1 Solution to Problem 6.40

(a) We will write $T_i = \tau_i(X_1, \ldots, X_n)$ where the $X_i$s are i.i.d. with PDF

$$f(x|\mu, \sigma) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right),$$

where the location parameter $\mu$ is any real number and the scale parameter $\sigma$ is $> 0$. Letting

$$Z_i = \frac{X_i - \mu}{\sigma}$$

we see that the $Z_i$ are i.i.d. with PDF $g(z)$. Thus,

$$\frac{T_1}{T_2} = \frac{\tau_1(X_1, \ldots, X_n)}{\tau_2(X_1, \ldots, X_n)} = \frac{\tau_1(\sigma Z_1 + \mu, \ldots, \sigma Z_n + \mu)}{\tau_2(\sigma Z_1 + \mu, \ldots, \sigma Z_n + \mu)} = \frac{\sigma \tau_1(Z_1, \ldots, Z_n)}{\sigma \tau_2(Z_1, \ldots, Z_n)} = \frac{\tau_1(Z_1, \ldots, Z_n)}{\tau_2(Z_1, \ldots, Z_n)}.$$

Since the distribution of the $Z_i$ don’t depend on the parameters $(\mu, \sigma)$, it follows that $T_1/T_2$ is ancillary.

(b) Trivial.

2 Solution to Problem 7.2

(a) Assuming $\alpha$ is known, we have the log likelihood,

$$\log L(\beta|X) = -n\alpha \log \beta - \beta^{-1} \sum_i X_i + C,$$

where $C$ is a constant that doesn’t depend on $\beta$. Taking derivatives and setting equal to 0 gives

$$-\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_i X_i = 0$$
and the solution is
\[ \hat{\beta}(\alpha) = \alpha^{-1} \bar{X}. \]

It is easy to check that the second derivative of the log likelihood is negative, so this gives us the maximum.

(b) The log likelihood with both \( \alpha \) and \( \beta \) unknown is
\[
\log L(\alpha, \beta | X) = -n \log \Gamma(\alpha) + \alpha \sum_i \log X_i - n\alpha \log \beta - \beta^{-1} \sum_i X_i + C.
\]

Substituting in the value of \( \beta \) that maximizes the likelihood for fixed \( \alpha \) gives a concentrated likelihood:
\[
\log L \left( \alpha, \hat{\beta}(\alpha) | X \right) = -n \log \Gamma(\alpha) + \alpha \sum_i \log X_i + n\alpha \log \alpha + n\alpha \log \bar{X} - n\alpha + C
\]
\[
= n \left\{ -\log \Gamma(\alpha) + \alpha \left[ -\log \bar{X} + \bar{L} - 1 + \log \alpha \right] \right\} + C
\]

where \( C \) doesn’t depend on \( \alpha \) and
\[
\bar{L} = \frac{1}{n} \sum_i \log X_i.
\]

Therefore, we can maximize \( \log L \left( \alpha, \hat{\beta}(\alpha) | X \right) \) over \( \alpha \) to find the MLE for \( \alpha \), then plug this in to \( \hat{\beta}(\alpha) \) to obtain the MLE for \( \beta \), i.e. \( \hat{\beta} = \hat{\beta}(\hat{\alpha}) \).

3 Solution to Problem 7.6

(a) The likelihood is
\[
L(\theta | x_1, \ldots, x_n) = \prod_i \theta x_i^{-2} \mathcal{I}_{\{\theta, \infty\}}(x_i)
\]
\[
= \theta^n \mathcal{I}_{\{\theta, \infty\}} \left( \min_i x_i \right) \prod_i x_i^{-2}
\]

so
\[
T = \min_i X_i,
\]
is a sufficient statistic.

(b) To find the MLE, it suffices to maximize the \( g(T, \theta) \) given above, i.e. to maximize

\[
L_1(\theta | T) = \theta^n I_{[\theta, \infty)}(T) = \theta^n I_{(0,T]}(\theta).
\]

(One can easily check that \( I_{[\theta, \infty)}(T) = I_{(0,T]}(\theta) \) for any \( 0 < \theta \leq T \).) The maximum is clearly \( T^n \) at \( \theta = T \), i.e. the MLE is \( \hat{\theta} = T = \min_i X_i \).

