j q_j)^2}\sum_i q_i\sigma_i^2 \\ &= \frac{1}{[\sum_j q_j^*(1+\lambda t_j)]^2}\sum_i q_i^{*2}(1+\lambda t_i)^2\sigma_i^2 \\ &= \frac{1}{[\sum_j q_j^*(1+\lambda t_j)]^2\sum_j(1/\sigma_j^2)}\sum_i q_i^*(1+\lambda t_i)^2,\end{aligned}$$

using the definition of $q_i^*$. Now write

$$\sum_i q_i^*(1+\lambda t_i)^2 = 1 + 2\lambda\sum_j q_j t_j + \lambda^2\sum_j q_j t_j^2 = [1+\lambda\sum_j q_j t_j]^2 + \lambda^2[\sum_j q_j t_j^2 - (\sum_j q_j t_j)^2],$$

where we used the fact that $\sum_j q_j^* = 1$. Now since

$$[\sum_j q_j^*(1+\lambda t_j)]^2 = [1+\lambda\sum_j q_j t_j]^2,$$

$$\begin{aligned}\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) &= \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[\sum_j q_j t_j^2 - (\sum_j q_j t_j)^2]}{[1+\lambda\sum_j q_j t_j]^2}\right] \\ &\le \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[1 - (\sum_j q_j t_j)^2]}{[1+\lambda\sum_j q_j t_j]^2}\right],\end{aligned}$$

since $\sum_j q_j t_j^2 \le 1$. Now let $T = \sum_j q_j t_j$, and

$$\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) \le \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[1-T^2]}{[1+\lambda T]^2}\right],$$

and the right hand side is maximized at $T = -\lambda$, with maximizing value

$$\operatorname{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) \le \frac{1}{\sum_j (1/\sigma_j^2)}\left[1 + \frac{\lambda^2[1-\lambda^2]}{[1-\lambda^2]^2}\right] = \operatorname{Var} W^* \frac{1}{1-\lambda^2}.$$

Bloch and Moses (1988) define $\lambda$ as the solution to

$$b_{\max}/b_{\min} = \frac{1+\lambda}{1-\lambda},$$

where $b_i/b_j$ are the ratio of the normalized weights which, in the present notation, is

$$b_i/b_j = (1+\lambda t_i)/(1+\lambda t_j).$$

The right hand side is maximized by taking $t_i$ as large as possible and $t_j$ as small as possible, and setting $t_i = 1$ and $t_j = -1$ (the extremes) yields the Bloch and Moses (1988) solution.

b.

$$b_i = \frac{1/k}{(1/\sigma_i^2)\Big/\left(\sum_j 1/\sigma_j^2\right)} = \frac{\sigma_i^2}{k}\sum_j 1/\sigma_j^2.$$

Thus,

$$b_{\max} = \frac{\sigma_{\max}^2}{k}\sum_j 1/\sigma_j^2 \qquad \text{and} \qquad b_{\min} = \frac{\sigma_{\min}^2}{k}\sum_j 1/\sigma_j^2$$

and $B = b_{\max}/b_{\min} = \sigma_{\max}^2/\sigma_{\min}^2$. Solving $B = (1+\lambda)/(1-\lambda)$ yields $\lambda = (B-1)/(B+1)$. Substituting this into Tukey's inequality yields

$$\frac{\operatorname{Var} W}{\operatorname{Var} W^*} \le \frac{(B+1)^2}{4B} = \frac{((\sigma_{\max}^2/\sigma_{\min}^2)+1)^2}{4(\sigma_{\max}^2/\sigma_{\min}^2)}.$$

7.44 $\sum_i X_i$ is a complete sufficient statistic for $\theta$ when $X_i \sim \mathrm{n}(\theta, 1)$. $\bar{X}^2 - 1/n$ is a function of $\sum_i X_i$. Therefore, by Theorem 7.3.23, $\bar{X}^2 - 1/n$ is the unique best unbiased estimator of its expectation.

$$\mathrm{E}\left(\bar{X}^2 - \frac{1}{n}\right) = \operatorname{Var}\bar{X} + (\mathrm{E}\,\bar{X})^2 - \frac{1}{n} = \frac{1}{n} + \theta^2 - \frac{1}{n} = \theta^2.$$

