CHAPTER 7

Section 7.1

1.

a. We use the sample mean, $\bar{x}$, to estimate the population mean μ: $\hat{\mu} = \bar{x} = \frac{\Sigma x_i}{n} = \frac{3753}{33} = 113.73$.

b. The quantity described is the median, $\tilde{\mu}$, which we estimate with the sample median: $\tilde{x}$ = the middle observation when arranged in ascending order = the 17th ordered observation = 113.

c. To estimate σ, we use the sample standard deviation, $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{33}(x_i - 113.73)^2}{33-1}} =$

$= \sqrt{162.39} = 12.74$. First-graders' IQ scores typically differ from the mean IQ of 113.73 by about ±12.74 points.

d. All but three of the 33 first graders have IQs above 100. With "success" = IQ greater than 100 and x = # of successes = 33, $\hat{p} = \frac{x}{n} = \frac{30}{33} = .9091$.

e. A sensible estimate of σ/μ is $\hat{\sigma}/\hat{\mu} = s/\bar{x}$ = 12.74/113.73 = .112.

3. You can calculate for this data set that $\bar{x} = 1.3481$ and $s = .3385$.

a. We use the sample mean, $\bar{x} = 1.3481$.

b. The estimated standard error of $\bar{x}$ is $\frac{s}{\sqrt{n}} = \frac{.3385}{\sqrt{16}} = .0846$.

c. Because we assume normality, the mean = median, so we also use the sample mean $\bar{x} = 1.3481$. We could also easily use the sample median.

d. For a normal distribution, the 90th percentile is equal to $\mu + 1.28\sigma$. An estimate of that population 90th percentile is $\hat{\mu} + (1.28)\hat{\sigma} = \bar{x} + 1.28s = 1.3481 + (1.28)(.3385) = 1.7814$.

e. Since we can assume normality, $P(X < 1.5) = \Phi\left(\frac{1.5-\mu}{\sigma}\right) \approx \Phi\left(\frac{1.5-\bar{x}}{s}\right) = \Phi\left(\frac{1.5-1.3481}{.3385}\right)$

$= \Phi(.45) = .6736$.

5. Let θ = the total audited value. Three potential estimators of θ are $\hat{\theta}_1 = N\overline{X}$, $\hat{\theta}_2 = T - N\overline{D}$, and $\hat{\theta}_3 = T \cdot \dfrac{\overline{X}}{\overline{Y}}$.

From the data, $\overline{y}$ = 374.6, $\overline{x}$ = 340.6, and $\overline{d}$ = 34.0. Knowing N = 5,000 and T = 1,761,300, the three corresponding estimates are $\hat{\theta}_1 = (5{,}000)(340.6) = 1{,}703{,}000$, $\hat{\theta}_2 = 1{,}761{,}300 - (5{,}000)(34.0) = 1{,}591{,}300$, and $\hat{\theta}_3 = 1{,}761{,}300\left(\dfrac{340.6}{374.6}\right) = 1{,}601{,}438.281$.

7.

a. $\hat{\mu} = \overline{x} = \dfrac{\sum x_i}{n} = \dfrac{1206}{10} = 120.6.$

b. Since $\tau = 10{,}000\mu$, $\hat{\tau} = 10{,}000\hat{\mu} = 10{,}000(120.6) = 1{,}206{,}000$.

c. 8 of 10 houses in the sample used at least 100 therms (the "successes"), so $\hat{p} = \frac{8}{10} = .80$.

d. The ordered sample values are 89, 99, 103, 109, 118, 122, 125, 138, 147, 156, from which the two middle values are 118 and 122, so $\hat{\eta} = \tilde{x} = (118 + 122)/2 = 120$.

9.

a. $E(\overline{X}) = \mu = E(X)$, so $\overline{X}$ is an unbiased estimator for the Poisson parameter μ. Since n = 150,
$\hat{\mu} = \overline{x} = \dfrac{\Sigma x_i}{n} = \dfrac{(0)(18) + (1)(37) + \ldots + (7)(1)}{150} = \dfrac{317}{150} = 2.11.$

b. $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{\sqrt{\mu}}{\sqrt{n}}$, so the estimated standard error is $\sqrt{\dfrac{\hat{\mu}}{n}} = \dfrac{\sqrt{2.11}}{\sqrt{150}} = .119$.

