7.8 $X \sim \mathcal{N}(\theta, 1)$ and $Y \sim \mathcal{N}(3\theta, 1)$ are independent random variables.

a. Since X and Y have common support for all values of $\theta \in \Theta = \mathcal{R}^1$, we try finding the maximum likelihood estimate by differentiating the log likelihood. The likelihood function is

$$L(\theta \,|\, X, Y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(X-\theta)^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(Y-3\theta)^2}{2}\right\},$$

and the log likelihood is

$$\mathcal{L}(\theta \,|\, X, Y) = -\log(2\pi) - \frac{(X-\theta)^2}{2} - \frac{(Y-3\theta)^2}{2}.$$

We thus find the score function

$$U(\theta) = (X-\theta) + 3(Y-3\theta),$$

and by setting $U(\hat{\theta}) = 0$, we find $\hat{\theta} = (X+3Y)/10$. (Note that the second derivative of the log likelihood is negative, so $\hat{\theta}$ is in fact the MLE.)

b. For $\tilde{\theta} = aX + bY$, simple laws of expectation tell us that $E[\tilde{\theta}] = a\theta + b3\theta = (a+3b)\theta$. For $\tilde{\theta}$ to be unbiased, we thus need $a + 3b = 1$.

c. For $\tilde{\theta} = aX + bY$, the independence of X and Y along with the properties of variance of scaled random variables and sums of random variables provides that $Var(\tilde{\theta}) = a^2 Var(X) + b^2 Var(Y) = a^2 + b^2$. So for unbiased $\tilde{\theta}$, $a = 1 - 3b$ and the variance is $Var(\tilde{\theta}) = 1 - 6b + 9b^2 + b^2 = 1 - 6b + 10b^2$. The derivative of this variance with respect to b is

$$\frac{dVar(\tilde{\theta})}{db} = -6 + 20b,$$

and setting the derivative equal to zero yields $b = 3/10$, which is a minimum, because the second derivative of the variance is positive. We then find $a = 1 - 3b = 1/10$. This choice of a and b corresponds exactly to the MLE found in part a, so we have shown that the MLE for this Normal linear regression problem is the best linear unbiased estimator.

An aside: Note that this result is easily generalized: Suppose that for $i = 1, \ldots, n$ we have independent random variables $X_i \sim (w_i\theta, \sigma^2)$ (not necessarily normal). Consider an unbiased linear estimator

$$\tilde{\theta} = \sum_{i=1}^{n} a_i X_i.$$

Because we can easily find the mean and variance of $\tilde{\theta}$ as

$$\tilde{\theta} \sim \left(\theta \sum_{i=1}^{n} a_i w_i, \sigma^2 \sum_{i=1}^{n} a_i^2 \right),$$

the unbiasedness of $\tilde{\theta}$ dictates that $\sum_{i=1}^{n} a_i w_i = 1$. So we minimize the variance subject to this constraint using the method of Lagrange multipliers. That is we minimize the function

$$g(\vec{a}, \lambda) = \sum_{i=1}^{n} a_i^2 + \lambda \left(1 - \sum_{i=1}^{n} a_i w_i \right),$$

where $\lambda > 0$ is the Lagrange multiplier. Now

$$\frac{\partial g}{\partial a_k} = 2a_k - \lambda w_k$$
$$\frac{\partial g}{\partial \lambda} = \left(1 - \sum_{i=1}^{n} a_i w_i \right),$$

and setting these $n+1$ equations simultaneously equal to zero, we find that

$$a_k = \frac{\lambda w_k}{2} \qquad \text{and} \qquad \lambda = \frac{2}{\sum_{i=1}^{n} w_i^2}.$$

Thus, the BLUE is

$$\tilde{\theta} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i^2}.$$

Now consider the least squares estimator found as the estimator $\hat{\theta}$ which minimizes the sum of the squared deviations

$$S(\theta) = \sum_{i=1}^{n} (X_i - E(X_i))^2 = \sum_{i=1}^{n} (X_i - w_i \theta)^2.$$