Therefore, $\bar{X}^2 - 1/n$ is the UMVUE of $\theta^2$. We will calculate

$$\operatorname{Var}\left(\bar{X}^2 - 1/n\right) = \operatorname{Var}(\bar{X}^2) = \mathrm{E}(\bar{X}^4) - [\mathrm{E}(\bar{X}^2)]^2, \quad \text{where } \bar{X} \sim \mathrm{n}(\theta, 1/n),$$

but first we derive some general formulas that will also be useful in later exercises. Let $Y \sim \mathrm{n}(\theta, \sigma^2)$. Then here are formulas for $\mathrm{E}\,Y^4$ and $\operatorname{Var} Y^2$.

$$\begin{aligned}
\mathrm{E}\,Y^4 &= \mathrm{E}[Y^3(Y-\theta+\theta)] = \mathrm{E}\,Y^3(Y-\theta) + \mathrm{E}\,Y^3\theta = E\,Y^3(Y-\theta) + \theta\mathrm{E}\,Y^3. \\
\mathrm{E}\,Y^3(Y-\theta) &= \sigma^2\mathrm{E}(3Y^2) = \sigma^2 3\left(\sigma^2+\theta^2\right) = 3\sigma^4 + 3\theta^2\sigma^2. \qquad \text{(Stein's Lemma)} \\
\theta\mathrm{E}\,Y^3 &= \theta\left(3\theta\sigma^2 + \theta^3\right) = 3\theta^2\sigma^2 + \theta^4. \qquad \text{(Example 3.6.6)} \\
\operatorname{Var} Y^2 &= 3\sigma^4 + 6\theta^2\sigma^2 + \theta^4 - (\sigma^2+\theta^2)^2 = 2\sigma^4 + 4\theta^2\sigma^2.
\end{aligned}$$

Thus,

$$\operatorname{Var}\left(\bar{X}^2 - \frac{1}{n}\right) = \operatorname{Var}\bar{X}^2 = 2\frac{1}{n^2} + 4\theta^2\frac{1}{n} > \frac{4\theta^2}{n}.$$

To calculate the Cramér-Rao lower bound, we have

$$\begin{aligned}\mathrm{E}_\theta\left(\frac{\partial^2 \log f(X|\theta)}{\partial\theta^2}\right) &= \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\frac{1}{\sqrt{2\pi}}e^{-(X-\theta)^2/2}\right)\\ &= \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\log(2\pi)^{-1/2}-\frac{1}{2}(X-\theta)^2\right]\right) = \mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}(X-\theta)\right) = -1,\end{aligned}$$

and $\tau(\theta)=\theta^2$, $[\tau'(\theta)]^2=(2\theta)^2=4\theta^2$ so the Cramér-Rao Lower Bound for estimating $\theta^2$ is

$$\frac{[\tau'(\theta)]^2}{-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} = \frac{4\theta^2}{n}.$$

Thus, the UMVUE of $\theta^2$ does not attain the Cramér-Rao bound. (However, the ratio of the variance and the lower bound $\to 1$ as $n\to\infty$.)

7.45 a. Because $\mathrm{E}\,S^2=\sigma^2$, $\mathrm{bias}(aS^2)=\mathrm{E}(aS^2)-\sigma^2=(a-1)\sigma^2$. Hence,

$$\mathrm{MSE}(aS^2)=\mathrm{Var}(aS^2)+\mathrm{bias}(aS^2)^2=a^2\mathrm{Var}(S^2)+(a-1)^2\sigma^4.$$

b. There were two typos in early printings; $\kappa=\mathrm{E}[X-\mu]^4/\sigma^4$ and

$$\mathrm{Var}(S^2)=\frac{1}{n}\left(\kappa-\frac{n-3}{n-1}\right)\sigma^4.$$

See Exercise 5.8b for the proof.

c. There was a typo in early printings; under normality $\kappa=3$. Under normality we have

$$\kappa=\frac{\mathrm{E}[X-\mu]^4}{\sigma^4}=\mathrm{E}\left[\frac{X-\mu}{\sigma}\right]^4=\mathrm{E}\,Z^4,$$

where $Z\sim \mathrm{n}(0,1)$. Now, using Lemma 3.6.5 with $g(z)=z^3$ we have

$$\kappa=\mathrm{E}\,Z^4=\mathrm{E}\,g(Z)Z=1\mathrm{E}(3Z^2)=3\mathrm{E}\,Z^2=3.$$

