Solution to Problems 8.6, 8.21 and 8.48

8.6. $X \sim \text{Binom}(n, p)$, so the likelihood function is

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } 0 < p < 1,$$

and the log-likelihood function is

$$l(p) = \log \binom{n}{x} + x \log p + (n-x)\log(1-p) \qquad \text{for } 0 < p < 1.$$

a) Evaluating the derivative

$$\begin{aligned} l'(p) &= \frac{x}{p} - \frac{n-x}{1-p}. \\ \Rightarrow l'(p) > 0 &\Leftrightarrow p < \frac{x}{n}, \end{aligned}$$

so it is clear that the root $p = x/n$ corresponds to a maximum of $l(p)$ and hence of $L(p)$; that is,

$$\hat{p}_{ML} = \frac{X}{n}.$$

Alternately,

$$l''(p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2},$$

so $l''(p) < 0$ for all permitted values of p, leading to the same conclusion.

b) The Fisher information for p in the observation X is given by:

$$\begin{aligned} \Rightarrow -E(l''(p)) &= \frac{E(X)}{p^2} + \frac{n - E(X)}{(1-p)^2} \\ &= \frac{n}{p} + \frac{n}{(1-p)} \\ &= \frac{n}{p(1-p)} \\ \Rightarrow \text{CRLB} &= \frac{p(1-p)}{n}. \end{aligned}$$

Direct calculation yields

$$\text{Var}(\hat{p}_{ML}) = \frac{\text{Var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Thus, in this example, the MLE achieves the Cramer-Rao lower bound exactly (no matter what the value of n) and not just asymptotically (as is always the case).

c) With $n = 10$ and $X = 5$, the likelihood function becomes

$$L(p) = 252\, p^5 (1-p)^5 \qquad \text{for } 0 < p < 1,$$

which is easy to plot using R. In particular, $L(p)$ is symmetric around $p = 1/2$, achieves its maximum at $p = 1/2$ and $L(p) \to 0$ as $p \to 0, 1$.

The log likelihood is

$$l(p) = \log(252) + 5 \log[p(1-p)] \qquad \text{for } 0 < p < 1,$$

which is also symmetric around $p = 1/2$, achieves its maximum at $p = 1/2$ and $\to -\infty$ as $p \to 0, 1$.

8.21. a) We need to evaluate the first population moment:

$$\mu_1 = E(X) = \int_\theta^\infty x e^{-(x-\theta)}\, dx.$$

This is easy to integrate directly, but let's evaluate the integral by relating it to a familiar density function. If we use the transformation $y = x - \theta$, the limits of integration will become 0 and ∞ and the integral will be related to a gamma density function. Let $y = x - \theta \Rightarrow dy = dx$, to obtain

$$\mu_1 = \int_0^\infty (y+\theta) e^{-y}\, dy = \int_0^\infty y e^{-y}\, dy + \theta \int_0^\infty e^{-y}\, dy = \Gamma(2) + \theta = 1 + \theta.$$

Re-expressing this, we have $\theta = \mu_1 - 1$, so we obtain $\hat{\theta}_{MM} = \bar{X} - 1$.

Although not requested, you might want to show $E(\hat{\theta}_{MM}) = E(\bar{X}) - 1 = E(X) - 1 = \theta$ (so $\hat{\theta}_{MM}$ is unbiased) and $\text{Var}(\hat{\theta}_{MM}) = \text{Var}(\bar{X}) = \text{Var}(X)/n = 1/n$, so that $MSE(\hat{\theta}_{MM}) = 1/n$. To get $\text{Var}(X)$, you can evaluate $E(X^2)$ along the lines above or (more easily) by noting that $X - \theta \sim \exp(1) = \text{gamma}(1,\ 1)$, from which you immediately get that $\text{Var}(X) = \text{Var}(X - \theta) = 1$, leading to the desired results. (Of course, you could have used the same device to evaluate $E(X) = \theta + 1$.)

b) Care is required here because the support of the density function depends on the parameter θ. In such cases — just like the $U(0,\ \theta)$ case done in class — you need to maximize the likelihood function directly; differentiating and setting to 0 to find roots — of either the likelihood function or the log likelihood function — is **not** going to work! The main point then is to make sure you get the likelihood function correct.

The likelihood function is given by

$$L(\theta) = \begin{cases} \prod_{i=1}^n e^{-(x_i - \theta)} & \text{provided } x_i \geq \theta \text{ for all } i, \\ 0 & \text{otherwise.} \end{cases}$$

If we set $x_{min} = \text{Min}(x_1, x_2, \ldots, x_n)$, then

$$L(\theta) = \begin{cases} e^{-n(\bar{x} - \theta)} & \text{provided } \theta \leq x_{min}, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, $L(\theta) = 0$ for $\theta > x_{min}$. On the other hand, for $\theta \le x_{min}$, $L(\theta)$ increases as θ increases ($L(\theta) \to 0$ as $\theta \to -\infty$). It follows immediately that:

$$\hat{\theta}_{ML} = X_{min}.$$

Although not requested, you might want to evaluate the exact Bias and MSE of $\hat{\theta}_{ML}$ so you can compare the performance of the $\hat{\theta}_{MM}$ and $\hat{\theta}_{ML}$ for any fixed value of n. To do this, you first need to determine the distribution of X_{min}. Of course, the values of X_{min} are limited to values greater than θ. For $x > \theta$, we have

$$P(X_{min} > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x) = \prod_{i=1}^{n} P(X_i > x) \text{ by independence.}$$