Taking the derivative with respect to θ,

$$\frac{dS}{d\theta} = -2 \sum_{i=1}^{n} w_i (X_i - w_i \theta),$$

and setting that derivative equal to zero yields

$$\hat{\theta} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i^2},$$

which is the BLUE found above. If we further knew that each of the X_i's was normally distributed, we would find that $S(\theta)$ is directly proportional to the score function: $U(\theta) = -S(\theta)/\sigma^2$, and the least squares estimate is the MLE as well as being BLUE. This problem is a simple linear regression problem in which the intercept is known to be zero. The results generalize to a nonzero intercept and to multiple linear regression models:

- For independent random variables having common variance and means that are a known linear function of an unknown parameter vector, the least squares estimate is the BLUE no matter what the shape of the distribution of the X_i's (they need not even have the same shape). (This is the Gauss-Markov Theorem.)
- If the independent random variables described above are normally distributed, the least squares estimate is also the MLE.

7.10 We are given that estimator $T_n(\vec{X}_n)$ has asymptotic distribution

$$n^\delta(T_n(\vec{X}_n) - \theta) \to_d \mathcal{N}\left(0, J^2(\theta)\right)$$

for some $\delta > 0$. Now $n^{-\delta} \to 0$, so a simple application of Slutsky's theorem provides that

$$n^{-\delta} n^\delta(T_n(\vec{X}_n) - \theta) \to_d 0\mathcal{N}\left(0, J^2(\theta)\right) = 0.$$

And when a random variable converges in distribution to a constant, it also converges in probability to that constant, proving the consistency we desire.

7.11 In the following, we will consider the concentration of the approximate asymptotic distributions in the following manner: For two estimators T_n and S_n, we define T_n asymptotically more efficient than S_n if for any given $r > 0$, there exists an n_r such that for all $n > n_r$, $Pr(|T_n - \theta| < r) - Pr(|S_n - \theta| < r) > 0$.

Suppose we are given that estimators $T_n(\vec{X}_n)$ and $S_n(\vec{X}_n)$ have asymptotic distributions

$$n^\delta(T_n(\vec{X}_n) - \theta) \to_d \mathcal{N}\left(0, J^2(\theta)\right)$$
$$n^\gamma(S_n(\vec{X}_n) - \theta) \to_d \mathcal{N}\left(0, V^2(\theta)\right)$$

for some $\delta > 0$ and $\gamma > 0$. Then, by the definition of convergence in distribution, for any $\epsilon > 0$ there exist $n_{\epsilon T}$ and $n_{\epsilon S}$ such that for all c and every $n > n_\epsilon = \max(n_{\epsilon T}, n_{\epsilon S})$

$$|Pr(n^\delta(T_n(\vec{X}_n) - \theta) < c) - \Phi(c/J(\theta))| < \epsilon$$
$$|Pr(n^\gamma(S_n(\vec{X}_n) - \theta) < c) - \Phi(c/V(\theta))| < \epsilon$$

where $\Phi(z)$ is the cumulative distribution function for the standard normal distribution. Now, for fixed n,

$$\begin{aligned} Pr(|T_n - \theta| < r) &= Pr((T_n - \theta) < r) - Pr((T_n - \theta) < -r) \\ &= Pr((T_n - \theta) < r) - Pr((T_n - \theta) < -r) \\ &= Pr(n^\delta(T_n - \theta) < n^\delta r) - Pr(n^\delta(T_n - \theta) < -n^\delta r) \end{aligned}$$

with an analogous result for $Pr(|S_n - \theta| < r)$. So for sufficiently large n,

$$|Pr(|T_n - \theta| < r) - (\Phi(n^\delta r/J(\theta)) - \Phi(-n^\delta r/J(\theta))| < 2\epsilon$$
$$|Pr(|S_n - \theta| < r) - (\Phi(n^\gamma r/V(\theta)) - \Phi(-n^\gamma r/V(\theta))| < 2\epsilon$$

