Solutions Manual for
Statistical Inference, Second Edition

George Casella
University of Florida

Roger L. Berger
North Carolina State University

Damaris Santana
University of Florida

"When I hear you give your reasons," I remarked, "the thing always appears to me to be so ridiculously simple that I could easily do it myself, though at each successive instance of your reasoning I am baffled until you explain your process."

Dr. Watson to Sherlock Holmes
A Scandal in Bohemia

0.1 Description

This solutions manual contains solutions for *all odd numbered problems* plus a large number of solutions for even numbered problems. Of the 624 exercises in *Statistical Inference, Second Edition*, this manual gives solutions for 484 (78%) of them. There is an obtuse pattern as to which solutions were included in this manual. We assembled all of the solutions that we had from the first edition, and filled in so that all odd-numbered problems were done. In the passage from the first to the second edition, problems were shuffled with no attention paid to numbering (hence no attention paid to minimize the new effort), but rather we tried to put the problems in logical order.

A major change from the first edition is the use of the computer, both symbolically through Mathematica^{tm} and numerically using R. Some solutions are given as code in either of these languages. Mathematica^{tm} can be purchased from Wolfram Research, and R is a free download from `http://www.r-project.org/`.

Here is a detailed listing of the solutions included.

Chapter	**Number of Exercises**	**Number of Solutions**	**Missing**
1	55	51	26, 30, 36, 42
2	40	37	34, 38, 40
3	50	42	4, 6, 10, 20, 30, 32, 34, 36
4	65	52	8, 14, 22, 28, 36, 40 48, 50, 52, 56, 58, 60, 62
5	69	46	2, 4, 12, 14, 26, 28 all even problems from 36 − 68
6	43	35	8, 16, 26, 28, 34, 36, 38, 42
7	66	52	4, 14, 16, 28, 30, 32, 34, 36, 42, 54, 58, 60, 62, 64
8	58	51	36, 40, 46, 48, 52, 56, 58
9	58	41	2, 8, 10, 20, 22, 24, 26, 28, 30 32, 38, 40, 42, 44, 50, 54, 56
10	48	26	all even problems except 4 and 32
11	41	35	4, 20, 22, 24, 26, 40
12	31	16	all even problems

0.2 Acknowledgement

Many people contributed to the assembly of this solutions manual. We again thank all of those who contributed solutions to the first edition – many problems have carried over into the second edition. Moreover, throughout the years a number of people have been in constant touch with us, contributing to both the presentations and solutions. We apologize in advance for those we forget to mention, and we especially thank Jay Beder, Yong Sung Joo, Michael Perlman, Rob Strawderman, and Tom Wehrly. Thank you all for your help.

And, as we said the first time around, although we have benefited greatly from the assistance and comments of others in the assembly of this manual, we are responsible for its ultimate correctness. To this end, we have tried our best but, as a wise man once said, “You pays your money and you takes your chances.”

George Casella
Roger L. Berger
Damaris Santana
December, 2001

Chapter 1

Probability Theory

"If any little problem comes your way, I shall be happy, if I can, to give you a hint or two as to its solution."

Sherlock Holmes
The Adventure of the Three Students

1.1 a. Each sample point describes the result of the toss (H or T) for each of the four tosses. So, for example THTT denotes T on 1st, H on 2nd, T on 3rd and T on 4th. There are $2^4 = 16$ such sample points.

b. The number of damaged leaves is a nonnegative integer. So we might use $S = \{0, 1, 2, \ldots\}$.

c. We might observe fractions of an hour. So we might use $S = \{t : t \geq 0\}$, that is, the half infinite interval $[0, \infty)$.

d. Suppose we weigh the rats in ounces. The weight must be greater than zero so we might use $S = (0, \infty)$. If we know no 10-day-old rat weighs more than 100 oz., we could use $S = (0, 100]$.

e. If n is the number of items in the shipment, then $S = \{0/n, 1/n, \ldots, 1\}$.

1.2 For each of these equalities, you must show containment in both directions.

a. $x \in A\backslash B \Leftrightarrow x \in A$ and $x \notin B \Leftrightarrow x \in A$ and $x \notin A \cap B \Leftrightarrow x \in A\backslash(A \cap B)$. Also, $x \in A$ and $x \notin B \Leftrightarrow x \in A$ and $x \in B^c \Leftrightarrow x \in A \cap B^c$.

b. Suppose $x \in B$. Then either $x \in A$ or $x \in A^c$. If $x \in A$, then $x \in B \cap A$, and, hence $x \in (B \cap A) \cup (B \cap A^c)$. Thus $B \subset (B \cap A) \cup (B \cap A^c)$. Now suppose $x \in (B \cap A) \cup (B \cap A^c)$. Then either $x \in (B \cap A)$ or $x \in (B \cap A^c)$. If $x \in (B \cap A)$, then $x \in B$. If $x \in (B \cap A^c)$, then $x \in B$. Thus $(B \cap A) \cup (B \cap A^c) \subset B$. Since the containment goes both ways, we have $B = (B \cap A) \cup (B \cap A^c)$. (Note, a more straightforward argument for this part simply uses the Distributive Law to state that $(B \cap A) \cup (B \cap A^c) = B \cap (A \cup A^c) = B \cap S = B$.)

c. Similar to part a).

d. From part b).
$A \cup B = A \cup [(B \cap A) \cup (B \cap A^c)] = A \cup (B \cap A) \cup A \cup (B \cap A^c) = A \cup [A \cup (B \cap A^c)] = A \cup (B \cap A^c)$.

1.3 a. $x \in A \cup B \;\Leftrightarrow\; x \in A$ or $x \in B \;\Leftrightarrow\; x \in B \cup A$

$x \in A \cap B \;\Leftrightarrow\; x \in A$ and $x \in B \;\Leftrightarrow\; x \in B \cap A$.

b. $x \in A \cup (B \cup C) \Leftrightarrow\; x \in A$ or $x \in B \cup C \Leftrightarrow\; x \in A \cup B$ or $x \in C \Leftrightarrow\; x \in (A \cup B) \cup C$.

(It can similarly be shown that $A \cup (B \cup C) = (A \cup C) \cup B$.)

$x \in A \cap (B \cap C) \Leftrightarrow\; x \in A$ and $x \in B$ and $x \in C \Leftrightarrow\; x \in (A \cap B) \cap C$.

c. $x \in (A \cup B)^c \Leftrightarrow\; x \notin A$ or $x \notin B \Leftrightarrow\; x \in A^c$ and $x \in B^c \Leftrightarrow\; x \in A^c \cap B^c$

$x \in (A \cap B)^c \Leftrightarrow\; x \notin A \cap B \Leftrightarrow\; x \notin A$ and $x \notin B \Leftrightarrow\; x \in A^c$ or $x \in B^c \Leftrightarrow\; x \in A^c \cup B^c$.

1.4 a. "A or B or both" is $A \cup B$. From Theorem 1.2.9b we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

b. "A or B but not both" is $(A \cap B^c) \cup (B \cap A^c)$. Thus we have

$$\begin{aligned} P((A \cap B^c) \cup (B \cap A^c)) &= P(A \cap B^c) + P(B \cap A^c) && \text{(disjoint union)} \\ &= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] && \text{(Theorem1.2.9a)} \\ &= P(A) + P(B) - 2P(A \cap B). \end{aligned}$$

c. "At least one of A or B" is $A \cup B$. So we get the same answer as in a).

d. "At most one of A or B" is $(A \cap B)^c$, and $P((A \cap B)^c) = 1 - P(A \cap B)$.

1.5 a. $A \cap B \cap C = \{$a U.S. birth results in identical twins that are female$\}$

b. $P(A \cap B \cap C) = \frac{1}{90} \times \frac{1}{3} \times \frac{1}{2}$

1.6

$$p_0 = (1-u)(1-w), \quad p_1 = u(1-w) + w(1-u), \quad p_2 = uw,$$

$$\begin{aligned} p_0 = p_2 &\quad\Rightarrow\quad u + w = 1 \\ p_1 = p_2 &\quad\Rightarrow\quad uw = 1/3. \end{aligned}$$

These two equations imply $u(1-u) = 1/3$, which has no solution in the real numbers. Thus, the probability assignment is not legitimate.

1.7 a.

$$P(\text{scoring } i \text{ points}) = \begin{cases} 1 - \frac{\pi r^2}{A} & \text{if } i = 0 \\ \frac{\pi r^2}{A}\left[\frac{(6-i)^2-(5-i)^2}{5^2}\right] & \text{if } i = 1, \ldots, 5. \end{cases}$$

b.

$$\begin{aligned} P(\text{scoring } i \text{ points}|\text{board is hit}) &= \frac{P(\text{scoring } i \text{ points} \cap \text{board is hit})}{P(\text{board is hit})} \\ P(\text{board is hit}) &= \frac{\pi r^2}{A} \\ P(\text{scoring } i \text{ points} \cap \text{board is hit}) &= \frac{\pi r^2}{A}\left[\frac{(6-i)^2 - (5-i)^2}{5^2}\right] \qquad i = 1, \ldots, 5. \end{aligned}$$

Therefore,

$$P(\text{scoring } i \text{ points}|\text{board is hit}) = \frac{(6-i)^2 - (5-i)^2}{5^2} \qquad i = 1, \ldots, 5$$

which is exactly the probability distribution of Example 1.2.7.

1.8 a. $P(\text{scoring exactly } i \text{ points}) = P(\text{inside circle } i) - P(\text{inside circle } i+1)$. Circle i has radius $(6-i)r/5$, so

$$P(s\text{scoring exactly } i \text{ points}) = \frac{\pi (6-i)^2 r^2}{5^2 \pi r^2} - \frac{\pi\left((6-(i+1))\right)^2 r^2}{5^2 \pi r^2} = \frac{(6-i)^2 - (5-i)^2}{5^2}.$$

b. Expanding the squares in part a) we find $P(\text{scoring exactly } i \text{ points}) = \frac{11-2i}{25}$, which is decreasing in i.

c. Let $P(i) = \frac{11-2i}{25}$. Since $i \leq 5$, $P(i) \geq 0$ for all i. $P(S) = P(\text{hitting the dartboard}) = 1$ by definition. Lastly, $P(i \cup j) = \text{area of } i \text{ ring} + \text{area of } j \text{ ring} = P(i) + P(j)$.

1.9 a. Suppose $x \in (\cup_\alpha A_\alpha)^c$, by the definition of complement $x \notin \cup_\alpha A_\alpha$, that is $x \notin A_\alpha$ for all $\alpha \in \Gamma$. Therefore $x \in A_\alpha^c$ for all $\alpha \in \Gamma$. Thus $x \in \cap_\alpha A_\alpha^c$ and, by the definition of intersection $x \in A_\alpha^c$ for all $\alpha \in \Gamma$. By the definition of complement $x \notin A_\alpha$ for all $\alpha \in \Gamma$. Therefore $x \notin \cup_\alpha A_\alpha$. Thus $x \in (\cup_\alpha A_\alpha)^c$.

b. Suppose $x \in (\cap_\alpha A_\alpha)^c$, by the definition of complement $x \notin (\cap_\alpha A_\alpha)$. Therefore $x \notin A_\alpha$ for some $\alpha \in \Gamma$. Therefore $x \in A_\alpha^c$ for some $\alpha \in \Gamma$. Thus $x \in \cup_\alpha A_\alpha^c$ and, by the definition of union, $x \in A_\alpha^c$ for some $\alpha \in \Gamma$. Therefore $x \notin A_\alpha$ for some $\alpha \in \Gamma$. Therefore $x \notin \cap_\alpha A_\alpha$. Thus $x \in (\cap_\alpha A_\alpha)^c$.

1.10 For $A_1, \ldots, A_n$

$$(i) \quad \left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c \qquad (ii) \quad \left(\bigcap_{i=1}^n A_i\right)^c = \bigcup_{i=1}^n A_i^c$$

Proof of (i): If $x \in (\cup A_i)^c$, then $x \notin \cup A_i$. That implies $x \notin A_i$ for any i, so $x \in A_i^c$ for every i and $x \in \cap A_i$.

Proof of (ii): If $x \in (\cap A_i)^c$, then $x \notin \cap A_i$. That implies $x \in A_i^c$ for some i, so $x \in \cup A_i^c$.

1.11 We must verify each of the three properties in Definition 1.2.1.

a. (1) The empty set $\emptyset \in \{\emptyset, S\}$. Thus $\emptyset \in \mathcal{B}$. (2) $\emptyset^c = S \in \mathcal{B}$ and $S^c = \emptyset \in \mathcal{B}$. (3) $\emptyset \cup S = S \in \mathcal{B}$.

b. (1) The empty set $\emptyset$ is a subset of any set, in particular, $\emptyset \subset S$. Thus $\emptyset \in \mathcal{B}$. (2) If $A \in \mathcal{B}$, then $A \subset S$. By the definition of complementation, A^c is also a subset of S, and, hence, $A^c \in \mathcal{B}$. (3) If $A_1, A_2, \ldots \in \mathcal{B}$, then, for each i, $A_i \subset S$. By the definition of union, $\cup A_i \subset S$. Hence, $\cup A_i \in \mathcal{B}$.

c. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be the two sigma algebras. (1) $\emptyset \in \mathcal{B}_1$ and $\emptyset \in \mathcal{B}_2$ since $\mathcal{B}_1$ and $\mathcal{B}_2$ are sigma algebras. Thus $\emptyset \in \mathcal{B}_1 \cap \mathcal{B}_2$. (2) If $A \in \mathcal{B}_1 \cap \mathcal{B}_2$, then $A \in \mathcal{B}_1$ and $A \in \mathcal{B}_2$. Since $\mathcal{B}_1$ and $\mathcal{B}_2$ are both sigma algebra $A^c \in \mathcal{B}_1$ and $A^c \in \mathcal{B}_2$. Therefore $A^c \in \mathcal{B}_1 \cap \mathcal{B}_2$. (3) If $A_1, A_2, \ldots \in \mathcal{B}_1 \cap \mathcal{B}_2$, then $A_1, A_2, \ldots \in \mathcal{B}_1$ and $A_1, A_2, \ldots \in \mathcal{B}_2$. Therefore, since $\mathcal{B}_1$ and $\mathcal{B}_2$ are both sigma algebra, $\cup_{i=1}^\infty A_i \in \mathcal{B}_1$ and $\cup_{i=1}^\infty A_i \in \mathcal{B}_2$. Thus $\cup_{i=1}^\infty A_i \in \mathcal{B}_1 \cap \mathcal{B}_2$.

1.12 First write

$$\begin{aligned}
P\left(\bigcup_{i=1}^\infty A_i\right) &= P\left(\bigcup_{i=1}^n A_i \cup \bigcup_{i=n+1}^\infty A_i\right) \\
&= P\left(\bigcup_{i=1}^n A_i\right) + P\left(\bigcup_{i=n+1}^\infty A_i\right) && (A_i\text{s are disjoint}) \\
&= \sum_{i=1}^n P(A_i) + P\left(\bigcup_{i=n+1}^\infty A_i\right) && (\text{finite additivity})
\end{aligned}$$

Now define $B_k = \bigcup_{i=k}^\infty A_i$. Note that $B_{k+1} \subset B_k$ and $B_k \to \phi$ as $k \to \infty$. (Otherwise the sum of the probabilities would be infinite.) Thus

$$P\left(\bigcup_{i=1}^\infty A_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^\infty A_i\right) = \lim_{n\to\infty}\left[\sum_{i=1}^n P(A_i) + P(B_{n+1})\right] = \sum_{i=1}^\infty P(A_i).$$

1.13 If A and B are disjoint, $P(A \cup B) = P(A) + P(B) = \frac{1}{3} + \frac{3}{4} = \frac{13}{12}$, which is impossible. More generally, if A and B are disjoint, then $A \subset B^c$ and $P(A) \le P(B^c)$. But here $P(A) > P(B^c)$, so A and B cannot be disjoint.

1.14 If $S = \{s_1, \ldots, s_n\}$, then any subset of S can be constructed by either including or excluding s_i, for each i. Thus there are 2^n possible choices.

1.15 Proof by induction. The proof for $k = 2$ is given after Theorem 1.2.14. Assume true for k, that is, the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways. For $k+1$, the $k+1$th task can be done in n_{k+1} ways, and for each one of these ways we can complete the job by performing the remaining k tasks. Thus for each of the n_{k+1} we have $n_1 \times n_2 \times \cdots \times n_k$ ways of completing the job by the induction hypothesis. Thus, the number of ways we can do the job is $\underbrace{(1 \times (n_1 \times n_2 \times \cdots \times n_k)) + \cdots + (1 \times (n_1 \times n_2 \times \cdots \times n_k))}_{n_{k+1}\text{terms}} = n_1 \times n_2 \times \cdots \times n_k \times n_{k+1}$.

1.16 a) 26^3. b) $26^3 + 26^2$. c) $26^4 + 26^3 + 26^2$.

1.17 There are $\binom{n}{2} = n(n-1)/2$ pieces on which the two numbers do not match. (Choose 2 out of n numbers without replacement.) There are n pieces on which the two numbers match. So the total number of different pieces is $n + n(n-1)/2 = n(n+1)/2$.

1.18 The probability is $\frac{\binom{n}{2}n!}{n^n} = \frac{(n-1)(n-1)!}{2n^{n-2}}$. There are many ways to obtain this. Here is one. The denominator is n^n because this is the number of ways to place n balls in n cells. The numerator is the number of ways of placing the balls such that exactly one cell is empty. There are n ways to specify the empty cell. There are $n-1$ ways of choosing the cell with two balls. There are $\binom{n}{2}$ ways of picking the 2 balls to go into this cell. And there are $(n-2)!$ ways of placing the remaining $n-2$ balls into the $n-2$ cells, one ball in each cell. The product of these is the numerator $n(n-1)\binom{n}{2}(n-2)! = \binom{n}{2}n!$.

1.19 a. $\binom{6}{4} = 15$.

b. Think of the n variables as n bins. Differentiating with respect to one of the variables is equivalent to putting a ball in the bin. Thus there are r unlabeled balls to be placed in n unlabeled bins, and there are $\binom{n+r-1}{r}$ ways to do this.

1.20 A sample point specifies on which day (1 through 7) each of the 12 calls happens. Thus there are 7^{12} equally likely sample points. There are several different ways that the calls might be assigned so that there is at least one call each day. There might be 6 calls one day and 1 call each of the other days. Denote this by 6111111. The number of sample points with this pattern is $7\binom{12}{6}6!$. There are 7 ways to specify the day with 6 calls. There are $\binom{12}{6}$ to specify which of the 12 calls are on this day. And there are 6! ways of assigning the remaining 6 calls to the remaining 6 days. We will now count another pattern. There might be 4 calls on one day, 2 calls on each of two days, and 1 call on each of the remaining four days. Denote this by 4221111. The number of sample points with this pattern is $7\binom{12}{4}\binom{6}{2}\binom{8}{2}\binom{6}{2}4!$. (7 ways to pick day with 4 calls, $\binom{12}{4}$ to pick the calls for that day, $\binom{6}{2}$ to pick two days with two calls, $\binom{8}{2}$ ways to pick two calls for lowered numbered day, $\binom{6}{2}$ ways to pick the two calls for higher numbered day, 4! ways to order remaining 4 calls.) Here is a list of all the possibilities and the counts of the sample points for each one.

pattern	number of sample points	
6111111	$7\binom{12}{6}6! =$	4,656,960
5211111	$7\binom{12}{5}6\binom{7}{2}5! =$	83,825,280
4221111	$7\binom{12}{4}\binom{6}{2}\binom{8}{2}\binom{6}{2}4! =$	523,908,000
4311111	$7\binom{12}{4}6\binom{8}{3}5! =$	139,708,800
3321111	$\binom{7}{2}\binom{12}{3}\binom{9}{3}5\binom{6}{2}4! =$	698,544,000
3222111	$7\binom{12}{3}\binom{6}{3}\binom{9}{3}\binom{7}{2}\binom{5}{2}3! =$	1,397,088,000
2222211	$\binom{7}{5}\binom{12}{2}\binom{10}{2}\binom{8}{2}\binom{6}{2}\binom{4}{2}2! =$	314,344,800
		3,162,075,840

The probability is the total number of sample points divided by 7^{12}, which is $\frac{3{,}162{,}075{,}840}{7^{12}} \approx .2285$.

1.21 The probability is $\frac{\binom{n}{2r}2^{2r}}{\binom{2n}{2r}}$. There are $\binom{2n}{2r}$ ways of choosing $2r$ shoes from a total of $2n$ shoes. Thus there are $\binom{2n}{2r}$ equally likely sample points. The numerator is the number of sample points for which there will be no matching pair. There are $\binom{n}{2r}$ ways of choosing $2r$ different shoes styles. There are two ways of choosing within a given shoe style (left shoe or right shoe), which gives 2^{2r} ways of arranging each one of the $\binom{n}{2r}$ arrays. The product of this is the numerator $\binom{n}{2r}2^{2r}$.

1.22 a) $\frac{\binom{31}{15}\binom{29}{15}\binom{31}{15}\binom{30}{15}\cdots\binom{31}{15}}{\binom{366}{180}}$ b) $\frac{\frac{336}{366}\frac{335}{365}\cdots\frac{316}{336}}{\binom{366}{30}}$.

1.23

$$\begin{aligned} P(\text{ same number of heads }) &= \sum_{x=0}^{n} P(1^{st} \text{ tosses } x, 2^{nd} \text{ tosses } x) \\ &= \sum_{x=0}^{n}\left[\binom{n}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{n-x}\right]^2 = \left(\frac{1}{4}\right)^{n}\sum_{x=0}^{n}\binom{n}{x}^2. \end{aligned}$$

1.24 a.

$$\begin{aligned} P(A \text{ wins}) &= \sum_{i=1}^{\infty} P(A \text{ wins on } i^{th} \text{ toss}) \\ &= \frac{1}{2}+\left(\frac{1}{2}\right)^2\frac{1}{2}+\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)+\cdots = \sum_{i=0}^{\infty}\left(\frac{1}{2}\right)^{2i+1} = 2/3. \end{aligned}$$

b. $P(A \text{ wins}) = p+(1-p)^2p+(1-p)^4p+\cdots = \sum_{i=0}^{\infty} p(1-p)^{2i} = \frac{p}{1-(1-p)^2}$.

c. $\frac{d}{dp}\left(\frac{p}{1-(1-p)^2}\right) = \frac{p^2}{[1-(1-p)^2]^2} > 0$. Thus the probability is increasing in p, and the minimum is at zero. Using L'Hôpital's rule we find $\lim_{p\to 0}\frac{p}{1-(1-p)^2} = 1/2$.

1.25 Enumerating the sample space gives $S' = \{(B,B),(B,G),(G,B),(G,G)\}$,with each outcome equally likely. Thus $P(\text{at least one boy}) = 3/4$ and $P(\text{both are boys}) = 1/4$, therefore

$$P(\text{ both are boys} \mid \text{at least one boy }) = 1/3.$$

An ambiguity may arise if order is not acknowledged, the space is $S' = \{(B,B),(B,G),(G,G)\}$, with each outcome equally likely.

1.27 a. For n odd the proof is straightforward. There are an even number of terms in the sum $(0,1,\cdots,n)$, and $\binom{n}{k}$ and $\binom{n}{n-k}$, which are equal, have opposite signs. Thus, all pairs cancel and the sum is zero. If n is even, use the following identity, which is the basis of Pascal's triangle: For $k>0$, $\binom{n}{k} = \binom{n-1}{k}+\binom{n-1}{k-1}$. Then, for n even

$$\begin{aligned} \sum_{k=0}^{n}(-1)^k\binom{n}{k} &= \binom{n}{0}+\sum_{k=1}^{n-1}(-1)^k\binom{n}{k}+\binom{n}{n} \\ &= \binom{n}{0}+\binom{n}{n}+\sum_{k=1}^{n-1}(-1)^k\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right] \\ &= \binom{n}{0}+\binom{n}{n}-\binom{n-1}{0}-\binom{n-1}{n-1} = 0. \end{aligned}$$

b. Use the fact that for $k>0$, $k\binom{n}{k} = n\binom{n-1}{k-1}$ to write

$$\sum_{k=1}^{n} k\binom{n}{k} = n\sum_{k=1}^{n}\binom{n-1}{k-1} = n\sum_{j=0}^{n-1}\binom{n-1}{j} = n2^{n-1}.$$

c. $\sum_{k=1}^{n}(-1)^{k+1}k\binom{n}{k} = \sum_{k=1}^{n}(-1)^{k+1}\binom{n-1}{k-1} = n\sum_{j=0}^{n-1}(-1)^{j}\binom{n-1}{j} = 0$ from part a).

1.28 The average of the two integrals is

$$\begin{aligned}\left[(n\log n - n) + ((n+1)\log(n+1) - n)\right]/2 &= \left[n\log n + (n+1)\log(n+1)\right]/2 - n \\ &\approx (n+1/2)\log n - n.\end{aligned}$$

Let $d_n = \log n! - [(n+1/2)\log n - n]$, and we want to show that $\lim_{n\to\infty} md_n = c$, a constant. This would complete the problem, since the desired limit is the exponential of this one. This is accomplished in an indirect way, by working with differences, which avoids dealing with the factorial. Note that

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right)\log\left(1 + \frac{1}{n}\right) - 1.$$

Differentiation will show that $((n+\frac{1}{2}))\log((1+\frac{1}{n}))$ is increasing in n, and has minimum value $(3/2)\log 2 = 1.04$ at $n = 1$. Thus $d_n - d_{n+1} > 0$. Next recall the Taylor expansion of $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \cdots$. The first three terms provide an upper bound on $\log(1+x)$, as the remaining adjacent pairs are negative. Hence

$$0 < d_n d_{n+1} < \left(n+\frac{1}{2}\right)\left(\frac{1}{n}\frac{1}{2n^2} + \frac{1}{3n^3}\right) - 1 = \frac{1}{12n^2} + \frac{1}{6n^3}.$$

It therefore follows, by the comparison test, that the series $\sum_1^\infty d_n - d_{n+1}$ converges. Moreover, the partial sums must approach a limit. Hence, since the sum telescopes,

$$\lim_{N\to\infty}\sum_1^N d_n - d_{n+1} = \lim_{N\to\infty} d_1 - d_{N+1} = c.$$

Thus $\lim_{n\to\infty} d_n = d_1 - c$, a constant.

1.29 a.

Unordered	Ordered
{4,4,12,12}	(4,4,12,12), (4,12,12,4), (4,12,4,12)
	(12,4,12,4), (12,4,4,12), (12,12,4,4)

Unordered	Ordered
	(2,9,9,12), (2,9,12,9), (2,12,9,9), (9,2,9,12)
{2,9,9,12}	(9,2,12,9), (9,9,2,12), (9,9,12,2), (9,12,2,9)
	(9,12,9,2), (12,2,9,9), (12,9,2,9), (12,9,9,2)

b. Same as (a).

c. There are 6^6 ordered samples with replacement from $\{1,2,7,8,14,20\}$. The number of ordered samples that would result in $\{2,7,7,8,14,14\}$ is $\frac{6!}{2!2!1!1!} = 180$ (See Example 1.2.20). Thus the probability is $\frac{180}{6^6}$.

d. If the k objects were distinguishable then there would be $k!$ possible ordered arrangements. Since we have $k_1,\ldots,k_m$ different groups of indistinguishable objects, once the positions of the objects are fixed in the ordered arrangement permutations within objects of the same group won't change the ordered arrangement. There are $k_1!k_2!\cdots k_m!$ of such permutations for each ordered component. Thus there would be $\frac{k!}{k_1!k_2!\cdots k_m!}$ different ordered components.

e. Think of the m distinct numbers as m bins. Selecting a sample of size k, with replacement, is the same as putting k balls in the m bins. This is $\binom{k+m-1}{k}$, which is the number of distinct bootstrap samples. Note that, to create all of the bootstrap samples, we do not need to know what the original sample was. We only need to know the sample size and the distinct values.

1.31 a. The number of ordered samples drawn with replacement from the set $\{x_1,\ldots,x_n\}$ is n^n. The number of ordered samples that make up the unordered sample $\{x_1,\ldots,x_n\}$ is $n!$. Therefore the outcome with average $\frac{x_1+x_2+\cdots+x_n}{n}$ that is obtained by the unordered sample $\{x_1,\ldots,x_n\}$

has probability $\frac{n!}{n^n}$. Any other unordered outcome from $\{x_1,\dots,x_n\}$, distinct from the unordered sample $\{x_1,\dots,x_n\}$, will contain m different numbers repeated $k_1,\dots,k_m$ times where $k_1+k_2+\cdots+k_m = n$ with at least one of the k_i's satisfying $2 \le k_i \le n$. The probability of obtaining the corresponding average of such outcome is

$$\frac{n!}{k_1!k_2!\cdots k_m!n^n} < \frac{n!}{n^n}, \text{ since } k_1!k_2!\cdots k_m! > 1.$$

Therefore the outcome with average $\frac{x_1+x_2+\cdots+x_n}{n}$ is the most likely.

b. Stirling's approximation is that, as $n\to\infty$, $n! \approx \sqrt{2\pi}n^{n+(1/2)}e^{-n}$, and thus

$$\left(\frac{n!}{n^n}\right)\Big/\left(\frac{\sqrt{2n\pi}}{e^n}\right) = \frac{n!e^n}{n^n\sqrt{2n\pi}} = \frac{\sqrt{2\pi}n^{n+(1/2)}e^{-n}e^n}{n^n\sqrt{2n\pi}} = 1.$$

c. Since we are drawing with replacement from the set $\{x_1,\dots,x_n\}$, the probability of choosing any x_i is $\frac{1}{n}$. Therefore the probability of obtaining an ordered sample of size n without x_i is $(1-\frac{1}{n})^n$. To prove that $\lim_{n\to\infty}(1-\frac{1}{n})^n = e^{-1}$, calculate the limit of the log. That is

$$\lim_{n\to\infty} n\log\left(1-\frac{1}{n}\right) = \lim_{n\to\infty}\frac{\log\left(1-\frac{1}{n}\right)}{1/n}.$$

L'Hôpital's rule shows that the limit is -1, establishing the result. See also Lemma 2.3.14.

1.32 This is most easily seen by doing each possibility. Let $P(i)$ = probability that the candidate hired on the ith trial is best. Then

$$P(1)=\frac{1}{N},\quad P(2)=\frac{1}{N-1},\quad \dots\quad ,P(i)=\frac{1}{N-i+1},\quad \dots\quad ,P(N)=1.$$

1.33 Using Bayes rule

$$P(M|CB) = \frac{P(CB|M)P(M)}{P(CB|M)P(M)+P(CB|F)P(F)} = \frac{.05\times\frac{1}{2}}{.05\times\frac{1}{2}+.0025\times\frac{1}{2}} = .9524.$$

1.34 a.

$$\begin{aligned}&P(\text{Brown Hair})\\ &= P(\text{Brown Hair}|\text{Litter 1})P(\text{Litter 1}) + P(\text{Brown Hair}|\text{Litter 2})P(\text{Litter 2})\\ &= \left(\frac{2}{3}\right)\left(\frac{1}{2}\right)+\left(\frac{3}{5}\right)\left(\frac{1}{2}\right) = \frac{19}{30}.\end{aligned}$$

b. Use Bayes Theorem

$$P(\text{Litter 1}|\text{Brown Hair}) = \frac{P(BH|L1)P(L1)}{P(BH|L1)P(L1)+P(BH|L2)P(L2} = \frac{\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)}{\frac{19}{30}} = \frac{10}{19}.$$

1.35 Clearly $P(\cdot|B)\ge 0$, and $P(S|B)=1$. If $A_1, A_2, \dots$ are disjoint, then

$$\begin{aligned}P\left(\bigcup_{i=1}^\infty A_i\middle| B\right) &= \frac{P\left(\bigcup_{i=1}^\infty A_i\cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^\infty (A_i\cap B)\right)}{P(B)}\\ &= \frac{\sum_{i=1}^\infty P(A_i\cap B)}{P(B)} = \sum_{i=1}^\infty P(A_i|B).\end{aligned}$$

1.37 a. Using the same events A, B, C and $\mathcal{W}$ as in Example 1.3.4, we have

$$\begin{aligned} P(\mathcal{W}) &= P(\mathcal{W}|A)P(A) + P(\mathcal{W}|B)P(B) + P(\mathcal{W}|C)P(C) \\ &= \gamma\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right) \quad = \quad \frac{\gamma+1}{3}. \end{aligned}$$

Thus, $P(A|\mathcal{W}) = \frac{P(A\cap\mathcal{W})}{P(\mathcal{W})} = \frac{\gamma/3}{(\gamma+1)/3} = \frac{\gamma}{\gamma+1}$ where,

$$\begin{cases} \frac{\gamma}{\gamma+1} = \frac{1}{3} & \text{if } \gamma = \frac{1}{2} \\ \frac{\gamma}{\gamma+1} < \frac{1}{3} & \text{if } \gamma < \frac{1}{2} \\ \frac{\gamma}{\gamma+1} > \frac{1}{3} & \text{if } \gamma > \frac{1}{2}. \end{cases}$$

b. By Exercise 1.35, $P(\cdot|\mathcal{W})$ is a probability function. A, B and C are a partition. So

$$P(A|\mathcal{W}) + P(B|\mathcal{W}) + P(C|\mathcal{W}) = 1.$$

But, $P(B|\mathcal{W}) = 0$. Thus, $P(A|\mathcal{W}) + P(C|\mathcal{W}) = 1$. Since $P(A|\mathcal{W}) = 1/3$, $P(C|\mathcal{W}) = 2/3$. (This could be calculated directly, as in Example 1.3.4.) So if A can swap fates with C, his chance of survival becomes 2/3.

1.38 a. $P(A) = P(A\cap B) + P(A\cap B^c)$ from Theorem 1.2.11a. But $(A\cap B^c) \subset B^c$ and $P(B^c) = 1 - P(B) = 0$. So $P(A\cap B^c) = 0$, and $P(A) = P(A\cap B)$. Thus,

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)}{1} = P(A)$$

.

b. $A \subset B$ implies $A\cap B = A$. Thus,

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

And also,

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)}{P(B)}.$$

c. If A and B are mutually exclusive, then $P(A\cup B) = P(A) + P(B)$ and $A\cap(A\cup B) = A$. Thus,

$$P(A|A\cup B) = \frac{P(A\cap(A\cup B))}{P(A\cup B)} = \frac{P(A)}{P(A)+P(B)}.$$

d. $P(A\cap B\cap C) = P(A\cap(B\cap C)) = P(A|B\cap C)P(B\cap C) = P(A|B\cap C)P(B|C)P(C)$.

1.39 a. Suppose A and B are mutually exclusive. Then $A\cap B = \emptyset$ and $P(A\cap B) = 0$. If A and B are independent, then $0 = P(A\cap B) = P(A)P(B)$. But this cannot be since $P(A) > 0$ and $P(B) > 0$. Thus A and B cannot be independent.

b. If A and B are independent and both have positive probability, then

$$0 < P(A)P(B) = P(A\cap B).$$

This implies $A\cap B \neq \emptyset$, that is, A and B are not mutually exclusive.

1.40 a. $P(A^c\cap B) = P(A^c|B)P(B) = [1 - P(A|B)]P(B) = [1 - P(A)]P(B) = P(A^c)P(B)$, where the third equality follows from the independence of A and B.

b. $P(A^c\cap B^c) = P(A^c) - P(A^c\cap B) = P(A^c) - P(A^c)P(B) = P(A^c)P(B^c)$.