So, if $F(\cdot)$ denotes the cumulative distribution function of X_{min}, we have

$$1 - F(x) = \prod_{i=1}^{n} \int_x^{\infty} e^{-(u-\theta)}\, du = \left[e^{-(x-\theta)}\right]^n = e^{-n(x-\theta)},$$

which yields the density function for X_{min} as:

$$f(x) = \begin{cases} ne^{-n(x-\theta)} & \text{for } x \ge \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Now you can evaluate $E(X_{min})$ and $E(X_{min}^2)$ directly along the lines in a). Alternately, the density evaluated for X_{min} implies that $X_{min} - \theta \sim \text{gamma}(1, n)$, so you can immediately obtain $E(X_{min} - \theta) = 1/n$ or $E(X_{min}) = \theta + 1/n$ (so $\text{Bias}(\hat{\theta}_{ML}) = 1/n$) and $\text{Var}(X_{min}) = \text{Var}(X_{min} - \theta) = 1/n^2$.

In summary, for this problem,

$$\text{MSE}(\hat{\theta}_{ML}) = 2/n^2 \quad \text{and} \quad \text{MSE}(\hat{\theta}_{MM}) = 1/n,$$

so

$$\text{MSE}(\hat{\theta}_{ML}) \le \text{MSE}(\hat{\theta}_{MM}) \Leftrightarrow n \ge 2.$$

Note that in this problem, the variance of the MLE equals $1/n^2$, whereas in typical problems the variance of the MLE behaves like $1/n$. This is directly related to the fact that the support of the density function of the X_i's depends on the parameter θ. In this problem, the "standard" asymptotic results for the MLE do **not** hold because the support of the density depends on the parameter which implies that one **cannot** interchange differentiation and integration the way we did when we derived those results.

8.48. Note that the method of estimating λ (using the proportion of 0's to estimate $P(X = 0)$) corresponds to using the method of moments when the only aspect of the data that is observed is whether the Poisson observations are 0 or not. We are interested in

$$\tilde{\lambda} = -\log\left(\frac{Y}{n}\right) = g(Y) \text{ say,}$$

where $g(y) = -\log(y/n) = -\log(y) + \log(n)$. Recall the second order Taylor series expansion

$$g(Y) \approx g(\mu_Y) + (Y - \mu_Y)g'(\mu_Y) + \frac{1}{2}(Y - \mu_Y)^2 g''(\mu_Y)$$

yields

$$E[g(Y)] \cong g(\mu_Y) + \frac{1}{2}\text{Var}(Y)\, g''(\mu_Y).$$

Now $Y \sim \text{Binom}(n, p_0) \;\Rightarrow\; \mu_Y = E(Y) = np_0$ and $\text{Var}(Y) = np_0(1 - p_0)$. Since

$$g'(y) = -\frac{1}{y} \quad \text{and} \quad g''(y) = \frac{1}{y^2},$$

the second order approximation yields

$$E(\tilde{\lambda}) \cong -\log\left(\frac{np_0}{n}\right) + \frac{1}{2}np_0(1 - p_0) \times \frac{1}{(np_0)^2} = -\log(p_0) + \frac{1 - p_0}{2np_0}.$$

Substituting $p_0 = e^{-\lambda}$ yields

$$E(\tilde{\lambda}) \cong \lambda + (e^{\lambda} - 1)/2n \;\Rightarrow\; \text{Bias}(\tilde{\lambda}) \cong (e^{\lambda} - 1)/2n,$$

so the estimator $\tilde{\lambda}$ is asymptotically unbiased.

Similarly, we evaluate the asymptotic variance of $\tilde{\lambda}$ using the first order approximation:

$$\text{A.Var}(\tilde{\lambda}) = \text{Var}(Y)\,[g'(\mu_Y)]^2 = np_0(1 - p_0)\left(-\frac{1}{np_0}\right)^2 = \frac{e^{\lambda} - 1}{n}.$$

The MLE of λ is given by $\bar{X}$ (see page 282) which has expectation λ (so is unbiased for any value of n) and variance λ/n. Since the contribution of the bias of $\tilde{\lambda}$ to its MSE is negligible for large values of n (relative to the magnitude of the variance), we focus on comparing the variances:

$$\frac{\text{A.Var}(\tilde{\lambda})}{\text{Var}(\hat{\lambda}_{ML})} = \frac{(e^{\lambda} - 1)/n}{\lambda/n} = \frac{e^{\lambda} - 1}{\lambda} > 1 \qquad \text{for } \lambda > 0.$$

Note that $(e^{\lambda} - 1)/\lambda \approx 1 + \lambda/2$ as $\lambda \to 0$, so the advantage of the MLE is not large when λ is small (which makes intuitive sense: when λ is small, many of the Poisson observations will be 0's, so estimating λ based simply on the proportion of 0's should work pretty well). However, when λ is larger, $e^{\lambda} - 1$ is much larger than λ and the MLE yields a much more precise estimate than $\tilde{\lambda}$ (which also makes intuitive sense: if only a small proportion of the Poisson observations are 0's, estimating λ based simply on the proportion of 0's is not going to lead to a very precise estimate).