(c) Presumably, we should compute the mean \( \mu(\theta) = E[X_i | \theta] \) and then solve for \( \theta \) in the equation \( \mu(\theta) = \bar{X} \). OK, so blindly following the recipe,

\[
E[X_i | \theta] = \int_{\theta}^{\infty} x \theta^1 x 2^{-1} dx = \theta \int_{\theta}^{\infty} x^{-1} dx = \infty.
\]

Oh, I get it. It’s a trick question. The method of moments estimator doesn’t exist, if we follow the usual recipe.

Of course, we could compute \( \gamma_\alpha(\theta) = E[X_i^\alpha | \theta] \) for any \( \alpha < 1 \), then set this equal to \( n^{-1} \sum_i X_i^{\alpha} \) and solve for \( \theta \):

\[
\gamma_\alpha(\theta) = \theta \int_{\theta}^{\infty} x^{-2+\alpha} dx = \frac{\theta^\alpha}{1-\alpha}.
\]

This yields an estimate

\[
\hat{\theta} = \left( \frac{1-\alpha}{n} \sum_i X_i^{\alpha} \right)^{1/\alpha}.
\]

Obviously different \( \alpha \)s will give different values. Which one should we choose?

### 4 Solution to Problem 7.9

The likelihood can be written

\[
L(\theta | x) = \theta^{-n} I_{[T, \infty)}(\theta),
\]

where \( T = \max_i X_i \) is the sufficient statistic. The MLE is clearly

\[
\hat{\theta}_{MLE} = T.
\]
To compute the mean and variance of this, we need the PDF of $T$ which is

$$f_T(t|\theta) = n\theta^{-n}t^{n-1}I_{[0,\theta]}(t).$$

Hence,

$$E[T^i|\theta] = n\theta^{-n}\int_0^\theta t^{n-1+i} \, dt = \frac{n}{n+i}\theta^i.$$

In particular,

$$E[T|\theta] = \frac{n}{n+1}\theta$$

$$\text{Var}[T|\theta] = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$  \hfill (1)

Finally, we can compute the Mean Squared Error using equation (7.3.1), p. 330:

$$\text{MSE} \left(\theta, \hat{\theta}_{MLE}\right) = \text{Bias}^2 \left(\theta, \hat{\theta}_{MLE}\right) + \text{Var}[\hat{\theta}_{MLE}|\theta]$$

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

$$= \frac{2}{(n+1)(n+2)}\theta^2.$$

Obviously,

$$\mu(\theta) = E[X_i|\theta] = \theta/2.$$  

So the Method of Moments estimator is

$$\hat{\theta}_{MME} = 2\bar{X}.$$  

Now $\bar{X}$ is always an unbiased estimator of the mean, so

$$E\left[\hat{\theta}_{MME}|\theta\right] = \theta,$$

i.e., $\hat{\theta}_{MME}$ is an unbiased estimator of $\theta$. It is easy to check that

$$\text{Var}[X_i|\theta] = \theta^2/12,$$
e.g. apply (1) above with \( n = 1 \). Hence,

\[
\text{Var} \left[ \hat{\theta}_{MME} | \theta \right] = 4 \text{Var} \left[ X | \theta \right] = \theta^2 / (3n).
\]

Since \( \hat{\theta}_{MME} \) is an unbiased estimator of \( \theta \), it follows that

\[
\text{MSE} \left( \theta, \hat{\theta}_{MME} \right) = \frac{\theta^2}{3n}.
\]

Clearly, both MSEs go to 0 as \( n \to \infty \), but the MSE for the MLE goes to 0 faster. One can check that

\[
\text{MSE} \left( \theta, \hat{\theta}_{MME} \right) - \text{MSE} \left( \theta, \hat{\theta}_{MLE} \right) = \frac{n^2 - 3n + 2}{3n(n+1)(n+2)} - \frac{(n-1)(n-2)}{3n(n+1)(n+2)},
\]

which is 0 for \( n = 1 \) and \( n = 2 \), but positive for all \( n > 2 \). Hence, the MLE is never worse than the MME in terms of MSE, and for any sample size \( \geq 3 \) is strictly better in terms of MSE, so we would prefer the MLE over the MME.

