6.32 In the Formal Likelihood Principle, take $E = E_1 = E_2 = E$. Then the conclusion is $Ev(E, x_1) = Ev(E, x_2)$ if $L(\theta|x_1)/L(\theta|x_2) = c$. Thus evidence is equal whenever the likelihood functions are equal, and this follows from Formal Sufficiency and Conditionality.

6.33 a. For all sample points except $(2, x_2^*)$ (but including $(1, x_1^*)$), $T(j, x_j) = (j, x_j)$. Hence,

$$g(T(j, x_j)|\theta)h(j, x_j) = g((j, x_j)|\theta)1 = f^*((j, x_j)|\theta).$$

For $(2, x_2^*)$ we also have

$$g(T(2, x_2^*)|\theta)h(2, x_2^*) = g((1, x_1^*)|\theta)C = f^*((1, x_1^*)|\theta)C = C\frac{1}{2}f_1(x_1^*|\theta)$$

$$= C\frac{1}{2}L(\theta|x_1^*) = \frac{1}{2}L(\theta|x_2^*) = \frac{1}{2}f_2(x_2^*|\theta) = f^*((2, x_2^*)|\theta).$$

By the Factorization Theorem, $T(J, X, J)$ is sufficient.

b. Equations 6.3.4 and 6.3.5 follow immediately from the two Principles. Combining them we have $Ev(E_1, x_1^*) = Ev(E_2, x_2^*)$, the conclusion of the Formal Likelihood Principle.

c. To prove the Conditionality Principle. Let one experiment be the $E^*$ experiment and the other $E_\beta$. Then

$$L(\theta|j, x_j) = f^*((j, x_j)|\theta) = \frac{1}{2}f_j(x_j|\theta) = \frac{1}{2}L(\theta|x_j).$$

Letting $(j, x_j)$ play the roles of $x_1^*$ and $x_2^*$ in the Formal Likelihood Principle we can conclude $Ev(E^*, (j, x_j)) = Ev(E_\beta, x_j)$, the Conditionality Principle. Now consider the Formal Sufficiency Principle. If $T(X)$ is sufficient and $T(x) = T(y)$, then $L(\theta|x) = CL(\theta|y)$, where $C = h(x)/h(y)$ and $h$ is the function from the Factorization Theorem. Hence, by the Formal Likelihood Principle, $Ev(E, x) = Ev(E, y)$, the Formal Sufficiency Principle.

6.35 Let 1 = success and 0 = failure. The four sample points are \{0, 10, 110, 111\}. From the likelihood principle, inference about $p$ is only through $L(p|x)$. The values of the likelihood are 1, $p$, $p^2$, and $p^3$, and the sample size does not directly influence the inference.

6.37 a. For one observation $(X, Y)$ we have

$$I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log f(X, Y|\theta)\right) = -E\left(-\frac{2Y}{\theta^2}\right) = \frac{2E Y}{\theta^2}.$$  But, $Y \sim \text{exponential}(\theta)$, and $E Y = \theta$. Hence, $I(\theta) = 2/\theta^2$ for a sample of size one, and $I(\theta) = 2n/\theta^2$ for a sample of size $n$.

b. (i) The cdf of $T$ is

$$P(T \leq t) = P \left(\frac{\sum Y_i}{\sum X_i} \leq t^2\right) = P \left(\frac{2\sum Y_i/\theta}{2\sum X_i/\theta} \leq t^2/\theta^2\right) = P(F_{2n, 2n} \leq t^2/\theta^2)$$
where $F_{2n,2n}$ is an $F$ random variable with $2n$ degrees of freedom in the numerator and denominator. This follows since $2Y_i/\theta$ and $2X_i\theta$ are all independent exponential(1), or $\chi^2_2$. Differentiating (in $t$) and simplifying gives the density of $T$ as
\begin{equation}
 f_T(t) = \frac{\Gamma(2n)}{\Gamma(n)^2} \frac{t^2}{(t^2 + \theta^2)} \left( \frac{\theta^2}{t^2 + \theta^2} \right)^n,
\end{equation}
and the second derivative (in $\theta$) of the log density is
\begin{equation}
 2n \frac{t^4 + 2t^2\theta^2 - \theta^4}{\theta^2(t^2 + \theta^2)^2} = 2n \frac{1}{\theta^2} \left( 1 - \frac{2}{(t^2/\theta^2 + 1)^2} \right),
\end{equation}
and the information in $T$ is
\begin{equation}
 \frac{2n}{\theta^2} \left[ 1 - 2E \left( \frac{1}{T^2/\theta^2 + 1} \right)^2 \right] = \frac{2n}{\theta^2} \left[ 1 - 2E \left( \frac{1}{F_{2n,2n}^2 + 1} \right)^2 \right].
\end{equation}
The expected value is
\begin{equation}
 E \left( \frac{1}{F_{2n,2n}^2 + 1} \right)^2 = \frac{\Gamma(2n)}{\Gamma(n)^2} \int_0^\infty \frac{1}{(1 + w)^2(1 + w)^{2n}} \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n + 2)} = \frac{n + 1}{2(2n + 1)}.
\end{equation}
Substituting this above gives the information in $T$ as
\begin{equation}
 \frac{2n}{\theta^2} \left[ 1 - 2 \frac{n + 1}{2(2n + 1)} \right] = I(\theta) \frac{n}{2n + 1},
\end{equation}
which is not the answer reported by Joshi and Nabar.