To minimize $\mathrm{MSE}(S^2)$ in general, write $\mathrm{Var}(S^2)=B\sigma^4$. Then minimizing $\mathrm{MSE}(S^2)$ is equivalent to minimizing $a^2B+(a-1)^2$. Set the derivative of this equal to 0 ($B$ is not a function of $a$) to obtain the minimizing value of $a$ is $1/(B+1)$. Using the expression in part (b), under normality the minimizing value of $a$ is

$$\frac{1}{B+1}=\frac{1}{\frac{1}{n}\left(3-\frac{n-3}{n-1}\right)+1}=\frac{n-1}{n+1}.$$

d. There was a typo in early printings; the minimizing $a$ is

$$a=\frac{n-1}{(n+1)+\frac{(\kappa-3)(n-1)}{n}}.$$

To obtain this simply calculate $1/(B+1)$ with (from part (b))

$$B=\frac{1}{n}\left(\kappa-\frac{n-3}{n-1}\right).$$

e. Using the expression for $a$ in part (d), if $\kappa = 3$ the second term in the denominator is zero and $a = (n-1)/(n+1)$, the normal result from part (c). If $\kappa < 3$, the second term in the denominator is negative. Because we are dividing by a smaller value, we have $a > (n-1)/(n+1)$. Because $\text{Var}(S^2) = B\sigma^4$, $B > 0$, and, hence, $a = 1/(B+1) < 1$. Similarly, if $\kappa > 3$, the second term in the denominator is positive. Because we are dividing by a larger value, we have $a < (n-1)/(n+1)$.

7.46 a. For the uniform$(\theta, 2\theta)$ distribution we have $\mathrm{E}\,X = (2\theta+\theta)/2 = 3\theta/2$. So we solve $3\theta/2 = \bar{X}$ for $\theta$ to obtain the method of moments estimator $\tilde{\theta} = 2\bar{X}/3$.

b. Let $x_{(1)}, \ldots, x_{(n)}$ denote the observed order statistics. Then, the likelihood function is

$$L(\theta|\mathbf{x}) = \frac{1}{\theta^n} I_{[x_{(n)}/2, x_{(1)}]}(\theta).$$

Because $1/\theta^n$ is decreasing, this is maximized at $\hat{\theta} = x_{(n)}/2$. So $\hat{\theta} = X_{(n)}/2$ is the MLE. Use the pdf of $X_{(n)}$ to calculate $\mathrm{E}\,X_{(n)} = \frac{2n+1}{n+1}\theta$. So $\mathrm{E}\,\hat{\theta} = \frac{2n+1}{2n+2}\theta$, and if $k = (2n+2)/(2n+1)$, $\mathrm{E}\,k\hat{\theta} = \theta$.

c. From Exercise 6.23, a minimal sufficient statistic for $\theta$ is $(X_{(1)}, X_{(n)})$. $\tilde{\theta}$ is not a function of this minimal sufficient statistic. So by the Rao-Blackwell Theorem, $\mathrm{E}(\tilde{\theta}|X_{(1)}, X_{(n)})$ is an unbiased estimator of $\theta$ ($\tilde{\theta}$ is unbiased) with smaller variance than $\tilde{\theta}$. The MLE is a function of $(X_{(1)}, X_{(n)})$, so it can not be improved with the Rao-Blackwell Theorem.

d. $\tilde{\theta} = 2(1.16)/3 = .7733$ and $\hat{\theta} = 1.33/2 = .6650$.

7.47 $X_i \sim \mathrm{n}(r, \sigma^2)$, so $\bar{X} \sim \mathrm{n}(r, \sigma^2/n)$ and $\mathrm{E}\,\bar{X}^2 = r^2 + \sigma^2/n$. Thus $\mathrm{E}\left[(\pi\bar{X}^2 - \pi\sigma^2/n)\right] = \pi r^2$ is best unbiased because $\bar{X}$ is a complete sufficient statistic. If $\sigma^2$ is unknown replace it with $s^2$ and the conclusion still holds.