**Note:** One may actually do better by correcting for the bias in the MLE; i.e., the estimator

\[
\hat{\theta}_{UMVUE} = \frac{n+1}{n} T,
\]

may actually be better than the MLE in terms of MSE.

### 5 Solution to Problem 7.19

(a) Note that the \( Y_i \) are independent with \( Y_i \sim N(\beta x_i, \sigma^2) \). Thus, the joint PDF is

\[
f \left( y_1, \ldots, y_n | \beta, \sigma^2 \right) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} \sum_i (y_i - \beta x_i)^2 \right]
\]

\[
= (\sigma^2)^{-n/2} \exp \left[ -\frac{\beta^2}{2\sigma^2} \sum_i x_i^2 \right] c(\beta, \sigma^2) \exp \left[ -\frac{1}{2\sigma^2} \sum_i y_i^2 \right] + \frac{\beta}{\sigma^2} \sum_i x_i y_i \left( 2\pi \right)^{-n/2}.
\]
Clearly this is an exponential family and the 2-D sufficient statistic is

\[ (T_1, T_2) = \left( \sum_i Y_i^2, \sum_i x_i Y_i \right). \]

(b) The MLE for \( \beta \) can be obtained by minimizing

\[ R(\beta) = \sum_i (Y_i - \beta x_i)^2. \]

It is also known as the Least Squares Estimator. Taking \( \frac{d}{d\beta} \) and setting equal to 0 gives the normal equation

\[ -2 \sum_i x_i (Y_i - \beta x_i) = 0. \]

The solution is

\[ \hat{\beta} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2}. \]

We have

\[ E \left[ \hat{\beta} \mid \beta, \sigma^2 \right] = \frac{\sum_i x_i E[Y_i \mid \beta, \sigma^2]}{\sum_i x_i^2} = \frac{\sum_i x_i (\beta x_i)}{\sum_i x_i^2} = \beta \frac{\sum_i x_i^2}{\sum_i x_i^2} = \beta. \]

This shows that \( \hat{\beta} \) is an unbiased estimator of \( \beta. \)

(c) Note that \( \hat{\beta} \) is a linear combination of independent normal RVs, so by a theorem stated in class, \( \hat{\beta} \) has a normal distribution. We’ve already computed the mean, so we need only compute the variance.

\[ \text{Var} \left[ \hat{\beta} \mid \beta, \sigma^2 \right] = \frac{\text{Var} \left[ \sum_i x_i Y_i \right]}{(\sum_i x_i^2)^2} = \frac{\sum_i x_i^2 \text{Var}[Y_i]}{(\sum_i x_i^2)^2} = \frac{\sum_i x_i^2 \sigma^2}{(\sum_i x_i^2)^2} = \frac{\sigma^2}{\sum_i x_i^2}. \]
Note that an analysis of units (e.g., suppose the $Y_i$s are in kilograms and the $x_i$s are in meters) doesn’t show any problem with this result.

6 Solution to Problem 7.20

(a)

\[
E \left[ \frac{\sum_i Y_i}{\sum_i x_i} \right] = \frac{\sum_i E[Y_i]}{\sum_i x_i} \\
= \frac{\sum_i \beta x_i}{\sum_i x_i} \\
= \beta,
\]

so the estimator is unbiased.

(b)

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

Now, we claim that

\[
\sum_i x_i^2 \geq \frac{1}{n} \left( \sum_i x_i \right)^2,
\]

which shows that the MLE from the previous exercise has MSE no larger than this estimator. Think of a random variable $X$ obtained by picking one of the $x_i$s at random. This is essentially the uniform distribution on \{x_1, \ldots, x_n\}, except if a value appears more than once in the set, then it gets probability proportional to the number of times it appears. Clearly

\[
\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{n} \sum_i x_i^2 - \left( \frac{1}{n} \sum_i x_i \right)^2 \geq 0.
\]

This proves our claim. In fact, we have strict inequality unless the $x_i$s are all equal, in which case the MLE and the estimator in this problem are in fact equal to $\bar{Y}/x$ where $x$ is the common value of the $x_i$s.