(ii) Let $W = \sum_i X_i$ and $V = \sum_i Y_i$. In each pair, $X_i$ and $Y_i$ are independent, so $W$ and $V$ are independent. $X_i \sim \text{exponential}(1/\theta)$; hence, $W \sim \text{gamma}(n, 1/\theta)$. $Y_i \sim \text{exponential}(\theta)$; hence, $V \sim \text{gamma}(n, \theta)$. Use this joint distribution of $(W, V)$ to derive the joint pdf of $(T, U)$ as
\begin{equation}
 f(t, u|\theta) = \frac{2}{\Gamma(n)^2} u^{2n-1} \exp \left( -\frac{u\theta}{t} - \frac{ut}{\theta} \right), \quad u > 0, \ t > 0.
\end{equation}
Now, the information in $(T, U)$ is
\begin{equation}
 -E \left( \frac{\partial^2}{\partial \theta^2} \log f(T, U|\theta) \right) = -E \left( -\frac{2UU}{\theta^3} \right) = E \left( \frac{2V}{\theta^3} \right) = \frac{2n\theta}{\theta^3} = \frac{2n}{\theta^2}.
\end{equation}

(iii) The pdf of the sample is $f(x, y) = \exp \left[ -\theta \left( \sum x_i - \sum y_i \right) / \theta \right]$. Hence, $(W, V)$ defined as in part (ii) is sufficient. $(T, U)$ is a one-to-one function of $(W, V)$, hence $(T, U)$ is also sufficient. But, $E U^2 = E W V = (n/\theta)(n\theta) = n^2$ does not depend on $\theta$. So $E(U^2 - n^2) = 0$ for all $\theta$, and $(T, U)$ is not complete.

6.39 a. The transformation from Celsius to Fahrenheit is $y = 9x/5 + 32$. Hence,
\begin{align*}
\frac{5}{9}(T^*(y) - 32) &= \frac{5}{9} \left( (.5)(y) + (.5)(212) - 32 \right) \\
&= \frac{5}{9} \left( (.5)(9x/5 + 32) + (.5)(212) - 32 \right) = (.5)x + 50 = T(x).
\end{align*}
b. $T(x) = (.5)x + 50 \neq (.5)x + 106 = T^*(x)$. Thus, we do not have equivariance.
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.

\[\begin{array}{cccccc}
x & 0 & 1 & 2 & 3 & 4 \\
\hat{\theta} & 1 & 1 & 2 & \text{or} & 3 \\
\end{array}\]

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.

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

\[
\log L(\beta|x) = -\log \Gamma(\alpha)^{n} - n\alpha \log \beta + (\alpha-1) \log \left[ \prod_{i=1}^{n} x_{i} \right] - \frac{\sum_{i=1}^{n} x_{i}}{\beta}
\]

\[
\frac{\partial \log L}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{\sum_{i=1}^{n} x_{i}}{\beta^2}
\]

Set the partial derivative equal to 0 and solve for $\beta$ to obtain $\hat{\beta} = \frac{\sum_{i=1}^{n} x_{i}}{(n\alpha)}$. To check that this is a maximum, calculate

\[
\frac{\partial^2 \log L}{\partial \beta^2} \bigg|_{\beta = \hat{\beta}} = \frac{n\alpha}{\beta^2} - 2 \left( \frac{\sum_{i=1}^{n} x_{i}}{\beta^3} \right) \bigg|_{\beta = \hat{\beta}} = \frac{(n\alpha)^2}{(\sum_{i=1}^{n} x_{i})^2} - \frac{2(n\alpha)^2}{(\sum_{i=1}^{n} x_{i})^2} = \frac{(n\alpha)^2}{(\sum_{i=1}^{n} x_{i})^2} < 0.
\]

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

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

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 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 $\sum_{i=1}^{n} x_{i}/(n\alpha)$. Substitute this into $L$. Then we just need to maximize the function of the one variable $\alpha$ given by

\[
\frac{1}{\Gamma(\alpha)^{n}(\sum_{i=1}^{n} x_{i}/(n\alpha))^{n\alpha}} \left[ \prod_{i=1}^{n} x_{i} \right]^{\alpha-1} e^{-\sum_{i=1}^{n} x_{i}/(\sum_{i=1}^{n} x_{i}/(n\alpha))}
\]

\[
= \frac{1}{\Gamma(\alpha)^{n}(\sum_{i=1}^{n} x_{i}/(n\alpha))^{n\alpha}} \left[ \prod_{i=1}^{n} x_{i} \right]^{\alpha-1} e^{-n\alpha}.
\]
For the given data, \( n = 14 \) and \( \sum x_i = 323.6 \). Many computer programs can be used to maximize this function. From PROC NLIN in SAS we obtain \( \hat{\alpha} = 514.219 \) and, hence, \( \hat{\beta} = \frac{323.6}{514.219} = .0450 \).

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

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

\[
(1 - p)^n = \prod_i (1 - x_i),
\]

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

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

Because the right-hand side of the first equation is decreasing in \( \hat{z} \), and because \( \hat{k} \leq 1/\hat{z} \) (so \( \hat{z} \leq 1/\hat{k} \)) 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 \( (\frac{1}{2})^4 = (1 - 2x)(1 - z)(1 - 19z) \), 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. \( f(x|\theta) = \prod_i \frac{\theta x_i^{-2} I(\theta, \infty)(x_i)}{\theta^\alpha I(\theta, \infty)(x_i)} \). Thus, \( X_{(1)} \) is a sufficient statistic for \( \theta \) by the Factorization Theorem.

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

c. \( E X = \int_0^\infty x \theta x^{-1} dx = \theta \log x \bigg|_0^\infty = \infty \). Thus the method of moments estimator of \( \theta \) does not exist. (This is the Pareto distribution with \( \alpha = \theta, \beta = 1 \).)