7.48 a. The Cramér-Rao Lower Bound for unbiased estimates of $p$ is

$$\frac{\left[\frac{d}{dp}p\right]^2}{-n\mathrm{E}\frac{d^2}{dp^2}\log L(p|X)} = \frac{1}{-n\mathrm{E}\left\{\frac{d^2}{dp^2}\log[p^X(1-p)^{1-X}]\right\}} = \frac{1}{-n\mathrm{E}\left\{-\frac{X}{p^2} - \frac{(1-X)}{(1-p)^2}\right\}} = \frac{p(1-p)}{n},$$

because $\mathrm{E}\,X = p$. The MLE of $p$ is $\hat{p} = \sum_i X_i/n$, with $\mathrm{E}\,\hat{p} = p$ and $\text{Var}\,\hat{p} = p(1-p)/n$. Thus $\hat{p}$ attains the CRLB and is the best unbiased estimator of $p$.

b. By independence, $\mathrm{E}(X_1X_2X_3X_4) = \prod_i \mathrm{E}\,X_i = p^4$, so the estimator is unbiased. Because $\sum_i X_i$ is a complete sufficient statistic, Theorems 7.3.17 and 7.3.23 imply that $\mathrm{E}(X_1X_2X_3X_4|\sum_i X_i)$ is the best unbiased estimator of $p^4$. Evaluating this yields

$$\begin{aligned}\mathrm{E}\left(X_1X_2X_3X_4\middle|\sum_i X_i = t\right) &= \frac{P(X_1 = X_2 = X_3 = X_4 = 1, \sum_{i=5}^n X_i = t-4)}{P(\sum_i X_i = t)}\\ &= \frac{p^4\binom{n-4}{t-4}p^{t-4}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \binom{n-4}{t-4}\Big/\binom{n}{t},\end{aligned}$$

for $t \geq 4$. For $t < 4$ one of the $X_i$s must be zero, so the estimator is $\mathrm{E}(X_1X_2X_3X_4|\sum_i X_i = t) = 0$.

7.49 a. From Theorem 5.5.9, $Y = X_{(1)}$ has pdf

$$f_Y(y) = \frac{n!}{(n-1)!}\frac{1}{\lambda}e^{-y/\lambda}\left[1-(1-e^{-y/\lambda})\right]^{n-1} = \frac{n}{\lambda}e^{-ny/\lambda}.$$

Thus $Y \sim \text{exponential}(\lambda/n)$ so $\mathrm{E}\,Y = \lambda/n$ and $nY$ is an unbiased estimator of $\lambda$.

b. Because $f_X(x)$ is in the exponential family, $\sum_i X_i$ is a complete sufficient statistic and $\mathrm{E}\left(nX_{(1)}|\sum_i X_i\right)$ is the best unbiased estimator of $\lambda$. Because $\mathrm{E}\left(\sum_i X_i\right) = n\lambda$, we must have $\mathrm{E}\left(nX_{(1)}|\sum_i X_i\right) = \sum_i X_i/n$ by completeness. Of course, any function of $\sum_i X_i$ that is an unbiased estimator of $\lambda$ is the best unbiased estimator of $\lambda$. Thus, we know directly that because $\mathrm{E}(\sum_i X_i) = n\lambda$, $\sum_i X_i/n$ is the best unbiased estimator of $\lambda$.

c. From part (a), $\hat{\lambda} = 601.2$ and from part (b) $\hat{\lambda} = 128.8$. Maybe the exponential model is not a good assumption.

7.50 a. $\mathrm{E}(a\bar{X} + (1-a)cS) = a\mathrm{E}\,\bar{X} + (1-a)\mathrm{E}(cS) = a\theta + (1-a)\theta = \theta$. So $a\bar{X} + (1-a)cS$ is an unbiased estimator of $\theta$.

b. Because $\bar{X}$ and $S^2$ are independent for this normal model, $\mathrm{Var}(a\bar{X}+(1-a)cS) = a^2V_1+(1-a)^2V_2$, where $V_1 = \mathrm{Var}\bar{X} = \theta^2/n$ and $V_2 = \mathrm{Var}(cS) = c^2\mathrm{E}\,S^2 - \theta^2 = c^2\theta^2 - \theta^2 = (c^2-1)\theta^2$. Use calculus to show that this quadratic function of $a$ is minimized at

$$a = \frac{V_2}{V_1+V_2} = \frac{(c^2-1)\theta^2}{((1/n)+c^2-1)\theta^2} = \frac{(c^2-1)}{((1/n)+c^2-1)}.$$

c. Use the factorization in Example 6.2.9, with the special values $\mu = \theta$ and $\sigma^2 = \theta^2$, to show that $(\bar{X}, S^2)$ is sufficient. $\mathrm{E}(\bar{X} - cS) = \theta - \theta = 0$, for all $\theta$. So $\bar{X} - cS$ is a nonzero function of $(\bar{X}, S^2)$ whose expected value is always zero. Thus $(\bar{X}, S^2)$ is not complete.