1.41 a.

$$\begin{aligned}
&P(\text{ dash sent} \mid \text{dash rec}) \\
&= \frac{P(\text{ dash rec} \mid \text{dash sent})P(\text{ dash sent})}{P(\text{ dash rec} \mid \text{dash sent})P(\text{ dash sent}) + P(\text{ dash rec} \mid \text{dot sent})P(\text{ dot sent})} \\
&= \frac{(2/3)(4/7)}{(2/3)(4/7)+(1/4)(3/7)} = 32/41.
\end{aligned}$$

b. By a similar calculation as the one in (a) $P(\text{dot sent}|\text{dot rec}) = 27/434$. Then we have $P(\text{ dash sent}|\text{dot rec}) = \frac{16}{43}$. Given that dot-dot was received, the distribution of the four possibilities of what was sent are

Event	Probability
dash-dash	$(16/43)^2$
dash-dot	$(16/43)(27/43)$
dot-dash	$(27/43)(16/43)$
dot-dot	$(27/43)^2$

1.43 a. For Boole's Inequality,

$$P(\cup_{i=1}^n) \le \sum_{i=1}^n P(A_i) - P_2 + P_3 + \cdots \pm P_n \le \sum_{i=1}^n P(A_i)$$

since $P_i \ge P_j$ if $i \le j$ and therefore the terms $-P_{2k} + P_{2k+1} \le 0$ for $k = 1, \ldots, \frac{n-1}{2}$ when n is odd. When n is even the last term to consider is $-P_n \le 0$. For Bonferroni's Inequality apply the inclusion-exclusion identity to the A_i^c, and use the argument leading to (1.2.10).

b. We illustrate the proof that the P_i are increasing by showing that $P_2 \ge P_3$. The other arguments are similar. Write

$$\begin{aligned}
P_2 = \sum_{1\le i<j\le n} P(A_i \cap A_j) &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P(A_i \cap A_j) \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\sum_{k=1}^{n} P(A_i \cap A_j \cap A_k) + P(A_i \cap A_j \cap (\cup_k A_k)^c)\right]
\end{aligned}$$

Now to get to P_3 we drop terms from this last expression. That is

$$\begin{aligned}
&\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\sum_{k=1}^{n} P(A_i \cap A_j \cap A_k) + P(A_i \cap A_j \cap (\cup_k A_k)^c)\right] \\
&\ge \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\sum_{k=1}^{n} P(A_i \cap A_j \cap A_k)\right] \\
&\ge \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) = \sum_{1\le i<j<k\le n} P(A_i \cap A_j \cap A_k) = P_3.
\end{aligned}$$

The sequence of bounds is improving because the bounds $P_1, P_1-P_2+P_3, P_1-P_2+P_3-P_4+P_5, \ldots$, are getting smaller since $P_i \ge P_j$ if $i \le j$ and therefore the terms $-P_{2k} + P_{2k+1} \le 0$. The lower bounds $P_1 - P_2, P_1 - P_2 + P_3 - P_4, P_1 - P_2 + P_3 - P_4 + P_5 - P_6, \ldots$, are getting bigger since $P_i \ge P_j$ if $i \le j$ and therefore the terms $P_{2k+1} - P_{2k} \ge 0$.

c. If all of the A_i are equal, all of the probabilities in the inclusion-exclusion identity are the same. Thus

$$P_1 = nP(A), \quad P_2 = \binom{n}{2}P(A), \quad \ldots \quad , P_j = \binom{n}{j}P(A),$$

and the sequence of upper bounds on $P(\cup_i A_i) = P(A)$ becomes

$$P_1 = nP(A), \quad P_1 - P_2 + P_3 = \left[n - \binom{n}{2} + \binom{n}{3}\right] P(A), \ldots$$

which eventually sum to one, so the last bound is exact. For the lower bounds we get

$$P_1 - P_2 = \left[n - \binom{n}{2}\right] P(A), \quad P_1 - P_2 + P_3 - P_4 = \left[n - \binom{n}{2} + \binom{n}{3} - \binom{n}{4}\right] P(A), \ldots$$

which start out negative, then become positive, with the last one equaling $P(A)$ (see Schwager 1984 for details).

1.44 $P(\text{at least 10 correct}|\text{guessing}) = \sum_{k=10}^{20} \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} = .01386$.

1.45 $\mathcal{X}$ is finite. Therefore $\mathcal{B}$ is the set of all subsets of $\mathcal{X}$. We must verify each of the three properties in Definition 1.2.4. (1) If $A \in \mathcal{B}$ then $P_X(A) = P(\cup_{x_i \in A}\{s_j \in S : X(s_j) = x_i\}) \geq 0$ since P is a probability function. (2) $P_X(\mathcal{X}) = P(\cup_{i=1}^m \{s_j \in S : X(s_j) = x_i\}) = P(S) = 1$. (3) If $A_1, A_2, \ldots \in \mathcal{B}$ and pairwise disjoint then

$$\begin{aligned} P_X(\cup_{k=1}^\infty A_k) &= P(\bigcup_{k=1}^\infty \{\cup_{x_i \in A_k}\{s_j \in S : X(s_j) = x_i\}\}) \\ &= \sum_{k=1}^\infty P(\cup_{x_i \in A_k}\{s_j \in S : X(s_j) = x_i\}) \quad = \quad \sum_{k=1}^\infty P_X(A_k), \end{aligned}$$

where the second inequality follows from the fact the P is a probability function.

1.46 This is similar to Exercise 1.20. There are 7^7 equally likely sample points. The possible values of X_3 are 0, 1 and 2. Only the pattern 331 (3 balls in one cell, 3 balls in another cell and 1 ball in a third cell) yields $X_3 = 2$. The number of sample points with this pattern is $\binom{7}{2}\binom{7}{3}\binom{4}{3}5 = 14{,}700$. So $P(X_3 = 2) = 14{,}700/7^7 \approx .0178$. There are 4 patterns that yield $X_3 = 1$. The number of sample points that give each of these patterns is given below.

pattern	number of sample points	
34	$7\binom{7}{3}6$	$= 1{,}470$
322	$7\binom{7}{3}\binom{6}{2}\binom{4}{2}\binom{2}{2}$	$= 22{,}050$
3211	$7\binom{7}{3}6\binom{4}{2}\binom{5}{2}2!$	$= 176{,}400$
31111	$7\binom{7}{3}\binom{6}{4}4!$	$= 88{,}200$
		288,120

So $P(X_3 = 1) = 288{,}120/7^7 \approx .3498$. The number of sample points that yield $X_3 = 0$ is $7^7 - 288{,}120 - 14{,}700 = 520{,}723$, and $P(X_3 = 0) = 520{,}723/7^7 \approx .6322$.

1.47 All of the functions are continuous, hence right-continuous. Thus we only need to check the limit, and that they are nondecreasing

a. $\lim_{x\to-\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{-\pi}{2}\right) = 0$, $\lim_{x\to\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2}\right) = 1$, and $\frac{d}{dx}\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)\right) = \frac{1}{1+x^2} > 0$, so $F(x)$ is increasing.

b. See Example 1.5.5.

c. $\lim_{x\to-\infty} e^{-e^{-x}} = 0$, $\lim_{x\to\infty} e^{-e^{-x}} = 1$, $\frac{d}{dx}e^{-e^{-x}} = e^{-x}e^{-e^{-x}} > 0$.

d. $\lim_{x\to-\infty}(1 - e^{-x}) = 0$, $\lim_{x\to\infty}(1 - e^{-x}) = 1$, $\frac{d}{dx}(1 - e^{-x}) = e^{-x} > 0$.

e. $\lim_{y\to-\infty} \frac{1-\epsilon}{1+e^{-y}} = 0$, $\lim_{y\to\infty} \epsilon + \frac{1-\epsilon}{1+e^{-y}} = 1$, $\frac{d}{dx}(\frac{1-\epsilon}{1+e^{-y}}) = \frac{(1-\epsilon)e^{-y}}{(1+e^{-y})^2} > 0$ and $\frac{d}{dx}(\epsilon + \frac{1-\epsilon}{1+e^{-y}}) > 0$, $F_Y(y)$ is continuous except on $y=0$ where $\lim_{y\downarrow 0}(\epsilon + \frac{1-\epsilon}{1+e^{-y}}) = F(0)$. Thus is $F_Y(y)$ right continuous.

1.48 If $F(\cdot)$ is a cdf, $F(x) = P(X \le x)$. Hence $\lim_{x\to\infty} P(X \le x) = 0$ and $\lim_{x\to-\infty} P(X \le x) = 1$. $F(x)$ is nondecreasing since the set $\{x : X \le x\}$ is nondecreasing in x. Lastly, as $x \downarrow x_0$, $P(X \le x) \to P(X \le x_0)$, so $F(\cdot)$ is right-continuous. (This is merely a consequence of defining $F(x)$ with "$\le$".)

1.49 For every t, $F_X(t) \le F_Y(t)$. Thus we have

$$P(X > t) = 1 - P(X \le t) = 1 - F_X(t) \ge 1 - F_Y(t) = 1 - P(Y \le t) = P(Y > t).$$

And for some t^*, $F_X(t^*) < F_Y(t^*)$. Then we have that

$$P(X > t^*) = 1 - P(X \le t^*) = 1 - F_X(t^*) > 1 - F_Y(t^*) = 1 - P(Y \le t^*) = P(Y > t^*).$$

1.50 Proof by induction. For $n = 2$

$$\sum_{k=1}^{2} t^{k-1} = 1 + t = \frac{1-t^2}{1-t}.$$

Assume true for n, this is $\sum_{k=1}^{n} t^{k-1} = \frac{1-t^n}{1-t}$. Then for $n+1$

$$\sum_{k=1}^{n+1} t^{k-1} = \sum_{k=1}^{n} t^{k-1} + t^n = \frac{1-t^n}{1-t} + t^n = \frac{1-t^n+t^n(1-t)}{1-t} = \frac{1-t^{n+1}}{1-t},$$

where the second inequality follows from the induction hypothesis.

1.51 This kind of random variable is called hypergeometric in Chapter 3. The probabilities are obtained by counting arguments, as follows.

x	$f_X(x) = P(X = x)$	
0	$\binom{5}{0}\binom{25}{4} \Big/ \binom{30}{4}$	$\approx .4616$
1	$\binom{5}{1}\binom{25}{3} \Big/ \binom{30}{4}$	$\approx .4196$
2	$\binom{5}{2}\binom{25}{2} \Big/ \binom{30}{4}$	$\approx .1095$
3	$\binom{5}{3}\binom{25}{1} \Big/ \binom{30}{4}$	$\approx .0091$
4	$\binom{5}{4}\binom{25}{0} \Big/ \binom{30}{4}$	$\approx .0002$

The cdf is a step function with jumps at $x = 0, 1, 2, 3$ and 4.

1.52 The function $g(\cdot)$ is clearly positive. Also,

$$\int_{x_0}^{\infty} g(x)dx = \int_{x_0}^{\infty} \frac{f(x)}{1-F(x_0)}dx = \frac{1-F(x_0)}{1-F(x_0)} = 1.$$

1.53 a. $\lim_{y\to-\infty} F_Y(y) = \lim_{y\to-\infty} 0 = 0$ and $\lim_{y\to\infty} F_Y(y) = \lim_{y\to\infty} 1 - \frac{1}{y^2} = 1$. For $y \le 1$, $F_Y(y) = 0$ is constant. For $y > 1$, $\frac{d}{dy}F_Y(y) = 2/y^3 > 0$, so F_Y is increasing. Thus for all y, F_Y is nondecreasing. Therefore F_Y is a cdf.

b. The pdf is $f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 2/y^3 & \text{if } y > 1 \\ 0 & \text{if } y \le 1. \end{cases}$

c. $F_Z(z) = P(Z \le z) = P(10(Y-1) \le z) = P(Y \le (z/10)+1) = F_Y((z/10)+1)$. Thus,

$$F_Z(z) = \begin{cases} 0 & \text{if } z \le 0 \\ 1 - \left(\frac{1}{[(z/10)+1]^2}\right) & \text{if } z > 0. \end{cases}$$

1.54 a. $\int_0^{\pi/2} \sin x dx = 1$. Thus, $c = 1/1 = 1$.

b. $\int_{-\infty}^{\infty} e^{-|x|} dx = \int_{-\infty}^{0} e^{x} dx + \int_0^{\infty} e^{-x} dx = 1 + 1 = 2$. Thus, $c = 1/2$.

1.55

$$P(V \le 5) = P(T < 3) = \int_0^3 \frac{1}{1.5} e^{-t/1.5}\, dt = 1 - e^{-2}.$$

For $v \ge 6$,

$$P(V \le v) = P(2T \le v) = P\left(T \le \frac{v}{2}\right) = \int_0^{\frac{v}{2}} \frac{1}{1.5} e^{-t/1.5}\, dt = 1 - e^{-v/3}.$$

Therefore,

$$P(V \le v) = \begin{cases} 0 & -\infty < v < 0, \\ 1 - e^{-2} & 0 \le v < 6\ , \\ 1 - e^{-v/3} & 6 \le v \end{cases} \quad .$$

Chapter 2

Transformations and Expectations

2.1 a. $f_x(x) = 42x^5(1-x)$, $0 < x < 1$; $y = x^3 = g(x)$, monotone, and $\mathcal{Y} = (0,1)$. Use Theorem 2.1.5.

$$\begin{aligned} f_Y(y) &= f_x(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| = f_x(y^{1/3})\frac{d}{dy}(y^{1/3}) = 42y^{5/3}(1-y^{1/3})(\frac{1}{3}y^{-2/3}) \\ &= 14y(1-y^{1/3}) = 14y - 14y^{4/3}, \quad 0 < y < 1. \end{aligned}$$

To check the integral,

$$\int_0^1 (14y - 14y^{4/3})dy = 7y^2 - 14\frac{y^{7/3}}{7/3}\Big|_0^1 = 7y^2 - 6y^{7/3}\Big|_0^1 = 1 - 0 = 1.$$

b. $f_x(x) = 7e^{-7x}$, $0 < x < \infty$, $y = 4x+3$, monotone, and $\mathcal{Y} = (3,\infty)$. Use Theorem 2.1.5.

$$f_Y(y) = f_x(\frac{y-3}{4})\left|\frac{d}{dy}(\frac{y-3}{4})\right| = 7e^{-(7/4)(y-3)}\left|\frac{1}{4}\right| = \frac{7}{4}e^{-(7/4)(y-3)}, \; 3 < y < \infty.$$

To check the integral,

$$\int_3^\infty \frac{7}{4}e^{-(7/4)(y-3)}dy = -e^{-(7/4)(y-3)}\Big|_3^\infty = 0 - (-1) = 1.$$

c. $F_Y(y) = P(0 \le X \le \sqrt{y}) = F_X(\sqrt{y})$. Then $f_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y})$. Therefore

$$f_Y(y) = \frac{1}{2\sqrt{y}}30(\sqrt{y})^2(1-\sqrt{y})^2 = 15y^{\frac{1}{2}}(1-\sqrt{y})^2, \quad 0 < y < 1.$$

To check the integral,

$$\int_0^1 15y^{\frac{1}{2}}(1-\sqrt{y})^2 dy = \int_0^1 (15y^{\frac{1}{2}} - 30y + 15y^{\frac{3}{2}})dy = 15(\frac{2}{3}) - 30(\frac{1}{2}) + 15(\frac{2}{5}) = 1.$$

2.2 In all three cases, Theorem 2.1.5 is applicable and yields the following answers.

a. $f_Y(y) = \frac{1}{2}y^{-1/2}$, $0 < y < 1$.
b. $f_Y(y) = \frac{(n+m+1)!}{n!m!}e^{-y(n+1)}(1-e^{-y})^m$, $0 < y < \infty$.
c. $f_Y(y) = \frac{1}{\sigma^2}\frac{\log y}{y}e^{-(1/2)((\log y)/\sigma)^2}$, $0 < y < \infty$.

2.3 $P(Y = y) = P(\frac{X}{X+1} = y) = P(X = \frac{y}{1-y}) = \frac{1}{3}(\frac{2}{3})^{y/(1-y)}$, where $y = 0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \frac{x}{x+1}, \ldots$.

2.4 a. $f(x)$ is a pdf since it is positive and

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^0 \frac{1}{2}\lambda e^{\lambda x}dx + \int_0^\infty \frac{1}{2}\lambda e^{-\lambda x}dx = \frac{1}{2} + \frac{1}{2} = 1.$$

b. Let X be a random variable with density $f(x)$.

$$P(X < t) = \begin{cases} \int_{-\infty}^{t} \frac{1}{2}\lambda e^{\lambda x} dx & \text{if } t < 0 \\ \int_{-\infty}^{0} \frac{1}{2}\lambda e^{\lambda x} dx + \int_{0}^{t} \frac{1}{2}\lambda e^{-\lambda x} dx & \text{if } t \geq 0 \end{cases}$$

where, $\int_{-\infty}^{t} \frac{1}{2}\lambda e^{\lambda x} dx = \frac{1}{2}e^{\lambda x}\big|_{-\infty}^{t} = \frac{1}{2}e^{\lambda t}$ and $\int_{0}^{t} \frac{1}{2}\lambda e^{-\lambda x} dx = -\frac{1}{2}e^{-\lambda x}\big|_{0}^{t} = -\frac{1}{2}e^{-\lambda t} + \frac{1}{2}$. Therefore,

$$P(X < t) = \begin{cases} \frac{1}{2}e^{\lambda t} & \text{if } t < 0 \\ 1 - \frac{1}{2}e^{-\lambda t} dx & \text{if } t \geq 0 \end{cases}$$

c. $P(|X| < t) = 0$ for $t < 0$, and for $t \geq 0$,

$$\begin{aligned} P(|X| < t) &= P(-t < X < t) = \int_{-t}^{0} \frac{1}{2}\lambda e^{\lambda x} dx + \int_{0}^{t} \frac{1}{2}\lambda e^{-\lambda x} dx \\ &= \frac{1}{2}\left[1 - e^{-\lambda t}\right] + \frac{1}{2}\left[-e^{-\lambda t} + 1\right] = 1 - e^{-\lambda t}. \end{aligned}$$

2.5 To apply Theorem 2.1.8. Let $A_0 = \{0\}$, $A_1 = (0, \frac{\pi}{2})$, $A_3 = (\pi, \frac{3\pi}{2})$ and $A_4 = (\frac{3\pi}{2}, 2\pi)$. Then $g_i(x) = \sin^2(x)$ on A_i for $i = 1, 2, 3, 4$. Therefore $g_1^{-1}(y) = \sin^{-1}(\sqrt{y})$, $g_2^{-1}(y) = \pi - \sin^{-1}(\sqrt{y})$, $g_3^{-1}(y) = \sin^{-1}(\sqrt{y}) + \pi$ and $g_4^{-1}(y) = 2\pi - \sin^{-1}(\sqrt{y})$. Thus

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi}\left|\frac{1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}\right| + \frac{1}{2\pi}\left|-\frac{1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}\right| + \frac{1}{2\pi}\left|\frac{1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}\right| + \frac{1}{2\pi}\left|-\frac{1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}\right| \\ &= \frac{1}{\pi\sqrt{y(1-y)}}, \quad 0 \leq y \leq 1 \end{aligned}$$

To use the cdf given in (2.1.6) we have that $x_1 = \sin^{-1}(\sqrt{y})$ and $x_2 = \pi - \sin^{-1}(\sqrt{y})$. Then by differentiating (2.1.6) we obtain that

$$\begin{aligned} f_Y(y) &= 2f_X(\sin^{-1}(\sqrt{y})\frac{d}{dy}(\sin^{-1}(\sqrt{y}) - 2f_X(\pi - \sin^{-1}(\sqrt{y})\frac{d}{dy}(\pi - \sin^{-1}(\sqrt{y}) \\ &= 2(\frac{1}{2\pi}\frac{1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}) - 2(\frac{1}{2\pi}\frac{-1}{\sqrt{1-y}}\frac{1}{2\sqrt{y}}) \\ &= \frac{1}{\pi\sqrt{y(1-y)}} \end{aligned}$$

2.6 Theorem 2.1.8 can be used for all three parts.

a. Let $A_0 = \{0\}$, $A_1 = (-\infty, 0)$ and $A_2 = (0, \infty)$. Then $g_1(x) = |x|^3 = -x^3$ on A_1 and $g_2(x) = |x|^3 = x^3$ on A_2. Use Theorem 2.1.8 to obtain

$$f_Y(y) = \frac{1}{3}e^{-y^{1/3}}y^{-2/3}, \quad 0 < y < \infty$$

.

b. Let $A_0 = \{0\}$, $A_1 = (-1, 0)$ and $A_2 = (0, 1)$. Then $g_1(x) = 1 - x^2$ on A_1 and $g_2(x) = 1 - x^2$ on A_2. Use Theorem 2.1.8 to obtain

$$f_Y(y) = \frac{3}{8}(1-y)^{-1/2} + \frac{3}{8}(1-y)^{1/2}, \quad 0 < y < 1$$

.

c. Let $A_0 = \{0\}$, $A_1 = (-1, 0)$ and $A_2 = (0, 1)$. Then $g_1(x) = 1 - x^2$ on A_1 and $g_2(x) = 1 - x$ on A_2. Use Theorem 2.1.8 to obtain

$$f_Y(y) = \frac{3}{16}(1 - \sqrt{1-y})^2 \frac{1}{\sqrt{1-y}} + \frac{3}{8}(2-y)^2, \quad 0 < y < 1$$

.

2.7 Theorem 2.1.8 does not directly apply.

a. Theorem 2.1.8 does not directly apply. Instead write

$$\begin{aligned} P(Y \le y) &= P(X^2 \le y) \\ &= \begin{cases} P(-\sqrt{y} \le X \le \sqrt{y}) & \text{if } |x| \le 1 \\ P(1 \le X \le \sqrt{y}) & \text{if } x \ge 1 \end{cases} \\ &= \begin{cases} \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx & \text{if } |x| \le 1 \\ \int_1^{\sqrt{y}} f_X(x)dx & \text{if } x \ge 1 \end{cases}. \end{aligned}$$

Differentiation gives

$$f_y(y) = \begin{cases} \frac{2}{9}\frac{1}{\sqrt{y}} & \text{if } y \le 1 \\ \frac{1}{9} + \frac{1}{9}\frac{1}{\sqrt{y}} & \text{if } y \ge 1 \end{cases}.$$

b. If the sets $B_1, B_2, \ldots, B_K$ are a partition of the range of Y, we can write

$$f_Y(y) = \sum_k f_Y(y) I(y \in B_k)$$

and do the transformation on each of the B_k. So this says that we can apply Theorem 2.1.8 on each of the B_k and add up the pieces. For $A_1 = (-1, 1)$ and $A_2 = (1, 2)$ the calculations are identical to those in part (a). (Note that on A_1 we are essentially using Example 2.1.7).

2.8 For each function we check the conditions of Theorem 1.5.3.

a. (i) $\lim_{x\to 0} F(x) = 1 - e^{-0} = 0$, $\lim_{x\to -\infty} F(x) = 1 - e^{-\infty} = 1$.
(ii) $1 - e^{-x}$ is increasing in x.
(iii) $1 - e^{-x}$ is continuous.
(iv) $F_x^{-1}(y) = -\log(1-y)$.

b. (i) $\lim_{x\to -\infty} F(x) = e^{-\infty}/2 = 0$, $\lim_{x\to\infty} F(x) = 1 - (e^{1-\infty}/2) = 1$.
(ii) $e^{-x/2}$ is increasing, $1/2$ is nondecreasing, $1 - (e^{1-x}/2)$ is increasing.
(iii) For continuity we only need check $x = 0$ and $x = 1$, and $\lim_{x\to 0} F(x) = 1/2$, $\lim_{x\to 1} F(x) = 1/2$, so F is continuous.
(iv)

$$F_X^{-1}(y) = \begin{cases} \log(2y) & 0 \le y < \frac{1}{2} \le y < 1, \\ 1 - \log(2(1-y)) & \frac{1}{2} \le y < 1 \end{cases}$$

c. (i) $\lim_{x\to -\infty} F(x) = e^{-\infty}/4 = 0$, $\lim_{x\to\infty} F(x) = 1 - e^{-\infty}/4 = 1$.
(ii) $e^{-x}/4$ and $1 - e^{-x}/4$ are both increasing in x.
(iii) $\lim_{x\downarrow 0} F(x) = 1 - e^{-0}/4 = \frac{3}{4} = F(0)$, so F is right-continuous.
(iv) $F_X^{-1}(y) = \begin{cases} \log(4y) & 0 \le y < \frac{1}{4} \\ -\log(4(1-y)) & \frac{1}{4} \le y < 1 \end{cases}$

2.9 From the probability integral transformation, Theorem 2.1.10, we know that if $u(x) = F_x(x)$, then $F_x(X) \sim \text{uniform}(0, 1)$. Therefore, for the given pdf, calculate

$$u(x) = F_x(x) = \begin{cases} 0 & \text{if } x \le 1 \\ (x-1)^2/4 & \text{if } 1 < x < 3 \\ 1 & \text{if } 3 \le x \end{cases}$$

.

2.10 a. We prove part b), which is equivalent to part a).

b. Let $A_y = \{x : F_x(x) \le y\}$. Since F_x is nondecreasing, A_y is a half infinite interval, either open, say $(-\infty, x_y)$, or closed, say $(-\infty, x_y]$. If A_y is closed, then

$$F_Y(y) = P(Y \le y) = P(F_x(X) \le y) = P(X \in A_y) = F_x(x_y) \le y.$$

The last inequality is true because $x_y \in A_y$, and $F_x(x) \le y$ for every $x \in A_y$. If A_y is open, then

$$F_Y(y) = P(Y \le y) = P(F_x(X) \le y) = P(X \in A_y),$$

as before. But now we have

$$P(X \in A_y) = P(X \in (-\infty, x_y)) = \lim_{x \uparrow y} P(X \in (-\infty, x]),$$

Use the Axiom of Continuity, Exercise 1.12, and this equals $\lim_{x\uparrow y} F_X(x) \le y$. The last inequality is true since $F_x(x) \le y$ for every $x \in A_y$, that is, for every $x < x_y$. Thus, $F_Y(y) \le y$ for every y. To get strict inequality for some y, let y be a value that is "jumped over" by F_x. That is, let y be such that, for some x_y,

$$\lim_{x\uparrow y} F_X(x) < y < F_X(x_y).$$

For such a y, $A_y = (-\infty, x_y)$, and $F_Y(y) = \lim_{x\uparrow y} F_X(x) < y$.

2.11 a. Using integration by parts with $u = x$ and $dv = xe^{\frac{-x^2}{2}} dx$ then

$$\mathrm{E}X^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{2\pi} e^{\frac{-x^2}{2}} dx = \frac{1}{2\pi} \left[-xe^{\frac{-x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx \right] = \frac{1}{2\pi}(2\pi) = 1.$$

Using example 2.1.7 let $Y = X^2$. Then

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left[\frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} + \frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} \right] = \frac{1}{\sqrt{2\pi y}} e^{\frac{-y}{2}}.$$

Therefore,

$$\mathrm{E}Y = \int_0^{\infty} \frac{y}{\sqrt{2\pi y}} e^{\frac{-y}{2}} dy = \frac{1}{\sqrt{2\pi}} \left[-2y^{\frac{1}{2}} e^{\frac{-y}{2}} \Big|_0^{\infty} + \int_0^{\infty} y^{\frac{-1}{2}} e^{\frac{-y}{2}} dy \right] = \frac{1}{\sqrt{2\pi}}(\sqrt{2\pi}) = 1.$$

This was obtained using integration by parts with $u = 2y^{\frac{1}{2}}$ and $dv = \frac{1}{2}e^{\frac{-y}{2}}$ and the fact the $f_Y(y)$ integrates to 1.

b. $Y = |X|$ where $-\infty < x < \infty$. Therefore $0 < y < \infty$. Then

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(|X| \le y) = P(-y \le X \le y) \\ &= P(x \le y) - P(X \le -y) = F_X(y) - F_X(-y). \end{aligned}$$

Therefore,

$$F_Y(y) = \frac{d}{dy}F_Y(y) = f_X(y) + f_X(-y) = \frac{1}{\sqrt{2\pi}}e^{\frac{-y}{2}} + \frac{1}{\sqrt{2\pi}}e^{\frac{-y}{2}} = \sqrt{\frac{2}{\pi}}e^{\frac{-y}{2}}.$$

Thus,

$$\mathrm{E}Y = \int_0^\infty y\sqrt{\frac{2}{\pi}}e^{\frac{-y}{2}}\,dy = \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-u}du = \sqrt{\frac{2}{\pi}}\left[-e^{-u}\Big|_0^\infty\right] = \sqrt{\frac{2}{\pi}},$$

where $u = \frac{y^2}{2}$.

$$\mathrm{E}Y^2 = \int_0^\infty y^2\sqrt{\frac{2}{\pi}}e^{\frac{-y}{2}}\,dy = \sqrt{\frac{2}{\pi}}\left[-ye^{\frac{-y}{2}}\Big|_0^\infty + \int_0^\infty e^{\frac{-y}{2}}\,dy\right] = \sqrt{\frac{2}{\pi}}\sqrt{\frac{\pi}{2}} = 1.$$

This was done using integration by part with $u = y$ and $dv = ye^{\frac{-y}{2}}dy$. Then $\mathrm{Var}(Y) = 1 - \frac{2}{\pi}$.

2.12 We have $\tan x = y/d$, therefore $\tan^{-1}(y/d) = x$ and $\frac{d}{dy}\tan^{-1}(y/d) = \frac{1}{1+(y/d)^2}\,\frac{1}{d}dy = dx$. Thus,

$$f_Y(y) = \frac{2}{\pi d}\,\frac{1}{1+(y/d)^2}, \quad 0 < y < \infty.$$

This is the Cauchy distribution restricted to $(0,\infty)$, and the mean is infinite.

2.13 $P(X = k) = (1-p)^k p + p^k(1-p)$, $k = 1, 2, \ldots$. Therefore,

$$\begin{aligned}\mathrm{E}X &= \sum_{k=1}^\infty k[(1-p)^k p + p^k(1-p)] = (1-p)p\left[\sum_{k=1}^\infty k(1-p)^{k-1} + \sum_{k=1}^\infty kp^{k-1}\right] \\ &= (1-p)p\left[\frac{1}{p^2} + \frac{1}{(1-p)^2}\right] = \frac{1-2p+2p^2}{p(1-p)}.\end{aligned}$$

2.14

$$\begin{aligned}\int_0^\infty (1 - F_X(x))dx &= \int_0^\infty P(X > x)dx \\ &= \int_0^\infty \int_x^\infty f_X(y)dydx \\ &= \int_0^\infty \int_0^y dx f_X(y)dy \\ &= \int_0^\infty y f_X(y)dy = \mathrm{E}X,\end{aligned}$$

where the last equality follows from changing the order of integration.

2.15 Assume without loss of generality that $X \le Y$. Then $X \vee Y = Y$ and $X \wedge Y = X$. Thus $X + Y = (X \wedge Y) + (X \vee Y)$. Taking expectations

$$\mathrm{E}[X+Y] = \mathrm{E}[(X\wedge Y) + (X \vee Y)] = \mathrm{E}(X\wedge Y) + \mathrm{E}(X\vee Y).$$

Therefore $\mathrm{E}(X\vee Y) = \mathrm{E}X + \mathrm{E}Y - \mathrm{E}(X\wedge Y)$.

2.16 From Exercise 2.14,

$$\mathrm{E}T = \int_0^\infty \left[ae^{-\lambda t} + (1-a)e^{-\mu t}\right]dt = \frac{-ae^{-\lambda t}}{\lambda} - \frac{(1-a)e^{-\mu t}}{\mu}\Big|_0^\infty = \frac{a}{\lambda} + \frac{1-a}{\mu}.$$

2.17 a. $\int_0^m 3x^2 dx = m^3 \stackrel{set}{=} \frac{1}{2} \quad \Rightarrow \quad m = \left(\frac{1}{2}\right)^{1/3} = .794.$

b. The function is symmetric about zero, therefore $m = 0$ as long as the integral is finite.

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = \frac{1}{\pi}\tan^{-1}(x)\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2}+\frac{\pi}{2}\right) = 1.$$

This is the Cauchy pdf.

2.18 $\mathrm{E}|X-a| = \int_{-\infty}^{\infty} |x-a| f(x)dx = \int_{-\infty}^{a} -(x-a)f(x)dx + \int_a^{\infty}(x-a)f(x)dx$. Then,

$$\frac{d}{da}\mathrm{E}|X-a| = \int_{-\infty}^{a} f(x)dx - \int_a^{\infty} f(x)dx \stackrel{\text{set}}{=} 0.$$

The solution to this equation is $a =$ median. This is a minimum since $d^2/da^2\mathrm{E}|X-a| = 2f(a) > 0$.

2.19

$$\begin{aligned}
\frac{d}{da}\mathrm{E}(X-a)^2 &= \frac{d}{da}\int_{-\infty}^{\infty}(x-a)^2 f_X(x)dx = \int_{-\infty}^{\infty}\frac{d}{da}(x-a)^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} -2(x-a)f_X(x)dx = -2\left[\int_{-\infty}^{\infty} x f_X(x)dx - a\int_{-\infty}^{\infty} f_X(x)dx\right] \\
&= -2[\mathrm{E}X - a].
\end{aligned}$$

Therefore if $\frac{d}{da}\mathrm{E}(X-a)^2 = 0$ then $-2[\mathrm{E}X-a] = 0$ which implies that $\mathrm{E}X = a$. If $\mathrm{E}X = a$ then $\frac{d}{da}\mathrm{E}(X-a)^2 = -2[\mathrm{E}X-a] = -2[a-a] = 0$. $\mathrm{E}X = a$ is a minimum since $d^2/da^2\mathrm{E}(X-a)^2 = 2 > 0$. The assumptions that are needed are the ones listed in Theorem 2.4.3.

2.20 From Example 1.5.4, if $X =$ number of children until the first daughter, then

$$P(X=k) = (1-p)^{k-1}p,$$

where p = probability of a daughter. Thus X is a geometric random variable, and

$$\begin{aligned}
\mathrm{E}X &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p - \sum_{k=1}^{\infty}\frac{d}{dp}(1-p)^k = -p\frac{d}{dp}\left[\sum_{k=0}^{\infty}(1-p)^k - 1\right] \\
&= -p\frac{d}{dp}\left[\frac{1}{p}-1\right] = \frac{1}{p}.
\end{aligned}$$

Therefore, if p = $\frac{1}{2}$,the expected number of children is two.

2.21 Since $g(x)$ is monotone

$$\mathrm{E}g(X) = \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_{-\infty}^{\infty} y f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)dy = \int_{-\infty}^{\infty} y f_Y(y)dy = \mathrm{E}Y,$$

where the second equality follows from the change of variable $y = g(x)$, $x = g^{-1}(y)$ and $dx = \frac{d}{dy}g^{-1}(y)dy$.

2.22 a. Using integration by parts with $u = x$ and $dv = xe^{-x^2/\beta^2}$ we obtain that

$$\int_0^{\infty} x^2 e^{-x^2/\beta^2} dx^2 = \frac{\beta^2}{2}\int_0^{\infty} e^{-x^2/\beta^2} dx.$$

The integral can be evaluated using the argument on pages 104-105 (see 3.3.14) or by transforming to a gamma kernel (use $y = -\lambda^2/\beta^2$). Therefore, $\int_0^{\infty} e^{-x^2/\beta^2} dx = \sqrt{\pi}\beta/2$ and hence the function integrates to 1.

b. $\mathrm{E}X = 2\beta/\sqrt{\pi}$ $\qquad \mathrm{E}X^2 = 3\beta^2/2$ $\qquad \mathrm{Var}X = \beta^2\left[\frac{3}{2} - \frac{4}{\pi}\right]$.