11. From the description $X_1 \sim \text{Bin}(n_1, p_1)$ and $X_2 \sim \text{Bin}(n_2, p_2)$.

a. $E\left(\hat{P}_1 - \hat{P}_2\right) = \dfrac{1}{n_1}E(X_1) - \dfrac{1}{n_2}E(X_2) = \dfrac{1}{n_1}(n_1 p_1) - \dfrac{1}{n_2}(n_2 p_2) = p_1 - p_2$. Hence, by definition, $\hat{P}_1 - \hat{P}_2$ is an unbiased estimator of $p_1 - p_2$.

b. $V\left(\hat{P}_1 - \hat{P}_2\right) = V\left(\dfrac{X_1}{n_1}\right) + (-1)^2 V\left(\dfrac{X_2}{n_2}\right) = \left(\dfrac{1}{n_1}\right)^2 V(X_1) + \left(\dfrac{1}{n_2}\right)^2 V(X_2) =$

$\dfrac{1}{n_1^2}(n_1 p_1 q_1) + \dfrac{1}{n_2^2}(n_2 p_2 q_2) = \dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}$, and the standard error is the square root of this quantity.

c. With $\hat{p}_1 = \dfrac{x_1}{n_1}$, $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{p}_2 = \dfrac{x_2}{n_2}$, $\hat{q}_2 = 1 - \hat{p}_2$, the estimated standard error is $\sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}$.

d. $\hat{p}_1 - \hat{p}_2 = \frac{127}{200} - \frac{176}{200} = .635 - .880 = -.245$.

e. $\sqrt{\frac{(.635)(.365)}{200} + \frac{(.880)(.120)}{200}} = .041$.

13.

a. First, the mgf of each X_i is $M_{X_i}(t) = \frac{\lambda}{\lambda - t}$. Then, using independence, $M_{\Sigma X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$. Finally, using $\overline{X} = \frac{1}{n}\Sigma X_i$ and the properties of mgfs, $M_{\overline{X}}(t) = M_{\Sigma X_i}(\frac{1}{n}t) = \left(\frac{\lambda}{\lambda - \frac{1}{n}t}\right)^n = \frac{1}{(1 - t/n\lambda)^n}$. This is precisely the mgf of the gamma distribution with $\alpha = n$ and $\beta = 1/(n\lambda)$, so by uniqueness of mgfs $\overline{X}$ has this distribution.

b. Use Equation (4.5): With $Y = \overline{X} \sim \text{Gamma}(n, 1/n\lambda)$,

$$E(\hat{\lambda}) = E\left(\frac{1}{Y}\right) = \int_0^\infty \frac{1}{y} \cdot \frac{1}{\Gamma(n)(1/n\lambda)^n} y^{n-1} e^{-y/[1/n\lambda]}\, dy = \frac{1}{\Gamma(n)(1/n\lambda)^n} \int_0^\infty y^{n-2} e^{-y/[1/n\lambda]}\, dy$$

$$= \frac{1}{\Gamma(n)(1/n\lambda)^n} \Gamma(n-1)(1/n\lambda)^{n-1} = \frac{\Gamma(n-1)(n\lambda)^n}{\Gamma(n)(n\lambda)^{n-1}} = \frac{n\lambda}{n-1}$$

In particular, since $n/(n-1) > 1$, $\hat{\lambda} = 1/\overline{X}$ is a biased-<u>high</u> estimator of λ.
Similarly,

$$E(\hat{\lambda}^2) = E\left(\frac{1}{Y^2}\right) = \frac{1}{\Gamma(n)(1/n\lambda)^n} \int_0^\infty y^{n-3} e^{-y/[1/n\lambda]}\, dy = \cdots = \frac{\Gamma(n-2)(n\lambda)^n}{\Gamma(n)(n\lambda)^{n-2}} = \frac{(n\lambda)^2}{(n-1)(n-2)},$$

from which $V(\hat{\lambda}) = E(\hat{\lambda}^2) - [E(\hat{\lambda})]^2 = \frac{(n\lambda)^2}{(n-1)(n-2)} - \left[\frac{n\lambda}{n-1}\right]^2 = \frac{n^2\lambda^2}{(n-1)^2(n-2)}$.

c. The standard error of $\hat{\lambda}$ is the square root of the variance expression from part **b**. Since that expression includes the unknown λ, we must estimate λ in the SE with $\hat{\lambda} = 1/\overline{x}$. The result is the estimated standard error

$$s_{\hat{\lambda}} = \sqrt{\frac{n^2\hat{\lambda}^2}{(n-1)^2(n-2)}} = \sqrt{\frac{n^2}{(n-1)^2(n-2)\overline{x}^2}}.$$

15. $\mu = E(X) = \int_{-1}^{1} x \cdot .5(1 + \theta x)dx = \left.\frac{x^2}{4} + \frac{\theta x^3}{6}\right|_{-1}^{1} = \frac{1}{3}\theta \Rightarrow \theta = 3\mu$. Hence,

$$\hat{\theta} = 3\overline{X} \Rightarrow E(\hat{\theta}) = E(3\overline{X}) = 3E(\overline{X}) = 3\mu = 3\left(\frac{1}{3}\right)\theta = \theta.$$

17.

a. $E(X^2) = 2\theta$ implies that $E\left(\frac{X^2}{2}\right) = \theta$. Consider $\hat{\theta} = \frac{\sum X_i^2}{2n}$. Then

$$E\left(\hat{\theta}\right) = E\left(\frac{\sum X_i^2}{2n}\right) = \frac{\sum E\left(X_i^2\right)}{2n} = \frac{\sum 2\theta}{2n} = \frac{2n\theta}{2n} = \theta$$, implying that $\hat{\theta}$ is an unbiased estimator for θ.

b. $\sum x_i^2 = 1490.1058$, so $\hat{\theta} = \frac{1490.1058}{2(10)} = 74.505$.