7.51 a. Straightforward calculation gives:

$$\mathrm{E}\left[\theta - (a_1\bar{X} + a_2cS)\right]^2 = a_1^2\mathrm{Var}\,\bar{X} + a_2^2c^2\mathrm{Var}\,S + \theta^2(a_1+a_2-1)^2.$$

Because $\mathrm{Var}\,\bar{X} = \theta^2/n$ and $\mathrm{Var}\,S = \mathrm{E}\,S^2 - (\mathrm{E}\,S)^2 = \theta^2\left(\frac{c^2-1}{c^2}\right)$, we have

$$\mathrm{E}\left[\theta - (a_1\bar{X} + a_2cS)\right]^2 = \theta^2\left[a_1^2\Big/n + a_2^2(c^2-1) + (a_1+a_2-1)^2\right],$$

and we only need minimize the expression in square brackets, which is independent of $\theta$. Differentiating yields $a_2 = \left[(n+1)c^2 - n\right]^{-1}$ and $a_1 = 1 - \left[(n+1)c^2 - n\right]^{-1}$.

b. The estimator $T^*$ has minimum MSE over a class of estimators that contain those in Exercise 7.50.

c. Because $\theta > 0$, restricting $T^* \geq 0$ will improve the MSE.

d. No. It does not fit the definition of either one.

7.52 a. Because the Poisson family is an exponential family with $t(x) = x$, $\sum_i X_i$ is a complete sufficient statistic. Any function of $\sum_i X_i$ that is an unbiased estimator of $\lambda$ is the unique best unbiased estimator of $\lambda$. Because $\bar{X}$ is a function of $\sum_i X_i$ and $\mathrm{E}\,\bar{X} = \lambda$, $\bar{X}$ is the best unbiased estimator of $\lambda$.

b. $S^2$ is an unbiased estimator of the population variance, that is, $\mathrm{E}\,S^2 = \lambda$. $\bar{X}$ is a one-to-one function of $\sum_i X_i$. So $\bar{X}$ is also a complete sufficient statistic. Thus, $\mathrm{E}(S^2|\bar{X})$ is an unbiased estimator of $\lambda$ and, by Theorem 7.3.23, it is also the unique best unbiased estimator of $\lambda$. Therefore $\mathrm{E}(S^2|\bar{X}) = \bar{X}$. Then we have

$$\mathrm{Var}\,S^2 = \mathrm{Var}\left(\mathrm{E}(S^2|\bar{X})\right) + \mathrm{E}\,\mathrm{Var}(S^2|\bar{X}) = \mathrm{Var}\,\bar{X} + \mathrm{E}\,\mathrm{Var}(S^2|\bar{X}),$$

so $\mathrm{Var}\,S^2 > \mathrm{Var}\,\bar{X}$.

c. We formulate a general theorem. Let $T(X)$ be a complete sufficient statistic, and let $T'(X)$ be any statistic other than $T(X)$ such that $\mathrm{E}\,T(X) = \mathrm{E}\,T'(X)$. Then $\mathrm{E}[T'(X)|T(X)] = T(X)$ and $\mathrm{Var}\,T'(X) > \mathrm{Var}\,T(X)$.

7.53 Let $a$ be a constant and suppose $\text{Cov}_{\theta_0}(W,U) > 0$. Then

$$\text{Var}_{\theta_0}(W + aU) = \text{Var}_{\theta_0} W + a^2 \text{Var}_{\theta_0} U + 2a\text{Cov}_{\theta_0}(W,U).$$

Choose $a \in \left(-2\text{Cov}_{\theta_0}(W,U)\Big/\text{Var}_{\theta_0}U, 0\right)$. Then $\text{Var}_{\theta_0}(W + aU) < \text{Var}_{\theta_0} W$, so $W$ cannot be best unbiased.

