6.17 The population pmf is $f(x|\theta) = \theta(1-\theta)^{x-1} = \frac{\theta}{1-\theta} e^{\log(1-\theta)x}$, an exponential family with $t(x) = x$. Thus, $\sum_i X_i$ is a complete, sufficient statistic by Theorems 6.2.10 and 6.2.25. $\sum_i X_i - n \sim$ negative binomial(n, θ).

6.22 a. The sample density is $\prod_i \theta x_i^{\theta-1} = \theta^n (\prod_i x_i)^{\theta-1}$, so $\prod_i X_i$ is sufficient for θ, not $\sum_i X_i$.

b. Because $\prod_i f(x_i|\theta) = \theta^n e^{(\theta-1)\log(\prod_i x_i)}$, $\log(\prod_i X_i)$ is complete and sufficient by Theorem 6.2.25. Because $\prod_i X_i$ is a one-to-one function of $\log(\prod_i X_i)$, $\prod_i X_i$ is also a complete sufficient statistic.

7.37 To find a best unbiased estimator of θ, 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 θ. So $g(\theta) = 0$, for all θ, 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 θ 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.)

7.48 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 $\mathrm{Var}\,\hat{p} = p(1-p)/n$. 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_is 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 $Y \sim$ exponential(λ/n) so $\mathrm{E}\, Y = \lambda/n$ and nY is an unbiased estimator of λ.

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 λ. 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 λ is the best unbiased estimator of λ. 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 λ.

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.

7.52 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 λ is the unique best unbiased estimator of λ. 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 λ.

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 λ and, by Theorem 7.3.23, it is also the unique best unbiased estimator of λ. 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 $\mathrm{Var}\,T'(X) > \mathrm{Var}\,T(X)$.