19.

a. $E(\hat{P}) = \sum_{x=r}^{\infty} \frac{r-1}{x-1} \cdot \binom{x-1}{r-1} \cdot p^r \cdot (1-p)^{x-r} = \sum_{x=r}^{\infty} \frac{(x-2)!}{(x-r)!(r-2)!} \cdot p^r \cdot (1-p)^{x-r} = \sum_{x=r}^{\infty} \binom{x-2}{r-2} \cdot p^r \cdot (1-p)^{x-r}$.

Make the suggested substitutions $y = x - 1$ and $s = r - 1$, i.e. $x = y + 1$ and $r = s + 1$:

$$E(\hat{P}) = \sum_{y=s}^{\infty} \binom{y-1}{s-1} p^{s+1} (1-p)^{y-s} = p \sum_{y=s}^{\infty} \binom{y-1}{s-1} p^{s} (1-p)^{y-s} = p \sum_{y=s}^{\infty} nb(y; s, p) = p \cdot 1 = p.$$

The last steps use the fact that the term inside the summation is the negative binomial pmf with parameters s and p, and all pmfs sum to 1.

b. For the given sequence, $x = 5$, so $\hat{p} = \frac{5-1}{5+5-1} = \frac{4}{9} = .444$.

21.

a. $\lambda = .5p + .15 \Rightarrow 2\lambda = p + .3$, so $p = 2\lambda - .3$ and $\hat{p} = 2\hat{\lambda} - .3 = 2\left(\frac{Y}{n}\right) - .3$; the estimate is

$2\left(\frac{20}{80}\right) - .3 = .2$.

b. $E(\hat{p}) = E\left(2\hat{\lambda} - .3\right) = 2E\left(\hat{\lambda}\right) - .3 = 2\lambda - .3 = p$, as desired.

c. Here $\lambda = .7p + (.3)(.3)$, so $p = \frac{10}{7}\lambda - \frac{9}{70}$ and $\hat{p} = \frac{10}{7}\left(\frac{Y}{n}\right) - \frac{9}{70}$.

23. As suggested, let $\mu = E(\hat{\theta})$ for notational ease. The left-hand side (the MSE) expands to

$E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 = E(\hat{\theta}^2) - 2\theta\mu + \theta^2$.

The right-hand side expands to

$V(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 = V(\hat{\theta}) + [\mu - \theta]^2 = E(\hat{\theta}^2) - [E(\hat{\theta})]^2 + \mu^2 - 2\mu\theta + \theta^2 = E(\hat{\theta}^2) - 2\theta\mu + \theta^2$.

These two expressions are the same, so the two original quantities are equal.

Section 7.2

25.

a. To find the mle of p, we'll take the derivative of the log-likelihood function $\ell(p) = \ln\left[\binom{n}{x} p^x (1-p)^{n-x}\right] = \ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p)$, set it equal to zero, and solve for p.

$\ell'(p) = \frac{d}{dp}\left[\ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p)\right] = \frac{x}{p} - \frac{n-x}{1-p} = 0 \Rightarrow x(1-p) = p(n-x) \Rightarrow p = x/n$, so the maximum likelihood estimator of p is $\hat{p} = \frac{X}{n}$, which is simply the sample proportion of successes. For $n = 20$ and $x = 3$, $\hat{p} = \frac{3}{20} = .15$.

b. Since X is binomial, $E(X) = np$, from which $E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$; thus, $\hat{p}$ is an unbiased estimator of p.

c. By the invariance principle, the mle of $(1-p)^5$ is just $(1-\hat{p})^5$. For $n = 20$ and $x = 3$, we have $(1 - .15)^5 = .4437$.

27.

a. $E(X) = \int_0^1 x(\theta+1)x^\theta dx = \frac{\theta+1}{\theta+2}$, so the moment estimator $\hat{\theta}$ is the solution to $\overline{X} = \frac{\hat{\theta}+1}{\hat{\theta}+2}$, yielding $\hat{\theta} = \frac{1}{1-\overline{X}} - 2$. Since $\overline{x} = .80, \hat{\theta} = 5 - 2 = 3$.

b. $f(x_1,...,x_n;\theta) = (\theta+1)^n (x_1x_2...x_n)^\theta$, so the log likelihood is $\ell(\theta) = n\ln(\theta+1) + \theta\sum\ln(x_i)$. Taking the derivative and equating to 0 yields $\frac{n}{\theta+1} = -\sum\ln(x_i)$, so $\hat{\theta} = -\frac{n}{\sum\ln(X_i)} - 1$. Taking $\ln(x_i)$ for each given x_i yields ultimately $\hat{\theta} = 3.12$.