We then know that for sufficiently large n, the quantity

$$(\Phi(n^\delta r/J(\theta)) - \Phi(-n^\delta r/J(\theta)) - (\Phi(n^\gamma r/V(\theta)) - \Phi(-n^\gamma r/V(\theta))$$

is within 4ϵ of the true value of

$$Pr(|T_n - \theta| < r) - Pr(|S_n - \theta| < r).$$

Now, for fixed $r > 0$ and $J(\theta)$ and $V(\theta)$, if $\delta > \gamma > 0$, then for any n such that

$$n > \left(\frac{J(\theta)}{V(\theta)}\right)^{1/(\delta-\gamma)}$$

we have

$$\Phi(n^\delta r/J(\theta)) - \Phi(n^\gamma r/V(\theta)) = \Phi(-n^\gamma r/V(\theta)) - \Phi(-n^\delta r/J(\theta)) = \Delta_n > 0$$

Note that Δ_n increases as a function of n for $\delta > \gamma$. Thus for any choice of $r > 0$ and any given $\delta > \gamma > 0$, we can choose an n_0 such that $\Delta_{n_0} > 0$. Then, choosing $\epsilon < \Delta_{n_0}/4$ will guarantee that for $n > n_r = \max(n_0, n_\epsilon)$ the comparison based on the approximate distributions will be of the same sign as the comparison based on the actual distributions.

Now consider a CAN estimator S_n, which by definition has $\gamma = 1/2$. Then

a. when $\delta > 1/2$, T_n is asymptotically more efficient than S_n,

b. when $\delta = 1/2$, T_n is also CAN allowing comparisons of asymptotic relative efficiency according to that definition, and

c. when $\delta < 1/2$, S_n is asymptotically less efficient than S_n.

7.13 We are given $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with μ known. The MLE of σ^2 is thus

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Now, based on the normality of X_i, we know that for $W_i = ((X_i - \mu)/\sigma)^2$ has the W_i's i.i.d. $W_i \sim \chi_1^2$ and thus $E(W_i) = 1$ and $Var(W_i) = 2$. By the CLT,

$$\sqrt{n}(\overline{W} - 1) \to_d \mathcal{N}(0, 2)$$

and because $\hat{\sigma^2} = \sigma^2\overline{W}$, we immediately obtain (via delta method or Slutsky's) that

$$\sqrt{n}(\hat{\sigma^2} - \sigma^2) \to_d \mathcal{N}(0, 2\sigma^4).$$

An aside: Note that we know that $\sum_{i=1}^{n} W_i$ is exactly distributed according to a chi square distribution with n degrees of freedom. Hence, we find that for large degrees of freedom, a chi square distribution is approximately normal.

7.14 We are given that independent $X_i \sim \mathcal{U}(\theta - 1/2, \theta + 1/2)$ for $i = 1, 2, \ldots$. Any estimator $a(X_{(1)} + 1/2) + (1-a)(X_{(n)-1/2})$ for $0 < a < 1$ is an MLE of θ. I will first find the exact distribution of the general MLE.

The density for the MLE is found by first finding the joint density for $U = a(X_{(1)} + 1/2) \in (a\theta, a(\theta+1))$ and $V = (1-a)(X_{(n)} - 1/2) \in ((1-a)(\theta-1), (1-a)\theta)$, and then finding the density for $T_n = U + V \in (\theta + a - 1, \theta + a)$ by convolution. Now for $\theta - 1/2 < u/a - 1/2 < v/(1-a) + 1/2 < \theta + 1/2$

$$
\begin{aligned}
Pr(U > u, V < v) &= Pr(X_{(1)} > \frac{u}{a} - \frac{1}{2}, X_{(n)} < \frac{v}{(1-a)} + \frac{1}{2}) \\
&= \prod_{i=1}^{n} Pr(\frac{u}{a} - \frac{1}{2} < X_i < \frac{v}{(1-a)} + \frac{1}{2}) \\
&= \left(\frac{v}{(1-a)} + \frac{1}{2} - \frac{u}{a} + \frac{1}{2} \right)^n
\end{aligned}
$$