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

7.8 a. \( E X^2 = Var X + \mu^2 = \sigma^2. \) Therefore \( X^2 \) is an unbiased estimator of \( \sigma^2 \).

b. \[
L(\sigma|x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/(2\sigma^2)}, \quad \log L(\sigma|x) = \log(2\pi)^{-1/2} - \log \sigma - x^2/(2\sigma^2).
\]

\[
\frac{\partial \log L}{\partial \sigma} = -\frac{x^2}{\sigma^3} \Rightarrow \hat{\sigma} = \sqrt{X^2} = |X|.
\]

\[
\frac{\partial^2 \log L}{\partial \sigma^2} = -\frac{3x^2}{\sigma^4} + \frac{1}{\sigma^2}, \text{ which is negative at } \hat{\sigma} = |x|.
\]

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 \( E X = 0 \) is known, just equate \( E X^2 = \sigma^2 = \frac{1}{n} \sum X_i^2 = X^2 \Rightarrow \hat{\sigma} = |X| \).

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

\[
E \hat{\theta} = 2\bar{X} \Rightarrow \bar{X} = \frac{\theta}{2} = \theta, \quad \text{and} \quad Var \hat{\theta} = 4Var \bar{X} = 4 \frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.
\]
The likelihood function is
\[ L(\theta|x) = \prod_{i=1}^{n} \frac{1}{\theta} f_{[0,\theta]}(x_i) = \frac{1}{\theta^n} f_{[0,\theta]}(x_{(n)}) f_{[0,\infty)}(x_{(1)}), \]
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
\[ E \hat{\theta} = \frac{n}{n+1} \theta, \quad E \hat{\theta}^2 = \frac{n}{n+2} \theta^2 \quad \text{and} \quad \text{Var} \hat{\theta} = \frac{n \theta^2}{(n+2)(n+1)^2}. \]
\( \hat{\theta} \) 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, \( \text{Var} \hat{\theta} < \text{Var} \theta \) for all \( \theta \). So, if \( n \) is large, \( \hat{\theta} \) is probably preferable to \( \theta \).

---

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

b. For any fixed \( \alpha \), \( L(\alpha, \beta|x) = 0 \) if \( \beta < x_{(n)} \), and \( L(\alpha, \beta|x) \) a decreasing function of \( \beta \) if \( \beta \geq x_{(n)} \). Thus, \( X_{(n)} \) is the MLE of \( \beta \). For the MLE of \( \alpha \) calculate
\[ \frac{\partial}{\partial \alpha} \log L = \frac{\partial}{\partial \alpha} \left[ n \log \alpha - n \log \beta + (\alpha-1) \log \prod_i x_i \right] = \frac{n}{\alpha} - n \log \beta + \log \prod_i x_i. \]
Set the derivative equal to zero and use \( \hat{\beta} = X_{(n)} \) to obtain
\[ \hat{\alpha} = \frac{n}{n \log X_{(n)} - \log \prod_i X_i} = \left[ \frac{1}{n} \sum_i (\log X_{(n)} - \log X_i) \right]^{-1}. \]
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.
\[ f(x|\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \left( \prod_{i=1}^{n} x_i \right)^{\theta-1} = L(\theta|x) \]
\[ \frac{d}{d\theta} \log L = \frac{d}{d\theta} \left[ n \log \theta + (\theta-1) \log \prod_i x_i \right] = \frac{n}{\theta} + \sum_i \log x_i. \]
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 \( -\sum_i \log X_i \sim \gamma(n, 1/\theta) \). Thus \( \hat{\theta} = n/T, \) where \( T \sim \gamma(n, 1/\theta) \). We can either calculate the first and second moments directly, or use the fact that \( \hat{\theta} \) is inverted gamma (page 51). We have
\[ E \frac{1}{T} = \frac{\theta^n}{\Gamma(n)} \int_0^\infty \frac{1}{t} t^{n-1} e^{-\theta t} \, dt = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}. \]
\[ E \frac{1}{T^2} = \frac{\theta^n}{\Gamma(n)} \int_0^\infty \frac{1}{t^2} t^{n-1} e^{-\theta t} \, dt = \frac{\theta^n}{\Gamma(n)} \frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}. \]
and thus
\[ E\hat{\theta} = \frac{n}{n-1} \theta \quad \text{and} \quad \text{Var}\hat{\theta} = \frac{n^2}{(n-1)^2(n-2)} \theta^2 \rightarrow 0 \text{ as } n \to \infty. \]

b. Because \( X \sim \text{beta}(\theta, 1) \), \( E X = \theta/(\theta + 1) \) and the method of moments estimator is the solution to
\[ \frac{1}{n} \sum_i X_i = \frac{\theta}{\theta + 1} \Rightarrow \hat{\theta} = \frac{\sum_i X_i}{n - \sum_i X_i}. \]

7.12 \( X_i \sim \text{iid Bernoulli}(\theta) \), \( 0 \leq \theta \leq 1/2 \).

a. method of moments:
\[ EX = \theta = \frac{1}{n} \sum_i X_i = \bar{X} \quad \Rightarrow \quad \hat{\theta} = \bar{X}. \]

MLE: In Example 7.2.7, we showed that \( L(\theta|x) \) is increasing for \( \theta \leq \bar{x} \) and is decreasing for \( \theta > \bar{x} \). Remember that \( 0 \leq \theta \leq 1/2 \) in this exercise. Therefore, when \( \bar{X} \leq 1/2, \) \( \bar{X} \) is the MLE of \( \theta \), because \( \bar{X} \) is the overall maximum of \( L(\theta|x) \). When \( \bar{X} > 1/2 \), \( L(\theta|x) \) 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 \( \hat{\theta} = \min\{\bar{X}, 1/2\} \).