29. The number of helmets examined, X, until r flawed helmets are found has a negative binomial distribution: $X \sim \text{NB}(r, p)$. To find the mle of p, we'll take the derivative of the log-likelihood function
$\ell(p) = \ln\left[\binom{x-1}{r-1} p^r (1-p)^{x-r}\right] = \ln\binom{x-1}{r-1} + r\ln(p) + (x-r)\ln(1-p)$, set it equal to zero, and solve for p.

$\ell'(p) = \frac{d}{dp}\left[\ln\binom{x-1}{r-1} + r\ln(p) + (x-r)\ln(1-p)\right] = \frac{r}{p} - \frac{x-r}{1-p} = 0 \Rightarrow r(1-p) = (x-r)p \Rightarrow p = r/x$, so the mle of p is $\hat{p} = \frac{r}{X}$. This is the number of successes over the total number of trials; with $r = 3$ and $x = 20$, $\hat{p} = .15$. Yes, this is the same as the mle of p based on the binomial model in Exercise 25.

In contrast, the unbiased estimator from Exercise 19 is $\hat{p} = \frac{r-1}{X-1}$, which is not the same as the maximum likelihood estimator. (With $r = 3$ and $x = 20$, the calculated value of the unbiased estimator is 2/19, rather than 3/20.)

31.

a. Since the X_i are independent, the likelihood function is

$$L(\theta) = f(x_1,...,x_n;\theta) = f(x_1;\theta)\cdots f(x_n;\theta) = \frac{1}{\sqrt{2\pi\theta}}e^{-x_1^2/2\theta}\cdots\frac{1}{\sqrt{2\pi\theta}}e^{-x_n^2/2\theta} = (2\pi\theta)^{-n/2}e^{-\Sigma x_i^2/2\theta} .$$

b. $\ell(\theta) = \ln[L(\theta)] = \ln\left[(2\pi\theta)^{-n/2}e^{-\Sigma x_i^2/2\theta}\right] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{\sum x_i^2}{2\theta}$.

c. $\ell'(\theta) = 0 - \frac{n}{2}\cdot\frac{1}{\theta} + \frac{\sum x_i^2}{2\theta^2} = 0 \Rightarrow \frac{n}{2\theta} = \frac{\sum x_i^2}{2\theta^2} \Rightarrow \theta = \frac{\sum x_i^2}{n}$. It's easy to show this is the local maximum of the log-likelihood function; hence, the mle of θ is $\hat{\theta} = \frac{\sum x_i^2}{n}$.

d. By the invariance principle, the mle of $\tau = 1/\theta$ is $\hat{\tau} = 1/\hat{\theta} = \frac{n}{\sum X_i^2}$.

33.

a. The likelihood function is $L(\theta) = f(x_1,...,x_n;\theta) = \prod_{i=1}^{n}\frac{x_i}{\theta}e^{-x_i^2/(2\theta)} = \frac{\Pi x_i}{\theta^n}e^{-\Sigma x_i^2/(2\theta)}$, so the log-likelihood function is $\ell(\theta) = \ln[L(\theta)] = \ln[\Pi x_i] - n\ln(\theta) - \frac{\Sigma x_i^2}{2\theta}$. To find the mle of θ, differentiate and set equal to zero: $0 = \ell'(\theta) = 0 - \frac{n}{\theta} + \frac{\Sigma x_i^2}{2\theta^2} \Rightarrow \theta = \frac{\Sigma x_i^2}{2n}$. Hence, the mle of θ is $\hat{\theta} = \frac{\Sigma x_i^2}{2n}$, identical to the unbiased estimator in Exercise 17. In particular, they share the same numerical value for the given data: $\hat{\theta}$ = 74.505.

b. The median η of the Rayleigh distribution satisfies $.5 = \int_0^{\eta}\frac{x}{\theta}e^{-x^2/(2\theta)}\,dx = -e^{-x^2/(2\theta)}\Big|_0^{\eta} = 1 - e^{-\eta^2/(2\theta)}$; solving for η gives $\eta = \sqrt{-2\ln(.5)\theta}$. (Since ln(.5) < 0, the quantity under the square root is positive.) By the invariance principle, the mle of η is $\hat{\eta} = \sqrt{-2\ln(.5)\hat{\theta}} = \sqrt{-\ln(.5)\Sigma x_i^2/n}$. For the given data, the maximum likelihood estimate of η is 10.163.

35. The likelihood is $f(y;n,p) = \binom{n}{y}p^y(1-p)^{n-y}$ where $p = P(X \ge 24) = 1 - \int_0^{24}\lambda e^{-\lambda x}dx = e^{-24\lambda}$. We know $\hat{p} = \frac{y}{n}$, so by the invariance principle $\hat{p} = e^{-24\hat{\lambda}} \Rightarrow \hat{\lambda} = -\frac{\ln\hat{p}}{24} = .0120$ for n = 20, y = 15.