2.23 a. Use Theorem 2.1.8 with $A_0 = \{0\}$, $A_1 = (-1, 0)$ and $A_2 = (0, 1)$. Then $g_1(x) = x^2$ on A_1 and $g_2(x) = x^2$ on A_2. Then

$$f_Y(y) = \frac{1}{2}y^{-1/2}, \quad 0 < y < 1.$$

b. $\mathrm{E}Y = \int_0^1 y f_Y(y)dy = \frac{1}{3}$ $\qquad \mathrm{E}Y^2 = \int_0^1 y^2 f_Y(y)dy = \frac{1}{5}$ $\qquad \mathrm{Var}Y = \frac{1}{5} - \left(\frac{1}{3}\right)^2 = \frac{4}{45}$.

2.24 a. $\mathrm{E}X = \int_0^1 x a x^{a-1}dx = \int_0^1 a x^a dx = \left.\frac{ax^{a+1}}{a+1}\right|_0^1 = \frac{a}{a+1}$.

$\mathrm{E}X^2 = \int_0^1 x^2 a x^{a-1}dx = \int_0^1 a x^{a+1}dx = \left.\frac{ax^{a+2}}{a+2}\right|_0^1 = \frac{a}{a+2}$.

$\mathrm{Var}X = \frac{a}{a+2} - \left(\frac{a}{a+1}\right)^2 = \frac{a}{(a+2)(a+1)^2}$.

b. $\mathrm{E}X = \sum_{x=1}^n \frac{x}{n} = \frac{1}{n}\sum_{x=1}^n x = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$.
$\mathrm{E}X^2 = \sum_{i=1}^n \frac{x^2}{n} = \frac{1}{n}\sum_{i=1}^n x^2 = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$.
$\mathrm{Var}X = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4} = \frac{n^2+1}{12}$.

c. $\mathrm{E}X = \int_0^2 x\frac{3}{2}(x-1)^2dx = \frac{3}{2}\int_0^2 (x^3 - 2x^2 + x)dx = 1$.
$\mathrm{E}X^2 = \int_0^2 x^2\frac{3}{2}(x-1)^2dx = \frac{3}{2}\int_0^2 (x^4 - 2x^3 + x^2)dx = \frac{8}{5}$.
$\mathrm{Var}X = \frac{8}{5} - 1^2 = \frac{3}{5}$.

2.25 a. $Y = -X$ and $g^{-1}(y) = -y$. Thus $f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| = f_X(-y)|-1| = f_X(y)$ for every y.

b. To show that $M_X(t)$ is symmetric about 0 we must show that $M_X(0+\epsilon) = M_X(0-\epsilon)$ for all $\epsilon > 0$.

$$\begin{aligned} M_X(0+\epsilon) &= \int_{-\infty}^{\infty} e^{(0+\epsilon)x} f_X(x)dx = \int_{-\infty}^{0} e^{\epsilon x} f_X(x)dx + \int_0^{\infty} e^{\epsilon x} f_X(x)dx \\ &= \int_0^{\infty} e^{\epsilon(-x)} f_X(-x)dx + \int_{-\infty}^{0} e^{\epsilon(-x)} f_X(-x)dx = \int_{-\infty}^{\infty} e^{-\epsilon x} f_X(x)dx \\ &= \int_{-\infty}^{\infty} e^{(0-\epsilon)x} f_X(x)dx = M_X(0-\epsilon). \end{aligned}$$

2.26 a. There are many examples; here are three. The standard normal pdf (Example 2.1.9) is symmetric about $a = 0$ because $(0-\epsilon)^2 = (0+\epsilon)^2$. The Cauchy pdf (Example 2.2.4) is symmetric about $a = 0$ because $(0-\epsilon)^2 = (0+\epsilon)^2$. The uniform$(0,1)$ pdf (Example 2.1.4) is symmetric about $a = 1/2$ because

$$f((1/2)+\epsilon) = f((1/2)-\epsilon) = \begin{cases} 1 & \text{if } 0 < \epsilon < \frac{1}{2} \\ 0 & \text{if } \frac{1}{2} \le \epsilon < \infty \end{cases}.$$

b.

$$\begin{aligned} \int_a^{\infty} f(x)dx &= \int_0^{\infty} f(a+\epsilon)d\epsilon && \text{(change variable, } \epsilon = x - a) \\ &= \int_0^{\infty} f(a-\epsilon)d\epsilon && (f(a+\epsilon) = f(a-\epsilon) \text{ for all } \epsilon > 0) \\ &= \int_{-\infty}^{a} f(x)dx. && \text{(change variable, } x = a - \epsilon) \end{aligned}$$

Since

$$\int_{-\infty}^{a} f(x)dx + \int_{a}^{\infty} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1,$$

it must be that

$$\int_{-\infty}^{a} f(x)\,dx = \int_{a}^{\infty} f(x)\,dx = 1/2.$$

Therefore, a is a median.

c.

$$\begin{aligned} \mathrm{E}X - a &= \mathrm{E}(X-a) = \int_{-\infty}^{\infty}(x-a)f(x)dx \\ &= \int_{-\infty}^{a}(x-a)f(x)dx + \int_{a}^{\infty}(x-a)f(x)dx \\ &= \int_{0}^{\infty}(-\epsilon)f(a-\epsilon)d\epsilon + \int_{0}^{\infty}\epsilon f(a+\epsilon)d\epsilon \end{aligned}$$

With a change of variable, $\epsilon = a - x$ in the first integral, and $\epsilon = x - a$ in the second integral we obtain that

$$\begin{aligned} \mathrm{E}X - a &= \mathrm{E}(X-a) \\ &= -\int_{0}^{\infty}\epsilon f(a-\epsilon)d\epsilon + \int_{0}^{\infty}\epsilon f(a-\epsilon)d\epsilon \qquad (f(a+\epsilon) = f(a-\epsilon) \text{ for all } \epsilon > 0) \\ &= 0. \qquad \text{(two integrals are same)} \end{aligned}$$

Therefore, $\mathrm{E}X = a$.

d. If $a > \epsilon > 0$,

$$f(a-\epsilon) = e^{-(a-\epsilon)} > e^{-(a+\epsilon)} = f(a+\epsilon).$$

Therefore, $f(x)$ is not symmetric about $a > 0$. If $-\epsilon < a \le 0$,

$$f(a-\epsilon) = 0 < e^{-(a+\epsilon)} = f(a+\epsilon).$$

Therefore, $f(x)$ is not symmetric about $a \le 0$, either.

e. The median of $X = \log 2 < 1 = \mathrm{E}X$.

2.27 a. The standard normal pdf.

b. The uniform on the interval $(0, 1)$.

c. For the case when the mode is unique. Let a be the point of symmetry and b be the mode. Let assume that a is not the mode and without loss of generality that $a = b+\epsilon > b$ for $\epsilon > 0$. Since b is the mode then $f(b) > f(b+\epsilon) \ge f(b+2\epsilon)$ which implies that $f(a-\epsilon) > f(a) \ge f(a+\epsilon)$ which contradict the fact the $f(x)$ is symmetric. Thus a is the mode.
For the case when the mode is not unique, there must exist an interval (x_1, x_2) such that $f(x)$ has the same value in the whole interval, i.e, $f(x)$ is flat in this interval and for all $b \in (x_1, x_2)$, b is a mode. Let assume that $a \notin (x_1, x_2)$, thus a is not a mode. Let also assume without loss of generality that $a = (b+\epsilon) > b$. Since b is a mode and $a = (b+\epsilon) \notin (x_1, x_2)$ then $f(b) > f(b+\epsilon) \ge f(b+2\epsilon)$ which contradict the fact the $f(x)$ is symmetric. Thus $a \in (x_1, x_2)$ and is a mode.

d. $f(x)$ is decreasing for $x \ge 0$, with $f(0) > f(x) > f(y)$ for all $0 < x < y$. Thus $f(x)$ is unimodal and 0 is the mode.

2.28 a.

$$\begin{aligned}
\mu_3 &= \int_{-\infty}^{\infty}(x-a)^3 f(x)dx \;=\; \int_{-\infty}^{a}(x-a)^3 f(x)dx + \int_{a}^{\infty}(x-a)^3 f(x)dx \\
&= \int_{-\infty}^{0} y^3 f(y+a)dy + \int_{0}^{\infty} y^3 f(y+a)dy \qquad \text{(change variable } y = x-a) \\
&= \int_{0}^{\infty} -y^3 f(-y+a)dy + \int_{0}^{\infty} y^3 f(y+a)dy \\
&= 0. \qquad\qquad (f(-y+a) = f(y+a))
\end{aligned}$$

b. For $f(x) = e^{-x}$, $\mu_1 = \mu_2 = 1$, therefore $\alpha_3 = \mu_3$.

$$\begin{aligned}
\mu_3 &= \int_0^{\infty}(x-1)^3 e^{-x}dx \;=\; \int_0^{\infty}(x^3 - 3x^2 + 3x - 1)e^{-x}dx \\
&= \Gamma(4) - 3\Gamma(3) + 3\Gamma(2) - \Gamma(1) \;=\; 3! - 3\times 2! + 3\times 1 - 1 \;=\; 3.
\end{aligned}$$

c. Each distribution has $\mu_1 = 0$, therefore we must calculate $\mu_2 = \mathrm{E}X^2$ and $\mu_4 = \mathrm{E}X^4$.

(i) $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $\quad \mu_2 = 1$, $\quad \mu_4 = 3$, $\quad \alpha_4 = 3$.
(ii) $f(x) = \frac{1}{2}$, $-1 < x < 1$, $\quad \mu_2 = \frac{1}{3}$, $\quad \mu_4 = \frac{1}{5}$, $\quad \alpha_4 = \frac{9}{5}$.
(iii) $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$, $\quad \mu_2 = 2$, $\quad \mu_4 = 24$, $\quad \alpha_4 = 6$.

As a graph will show, (iii) is most peaked, (i) is next, and (ii) is least peaked.

2.29 a. For the binomial

$$\begin{aligned}
\mathrm{E}X(X-1) &= \sum_{x=2}^{n} x(x-1)\binom{n}{x}p^x(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n}\binom{n-2}{x}p^{x-2}(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{y=0}^{n-2}\binom{n-2}{y}p^{y}(1-p)^{n-2-y} = n(n-1)p^2,
\end{aligned}$$

where we use the identity $x(x-1)\binom{n}{x} = n(n-1)\binom{n-2}{x}$, substitute $y = x-2$ and recognize that the new sum is equal to 1. Similarly, for the Poisson

$$\mathrm{E}X(X-1) = \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2 \sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!} = \lambda^2,$$

where we substitute $y = x-2$.

b. $\mathrm{Var}(X) = \mathrm{E}[X(X-1)] + \mathrm{E}X - (\mathrm{E}X)^2$. For the binomial

$$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

For the Poisson

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

c.

$$\mathrm{E}Y = \sum_{y=0}^{n} y\frac{a}{y+a}\binom{n}{y}\frac{\binom{a+b-1}{a}}{\binom{n+a+b-1}{y+a}} = \sum_{y=1}^{n} n\frac{a}{(y-1)+(a+1)}\binom{n-1}{y-1}\frac{\binom{a+b-1}{a}}{\binom{(n-1)+(a+1)+b-1}{(y-1)+(a+1)}}$$

$$
\begin{aligned}
&= \sum_{y=1}^{n} n \frac{a}{(y-1)+(a+1)} \binom{n-1}{y-1} \frac{\binom{a+b-1}{a}}{\binom{(n-1)+(a+1)+b-1}{(y-1)+(a+1)}} \\
&= \frac{\frac{na}{a+1}\binom{a+b-1}{a}}{\binom{a+1+b-1}{a+1}} \sum_{y=1}^{n} \frac{a+1}{(y-1)+(a+1)} \binom{n-1}{y-1} \frac{\binom{a+1+b-1}{a+1}}{\binom{(n-1)+(a+1)+b-1}{(y-1)+(a+1)}} \\
&= \frac{na}{a+b} \sum_{j=0}^{n-1} \frac{a+1}{j+(a+1)} \binom{n-1}{j} \frac{\binom{a+1+b-1}{a+1}}{\binom{(n-1)+(a+1)+b-1}{(j+(a+1))}} = \frac{na}{a+b},
\end{aligned}
$$

since the last summation is 1, being the sum over all possible values of a beta-binomial$(n-1, a+1, b)$. $\mathrm{E}[Y(Y-1)] = \frac{n(n-1)a(a+1)}{(a+b)(a+b+1)}$ is calculated similar to EY, but using the identity $y(y-1)\binom{n}{y} = n(n-1)\binom{n-2}{y-2}$ and adding 2 instead of 1 to the parameter a. The sum over all possible values of a beta-binomial$(n-2, a+2, b)$ will appear in the calculation. Therefore

$$
\mathrm{Var}(Y) = \mathrm{E}[Y(Y-1)] + \mathrm{E}Y - (\mathrm{E}Y)^2 = \frac{nab(n+a+b)}{(a+b)^2(a+b+1)}.
$$

2.30 a. $\mathrm{E}(e^{tX}) = \int_0^c e^{tx}\frac{1}{c}dx = \frac{1}{ct}e^{tx}\Big|_0^c = \frac{1}{ct}e^{tc} - \frac{1}{ct}1 = \frac{1}{ct}(e^{tc}-1)$.

b. $\mathrm{E}(e^{tX}) = \int_0^c \frac{2x}{c^2}e^{tx}dx = \frac{2}{c^2t^2}(cte^{tc} - e^{tc} + 1)$. (integration-by-parts)

c.

$$
\begin{aligned}
\mathrm{E}(e^{tx}) &= \int_{-\infty}^{\alpha} \frac{1}{2\beta} e^{(x-\alpha)/\beta} e^{tx} dx + \int_{\alpha}^{\infty} \frac{1}{2\beta} e^{-(x-\alpha)/\beta} e^{tx} dx \\
&= \frac{e^{-\alpha/\beta}}{2\beta} \frac{1}{(\frac{1}{\beta}+t)} e^{x(\frac{1}{\beta}+t)}\Bigg|_{-\infty}^{\alpha} + -\frac{e^{\alpha/\beta}}{2\beta} \frac{1}{(\frac{1}{\beta}-t)} e^{-x(\frac{1}{\beta}-t)}\Bigg|_{\alpha}^{\infty} \\
&= \frac{4e^{\alpha t}}{4-\beta^2 t^2}, \quad -2/\beta < t < 2/\beta.
\end{aligned}
$$

d. $\mathrm{E}\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx}\binom{r+x-1}{x}p^r(1-p)^x = p^r\sum_{x=0}^{\infty}\binom{r+x-1}{x}\left((1-p)e^t\right)^x$. Now use the fact that $\sum_{x=0}^{\infty}\binom{r+x-1}{x}\left((1-p)e^t\right)^x\left(1-(1-p)e^t\right)^r = 1$ for $(1-p)e^t < 1$, since this is just the sum of this pmf, to get $\mathrm{E}(e^{tX}) = \left(\frac{p}{1-(1-p)e^t}\right)^r$, $t < -\log(1-p)$.

2.31 Since the mgf is defined as $M_X(t) = \mathrm{E}e^{tX}$, we necessarily have $M_X(0) = \mathrm{E}e^0 = 1$. But $t/(1-t)$ is 0 at $t=0$, therefore it cannot be an mgf.

2.32

$$
\frac{d}{dt}S(t)\Bigg|_{t=0} = \frac{d}{dt}\left(\log(M_x(t))\right)\Bigg|_{t=0} = \frac{\frac{d}{dt}M_x(t)}{M_x(t)}\Bigg|_{t=0} = \frac{\mathrm{E}X}{1} = \mathrm{E}X \quad \left(\text{since } M_X(0) = \mathrm{E}e^0 = 1\right)
$$

$$
\begin{aligned}
\frac{d^2}{dt^2}S(t)\Bigg|_{t=0} &= \frac{d}{dt}\left(\frac{M_x'(t)}{M_x(t)}\right)\Bigg|_{t=0} = \frac{M_x(t)M_x''(t) - [M_x'(t)]^2}{[M_x(t)]^2}\Bigg|_{t=0} \\
&= \frac{1\cdot\mathrm{E}X^2 - (\mathrm{E}X)^2}{1} = \mathrm{Var}X.
\end{aligned}
$$

2.33 a. $M_X(t) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{(e^t\lambda)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$.

$\mathrm{E}X = \frac{d}{dt}M_x(t)\big|_{t=0} = e^{\lambda(e^t-1)}\lambda e^t\Big|_{t=0} = \lambda$.

$\mathrm{E}X^2 = \frac{d^2}{dt^2}M_x(t)\Big|_{t=0} = \lambda e^t e^{\lambda(e^t-1)}\lambda e^t + \lambda e^t e^{\lambda(e^t-1)}\Big|_{t=0} = \lambda^2+\lambda.$
$\mathrm{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \lambda^2+\lambda-\lambda^2 = \lambda.$

b.

$$\begin{aligned}
M_x(t) &= \sum_{x=0}^{\infty} e^{tx}p(1-p)^x = p\sum_{x=0}^{\infty}((1-p)e^t)^x \\
&= p\frac{1}{1-(1-p)e^t} = \frac{p}{1-(1-p)e^t}, \quad t < -\log(1-p). \\
\mathrm{E}X &= \frac{d}{dt}M_x(t)\Big|_{t=0} = \frac{-p}{(1-(1-p)e^t)^2}\left(-(1-p)e^t\right)\Big|_{t=0} \\
&= \frac{p(1-p)}{p^2} = \frac{1-p}{p}. \\
\mathrm{E}X^2 &= \frac{d^2}{dt^2}M_x(t)\Big|_{t=0} \\
&= \left.\frac{\left(1-(1-p)e^t\right)^2\left(p(1-p)e^t\right) + p(1-p)e^t 2\left(1-(1-p)e^t\right)(1-p)e^t}{(1-(1-p)e^t)^4}\right|_{t=0} \\
&= \frac{p^3(1-p)+2p^2(1-p)^2}{p^4} = \frac{p(1-p)+2(1-p)^2}{p^2}. \\
\mathrm{Var}X &= \frac{p(1-p)+2(1-p)^2}{p^2} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

c. $M_x(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x^2-2\mu x-2\sigma^2 tx+\mu^2)/2\sigma^2}dx$. Now complete the square in the numerator by writing

$$\begin{aligned}
x^2-2\mu x-2\sigma^2 tx+\mu^2 &= x^2-2(\mu+\sigma^2 t)x \pm (\mu+\sigma^2 t)^2+\mu^2 \\
&= (x-(\mu+\sigma^2 t))^2-(\mu+\sigma^2 t)^2+\mu^2 \\
&= (x-(\mu+\sigma^2 t))^2-[2\mu\sigma^2 t+(\sigma^2 t)^2].
\end{aligned}$$

Then we have $M_x(t) = e^{[2\mu\sigma^2 t+(\sigma^2 t)^2]/2\sigma^2}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\mu+\sigma^2 t))^2}dx = e^{\mu t+\frac{\sigma^2 t^2}{2}}$.
$\mathrm{E}X = \frac{d}{dt}M_x(t)\big|_{t=0} = (\mu+\sigma^2 t)e^{\mu t+\sigma^2 t^2/2}\Big|_{t=0} = \mu.$
$\mathrm{E}X^2 = \frac{d^2}{dt^2}M_x(t)\Big|_{t=0} = (\mu+\sigma^2 t)^2 e^{\mu t+\sigma^2 t^2/2} + \sigma^2 e^{\mu t+\sigma^2 t/2}\Big|_{t=0} = \mu^2+\sigma^2.$
$\mathrm{Var}X = \mu^2+\sigma^2-\mu^2 = \sigma^2.$

2.35 a.

$$\begin{aligned}
\mathrm{E}X_1^r &= \int_0^{\infty} x^r\frac{1}{\sqrt{2\pi}x}e^{-(\log x)^2/2}dx && (f_1 \text{ is lognormal with } \mu=0,\ \sigma_2=1) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{y(r-1)}e^{-y^2/2}e^y dy && (\text{substitute } y=\log x,\ dy=(1/x)dx) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2+ry}dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\left(y^2-2ry+r^2\right)/2}e^{r^2/2}dy \\
&= e^{r^2/2}.
\end{aligned}$$

b.

$$\begin{aligned}
\int_0^\infty x^r f_1(x)\sin(2\pi\log x)dx &= \int_0^\infty x^r \frac{1}{\sqrt{2\pi}x}e^{-(\log x)^2/2}\sin(2\pi\log x)dx \\
&= \int_{-\infty}^\infty e^{(y+r)r}\frac{1}{\sqrt{2\pi}}e^{-(y+r)^2/2}\sin(2\pi y+2\pi r)dy \\
&\qquad\qquad (\text{substitute } y=\log x,\ dy=(1/x)dx) \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{(r^2-y^2)/2}\sin(2\pi y)dy \\
&\qquad\qquad (\sin(a+2\pi r)=\sin(a) \text{ if } r=0,1,2,\ldots) \\
&= 0,
\end{aligned}$$

because $e^{(r^2-y^2)/2}\sin(2\pi y) = -e^{(r^2-(-y)^2)/2}\sin(2\pi(-y))$; the integrand is an odd function so the negative integral cancels the positive one.

2.36 First, it can be shown that

$$\lim_{x\to\infty} e^{tx-(\log x)^2} = \infty$$

by using l'Hôpital's rule to show

$$\lim_{x\to\infty}\frac{tx-(\log x)^2}{tx} = 1,$$

and, hence,

$$\lim_{x\to\infty} tx-(\log x)^2 = \lim_{x\to\infty} tx = \infty.$$

Then for any $k>0$, there is a constant c such that

$$\int_k^\infty \frac{1}{x}e^{tx}e^{(\log x)^2/2}dx \ge c\int_k^\infty \frac{1}{x}dx = c\log x|_k^\infty = \infty.$$

Hence $M_x(t)$ does not exist.

2.37 a. The graph looks very similar to Figure 2.3.2 except that f_1 is symmetric around 0 (since it is standard normal).

b. The functions look like $t^2/2$ – it is impossible to see any difference.

c. The mgf of f_1 is $e^{K_1(t)}$. The mgf of f_2 is $e^{K_2(t)}$.

d. Make the transformation $y=e^x$ to get the densities in Example 2.3.10.

2.39 a. $\frac{d}{dx}\int_0^x e^{-\lambda t}dt = e^{-\lambda x}$. Verify

$$\frac{d}{dx}\left[\int_0^x e^{-\lambda t}dt\right] = \frac{d}{dx}\left[-\frac{1}{\lambda}e^{-\lambda t}\Big|_0^x\right] = \frac{d}{dx}\left(-\frac{1}{\lambda}e^{-\lambda x}+\frac{1}{\lambda}\right) = e^{-\lambda x}.$$

b. $\frac{d}{d\lambda}\int_0^\infty e^{-\lambda t}dt = \int_0^\infty \frac{d}{d\lambda}e^{-\lambda t}dt = \int_0^\infty -te^{-\lambda t}dt = -\frac{\Gamma(2)}{\lambda^2} = -\frac{1}{\lambda^2}$. Verify

$$\frac{d}{d\lambda}\int_0^\infty e^{-\lambda t}dt = \frac{d}{d\lambda}\frac{1}{\lambda} = -\frac{1}{\lambda^2}.$$

c. $\frac{d}{dt}\int_t^1 \frac{1}{x^2}dx = -\frac{1}{t^2}$. Verify

$$\frac{d}{dt}\left[\int_t^1 \frac{1}{x^2}dx\right] = \frac{d}{dt}\left(-\frac{1}{x}\Big|_t^1\right) = \frac{d}{dt}\left(-1+\frac{1}{t}\right) = -\frac{1}{t^2}.$$

d. $\frac{d}{dt}\int_1^\infty \frac{1}{(x-t)^2}dx = \int_1^\infty \frac{d}{dt}\left(\frac{1}{(x-t)^2}\right)dx = \int_1^\infty 2(x-t)^{-3}dx = -(x-t)^{-2}\Big|_1^\infty = \frac{1}{(1-t)^2}$. Verify

$$\frac{d}{dt}\int_1^\infty (x-t)^{-2}dx = \frac{d}{dt}\left[-(x-t)^{-1}\Big|_1^\infty\right] = \frac{d}{dt}\frac{1}{1-t} = \frac{1}{(1-t)^2}.$$

Chapter 3

Common Families of Distributions

3.1 The pmf of X is $f(x) = \frac{1}{N_1-N_0+1}$, $x = N_0, N_0+1, \ldots, N_1$. Then

$$\begin{aligned}
\mathrm{E}X &= \sum_{x=N_0}^{N_1} x \frac{1}{N_1-N_0+1} = \frac{1}{N_1-N_0+1}\left(\sum_{x=1}^{N_1} x - \sum_{x=1}^{N_0-1} x\right) \\
&= \frac{1}{N_1-N_0+1}\left(\frac{N_1(N_1+1)}{2} - \frac{(N_0-1)(N_0-1+1)}{2}\right) \\
&= \frac{N_1+N_0}{2}.
\end{aligned}$$

Similarly, using the formula for $\sum_1^N x^2$, we obtain

$$\begin{aligned}
\mathrm{E}x^2 &= \frac{1}{N_1-N_0+1}\left(\frac{N_1(N_1+1)(2N_1+1) - N_0(N_0-1)(2N_0-1)}{6}\right) \\
\mathrm{Var}X &= \mathrm{E}X^2 - \mathrm{E}X = \frac{(N_1-N_0)(N_1-N_0+2)}{12}.
\end{aligned}$$

3.2 Let X = number of defective parts in the sample. Then $X \sim$ hypergeometric$(N = 100, M, K)$ where M = number of defectives in the lot and K = sample size.

a. If there are 6 or more defectives in the lot, then the probability that the lot is accepted ($X = 0$) is at most

$$P(X = 0 \mid M = 100, N = 6, K) = \frac{\binom{6}{0}\binom{94}{K}}{\binom{100}{K}} = \frac{(100-K)\cdot \cdots \cdot(100-K-5)}{100\cdot\cdots\cdot 95}.$$

By trial and error we find $P(X = 0) = .10056$ for $K = 31$ and $P(X = 0) = .09182$ for $K = 32$. So the sample size must be at least 32.

b. Now $P(\text{accept lot}) = P(X = 0 \text{ or } 1)$, and, for 6 or more defectives, the probability is at most

$$P(X = 0 \text{ or } 1 \mid M = 100, N = 6, K) = \frac{\binom{6}{0}\binom{94}{K}}{\binom{100}{K}} + \frac{\binom{6}{1}\binom{94}{K-1}}{\binom{100}{K}}.$$

By trial and error we find $P(X = 0 \text{ or } 1) = .10220$ for $K = 50$ and $P(X = 0 \text{ or } 1) = .09331$ for $K = 51$. So the sample size must be at least 51.

3.3 In the seven seconds for the event, no car must pass in the last three seconds, an event with probability $(1-p)^3$. The only occurrence in the first four seconds, for which the pedestrian does not wait the entire four seconds, is to have a car pass in the first second and no other car pass. This has probability $p(1-p)^3$. Thus the probability of waiting exactly four seconds before starting to cross is $[1-p(1-p)^3](1-p)^3$.

3.5 Let X = number of effective cases. If the new and old drugs are equally effective, then the probability that the new drug is effective on a case is .8. If the cases are independent then $X \sim$ binomial$(100, .8)$, and

$$P(X \geq 85) = \sum_{x=85}^{100} \binom{100}{x} .8^x .2^{100-x} = .1285.$$

So, even if the new drug is no better than the old, the chance of 85 or more effective cases is not too small. Hence, we cannot conclude the new drug is better. Note that using a normal approximation to calculate this binomial probability yields $P(X \geq 85) \approx P(Z \geq 1.125) =$ .1303.

3.7 Let $X \sim$ Poisson(λ). We want $P(X \geq 2) \geq .99$, that is,

$$P(X \leq 1) = e^{-\lambda} + \lambda e^{-\lambda} \leq .01.$$

Solving $e^{-\lambda} + \lambda e^{-\lambda} = .01$ by trial and error (numerical bisection method) yields $\lambda = 6.6384$.

3.8 a. We want $P(X > N) < .01$ where $X \sim$ binomial$(1000, 1/2)$. Since the 1000 customers choose randomly, we take $p = 1/2$. We thus require

$$P(X > N) = \sum_{x=N+1}^{1000} \binom{1000}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{1000-x} < .01$$

which implies that

$$\left(\frac{1}{2}\right)^{1000} \sum_{x=N+1}^{1000} \binom{1000}{x} < .01.$$

This last inequality can be used to solve for N, that is, N is the smallest integer that satisfies

$$\left(\frac{1}{2}\right)^{1000} \sum_{x=N+1}^{1000} \binom{1000}{x} < .01.$$

The solution is $N = 537$.

b. To use the normal approximation we take $X \sim$ n$(500, 250)$, where we used $\mu = 1000(\frac{1}{2}) = 500$ and $\sigma^2 = 1000(\frac{1}{2})(\frac{1}{2}) = 250$.Then

$$P(X > N) = P\left(\frac{X - 500}{\sqrt{250}} > \frac{N - 500}{\sqrt{250}}\right) < .01$$

thus,

$$P\left(Z > \frac{N - 500}{\sqrt{250}}\right) < .01$$

where $Z \sim$ n$(0, 1)$. From the normal table we get

$$\begin{aligned} P(Z > 2.33) \approx .0099 < .01 \quad &\Rightarrow \quad \frac{N - 500}{\sqrt{250}} = 2.33 \\ &\Rightarrow \quad N \approx 537. \end{aligned}$$

Therefore, each theater should have at least 537 seats, and the answer based on the approximation equals the exact answer.

3.9 a. We can think of each one of the 60 children entering kindergarten as 60 independent Bernoulli trials with probability of success (a twin birth) of approximately $\frac{1}{90}$. The probability of having 5 or more successes approximates the probability of having 5 or more sets of twins entering kindergarten. Then $X \sim \text{binomial}(60, \frac{1}{90})$ and

$$P(X \geq 5) = 1 - \sum_{x=0}^{4} \binom{60}{x} \left(\frac{1}{90}\right)^x \left(1 - \frac{1}{90}\right)^{60-x} = .0006,$$

which is small and may be rare enough to be newsworthy.

b. Let X be the number of elementary schools in New York state that have 5 or more sets of twins entering kindergarten. Then the probability of interest is $P(X \geq 1)$ where $X \sim \text{binomial}(310,.0006)$. Therefore $P(X \geq 1) = 1 - P(X = 0) = .1698$.

c. Let X be the number of States that have 5 or more sets of twins entering kindergarten during any of the last ten years. Then the probability of interest is $P(X \geq 1)$ where $X \sim \text{binomial}(500, .1698)$. Therefore $P(X \geq 1) = 1 - P(X = 0) = 1 - 3.90 \times 10^{-41} \approx 1$.

3.11 a.

$$\begin{aligned} &\lim_{M/N\to p, M\to\infty, N\to\infty} \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\ &= \frac{K!}{x!(K-x)!} \lim_{M/N\to p, M\to\infty, N\to\infty} \frac{M!(N-M)!(N-K)!}{N!(M-x)!(N-M-(K-x))!} \end{aligned}$$

In the limit, each of the factorial terms can be replaced by the approximation from Stirling's formula because, for example,

$$M! = (M!/(\sqrt{2\pi}M^{M+1/2}e^{-M}))\sqrt{2\pi}M^{M+1/2}e^{-M}$$

and $M!/(\sqrt{2\pi}M^{M+1/2}e^{-M}) \to 1$. When this replacement is made, all the $\sqrt{2\pi}$ and exponential terms cancel. Thus,

$$\begin{aligned} &\lim_{M/N\to p, M\to\infty, N\to\infty} \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\ &= \binom{K}{x} \lim_{M/N\to p, M\to\infty, N\to\infty} \frac{M^{M+1/2}(N-M)^{N-M+1/2}(N-K)^{N-K+1/2}}{N^{N+1/2}(M-x)^{M-x+1/2}(N-M-K+x)^{N-M-(K-x)+1/2}}. \end{aligned}$$

We can evaluate the limit by breaking the ratio into seven terms, each of which has a finite limit we can evaluate. In some limits we use the fact that $M \to \infty$, $N \to \infty$ and $M/N \to p$ imply $N - M \to \infty$. The first term (of the seven terms) is

$$\lim_{M\to\infty} \left(\frac{M}{M-x}\right)^M = \lim_{M\to\infty} \frac{1}{\left(\frac{M-x}{M}\right)^M} = \lim_{M\to\infty} \frac{1}{\left(1+\frac{-x}{M}\right)^M} = \frac{1}{e^{-x}} = e^x.$$

Lemma 2.3.14 is used to get the penultimate equality. Similarly we get two more terms,

$$\lim_{N-M\to\infty} \left(\frac{N-M}{N-M-(K-x)}\right)^{N-M} = e^{K-x}$$

and

$$\lim_{N\to\infty} \left(\frac{N-K}{N}\right)^N = e^{-K}.$$

Note, the product of these three limits is one. Three other terms are

$$\begin{aligned}
\lim M \to \infty \left(\frac{M}{M-x}\right)^{1/2} &= 1 \\
\lim_{N-M\to\infty} \left(\frac{N-M}{N-M-(K-x)}\right)^{1/2} &= 1
\end{aligned}$$

and

$$\lim_{N\to\infty} \left(\frac{N-K}{N}\right)^{1/2} = 1.$$

The only term left is

$$\begin{aligned}
&\lim_{M/N\to p, M\to\infty, N\to\infty} \frac{(M-x)^x (N-M-(K-x))^{K-x}}{(N-K)^K} \\
&= \lim_{M/N\to p, M\to\infty, N\to\infty} \left(\frac{M-x}{N-K}\right)^x \left(\frac{N-M-(K-x)}{N-K}\right)^{K-x} \\
&= p^x(1-p)^{K-x}.
\end{aligned}$$

b. If in (a) we in addition have $K \to \infty$, $p \to 0$, $MK/N \to pK \to \lambda$, by the Poisson approximation to the binomial, we heuristically get

$$\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \to \binom{K}{x} p^x (1-p)^{K-x} \to \frac{e^{-\lambda}\lambda^x}{x!}.$$

c. Using Stirling's formula as in (a), we get

$$\begin{aligned}
&\lim_{N,M,K\to\infty, \frac{M}{N}\to 0, \frac{KM}{N}\to\lambda} \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\
&= \lim_{N,M,K\to\infty, \frac{M}{N}\to 0, \frac{KM}{N}\to\lambda} \frac{e^{-x}}{x!} \frac{K^x e^x M^x e^x (N-M)^{K-x} e^{K-x}}{N^K e^K} \\
&= \frac{1}{x!} \lim_{N,M,K\to\infty, \frac{M}{N}\to 0, \frac{KM}{N}\to\lambda} \left(\frac{KM}{N}\right)^x \left(\frac{N-M}{N}\right)^{K-x} \\
&= \frac{1}{x!}\lambda^x \lim_{N,M,K\to\infty, \frac{M}{N}\to 0, \frac{KM}{N}\to\lambda} \left(1 - \frac{\frac{MK}{N}}{K}\right)^K \\
&= \frac{e^{-\lambda}\lambda^x}{x!}.
\end{aligned}$$

3.12 Consider a sequence of Bernoulli trials with success probability p. Define X = number of successes in first n trials and Y = number of failures before the rth success. Then X and Y have the specified binomial and hypergeometric distributions, respectively. And we have

$$\begin{aligned}
F_x(r-1) &= P(X \le r-1) \\
&= P(r\text{th success on } (n+1)\text{st or later trial}) \\
&= P(\text{at least } n+1-r \text{ failures before the } r\text{th success}) \\
&= P(Y \ge n-r+1) \\
&= 1 - P(Y \le n-r) \\
&= 1 - F_Y(n-r).
\end{aligned}$$

3.13 For any X with support $0,1,\ldots$, we have the mean and variance of the 0−*truncated* X_T are given by

$$\begin{aligned}
\mathrm{E}X_T &= \sum_{x=1}^{\infty} xP(X_T=x) = \sum_{x=1}^{\infty} x\frac{P(X=x)}{P(X>0)} \\
&= \frac{1}{P(X>0)}\sum_{x=1}^{\infty} xP(X=x) = \frac{1}{P(X>0)}\sum_{x=0}^{\infty} xP(X=x) = \frac{\mathrm{E}X}{P(X>0)}.
\end{aligned}$$

In a similar way we get $\mathrm{E}X_T^2 = \frac{\mathrm{E}X^2}{P(X>0)}$. Thus,

$$\mathrm{Var}X_T = \frac{\mathrm{E}X^2}{P(X>0)} - \left(\frac{\mathrm{E}X}{P(X>0)}\right)^2.$$

a. For Poisson(λ), $P(X>0) = 1-P(X=0) = 1-\frac{e^{-\lambda}\lambda^0}{0!} = 1-e^{-\lambda}$, therefore

$$\begin{aligned}
P(X_T=x) &= \frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})} \quad x=1,2,\ldots \\
\mathrm{E}X_T &= \lambda/(1-e^{-\lambda}) \\
\mathrm{Var}X_T &= (\lambda^2+\lambda)/(1-e^{-\lambda}) - (\lambda/(1-e^{-\lambda}))^2.
\end{aligned}$$

b. For negative binomial(r,p), $P(X>0) = 1-P(X=0) = 1-\binom{r-1}{0}p^r(1-p)^0 = 1-p^r$. Then

$$\begin{aligned}
P(X_T=x) &= \frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r}, \quad x=1,2,\ldots \\
\mathrm{E}X_T &= \frac{r(1-p)}{p(1-p^r)} \\
\mathrm{Var}X_T &= \frac{r(1-p)+r^2(1-p)^2}{p^2(1-p^r)} - \left[\frac{r(1-p)}{p(1-p^r)^2}\right].
\end{aligned}$$

3.14 a. $\sum_{x=1}^{\infty}\frac{-(1-p)^x}{x\log p} = \frac{1}{\log p}\sum_{x=1}^{\infty}\frac{-(1-p)^x}{x} = 1$, since the sum is the Taylor series for $\log p$.

b.

$$\mathrm{E}X = \frac{-1}{\log p}\left[\sum_{x=1}^{\infty}(1-p)^x\right] = \frac{-1}{\log p}\left[\sum_{x=0}^{\infty}(1-p)^x - 1\right] == \frac{-1}{\log p}\left[\frac{1}{p}-1\right] = \frac{-1}{\log p}\left(\frac{1-p}{p}\right).$$

Since the geometric series converges uniformly,

$$\begin{aligned}
\mathrm{E}X^2 &= \frac{-1}{\log p}\sum_{x=1}^{\infty}x(1-p)^x = \frac{(1-p)}{\log p}\sum_{x=1}^{\infty}\frac{d}{dp}(1-p)^x \\
&= \frac{(1-p)}{\log p}\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^x = \frac{(1-p)}{\log p}\frac{d}{dp}\left[\frac{1-p}{p}\right] = \frac{-(1-p)}{p^2\log p}.
\end{aligned}$$

Thus

$$\mathrm{Var}X = \frac{-(1-p)}{p^2\log p}\left[1+\frac{(1-p)}{\log p}\right].$$

Alternatively, the mgf can be calculated,

$$M_x(t) = \frac{-1}{\log p}\sum_{x=1}^{\infty}\left[(1-p)e^t\right]^x = \frac{\log(1+pe^t-e^t)}{\log p}$$

and can be differentiated to obtain the moments.