b. The MSE of \( \hat{\theta} \) is \( \text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2 = (\theta(1 - \theta)/n) + 0^2 = \theta(1 - \theta)/n. \) There is no simple formula for \( \text{MSE}(\hat{\theta}) \), but an expression is
\[
\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = \sum_{y=0}^{[n/2]} (\frac{y}{n} - \theta)^2 \binom{n}{y} \theta^y(1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n} \frac{1}{2} (\frac{y}{n} - \theta)^2 \binom{n}{y} \theta^y(1-\theta)^{n-y},
\]
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
\[ \text{MSE}(\hat{\theta}) = E(\bar{X} - \theta)^2 = \sum_{y=0}^{n} (\frac{y}{n} - \theta)^2 \binom{n}{y} \theta^y(1-\theta)^{n-y}. \]

Therefore,
\[
\text{MSE}(\bar{X}) - \text{MSE}(\hat{\theta}) = \sum_{y=[n/2]+1}^{n} \left( \frac{y}{n} - \theta \right)^2 \binom{n}{y} \theta^y(1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n} \left( \frac{y}{n} - \frac{1}{2} \right) \binom{n}{y} \theta^y(1-\theta)^{n-y}.
\]

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

7.13 \( L(\theta|x) = \prod_i \frac{1}{2} e^{-\frac{1}{2} |x_i - \theta|} = \frac{1}{2^n} e^{-\frac{1}{2} \sum_i |x_i - \theta|} \), 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 \( x_{(j)} \leq \theta \leq x_{(j+1)} \),
\[
\sum_{i=1}^{n} |x_{(i)} - \theta| = \sum_{i=1}^{j} (\theta - x_{(i)}) + \sum_{i=j+1}^{n} (x_{(i)} - \theta) = (2j - n) \theta - \sum_{i=1}^{j} x_{(i)} + \sum_{i=j+1}^{n} x_{(i)}.
\]
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

\[
\bar{x} = \mathbb{E}X = \mu_X, \quad \frac{1}{n} \sum_i x_i^2 = \mathbb{E}X^2 = \sigma_X^2 + \mu_X^2,
\]

\[
\bar{y} = \mathbb{E}Y = \mu_Y, \quad \frac{1}{n} \sum_i y_i^2 = \mathbb{E}Y^2 = \sigma_Y^2 + \mu_Y^2.
\]

We also need an equation involving $\rho$.

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

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

\[
\frac{1}{n} \sum_i x_i^2 - \bar{x}^2 = \frac{\sum_i x_i^2 - (\sum_i x_i)^2/n}{n} = \frac{\sum_i (x_i - \bar{x})^2}{n}
\]

are used.

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

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

the solutions to the $k$ equations $\sum_{j=1}^{n} t_i(x_j) = \mathbb{E}_\theta \left(\sum_{j=1}^{n} t_i(X_j)\right) = n\mathbb{E}_\theta t_i(X_1), i = 1, \ldots, k$, provide the unique MLE for $\theta$. 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

\[
\sum_{i} x_i = n\mu_X, \quad \sum_{i} x_i^2 = n(\mu_X^2 + \sigma_X^2),
\]

\[
\sum_{i} y_i = n\mu_Y, \quad \sum_{i} y_i^2 = n(\mu_Y^2 + \sigma_Y^2),
\]

\[
\sum_{i} x_i y_i = \sum_{i} [\text{Cov}(X, Y) + \mu_X \mu_Y] = n(\rho \sigma_X \sigma_Y + \mu_X \mu_Y).
\]

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

\[
L(\theta|x, y) = L(\theta|x)L(\theta|x|y)
\]

\[
= (2\pi \sigma_X^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_X^2} \sum_i (x_i - \mu_X)^2\right\}
\]