37.

a. The pdf is symmetric about θ, so $E(X) = \theta$. Hence the mme of θ is $\hat{\theta} = \bar{X}$.

b. $L(\theta) = e^{-|x_1-\theta|} \cdots e^{-|x_n-\theta|} = e^{-\Sigma|x_i-\theta|}$. While this isn't a differentiable function with respect to θ, we can exploit the hint. The function $e^{-\Sigma|x_i-\theta|}$ is *maximized* precisely when $\Sigma|x_i - \theta|$ is *minimized* (because of the negative sign), and $\Sigma|x_i - \theta|$ is minimized by $\theta = \tilde{x}$. Therefore, the maximum likelihood estimator of θ is $\hat{\theta} = \tilde{X}$.

Section 7.3

39. Each $X_i \sim \text{Bin}(k, p)$ and they're independent, so $T \sim \text{Bin}(nk, p)$. The question is whether T is sufficient for p. Let's find out: $P(\mathbf{X} = (x_1,\ldots,x_n) \mid T = \Sigma x_i) = \dfrac{P(X_1 = x_1,\ldots,X_n = x_n)}{P(T = \Sigma x_i)} =$

$$\frac{\binom{k}{x_1}p^{x_1}q^{k-x_1}\cdots\binom{k}{x_n}p^{x_n}q^{k-x_n}}{\binom{nk}{\Sigma x_i}p^{\Sigma x_i}q^{nk-\Sigma x_i}} = \frac{\binom{k}{x_1}\cdots\binom{k}{x_n}p^{\Sigma x_i}q^{nk-\Sigma x_i}}{\binom{nk}{\Sigma x_i}p^{\Sigma x_i}q^{nk-\Sigma x_i}} = \frac{\binom{k}{x_1}\cdots\binom{k}{x_n}}{\binom{nk}{\Sigma x_i}}$$

. This conditional distribution does not depend on p, so T is sufficient for p. That is, statistician A really doesn't have more information about p than statistician B.

41. Re-write the joint pdf: $f(x_1,..,x_n;\alpha,\beta) = \prod_{i=1}^{n} \dfrac{x_i^{\alpha-1}e^{-x_i/\beta}}{\Gamma(\alpha)\beta^\alpha} = \dfrac{[\Pi x_i]^{\alpha-1}e^{-\Sigma x_i/\beta}}{[\Gamma(\alpha)\beta^\alpha]^n}$. Let $g(\Pi x_i, \Sigma x_i;\alpha,\beta)$ be this entire expression (and $h = 1$ vacuously). Then, by the factorization theorem, ΠX_i and $\sum X_i$ are jointly sufficient for α and β.

43. The joint pmf is $p(x_1,\ldots,x_n;p) = \prod \binom{x_i-1}{r-1} p^r(1-p)^{x_i-r} = \left[\prod \binom{x_i-1}{r-1}\right] p^{nr}(1-p)^{\Sigma x_i - nr}$. (Remember that both n and r are known.) Let $g(\Sigma x_i, p) = p^{nr}(1-p)^{\Sigma x_i - nr}$ and $h(x_1,\ldots,x_n) = \prod \binom{x_i-1}{r-1}$, which does not depend on p. Then, by the factorization theorem, $\sum X_i$ is sufficient for p.

45. Let $I(A)$ denote the indicator of an event. Then

$$f(x_1,\ldots,x_n;\theta_1) = \prod_{i=1}^{n} \frac{1}{2\theta-\theta} I(\theta < x_i < 2\theta) = \theta^{-n} I(\theta < x_1,\ldots,x_n < 2\theta)$$
$$= \theta^{-n} I(\theta < \min\{x_i\} \cap \max\{x_i\} < 2\theta)$$

Set g equal to this entire expression and $h = 1$. By the factorization theorem, $(\min\{X_i\}, \max\{X_i\})$ are jointly sufficient for θ_1 and θ_2.

47. Let Y be the number of items in your sample of 2 that work, so that $Y \sim \text{Bin}(2, p)$, and define $U = I(Y = 1)$. Then $E[U] = P(Y = 1) = 2pq$. Applying the Rao-Blackwell Theorem, condition on the sufficient statistic $X = x$ to give the improved estimator $U^* = E[U \mid X = x] = P(Y = 1 \mid X = x)$. Let's determine U^* explicitly.

$X - Y$ is the number of working items in the last $n - 2$ components, so $X - Y \sim \text{Bin}(n - 2, p)$ and $X - Y$ is independent of Y. Therefore,

$$U^* = P(Y=1 \mid X=x) = \frac{P(Y=1 \cap X=x)}{P(X=x)} = \frac{P(Y=1 \cap X-Y=x-1)}{P(X=x)}$$

$$= \frac{P(Y=1)P(X-Y=x-1)}{P(X=x)} \quad \text{independence}$$

$$= \frac{\binom{2}{1}p^1q^1\binom{n-2}{x-1}p^{x-1}q^{n-2-(x-1)}}{\binom{n}{x}p^xq^{n-x}} = \frac{\binom{2}{1}\binom{n-2}{x-1}}{\binom{n}{x}} = \frac{2x(n-x)}{n(n-1)}.$$

49. The Rao-Blackwell Theorem implies that a sufficient statistic has minimum variance among all unbiased estimators. Any statistic *not* purely a function of the sufficient statistic must necessarily have greater variance. Since $\overline{X}$ is sufficient for μ (while S^2 isn't), $\overline{X}$ must have the least variance among the unbiased estimators $\overline{X}$, S^2, and $\hat{\mu} = (\overline{X} + S^2)/2$. Notice we can determine this *without* knowing the variances of the last two estimators (which cannot be easily found)!