The joint density is then found as the negative derivative of the joint probability

$$
\begin{aligned}
f_{U,V}(u,v) &= -\frac{\partial^2}{\partial u \partial v} Pr(U > u, V < v) \\
&= \frac{n(n-1)}{a(1-a)} \left(\frac{v}{(1-a)} - \frac{u}{a} + 1 \right)^{n-2} \mathbf{1}_{[\theta < u/a < v/(1-a)+1 < \theta+1]}
\end{aligned}
$$

The density for $T_n = U + V$ is thus (for $\theta + a - 1 < t < \theta + a$)

$$
\begin{aligned}
f_T(t) &= \mathbf{1}_{[\theta+a-1<t<\theta+a]} \int_{-\infty}^{\infty} f_{U,V}(t-v, v) dv \\
&= \mathbf{1}_{[\theta+a-1<t<\theta+a]} \int_{L_t}^{H_t} \frac{n(n-1)}{a(1-a)} \left(\frac{v}{(1-a)} - \frac{t-v}{a} + 1 \right)^{n-2} dv \\
&= n \left(\frac{H_t}{a(1-a)} - \frac{t}{a} + 1 \right)^{n-1} - n \left(\frac{L_t}{a(1-a)} - \frac{t}{a} + 1 \right)^{n-1}
\end{aligned}
$$

where the limits of integration L_t and H_t will be determined by the indicator function in the formula for $f_{U,V}(t-v, v)$. We need to integrate over those values of v such that

$$
\theta < \frac{t-v}{a} < \frac{v}{1-a} + 1 < \theta + 1,
$$

which when taking each successive pair of terms dictates that

$$
\begin{aligned}
v &< t - a\theta \\
(1-a)(t-a) &< v \\
v &< (1-a)\theta
\end{aligned}
$$

The lower limit of integration is thus clearly $(1-a)t$ for all $t \in (\theta+a-1, \theta+a)$, and we will have to consider cases for H_t. The upper limit will be $t - a\theta$ when

$$t - a\theta < (1-a)\theta \quad \text{or} \quad t < \theta,$$

and the upper limit will be $(1-a)\theta$ otherwise. Hence, defining sets $A = (\theta+a-1, \theta)$, $B = [\theta, \theta+a)$, and $C = [\theta+a, \infty)$, we find the density for T_n as

$$f_T(t) = n\left(\frac{t-\theta}{1-a}+1\right)^{n-1} \mathbf{1}_A(t) + n\left(\frac{\theta-t}{a}+1\right)^{n-1} \mathbf{1}_B(t).$$

We then find the cumulative distribution function by integrating the density.

$$\begin{aligned} F_T(t) &= \int_{-\infty}^{t} f_T(y)\,dy \\ &= (1-a)\left(\frac{t-\theta}{1-a}+1\right)^{n} \mathbf{1}_A(t) + \left(1 - a\left(\frac{\theta-t}{a}+1\right)^{n}\right) \mathbf{1}_B(t) + \mathbf{1}_C(t) \end{aligned}$$

Now we want to consider the distribution of $Y_n = n(T_n - \theta)$.

$$\begin{aligned} Pr(n(T-\theta) < y) &= Pr(T < \frac{y}{n} + \theta) \\ &= (1-a)\left(\frac{y}{n(1-a)}+1\right)^{n} \mathbf{1}_A(\frac{y}{n}+\theta) + \\ &\quad \left(1 - a\left(-\frac{y}{na}+1\right)^{n}\right) \mathbf{1}_B(\frac{y}{n}+\theta) + \mathbf{1}_C(\frac{y}{n}+\theta) \end{aligned}$$