\[
\times (2\pi \sigma_Y^2(1-\rho^2))^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)} \sum_i (y_i - (\mu_Y + \rho \sigma_Y \sigma_X (x_i - \mu_X)))^2\right\}
\]
We know that \( \bar{x} \) and \( \sigma_X^2 = \frac{\sum(x_i - \bar{x})^2}{n} \) maximizes \( A \); the question is whether given \( \sigma_Y \), \( \mu_Y \), and \( \rho \), does \( \bar{x} \), \( \sigma_X^2 \) maximize \( B \)? Let us first fix \( \sigma_X^2 \) and look for \( \mu_X \), that maximizes \( B \). We have
\[
\frac{\partial \log B}{\partial \mu_X} \propto -2 \left( \sum_i \left[ (y_i - \mu_Y) - \frac{\rho \sigma_Y}{\sigma_X} (x_i - \mu_X) \right] \right) \frac{\rho \sigma_Y}{\sigma_X} \equiv 0
\]
\[
\Rightarrow \sum_i (y_i - \mu_Y) = \rho \frac{\sigma_Y}{\sigma_X} \sum (x_i - \mu_X).
\]
Similarly do the same procedure for \( L(\theta|y)L(\theta, y|x) \) This implies \( \sum (x_i - \mu_X) = \frac{\rho \sigma_X}{\sigma_Y} \sum (y_i - \mu_Y) \). The solutions \( \hat{\mu_X} \) and \( \hat{\mu_Y} \) therefore must satisfy both equations. If \( \sum (y_i - \mu_Y) = 0 \) or \( \sum (x_i - \mu_X) = 0 \), we will get \( \rho = 1/\rho \), so we need \( \sum_i (y_i - \mu_Y) = 0 \) and \( \sum_i (x_i - \mu_X) = 0 \). This implies \( \mu_X = \bar{x} \) and \( \mu_Y = \bar{y} \). \( \frac{\partial \log B}{\partial \mu_X} < 0 \). Therefore it is maximum. To get \( \frac{\partial \log B}{\partial \sigma_X^2} \) take
\[
\frac{\partial \log B}{\partial \sigma_X^2} \propto \sum_i \frac{\rho \sigma_Y}{\sigma_X} (x_i - \mu_X) \left[ (y_i - \mu_Y) - \frac{\rho \sigma_Y}{\sigma_X} (x_i - \mu_X) \right] \equiv 0
\]
\[
\Rightarrow \sum_i (x_i - \mu_X)(y_i - \mu_Y) = \rho \frac{\sigma_Y}{\sigma_X} \sum (x_i - \mu_X)^2.
\]
Similarly, \( \sum_i (x_i - \mu_X)(y_i - \mu_Y) = \frac{\sigma_X^2}{\sigma_Y} \sum_i (y_i - \mu_Y)^2 \). Thus \( \frac{\partial \log B}{\partial \sigma_X^2} \) and \( \frac{\partial \log B}{\partial \sigma_Y^2} \) must satisfy the above two equations with \( \hat{\mu_X} = \bar{x}, \hat{\mu_Y} = \bar{y} \). This implies
\[
\frac{\partial \sigma_Y}{\partial \mu_X} \sum_i (x_i - \bar{x})^2 = \frac{\partial \sigma_X}{\partial \mu_Y} \sum_i (y_i - \bar{y})^2 \Rightarrow \sum_i (x_i - \bar{x})^2 = \sum_i (y_i - \bar{y})^2
\]
Therefore, \( \sigma_X^2 = \alpha \sum_i (x_i - \bar{x})^2 \), \( \sigma_Y^2 = \alpha \sum_i (y_i - \bar{y})^2 \) where \( \alpha \) is a constant. Combining the knowledge that \( (\bar{x}, \bar{y}, \sum_i (x_i - \bar{x})^2) = (\hat{\mu_X}, \hat{\sigma_X}^2) \) maximizes \( A \), we conclude that \( \alpha = 1/n \).
Lastly, we find \( \hat{\rho} \), the MLE of \( \rho \). Write
\[
\log L(\bar{x}, \bar{y}, \sigma_X^2, \sigma_Y^2, \rho|x,y) = -\frac{n}{2} \log(1 - \rho^2) - \frac{1}{2(1-\rho^2)} \sum_i \left[ \frac{(x_i - \bar{x})^2}{\sigma_X^2} - \frac{2\rho(x_i - \bar{x})(y_i - \bar{y})}{\sigma_X \sigma_Y} + \frac{(y_i - \bar{y})^2}{\sigma_Y^2} \right]
\]
\[
= -\frac{n}{2} \log(1 - \rho^2) - \frac{1}{2(1-\rho^2)} \left[ 2n - 2\rho \sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_X \sigma_Y} \right]
\]
because \( \sigma_X^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2 \) and \( \sigma_Y^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2 \). Now
\[
\log L = -\frac{n}{2} \log(1 - \rho^2) - \frac{n}{1 - \rho^2} \frac{\rho A}{1 - \rho^2}
\]
and
\[
\frac{\partial \log L}{\partial \rho} = -\frac{n}{1 - \rho^2} \frac{\rho A}{(1-\rho^2)^2} + A(1-\rho^2) + 2A \rho^2 \frac{\rho A}{(1-\rho^2)^2} \equiv 0
\]
This implies
\[
\frac{A + A \rho^2 - n \rho - n \rho^3}{(1-\rho^2)^2} = 0 \Rightarrow A(1 + \rho^2) = n \rho(1 + \rho^2)
\]
\[
\Rightarrow \hat{\rho} = \frac{A}{n} = \frac{1}{n} \sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_X \sigma_Y}.
\]
7.21 a.

\[ E \frac{1}{n} \sum_i Y_i x_i = \frac{1}{n} \sum_i EY_i x_i = \frac{1}{n} \sum_i \beta x_i x_i = \beta. \]

b.

\[ \text{Var} \frac{1}{n} \sum_i Y_i x_i = \frac{1}{n^2} \sum_i \text{Var} Y_i x_i = \frac{\sigma^2}{n^2} \sum_i \frac{1}{x_i^2}. \]

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,

\[ \text{Var} \beta = \frac{\sigma^2}{\sum_i x_i^2} \leq \frac{\sigma^2}{n^2} \sum_i \frac{1}{x_i^2} = \text{Var} \frac{1}{n} \sum_i Y_i x_i. \]

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

\[ \frac{1}{x_i^2} \leq \frac{1}{n} \sum_i \frac{1}{x_i^2}. \]

Thus,

\[ \text{Var} \left( \frac{\sum_i Y_i x_i}{\sum_i x_i} \right) = \frac{\sigma^2}{n(x^2)} \leq \frac{\sigma^2}{n^2} \sum_i \frac{1}{x_i^2} = \text{Var} \frac{1}{n} \sum_i Y_i x_i. \]

7.22 a.

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

b. Factor the exponent in part (a) as

\[ -\frac{n}{2\sigma^2} (\overline{x} - \theta)^2 - \frac{1}{2\tau^2} (\overline{y} - \mu)^2 = -\frac{1}{2\nu^2} (\overline{y} - \delta(x))^2 - \frac{1}{\tau^2 + \sigma^2/n} (\overline{x} - \mu)^2, \]

where \( \delta(x) = (\tau^2 + (\sigma^2/n)\mu)/(\tau^2 + \sigma^2/n) \) and \( \nu = (\sigma^2/\tau^2/n)/\tau^2 + \sigma^2/n \). Let \( n(a, b) \) denote the pdf of a normal distribution with mean \( a \) and variance \( b \). The above factorization shows that

\[ f(x, \theta) = n(\theta, \sigma^2/n) \times n(\mu, \tau^2) = n(\delta(x), \nu^2) \times n(\mu, \tau^2 + \sigma^2/n), \]

where the marginal distribution of \( \overline{X} \) is \( n(\mu, \tau^2 + \sigma^2/n) \) and the posterior distribution of \( \theta | x \) is \( n(\delta(x), \nu^2) \). 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