7.55 All three parts can be solved by this general method. Suppose $X \sim f(x|\theta) = c(\theta)m(x)$, $a < x < \theta$. Then $1/c(\theta) = \int_a^\theta m(x)\,dx$, and the cdf of $X$ is $F(x) = c(\theta)/c(x)$, $a < x < \theta$. Let $Y = X_{(n)}$ be the largest order statistic. Arguing as in Example 6.2.23 we see that $Y$ is a complete sufficient statistic. Thus, any function $T(Y)$ that is an unbiased estimator of $h(\theta)$ is the best unbiased estimator of $h(\theta)$. By Theorem 5.4.4 the pdf of $Y$ is $g(y|\theta) = nm(y)c(\theta)^n/c(y)^{n-1}$, $a < y < \theta$. Consider the equations

$$\int_a^\theta f(x|\theta)\,dx = 1 \qquad \text{and} \qquad \int_a^\theta T(y)g(y|\theta)\,dy = h(\theta),$$

which are equivalent to

$$\int_a^\theta m(x)\,dx = \frac{1}{c(\theta)} \qquad \text{and} \qquad \int_a^\theta \frac{T(y)nm(y)}{c(y)^{n-1}}\,dy = \frac{h(\theta)}{c(\theta)^n}.$$

Differentiating both sides of these two equations with respect to $\theta$ and using the Fundamental Theorem of Calculus yields

$$m(\theta) = -\frac{c'(\theta)}{c(\theta)^2} \qquad \text{and} \qquad \frac{T(\theta)nm(\theta)}{c(\theta)^{n-1}} = \frac{c(\theta)^n h'(\theta) - h(\theta)nc(\theta)^{n-1}c'(\theta)}{c(\theta)^{2n}}.$$

Change $\theta$s to $y$s and solve these two equations for $T(y)$ to get the best unbiased estimator of $h(\theta)$ is

$$T(y) = h(y) + \frac{h'(y)}{nm(y)c(y)}.$$

For $h(\theta) = \theta^r$, $h'(\theta) = r\theta^{r-1}$.

a. For this pdf, $m(x) = 1$ and $c(\theta) = 1/\theta$. Hence

$$T(y) = y^r + \frac{ry^{r-1}}{n(1/y)} = \frac{n+r}{n}y^r.$$

b. If $\theta$ is the lower endpoint of the support, the smallest order statistic $Y = X_{(1)}$ is a complete sufficient statistic. Arguing as above yields the best unbiased estimator of $h(\theta)$ is

$$T(y) = h(y) - \frac{h'(y)}{nm(y)c(y)}.$$

For this pdf, $m(x) = e^{-x}$ and $c(\theta) = e^\theta$. Hence

$$T(y) = y^r - \frac{ry^{r-1}}{ne^{-y}e^y} = y^r - \frac{ry^{r-1}}{n}.$$

c. For this pdf, $m(x) = e^{-x}$ and $c(\theta) = 1/(e^{-\theta} - e^{-b})$. Hence

$$T(y) = y^r - \frac{ry^{r-1}}{ne^{-y}}(e^{-y} - e^{-b}) = y^r - \frac{ry^{r-1}(1 - e^{-(b-y)})}{n}.$$

7.56 Because $T$ is sufficient, $\phi(T) = \mathrm{E}[h(X_1, \ldots, X_n)|T]$ is a function only of $T$. That is, $\phi(T)$ is an estimator. If $\mathrm{E}\, h(X_1, \ldots, X_n) = \tau(\theta)$, then

$$\mathrm{E}\, h(X_1, \cdots, X_n) = \mathrm{E}\left[\mathrm{E}\left(h(X_1, \ldots, X_n)|\, T\right)\right] = \tau(\theta),$$

so $\phi(T)$ is an unbiased estimator of $\tau(\theta)$. By Theorem 7.3.23, $\phi(T)$ is the best unbiased estimator of $\tau(\theta)$.