Taking the limit as $n \to \infty$, we note that

$$\begin{aligned} \mathbf{1}_A(y/n+\theta) &\to \mathbf{1}_{[y<0]} \\ \mathbf{1}_B(y/n+\theta) &\to \mathbf{1}_{[y\geq 0]} \\ \mathbf{1}_C(y/n+\theta) &\to 0 \\ \left(1 + \frac{y}{n(1-a)}\right)^{n} &\to e^{y/(1-a)} \\ \left(1 - \frac{y}{na}\right)^{n} &\to e^{-y/a} \end{aligned}$$

so

$$F_{Y_n}(y) \to (1-a)e^{y/(1-a)}\mathbf{1}_{[y<0]} + \left(1 - ae^{-y/a}\right)\mathbf{1}_{[y\geq 0]}.$$

a. Now, for $a = 1/2$, we find that the asymptotic distribution for Y_n is the double exponential distribution

$$F_{Y_n}(y) \to \frac{1}{2}e^{2y}\mathbf{1}_{[y<0]} + \left(1 - \frac{1}{2}e^{-2y}\right)\mathbf{1}_{[y\geq 0]}.$$

We can easily find that the density for this asymptotic distribution is

$$f(y) = e^{-2|y|},$$

and the mean and variance of this asymptotic distribution are 0 and 1/2, respectively. Hence, the approximate asymptotic distribution of $T = Y/n + \theta$ will have mean θ and variance $1/(2n^2)$.

b. We now consider an estimator $\overline{X}_n$ which by the CLT is easily shown to be CAN with approximate asymptotic distribution

$$\overline{X}_n \dot{\sim} \mathcal{N}\left(\theta, \frac{1}{12n}\right).$$

Now for large n, the probability that $\overline{X}_n$ will be greater than $r\sqrt{12n}$ units away from θ is approximately

$$Pr(|\overline{X}_n - \theta| > r\sqrt{12n}) \doteq 2\Phi\left(-r\right),$$

while by Chebyshev's we know that the approximate distribution would suggest

$$Pr(|T_n - \theta| > r\sqrt{12n}) \leq \frac{1}{24r^2n^3}.$$

Thus, for any fixed r, it is relatively easy to find n such that

$$\frac{1}{24r^2n^3} <<< 2\Phi\left(-r\right).$$

Hence, by similar arguments, we prefer T_n to any CAN estimator in large samples.

c. The asymptotic distribution for the general case was found above.

Supplemental Problems

1. Let Y_i, $i = 1, \ldots, n$ be independent exponential random variables with $Y_i \sim \mathcal{E}(\log(2)/\theta)$ (so $F_Y(y) = 1 - \exp(-\log(2)y/\theta)$ for $0 < y < \infty$.

   a. Find the parametric MLE of the median of the distribution of Y_i. Derive its asymptotic distribution.

Ans: The median of the exponential distribution is that value ω such that $F_Y(\omega) = 1 - \exp(-\log(2)\omega/\theta) = 1/2$. Straightforward substitution finds that $\omega = \theta$. Now, we find the MLE $\hat{\theta}$

$$\begin{aligned}
L_i(\theta) &= \frac{\log(2)}{\theta}\exp(-\log(2)Y_i/\theta) \\
\mathcal{L}_i(\theta) &= -\log(\theta) - \frac{\log(2)Y_i}{\theta} \\
U_i(\theta) &= -\frac{1}{\theta} + \frac{\log(2)Y_i}{\theta^2} \\
U(\theta) &= -\frac{n}{\theta} + \frac{\log(2)n\overline{Y}}{\theta^2}
\end{aligned}$$

Then, the score equation $U(\hat{\theta}) = 0$ dictates that $\hat{\theta} = \log(2)\overline{Y}$. As this involves the sample mean, the asymptotic distribution is easily found via the CLT and Slutsky's theorem (or the delta method), by noting that the mean and variance of the exponential are $\theta/\log(2)$ and $(\theta/\log(2))^2$, respectively. Hence, the asymptotic distribution of the sample mean is

$$\sqrt{n}\left(\overline{Y} - \frac{\theta}{\log(2)}\right) \to_d \mathcal{N}\left(0, \frac{\theta^2}{\log^2(2)}\right),$$

and since $\hat{\theta} = \overline{Y}\log(2)$

$$\sqrt{n}(\hat{\theta} - \theta) \to_d \mathcal{N}(0, \theta^2).$$

b. Find the asymptotic distribution of the sample median.