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

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

\[ \pi(\theta | t) \propto \left[ \frac{1}{\theta} \right]^{(n-1)/2-1} e^{-(n-1)t/2\theta} \left[ \frac{1}{\theta^{n+1} e^{-1/\theta \theta}} \right] \]

\[ \propto \left( \frac{1}{\theta} \right)^{(n-1)/2 + \alpha + 1} \exp \left\{ -\frac{1}{\theta} \left[ \frac{(n-1)t}{2} + \frac{1}{\beta} \right] \right\}, \]
which we recognize as the kernel of an inverted gamma pdf, \( \text{IG}(a, b) \), with

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

Direct calculation shows that the mean of an \( \text{IG}(a, b) \) is \( 1/((a-1)b) \), so

\[
E(\theta|t) = \frac{n/2 + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1} = \frac{n/2 + \frac{1}{\beta}}{n/2 + \alpha - 1}.
\]

This is a Bayes estimator of \( \sigma^2 \).

---

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

a. The marginal pmf of \( Y \) is

\[
m(y) = \int_0^\infty \frac{(n \lambda)^y e^{-n \lambda}}{y!} \frac{1}{\Gamma(a) \beta^a} \lambda^{a-1} e^{-\lambda/\beta} d\lambda
\]

\[
= \frac{n^y}{y! \Gamma(a) \beta^a} \int_0^\infty \lambda^{y+a-1} e^{-\lambda/\beta \Gamma(a, n \lambda/\beta)} d\lambda
\]

\[
= \frac{n^y}{y! \Gamma(a) \beta^a} \Gamma(y + a) \left( \frac{\beta}{n \beta + 1} \right)^{y+a}.
\]

Thus,

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

b.

\[
E(\lambda|y) = (y + \alpha) \frac{\beta}{n \beta + 1} = \frac{\beta}{n \beta + 1} y + \frac{1}{n \beta + 1} (\alpha \beta).
\]

\[
\text{Var}(\lambda|y) = (y + \alpha) \left( \frac{\beta^2}{(n \beta + 1)^2} \right).
\]

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

\[
m(x|\mu, \sigma^2, \tau^2) = \int_{-\infty}^\infty f(x|\theta, \sigma^2) \pi(\theta|\mu, \tau^2) d\theta = \int_{-\infty}^\infty \frac{1}{2\pi \sigma \tau} \exp \left( -\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2} \right) d\theta.
\]

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

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

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

\[
m(x|\mu, \sigma^2, \tau^2) = \frac{1}{2\pi \sigma \tau} \sqrt{2\pi} \frac{\sigma \tau}{\sqrt{\sigma^2 + \tau^2}} \exp \left\{ -\frac{(x - \mu)^2}{2(\sigma^2 + \tau^2)} \right\}
\]

\[
= \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2 + \tau^2}} \exp \left\{ -\frac{(x - \mu)^2}{2(\sigma^2 + \tau^2)} \right\},
\]

a \( n(\mu, \sigma^2 + \tau^2) \) pdf.
b. 
\[ \prod_i f(x_i \mid t) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_i (x_i - t)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} n(t - t)^2} e^{-\frac{1}{2} (n-1) \sum_i (x_i - t)^2}, \]
so 
\[ \delta_p(x) = \frac{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} t e^{-\frac{1}{2} n(t - t)^2} dt}{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} e^{-\frac{1}{2} n(x - t)^2} dt} = \frac{x}{1} = x. \]

c. 
\[ \prod_i f(x_i \mid t) = \prod_i I \left(t - \frac{1}{2} \leq x_i \leq t + \frac{1}{2}\right) = I \left(x_{(n)} - \frac{1}{2} \leq t \leq x_{(1)} + \frac{1}{2}\right), \]
so 
\[ \delta_p(x) = \frac{\int_{x_{(n)} + 1/2}^{x_{(1)} + 1/2} t dt}{\int_{x_{(1)} + 1/2}^{x_{(1)} + 1/2} 1 dt} = \frac{x_{(1)} + x_{(n)}}{2}. \]

--- 7.37 To find a best unbiased estimator of \( \theta \), first find a complete sufficient statistic. The joint pdf is 
\[ f(x \mid \theta) = \left(\frac{1}{2\theta}\right)^n \prod_i I_{(-\infty,\theta)}(x_i) = \left(\frac{1}{2\theta}\right)^n I_{[0,\theta]}(\max_i |x_i|). \]
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 
\[ E g(Y) = \int_0^\theta \frac{ny^{n-1}}{\theta^n} g(y) dy = 0, \] 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 
\[ E Y = \int_0^\theta \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1} \theta \Rightarrow E \left( \frac{n+1}{n} Y \right) = \theta. \]
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 \( (X_{(1)}, X_{(n)}) \) 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. 
\[ \frac{\partial}{\partial \theta} \log L(\theta \mid x) = \frac{\partial}{\partial \theta} \log \prod_i \theta x_i^{\theta-1} = \frac{\partial}{\partial \theta} \sum_i \log \theta + (\theta-1) \log x_i \]
\[ = \sum_i \left[ \frac{1}{\theta} + \log x_i \right] = -n \left[ -\sum_i \frac{\log x_i - 1}{\theta} \right]. \]
Thus, \( -\sum_i \log X_i/n \) is the UMVUE of \( 1/\theta \) and attains the Cramér-Rao bound.
b. 

\[
\frac{\partial}{\partial \theta} \log L(\theta | x) = \frac{\partial}{\partial \theta} \log \prod_{i}^{l} \log_\theta \frac{\theta}{\theta - 1} = \frac{\partial}{\partial \theta} \sum_{i} \left[ \log \log \theta - \log(\theta - 1) + x_i \log \theta \right]
\]

\[= \sum_{i} \left( \frac{1}{\log \theta} - \frac{1}{\theta - 1} \right) + \frac{n}{\theta} \sum_{i} x_i = \frac{n}{\theta} - \frac{n}{\theta - 1} + \frac{n \bar{x}}{\theta} \]

\[= \frac{n}{\theta} \left[ 2 - \left( \frac{\theta}{\theta - 1} - \frac{1}{\log \theta} \right) \right].\]