3.15 The moment generating function for the negative binomial is

$$M(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r=\left(1+\frac{1}{r}\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\right)^r,$$

the term

$$\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\rightarrow\frac{\lambda(e^t-1)}{1}=\lambda(e^t-1)\quad\text{as } r\rightarrow\infty,\ p\rightarrow 1 \text{ and } r(p-1)\rightarrow\lambda.$$

Thus by Lemma 2.3.14, the negative binomial moment generating function converges to $e^{\lambda(e^t-1)}$, the Poisson moment generating function.

3.16 a. Using integration by parts with, $u=t^\alpha$ and $dv=e^{-t}dt$, we obtain

$$\Gamma(\alpha+1)=\int_0^\infty t^{(\alpha+1)-1}e^{-t}dt=t^\alpha(-e^{-t})\Big|_0^\infty-\int_0^\infty \alpha t^{\alpha-1}(-e^{-t})dt=0+\alpha\Gamma(\alpha)=\alpha\Gamma(\alpha).$$

b. Making the change of variable $z=\sqrt{2t}$, i.e., $t=z^2/2$, we obtain

$$\Gamma(1/2)=\int_0^\infty t^{-1/2}e^{-t}dt=\int_0^\infty\frac{\sqrt{2}}{z}e^{-z^2/2}zdz=\sqrt{2}\int_0^\infty e^{-z^2/2}dz=\sqrt{2}\frac{\sqrt{\pi}}{\sqrt{2}}=\sqrt{\pi}.$$

where the penultimate equality uses (3.3.14).

3.17

$$\begin{aligned}\mathrm{E}X^\nu &= \int_0^\infty x^\nu\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{(\nu+\alpha)-1}e^{-x/\beta}dx\\ &= \frac{\Gamma(\nu+\alpha)\beta^{\nu+\alpha}}{\Gamma(\alpha)\beta^\alpha} = \frac{\beta^\nu\Gamma(\nu+\alpha)}{\Gamma(\alpha)}.\end{aligned}$$

Note, this formula is valid for all $\nu>-\alpha$. The expectation does not exist for $\nu\le-\alpha$.

3.18 If $Y\sim$ negative binomial(r,p), its moment generating function is $M_Y(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r$, and, from Theorem 2.3.15, $M_{pY}(t)=\left(\frac{p}{1-(1-p)e^{pt}}\right)^r$. Now use L'Hôpital's rule to calculate

$$\lim_{p\rightarrow 0}\left(\frac{p}{1-(1-p)e^{pt}}\right)=\lim_{p\rightarrow 0}\frac{1}{(p-1)te^{pt}+e^{pt}}=\frac{1}{1-t},$$

so the moment generating function converges to $(1-t)^{-r}$, the moment generating function of a gamma$(r,1)$.

3.19 Repeatedly apply the integration-by-parts formula

$$\frac{1}{\Gamma(n)}\int_x^\infty z^{n-1}z^{-z}dz=\frac{x^{n-1}e^{-x}}{(n-1)!}+\frac{1}{\Gamma(n-1)}\int_x^\infty z^{n-2}z^{-z}dz,$$

until the exponent on the second integral is zero. This will establish the formula. If $X\sim$ gamma$(\alpha,1)$ and $Y\sim$ Poisson(x). The probabilistic relationship is $P(X\ge x)=P(Y\le\alpha-1)$.

3.21 The moment generating function would be defined by $\frac{1}{\pi}\int_{-\infty}^\infty\frac{e^{tx}}{1+x^2}dx$. On $(0,\infty)$, $e^{tx}>x$, hence

$$\int_0^\infty\frac{e^{tx}}{1+x^2}dx>\int_0^\infty\frac{x}{1+x^2}dx=\infty,$$

thus the moment generating function does not exist.

3.22 a.

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \quad (\text{let } y = x-2) \\
&= e^{-\lambda}\lambda^2 \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \;=\; e^{-\lambda}\lambda^2 e^{\lambda} \;=\; \lambda^2 \\
\mathrm{E}X^2 &= \lambda^2 + \mathrm{E}X \;=\; \lambda^2 + \lambda \\
\mathrm{Var}X &= \mathrm{E}X^2 - (\mathrm{E}X)^2 \;=\; \lambda^2 + \lambda - \lambda^2 \;=\; \lambda.
\end{aligned}$$

b.

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1)\binom{r+x-1}{x} pr(1-p)^x \\
&= \sum_{x=2}^{\infty} r(r+1)\binom{r+x-1}{x-2} pr(1-p)^x \\
&= r(r+1)\frac{(1-p)^2}{p^2} \sum_{y=0}^{\infty} \binom{r+2+y-1}{y} pr+2(1-p)^y \\
&= r(r-1)\frac{(1-p)^2}{p^2},
\end{aligned}$$

where in the second equality we substituted $y = x-2$, and in the third equality we use the fact that we are summing over a negative binomial$(r+2, p)$ pmf. Thus,

$$\begin{aligned}
\mathrm{Var}X &= \mathrm{E}X(X-1) + \mathrm{E}X - (\mathrm{E}X)^2 \\
&= r(r+1)\frac{(1-p)^2}{p^2} + \frac{r(1-p)}{p} - \frac{r^2(1-p)^2}{p^2} \\
&= \frac{r(1-p)}{p^2}.
\end{aligned}$$

c.

$$\begin{aligned}
\mathrm{E}X^2 &= \int_0^{\infty} x^2 \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx \;=\; \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha+1} e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Gamma(\alpha+2)\beta^{\alpha+2} \;=\; \alpha(\alpha+1)\beta^2. \\
\mathrm{Var}X &= \mathrm{E}X^2 - (\mathrm{E}X)^2 \;=\; \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 \;=\; \alpha\beta^2.
\end{aligned}$$

d. (Use 3.3.18)

$$\begin{aligned}
\mathrm{E}X &= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)} \;=\; \frac{\alpha\Gamma(\alpha)\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} \;=\; \frac{\alpha}{\alpha+\beta}. \\
\mathrm{E}X^2 &= \frac{\Gamma(\alpha+2)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)\Gamma(\alpha)} \;=\; \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\alpha+\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} \;=\; \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}. \\
\mathrm{Var}X &= \mathrm{E}X^2 - (\mathrm{E}X)^2 \;=\; \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \;=\; \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

e. The double exponential(μ,σ) pdf is symmetric about μ. Thus, by Exercise 2.26, $\mathrm{E}X=\mu$.

$$\begin{aligned}\mathrm{Var}X &= \int_{-\infty}^{\infty}(x-\mu)^2\frac{1}{2\sigma}e^{-|x-\mu|/\sigma}dx \;=\; \int_{-\infty}^{\infty}\sigma z^2\frac{1}{2}e^{-|z|}\sigma dz\\ &= \sigma^2\int_0^{\infty}z^2e^{-z}dz \;=\; \sigma^2\Gamma(3) \;=\; 2\sigma^2.\end{aligned}$$

3.23 a.

$$\int_{\alpha}^{\infty}x^{-\beta-1}dx=\frac{-1}{\beta}x^{-\beta}\bigg|_{\alpha}^{\infty}=\frac{1}{\beta\alpha^{\beta}},$$

thus $f(x)$ integrates to 1 .

b. $\mathrm{E}X^n=\frac{\beta\alpha^n}{(n-\beta)}$, therefore

$$\begin{aligned}\mathrm{E}X &= \frac{\alpha\beta}{(1-\beta)}\\ \mathrm{E}X^2 &= \frac{\alpha\beta^2}{(2-\beta)}\\ \mathrm{Var}X &= \frac{\alpha\beta^2}{2-\beta}-\frac{(\alpha\beta)^2}{(1-\beta)^2}\end{aligned}$$

c. If $\beta<2$ the integral of the second moment is infinite.

3.24 a. $f_x(x)=\frac{1}{\beta}e^{-x/\beta}$, $x>0$. For $Y=X^{1/\gamma}$, $f_Y(y)=\frac{\gamma}{\beta}e^{-y^{\gamma}/\beta}y^{\gamma-1}$, $y>0$. Using the transformation $z=y^{\gamma}/\beta$, we calculate

$$\mathrm{E}Y^n=\frac{\gamma}{\beta}\int_0^{\infty}y^{\gamma+n-1}e^{-y^{\gamma}/\beta}dy=\beta^{n/\gamma}\int_0^{\infty}z^{n/\gamma}e^{-z}dz=\beta^{n/\gamma}\Gamma\left(\frac{n}{\gamma}+1\right).$$

Thus $\mathrm{E}Y=\beta^{1/\gamma}\Gamma(\frac{1}{\gamma}+1)$ and $\mathrm{Var}Y=\beta^{2/\gamma}\left[\Gamma\left(\frac{2}{\gamma}+1\right)-\Gamma^2\left(\frac{1}{\gamma}+1\right)\right]$.

b. $f_x(x)=\frac{1}{\beta}e^{-x/\beta}$, $x>0$. For $Y=(2X/\beta)^{1/2}$, $f_Y(y)=ye^{-y^2/2}$, $y>0$. We now notice that

$$\mathrm{E}Y=\int_0^{\infty}y^2e^{-y^2/2}dy=\frac{\sqrt{2\pi}}{2}$$

since $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^2e^{-y^2/2}=1$, the variance of a standard normal, and the integrand is symmetric. Use integration-by-parts to calculate the second moment

$$\mathrm{E}Y^2=\int_0^{\infty}y^3e^{-y^2/2}dy=2\int_0^{\infty}ye^{-y^2/2}dy=2,$$

where we take $u=y^2$, $dv=ye^{-y^2/2}$. Thus $\mathrm{Var}Y=2(1-\pi/4)$.

c. The gamma(a,b) density is

$$f_X(x)=\frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}.$$

Make the transformation $y=1/x$ with $dx=-dy/y^2$ to get

$$f_Y(y)=f_X(1/y)|1/y^2|=\frac{1}{\Gamma(a)b^a}\left(\frac{1}{y}\right)^{a+1}e^{-1/by}.$$

The first two moments are

$$\begin{aligned} \mathrm{E}Y &= \frac{1}{\Gamma(a)b^a}\int_0^\infty \left(\frac{1}{y}\right)^a e^{-1/by} &= \frac{\Gamma(a-1)b^{a-1}}{\Gamma(a)b^a} &= \frac{1}{(a-1)b} \\ \mathrm{E}Y^2 &= \frac{\Gamma(a-2)b^{a-2}}{\Gamma(a)b^a} &= \frac{1}{(a-1)(a-2)b^2}, \end{aligned}$$

and so $\mathrm{Var}Y = \frac{1}{(a-1)^2(a-2)b^2}$.

d. $f_x(x) = \frac{1}{\Gamma(3/2)\beta^{3/2}}x^{3/2-1}e^{-x/\beta}$, $x>0$. For $Y=(X/\beta)^{1/2}$, $f_Y(y) = \frac{2}{\Gamma(3/2)}y^2e^{-y^2}$, $y>0$. To calculate the moments we use integration-by-parts with $u=y^2$, $dv = ye^{-y^2}$ to obtain

$$\mathrm{E}Y = \frac{2}{\Gamma(3/2)}\int_0^\infty y^3 e^{-y^2}dy \;=\; \frac{2}{\Gamma(3/2)}\int_0^\infty ye^{-y^2}dy \;=\; \frac{1}{\Gamma(3/2)}$$

and with $u=y^3, dv = ye^{-y^2}$ to obtain

$$\mathrm{E}Y^2 = \frac{2}{\Gamma(3/2)}\int_0^\infty y^4 e^{-y^2}dy \;=\; \frac{3}{\Gamma(3/2)}\int_0^\infty y^2e^{-y^2}dy \;=\; \frac{3}{\Gamma(3/2)}\sqrt{\pi}.$$

Using the fact that $\frac{1}{2\sqrt{\pi}}\int_{-\infty}^\infty y^2e^{-y^2}=1$, since it is the variance of a $\mathrm{n}(0,2)$, symmetry yields $\int_0^\infty y^2e^{-y^2}dy = \sqrt{\pi}$. Thus, $\mathrm{Var}Y = 6-4/\pi$, using $\Gamma(3/2)=\frac{1}{2}\sqrt{\pi}$.

e. $f_x(x) = e^{-x}$, $x>0$. For $Y = \alpha - \gamma\log X$, $f_Y(y) = e^{-e^{\frac{\alpha-y}{\gamma}}}e^{\frac{\alpha-y}{\gamma}}\frac{1}{\gamma}$, $-\infty<y<\infty$. Calculation of $\mathrm{E}Y$ and $\mathrm{E}Y^2$ cannot be done in closed form. If we define

$$I_1 = \int_0^\infty \log x e^{-x}dx, \qquad I_2 = \int_0^\infty (\log x)^2 e^{-x}dx,$$

then $\mathrm{E}Y = \mathrm{E}(\alpha-\gamma\log x) = \alpha-\gamma I_1$, and $\mathrm{E}Y^2 = \mathrm{E}(\alpha-\gamma\log x)^2 = \alpha^2 - 2\alpha\gamma I_1 + \gamma^2 I_2$.The constant $I_1 = .5772157$ is called Euler's constant.

3.25 Note that if T is continuous then,

$$\begin{aligned} P(t\le T\le t+\delta | t\le T) &= \frac{P(t\le T\le t+\delta, t\le T)}{P(t\le T)} \\ &= \frac{P(t\le T\le t+\delta)}{P(t\le T)} \\ &= \frac{F_T(t+\delta)-F_T(t)}{1-F_T(t)}. \end{aligned}$$

Therefore from the definition of derivative,

$$h_T(t) = \frac{1}{1-F_T(t)} = \lim_{\delta\to 0}\frac{F_T(t+\delta)-F_T(t)}{\delta} = \frac{F_T'(t)}{1-F_T(t)} = \frac{f_T(t)}{1-F_T(t)}.$$

Also,

$$-\frac{d}{dt}(\log[1-F_T(t)]) = -\frac{1}{1-F_T(t)}(-f_T(t)) = h_T(t).$$

3.26 a. $f_T(t) = \frac{1}{\beta}e^{-t/\beta}$ and $F_T(t) = \int_0^t \frac{1}{\beta}e^{-x/\beta}dx = -\left.e^{-x/\beta}\right|_0^t = 1-e^{-t/\beta}$. Thus,

$$h_T(t) = \frac{f_T(t)}{1-F_T(t)} = \frac{(1/\beta)e^{-t/\beta}}{1-(1-e^{-t/\beta})} = \frac{1}{\beta}.$$

b. $f_T(t) = \frac{\gamma}{\beta} t^{\gamma-1} e^{-t^\gamma/\beta}, t \geq 0$ and $F_T(t) = \int_0^t \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^\gamma/\beta} dx = \int_0^{t^{\gamma/\beta}} e^{-u} du = -\left. e^{-u} \right|_0^{t^{\gamma/\beta}} = 1 - e^{-t^\gamma/\beta}$, where $u = x^{\gamma/\beta}$. Thus,

$$h_T(t) = \frac{(\gamma/\beta) t^{\gamma-1} e^{-t^\gamma/\beta}}{e^{-t^\gamma/\beta}} = \frac{\gamma}{\beta} t^{\gamma-1}.$$

c. $F_T(t) = \frac{1}{1+e^{-(t-\mu)/\beta}}$ and $f_T(t) = \frac{e^{-(t-\mu)/\beta}}{\left(1+e^{-(t-\mu)/\beta}\right)^2}$. Thus,

$$h_T(t) = \frac{1}{\beta} e^{-(t-\mu)/\beta \left(1+e^{-(t-\mu)/\beta}\right)^2} \frac{1}{\frac{e^{-(t-\mu)/\beta}}{1+e^{-(t-\mu)/\beta}}} = \frac{1}{\beta} F_T(t).$$

3.27 a. The uniform pdf satisfies the inequalities of Exercise 2.27, hence is unimodal.

b. For the gamma(α, β) pdf $f(x)$, ignoring constants, $\frac{d}{dx} f(x) = \frac{x^{\alpha-2} e^{-x/\beta}}{\beta} \left[\beta(\alpha-1) - x\right]$, which only has one sign change. Hence the pdf is unimodal with mode $\beta(\alpha - 1)$.

c. For the n(μ, σ^2) pdf $f(x)$, ignoring constants, $\frac{d}{dx} f(x) = \frac{x-\mu}{\sigma^2} e^{-(-x/\beta)^2/2\sigma^2}$, which only has one sign change. Hence the pdf is unimodal with mode μ.

d. For the beta(α, β) pdf $f(x)$, ignoring constants,

$$\frac{d}{dx} f(x) = x^{\alpha-2} (1-x)^{\beta-2} \left[(\alpha-1) - x(\alpha+\beta-2)\right],$$

which only has one sign change. Hence the pdf is unimodal with mode $\frac{\alpha-1}{\alpha+\beta-2}$.

3.28 a. (i) μ known,

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right),$$

$h(x) = 1, \quad c(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} I_{(0,\infty)}(\sigma^2), \quad w_1(\sigma^2) = -\frac{1}{2\sigma^2}, \quad t_1(x) = (x-\mu)^2.$

(ii) σ^2 known,

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\mu \frac{x}{\sigma^2}\right),$$

$h(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right), \quad c(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\mu^2}{2\sigma^2}\right), \quad w_1(\mu) = \mu, \quad t_1(x) = \frac{x}{\sigma^2}.$

b. (i) α known,

$$f(x|\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{\frac{-x}{\beta}},$$

$h(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}, \ x > 0, \quad c(\beta) = \frac{1}{\beta^\alpha}, \quad w_1(\beta) = \frac{1}{\beta}, \quad t_1(x) = -x.$

(ii) β known,

$$f(x|\alpha) = e^{-x/\beta} \frac{1}{\Gamma(\alpha)\beta^\alpha} \exp((\alpha-1)\log x),$$

$h(x) = e^{-x/\beta}, \ x > 0, \quad c(\alpha) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \quad w_1(\alpha) = \alpha - 1, \quad t_1(x) = \log x.$

(iii) α, β unknown,

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \exp((\alpha-1)\log x - \frac{x}{\beta}),$$

$h(x) = I_{\{x>0\}}(x), \quad c(\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}, \quad w_1(\alpha) = \alpha - 1, \quad t_1(x) = \log x,$
$w_2(\alpha, \beta) = -1/\beta, \quad t_2(x) = x.$

c. (i) α known, $h(x) = x^{\alpha-1} I_{[0,1]}(x), \quad c(\beta) = \frac{1}{B(\alpha,\beta)}, \quad w_1(\beta) = \beta - 1, \quad t_1(x) = \log(1-x).$

(ii) β known, $h(x) = (1-x)^{\beta-1} I_{[0,1]}(x), \quad c(\alpha) = \frac{1}{B(\alpha,\beta)}, \quad w_1(x) = \alpha - 1, \quad t_1(x) = \log x.$

(iii) α, β unknown,
$h(x) = I_{[0,1]}(x),\ c(\alpha,\beta) = \frac{1}{B(\alpha,\beta)},\quad w_1(\alpha) = \alpha - 1,\quad t_1(x) = \log x,$
$w_2(\beta) = \beta - 1,\ t_2(x) = \log(1-x).$

d. $h(x) = \frac{1}{x!} I_{\{0,1,2,\ldots\}}(x),\quad c(\theta) = e^{-\theta},\quad w_1(\theta) = \log\theta,\quad t_1(x) = x.$

e. $h(x) = \binom{x-1}{r-1} I_{\{r,r+1,\ldots\}}(x),\quad c(p) = \left(\frac{p}{1-p}\right)^r,\quad w_1(p) = \log(1-p),\quad t_1(x) = x.$

3.29 a. For the $\mathrm{n}(\mu,\sigma^2)$

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)\left(\frac{e^{-\mu^2/2\sigma^2}}{\sigma}\right)\left(e^{-x^2/2\sigma^2 + x\mu/\sigma^2}\right),$$

so the natural parameter is $(\eta_1,\eta_2) = (-1/2\sigma^2, \mu/\sigma^2)$ with natural parameter space $\{(\eta_1,\eta_2){:}\eta_1 < 0, -\infty < \eta_2 < \infty\}$.

b. For the gamma(α,β),

$$f(x) = \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)\left(e^{(\alpha-1)\log x - x/\beta}\right),$$

so the natural parameter is $(\eta_1,\eta_2) = (\alpha-1, -1/\beta)$ with natural parameter space $\{(\eta_1,\eta_2){:}\eta_1 > -1, \eta_2 < 0\}$.

c. For the beta(α,β),

$$f(x) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)\left(e^{(\alpha-1)\log x + (\beta-1)\log(1-x)}\right),$$

so the natural parameter is $(\eta_1,\eta_2) = (\alpha-1, \beta-1)$ and the natural parameter space is $\{(\eta_1,\eta_2){:}\eta_1 > -1, \eta_2 > -1\}$.

d. For the Poisson

$$f(x) = \left(\frac{1}{x!}\right)\left(e^{-\theta}\right) e^{x\log\theta}$$

so the natural parameter is $\eta = \log\theta$ and the natural parameter space is $\{\eta{:}-\infty < \eta < \infty\}$.

e. For the negative binomial(r,p), r known,

$$P(X=x) = \binom{r+x-1}{x}\left(p^r\right)\left(e^{x\log(1-p)}\right),$$

so the natural parameter is $\eta = \log(1-p)$ with natural parameter space $\{\eta{:}\eta < 0\}$.

3.31 a.

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\int h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&= \int h(x)c'(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&\quad + \int h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)dx \\
&= \int h(x)\left[\frac{\partial}{\partial\theta_j}\log c(\theta)\right]c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx + \mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right] \\
&= \frac{\partial}{\partial\theta_j}\log c(\theta) + \mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right]
\end{aligned}$$

Therefore $\mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right] = -\frac{\partial}{\partial\theta_j}\log c(\theta).$

b.

$$
\begin{aligned}
0 &= \frac{\partial^2}{\partial\theta^2}\int h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&= \int h(x)c''(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&\quad+\int h(x)c'(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)dx \\
&\quad+\int h(x)c'(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)dx \\
&\quad+\int h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)^2dx \\
&\quad+\int h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)\left(\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(x)\right)dx \\
&= \int h(x)\left[\frac{\partial^2}{\partial\theta_j^2}\log c(\theta)\right]c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&\quad+\int h(x)\left[\frac{c'(\theta)}{c(\theta)}\right]^2c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)dx \\
&\quad+2\left(\frac{\partial}{\partial\theta_j}\log c(\theta)\right)\mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right] \\
&\quad+\mathrm{E}\left[(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x))^2\right]+\mathrm{E}\left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(x)\right] \\
&= \frac{\partial^2}{\partial\theta_j^2}\log c(\theta)+\left[\frac{\partial}{\partial\theta_j}\log c(\theta)\right]^2 \\
&\quad-2\mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right]\mathrm{E}\left[\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right] \\
&\quad+\mathrm{E}\left[(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x))^2\right]+\mathrm{E}\left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(x)\right] \\
&= \frac{\partial^2}{\partial\theta_j^2}\log c(\theta)+\mathrm{Var}\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)+\mathrm{E}\left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(x)\right].
\end{aligned}
$$

Therefore $\mathrm{Var}\left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial\theta_j}t_i(x)\right)=-\frac{\partial^2}{\partial\theta_j^2}\log c(\theta)-\mathrm{E}\left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(x)\right]$.

3.33 a. (i) $h(x)=e^xI_{\{-\infty<x<\infty\}}(x)$, $c(\theta)=\frac{1}{\sqrt{2\pi\theta}}\exp(\frac{-\theta}{2})\theta>0$, $w_1(\theta)=\frac{1}{2\theta}$, $t_1(x)=-x^2$.

(ii) The nonnegative real line.

b. (i) $h(x)=I_{\{-\infty<x<\infty\}}(x)$, $c(\theta)=\frac{1}{\sqrt{2\pi a\theta^2}}\exp(\frac{-1}{2a})-\infty<\theta<\infty, a>0$, $w_1(\theta)=\frac{1}{2a\theta^2}$, $w_2(\theta)=\frac{1}{a\theta}$, $t_1(x)=-x^2$, $t_2(x)=x$.

(ii) A parabola.

c. (i) $h(x) = \frac{1}{x} I_{\{0<x<\infty\}}(x)$, $c(\alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \alpha > 0$, $w_1(\alpha) = \alpha$, $w_2(\alpha) = \alpha$, $t_1(x) = \log(x)$, $t_2(x) = -x$.

(ii) A line.

d. (i) $h(x) = C \exp(x^4) I_{\{-\infty<x<\infty\}}(x)$, $c(\theta) = \exp(\theta^4) - \infty < \theta < \infty$, $w_1(\theta) = \theta$, $w_2(\theta) = \theta^2$, $w_3(\theta) = \theta^3$, $t_1(x) = -4x^3$, $t_2(x) = 6x^2$, $t_3(x) = -4x$.

(ii) The curve is a spiral in 3-space.

(iii) A good picture can be generated with the Mathematica statement

```
ParametricPlot3D[{t, t^2, t^3}, {t, 0, 1}, ViewPoint -> {1, -2, 2.5}].
```

3.35 a. In Exercise 3.34(a) $w_1(\lambda) = \frac{1}{2\lambda}$ and for a $\mathrm{n}(e^\theta, e^\theta)$, $w_1(\theta) = \frac{1}{2e^\theta}$.

b. $\mathrm{E}X = \mu = \alpha\beta$, then $\beta = \frac{\mu}{\alpha}$. Therefore $h(x) = \frac{1}{x} I_{\{0<x<\infty\}}(x)$, $c(\alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)(\frac{\mu}{\alpha})^\alpha}, \alpha > 0$, $w_1(\alpha) = \alpha$, $w_2(\alpha) = \frac{\alpha}{\mu}$, $t_1(x) = \log(x)$, $t_2(x) = -x$.

c. From (b) then $(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) = (\alpha_1, \ldots, \alpha_n, \frac{\alpha_1}{\mu}, \ldots, \frac{\alpha_n}{\mu})$

3.37 The pdf $(\frac{1}{\sigma}) f(\frac{(x-\mu)}{\sigma})$ is symmetric about μ because, for any $\epsilon > 0$,

$$\frac{1}{\sigma} f\left(\frac{(\mu+\epsilon)-\mu}{\sigma}\right) = \frac{1}{\sigma} f\left(\frac{\epsilon}{\sigma}\right) = \frac{1}{\sigma} f\left(-\frac{\epsilon}{\sigma}\right) = \frac{1}{\sigma} f\left(\frac{(\mu-\epsilon)-\mu}{\sigma}\right).$$

Thus, by Exercise 2.26b, μ is the median.

3.38 $P(X > x_\alpha) = P(\sigma Z + \mu > \sigma z_\alpha + \mu) = P(Z > z_\alpha)$ by Theorem 3.5.6.

3.39 First take $\mu = 0$ and $\sigma = 1$.

a. The pdf is symmetric about 0, so 0 must be the median. Verifying this, write

$$P(Z \geq 0) = \int_0^\infty \frac{1}{\pi} \frac{1}{1+z^2} dz = \frac{1}{\pi} \tan^{-1}(z) \Big|_0^\infty = \frac{1}{\pi} \left(\frac{\pi}{2} - 0\right) = \frac{1}{2}.$$

b. $P(Z \geq 1) = \frac{1}{\pi} \tan^{-1}(z) \Big|_1^\infty = \frac{1}{\pi} \left(\frac{\pi}{2} - \frac{\pi}{4}\right) = \frac{1}{4}$. By symmetry this is also equal to $P(Z \leq -1)$. Writing $z = (x - \mu)/\sigma$ establishes $P(X \geq \mu) = \frac{1}{2}$ and $P(X \geq \mu + \sigma) = \frac{1}{4}$.

3.40 Let $X \sim f(x)$ have mean μ and variance σ^2. Let $Z = \frac{X-\mu}{\sigma}$. Then

$$\mathrm{E}Z = \left(\frac{1}{\sigma}\right) \mathrm{E}(X - \mu) = 0$$

and

$$\mathrm{Var}Z = \mathrm{Var}\left(\frac{X - \mu}{\sigma}\right) = \left(\frac{1}{\sigma^2}\right) \mathrm{Var}(X - \mu) = \left(\frac{1}{\sigma^2}\right) \mathrm{Var}X = \frac{\sigma^2}{\sigma^2} = 1.$$

Then compute the pdf of Z, $f_Z(z) = f_x(\sigma z + \mu) \cdot \sigma = \sigma f_x(\sigma z + \mu)$ and use $f_Z(z)$ as the standard pdf.

3.41 a. This is a special case of Exercise 3.42a.

b. This is a special case of Exercise 3.42b.

3.42 a. Let $\theta_1 > \theta_2$. Let $X_1 \sim f(x - \theta_1)$ and $X_2 \sim f(x - \theta_2)$. Let $F(z)$ be the cdf corresponding to $f(z)$ and let $Z \sim f(z)$.Then

$$\begin{aligned} F(x \mid \theta_1) &= P(X_1 \leq x) = P(Z + \theta_1 \leq x) = P(Z \leq x - \theta_1) = F(x - \theta_1) \\ &\leq F(x - \theta_2) = P(Z \leq x - \theta_2) = P(Z + \theta_2 \leq x) = P(X_2 \leq x) \\ &= F(x \mid \theta_2). \end{aligned}$$

The inequality is because $x-\theta_2 > x-\theta_1$, and F is nondecreasing. To get strict inequality for some x, let $(a,b]$ be an interval of length $\theta_1-\theta_2$ with $P(a<Z\le b)=F(b)-F(a)>0$. Let $x=a+\theta_1$. Then

$$\begin{aligned} F(x\mid\theta_1) &= F(x-\theta_1) = F(a+\theta_1-\theta_1) = F(a) \\ &< F(b) = F(a+\theta_1-\theta_2) = F(x-\theta_2) = F(x\mid\theta_2). \end{aligned}$$

b. Let $\sigma_1>\sigma_2$. Let $X_1\sim f(x/\sigma_1)$ and $X_2\sim f(x/\sigma_2)$. Let $F(z)$ be the cdf corresponding to $f(z)$ and let $Z\sim f(z)$. Then, for $x>0$,

$$\begin{aligned} F(x\mid\sigma_1) &= P(X_1\le x) = P(\sigma_1 Z\le x) = P(Z\le x/\sigma_1) = F(x/\sigma_1) \\ &\le F(x/\sigma_2) = P(Z\le x/\sigma_2) = P(\sigma_2 Z\le x) = P(X_2\le x) \\ &= F(x\mid\sigma_2). \end{aligned}$$

The inequality is because $x/\sigma_2 > x/\sigma_1$ (because $x>0$ and $\sigma_1>\sigma_2>0$), and F is nondecreasing. For $x\le 0$, $F(x\mid\sigma_1)=P(X_1\le x)=0=P(X_2\le x)=F(x\mid\sigma_2)$. To get strict inequality for some x, let $(a,b]$ be an interval such that $a>0$, $b/a=\sigma_1/\sigma_2$ and $P(a<Z\le b)=F(b)-F(a)>0$. Let $x=a\sigma_1$. Then

$$\begin{aligned} F(x\mid\sigma_1) &= F(x/\sigma_1) = F(a\sigma_1/\sigma_1) = F(a) \\ &< F(b) = F(a\sigma_1/\sigma_2) = F(x/\sigma_2) \\ &= F(x\mid\sigma_2). \end{aligned}$$

3.43 a. $F_Y(y|\theta)=1-F_X(\frac{1}{y}|\theta)$ $y>0$, by Theorem 2.1.3. For $\theta_1>\theta_2$,

$$F_Y(y|\theta_1)=1-F_X\left(\frac{1}{y}\middle|\,\theta_1\right)\le 1-F_X\left(\frac{1}{y}\middle|\,\theta_2\right)=F_Y(y|\theta_2)$$

for all y, since $F_X(x|\theta)$ is stochastically increasing and if $\theta_1>\theta_2$, $F_X(x|\theta_2)\le F_X(x|\theta_1)$ for all x. Similarly, $F_Y(y|\theta_1)=1-F_X(\frac{1}{y}|\theta_1)<1-F_X(\frac{1}{y}|\theta_2)=F_Y(y|\theta_2)$ for some y, since if $\theta_1>\theta_2$, $F_X(x|\theta_2)<F_X(x|\theta_1)$ for some x. Thus $F_Y(y|\theta)$ is stochastically decreasing in θ.

b. $F_X(x|\theta)$ is stochastically increasing in θ. If $\theta_1>\theta_2$ and $\theta_1,\theta_2>0$ then $\frac{1}{\theta_2}>\frac{1}{\theta_1}$. Therefore $F_X(x|\frac{1}{\theta_1})\le F_X(x|\frac{1}{\theta_2})$ for all x and $F_X(x|\frac{1}{\theta_1})<F_X(x|\frac{1}{\theta_2})$ for some x. Thus $F_X(x|\frac{1}{\theta})$ is stochastically decreasing in θ.