Ans: For notational convenience, let $\tilde{\theta}$ be the sample median. Now we know that the asymptotic distribution for sample medians of continuous distributions is

$$\sqrt{n}(\tilde{\theta} - \theta) \to_d \mathcal{N}\left(0, \frac{1}{4f^2(\theta)}\right).$$

For the exponential distribution, $f(\theta) = \log(2)/(2\theta)$, so we have

$$\sqrt{n}(\tilde{\theta} - \theta) \to_d \mathcal{N}\left(0, \frac{\theta^2}{\log^2(2)}\right).$$

c. What is the asymptotic relative efficiency of the two estimators found in parts 1a and 1b?

Ans: The asymptotic relative efficiency is the ratio of the variances of the asymptotic distributions

$$e(\hat{\theta}, \tilde{\theta}) = \log^2(2) \doteq 0.5,$$

so the parametric estimator is more efficient.

d. Now suppose that the true distribution of the independent Y_i's is as lognormal $Y_i \sim LN(\mu, \sigma^2)$, having density

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right) \mathbf{1}_{[y>0]}.$$

Further suppose $\mu = \log(\theta)$. For what function of θ is the estimator you found in part a consistent? What is the asymptotic distribution of the estimator from part 1a under this new distribution for Y?

Ans: For the lognormal distribution, $E[Y_i] = \exp(\mu + \sigma^2/2)$ and $Var(Y_i) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$. By the CLT, we know therefore that

$$\sqrt{n}\left(\overline{Y} - \exp(\mu + \sigma^2/2)\right) \to_d \mathcal{N}\left(0, (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\right),$$

and since $\hat{\theta} = \overline{Y}\log(2)$

$$\sqrt{n}\left(\hat{\theta} - \log(2)\exp(\mu + \sigma^2/2)\right) \to_d \mathcal{N}\left(0, \log^2(2)(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\right).$$

Immediately we see that $\hat{\theta}$ is consistent for $\log(2)\exp(\mu + \sigma^2/2)$, which is only the median of the distribution of Y if $\log(2)\exp(\sigma^2/2) = 1$, since the median of the lognormal distribution is $\exp(\mu)$.

2. Let Y_i, $i = 1, \ldots, n$ be independent lognormal random variables with $Y_i \sim LN(\log(\theta), \sigma^2)$.

   a. Find the parametric MLE of the median of the distribution of Y_i. Derive its asymptotic distribution.

   Ans: The parametric MLE in this case is the geometric mean

   $$\check{\theta} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(Y_i)\right).$$

   The asymptotic distribution is easily found via the CLT for the $\log(Y_i)$'s (which are normally distributed), and then using the delta method.

   b. Find the asymptotic distribution of the sample median.

   Ans: Use the asymptotic result in a manner analogous to the above problem.

   c. What is the asymptotic relative efficiency of the two estimators found in parts 2a and 2b?

   d. Now suppose that the true distribution of the independent Y_i's is as exponential $Y_i \sim \mathcal{E}(\log(2)/\theta)$ as in problem 1. For what function of θ is the estimator you found in part 2a consistent? What is the asymptotic distribution of the estimator from part 2a under this new distribution for Y?

   Ans: In this case, $\check{\theta}$ will be consistent for the geometric mean of the exponential distribution. The asymptotic distribution is found via the mean and variance for log transformed exponential random variables.

3. Discuss the relative merits of using parametric versus nonparametric estimators relative to your results in problems 1 and 2.

   Ans: In each case, the parametric estimators were most efficient. However, if we were wrong about the true distribution of the data, neither parametric estimator turned out to be consistent for the median. The nonparametric estimator was "distribution-free", in the sense that it was consistent for the median no matter what the underlying distribution. So what do you want: efficiency or consistency? If it is of paramount importance to estimate the median, it seems to me that inference based on parametric estimation is treacherous.