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

Note: We claim that \(\frac{\partial}{\partial \theta} \log L(\theta | X) = a(\theta)[W(X) - \tau(\theta)]\), then \(E W(X) = \tau(\theta)\), because under the condition of the Cramér-Rao Theorem, \(E \frac{\partial}{\partial \theta} \log L(\theta | x) = 0\). 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

\[E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right] = E_{\theta} \left[ \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta} \log f(X|\theta) \right) \right] = E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} f(X|\theta) \right] = E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} f(X|\theta) \right] - \left( \frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2.\]

Now consider the first term:

\[E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} f(X|\theta) \right] = \int \left[ \frac{\partial^2}{\partial \theta^2} f(x|\theta) \right] dx = \frac{d}{d\theta} \int \frac{\partial}{\partial \theta} f(x|\theta) dx \quad \text{ (assumption)} \]

\[= \frac{d}{d\theta} E_{\theta} \left[ \frac{\partial}{\partial \theta} \log f(X|\theta) \right] = 0, \quad \text{(7.38)}\]

and the identity is proved.

7.40

\[\frac{\partial}{\partial \theta} \log L(\theta | x) = \frac{\partial}{\partial p} \log \prod_{i}^{l} \frac{p_{X_i}(1 - p)^{1-x_i}}{1 - p} = \frac{\partial}{\partial p} \sum_{i} x_i \log p + (1 - x_i) \log(1 - p) \]

\[= \sum_{i} \left[ \frac{x_i}{p} - \frac{(1 - x_i)}{1 - p} \right] = \frac{n \bar{x}}{p} - \frac{n - n \bar{x}}{1 - p} = \frac{n}{p(1 - p)}[\bar{x} - p].\]

By Corollary 7.3.15, \(X\) is the UMVUE of \(p\) and attains the Cramér-Rao lower bound. Alternatively, we could calculate

\[-n E_{\theta} \left( \frac{\partial^2}{\partial p^2} \log f(X|\theta) \right) = -n E_{\theta} \left( \frac{\partial^2}{\partial p^2} \log \left[ p X (1 - p)^{1 - X} \right] \right) = -n E_{\theta} \left( \frac{\partial^2}{\partial p^2} \log X \log p + (1 - X) \log(1 - p) \right) \]

\[-n E_{\theta} \left( \frac{\partial}{\partial p} \log \left[ \frac{X}{p} - \frac{1 - X}{1 - p} \right] \right) = -n E_{\theta} \left( \frac{-X}{p^2} - \frac{1 - X}{(1 - p)^2} \right) \]

\[= -n \left( \frac{1}{p} - \frac{1}{1 - p} \right) = \frac{n}{p(1 - p)}.\]
and the right hand side is maximized at $T = -\lambda$, with maximizing value

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

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

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

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

$$b_i/b_j = (1 + \lambda \ell_i)/(1 + \lambda \ell_j).$$

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.

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

Thus,

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

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

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

$\sum_i X_i$ is a complete sufficient statistic for $\theta$ when $X_i \sim n(\theta, 1)$. $\bar{X}^2 - 1/n$ 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.

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

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

$$\text{Var} (\bar{X}^2 - 1/n) = \text{Var}(\bar{X}^2) = E(\bar{X}^4) - (E(\bar{X}^2))^2,$$

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

$$E Y^4 = E[Y^3(Y - \theta + \theta)] = E Y^3(Y - \theta) + E Y^2 \theta = E Y^3(Y - \theta) + \theta E Y^3.$$  

$$E Y^3(Y - \theta) = \sigma^2 E(3Y^2) = \sigma^2 3 (\sigma^2 + \theta^2) = 3\sigma^4 + 3\theta^2 \sigma^2.$$  

$$\theta E Y^3 = \theta (3\theta \sigma^2 + \theta^3) = 3\theta^2 \sigma^2 + \theta^4.$$  

$$\text{Var} Y^2 = 3\sigma^4 + 6\theta^2 \sigma^2 + \theta^4 - (\sigma^2 + \theta^2)^2 = 2\sigma^4 + 4\theta^2 \sigma^2.$$  

Thus,

$$\text{Var} \left( \frac{\bar{X}^2 - 1}{n} \right) = \text{Var} \bar{X}^2 = 2\frac{1}{n^2} + 4\theta^2 \frac{1}{n} > \frac{4\theta^2}{n}.$$
To calculate the Cramér-Rao lower bound, we have

\[
E_\theta \left( \frac{\partial^2 \log f(X|\theta)}{\partial \theta^2} \right) = E_\theta \left( \frac{\partial^2}{\partial \theta^2} \log \frac{1}{\sqrt{2\pi}} e^{-(X-\theta)^2/2} \right) = E_\theta \left( \frac{\partial}{\partial \theta} (X-\theta) \right) = -1,
\]

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

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

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 \( E S^2 = \sigma^2 \), \( \text{bias}(aS^2) = E(aS^2) - \sigma^2 = (a - 1)\sigma^2 \). Hence,

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

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

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

See Exercise 5.8b for the proof.