3.44 The function $g(x)=|x|$ is a nonnegative function. So by Chebychev's Inequality,

$$P(|X|\ge b)\le \mathrm{E}|X|/b.$$

Also, $P(|X|\ge b)=P(X^2\ge b^2)$. Since $g(x)=x^2$ is also nonnegative, again by Chebychev's Inequality we have

$$P(|X|\ge b)=P(X^2\ge b^2)\le \mathrm{E}X^2/b^2.$$

For $X\sim$ exponential(1), $\mathrm{E}|X|=\mathrm{E}X=1$ and $\mathrm{E}X^2=\mathrm{Var}X+(\mathrm{E}X)^2=2$. For $b=3$,

$$\mathrm{E}|X|/b=1/3>2/9=\mathrm{E}X^2/b^2.$$

Thus $\mathrm{E}X^2/b^2$ is a better bound. But for $b=\sqrt{2}$,

$$\mathrm{E}|X|/b=1/\sqrt{2}<1=\mathrm{E}X^2/b^2.$$

Thus $\mathrm{E}|X|/b$ is a better bound.

3.45 a.

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \geq \int_a^{\infty} e^{tx} f_X(x)dx \\ &\geq e^{ta} \int_a^{\infty} f_X(x)dx = e^{ta} P(X \geq a), \end{aligned}$$

where we use the fact that e^{tx} is increasing in x for $t > 0$.

b.

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \geq \int_{-\infty}^{a} e^{tx} f_X(x)dx \\ &\geq e^{ta} \int_{-\infty}^{a} f_X(x)dx = e^{ta} P(X \leq a), \end{aligned}$$

where we use the fact that e^{tx} is decreasing in x for $t < 0$.

c. $h(t, x)$ must be nonnegative.

3.46 For $X \sim$ uniform$(0, 1)$, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$, thus

$$P(|X - \mu| > k\sigma) = 1 - P\left(\frac{1}{2} - \frac{k}{\sqrt{12}} \leq X \leq \frac{1}{2} + \frac{k}{\sqrt{12}}\right) = \begin{cases} 1 - \frac{2k}{\sqrt{12}} & k < \sqrt{3}, \\ 0 & k \geq \sqrt{3}, \end{cases}$$

For $X \sim$ exponential(λ), $\mu = \lambda$ and $\sigma^2 = \lambda^2$, thus

$$P(|X - \mu| > k\sigma) = 1 - P(\lambda - k\lambda \leq X \leq \lambda + k\lambda) = \begin{cases} 1 + e^{-(k+1)} - e^{k-1} & k \leq 1 \\ e^{-(k+1)} & k > 1. \end{cases}$$

From Example 3.6.2, Chebychev's Inequality gives the bound $P(|X - \mu| > k\sigma) \leq 1/k^2$.

Comparison of probabilities

k	u(0,1) exact	exp(λ) exact	Chebychev
.1	.942	.926	100
.5	.711	.617	4
1	.423	.135	1
1.5	.134	.0821	.44
$\sqrt{3}$	0	0.0651	.33
2	0	0.0498	.25
4	0	0.00674	.0625
10	0	0.0000167	.01

So we see that Chebychev's Inequality is quite conservative.

3.47

$$\begin{aligned} P(|Z| > t) &= 2P(Z > t) = 2\frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2}dx \\ &= \sqrt{\frac{2}{\pi}} \int_t^{\infty} \frac{1+x^2}{1+x^2} e^{-x^2/2}dx \\ &= \sqrt{\frac{2}{\pi}} \left[\int_t^{\infty} \frac{1}{1+x^2} e^{-x^2/2}dx + \int_t^{\infty} \frac{x^2}{1+x^2} e^{-x^2/2}dx\right]. \end{aligned}$$

To evaluate the second term, let $u = \frac{x}{1+x^2}$, $dv = xe^{-x^2/2}dx$, $v = -e^{-x^2/2}$, $du = \frac{1-x^2}{(1+x^2)^2}$, to obtain

$$\begin{aligned}\int_t^\infty \frac{x^2}{1+x^2}e^{-x^2/2}dx &= \frac{x}{1+x^2}(-e^{-x^2/2})\Big|_t^\infty - \int_t^\infty \frac{1-x^2}{(1+x^2)^2}(-e^{-x^2/2})dx \\ &= \frac{t}{1+t^2}e^{-t^2/2} + \int_t^\infty \frac{1-x^2}{(1+x^2)^2}e^{-x^2/2}dx.\end{aligned}$$

Therefore,

$$\begin{aligned}P(Z \geq t) &= \sqrt{\frac{2}{\pi}}\frac{t}{1+t^2}e^{-t^2/2} + \sqrt{\frac{2}{\pi}}\int_t^\infty \left(\frac{1}{1+x^2} + \frac{1-x^2}{(1+x^2)^2}\right)e^{-x^2/2}dx \\ &= \sqrt{\frac{2}{\pi}}\frac{t}{1+t^2}e^{-t^2/2} + \sqrt{\frac{2}{\pi}}\int_t^\infty \left(\frac{2}{(1+x^2)^2}\right)e^{-x^2/2}dx \\ &\geq \sqrt{\frac{2}{\pi}}\frac{t}{1+t^2}e^{-t^2/2}.\end{aligned}$$

3.48 For the negative binomial

$$P(X = x+1) = \binom{r+x+1-1}{x+1}p^r(1-p)^{x+1} = \left(\frac{r+x}{x+1}\right)(1-p)P(X=x).$$

For the hypergeometric

$$P(X = x+1) = \begin{cases} \frac{(M-x)(k-x+x+1)(x+1)}{P(X=x)} & \text{if } x < k,\ x < M,\ x \geq M-(N-k) \\ \frac{\binom{M}{x+1}\binom{N-M}{k-x-1}}{\binom{N}{k}} & \text{if } x = M-(N-k)-1 \\ 0 & \text{otherwise.}\end{cases}$$

3.49 a.

$$\mathrm{E}(g(X)(X-\alpha\beta)) = \int_0^\infty g(x)(x-\alpha\beta)\frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-x/\beta}dx.$$

Let $u = g(x)$, $du = g'(x)$, $dv = (x-\alpha\beta)x^{\alpha-1}e^{-x/\beta}$, $v = -\beta x^\alpha e^{-x/\beta}$. Then

$$\mathrm{E}g(X)(X-\alpha\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left[-g(x)\beta x^\alpha e^{-x/\beta}\Big|_0^\infty + \beta\int_0^\infty g'(x)x^\alpha e^{-x/\beta}dx\right].$$

Assuming $g(x)$ to be differentiable, $\mathrm{E}|Xg'(X)| < \infty$ and $\lim_{x\to\infty} g(x)x^\alpha e^{-x/\beta} = 0$, the first term is zero, and the second term is $\beta\mathrm{E}(Xg'(X))$.

b.

$$\mathrm{E}\left[g(X)\left(\beta-(\alpha-1)\frac{1-X}{x}\right)\right] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 g(x)\left(\beta-(\alpha-1)\frac{1-x}{x}\right)x^{\alpha-1}(1-x)^{\beta-1}dx.$$

Let $u = g(x)$ and $dv = (\beta-(\alpha-1)\frac{1-x}{x})x^{\alpha-1}(1-x)^\beta$. The expectation is

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left[g(x)x^{\alpha-1}(1-x)^\beta\Big|_0^1 + \int_0^1 (1-x)g'(x)x^{\alpha-1}(1-x)^{\beta-1}dx\right] = \mathrm{E}((1-X)g'(X)),$$

assuming the first term is zero and the integral exists.

3.50 The proof is similar to that of part a) of Theorem 3.6.8. For $X \sim$ negative binomial(r, p),

$$
\begin{aligned}
&\mathrm{E}g(X) \\
&= \sum_{x=0}^{\infty} g(x) \binom{r+x-1}{x} p^r (1-p)^x \\
&= \sum_{y=1}^{\infty} g(y-1) \binom{r+y-2}{y-1} p^r (1-p)^{y-1} \qquad \text{(set } y = x+1\text{)} \\
&= \sum_{y=1}^{\infty} g(y-1) \left(\frac{y}{r+y-1}\right) \binom{r+y-1}{y} p^r (1-p)^{y-1} \\
&= \sum_{y=0}^{\infty} \left[\frac{y}{r+y-1} \frac{g(y-1)}{1-p}\right] \left[\binom{r+y-1}{y} p^r (1-p)^y\right] \qquad \text{(the summand is zero at y = 0)} \\
&= \mathrm{E}\left(\frac{X}{r+X-1} \frac{g(X-1)}{1-p}\right),
\end{aligned}
$$

where in the third equality we use the fact that $\binom{r+y-2}{y-1} = \left(\frac{y}{r+y-1}\right)\binom{r+y-1}{y}$.

Chapter 4

Multiple Random Variables

4.1 Since the distribution is uniform, the easiest way to calculate these probabilities is as the ratio of areas, the total area being 4.

a. The circle $x^2 + y^2 \le 1$ has area π, so $P(X^2 + Y^2 \le 1) = \frac{\pi}{4}$.
b. The area below the line $y = 2x$ is half of the area of the square, so $P(2X - Y > 0) = \frac{2}{4}$.
c. Clearly $P(|X + Y| < 2) = 1$.

4.2 These are all fundamental properties of integrals. The proof is the same as for Theorem 2.2.5 with bivariate integrals replacing univariate integrals.

4.3 For the experiment of tossing two fair dice, each of the points in the 36-point sample space are equally likely. So the probability of an event is (number of points in the event)/36. The given probabilities are obtained by noting the following equivalences of events.

$$\begin{aligned} P(\{X=0,Y=0\}) &= P(\{(1,1),(2,1),(1,3),(2,3),(1,5),(2,5)\}) = \frac{6}{36} = \frac{1}{6} \\ P(\{X=0,Y=1\}) &= P(\{(1,2),(2,2),(1,4),(2,4),(1,6),(2,6)\}) = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

$$\begin{aligned} &P(\{X=1,Y=0\}) \\ &= P(\{(3,1),(4,1),(5,1),(6,1),(3,3),(4,3),(5,3),(6,3),(3,5),(4,5),(5,5),(6,5)\}) \\ &= \frac{12}{36} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} &P(\{X=1,Y=1\}) \\ &= P(\{(3,2),(4,2),(5,2),(6,2),(3,4),(4,4),(5,4),(6,4),(3,6),(4,6),(5,6),(6,6)\}) \\ &= \frac{12}{36} = \frac{1}{3} \end{aligned}$$

4.4 a. $\int_0^1 \int_0^2 C(x + 2y)dxdy = 4C = 1$, thus $C = \frac{1}{4}$.

b. $f_X(x) = \begin{cases} \int_0^1 \frac{1}{4}(x + 2y)dy = \frac{1}{4}(x + 1) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$

c. $F_{XY}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(v, u)dvdu$. The way this integral is calculated depends on the values of x and y. For example, for $0 < x < 2$ and $0 < y < 1$,

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)dvdu = \int_0^x \int_0^y \frac{1}{4}(u + 2v)dvdu = \frac{x^2 y}{8} + \frac{y^2 x}{4}.$$

But for $0 < x < 2$ and $1 \le y$,

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)dvdu = \int_0^x \int_0^1 \frac{1}{4}(u + 2v)dvdu = \frac{x^2}{8} + \frac{x}{4}.$$

The complete definition of F_{XY} is

$$F_{XY}(x,y)=\begin{cases}0 & x\le 0 \text{ or } y\le 0\\ x^2y/8+y^2x/4 & 0<x<2 \text{ and } 0<y<1\\ y/2+y^2/2 & 2\le x \text{ and } 0<y<1\\ x^2/8+x/4 & 0<x<2 \text{ and } 1\le y\\ 1 & 2\le x \text{ and } 1\le y\end{cases}.$$

d. The function $z=g(x)=9/(x+1)^2$ is monotone on $0<x<2$, so use Theorem 2.1.5 to obtain $f_Z(z)=9/(8z^2)$, $1<z<9$.

4.5 a. $P(X>\sqrt{Y})=\int_0^1\int_{\sqrt{y}}^1(x+y)dxdy=\frac{7}{20}$.

b. $P(X^2<Y<X)=\int_0^1\int_y^{\sqrt{y}}2xdxdy=\frac{1}{6}$.

4.6 Let $A=$ time that A arrives and $B=$ time that B arrives. The random variables A and B are independent uniform$(1,2)$ variables. So their joint pdf is uniform on the square $(1,2)\times(1,2)$. Let $X=$ amount of time A waits for B. Then, $F_X(x)=P(X\le x)=0$ for $x<0$, and $F_X(x)=P(X\le x)=1$ for $1\le x$. For $x=0$, we have

$$F_X(0)=P(X\le 0)=P(X=0)=P(B\le A)=\int_1^2\int_1^a 1dbda=\frac{1}{2}.$$

And for $0<x<1$,

$$F_X(x)=P(X\le x)=1-P(X>x)=1-P(B-A>x)=1-\int_1^{2-x}\int_{a+x}^2 1dbda=\frac{1}{2}+x-\frac{x^2}{2}.$$

4.7 We will measure time in minutes past 8 A.M. So $X\sim$ uniform$(0,30)$, $Y\sim$ uniform$(40,50)$ and the joint pdf is $1/300$ on the rectangle $(0,30)\times(40,50)$.

$$P(\text{arrive before 9 A.M.})=P(X+Y<60)=\int_{40}^{50}\int_0^{60-y}\frac{1}{300}dxdy=\frac{1}{2}.$$

4.9

$$\begin{aligned}&P(a\le X\le b, c\le Y\le d)\\ &= P(X\le b, c\le Y\le d)-P(X\le a, c\le Y\le d)\\ &= P(X\le b, Y\le d)-P(X\le b, Y\le c)-P(X\le a, Y\le d)+P(X\le a, Y\le c)\\ &= F(b,d)-F(b,c)-F(a,d)-F(a,c)\\ &= F_X(b)F_Y(d)-F_X(b)F_Y(c)-F_X(a)F_Y(d)-F_X(a)F_Y(c)\\ &= P(X\le b)\left[P(Y\le d)-P(Y\le c)\right]-P(X\le a)\left[P(Y\le d)-P(Y\le c)\right]\\ &= P(X\le b)P(c\le Y\le d)-P(X\le a)P(c\le Y\le d)\\ &= P(a\le X\le b)P(c\le Y\le d).\end{aligned}$$

4.10 a. The marginal distribution of X is $P(X=1)=P(X=3)=\frac{1}{4}$ and $P(X=2)=\frac{1}{2}$. The marginal distribution of Y is $P(Y=2)=P(Y=3)=P(Y=4)=\frac{1}{3}$. But

$$P(X=2,Y=3)=0\ne(\frac{1}{2})(\frac{1}{3})=P(X=2)P(Y=3).$$

Therefore the random variables are not independent.

b. The distribution that satisfies $P(U=x,V=y)=P(U=x)P(V=y)$ where $U\sim X$ and $V\sim Y$ is

		U		
		1	2	3
V	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	3	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	4	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$

4.11 The support of the distribution of (U, V) is $\{(u, v) : u = 1, 2, \ldots; v = u+1, u+2, \ldots\}$. This is not a cross-product set. Therefore, U and V are not independent. More simply, if we know $U = u$, then we know $V > u$.

4.12 One interpretation of "a stick is broken at random into three pieces" is this. Suppose the length of the stick is 1. Let X and Y denote the two points where the stick is broken. Let X and Y both have uniform$(0, 1)$ distributions, and assume X and Y are independent. Then the joint distribution of X and Y is uniform on the unit square. In order for the three pieces to form a triangle, the sum of the lengths of any two pieces must be greater than the length of the third. This will be true if and only if the length of each piece is less than $1/2$. To calculate the probability of this, we need to identify the sample points (x, y) such that the length of each piece is less than $1/2$. If $y > x$, this will be true if $x < 1/2$, $y - x < 1/2$ and $1 - y < 1/2$. These three inequalities define the triangle with vertices $(0, 1/2)$, $(1/2, 1/2)$ and $(1/2, 1)$. (Draw a graph of this set.) Because of the uniform distribution, the probability that (X, Y) falls in the triangle is the area of the triangle, which is $1/8$. Similarly, if $x > y$, each piece will have length less than $1/2$ if $y < 1/2$, $x - y < 1/2$ and $1 - x < 1/2$. These three inequalities define the triangle with vertices $(1/2, 0)$, $(1/2, 1/2)$ and $(1, 1/2)$. The probability that (X, Y) is in this triangle is also $1/8$. So the probability that the pieces form a triangle is $1/8 + 1/8 = 1/4$.

4.13 a.

$$\begin{aligned}
&\mathrm{E}(Y - g(X))^2 \\
&= \mathrm{E}\left((Y - \mathrm{E}(Y \mid X)) + (\mathrm{E}(Y \mid X) - g(X))\right)^2 \\
&= \mathrm{E}(Y - \mathrm{E}(Y \mid X))^2 + \mathrm{E}(\mathrm{E}(Y \mid X) - g(X))^2 + 2\mathrm{E}\left[(Y - \mathrm{E}(Y \mid X))(\mathrm{E}(Y \mid X) - g(X))\right].
\end{aligned}$$

The cross term can be shown to be zero by iterating the expectation. Thus

$$\mathrm{E}(Y - g(X))^2 = \mathrm{E}(Y - \mathrm{E}(Y \mid X))^2 + \mathrm{E}(\mathrm{E}(Y \mid X) - g(X))^2 \geq \mathrm{E}(Y - \mathrm{E}(Y \mid X))^2, \text{ for all g}(\cdot).$$

The choice $g(X) = \mathrm{E}(Y \mid X)$ will give equality.

b. Equation (2.2.3) is the special case of a) where we take the random variable X to be a constant. Then, $g(X)$ is a constant, say b, and $\mathrm{E}(Y \mid X) = \mathrm{E}Y$.

4.15 We will find the conditional distribution of $Y|X + Y$. The derivation of the conditional distribution of $X|X + Y$ is similar. Let $U = X + Y$ and $V = Y$. In Example 4.3.1, we found the joint pmf of (U, V). Note that for fixed u, $f(u, v)$ is positive for $v = 0, \ldots, u$. Therefore the conditional pmf is

$$f(v|u) = \frac{f(u, v)}{f(u)} = \frac{\frac{\theta^{u-v}e^{-\theta}}{(u-v)!}\frac{\lambda^v e^{-\lambda}}{v!}}{\frac{(\theta+\lambda)^u e^{-(\theta+\lambda)}}{u!}} = \binom{u}{v}\left(\frac{\lambda}{\theta+\lambda}\right)^v\left(\frac{\theta}{\theta+\lambda}\right)^{u-v}, \quad v = 0, \ldots, u.$$

That is $V|U \sim$ binomial$(U, \lambda/(\theta + \lambda))$.

4.16 a. The support of the distribution of (U, V) is $\{(u, v) : u = 1, 2, \ldots; v = 0, \pm 1, \pm 2, \ldots\}$. If $V > 0$, then $X > Y$. So for $v = 1, 2, \ldots$, the joint pmf is

$$\begin{aligned}
f_{U,V}(u, v) &= P(U = u, V = v) = P(Y = u, X = u + v) \\
&= p(1-p)^{u+v-1}p(1-p)^{u-1} = p^2(1-p)^{2u+v-2}.
\end{aligned}$$

If $V<0$, then $X<Y$. So for $v=-1,-2,\ldots$, the joint pmf is

$$\begin{aligned} f_{U,V}(u,v) &= P(U=u,V=v) &= P(X=u,Y=u-v) \\ &= p(1-p)^{u-1}p(1-p)^{u-v-1} &= p^2(1-p)^{2u-v-2}. \end{aligned}$$

If $V=0$, then $X=Y$. So for $v=0$, the joint pmf is

$$f_{U,V}(u,0)=P(U=u,V=0)=P(X=Y=u)=p(1-p)^{u-1}p(1-p)^{u-1}=p^2(1-p)^{2u-2}.$$

In all three cases, we can write the joint pmf as

$$f_{U,V}(u,v)=p^2(1-p)^{2u+|v|-2}=\left(p^2(1-p)^{2u}\right)(1-p)^{|v|-2},\ u=1,2,\ldots;v=0,\pm1,\pm2,\ldots.$$

Since the joint pmf factors into a function of u and a function of v, U and V are independent.

b. The possible values of Z are all the fractions of the form r/s, where r and s are positive integers and $r<s$. Consider one such value, r/s, where the fraction is in reduced form. That is, r and s have no common factors. We need to identify all the pairs (x,y) such that x and y are positive integers and $x/(x+y)=r/s$. All such pairs are $(ir,i(s-r))$, $i=1,2,\ldots$. Therefore,

$$\begin{aligned} P\left(Z=\frac{r}{s}\right) &= \sum_{i=1}^{\infty}P(X=ir,Y=i(s-r)) &= \sum_{i=1}^{\infty}p(1-p)^{ir-1}p(1-p)^{i(s-r)-1} \\ &= \frac{p^2}{(1-p)^2}\sum_{i=1}^{\infty}((1-p)^s)^i &= \frac{p^2}{(1-p)^2}\frac{(1-p)^s}{1-(1-p)^s} = \frac{p^2(1-p)^{s-2}}{1-(1-p)^s}. \end{aligned}$$

c.

$$P(X=x,X+Y=t)=P(X=x,Y=t-x)=P(X=x)P(Y=t-x)=p^2(1-p)^{t-2}.$$

4.17 a. $P(Y=i+1)=\int_i^{i+1}e^{-x}dx=e^{-i}(1-e^{-1})$, which is geometric with $p=1-e^{-1}$.

b. Since $Y\geq 5$ if and only if $X\geq 4$,

$$P(X-4\leq x|Y\geq 5)=P(X-4\leq x|X\geq 4)=P(X\leq x)=e^{-x},$$

since the exponential distribution is memoryless.

4.18 We need to show $f(x,y)$ is nonnegative and integrates to 1. $f(x,y)\geq 0$, because the numerator is nonnegative since $g(x)\geq 0$, and the denominator is positive for all $x>0$, $y>0$. Changing to polar coordinates, $x=r\cos\theta$ and $y=r\sin\theta$, we obtain

$$\int_0^{\infty}\int_0^{\infty}f(x,y)dxdy=\int_0^{\pi/2}\int_0^{\infty}\frac{2g(r)}{\pi r}rdrd\theta=\frac{2}{\pi}\int_0^{\pi/2}\int_0^{\infty}g(r)drd\theta=\frac{2}{\pi}\int_0^{\pi/2}1d\theta=1.$$

4.19 a. Since $(X_1-X_2)\big/\sqrt{2}\sim \mathrm{n}(0,1)$, $(X_1-X_2)^2\big/2\sim\chi^2_1$ (see Example 2.1.9).

b. Make the transformation $y_1=\frac{x_1}{x_1+x_2}$, $y_2=x_1+x_2$ then $x_1=y_1y_2$, $x_2=y_2(1-y_1)$ and $|J|=y_2$. Then

$$f(y_1,y_2)=\left[\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}\right]\left[\frac{1}{\Gamma(\alpha_1+\alpha_2)}y_2^{\alpha_1+\alpha_2-1}e^{-y_2}\right],$$

thus $Y_1\sim\mathrm{beta}(\alpha_1,\alpha_2)$, $Y_2\sim\mathrm{gamma}(\alpha_1+\alpha_1,1)$ and are independent.

4.20 a. This transformation is not one-to-one because you cannot determine the sign of X_2 from Y_1 and Y_2. So partition the support of (X_1, X_2) into $\mathcal{A}_0 = \{-\infty < x_1 < \infty, x_2 = 0\}$, $\mathcal{A}_1 = \{-\infty < x_1 < \infty, x_2 > 0\}$ and $\mathcal{A}_2 = \{-\infty < x_1 < \infty, x_2 < 0\}$. The support of (Y_1, Y_2) is $\mathcal{B} = \{0 < y_1 < \infty, -1 < y_2 < 1\}$. The inverse transformation from $\mathcal{B}$ to $\mathcal{A}_1$ is $x_1 = y_2\sqrt{y_1}$ and $x_2 = \sqrt{y_1 - y_1 y_2^2}$ with Jacobian

$$J_1 = \begin{vmatrix} \frac{1}{2}\frac{y_2}{\sqrt{y_1}} & \sqrt{y_1} \\ \frac{1}{2}\frac{\sqrt{1-y_2^2}}{\sqrt{y_1}} & \frac{y_2\sqrt{y_1}}{\sqrt{1-y_2^2}} \end{vmatrix} = \frac{1}{2\sqrt{1-y_2^2}}.$$

The inverse transformation from $\mathcal{B}$ to $\mathcal{A}_2$ is $x_1 = y_2\sqrt{y_1}$ and $x_2 = -\sqrt{y_1 - y_1 y_2^2}$ with $J_2 = -J_1$. From (4.3.6), $f_{Y_1,Y_2}(y_1, y_2)$ is the sum of two terms, both of which are the same in this case. Then

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= 2\left[\frac{1}{2\pi\sigma^2}e^{-y_1/(2\sigma^2)}\frac{1}{2\sqrt{1-y_2^2}}\right] \\ &= \frac{1}{2\pi\sigma^2}e^{-y_1/(2\sigma^2)}\frac{1}{\sqrt{1-y_2^2}}, \qquad 0 < y_1 < \infty, -1 < y_2 < 1. \end{aligned}$$

b. We see in the above expression that the joint pdf factors into a function of y_1 and a function of y_2. So Y_1 and Y_2 are independent. Y_1 is the square of the distance from (X_1, X_2) to the origin. Y_2 is the cosine of the angle between the positive x_1-axis and the line from (X_1, X_2) to the origin. So independence says the distance from the origin is independent of the orientation (as measured by the angle).

4.21 Since R and θ are independent, the joint pdf of $T = R^2$ and θ is

$$f_{T,\theta}(t, \theta) = \frac{1}{4\pi}e^{-t/2}, \quad 0 < t < \infty, \quad 0 < \theta < 2\pi.$$

Make the transformation $x = \sqrt{t}\cos\theta$, $y = \sqrt{t}\sin\theta$. Then $t = x^2 + y^2$, $\theta = \tan^{-1}(y/x)$, and

$$J = \begin{vmatrix} 2x & 2y \\ \frac{-y}{x^2+y^2} & \frac{-x}{x^2+y^2} \end{vmatrix} = 2.$$

Therefore

$$f_{X,Y}(x, y) = \frac{2}{4\pi}e^{-\frac{1}{2}(x^2+y^2)}, \quad 0 < x^2 + y^2 < \infty, \quad 0 < \tan^{-1} y/x < 2\pi.$$

Thus,

$$f_{X,Y}(x, y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}, \quad -\infty < x, y < \infty.$$

So X and Y are independent standard normals.

4.23 a. Let $y = v$, $x = u/y = u/v$ then

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}.$$

$$f_{U,V}(u, v) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}\left(\frac{u}{v}\right)^{\alpha-1}\left(1-\frac{u}{v}\right)^{\beta-1}v^{\alpha+\beta-1}(1-v)^{\gamma-1}\frac{1}{v}, 0 < u < v < 1.$$

Then,

$$\begin{aligned}
f_U(u) &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}\int_u^1 v^{\beta-1}(1-v)^{\gamma-1}(\frac{v-u}{v})^{\beta-1}dv \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}\int_0^1 y^{\beta-1}(1-y)^{\gamma-1}dy \left(y=\frac{v-u}{1-u}, dy=\frac{dv}{1-u}\right) \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}\frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)} \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}, \qquad 0<u<1.
\end{aligned}$$

Thus, $U \sim \text{gamma}(\alpha, \beta+\gamma)$.

b. Let $x=\sqrt{uv}$, $y=\sqrt{\frac{u}{v}}$ then

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial x}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2}v^{1/2}u^{-1/2} & \frac{1}{2}u^{1/2}v^{-1/2} \\ \frac{1}{2}v^{-1/2}u^{-1/2} & -\frac{1}{2}u^{1/2}v^{-3/2} \end{vmatrix} = \frac{1}{2v}.$$

$$f_{U,V}(u,v) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}(\sqrt{uv}^{\alpha-1}(1-\sqrt{uv})^{\beta-1}\left(\sqrt{\frac{u}{v}}\right)^{\alpha+\beta-1}\left(1-\sqrt{\frac{u}{v}}\right)^{\gamma-1}\frac{1}{2v}.$$

The set $\{0<x<1, 0<y<1\}$ is mapped onto the set $\{0<u<v<\frac{1}{u}, 0<u<1\}$. Then,

$$\begin{aligned}
&f_U(u) \\
&= \int_u^{1/u} f_{U,V}(u,v)dv \\
&= \underbrace{\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}}_{\text{Call it A}}\int_u^{1/u}\left(\frac{1-\sqrt{uv}}{1-u}\right)^{\beta-1}\left(\frac{1-\sqrt{u/v}}{1-u}\right)^{\gamma-1}\frac{(\sqrt{u/v})^{\beta}}{2v(1-u)}dv.
\end{aligned}$$

To simplify, let $z=\frac{\sqrt{u/v}-u}{1-u}$. Then $v=u \Rightarrow z=1$, $v=1/u \Rightarrow z=0$ and $dz=-\frac{\sqrt{u/v}}{2(1-u)v}dv$. Thus,

$$\begin{aligned}
f_U(u) &= A\int z^{\beta-1}(1-z)^{\gamma-1}dz \qquad (\text{ kernel of beta}(\beta,\gamma)) \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}\frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)} \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)}u^{\alpha-1}(1-u)^{\beta+\gamma-1}, \qquad 0<u<1.
\end{aligned}$$

That is, $U \sim \text{beta}(\alpha, \beta+\gamma)$, as in a).

4.24 Let $z_1 = x+y$, $z_2=\frac{x}{x+y}$, then $x=z_1z_2$, $y=z_1(1-z_2)$ and

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial z_1} & \frac{\partial x}{\partial z_2} \\ \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2} \end{vmatrix} = \begin{vmatrix} z_2 & z_1 \\ 1-z_2 & -z_1 \end{vmatrix} = z_1.$$

The set $\{x>0, y>0\}$ is mapped onto the set $\{z_1>0, 0<z_2<1\}$.

$$\begin{aligned}
f_{Z_1,Z_2}(z_1,z_2) &= \frac{1}{\Gamma(r)}(z_1z_2)^{r-1}e^{-z_1z_2}\cdot\frac{1}{\Gamma(s)}(z_1-z_1z_2)^{s-1}e^{-z_1+z_1z_2}z_1 \\
&= \frac{1}{\Gamma(r+s)}z_1^{r+s-1}e^{-z_1}\cdot\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}z_2^{r-1}(1-z_2)^{s-1}, \qquad 0<z_1, 0<z_2<1.
\end{aligned}$$

$f_{Z_1,Z_2}(z_1,z_2)$ can be factored into two densities. Therefore Z_1 and Z_2 are independent and $Z_1 \sim \text{gamma}(r+s,1)$, $Z_2 \sim \text{beta}(r,s)$.

4.25 For X and Z independent, and $Y = X+Z$, $f_{XY}(x,y) = f_X(x)f_Z(y-x)$. In Example 4.5.8,

$$f_{XY}(x,y) = I_{(0,1)}(x)\frac{1}{10}I_{(0,1/10)}(y-x).$$

In Example 4.5.9, $Y = X^2 + Z$ and

$$f_{XY}(x,y) = f_X(x)f_Z(y-x^2) = \frac{1}{2}I_{(-1,1)}(x)\frac{1}{10}I_{(0,1/10)}(y-x^2).$$

4.26 a.

$$\begin{aligned} P(Z\le z, W=0) &= P(\min(X,Y)\le z, Y\le X) \quad = \quad P(Y\le z, Y\le X) \\ &= \int_0^z\int_y^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dxdy \\ &= \frac{\lambda}{\mu+\lambda}\left(1-\exp\left\{-\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)z\right\}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z\le z, W=1) &= P(\min(X,Y)\le z, X\le Y) \quad = \quad P(X\le z, X\le Y) \\ &= \int_0^z\int_x^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dydx \quad = \quad \frac{\mu}{\mu+\lambda}\left(1-\exp\left\{-\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)z\right\}\right). \end{aligned}$$

b.

$$P(W=0) = P(Y\le X) = \int_0^\infty\int_y^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dxdy = \frac{\lambda}{\mu+\lambda}.$$

$$P(W=1) = 1-P(W=0) = \frac{\mu}{\mu+\lambda}.$$

$$P(Z\le z) = P(Z\le z, W=0) + P(Z\le z, W=1) = 1-\exp\left\{-\left(\frac{1}{\mu}+\frac{1}{\lambda}\right)z\right\}.$$

Therefore, $P(Z\le z, W=i) = P(Z\le z)P(W=i)$, for $i=0,1$, $z>0$. So Z and W are independent.

4.27 From Theorem 4.2.14 we know $U \sim \text{n}(\mu+\gamma, 2\sigma^2)$ and $V \sim \text{n}(\mu-\gamma, 2\sigma^2)$. It remains to show that they are independent. Proceed as in Exercise 4.24.

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left[(x-\mu)^2+(y-\gamma)^2\right]} \qquad \text{(by independence, so} f_{XY} = f_X f_Y\text{)}$$

Let $u = x+y$, $v = x-y$, then $x = \frac{1}{2}(u+v)$, $y = \frac{1}{2}(u-v)$ and

$$|J| = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = \frac{1}{2}.$$

The set $\{-\infty < x < \infty, -\infty < y < \infty\}$ is mapped onto the set $\{-\infty < u < \infty, -\infty < v < \infty\}$. Therefore

$$\begin{aligned} f_{UV}(u,v) &= \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left[\left(\left(\frac{u+v}{2}\right)-\mu\right)^2+\left(\left(\frac{u-v}{2}\right)-\gamma\right)^2\right]}\cdot\frac{1}{2} \\ &= \frac{1}{4\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left[2\left(\frac{u}{2}\right)^2-u(\mu+\gamma)+\frac{(\mu+\gamma)^2}{2}+2\left(\frac{v}{2}\right)^2-v(\mu-\gamma)+\frac{(\mu+\gamma)^2}{2}\right]} \\ &= g(u)\frac{1}{4\pi\sigma^2}e^{-\frac{1}{2(2\sigma^2)}\left(u-(\mu+\gamma)\right)^2}\cdot h(v)e^{-\frac{1}{2(2\sigma^2)}\left(v-(\mu-\gamma)\right)^2}. \end{aligned}$$

By the factorization theorem, U and V are independent.