Section 7.4

51.

a. $f(x;p) = (1-p)^{x-1}p \Rightarrow \ell(p) = \ln[f(X;p)] = (X-1)\ln(1-p) + \ln(p) \Rightarrow \ell'(p) = -\dfrac{X-1}{1-p} + \dfrac{1}{p} \Rightarrow$

$\ell''(p) = -\dfrac{X-1}{(1-p)^2} - \dfrac{1}{p^2} \Rightarrow I(p) = E[-\ell''(p)] = \dfrac{E(X)-1}{(1-p)^2} + \dfrac{1}{p^2} = \dfrac{1/p-1}{(1-p)^2} + \dfrac{1}{p^2} = \dfrac{1}{p^2(1-p)}$. That's using the definition (7.5); using (7.6) instead, $I(p) = V(\ell'(p)) = V\left(-\dfrac{X-1}{1-p} + \dfrac{1}{p}\right) = \left(-\dfrac{1}{1-p}\right)^2 V(X) =$ $\dfrac{1}{(1-p)^2} \cdot \dfrac{1-p}{p^2} = \dfrac{1}{p^2(1-p)}$. In this case, (7.6) is more straightforward.

b. By the additive principle of information, $I_n(p) = n \cdot I(p) = \dfrac{n}{p^2(1-p)}$.

c. The C-R lower bound for the variance of an unbiased estimator of p is $\dfrac{1}{I_n(p)} = \dfrac{p^2(1-p)}{n}$.

53.

a. If we <u>ignore the boundary</u> and say $f(x; \theta) = 1/\theta$, then
$\ell(\theta) = -\ln(\theta), \ell'(\theta) = -1/\theta$, and $I(\theta) = E[(\ell'(\theta))^2] = E[1/\theta^2] = 1/\theta^2$.

b. The Cramér-Rao lower bound is $\dfrac{1}{nI(\theta)} = \dfrac{\theta^2}{n}$.

c. $\theta^2/[n(n+2)] < \theta^2/n$ and $\theta^2/(3n) < \theta^2/n$. This does not violate the Cramér-Rao theorem, however, because the boundaries of the uniform variable X include θ itself! In these circumstances, Fisher information is not well-defined, and the theorem does not apply. (Note, for example, that if we used the $V(\ell'(\theta))$ version of Fisher information in **a**, we'd get zero because $\ell'(\theta)$ is constant.)

55.

a. With σ known, $\ell(\mu) = C - \sum(x_i - \mu)^2/2\sigma^2$, so $\ell'(\mu) = 2\sum(x_i - \mu)^1/2\sigma^2 = 0 \Rightarrow$
$\sum(x_i - \mu) = \sum x_i - n\mu = 0 \Rightarrow \mu = \sum x_i/n = \bar{x}$ (unsurprisingly).

b. Since the original X's are normal, we know that $\bar{X}$ is normal, with mean μ and variance σ^2/n.

c. For a single observation, our work in **a** shows that $\ell'(\mu) = (X - \mu)/\sigma^2$, so
$I(\mu) = V((X-\mu)/\sigma^2) = V(X)/\sigma^4 = \sigma^2/\sigma^4 = 1/\sigma^2$. Hence the Cramér-Rao lower bound is
$\dfrac{1}{nI(\mu)} = \dfrac{\sigma^2}{n}$, which is precisely $V(\bar{X})$. So $\bar{X}$ is indeed efficient.

d. The answer to **b** and the suggested asymptotic distribution agree.

57.

a. In terms of x and σ, $\ln[f(x; \sigma)] = C - \ln(\sigma) - (x-\mu)^2/2\sigma^2 \Rightarrow \ell'(\sigma) = -1/\sigma + (X-\mu)^2/\sigma^3 \Rightarrow$
$\ell''(\sigma) = 1/\sigma^2 - 3(X-\mu)^2/\sigma^4 \Rightarrow I(\sigma) = -1/\sigma^2 + 3E[(X-\mu)^2]/\sigma^4 = -1/\sigma^2 + 3\sigma^2/\sigma^4 = 2/\sigma^2$.

b. Yes, Fisher information does depend on the parameterization: the answer in **a** is different from the answer, $1/(2\sigma^4)$, from the previous exercise.