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

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

where \( Z \sim N(0,1) \). Now, using Lemma 3.6.5 with \( g(z) = z^4 \) we have

\[
\kappa = E Z^4 = E g(Z) Z = 1E(3Z^2) = 3E Z^2 = 3.
\]

To minimize \( \text{MSE}(S^2) \) in general, write \( \text{Var}(S^2) = B \sigma^4 \). Then minimizing \( \text{MSE}(S^2) \) is equivalent to minimizing \( a^2 B + (a - 1)^2 \). 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

\[
a = \frac{n-1}{(n+1) + \left( \kappa - 3(n-1) \right)}.
\]

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 (e). 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 \( E X = (2\theta + \theta)/2 = 3\theta/2 \). So we solve \( 3\theta/2 = \bar{X} \) for \( \theta \) to obtain the method of moments estimator \( \hat{\theta} = 2\bar{X}/3 \).

b. Let \( x_1, \ldots, x_n \) denote the observed order statistics. Then, the likelihood function is

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

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 \( E X_{(n)} = \frac{2n+1}{n+1} \theta \). So \( E \hat{\theta} = \frac{2n+1}{2n+2} \theta \), and if \( k = (2n + 2)/(2n + 1) \),

\[
E (k \hat{\theta}) = \theta.
\]

c. From Exercise 6.23, a minimal sufficient statistic for \( \theta \) is \((X_1, X_{(n)}) \). \( \hat{\theta} \) is not a function of this minimal sufficient statistic. So by the Rao-Blackwell Theorem, \( E(\hat{\theta}|X_1, X_{(n)}) \) is an unbiased estimator of \( \theta \) (\( \hat{\theta} \) is unbiased) with smaller variance than \( \hat{\theta} \). The MLE is a function of \((X_1, X_{(n)}) \), so it can not be improved with the Rao-Blackwell Theorem.

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

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

7.48 a. The Cramer-Rao Lower Bound for unbiased estimates of \( p \) is

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

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

d. By independence, \( E(X_1X_2X_3X_4) = \prod E X_i = p^4 \), so the estimator is unbiased. Because \( \sum X_i \) is a complete sufficient statistic, Theorems 7.3.17 and 7.3.23 imply that \( E(X_1X_2X_3X_4|\sum X_i) \) is the best unbiased estimator of \( p^4 \). Evaluating this yields

\[
E \left( X_1X_2X_3X_4 | \sum X_i = t \right) = \frac{P(X_1 = X_2 = X_3 = X_4 = 1, \sum_{i=5}^{n} X_i = t - 4)}{P(\sum X_i = t)} = \frac{p^4(n-4)!p^{t-4}(1-p)^{n-t}}{(n-4)!p^t(1-p)^{n-t}} = \frac{(n-4)!}{(t-4)!} \frac{1}{n},
\]

for \( t \geq 4 \). For \( t < 4 \) one of the \( X_i \)'s must be zero, so the estimator is \( E(X_1X_2X_3X_4 | \sum X_i = t) = 0 \).

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 \( EY = \lambda/n \) and \( nY \) is an unbiased estimator of \( \lambda \).
b. Because \( f_X(x) \) is in the exponential family, \( \sum X_i \) is a complete sufficient statistic and 
\( E(nX_i|\sum X_i) \) is the best unbiased estimator of \( \lambda \). Because \( E(\sum X_i) = n\lambda \), we must have 
\( E(nX_i|\sum X_i) = \sum X_i/n \) by completeness. Of course, any function of \( \sum X_i \) that 
is an unbiased estimator of \( \lambda \) is the best unbiased estimator of \( \lambda \). Thus, we know directly 
that because \( E(\sum X_i) = n\lambda \), \( \sum X_i/n \) 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. \( E(a\bar{X} + (1-a)cS) = aE\bar{X} + (1-a)E(cS) = a\theta + (1-a)\theta = \theta \). 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, \( \text{Var}(a\bar{X} + (1-a)cS) = a^2V_1 + (1-a)^2V_2 \), where 
\( V_1 = \text{Var}\bar{X} = \theta^2/n \) and \( V_2 = \text{Var}(cS) = c^2E(S^2) = \theta^2 \). 
Use calculus to show that this quadratic function of \( a \) is minimized at 
\[
a = \frac{V_2}{V_1 + V_2} = \frac{(c^2-1)\theta^2}{(1/n) + (c^2-1)\theta^2} = \frac{(c^2-1)(1/n)}{(1/n) + (c^2-1)}.
\]

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. \( E(\bar{X} - cS) = \theta - \theta = 0 \), 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:
\[
E[\theta - (a_1\bar{X} + a_2cS)]^2 = a_1^2\text{Var}\bar{X} + a_2^2c^2\text{Var}S + \theta^2(a_1 + a_2 - 1)^2.
\]

Because \( \text{Var}\bar{X} = \theta^2/n \) and \( \text{Var}S = ES^2 - (ES)^2 = \theta^2 \left( \frac{c^2-1}{c^2} \right) \), we have
\[
E[\theta - (a_1\bar{X} + a_2cS)]^2 = \theta^2 \left[ a_1^2/n + a_2^2(c^2-1) + (a_1 + a_2 - 1)^2 \right],
\]
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 \( t(x) = x, \sum X_i \) is a complete 
sufficient statistic. Any function of \( \sum X_i \) that is an unbiased estimator of \( \lambda \) is the unique 
best unbiased estimator of \( \lambda \). Because \( \bar{X} \) is a function of \( \sum X_i \) and \( E\bar{X} = \lambda, \bar{X} \) is the best 
unbiased estimator of \( \lambda \).

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

Therefore \( E(S^2|\bar{X}) = \bar{X} \). Then we have
\[
\text{Var} S^2 = \text{Var} (E(S^2|\bar{X})) + E \text{Var}(S^2|\bar{X}) = \text{Var} \bar{X} + E \text{Var}(S^2|\bar{X}),
\]
so \( \text{Var} S^2 > \text{Var} \bar{X} \).

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 \( ET(X) = ET'(X) \). Then \( E[T'(X)|T(X)] = T(X) \) 
and \( \text{Var} T'(X) > \text{Var} T(X) \).