4.29 a. $\frac{X}{Y} = \frac{R\cos\theta}{R\sin\theta} = \cot\theta$. Let $Z = \cot\theta$. Let $A_1 = (0,\pi)$, $g_1(\theta) = \cot\theta$, $g_1^{-1}(z) = \cot^{-1} z$, $A_2 = (\pi, 2\pi), g_2(\theta) = \cot\theta$, $g_2^{-1}(z) = \pi + \cot^{-1} z$. By Theorem 2.1.8

$$f_Z(z) = \frac{1}{2\pi}\left|\frac{-1}{1+z^2}\right| + \frac{1}{2\pi}\left|\frac{-1}{1+z^2}\right| = \frac{1}{\pi}\frac{1}{1+z^2}, \quad -\infty < z < \infty.$$

b. $XY = R^2\cos\theta\sin\theta$ then $2XY = R^2 2\cos\theta\sin\theta = R^2\sin 2\theta$. Therefore $\frac{2XY}{R} = R\sin 2\theta$. Since $R = \sqrt{X^2+Y^2}$ then $\frac{2XY}{\sqrt{X^2+Y^2}} = R\sin 2\theta$. Thus $\frac{2XY}{\sqrt{X^2+Y^2}}$ is distributed as $\sin 2\theta$ which is distributed as $\sin\theta$. To see this let $\sin\theta \sim f_{\sin\theta}$. For the function $\sin 2\theta$ the values of the function $\sin\theta$ are repeated over each of the 2 intervals $(0,\pi)$ and $(\pi, 2\pi)$. Therefore the distribution in each of these intervals is the distribution of $\sin\theta$. The probability of choosing between each one of these intervals is $\frac{1}{2}$. Thus $f_{2\sin\theta} = \frac{1}{2}f_{\sin\theta} + \frac{1}{2}f_{\sin\theta} = f_{\sin\theta}$. Therefore $\frac{2XY}{\sqrt{X^2+Y^2}}$ has the same distribution as $Y = \sin\theta$. In addition, $\frac{2XY}{\sqrt{X^2+Y^2}}$ has the same distribution as $X = \cos\theta$ since $\sin\theta$ has the same distribution as $\cos\theta$. To see this let consider the distribution of $W = \cos\theta$ and $V = \sin\theta$ where $\theta \sim$ uniform$(0, 2\pi)$. To derive the distribution of $W = \cos\theta$ let $A_1 = (0,\pi)$, $g_1(\theta) = \cos\theta$, $g_1^{-1}(w) = \cos^{-1} w$, $A_2 = (\pi, 2\pi)$, $g_2(\theta) = \cos\theta$, $g_2^{-1}(w) = 2\pi - \cos^{-1} w$. By Theorem 2.1.8

$$f_W(w) = \frac{1}{2\pi}\left|\frac{-1}{\sqrt{1-w^2}}\right| + \frac{1}{2\pi}\left|\frac{1}{\sqrt{1-w^2}}\right| = \frac{1}{\pi}\frac{1}{\sqrt{1-w^2}}, -1 \le w \le 1.$$

To derive the distribution of $V = \sin\theta$, first consider the interval $(\frac{\pi}{2}, \frac{3\pi}{2})$. Let $g_1(\theta) = \sin\theta$, $4g_1^{-1}(v) = \pi - \sin^{-1} v$, then

$$f_V(v) = \frac{1}{\pi}\frac{1}{\sqrt{1-v^2}}, \quad -1 \le v \le 1.$$

Second, consider the set $\{(0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)\}$, for which the function $\sin\theta$ has the same values as it does in the interval $(\frac{-\pi}{2}, \frac{\pi}{2})$. Therefore the distribution of V in $\{(0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)\}$ is the same as the distribution of V in $(\frac{-\pi}{2}, \frac{\pi}{2})$ which is $\frac{1}{\pi}\frac{1}{\sqrt{1-v^2}}$, $-1 \le v \le 1$. On $(0, 2\pi)$ each of the sets $(\frac{\pi}{2}, \frac{3\pi}{2})$, $\{(0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)\}$ has probability $\frac{1}{2}$ of being chosen. Therefore

$$f_V(v) = \frac{1}{2}\frac{1}{\pi}\frac{1}{\sqrt{1-v^2}} + \frac{1}{2}\frac{1}{\pi}\frac{1}{\sqrt{1-v^2}} = \frac{1}{\pi}\frac{1}{\sqrt{1-v^2}}, \quad -1 \le v \le 1.$$

Thus W and V has the same distribution.
Let X and Y be iid n$(0,1)$. Then $X^2 + Y^2 \sim \chi^2_2$ is a positive random variable. Therefore with $X = R\cos\theta$ and $Y = R\sin\theta$, $R = \sqrt{X^2+Y^2}$ is a positive random variable and $\theta = \tan^{-1}(\frac{Y}{X}) \sim$ uniform$(0,1)$. Thus $\frac{2XY}{\sqrt{X^2+Y^2}} \sim X \sim$ n$(0,1)$.

4.30 a.

$$\begin{aligned}
\mathrm{E}Y &= \mathrm{E}\left\{\mathrm{E}(Y|X)\right\} = \mathrm{E}X = \frac{1}{2}. \\
\mathrm{Var}Y &= \mathrm{Var}\left(\mathrm{E}(Y|X)\right) + \mathrm{E}\left(\mathrm{Var}(Y|X)\right) = \mathrm{Var}X + \mathrm{E}X^2 = \frac{1}{12} + \frac{1}{3} = \frac{5}{12}. \\
\mathrm{E}XY &= \mathrm{E}[\mathrm{E}(XY|X)] = \mathrm{E}[X\mathrm{E}(Y|X)] = \mathrm{E}X^2 = \frac{1}{3} \\
\mathrm{Cov}(X,Y) &= \mathrm{E}XY - \mathrm{E}X\mathrm{E}Y = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.
\end{aligned}$$

b. The quick proof is to note that the distribution of $Y|X = x$ is n$(1,1)$, hence is independent of X. The bivariate transformation $t = y/x$, $u = x$ will also show that the joint density factors.

4.31 a.

$$\mathrm{E}Y = \mathrm{E}\{\mathrm{E}(Y|X)\} = \mathrm{E}nX = \frac{n}{2}.$$

$$\mathrm{Var}Y = \mathrm{Var}\left(\mathrm{E}(Y|X)\right) + \mathrm{E}\left(\mathrm{Var}(Y|X)\right) = \mathrm{Var}(nX) + \mathrm{E}nX(1-X) = \frac{n^2}{12} + \frac{n}{6}.$$

b.

$$P(Y = y, X \le x) = \binom{n}{y} x^y (1-x)^{n-y}, \quad y = 0, 1, \ldots, n, \quad 0 < x < 1.$$

c.

$$P(y = y) = \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}.$$

4.32 a. The pmf of Y, for $y = 0, 1, \ldots$, is

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_Y(y|\lambda) f_\Lambda(\lambda) d\lambda = \int_0^\infty \frac{\lambda^y e^{-\lambda}}{y!} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} d\lambda \\ &= \frac{1}{y!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{(y+\alpha)-1} \exp\left\{ \frac{-\lambda}{\left(\frac{\beta}{1+\beta}\right)} \right\} d\lambda \\ &= \frac{1}{y!\Gamma(\alpha)\beta^\alpha} \Gamma(y+\alpha) \left(\frac{\beta}{1+\beta}\right)^{y+\alpha}. \end{aligned}$$

If α is a positive integer,

$$f_Y(y) = \binom{y+\alpha-1}{y} \left(\frac{\beta}{1+\beta}\right)^y \left(\frac{1}{1+\beta}\right)^\alpha,$$

the negative binomial$(\alpha, 1/(1+\beta))$ pmf. Then

$$\begin{aligned} \mathrm{E}Y &= \mathrm{E}(\mathrm{E}(Y|\Lambda)) = \mathrm{E}\Lambda = \alpha\beta \\ \mathrm{Var}Y &= \mathrm{Var}(\mathrm{E}(Y|\Lambda)) + \mathrm{E}(\mathrm{Var}(Y|\Lambda)) = \mathrm{Var}\Lambda + \mathrm{E}\Lambda = \alpha\beta^2 + \alpha\beta = \alpha\beta(\beta+1). \end{aligned}$$

b. For $y = 0, 1, \ldots$, we have

$$\begin{aligned} P(Y = y|\lambda) &= \sum_{n=y}^\infty P(Y = y|N = n, \lambda) P(N = n|\lambda) \\ &= \sum_{n=y}^\infty \binom{n}{y} p^y (1-p)^{n-y} \frac{e^{-\lambda}\lambda^n}{n!} \\ &= \sum_{n=y}^\infty \frac{1}{y!(n-y)!} \left(\frac{p}{1-p}\right)^y [(1-p)\lambda]^n e^{-\lambda} \\ &= e^{-\lambda} \sum_{m=0}^\infty \frac{1}{y!m!} \left(\frac{p}{1-p}\right)^y [(1-p)\lambda]^{m+y} \qquad \text{(let } m = n-y\text{)} \\ &= \frac{e^{-\lambda}}{y!} \left(\frac{p}{1-p}\right)^y [(1-p)\lambda]^y \left[\sum_{m=0}^\infty \frac{[(1-p)\lambda]^m}{m!} \right] \\ &= e^{-\lambda} (p\lambda)^y e^{(1-p)\lambda} \\ &= \frac{(p\lambda)^y e^{-p\lambda}}{y!}, \end{aligned}$$

the Poisson($p\lambda$) pmf. Thus $Y|\Lambda \sim$ Poisson($p\lambda$). Now calculations like those in a) yield the pmf of Y, for $y = 0, 1, \ldots$, is

$$f_Y(y) = \frac{1}{\Gamma(\alpha)y!(p\beta)^\alpha}\Gamma(y+\alpha)\left(\frac{p\beta}{1+p\beta}\right)^{y+\alpha}.$$

Again, if α is a positive integer, $Y \sim$ negative binomial($\alpha, 1/(1+p\beta)$).

4.33 We can show that H has a negative binomial distribution by computing the mgf of H.

$$\mathrm{E}e^{Ht} = \mathrm{EE}\left(e^{Ht}\middle| N\right) = \mathrm{EE}\left(e^{(X_1+\cdots+X_N)t}\middle| N\right) = \mathrm{E}\left\{\left[\mathrm{E}\left(e^{X_1t}\middle| N\right)\right]^N\right\},$$

because, by Theorem 4.6.7, the mgf of a sum of independent random variables is equal to the product of the individual mgfs. Now,

$$\mathrm{E}e^{X_1t} = \sum_{x_1=1}^{\infty} e^{x_1t}\frac{-1}{\log p}\frac{(1-p)^{x_1}}{x_1} = \frac{-1}{\log p}\sum_{x_1=1}^{\infty}\frac{(e^t(1-p))^{x_1}}{x_1} = \frac{-1}{\log p}\left(-\log\left\{1-e^t(1-p)\right\}\right).$$

Then

$$\begin{aligned}
\mathrm{E}\left(\frac{\log\{1-e^t(1-p)\}}{\log p}\right)^N &= \sum_{n=0}^{\infty}\left(\frac{\log\{1-e^t(1-p)\}}{\log p}\right)^n\frac{e^{-\lambda}\lambda^n}{n!} \qquad \text{(since } N \sim \text{Poisson)}\\
&= e^{-\lambda}e^{\frac{\lambda\log(1-e^t(1-p))}{\log p}}\sum_{n=0}^{\infty}\frac{e^{\frac{-\lambda\log(1-e^t(1-p))}{\log p}}\left(\frac{\lambda\log(1-e^t(1-p))}{\log p}\right)^n}{n!}.
\end{aligned}$$

The sum equals 1. It is the sum of a Poisson$\Big([\lambda\log(1-e^t(1-p))]/[\log p]\Big)$ pmf. Therefore,

$$\begin{aligned}
\mathrm{E}(e^{Ht}) &= e^{-\lambda}\left[e^{\log(1-e^t(1-p))}\right]^{\lambda/\log p} = \left(e^{\log p}\right)^{-\lambda/\log p}\left(\frac{1}{1-e^t(1-p)}\right)^{-\lambda/\log p}\\
&= \left(\frac{p}{1-e^t(1-p)}\right)^{-\lambda/\log p}.
\end{aligned}$$

This is the mgf of a negative binomial(r, p), with $r = -\lambda/\log p$, if r is an integer.

4.34 a.

$$\begin{aligned}
P(Y=y) &= \int_0^1 P(Y=y|p)f_p(p)dp\\
&= \int_0^1 \binom{n}{y}p^y(1-p)^{n-y}\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp\\
&= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 p^{y+\alpha-1}(1-p)^{n+\beta-y-1}dp\\
&= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(y+\alpha)\Gamma(n+\beta-y)}{\Gamma(\alpha+n+\beta)}, \qquad y = 0, 1, \ldots, n.
\end{aligned}$$

b.

$$\begin{aligned}
P(X=x) &= \int_0^1 P(X=x|p)f_P(p)dp\\
&= \int_0^1 \binom{r+x-1}{x}p^r(1-p)^x\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp
\end{aligned}$$

$$\begin{aligned} &= \binom{r+x-1}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 p^{(r+\alpha)-1}(1-p)^{(x+\beta)-1}dp \\ &= \binom{r+x-1}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(r+\alpha)\Gamma(x+\beta)}{\Gamma(r+x+\alpha+\beta)} \qquad x=0,1,\ldots \end{aligned}$$

Therefore,

$$\mathrm{E}X = \mathrm{E}[\mathrm{E}(X|P)] = \mathrm{E}\left[\frac{r(1-P)}{P}\right] = \frac{r\beta}{\alpha-1},$$

since

$$\begin{aligned} \mathrm{E}\left[\frac{1-P}{P}\right] &= \int_0^1 \left(\frac{1-P}{P}\right)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 p^{(\alpha-1)-1}(1-p)^{(\beta+1)-1}dp \quad = \quad \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} \\ &= \frac{\beta}{\alpha-1}. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}(\mathrm{Var}(X|P)) + \mathrm{Var}(\mathrm{E}(X|P)) \quad = \quad \mathrm{E}\left[\frac{r(1-P)}{P^2}\right] + \mathrm{Var}\left(\frac{r(1-P)}{P}\right) \\ &= r\frac{(\beta+1)(\alpha+\beta)}{\alpha(\alpha-1)} + r^2\frac{\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}, \end{aligned}$$

since

$$\begin{aligned} \mathrm{E}\left[\frac{1-P}{P^2}\right] &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{(\alpha-2)-1}(1-p)^{(\beta+1)-1}dp \quad = \quad \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha-2)\Gamma(\beta+1)}{\Gamma(\alpha+\beta-1)} \\ &= \frac{(\beta+1)(\alpha+\beta)}{\alpha(\alpha-1)} \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}\left(\frac{1-P}{P}\right) &= \mathrm{E}\left[\left(\frac{1-P}{P}\right)^2\right] - \left(\mathrm{E}\left[\frac{1-P}{P}\right]\right)^2 \quad = \quad \frac{\beta(\beta+1)}{(\alpha-2)(\alpha-1)} - (\frac{\beta}{\alpha-1})^2 \\ &= \frac{\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}, \end{aligned}$$

where

$$\begin{aligned} \mathrm{E}\left[\left(\frac{1-P}{P}\right)^2\right] &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{(\alpha-2)-1}(1-p)^{(\beta+2)-1}dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha-2)\Gamma(\beta+2)}{\Gamma(\alpha-2+\beta+2)} \quad = \quad \frac{\beta(\beta+1)}{(\alpha-2)(\alpha-1)}. \end{aligned}$$

4.35 a. $\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X|P)) + \mathrm{Var}(\mathrm{E}(X|P))$. Therefore,

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}[nP(1-P)] + \mathrm{Var}(nP) \\ &= n\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + n^2\mathrm{Var}P \\ &= n\frac{\alpha\beta(\alpha+\beta+1-1)}{(\alpha+\beta^2)(\alpha+\beta+1)} + n^2\mathrm{Var}P \end{aligned}$$

$$\begin{aligned}
&= \frac{n\alpha\beta(\alpha+\beta+1)}{(\alpha+\beta^2)(\alpha+\beta+1)} - \frac{n\alpha\beta}{(\alpha+\beta^2)(\alpha+\beta+1)} + n^2\mathrm{Var}P \\
&= n\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta} - n\mathrm{Var}P + n^2\mathrm{Var}P \\
&= n\mathrm{E}P(1-\mathrm{E}P) + n(n-1)\mathrm{Var}P.
\end{aligned}$$

b. $\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y|\Lambda)) + \mathrm{Var}(\mathrm{E}(Y|\Lambda)) = \mathrm{E}\Lambda + \mathrm{Var}(\Lambda) = \mu + \frac{1}{\alpha}\mu^2$ since $\mathrm{E}\Lambda = \mu = \alpha\beta$ and $\mathrm{Var}(\Lambda) = \alpha\beta^2 = \frac{(\alpha\beta)^2}{\alpha} = \frac{\mu^2}{\alpha}$. The "extra-Poisson" variation is $\frac{1}{\alpha}\mu^2$.

4.37 a. Let $Y = \sum X_i$.

$$\begin{aligned}
P(Y=k) &= P(Y=k, \frac{1}{2} < c = \frac{1}{2}(1+p) < 1) \\
&= \int_0^1 (Y=k|c=\frac{1}{2}(1+p))P(P=p)dp \\
&= \int_0^1 \binom{n}{k}[\frac{1}{2}(1+p)]^k[1-\frac{1}{2}(1+p)]^{n-k}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}dp \\
&= \int_0^1 \binom{n}{k}\frac{(1+p)^k}{2^k}\frac{(1-p)^{n-k}}{2^{n-k}}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}dp \\
&= \binom{n}{k}\frac{\Gamma(a+b)}{2^n\Gamma(a)\Gamma(b)}\sum_{j=0}^{k}\int_0^1 p^{k+a-1}(1-p)^{n-k+b-1}dp \\
&= \binom{n}{k}\frac{\Gamma(a+b)}{2^n\Gamma(a)\Gamma(b)}\sum_{j=0}^{k}\binom{k}{j}\frac{\Gamma(k+a)\Gamma(n-k+b)}{\Gamma(n+a+b)} \\
&= \sum_{j=0}^{k}\left[\left(\frac{\binom{k}{j}}{2^n}\right)\left(\binom{n}{k}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(k+a)\Gamma(n-k+b)}{\Gamma(n+a+b)}\right)\right].
\end{aligned}$$

A mixture of beta-binomial.

b.

$$\mathrm{E}Y = \mathrm{E}(\mathrm{E}(Y|c)) = \mathrm{E}[nc] = \mathrm{E}\left[n\left(\frac{1}{2}(1+p)\right)\right] = \frac{n}{2}\left(1+\frac{a}{a+b}\right).$$

Using the results in Exercise 4.35(a),

$$\mathrm{Var}(Y) = n\mathrm{E}C(1-\mathrm{E}C) + n(n-1)\mathrm{Var}C.$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(Y) &= n\mathrm{E}\left[\frac{1}{2}(1+P)\right]\left(1-\mathrm{E}\left[\frac{1}{2}(1+P)\right]\right) + n(n-1)\mathrm{Var}\left(\frac{1}{2}(1+P)\right) \\
&= \frac{n}{4}(1+\mathrm{E}P)(1-\mathrm{E}P) + \frac{n(n-1)}{4}\mathrm{Var}P \\
&= \frac{n}{4}\left(1-\left(\frac{a}{a+b}\right)^2\right) + \frac{n(n-1)}{4}\frac{ab}{(a+b)^2(a+b+1)}.
\end{aligned}$$

4.38 a. Make the transformation $u = \frac{x}{\nu} - \frac{x}{\lambda}$, $du = \frac{-x}{\nu^2}d\nu$, $\frac{\nu}{\lambda-\nu} = \frac{x}{\lambda u}$. Then

$$\int_0^\lambda \frac{1}{\nu}e^{-x/\nu}\frac{1}{\Gamma(r)\Gamma(1-r)}\frac{\nu^{r-1}}{(\lambda-\nu)^r}d\nu$$

$$= \frac{1}{\Gamma(r)\Gamma(1-r)} \int_0^\infty \frac{1}{x} \left(\frac{x}{\lambda u}\right)^r e^{-(u+x/\lambda)} du$$
$$= \frac{x^{r-1}e^{-x/\lambda}}{\lambda^r \Gamma(r)\Gamma(1-r)} \int_0^\infty \left(\frac{1}{u}\right)^r e^{-u} du = \frac{x^{r-1}e^{-x/\lambda}}{\Gamma(r)\lambda^r},$$

since the integral is equal to $\Gamma(1-r)$ if $r < 1$.

b. Use the transformation $t = \nu/\lambda$ to get

$$\int_0^\lambda p_\lambda(\nu)d\nu = \frac{1}{\Gamma(r)\Gamma(1-r)} \int_0^\lambda \nu^{r-1}(\lambda-\nu)^{-r} d\nu = \frac{1}{\Gamma(r)\Gamma(1-r)} \int_0^1 t^{r-1}(1-t)^{-r} dt = 1,$$

since this is a beta$(r, 1-r)$.

c.

$$\frac{d}{dx} \log f(x) = \frac{d}{dx}\left[\log \frac{1}{\Gamma(r)\lambda^r} + (r-1)\log x - x/\lambda\right] = \frac{r-1}{x} - \frac{1}{\lambda} > 0$$

for some x, if $r > 1$. But,

$$\frac{d}{dx}\left[\log \int_0^\infty \frac{e^{-x/\nu}}{\nu} q_\lambda(\nu)d\nu\right] = \frac{-\int_0^\infty \frac{1}{\nu^2} e^{-x/\nu} q_\lambda(\nu)d\nu}{\int_0^\infty \frac{1}{\nu} e^{-x/\nu} q_\lambda(\nu)d\nu} < 0 \quad \forall x.$$

4.39 a. Without loss of generality lets assume that $i < j$. From the discussion in the text we have that

$$f(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n | x_j)$$
$$= \frac{(m-x_j)!}{x_1! \cdot \cdots \cdot x_{j-1}! \cdot x_{j+1}! \cdot \cdots \cdot x_n!}$$
$$\times \left(\frac{p_1}{1-p_j}\right)^{x_1} \cdot \cdots \cdot \left(\frac{p_{j-1}}{1-p_j}\right)^{x_{j-1}} \left(\frac{p_{j+1}}{1-p_j}\right)^{x_{j+1}} \cdot \cdots \cdot \left(\frac{p_n}{1-p_j}\right)^{x_n}.$$

Then,

$$f(x_i|x_j)$$
$$= \sum_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{j-1},x_{j+1},\ldots,x_n)} f(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n | x_j)$$
$$= \sum_{(x_k \neq x_i, x_j)} \frac{(m-x_j)!}{x_1! \cdot \cdots \cdot x_{j-1}! \cdot x_{j+1}! \cdot \cdots \cdot x_n!}$$
$$\times \left(\frac{p_1}{1-p_j}\right)^{x_1} \cdot \cdots \cdot \left(\frac{p_{j-1}}{1-p_j}\right)^{x_{j-1}} \left(\frac{p_{j+1}}{1-p_j}\right)^{x_{j+1}} \cdot \cdots \cdot \left(\frac{p_n}{1-p_j}\right)^{x_n}$$
$$\times \frac{(m-x_i-x_j)!\left(1-\frac{p_i}{1-p_j}\right)^{m-x_i-x_j}}{(m-x_i-x_j)!\left(1-\frac{p_i}{1-p_j}\right)^{m-x_i-x_j}}$$
$$= \frac{(m-x_j)!}{x_i!(m-x_i-x_j)!}\left(\frac{p_i}{1-p_j}\right)^{x_i}\left(1-\frac{p_i}{1-p_j}\right)^{m-x_i-x_j}$$
$$\times \sum_{(x_k \neq x_i, x_j)} \frac{(m-x_i-x_j)!}{x_1! \cdot \cdots \cdot x_{i-1}!, x_{i+1}! \cdot \cdots \cdot x_{j-1}!, x_{j+1}! \cdot \cdots \cdot x_n!}$$
$$\times \left(\frac{p_1}{1-p_j-p_i}\right)^{x_1} \cdot \cdots \cdot \left(\frac{p_{i-1}}{1-p_j-p_i}\right)^{x_{i-1}} \left(\frac{p_{i+1}}{1-p_j-p_i}\right)^{x_{i+1}}$$

$$\times \left(\frac{p_{j-1}}{1-p_j-p_i}\right)^{x_{j-1}}\left(\frac{p_{j+1}}{1-p_j-p_i}\right)^{x_{j+1}}\cdots\cdot\left(\frac{p_n}{1-p_j-p_i}\right)^{x_n}$$

$$= \frac{(m-x_j)!}{x_i!(m-x_i-x_j)!}\left(\frac{p_i}{1-p_j}\right)^{x_i}\left(1-\frac{p_i}{1-p_j}\right)^{m-x_i-x_j}.$$

Thus $X_i|X_j = x_j \sim \text{binomial}(m-x_j, \frac{p_i}{1-p_j})$.

b.

$$f(x_i,x_j) = f(x_i|x_j)f(x_j) = \frac{m!}{x_i!x_j!(m-x_j-x_i)!}p_i^{x_i}p_j^{x_j}(1-p_j-p_i)^{m-x_j-x_i}.$$

Using this result it can be shown that $X_i + X_j \sim \text{binomial}(m, p_i+p_j)$. Therefore,

$$\text{Var}(X_i+X_j) = m(p_i+p_j)(1-p_i-p_j).$$

By Theorem 4.5.6 $\text{Var}(X_i+X_j) = \text{Var}(X_i)+\text{Var}(X_j)+2\text{Cov}(X_i,X_j)$. Therefore,

$$\text{Cov}(X_i,X_j) = \frac{1}{2}[m(p_i+p_j)(1-p_i-p_j)-mp_i(1-p_i)-mp_i(1-p_i)] = \frac{1}{2}(-2mp_ip_j) = -mp_ip_j.$$

4.41 Let a be a constant. $\text{Cov}(a,X) = \text{E}(aX) - \text{E}a\text{E}X = a\text{E}X - a\text{E}X = 0$.

4.42

$$\rho_{XY,Y} = \frac{\text{Cov}(XY,Y)}{\sigma_{XY}\sigma_Y} = \frac{\text{E}(XY^2)-\mu_{XY}\mu_Y}{\sigma_{XY}\sigma_Y} = \frac{\text{E}X\text{E}Y^2-\mu_X\mu_Y\mu_Y}{\sigma_{XY}\sigma_Y},$$

where the last step follows from the independence of X and Y. Now compute

$$\begin{aligned}\sigma^2_{XY} &= \text{E}(XY)^2 - [\text{E}(XY)]^2 = \text{E}X^2\text{E}Y^2 - (\text{E}X)^2(\text{E}Y)^2 \\ &= (\sigma_X^2+\mu_X^2)(\sigma_Y^2+\mu_Y^2) - \mu_X^2\mu_Y^2 = \sigma_X^2\sigma_Y^2+\sigma_X^2\mu_Y^2+\sigma_Y^2\mu_X^2.\end{aligned}$$

Therefore,

$$\rho_{XY,Y} = \frac{\mu_X(\sigma_Y^2+\mu_Y^2)-\mu_X\mu_Y^2}{(\sigma_X^2\sigma_Y^2+\sigma_X^2\mu_Y^2+\sigma_Y^2\mu_X^2)^{1/2}\sigma_Y} = \frac{\mu_X\sigma_Y}{(\mu_X^2\sigma_Y^2+\mu_Y^2\sigma_X^2+\sigma_X^2\sigma_Y^2)^{1/2}}.$$

4.43

$$\begin{aligned}\text{Cov}(X_1+X_2,X_2+X_3) &= \text{E}(X_1+X_2)(X_2+X_3) - \text{E}(X_1+X_2)\text{E}(X_2+X_3) \\ &= (4\mu^2+\sigma^2) - 4\mu^2 = \sigma^2 \\ \text{Cov}(X_1+X_2)(X_1-X_2) &= \text{E}(X_1+X_2)(X_1-X_2) = \text{E}X_1^2 - X_2^2 = 0.\end{aligned}$$

4.44 Let $\mu_i = \text{E}(X_i)$. Then

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^n X_i\right) &= \text{Var}\left(X_1+X_2+\cdots+X_n\right) \\ &= \text{E}\left[(X_1+X_2+\cdots+X_n)-(\mu_1+\mu_2+\cdots+\mu_n)\right]^2 \\ &= \text{E}\left[(X_1-\mu_1)+(X_2-\mu_2)+\cdots+(X_n-\mu_n)\right]^2 \\ &= \sum_{i=1}^n \text{E}(X_i-\mu_i)^2 + 2\sum_{1\le i<j\le n}\text{E}(X_i-\mu_i)(X_j-\mu_j) \\ &= \sum_{i=1}^n \text{Var}X_i + 2\sum_{1\le i<j\le n}\text{Cov}(X_i,X_j).\end{aligned}$$

4.45 a. We will compute the marginal of X. The calculation for Y is similar. Start with

$$\begin{aligned} f_{XY}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\times \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 -2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right] \end{aligned}$$

and compute

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\omega^2-2\rho\omega z+z^2\right)}\sigma_Y\,dz,$$

where we make the substitution $z=\frac{y-\mu_Y}{\sigma_Y}$, $dy=\sigma_Y dz$, $\omega=\frac{x-\mu_X}{\sigma_X}$. Now the part of the exponent involving ω^2 can be removed from the integral, and we complete the square in z to get

$$\begin{aligned} f_X(x) &= \frac{e^{-\frac{\omega^2}{2(1-\rho^2)}}}{2\pi\sigma_X\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[(z^2-2\rho\omega z+\rho^2\omega^2)-\rho^2\omega^2\right]}dz \\ &= \frac{e^{-\omega^2/2(1-\rho^2)}e^{\rho^2\omega^2/2(1-\rho^2)}}{2\pi\sigma_X\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(z-\rho\omega)^2}dz. \end{aligned}$$

The integrand is the kernel of normal pdf with $\sigma^2=(1-\rho^2)$, and $\mu=\rho\omega$, so it integrates to $\sqrt{2\pi}\sqrt{1-\rho^2}$. Also note that $e^{-\omega^2/2(1-\rho^2)}e^{\rho^2\omega^2/2(1-\rho^2)}=e^{-\omega^2/2}$. Thus,

$$f_X(x)=\frac{e^{-\omega^2/2}}{2\pi\sigma_X\sqrt{1-\rho^2}}\sqrt{2\pi}\sqrt{1-\rho^2}=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2},$$

the pdf of $\mathrm{n}(\mu_X,\sigma_X^2)$.

b.

$$\begin{aligned} & f_{Y|X}(y|x) \\ &= \frac{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}}{\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-(1-\rho^2)\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2\sigma_Y^2\sqrt{(1-\rho^2}}\left[(y-\mu_Y)-\left(\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right]^2}, \end{aligned}$$

which is the pdf of $\mathrm{n}\Big((\mu_Y-\rho(\sigma_Y/\sigma_X)(x-\mu_X)\,,\sigma_Y\sqrt{1-\rho^2}\Big)$.

c. The mean is easy to check,

$$\mathrm{E}(aX+bY)=a\mathrm{E}X+b\mathrm{E}Y=a\mu_X+b\mu_Y,$$

as is the variance,

$$\text{Var}(aX+bY) = a^2\text{Var}X + b^2\text{Var}Y + 2ab\text{Cov}(X,Y) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y.$$

To show that $aX+bY$ is normal we have to do a bivariate transform. One possibility is $U = aX+bY$, $V = Y$, then get $f_{U,V}(u,v)$ and show that $f_U(u)$ is normal. We will do this in the standard case. Make the indicated transformation and write $x = \frac{1}{a}(u-bv)$, $y = v$ and obtain

$$|J| = \begin{vmatrix} 1/a & -b/a \\ 0 & 1 \end{vmatrix} = \frac{1}{a}.$$

Then

$$f_{UV}(u,v) = \frac{1}{2\pi a\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left[\frac{1}{a}(u-bv)\right]^2 - 2\frac{\rho}{a}(u-bv)+v^2\right]}.$$

Now factor the exponent to get a square in u. The result is

$$-\frac{1}{2(1-\rho^2)}\left[\frac{b^2+2\rho ab+a^2}{a^2}\right]\left[\frac{u^2}{b^2+2\rho ab+a^2} - 2\left(\frac{b+a\rho}{b^2+2\rho ab+a^2}\right)uv + v^2\right].$$

Note that this is joint bivariate normal form since $\mu_U = \mu_V = 0$, $\sigma_v^2 = 1$, $\sigma_u^2 = a^2+b^2+2ab\rho$ and

$$\rho^* = \frac{\text{Cov}(U,V)}{\sigma_U\sigma_V} = \frac{\text{E}(aXY+bY^2)}{\sigma_U\sigma_V} = \frac{a\rho+b}{\sqrt{a^2+b^2+ab\rho}},$$

thus

$$(1-\rho^{*2}) = 1 - \frac{a^2\rho^2+ab\rho+b^2}{a^2+b^2+2ab\rho} = \frac{(1-\rho^2)a^2}{a^2+b^2+2ab\rho} = \frac{(1-\rho^2)a^2}{\sigma_u^2}$$

where $a\sqrt{1-\rho^2} = \sigma_U\sqrt{1-\rho^{*2}}$. We can then write

$$f_{UV}(u,v) = \frac{1}{2\pi\sigma_U\sigma_V\sqrt{1-\rho^{*2}}}\exp\left[-\frac{1}{2\sqrt{1-\rho^{*2}}}\left(\frac{u^2}{\sigma_U^2} - 2\rho\frac{uv}{\sigma_U\sigma_V} + \frac{v^2}{\sigma_V^2}\right)\right],$$

which is in the exact form of a bivariate normal distribution. Thus, by part a), U is normal.