7.57 a. $T$ is a Bernoulli random variable. Hence,

$$\mathrm{E}_p T = P_p(T = 1) = P_p\left(\sum_{i=1}^{n} X_i > X_{n+1}\right) = h(p).$$

b. $\sum_{i=1}^{n+1} X_i$ is a complete sufficient statistic for $\theta$, so $\mathrm{E}\left(T \middle| \sum_{i=1}^{n+1} X_i\right)$ is the best unbiased estimator of $h(p)$. We have

$$\begin{aligned} \mathrm{E}\left(T \middle| \sum_{i=1}^{n+1} X_i = y\right) &= P\left(\sum_{i=1}^{n} X_i > X_{n+1} \middle| \sum_{i=1}^{n+1} X_i = y\right) \\ &= P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y\right) \Big/ P\left(\sum_{i=1}^{n+1} X_i = y\right). \end{aligned}$$

The denominator equals $\binom{n+1}{y} p^y (1-p)^{n+1-y}$. If $y = 0$ the numerator is

$$P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = 0\right) = 0.$$

If $y > 0$ the numerator is

$$P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 0\right) + P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 1\right)$$

which equals

$$P\left(\sum_{i=1}^{n} X_i > 0, \sum_{i=1}^{n} X_i = y\right) P(X_{n+1} = 0) + P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) P(X_{n+1} = 1).$$

For all $y > 0$,

$$P\left(\sum_{i=1}^{n} X_i > 0, \sum_{i=1}^{n} X_i = y\right) = P\left(\sum_{i=1}^{n} X_i = y\right) = \binom{n}{y} p^y (1-p)^{n-y}.$$

If $y = 1$ or $2$, then

$$P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) = 0.$$

And if $y > 2$, then

$$P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) = P\left(\sum_{i=1}^{n} X_i = y - 1\right) = \binom{n}{y-1} p^{y-1} (1-p)^{n-y+1}.$$

Therefore, the UMVUE is

$$\mathrm{E}\left(T\left|\sum_{i=1}^{n+1} X_i = y\right.\right) = \begin{cases} 0 & \text{if } y=0 \\ \frac{\binom{n}{y}p^y(1-p)^{n-y}(1-p)}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}}{\binom{n+1}{y}} = \frac{1}{(n+1)(n+1-y)} & \text{if } y=1 \text{ or } 2 \\ \frac{\left(\binom{n}{y}+\binom{n}{y-1}\right)p^y(1-p)^{n-y+1}}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}+\binom{n}{y-1}}{\binom{n+1}{y}} = 1 & \text{if } y>2. \end{cases}$$

7.59 We know $T = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Then

$$\mathrm{E}\,T^{p/2} = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\int_0^\infty t^{\frac{p+n-1}{2}-1}e^{-\frac{t}{2}}\,dt = \frac{2^{\frac{p}{2}}\Gamma\left(\frac{p+n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} = C_{p,n}.$$

Thus

$$\mathrm{E}\left(\frac{(n-1)S^2}{\sigma^2}\right)^{p/2} = C_{p,n},$$

so $(n-1)^{p/2}S^p\big/C_{p,n}$ is an unbiased estimator of $\sigma^p$. From Theorem 6.2.25, $(\bar{X}, S^2)$ is a complete, sufficient statistic. The unbiased estimator $(n-1)^{p/2}S^p\big/C_{p,n}$ is a function of $(\bar{X}, S^2)$. Hence, it is the best unbiased estimator.

7.61 The pdf for $Y \sim \chi^2_\nu$ is

$$f(y) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}y^{\nu/2-1}e^{-y/2}.$$

Thus the pdf for $S^2 = \sigma^2 Y/\nu$ is

$$g(s^2) = \frac{\nu}{\sigma^2}\frac{1}{\Gamma(\nu/2)2^{\nu/2}}\left(\frac{s^2\nu}{\sigma^2}\right)^{\nu/2-1}e^{-s^2\nu/(2\sigma^2)}.$$

Thus, the log-likelihood has the form (gathering together constants that do not depend on $s^2$ or $\sigma^2$)

$$\log L(\sigma^2|s^2) = \log\left(\frac{1}{\sigma^2}\right) + K\log\left(\frac{s^2}{\sigma^2}\right) - K'\frac{s^2}{\sigma^2} + K'',$$

where $K > 0$ and $K' > 0$.

The loss function in Example 7.3.27 is

$$L(\sigma^2, a) = \frac{a}{\sigma^2} - \log\left(\frac{a}{\sigma^2}\right) - 1,$$

so the loss of an estimator is the negative of its likelihood.