59.

a. For the geometric distribution, $\mu = 1/p$ and $\sigma^2 = (1-p)/p^2$. Thus $E(\bar{X}) = \mu = 1/p$ and
$V(\bar{X}) = \sigma^2/n = (1-p)/np^2$.

b. From Exercise 51, Fisher information from a random sample is $n/p^2(1-p)$. For any statistic whose expectation is $h(p) = 1/p$, the Cramér-Rao lower bound on the variance is given by
$\dfrac{[h'(p)]^2}{I_n(p)} = \dfrac{[-1/p^2]^2}{n/p^2(1-p)} = \dfrac{1}{p^4}\cdot\dfrac{p^2(1-p)}{n} = \dfrac{(1-p)}{np^2}$.

c. Yes: Since $V(\bar{X})$ exactly matches the Cramér-Rao lower bound from part **b**, $\bar{X}$ is an efficient estimator of $1/p$.

Supplementary Exercises

61. Let x_1 = the time until the first birth, x_2 = the elapsed time between the first and second births, and so on. Then $f(x_1,...,x_n;\lambda) = \lambda e^{-\lambda x_1}\cdot(2\lambda)e^{-2\lambda x_2}...(n\lambda)e^{-n\lambda x_n} = n!\lambda^n e^{-\lambda\Sigma k x_k}$. Thus the log likelihood is $\ln(n!) + n\ln(\lambda) - \lambda\Sigma k x_k$. Taking $\frac{d}{d\lambda}$ and equating to 0 yields $\hat{\lambda} = \frac{n}{\Sigma k x_k}$.

For the given sample, n = 6, x_1 = 25.2, x_2 = 41.7 – 25.2 = 16.5, x_3 = 9.5, x_4 = 4.3, x_5 = 4.0, x_6 = 2.3; so $\sum_{k=1}^{6} k x_k = (1)(25.2) + (2)(16.5) + ... + (6)(2.3) = 137.7$ and $\hat{\lambda} = \frac{6}{137.7} = .0436$.

63. The first moment of the Beta distribution is $E(X) = \frac{\alpha}{\alpha+\beta}$, while the second moment is more complicated:

$E(X^2) = V(X) + [E(X)]^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \left[\frac{\alpha}{\alpha+\beta}\right]^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$. The first and second *sample* moments are $\bar{X} = \frac{1}{6}(.873 + \cdots + .618) = .565$ and $\frac{1}{n}\sum x_i^2 = \frac{1}{6}(.873^2 + \cdots + .618^2) = .359161$.

To determine the method of moments estimates, let $c = \alpha + \beta$ and solve $\alpha / c = .565$, $\alpha(\alpha+1)/c(c+1)$ = .359161. The solutions are α = 2.912 and c = 5.154, from which $\beta = c - \alpha$ = 2.242. Therefore, the method of moments estimates are $\hat{\alpha}$ = 2.912 and $\hat{\beta}$ = 2.242.

65. Example 7.8 shows that $E(\hat{\sigma}^2) = c(n-1)\sigma^2$ and $V(\hat{\sigma}^2) = 2c^2(n-1)\sigma^4$. From these,

$$\begin{aligned}\text{MSE}(\hat{\sigma}^2) &= V(\hat{\sigma}^2) + [E(\hat{\sigma}^2) - \sigma^2]^2 = 2c^2(n-1)\sigma^4 + [c(n-1)\sigma^2 - \sigma^2]^2 \\ &= 2c^2(n-1)\sigma^4 + c^2(n-1)^2\sigma^4 - 2c(n-1)\sigma^4 + \sigma^4 \\ &= [(n^2-1)c^2 - 2(n-1)c + 1]\sigma^4\end{aligned}$$

To minimize the MSE, differentiate the expression in brackets with respect to c and solve for c:

$2(n^2-1)c^1 - 2(n-1) + 0 = 0 \Rightarrow c = \frac{2(n-1)}{2(n^2-1)} = \frac{1}{n+1}$, as claimed.

67. The median of the 16 values in Example 7.2 is $\tilde{x}$ = 985. The values of $|x_i - \tilde{x}|$ are 29, 11, 5, 5, 3, 2, 2, 0, 0, 0, 2, 2, 10, 14, 15, 22. When these 16 values are sorted, the middle two are 3 and 5, so the median *of these absolute differences* is 4, and $\hat{\sigma}$ = 4/.6745 = 5.93. The sample standard deviation of the original 16 values is substantially larger, at s = 11.66. (The unusually low value 956 may be affecting s.)

69.

a. The likelihood is $\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\mu_i)}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_i-\mu_i)}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^n}e^{-\frac{(\Sigma(x_i-\mu_i)^2+\Sigma(y_i-\mu_i)^2)}{2\sigma^2}}$. The log likelihood is thus $-n\ln(2\pi\sigma^2) - \frac{(\Sigma(x_i-\mu_i)^2+\Sigma(y_i-\mu_i)^2)}{2\sigma^2}$. Taking $\frac{d}{d\mu_i}$ and equating to zero gives $\hat{\mu}_i = \frac{x_i+y_i}{2}$.