4.46 a.

$$\begin{aligned}
\text{E}X &= a_X\text{E}Z_1 + b_X\text{E}Z_2 + \text{E}c_X = a_X0 + b_X0 + c_X = c_X \\
\text{Var}X &= a_X^2\text{Var}Z_1 + b_X^2\text{Var}Z_2 + \text{Var}c_X = a_X^2 + b_X^2 \\
\text{E}Y &= a_Y0 + b_Y0 + c_Y = c_Y \\
\text{Var}Y &= a_Y^2\text{Var}Z_1 + b_Y^2\text{Var}Z_2 + \text{Var}c_Y = a_Y^2 + b_Y^2 \\
\text{Cov}(X,Y) &= \text{E}XY - \text{E}X\cdot\text{E}Y \\
&= \text{E}[(a_Xa_YZ_1^2 + b_Xb_YZ_2^2 + c_Xc_Y + a_Xb_YZ_1Z_2 + a_Xc_YZ_1 + b_Xa_YZ_2Z_1 \\
&\qquad + b_Xc_YZ_2 + c_Xa_YZ_1 + c_Xb_YZ_2) - c_Xc_Y] \\
&= a_Xa_Y + b_Xb_Y,
\end{aligned}$$

since $\text{E}Z_1^2 = \text{E}Z_2^2 = 1$, and expectations of other terms are all zero.

b. Simply plug the expressions for a_X, b_X, etc. into the equalities in a) and simplify.

c. Let $D = a_Xb_Y - a_Yb_X = -\sqrt{1-\rho^2}\sigma_X\sigma_Y$ and solve for Z_1 and Z_2,

$$\begin{aligned}
Z_1 &= \frac{b_Y(X-c_X) - b_X(Y-c_Y)}{D} = \frac{\sigma_Y(X-\mu_X)+\sigma_X(Y-\mu_Y)}{\sqrt{2(1+\rho)}\sigma_X\sigma_Y} \\
Z_2 &= \frac{\sigma_Y(X-\mu_X)+\sigma_X(Y-\mu_Y)}{\sqrt{2(1-\rho)}\sigma_X\sigma_Y}.
\end{aligned}$$

Then the Jacobian is

$$J = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial y} \\ \frac{\partial z_2}{\partial x} & \frac{\partial z_2}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{b_Y}{D} & \frac{-b_X}{D} \\ \frac{-a_Y}{D} & \frac{a_X}{D} \end{pmatrix} = \frac{a_X b_Y}{D^2} - \frac{a_Y b_X}{D^2} = \frac{1}{D} = \frac{1}{-\sqrt{1-\rho^2}\sigma_X\sigma_Y},$$

and we have that

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\sigma_Y(x-\mu_X)+\sigma_X(y-\mu_Y))^2}{2(1+\rho)\sigma_X^2\sigma_Y^2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\sigma_Y(x-\mu_X)+\sigma_X(y-\mu_Y))^2}{2(1-\rho)\sigma_X^2\sigma_Y^2}} \frac{1}{\sqrt{1-\rho^2}\sigma_X\sigma_Y} \\ &= (2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2})^{-1}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \\ &- 2\rho\frac{x-\mu_X}{\sigma_X}\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2, \quad -\infty < x < \infty,\ -\infty < y < \infty, \end{aligned}$$

a bivariate normal pdf.

d. Another solution is

$$\begin{aligned} a_X &= \rho\sigma_X b_X = \sqrt{(1-\rho^2)}\sigma_X \\ a_Y &= \sigma_Y b_Y = 0 \\ c_X &= \mu_X \\ c_Y &= \mu_Y. \end{aligned}$$

There are an infinite number of solutions. Write $b_X = \pm\sqrt{\sigma_X^2 - a_X^2}$,$b_Y = \pm\sqrt{\sigma_Y^2 - a_Y^2}$, and substitute b_X,b_Y into $a_X a_Y = \rho\sigma_X\sigma_Y$. We get

$$a_X a_Y + \left(\pm\sqrt{\sigma_X^2 - a_X^2}\right)\left(\pm\sqrt{\sigma_Y^2 - a_Y^2}\right) = \rho\sigma_X\sigma_Y.$$

Square both sides and simplify to get

$$(1-\rho^2)\sigma_X^2\sigma_Y^2 = \sigma_X^2 a_Y^2 - 2\rho\sigma_X\sigma_Y a_X a_Y + \sigma_Y^2 a_X^2.$$

This is an ellipse for $\rho \neq \pm 1$, a line for $\rho = \pm 1$. In either case there are an infinite number of points satisfying the equations.

4.47 a. By definition of Z, for $z < 0$,

$$\begin{aligned} P(Z \le z) &= P(X \le z \text{ and } XY > 0) + P(-X \le z \text{ and } XY < 0) \\ &= P(X \le z \text{ and } Y < 0) + P(X \ge -z \text{ and } Y < 0) \qquad \text{(since } z < 0\text{)} \\ &= P(X \le z)P(Y < 0) + P(X \ge -z)P(Y < 0) \qquad \text{(independence)} \\ &= P(X \le z)P(Y < 0) + P(X \le z)P(Y > 0) \qquad \text{(symmetry of } X \text{and } Y\text{)} \\ &= P(X \le z)(P(Y < 0) + P(Y > 0)) \\ &= P(X \le z). \end{aligned}$$

By a similar argument, for $z > 0$, we get $P(Z > z) = P(X > z)$, and hence, $P(Z \le z) = P(X \le z)$. Thus, $Z \sim X \sim \mathrm{n}(0,1)$.

b. By definition of Z, $Z > 0 \Leftrightarrow$ either (i)$X < 0$ and $Y > 0$ or (ii)$X > 0$ and $Y > 0$. So Z and Y always have the same sign, hence they cannot be bivariate normal.

4.49 a.

$$\begin{aligned} f_X(x) &= \int (af_1(x)g_1(y) + (1-a)f_2(x)g_2(y))dy \\ &= af_1(x)\int g_1(y)dy + (1-a)f_2(x)\int g_2(y)dy \\ &= af_1(x) + (1-a)f_2(x). \\ f_Y(y) &= \int (af_1(x)g_1(y) + (1-a)f_2(x)g_2(y))dx \\ &= ag_1(y)\int f_1(x)dx + (1-a)g_2(y)\int f_2(x)dx \\ &= ag_1(y) + (1-a)g_2(y). \end{aligned}$$

b. ($\Rightarrow$) If X and Y are independent then $f(x,y) = f_X(x)f_Y(y)$. Then,

$$\begin{aligned} &f(x,y) - f_X(x)f_Y(y) \\ &\quad= af_1(x)g_1(y) + (1-a)f_2(x)g_2(y) \\ &\qquad - [af_1(x) + (1-a)f_2(x)][ag_1(y) + (1-a)g_2(y)] \\ &\quad= a(1-a)[f_1(x)g_1(y) - f_1(x)g_2(y) - f_2(x)g_1(y) + f_2(x)g_2(y)] \\ &\quad= a(1-a)[f_1(x) - f_2(x)][g_1(y) - g_2(y)] \\ &\quad= 0. \end{aligned}$$

Thus $[f_1(x) - f_2(x)][g_1(y) - g_2(y)] = 0$ since $0 < a < 1$.
($\Leftarrow$) if $[f_1(x) - f_2(x)][g_1(y) - g_2(y)] = 0$ then

$$f_1(x)g_1(y) + f_2(x)g_2(y) = f_1(x)g_2(y) + f_2(x)g_1(y).$$

Therefore

$$\begin{aligned} &f_X(x)f_Y(y) \\ &\quad= a^2f_1(x)g_1(y) + a(1-a)f_1(x)g_2(y) + a(1-a)f_2(x)g_1(y) + (1-a)^2f_2(x)g_2(y) \\ &\quad= a^2f_1(x)g_1(y) + a(1-a)[f_1(x)g_2(y) + f_2(x)g_1(y)] + (1-a)^2f_2(x)g_2(y) \\ &\quad= a^2f_1(x)g_1(y) + a(1-a)[f_1(x)g_1(y) + f_2(x)g_2(y)] + (1-a)^2f_2(x)g_2(y) \\ &\quad= af_1(x)g_1(y) + (1-a)f_2(x)g_2(y) = f(x,y). \end{aligned}$$

Thus X and Y are independent.

c.

$$\begin{aligned} \text{Cov}(X,Y) &= a\mu_1\xi_1 + (1-a)\mu_2\xi_2 - [a\mu_1 + (1-a)\mu_2][a\xi_1 + (1-a)\xi_2] \\ &= a(1-a)[\mu_1\xi_1 - \mu_1\xi_2 - \mu_2\xi_1 + \mu_2\xi_2] \\ &= a(1-a)[\mu_1 - \mu_2][\xi_1 - \xi_2]. \end{aligned}$$

To construct dependent uncorrelated random variables let $(X,Y) \sim af_1(x)g_1(y) + (1-a)f_2(x)g_2(y)$ where f_1, f_2, g_1, g_2 are such that $f_1 - f_2 \neq 0$ and $g_1 - g_2 \neq 0$ with $\mu_1 = \mu_2$ or $\xi_1 = \xi_2$.

d. (i) $f_1 \sim$ binomial(n,p), $f_2 \sim$ binomial(n,p), $g_1 \sim$ binomial(n,p), $g_2 \sim$ binomial$(n,1-p)$.
(ii) $f_1 \sim$ binomial(n,p_1), $f_2 \sim$ binomial(n,p_2), $g_1 \sim$ binomial(n,p_1), $g_2 \sim$ binomial(n,p_2).
(iii) $f_1 \sim$ binomial$(n_1,\frac{p}{n_1})$, $f_2 \sim$ binomial$(n_2,\frac{p}{n_2})$, $g_1 \sim$ binomial(n_1,p), $g_2 \sim$ binomial(n_2,p).

4.51 a.

$$\begin{aligned} P(X/Y \le t) &= \begin{cases} \frac{1}{2}t & t > 1 \\ \frac{1}{2} + (1-t) & t \le 1 \end{cases} \\ P(XY \le t) &= t - t\log t \qquad 0 < t < 1. \end{aligned}$$

b.

$$\begin{aligned} P(XY/Z \le t) &= \int_0^1 P(XY \le zt)dz \\ &= \begin{cases} \int_0^1 \left[\frac{zt}{2} + (1-zt)\right] dz & \text{if } t \le 1 \\ \int_0^{\frac{1}{t}} \left[\frac{zt}{2} + (1-zt)\right] dz + \int_{\frac{1}{t}}^1 \frac{1}{2zt}dz & \text{if } t \le 1 \end{cases} \\ &= \begin{cases} 1 - t/4 & \text{if } t \le 1 \\ t - \frac{1}{4t} + \frac{1}{2t}\log t & \text{if } t > 1 \end{cases}. \end{aligned}$$

4.53

$$\begin{aligned} P(\text{Real Roots}) &= P(B^2 > 4AC) \\ &= P(2\log B > \log 4 + \log A + \log C) \\ &= P(-2\log B \le -\log 4 - \log A - \log C) \\ &= P\left(-2\log B \le -\log 4 + (-\log A - \log C)\right). \end{aligned}$$

Let $X = -2\log B$, $Y = -\log A - \log C$. Then $X \sim$ exponential(2), $Y \sim$ gamma(2, 1), independent, and

$$\begin{aligned} P(\text{Real Roots}) &= P(X < -\log 4 + Y) \\ &= \int_{\log 4}^{\infty} P(X < -\log 4 + y) f_Y(y)dy \\ &= \int_{\log 4}^{\infty} \int_0^{-\log 4 + y} \frac{1}{2} e^{-x/2} dx y e^{-y} dy \\ &= \int_{\log 4}^{\infty} \left(1 - e^{-\frac{1}{2}\log 4} e^{-y/2}\right) y e^{-y} dy. \end{aligned}$$

Integration-by-parts will show that $\int_a^\infty y e^{-y/b} = b(a+b)e^{-a/b}$ and hence

$$P(\text{Real Roots}) = \frac{1}{4}(1 + \log 4) - \frac{1}{24}\left(\frac{2}{3} + \log 4\right) = .511.$$

4.54 Let $Y = \prod_{i=1}^n X_i$. Then $P(Y \le y) = P(\prod_{i=1}^n X_i \le y) = P(\sum_{i=1}^n -\log X_i \ge -\log y)$. Now, $-\log X_i \sim$ exponential(1) = gamma(1, 1). By Example 4.6.8, $\sum_{i=1}^n -\log X_i \sim$ gamma(n, 1). Therefore,

$$P(Y \le y) = \int_{-\log y}^{\infty} \frac{1}{\Gamma(n)} z^{n-1} e^{-z} dz,$$

and

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} \int_{-\log y}^{\infty} \frac{1}{\Gamma(n)} z^{n-1} e^{-z} dz \\ &= -\frac{1}{\Gamma(n)} (-\log y)^{n-1} e^{-(-\log y)} \frac{d}{dy}(-\log y) \\ &= \frac{1}{\Gamma(n)} (-\log y)^{n-1}, \qquad 0 < y < 1. \end{aligned}$$

4.55 Let X_1, X_2, X_3 be independent exponential(λ) random variables, and let $Y = \max(X_1, X_2, X_3)$, the lifetime of the system. Then

$$\begin{aligned} P(Y \le y) &= P\left(\max(X_1, X_2, X_3) \le y\right) \\ &= P\left(X_1 \le y \text{ and } X_2 \le y \text{ and } X_3 \le y\right) \\ &= P(X_1 \le y)P(X_2 \le y)P(X_3 \le y). \end{aligned}$$

by the independence of X_1, X_2 and X_3. Now each probability is $P(X_1 \le y) = \int_0^y \frac{1}{\lambda} e^{-x/\lambda} dx = 1 - e^{-y/\lambda}$, so

$$P(Y \le y) = \left(1 - e^{-y/\lambda}\right)^3, \qquad 0 < y < \infty,$$

and the pdf is

$$f_Y(y) = \begin{cases} 3\left(1 - e^{-y/\lambda}\right)^2 e^{-y/\lambda} & y > 0 \\ 0 & y \le 0. \end{cases}$$

4.57 a.

$$\begin{aligned} A_1 &= [\frac{1}{n}\sum_{x=1}^n x_i^1]^{\frac{1}{1}} = \frac{1}{n}\sum_{x=1}^n x_i, \quad \text{the arithmetic mean.} \\ A_{-1} &= [\frac{1}{n}\sum_{x=1}^n x_i^{-1}]^{-1} = \frac{1}{\frac{1}{n}(\frac{1}{x_1} + \cdots + \frac{1}{x_n})}, \quad \text{the harmonic mean.} \end{aligned}$$

$$\begin{aligned} \lim_{r\to 0} \log A_r &= \lim_{r\to 0} \log[\frac{1}{n}\sum_{x=1}^n x_i^r]^{\frac{1}{r}} = \lim_{r\to 0} \frac{1}{r} \log[\frac{1}{n}\sum_{x=1}^n x_i^r] = \lim_{r\to 0} \frac{\frac{1}{n}\sum_{i=1}^n r x_i^{r-1}}{\frac{1}{n}\sum_{i=1}^n x_i^r} \\ &= \lim_{r\to 0} \frac{\frac{1}{n}\sum_{i=1}^n x_i^r \log x_i}{\frac{1}{n}\sum_{i=1}^n x_i^r} = \frac{1}{n}\sum_{i=1}^n \log x_i = \frac{1}{n}\log(\prod_{i=1}^n x_i). \end{aligned}$$

Thus $A_0 = \lim_{r\to 0} A_r = \exp(\frac{1}{n}\log(\prod_{i=1}^n x_i)) = (\prod_{i=1}^n x_i)^{\frac{1}{n}}$, the geometric mean. The term $rx_i^{r-1} = x_i^r \log x_i$ since $rx_i^{r-1} = \frac{d}{dr}x_i^r = \frac{d}{dr}\exp(r\log x_i) = \exp(r\log x_i)\log x_i = x_i^r \log x_i$.

b. (i) if $\log A_r$ is nondecreasing then for $r \le r'$ $\log A_r \le \log A_{r'}$, then $e^{\log A_r} \le e^{\log A_{r'}}$. Therefore $A_r \le A_{r'}$. Thus A_r is nondecreasing in r.

(ii) $\frac{d}{dr}\log A_r = \frac{-1}{r^2}\log(\frac{1}{n}\sum_{x=1}^n x_i^r) + \frac{1}{r}\frac{\frac{1}{n}\sum_{i=1}^n r x_i^{r-1}}{\frac{1}{n}\sum_{i=1}^n x_i^r} = \frac{1}{r^2}\left[\frac{r\sum_{i=1}^n x_i^r \log x_i}{\sum_{x=1}^n x_i^r} - \log(\frac{1}{n}\sum_{x=1}^n x_i^r)\right]$, where we use the identity for rx_i^{r-1} showed in a).

(iii)

$$\begin{aligned} &\frac{r\sum_{i=1}^n x_i^r \log x_i}{\sum_{x=1}^n x_i^r} - \log(\frac{1}{n}\sum_{x=1}^n x_i^r) \\ &= \log(n) + \frac{r\sum_{i=1}^n x_i^r \log x_i}{\sum_{x=1}^n x_i^r} - \log(\sum_{x=1}^n x_i^r) \\ &= \log(n) + \sum_{i=1}^n \left[\frac{x_i^r}{\sum_{i=1}^n x_i^r} r\log x_i - \frac{x_i^r}{\sum_{i=1}^n x_i^r}\log(\sum_{x=1}^n x_i^r)\right] \\ &= \log(n) + \sum_{i=1}^n \left[\frac{x_i^r}{\sum_{i=1}^n x_i^r}(r\log x_i - \log(\sum_{x=1}^n x_i^r))\right] \\ &= \log(n) - \sum_{i=1}^n \frac{x_i^r}{\sum_{i=1}^n x_i^r}\log(\frac{\sum_{x=1}^n x_i^r}{x_i^r}) = \log(n) - \sum_{i=1}^n a_i \log(\frac{1}{a_i}). \end{aligned}$$

We need to prove that $\log(n) \geq \sum_{i=1}^{n} a_i \log(\frac{1}{a_i})$. Using Jensen inequality we have that $\mathrm{E}\log(\frac{1}{a}) = \sum_{i=1}^{n} a_i \log(\frac{1}{a_i}) \leq \log(\mathrm{E}\frac{1}{a}) = \log(\sum_{i=1}^{n} a_i \frac{1}{a_i}) = \log(n)$ which establish the result.

4.59 Assume that $\mathrm{E}X = 0$, $\mathrm{E}Y = 0$, and $\mathrm{E}Z = 0$. This can be done without loss of generality because we could work with the quantities $X - \mathrm{E}X$, etc. By iterating the expectation we have

$$\mathrm{Cov}(X, Y) = \mathrm{E}XY = \mathrm{E}[\mathrm{E}(XY|Z)].$$

Adding and subtracting $\mathrm{E}(X|Z)\mathrm{E}(Y|Z)$ gives

$$\mathrm{Cov}(X, Y) = \mathrm{E}[\mathrm{E}(XY|Z) - \mathrm{E}(X|Z)\mathrm{E}(Y|Z)] + \mathrm{E}[\mathrm{E}(X|Z)\mathrm{E}(Y|Z)].$$

Since $\mathrm{E}[\mathrm{E}(X|Z)] = \mathrm{E}X = 0$, the second term above is $\mathrm{Cov}[\mathrm{E}(X|Z)\mathrm{E}(Y|Z)]$. For the first term write

$$\mathrm{E}[\mathrm{E}(XY|Z) - \mathrm{E}(X|Z)\mathrm{E}(Y|Z)] = \mathrm{E}\left[\mathrm{E}\left\{XY - \mathrm{E}(X|Z)\mathrm{E}(Y|Z)|\, Z\right\}\right]$$

where we have brought $\mathrm{E}(X|Z)$ and $\mathrm{E}(Y|Z)$ inside the conditional expectation. This can now be recognized as $\mathrm{E}\mathrm{Cov}(X, Y|Z)$, establishing the identity.

4.61 a. To find the distribution of $f(X_1|Z)$, let $U = \frac{X_2-1}{X_1}$ and $V = X_1$. Then $x_2 = h_1(u, v) = uv+1$, $x_1 = h_2(u, v) = v$. Therefore

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J| = e^{-(uv+1)}e^{-v}v,$$

and

$$f_U(u) = \int_0^\infty ve^{-(uv+1)}e^{-v}dv = \frac{e^{-1}}{(u+1)^2}.$$

Thus $V|U = 0$ has distribution ve^{-v}. The distribution of $X_1|X_2$ is e^{-x_1} since X_1 and X_2 are independent.

b. The following Mathematica code will draw the picture; the solid lines are B_1 and the dashed lines are B_2. Note that the solid lines increase with $x1$, while the dashed lines are constant. Thus B_1 is informative, as the range of X_2 changes.

```
e = 1/10;
Plot[{-e*x1 + 1, e*x1 + 1, 1 - e, 1 + e}, {x1, 0, 5},
PlotStyle -> {Dashing[{}], Dashing[{}],Dashing[{0.15, 0.05}],
  Dashing[{0.15, 0.05}]}]
```

c.

$$\begin{aligned} P(X_1 \leq x|B_1) &= P(V \leq v^*| -\epsilon < U < \epsilon) = \frac{\int_0^{v^*}\int_{-\epsilon}^{\epsilon} ve^{-(uv+1)}e^{-v}dudv}{\int_0^{\infty}\int_{-\epsilon}^{\epsilon} ve^{-(uv+1)}e^{-v}dudv} \\ &= \frac{e^{-1}\left[\frac{e^{-v^*(1+\epsilon)}}{1+\epsilon} - \frac{1}{1+\epsilon} - \frac{e^{-v^*(1-\epsilon)}}{1-\epsilon} + \frac{1}{1-\epsilon}\right]}{e^{-1}\left[-\frac{1}{1+\epsilon} + \frac{1}{1-\epsilon}\right]}. \end{aligned}$$

Thus $\lim_{\epsilon\to 0} P(X_1 \leq x|B_1) = 1 - e^{-v^*} - v^*e^{-v^*} = \int_0^{v^*} ve^{-v}dv = P(V \leq v^*|U = 0)$.

$$P(X_1 \leq x|B_2) = \frac{\int_0^x \int_0^{1+\epsilon} e^{-(x_1+x_2)}dx_2dx_1}{\int_0^{1+\epsilon} e^{-x_2}dx_2} = \frac{e^{-(x+1+\epsilon)} - e^{-(1+\epsilon)} - e^{-x} + 1}{1 - e^{-(1+\epsilon)}}.$$

Thus $\lim_{\epsilon\to 0} P(X_1 \leq x|B_2) = 1 - e^x = \int_0^x e^{x_1}dx_1 = P(X_1 \leq x|X_2 = 1)$.

4.63 Since $X = e^Z$ and $g(z) = e^z$ is convex, by Jensen's Inequality $\mathrm{E}X = \mathrm{E}g(Z) \geq g(\mathrm{E}Z) = e^0 = 1$. In fact, there is equality in Jensen's Inequality if and only if there is an interval I with $P(Z \in I) = 1$ and $g(z)$ is linear on I. But e^z is linear on an interval only if the interval is a single point. So $\mathrm{E}X > 1$, unless $P(Z = \mathrm{E}Z = 0) = 1$.

4.64 a. Let a and b be real numbers. Then,

$$|a+b|^2 = (a+b)(a+b) = a^2 + 2ab + b^2 \leq |a|^2 + 2|ab| + |b|^2 = (|a| + |b|)^2.$$

Take the square root of both sides to get $|a+b| \leq |a| + |b|$.

b. $|X+Y| \leq |X| + |Y| \Rightarrow \mathrm{E}|X+Y| \leq \mathrm{E}(|X| + |Y|) = \mathrm{E}|X| + \mathrm{E}|Y|$.

4.65 Without loss of generality let us assume that $\mathrm{E}g(X) = \mathrm{E}h(X) = 0$. For part (a)

$$\begin{aligned}
\mathrm{E}(g(X)h(X)) &= \int_{-\infty}^{\infty} g(x)h(x)f_X(x)dx \\
&= \int_{\{x:h(x)\leq 0\}} g(x)h(x)f_X(x)dx + \int_{\{x:h(x)\geq 0\}} g(x)h(x)f_X(x)dx \\
&\leq g(x_0)\int_{\{x:h(x)\leq 0\}} h(x)f_X(x)dx + g(x_0)\int_{\{x:h(x)\geq 0\}} h(x)f_X(x)dx \\
&= \int_{-\infty}^{\infty} h(x)f_X(x)dx \\
&= g(x_0)\mathrm{E}h(X) = 0.
\end{aligned}$$

where x_0 is the number such that $h(x_0) = 0$. Note that $g(x_0)$ is a maximum in $\{x : h(x) \leq 0\}$ and a minimum in $\{x : h(x) \geq 0\}$ since $g(x)$ is nondecreasing. For part (b) where $g(x)$ and $h(x)$ are both nondecreasing

$$\begin{aligned}
\mathrm{E}(g(X)h(X)) &= \int_{-\infty}^{\infty} g(x)h(x)f_X(x)dx \\
&= \int_{\{x:h(x)\leq 0\}} g(x)h(x)f_X(x)dx + \int_{\{x:h(x)\geq 0\}} g(x)h(x)f_X(x)dx \\
&\geq g(x_0)\int_{\{x:h(x)\leq 0\}} h(x)f_X(x)dx + g(x_0)\int_{\{x:h(x)\geq 0\}} h(x)f_X(x)dx \\
&= \int_{-\infty}^{\infty} h(x)f_X(x)dx \\
&= g(x_0)\mathrm{E}h(X) = 0.
\end{aligned}$$

The case when $g(x)$ and $h(x)$ are both nonincreasing can be proved similarly.

Chapter 5

Properties of a Random Sample

5.1 Let $X = \#$ color blind people in a sample of size n. Then $X \sim \text{binomial}(n,p)$, where $p = .01$. The probability that a sample contains a color blind person is $P(X > 0) = 1 - P(X = 0)$, where $P(X = 0) = \binom{n}{0}(.01)^0(.99)^n = .99^n$. Thus,

$$P(X > 0) = 1 - .99^n > .95 \Leftrightarrow n > \log(.05)/\log(.99) \approx 299.$$

5.3 Note that $Y_i \sim$ Bernoulli with $p_i = P(X_i \geq \mu) = 1 - F(\mu)$ for each i. Since the Y_i's are iid Bernoulli, $\sum_{i=1}^n Y_i \sim \text{binomial}(n, p = 1 - F(\mu))$.

5.5 Let $Y = X_1 + \cdots + X_n$. Then $\bar{X} = (1/n)Y$, a scale transformation. Therefore the pdf of $\bar{X}$ is $f_{\bar{X}}(x) = \frac{1}{1/n} f_Y\left(\frac{x}{1/n}\right) = n f_Y(nx)$.

5.6 a. For $Z = X - Y$, set $W = X$. Then $Y = W - Z$, $X = W$, and $|J| = \begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} = 1$. Then $f_{Z,W}(z,w) = f_X(w) f_Y(w - z) \cdot 1$, thus $f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(w - z) dw$.

b. For $Z = XY$, set $W = X$. Then $Y = Z/W$ and $|J| = \begin{vmatrix} 0 & 1 \\ 1/w & -z/w^2 \end{vmatrix} = -1/w$. Then $f_{Z,W}(z,w) = f_X(w) f_Y(z/w) \cdot |-1/w|$, thus $f_Z(z) = \int_{-\infty}^{\infty} |-1/w| f_X(w) f_Y(z/w) dw$.

c. For $Z = X/Y$, set $W = X$. Then Y=W/Z and $|J| = \begin{vmatrix} 0 & 1 \\ -w/z^2 & 1/z \end{vmatrix} = w/z^2$. Then $f_{Z,W}(z,w) = f_X(w) f_Y(w/z) \cdot |w/z^2|$, thus $f_Z(z) = \int_{-\infty}^{\infty} |w/z^2| f_X(w) f_Y(w/z) dw$.

5.7 It is, perhaps, easiest to recover the constants by doing the integrations. We have

$$\int_{-\infty}^{\infty} \frac{B}{1+\left(\frac{\omega}{\sigma}\right)^2} d\omega \quad = \quad \sigma\pi B, \qquad \int_{-\infty}^{\infty} \frac{D}{1+\left(\frac{\omega-z}{\tau}\right)^2} d\omega \quad = \quad \tau\pi D$$

and

$$\begin{aligned} &\int_{-\infty}^{\infty} \left[\frac{A\omega}{1+\left(\frac{\omega}{\sigma}\right)^2} - \frac{C\omega}{1+\left(\frac{\omega-z}{\tau}\right)^2}\right] d\omega \\ &= \int_{-\infty}^{\infty} \left[\frac{A\omega}{1+\left(\frac{\omega}{\sigma}\right)^2} - \frac{C(\omega-z)}{1+\left(\frac{\omega-z}{\tau}\right)^2}\right] d\omega - Cz \int_{-\infty}^{\infty} \frac{1}{1+\left(\frac{\omega-z}{\tau}\right)^2} d\omega \\ &= A\frac{\sigma^2}{2} \log\left[1+\left(\frac{\omega}{\sigma}\right)^2\right] - \frac{C\tau^2}{2} \log\left[1+\left(\frac{\omega-z}{\tau}\right)^2\right]\Bigg|_{-\infty}^{\infty} - \tau\pi C z. \end{aligned}$$

The integral is finite and equal to zero if $A = M\frac{2}{\sigma^2}$, $C = M\frac{2}{\tau^2}$ for some constant M. Hence

$$f_Z(z) = \frac{1}{\pi^2\sigma\tau}\left[\sigma\pi B - \tau\pi D - \frac{2\pi M z}{\tau}\right] = \frac{1}{\pi(\sigma+\tau)} \, \frac{1}{1+(z/(\sigma+\tau))^2},$$

if $B = \frac{\tau}{\sigma+\tau}$, $D = \frac{\sigma}{\sigma+\tau)}$, $M = \frac{-\sigma\tau^2}{2z(\sigma+\tau)} \frac{1}{1+\left(\frac{z}{\sigma+\tau}\right)^2}$.

5.8 a.

$$
\begin{aligned}
&\frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2 \\
&= \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\bar{X}+\bar{X}-X_j)^2 \\
&= \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(X_i-\bar{X})^2-2(X_i-\bar{X})(X_j-\bar{X})+(X_j-\bar{X})^2\right] \\
&= \frac{1}{2n(n-1)}\left[\sum_{i=1}^{n}n(X_i-\bar{X})^2-2\sum_{i=1}^{n}(X_i-\bar{X})\underbrace{\sum_{j=1}^{n}(X_j-\bar{X})}_{=0}+n\sum_{j=1}^{n}(X_j-\bar{X})^2\right] \\
&= \frac{n}{2n(n-1)}\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{n}{2n(n-1)}\sum_{j=1}^{n}(X_j-\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad = \quad S^2.
\end{aligned}
$$

b. Although all of the calculations here are straightforward, there is a tedious amount of bookkeeping needed. It seems that induction is the easiest route. (Note: Without loss of generality we can assume $\theta_1=0$, so $\mathrm{E}X_i=0$.)

(i) Prove the equation for $n=4$. We have $S^2=\frac{1}{24}\sum_{i=1}^{4}\sum_{j=1}^{4}(X_i-X_j)^2$, and to calculate $\mathrm{Var}(S^2)$ we need to calculate $\mathrm{E}(S^2)^2$ and $\mathrm{E}(S^2)$. The latter expectation is straightforward and we get $\mathrm{E}(S^2)=24\theta_2$. The expected value $\mathrm{E}(S^2)^2=\mathrm{E}(S^4)$ contains $256(=4^4)$ terms of which $112(=4\times16+4\times16-4^2)$ are zero, whenever $i=j$. Of the remaining terms,

- 24 are of the form $\mathrm{E}(X_i-X_j)^4=2(\theta_4+3\theta_2^2)$
- 96 are of the form $\mathrm{E}(X_i-X_j)^2(X_i-X_k)^2=\theta_4+3\theta_2^2$
- 24 are of the form $\mathrm{E}(X_i-X_j)^2(X_k-X_\ell)^2=4\theta_2^2$

Thus,

$$
\mathrm{Var}(S^2)=\frac{1}{24^2}\left[24\times2(\theta_4+3\theta_2^2)+96(\theta_4+3\theta_2^2)+24\times4\theta_4-(24\theta_2)^2\right]=\frac{1}{4}\left[\theta_4-\frac{1}{3}\theta_2^2\right].
$$

(ii) Assume that the formula holds for n, and establish it for $n+1$. (Let S_n denote the variance based on n observations.) Straightforward algebra will establish

$$
\begin{aligned}
S_{n+1}^2 &= \frac{1}{2n(n+1)}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2+2\sum_{k=1}^{n}(X_k-X_{n+1})^2\right] \\
&\overset{\text{def'n}}{=} \frac{1}{2n(n+1)}[A+2B]
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{Var}(A) &= 4n(n-1)^2\left[\theta_4-\frac{n-3}{n-1}\theta_2^2\right] && \text{(induction hypothesis)} \\
\mathrm{Var}(B) &= n(n+1)\theta_4-n(n-3)\theta_2^2 && (X_k \text{ and } X_{n+1} \text{ are independent}) \\
\mathrm{Cov}(A,B) &= 2n(n-1)\left[\theta_4-\theta_2^2\right] && \text{(some minor bookkeeping needed)}
\end{aligned}
$$

Hence,

$$\mathrm{Var}(S^2_{n+1}) = \frac{1}{4n^2(n+1)^2}\left[\mathrm{Var}(A) + 4\mathrm{Var}(B) + 4\mathrm{Cov}(A,B)\right] = \frac{1}{n+1}\left[\theta_4 - \frac{n-2}{n}\theta_2^2\right],$$

establishing the induction and verifying the result.

c. Again assume that $\theta_1 = 0$. Then

$$\mathrm{Cov}(\bar{X}, S^2) = \frac{1}{2n^2(n-1)}\mathrm{E}\left\{\sum_{k=1}^{n} X_k \sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2\right\}.$$

The double sum over i and j has $n(n-1)$ nonzero terms. For each of these, the entire expectation is nonzero for only two values of k (when k matches either i or j). Thus

$$\mathrm{Cov}(\bar{X}, S^2) = \frac{2n(n-1)}{2n^2(n-1)}\mathrm{E}X_i(X_i - X_j)^2 = \frac{1}{n}\theta_3,$$

and $\bar{X}$ and S^2 are uncorrelated if $\theta_3 = 0$.

5.9 To establish the Lagrange Identity consider the case when $n = 2$,

$$\begin{aligned}
(a_1b_2 - a_2b_1)^2 &= a_1^2b_2^2 + a_2^2b_1^2 - 2a_1b_2a_2b_1 \\
&= a_1^2b_2^2 + a_2^2b_1^2 - 2a_1b_2a_2b_1 + a_1^2b_1^2 + a_2^2b_2^2 - a_1^2b_1^2 - a_2^2b_2^2 \\
&= (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1b_1 + a_2b_2)^2.
\end{aligned}$$

Assume that is true for n, then

$$\begin{aligned}
&\left(\sum_{i=1}^{n+1} a_i^2\right)\left(\sum_{i=1}^{n+1} b_i^2\right) - \left(\sum_{i=1}^{n+1} a_ib_i\right)^2 \\
&= \left(\sum_{i=1}^{n} a_i^2 + a_{n+1}^2\right)\left(\sum_{i=1}^{n} b_i^2 + b_{n+1}^2\right) - \left(\sum_{i=1}^{n} a_ib_i + a_{n+1}b_{n+1}\right)^2 \\
&= \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) - \left(\sum_{i=1}^{n} a_ib_i\right)^2 \\
&\quad + \left(\sum_{i=1}^{n} a_i^2\right)b_{n+1}^2 + a_{n+1}^2\left(\sum_{i=1}^{n} b_i^2\right) - 2\left(\sum_{i=1}^{n} a_ib_i\right)a_{n+1}b_{n+1} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(a_ib_j - a_jb_i)^2 + \sum_{i=1}^{n}(a_ib_{n+1} - a_{n+1}b_i)^2 \\
&= \sum_{i=1}^{n}\sum_{j=i+1}^{n+1}(a_ib_j - a_jb_i)^2.
\end{aligned}$$

If all the points lie on a straight line then $Y - \mu_y = c(X - \mu_x)$, for some constant $c \neq 0$. Let $b_i = Y - \mu_y$ and $a_i = (X - \mu_x)$, then $b_i = ca_i$. Therefore $\sum_{i=1}^{n}\sum_{j=i+1}^{n+1}(a_ib_j - a_jb_i)^2 = 0$. Thus the correlation coefficient is equal to 1.