7.63 Let $a = \tau^2/(\tau^2+1)$, so the Bayes estimator is $\delta^\pi(x) = ax$. Then $R(\mu, \delta^\pi) = (a-1)^2\mu^2 + a^2$. As $\tau^2$ increases, $R(\mu, \delta^\pi)$ becomes flatter.

7.65 a. Figure omitted.

b. The posterior expected loss is $\mathrm{E}\left(L(\theta,a)|x\right) = e^{ca}\mathrm{E}\,e^{-c\theta} - c\mathrm{E}(a-\theta) - 1$, where the expectation is with respect to $\pi(\theta|x)$. Then

$$\frac{d}{da}\mathrm{E}\left(L(\theta,a)|x\right) = ce^{ca}\mathrm{E}\,e^{-c\theta} - c \stackrel{\text{set}}{=} 0,$$

and $a = -\frac{1}{c}\log \mathrm{E}\,e^{-c\theta}$ is the solution. The second derivative is positive, so this is the minimum.

c. $\pi(\theta|x) = \mathrm{n}(\bar{x}, \sigma^2/n)$. So, substituting into the formula for a normal mgf, we find $\mathrm{E}\, e^{-c\theta} = e^{-c\bar{x}+\sigma^2 c^2/2n}$, and the LINEX posterior loss is

$$\mathrm{E}\left(L(\theta,a)|x\right) = e^{c(a-\bar{x})+\sigma^2c^2/2n} - c(a-\bar{x}) - 1.$$

Substitute $\mathrm{E}\, e^{-c\theta} = e^{-c\bar{x}+\sigma^2c^2/2n}$ into the formula in part (b) to find the Bayes rule is $\bar{x} - c\sigma^2/2n$.

d. For an estimator $\bar{X}+b$, the LINEX posterior loss (from part (c)) is

$$\mathrm{E}\left(L(\theta,\bar{x}+b)|x\right) = e^{cb}e^{c^2\sigma^2/2n} - cb - 1.$$

For $\bar{X}$ the expected loss is $e^{c^2\sigma^2/2n} - 1$, and for the Bayes estimator ($b = -c\sigma^2/2n$) the expected loss is $c^2\sigma^2/2n$. The marginal distribution of $\bar{X}$ is $m(\bar{x}) = 1$, so the Bayes risk is infinite for any estimator of the form $\bar{X}+b$.

e. For $\bar{X}+b$, the squared error risk is $\mathrm{E}\left[(\bar{X}+b)-\theta\right]^2 = \sigma^2/n + b^2$, so $\bar{X}$ is better than the Bayes estimator. The Bayes risk is infinite for both estimators.

7.66 Let $S = \sum_i X_i \sim \text{binomial}(n,\theta)$.

a. $\mathrm{E}\,\hat{\theta}^2 = \mathrm{E}\frac{S^2}{n^2} = \frac{1}{n^2}\mathrm{E}\, S^2 = \frac{1}{n^2}(n\theta(1-\theta)+(n\theta)^2) = \frac{\theta}{n} + \frac{n-1}{n}\theta^2$.

b. $T_n^{(i)} = \left(\sum_{j\neq i} X_j\right)^2 \Big/ (n-1)^2$. For $S$ values of $i$, $T_n^{(i)} = (S-1)^2/(n-1)^2$ because the $X_i$ that is dropped out equals 1. For the other $n-S$ values of $i$, $T_n^{(i)} = S^2/(n-1)^2$ because the $X_i$ that is dropped out equals 0. Thus we can write the estimator as

$$\mathrm{JK}(T_n) = n\frac{S^2}{n^2} - \frac{n-1}{n}\left(S\frac{(S-1)^2}{(n-1)^2} + (n-S)\frac{S^2}{(n-1)^2}\right) = \frac{S^2-S}{n(n-1)}.$$

c. $\mathrm{E}\,\mathrm{JK}(T_n) = \frac{1}{n(n-1)}(n\theta(1-\theta)+(n\theta)^2 - n\theta) = \frac{n^2\theta^2 - n\theta^2}{n(n-1)} = \theta^2$.

d. For this binomial model, $S$ is a complete sufficient statistic. Because $\mathrm{JK}(T_n)$ is a function of $S$ that is an unbiased estimator of $\theta^2$, it is the best unbiased estimator of $\theta^2$.