Substituting these estimates of the $\hat{\mu}_i$'s into the log likelihood gives

$-n\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\left(\sum\left(x_i-\frac{x_i+y_i}{2}\right)^2+\sum\left(y_i-\frac{x_i+y_i}{2}\right)^2\right)=-n\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\left(\frac{1}{2}\Sigma\left(x_i-y_i\right)^2\right)$. Now taking $\frac{d}{d\sigma^2}$, equating to zero, and solving for σ^2 gives the desired result.

b. $E\left(\hat{\sigma}^2\right)=\frac{1}{4n}E\left(\Sigma\left(X_i-Y_i\right)^2\right)=\frac{1}{4n}\cdot\Sigma E\left(X_i-Y_i\right)^2$, but
$E\left(X_i-Y_i\right)^2=V\left(X_i-Y_i\right)+\left[E\left(X_i-Y_i\right)\right]^2=2\sigma^2+0=2\sigma^2$. Thus
$E\left(\hat{\sigma}^2\right)=\frac{1}{4n}\Sigma\left(2\sigma^2\right)=\frac{1}{4n}2n\sigma^2=\frac{\sigma^2}{2}$, so the mle is definitely not unbiased; in fact, the expected value of the estimator is only half the value of what is being estimated! An unbiased estimator is easily obtained: $2\hat{\sigma}^2$.

71. Given $Y = y$, the investigator must have tested y individuals among whom r are allergic and $y - r$ aren't. Let **x** be any sequence of y 0's and 1's with exactly $(y - r)$ 0's and r 1's. Then
$P((X_1,\ldots,X_y)=\mathbf{x}\mid Y=y)=\frac{P((X_1,\ldots,X_y)=\mathbf{x})}{P(Y=y)}$. By independence of the X_i's, the numerator is just the product of exactly $(y - r)$ q's and exactly r p's. The denominator is a negative binomial probability. Continuing,
$P((X_1,\ldots,X_y)=\mathbf{x}\mid Y=y)=\frac{p^rq^{y-r}}{\binom{y-1}{r-1}p^rq^{y-r}}=\frac{1}{\binom{y-1}{r-1}}$, which does not depend on p. Therefore, by definition Y is sufficient for p. Knowing the order in which allergy and non-allergy sufferers arrive does not help estimate p.

73. Be careful here: $\hat{\sigma}$ is the MLE of σ and *not* the sample standard deviation! In other words, use $n = 3$ rather than $n - 1 = 2$ in your denominator. With the information provided, $c = 150$, $\hat{\mu}=\bar{x}$ = 150.40, $\hat{\sigma}$ = 3.06, $k = \sqrt{3/2}$, $w = -.1307$, and $kw = -.16$. Hence, the MVUE for θ is
$P(T < -.16(1)/\sqrt{1-(-.16)^2}\ ) = P(T < -.1621)$, where $T \sim t_1$. Software gives .448 for this probability.
In contrast, we may also write $\theta = P(X \le c) = \Phi((c - \mu)/\sigma)$, from which, by the invariance principle, the MLE of θ is $\Phi((c-\hat{\mu})/\hat{\sigma}) = \Phi(w) = \Phi(-.16) = .4364$.

75. $E[d(X)]=\sum d(x)\frac{e^{-\mu}\mu^x}{x!}=e^{-2\mu}\Rightarrow\sum\frac{d(x)\mu^x}{x!}=e^{-\mu}\Rightarrow\sum\frac{d(x)\mu^x}{x!}=\sum\frac{(-\mu)^x}{x!}$. From the uniqueness of Taylor series, these can only be equal if $d(X) = (-1)^X$. While unbiased, this estimator is ridiculous: if X happens to be even, we estimate the probability θ to be 1 (no matter whether $X = 0$ or $X = 200$). If X happens to be odd, we estimate θ to be -1!!

77.

a. The points do not fall <u>perfectly</u> on a straight line through the origin, but they come very close to fitting the line $y = 30x$.

b. The joint pdf here is $(2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2\right]$, and so the log-likelihood function is $-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2 = C - n\ln(\sigma)-\frac{\sum(y_i-\beta x_i)^2}{2\sigma^2}$. First, differentiate with respect to β and solve: $\frac{\sum 2(y_i-\beta x_i)(-\beta x_i)}{2\sigma^2}=0 \Rightarrow \hat{\beta}=\frac{\Sigma x_i y_i}{\Sigma x_i^2}$. Next, differentiate with respect to σ and solve: $-\frac{n}{\sigma}+\frac{\sum(y_i-\beta x_i)^2}{\sigma^3} \Rightarrow \hat{\sigma}^2=\frac{\sum(y_i-\hat{\beta}x_i)^2}{n}$. For the data provided, $\hat{\beta} = 30.040$, the estimated minutes per item, and $\hat{\sigma}^2=\frac{1}{n}\sum(y_i-\hat{\beta}x_i)^2 = 16.912$. When $x = 25$, we predict $y = \hat{\beta}(25)$ = 571.