5.10 a.

$$\theta_1 = \mathrm{E}X_i = \mu$$

$$\begin{aligned}
\theta_2 &= \mathrm{E}(X_i-\mu)^2 \;=\; \sigma^2\\
\theta_3 &= \mathrm{E}(X_i-\mu)^3\\
&= \mathrm{E}(X_i-\mu)^2(X_i-\mu) \qquad \text{(Stein's lemma: } \mathrm{E}g(X)(X-\theta)=\sigma^2\mathrm{E}g'(X))\\
&= 2\sigma^2\mathrm{E}(X_i-\mu) \;=\; 0\\
\theta_4 &= \mathrm{E}(X_i-\mu)^4 \;=\; \mathrm{E}(X_i-\mu)^3(X_i-\mu) \;=\; 3\sigma^2\mathrm{E}(X_i-\mu)^2 \;=\; 3\sigma^4.
\end{aligned}$$

b. $\mathrm{Var}S^2 = \frac{1}{n}(\theta_4 - \frac{n-3}{n-1}\theta_2^2) = \frac{1}{n}(3\sigma^4 - \frac{n-3}{n-1}\sigma^4) = \frac{2\sigma^4}{n-1}$.

c. Use the fact that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ and $\mathrm{Var}\chi^2_{n-1} = 2(n-1)$ to get

$$\mathrm{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

which implies $(\frac{(n-1)^2}{\sigma^4})\mathrm{Var}S^2 = 2(n-1)$ and hence

$$\mathrm{Var}S^2 = \frac{2(n-1)}{(n-1)^2/\sigma^4} = \frac{2\sigma^4}{n-1}.$$

Remark: Another approach to b), not using the χ^2 distribution, is to use linear model theory. For any matrix A $\mathrm{Var}(X'AX) = 2\mu_2^2\mathrm{tr}A^2 + 4\mu_2\theta'A\theta$, where μ_2 is σ^2, $\theta = \mathrm{E}X = \mu 1$. Write $S^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X}) = \frac{1}{n-1}X'(I-\bar{J}_n)X$.Where

$$I-\bar{J}_n = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n}\\ -\frac{1}{n} & 1-\frac{1}{n} & & \vdots\\ \vdots & & \ddots & \vdots\\ -\frac{1}{n} & \cdots & \cdots & 1-\frac{1}{n}\end{pmatrix}.$$

Notice that $\mathrm{tr}A^2 = \mathrm{tr}A = n-1$, $A\theta = 0$. So

$$\mathrm{Var}S^2 = \frac{1}{(n-1)^2}\mathrm{Var}(X'AX) = \frac{1}{(n-1)^2}\left(2\sigma^4(n-1)+0\right) = \frac{2\sigma^4}{n-1}.$$

5.11 Let $g(s) = s^2$. Since $g(\cdot)$ is a convex function, we know from Jensen's inequality that $\mathrm{E}g(S) \geq g(\mathrm{E}S)$, which implies $\sigma^2 = \mathrm{E}S^2 \geq (\mathrm{E}S)^2$. Taking square roots, $\sigma \geq \mathrm{E}S$. From the proof of Jensen's Inequality, it is clear that, in fact, the inequality will be strict unless there is an interval I such that g is linear on I and $P(X \in I) = 1$. Since s^2 is "linear" only on single points, we have $\mathrm{E}T^2 > (\mathrm{E}T)^2$ for any random variable T, unless $P(T = \mathrm{E}T) = 1$.

5.13

$$\begin{aligned}
\mathrm{E}\left(c\sqrt{S^2}\right) &= c\sqrt{\frac{\sigma^2}{n-1}}\mathrm{E}\left(\sqrt{\frac{S^2(n-1)}{\sigma^2}}\right)\\
&= c\sqrt{\frac{\sigma^2}{n-1}}\int_0^\infty \sqrt{q}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}q^{\left(\frac{n-1}{2}\right)-1}e^{-q/2}dq,
\end{aligned}$$

Since $\sqrt{S^2(n-1)/\sigma^2}$ is the square root of a χ^2 random variable. Now adjust the integrand to be another χ^2 pdf and get

$$\mathrm{E}\left(c\sqrt{S^2}\right) = c\sqrt{\frac{\sigma^2}{n-1}}\cdot\frac{\Gamma(n/2)2^{n/2}}{\Gamma((n-1)/2)2^{((n-1)/2}}\underbrace{\int_0^\infty \frac{1}{\Gamma(n/2)2^{n/2}}q^{(n-1)/2}-\frac{1}{2}e^{-q/2}dq}_{=1 \text{ since } \chi^2_n \text{ pdf}}.$$

So $c = \frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}$ gives $\mathrm{E}(cS) = \sigma$.

5.15 a.

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1} = \frac{X_{n+1} + \sum_{i=1}^{n} X_i}{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}.$$

b.

$$\begin{aligned}
nS_{n+1}^2 &= \frac{n}{(n+1)-1} \sum_{i=1}^{n+1} \left(X_i - \bar{X}_{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1} \left(X_i - \frac{X_{n+1} + n\bar{X}_n}{n+1}\right)^2 \qquad \text{(use (a))} \\
&= \sum_{i=1}^{n+1} \left(X_i - \frac{X_{n+1}}{n+1} - \frac{n\bar{X}_n}{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1} \left[\left(X_i - \bar{X}_n\right) - \left(\frac{X_{n+1}}{n+1} - \frac{\bar{X}_n}{n+1}\right)\right]^2 \qquad (\pm \bar{X}_n) \\
&= \sum_{i=1}^{n+1} \left[\left(X_i - \bar{X}_n\right)^2 - 2\left(X_i - \bar{X}_n\right)\left(\frac{X_{n+1} - \bar{X}_n}{n+1}\right) + \frac{1}{(n+1)^2}\left(X_{n+1} - \bar{X}_n\right)^2\right] \\
&= \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 + \left(X_{n+1} - \bar{X}_n\right)^2 - 2\frac{\left(X_{n+1} - \bar{X}_n\right)^2}{n+1} + \frac{n+1}{(n+1)^2}\left(X_{n+1} - \bar{X}_n\right)^2 \\
&\qquad \left(\text{since} \sum_{1}^{n} (X_i - \bar{X}_n) = 0\right) \\
&= (n-1)S_n^2 + \frac{n}{n+1}\left(X_{n+1} - \bar{X}_n\right)^2.
\end{aligned}$$

5.16 a. $\sum_{i=1}^{3} \left(\frac{X_i - i}{i}\right)^2 \sim \chi_3^2$

b. $\left(\frac{X_i - 1}{i}\right) \Big/ \sqrt{\sum_{i=2}^{3} \left(\frac{X_i - i}{i}\right)^2 \Big/ 2} \sim t_2$

c. Square the random variable in part b).

5.17 a. Let $U \sim \chi_p^2$ and $V \sim \chi_q^2$, independent. Their joint pdf is

$$\frac{1}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)2^{(p+q)/2}} u^{\frac{p}{2}-1} v^{\frac{q}{2}-1} e^{\frac{-(u+v)}{2}}.$$

From Definition 5.3.6, the random variable $X = (U/p)/(V/q)$ has an F distribution, so we make the transformation $x = (u/p)/(v/q)$ and $y = u + v$. (Of course, many choices of y will do, but this one makes calculations easy. The choice is prompted by the exponential term in the pdf.) Solving for u and v yields

$$u = \frac{\frac{p}{q}xy}{1 + \frac{q}{p}x}, \quad v = \frac{y}{1 + \frac{q}{p}x}, \text{ and } |J| = \frac{\frac{q}{p}y}{\left(1 + \frac{q}{p}x\right)^2}.$$

We then substitute into $f_{U,V}(u,v)$ to obtain

$$f_{X,Y}(x,y) = \frac{1}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)2^{(p+q)/2}} \left(\frac{\frac{p}{q}xy}{1 + \frac{q}{p}x}\right)^{\frac{p}{2}-1} \left(\frac{y}{1 + \frac{q}{p}x}\right)^{\frac{q}{2}-1} e^{\frac{-y}{2}} \frac{\frac{q}{p}y}{\left(1 + \frac{q}{p}x\right)^2}.$$

Note that the pdf factors, showing that X and Y are independent, and we can read off the pdfs of each: X has the F distribution and Y is χ^2_{p+q}. If we integrate out y to recover the proper constant, we get the F pdf

$$f_X(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\left(\frac{q}{p}\right)^{p/2}\frac{x^{p/2-1}}{\left(1+\frac{q}{p}x\right)^{\frac{p+q}{2}}}.$$

b. Since $F_{p,q} = \frac{\chi^2_p/p}{\chi^2_q/q}$, let $U \sim \chi^2_p$, $V \sim \chi^2_q$ and U and V are independent. Then we have

$$\begin{aligned} \mathrm{E}F_{p,q} &= \mathrm{E}\left(\frac{U/p}{V/q}\right) = \mathrm{E}\left(\frac{U}{p}\right)\mathrm{E}\left(\frac{q}{V}\right) && \text{(by independence)} \\ &= \frac{p}{p}q\mathrm{E}\left(\frac{1}{V}\right) && (\mathrm{E}U = p). \end{aligned}$$

Then

$$\begin{aligned} \mathrm{E}\left(\frac{1}{V}\right) &= \int_0^\infty \frac{1}{v}\frac{1}{\Gamma\left(\frac{q}{2}\right)2^{q/2}}v^{\frac{q}{2}-1}e^{-\frac{v}{2}}dv = \frac{1}{\Gamma\left(\frac{q}{2}\right)2^{q/2}}\int_0^\infty v^{\frac{q-2}{2}-1}e^{-\frac{v}{2}}dv \\ &= \frac{1}{\Gamma\left(\frac{q}{2}\right)2^{q/2}}\Gamma\left(\frac{q-2}{2}\right)2^{(q-2)/2} = \frac{\Gamma\left(\frac{q-2}{2}\right)2^{(q-2)/2}}{\Gamma\left(\frac{q-2}{2}\right)\left(\frac{q-2}{2}\right)2^{q/2}} = \frac{1}{q-2}. \end{aligned}$$

Hence, $\mathrm{E}F_{p,q} = \frac{p}{p}\frac{q}{q-2} = \frac{q}{q-2}$, if $q > 2$. To calculate the variance, first calculate

$$\mathrm{E}(F^2_{p,q}) = \mathrm{E}\left(\frac{U^2}{p^2}\frac{q^2}{V^2}\right) = \frac{q^2}{p^2}\mathrm{E}(U^2)\mathrm{E}\left(\frac{1}{V^2}\right).$$

Now

$$\mathrm{E}(U^2) = \mathrm{Var}(U) + (\mathrm{E}U)^2 = 2p + p^2$$

and

$$\mathrm{E}\left(\frac{1}{V^2}\right) = \int_0^\infty \frac{1}{v^2}\frac{1}{\Gamma(q/2)2^{q/2}}v^{(q/2)-1}e^{-v/2}dv = \frac{1}{(q-2)(q-4)}.$$

Therefore,

$$\mathrm{E}F^2_{p,q} = \frac{q^2}{p^2}p(2+p)\frac{1}{(q-2)(q-4)} = \frac{q^2}{p}\frac{(p+2)}{(q-2)(q-4)},$$

and, hence

$$\mathrm{Var}(F_{p,q}) = \frac{q^2(p+2)}{p(q-2)(q-4)} - \frac{q^2}{(q-2)^2} = 2\left(\frac{q}{q-2}\right)^2\left(\frac{q+p-2}{p(q-4)}\right),\ q > 4.$$

c. Write $X = \frac{U/p}{V/p}$ then $\frac{1}{X} = \frac{V/q}{U/p} \sim F_{q,p}$, since $U \sim \chi^2_p$, $V \sim \chi^2_q$ and U and V are independent.

d. Let $Y = \frac{(p/q)X}{1+(p/q)X} = \frac{pX}{q+pX}$, so $X = \frac{qY}{p(1-Y)}$ and $\left|\frac{dx}{dy}\right| = \frac{q}{p}(1-y)^{-2}$. Thus, Y has pdf

$$\begin{aligned} f_Y(y) &= \frac{\Gamma\left(\frac{q+p}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\left(\frac{p}{q}\right)^{\frac{p}{2}}\frac{\left(\frac{qy}{p(1-y)}\right)^{\frac{p-2}{2}}}{\left(1+\frac{p}{q}\frac{qy}{p(1-y)}\right)^{\frac{p+q}{2}}}\frac{q}{p(1-y)^2} \\ &= \left[B\left(\frac{p}{2},\frac{q}{2}\right)\right]^{-1}y^{\frac{p}{2}-1}(1-y)^{\frac{q}{2}-1} \sim \mathrm{beta}\left(\frac{p}{2},\frac{q}{2}\right). \end{aligned}$$

5.18 If $X \sim t_p$, then $X = Z/\sqrt{V/p}$ where $Z \sim \mathrm{n}(0,1)$, $V \sim \chi^2_p$ and Z and V are independent.

a. $\mathrm{E}X = \mathrm{E}Z/\sqrt{V/p} = (\mathrm{E}Z)(\mathrm{E}1/\sqrt{V/p}) = 0$, since $\mathrm{E}Z = 0$, as long as the other expectation is finite. This is so if $p > 1$. From part b), $X^2 \sim F_{1,p}$. Thus $\mathrm{Var}X = \mathrm{E}X^2 = p/(p-2)$, if $p > 2$ (from Exercise 5.17b).

b. $X^2 = Z^2/(V/p)$. $Z^2 \sim \chi^2_1$, so the ratio is distributed $F_{1,p}$.

c. The pdf of X is

$$f_X(x) = \left[\frac{\Gamma(\frac{p+1}{2})}{\Gamma(p/2)\sqrt{p\pi}}\right] \frac{1}{(1 + x^2/p)^{(p+1)/2}}.$$

Denote the quantity in square brackets by C_p. From an extension of Stirling's formula (Exercise 1.28) we have

$$\begin{aligned}
\lim_{p\to\infty} C_p &= \lim_{p\to\infty} \frac{\sqrt{2\pi}\left(\frac{p-1}{2}\right)^{\frac{p-1}{2}+\frac{1}{2}} e^{-\frac{p-1}{2}}}{\sqrt{2\pi}\left(\frac{p-2}{2}\right)^{\frac{p-2}{2}+\frac{1}{2}} e^{-\frac{p-2}{2}}} \frac{1}{\sqrt{p\pi}} \\
&= \frac{e^{-1/2}}{\sqrt{\pi}} \lim_{p\to\infty} \frac{\left(\frac{p-1}{2}\right)^{\frac{p-1}{2}+\frac{1}{2}}}{\left(\frac{p-2}{2}\right)^{\frac{p-2}{2}+\frac{1}{2}} \sqrt{p}} \quad = \quad \frac{e^{-1/2}}{\sqrt{\pi}} \frac{e^{1/2}}{\sqrt{2}},
\end{aligned}$$

by an application of Lemma 2.3.14. Applying the lemma again shows that for each x

$$\lim_{p\to\infty} \left(1+x^2/p\right)^{(p+1)/2} = e^{x^2/2},$$

establishing the result.

d. As the random variable $F_{1,p}$ is the square of a t_p, we conjecture that it would converge to the square of a $\mathrm{n}(0,1)$ random variable, a χ^2_1.

e. The random variable $qF_{q,p}$ can be thought of as the sum of q random variables, each a t_p squared. Thus, by all of the above, we expect it to converge to a χ^2_q random variable as $p \to \infty$.

5.19 a. $\chi^2_p \sim \chi^2_q + \chi^2_d$ where χ^2_q and χ^2_d are independent χ^2 random variables with q and $d = p - q$ degrees of freedom. Since χ^2_d is a positive random variable, for any $a > 0$,

$$P(\chi_p > a) = P(\chi^2_q + \chi^2_d > a) > P(\chi^2_q > a).$$

b. For $k_1 > k_2$, $k_1 \mathrm{F}_{k_1,\nu} \sim (U+V)/(W/\nu)$, where U, V and W are independent and $U \sim \chi^2_{k_2}$, $V \sim \chi^2_{k_1-k_2}$ and $W \sim \chi^2_\nu$. For any $a > 0$, because $V/(W/\nu)$ is a positive random variable, we have

$$P(k_1 F_{k_1,\nu} > a) = P((U+V)/(W/\nu) > a) > P(U/(W/\nu) > a) = P(k_2 F_{k_2,\nu} > a).$$

c. $\alpha = P(F_{k,\nu} > F_{\alpha,k,\nu}) = P(kF_{k,\nu} > kF_{\alpha,k,\nu})$. So, $kF_{\alpha,k,\nu}$ is the α cutoff point for the random variable $kF_{k,\nu}$. Because $kF_{k,\nu}$ is stochastically larger that $(k-1)F_{k-1,\nu}$, the α cutoff for $kF_{k,\nu}$ is larger than the α cutoff for $(k-1)F_{k-1,\nu}$, that is $kF_{\alpha,k,\nu} > (k-1)F_{\alpha,k-1,\nu}$.

5.20 a. The given integral is

$$\begin{aligned}
&\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2x/2} \nu\sqrt{x} \frac{1}{\Gamma(\nu/2)2^{\nu/2}} (\nu x)^{(\nu/2)-1} e^{-\nu x/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \frac{\nu^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}} \int_0^\infty e^{-t^2x/2} x^{((\nu+1)/2)-1} e^{-\nu x/2} dx
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}}\frac{\nu^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}}\int_0^\infty x^{((\nu+1)/2)-1}e^{-(\nu+t^2)x/2}dx \qquad \left(\begin{array}{c}\text{integrand is kernel of}\\ \text{gamma}((\nu+1)/2, 2/(\nu+t^2)\end{array}\right)\\
&= \frac{1}{\sqrt{2\pi}}\frac{\nu^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}}\Gamma((\nu+1)/2)\left(\frac{2}{\nu+t^2}\right)^{(\nu+1)/2}\\
&= \frac{1}{\sqrt{\nu\pi}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}},
\end{aligned}$$

the pdf of a t_ν distribution.

b. Differentiate both sides with respect to t to obtain

$$\nu f_F(\nu t) = \int_0^\infty y f_1(ty) f_\nu(y)dy,$$

where f_F is the F pdf. Now write out the two chi-squared pdfs and collect terms to get

$$\begin{aligned}
\nu f_F(\nu t) &= \frac{t^{-1/2}}{\Gamma(1/2)\Gamma(\nu/2)2^{(\nu+1)/2}}\int_0^\infty y^{(\nu-1)/2}e^{-(1+t)y/2}dy\\
&= \frac{t^{-1/2}}{\Gamma(1/2)\Gamma(\nu/2)2^{(\nu+1)/2}}\frac{\Gamma(\frac{\nu+1}{2})2^{(\nu+1)/2}}{(1+t)^{(\nu+1)/2}}.
\end{aligned}$$

Now define $y = \nu t$ to get

$$f_F(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\nu\Gamma(1/2)\Gamma(\nu/2)}\frac{(y/\nu)^{-1/2}}{(1+y/\nu)^{(\nu+1)/2}},$$

the pdf of an $F_{1,\nu}$.

c. Again differentiate both sides with respect to t, write out the chi-squared pdfs, and collect terms to obtain

$$(\nu/m)f_F((\nu/m)t) = \frac{t^{-m/2}}{\Gamma(m/2)\Gamma(\nu/2)2^{(\nu+m)/2}}\int_0^\infty y^{(m+\nu-2)/2}e^{-(1+t)y/2}dy.$$

Now, as before, integrate the gamma kernel, collect terms, and define $y = (\nu/m)t$ to get

$$f_F(y) = \frac{\Gamma(\frac{\nu+m}{2})}{\Gamma(m/2)\Gamma(\nu/2)}\left(\frac{m}{\nu}\right)^{m/2}\frac{y^{m/2-1}}{(1+(m/\nu)y)^{(\nu+m)/2}},$$

the pdf of an $F_{m,\nu}$.

5.21 Let m denote the median. Then, for general n we have

$$\begin{aligned}
P\left(\max(X_1,\ldots,X_n) > m\right) &= 1 - P\left(X_i \le m \text{ for } i = 1,2,\ldots,n\right)\\
&= 1 - \left[P\left(X_1 \le m\right)\right]^n \quad = \quad 1 - \left(\frac{1}{2}\right)^n.
\end{aligned}$$

5.22 Calculating the cdf of Z^2, we obtain

$$\begin{aligned}
F_{Z^2}(z) &= P((\min(X,Y))^2 \le z) = P(-z \le \min(X,Y) \le \sqrt{z})\\
&= P(\min(X,Y) \le \sqrt{z}) - P(\min(X,Y) \le -\sqrt{z})\\
&= [1 - P(\min(X,Y) > \sqrt{z})] - [1 - P(\min(X,Y) > -\sqrt{z})]\\
&= P(\min(X,Y) > -\sqrt{z}) - P(\min(X,Y) > \sqrt{z})\\
&= P(X > -\sqrt{z})P(Y > -\sqrt{z}) - P(X > \sqrt{z})P(Y > \sqrt{z}),
\end{aligned}$$

where we use the independence of X and Y. Since X and Y are identically distributed, $P(X > a) = P(Y > a) = 1 - F_X(a)$, so

$$F_{Z^2}(z) = (1 - F_X(-\sqrt{z}))^2 - (1 - F_X(\sqrt{z}))^2 = 1 - 2F_X(-\sqrt{z}),$$

since $1 - F_X(\sqrt{z}) = F_X(-\sqrt{z})$. Differentiating and substituting gives

$$f_{Z^2}(z) = \frac{d}{dz}F_{Z^2}(z) = f_X(-\sqrt{z})\frac{1}{\sqrt{z}} = \frac{1}{\sqrt{2\pi}}e^{-z/2}z^{-1/2},$$

the pdf of a χ_1^2 random variable. Alternatively,

$$\begin{aligned}
P(Z^2 \le z) &= P\left([\min(X,Y)]^2 \le z\right) \\
&= P(-\sqrt{z} \le \min(X,Y) \le \sqrt{z}) \\
&= P(-\sqrt{z} \le X \le \sqrt{z}, X \le Y) + P(-\sqrt{z} \le Y \le \sqrt{z}, Y \le X) \\
&= P(-\sqrt{z} \le X \le \sqrt{z} | X \le Y)P(X \le Y) \\
&\quad +P(-\sqrt{z} \le Y \le \sqrt{z} | Y \le X)P(Y \le X) \\
&= \frac{1}{2}P(-\sqrt{z} \le X \le \sqrt{z}) + \frac{1}{2}P(-\sqrt{z} \le Y \le \sqrt{z}),
\end{aligned}$$

using the facts that X and Y are independent, and $P(Y \le X) = P(X \le Y) = \frac{1}{2}$. Moreover, since X and Y are identically distributed

$$P(Z^2 \le z) = P(-\sqrt{z} \le X \le \sqrt{z})$$

and

$$\begin{aligned}
f_{Z^2}(z) &= \frac{d}{dz}P(-\sqrt{z} \le X \le \sqrt{z}) = \frac{1}{\sqrt{2\pi}}(e^{-z/2}\frac{1}{2}z^{-1/2} + e^{-z/2}\frac{1}{2}z^{-1/2}) \\
&= \frac{1}{\sqrt{2\pi}}z^{-1/2}e^{-z/2},
\end{aligned}$$

the pdf of a χ_1^2.

5.23

$$\begin{aligned}
P(Z > z) &= \sum_{x=1}^{\infty} P(Z > z|x)P(X = x) = \sum_{x=1}^{\infty} P(U_1 > z, \ldots, U_x > z|x)P(X = x) \\
&= \sum_{x=1}^{\infty}\prod_{i=1}^{x} P(U_i > z)P(X = x) \qquad \text{(by independence of the } U_i\text{'s)} \\
&= \sum_{x=1}^{\infty} P(U_i > z)^x P(X = x) = \sum_{x=1}^{\infty}(1-z)^x \frac{1}{(e-1)x!} \\
&= \frac{1}{(e-1)}\sum_{x=1}^{\infty}\frac{(1-z)^x}{x!} = \frac{e^{1-z}-1}{e-1} \qquad 0 < z < 1.
\end{aligned}$$

5.24 Use $f_X(x) = 1/\theta$, $F_X(x) = x/\theta$, $0 < x < \theta$. Let $Y = X_{(n)}$, $Z = X_{(1)}$. Then, from Theorem 5.4.6,

$$f_{Z,Y}(z,y) = \frac{n!}{0!(n-2)!0!}\frac{1}{\theta}\frac{1}{\theta}\left(\frac{z}{\theta}\right)^0\left(\frac{y-z}{\theta}\right)^{n-2}\left(1-\frac{y}{\theta}\right)^0 = \frac{n(n-1)}{\theta^n}(y-z)^{n-2},\ 0 < z < y < \theta.$$

Now let $W = Z/Y$, $Q = Y$. Then $Y = Q$, $Z = WQ$, and $|J| = q$. Therefore

$$f_{W,Q}(w,q) = \frac{n(n-1)}{\theta^n}(q-wq)^{n-2}q = \frac{n(n-1)}{\theta^n}(1-w)^{n-2}q^{n-1},\ 0<w<1, 0<q<\theta.$$

The joint pdf factors into functions of w and q, and, hence, W and Q are independent.

5.25 The joint pdf of $X_{(1)},\ldots,X_{(n)}$ is

$$f(u_1,\ldots,u_n) = \frac{n!a^n}{\theta^{an}}u_1^{a-1}\cdots u_n^{a-1}, \quad 0<u_1<\cdots<u_n<\theta.$$

Make the one-to-one transformation to $Y_1 = X_{(1)}/X_{(2)},\ldots,Y_{n-1} = X_{(n-1)}/X_{(n)}, Y_n = X_{(n)}$. The Jacobian is $J = y_2y_3^2\cdots y_n^{n-1}$. So the joint pdf of $Y_1,\ldots,Y_n$ is

$$\begin{aligned} f(y_1,\ldots,y_n) &= \frac{n!a^n}{\theta^{an}}(y_1\cdots y_n)^{a-1}(y_2\cdots y_n)^{a-1}\cdots(y_n)^{a-1}(y_2y_3^2\cdots y_n^{n-1}) \\ &= \frac{n!a^n}{\theta^{an}}y_1^{a-1}y_2^{2a-1}\cdots y_n^{na-1}, \quad 0<y_i<1; i=1,\ldots,n-1, \quad 0<y_n<\theta. \end{aligned}$$

We see that $f(y_1,\ldots,y_n)$ factors so $Y_1,\ldots,Y_n$ are mutually independent. To get the pdf of Y_1, integrate out the other variables and obtain that $f_{Y_1}(y_1) = c_1y_1^{a-1}$, $0<y_1<1$, for some constant c_1. To have this pdf integrate to 1, it must be that $c_1 = a$. Thus $f_{Y_1}(y_1) = ay_1^{a-1}$, $0<y_1<1$. Similarly, for $i = 2,\ldots,n-1$, we obtain $f_{Y_i}(y_i) = iay_i^{ia-1}, 0<y_i<1$. From Theorem 5.4.4, the pdf of Y_n is $f_{Y_n}(y_n) = \frac{na}{\theta^{na}}y_n^{na-1}$, $0<y_n<\theta$. It can be checked that the product of these marginal pdfs is the joint pdf given above.

5.27 a. $f_{X_{(i)}|X_{(j)}}(u|v) = f_{X_{(i)},X_{(j)}}(u,v)/f_{X_{(j)}}(v)$. Consider two cases, depending on which of i or j is greater. Using the formulas from Theorems 5.4.4 and 5.4.6, and after cancellation, we obtain the following.

(i) If $i<j$,

$$\begin{aligned} f_{X_{(i)}|X_{(j)}}(u|v) &= \frac{(j-1)!}{(i-1)!(j-1-i)!}f_X(u)F_X^{i-1}(u)[F_X(v)-F_X(u)]^{j-i-1}F_X^{1-j}(v) \\ &= \frac{(j-1)!}{(i-1)!(j-1-i)!}\frac{f_X(u)}{F_X(v)}\left[\frac{F_X(u)}{F_X(v)}\right]^{i-1}\left[1-\frac{F_X(u)}{F_X(v)}\right]^{j-i-1}, \quad u<v. \end{aligned}$$

Note this interpretation. This is the pdf of the ith order statistic from a sample of size $j-1$, from a population with pdf given by the truncated distribution, $f(u) = f_X(u)/F_X(v)$, $u<v$.

(ii) If $j<i$ and $u>v$,

$$\begin{aligned} &f_{X_{(i)}|X_{(j)}}(u|v) \\ &= \frac{(n-j)!}{(n-1)!(i-1-j)!}f_X(u)\left[1-F_X(u)\right]^{n-i}\left[F_X(u)-F_X(v)\right]^{i-1-j}\left[1-F_X(v)\right]^{j-n} \\ &= \frac{(n-j)!}{(i-j-1)!(n-i)!}\frac{f_X(u)}{1-F_X(v)}\left[\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{i-j-1}\left[1-\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{n-i}. \end{aligned}$$

This is the pdf of the $(i-j)$th order statistic from a sample of size $n-j$, from a population with pdf given by the truncated distribution, $f(u) = f_X(u)/(1-F_X(v))$, $u>v$.

b. From Example 5.4.7,

$$f_{V|R}(v|r) = \frac{n(n-1)r^{n-2}/a^n}{n(n-1)r^{n-2}(a-r)/a^n} = \frac{1}{a-r}, \quad r/2<v<a-r/2.$$

5.29 Let X_i = weight of ith booklet in package. The X_is are iid with $EX_i = 1$ and $\text{Var}X_i = .05^2$. We want to approximate $P\left(\sum_{i=1}^{100} X_i > 100.4\right) = P\left(\sum_{i=1}^{100} X_i/100 > 1.004\right) = P(\bar{X} > 1.004)$. By the CLT, $P(\bar{X} > 1.004) \approx P(Z > (1.004 - 1)/(.05/10)) = P(Z > .8) = .2119$.

5.30 From the CLT we have, approximately, $\bar{X}_1 \sim \text{n}(\mu, \sigma^2/n)$, $\bar{X}_2 \sim \text{n}(\mu, \sigma^2/n)$. Since $\bar{X}_1$ and $\bar{X}_2$ are independent, $\bar{X}_1 - \bar{X}_2 \sim \text{n}(0, 2\sigma^2/n)$. Thus, we want

$$\begin{aligned} .99 &\approx P\left(\left|\bar{X}_1 - \bar{X}_2\right| < \sigma/5\right) \\ &= P\left(\frac{-\sigma/5}{\sigma/\sqrt{n/2}} < \frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n/2}} < \frac{\sigma/5}{\sigma/\sqrt{n/2}}\right) \\ &\approx P\left(-\frac{1}{5}\sqrt{\frac{n}{2}} < Z < \frac{1}{5}\sqrt{\frac{n}{2}}\right), \end{aligned}$$

where $Z \sim \text{n}(0,1)$. Thus we need $P(Z \geq \sqrt{n}/5(\sqrt{2})) \approx .005$. From Table 1, $\sqrt{n}/5\sqrt{2} = 2.576$, which implies $n = 50(2.576)^2 \approx 332$.

5.31 We know that $\sigma^2_{\bar{X}} = 9/100$. Use Chebyshev's Inequality to get

$$P\left(-3k/10 < \bar{X} - \mu < 3k/10\right) \geq 1 - 1/k^2.$$

We need $1 - 1/k^2 \geq .9$ which implies $k \geq \sqrt{10} = 3.16$ and $3k/10 = .9487$. Thus

$$P(-.9487 < \bar{X} - \mu < .9487) \geq .9$$

by Chebychev's Inequality. Using the CLT, $\bar{X}$ is approximately $\text{n}\left(\mu, \sigma^2_{\bar{X}}\right)$ with $\sigma_{\bar{X}} = \sqrt{.09} = .3$ and $(\bar{X} - \mu)/.3 \sim \text{n}(0,1)$. Thus

$$.9 = P\left(-1.645 < \frac{\bar{X} - \mu}{.3} < 1.645\right) = P(-.4935 < \bar{X} - \mu < .4935).$$

Thus, we again see the conservativeness of Chebychev's Inequality, yielding bounds on $\bar{X} - \mu$ that are almost twice as big as the normal approximation. Moreover, with a sample of size 100, $\bar{X}$ is probably very close to normally distributed, even if the underlying X distribution is not close to normal.

5.32 a. For any $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sqrt{X_n} - \sqrt{a}\right| > \epsilon\right) &= P\left(\left|\sqrt{X_n} - \sqrt{a}\right|\left|\sqrt{X_n} + \sqrt{a}\right| > \epsilon\left|\sqrt{X_n} + \sqrt{a}\right|\right) \\ &= P\left(|X_n - a| > \epsilon\left|\sqrt{X_n} + \sqrt{a}\right|\right) \\ &\leq P\left(|X_n - a| > \epsilon\sqrt{a}\right) \to 0, \end{aligned}$$

as $n \to \infty$, since $X_n \to$ a in probability. Thus $\sqrt{X_n} \to \sqrt{a}$ in probability.

b. For any $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{a}{X_n} - 1\right| \leq \epsilon\right) &= P\left(\frac{a}{1+\epsilon} \leq X_n \leq \frac{a}{1-\epsilon}\right) \\ &= P\left(a - \frac{a\epsilon}{1+\epsilon} \leq X_n \leq a + \frac{a\epsilon}{1-\epsilon}\right) \\ &\geq P\left(a - \frac{a\epsilon}{1+\epsilon} \leq X_n \leq a + \frac{a\epsilon}{1+\epsilon}\right) \qquad \left(a + \frac{a\epsilon}{1+\epsilon} < a + \frac{a\epsilon}{1-\epsilon}\right) \\ &= P\left(|X_n - a| \leq \frac{a\epsilon}{1+\epsilon}\right) \to 1, \end{aligned}$$

as $n \to \infty$, since $X_n \to$ a in probability. Thus $a/X_n \to 1$ in probability.

c. $S_n^2 \to \sigma^2$ in probability. By a), $S_n = \sqrt{S_n^2} \to \sqrt{\sigma^2} = \sigma$ in probability. By b), $\sigma/S_n \to 1$ in probability.

5.33 For all $\epsilon > 0$ there exist N such that if $n > N$, then $P(X_n + Y_n > c) > 1 - \epsilon$. Choose N_1 such that $P(X_n > -m) > 1 - \epsilon/2$ and N_2 such that $P(Y_n > c + m) > 1 - \epsilon/2$. Then

$$P(X_n + Y_n > c) \geq P(X_n > -m, +Y_n > c + m) \geq P(X_n > -m) + P(Y_n > c + m) - 1 = 1 - \epsilon.$$

5.34 Using $\mathrm{E}\bar{X}_n = \mu$ and $\mathrm{Var}\bar{X}_n = \sigma^2/n$, we obtain

$$\mathrm{E}\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}}{\sigma}\mathrm{E}(\bar{X}_n - \mu) = \frac{\sqrt{n}}{\sigma}(\mu - \mu) = 0.$$

$$\mathrm{Var}\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{n}{\sigma^2}\mathrm{Var}(\bar{X}_n - \mu) = \frac{n}{\sigma^2}\mathrm{Var}\bar{X} = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1.$$

5.35 a. $X_i \sim$ exponential(1). $\mu_X = 1$, $\mathrm{Var}X = 1$. From the CLT, $\bar{X}_n$ is approximately $\mathrm{n}(1, 1/n)$. So

$$\frac{\bar{X}_n - 1}{1/\sqrt{n}} \to Z \sim \mathrm{n}(0,1) \quad \text{and} \quad P\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \leq x\right) \to P(Z \leq x).$$

b.

$$\frac{d}{dx}P(Z \leq x) = \frac{d}{dx}F_Z(x) = f_Z(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

$$\begin{aligned} &\frac{d}{dx}P\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \leq x\right) \\ &= \frac{d}{dx}\left(\sum_{i=1}^n X_i \leq x\sqrt{n} + n\right) \qquad \left(W = \sum_{i=1}^n X_i \sim \mathrm{gamma}(n,1)\right) \\ &= \frac{d}{dx}F_W(x\sqrt{n} + n) = f_W(x\sqrt{n} + n) \cdot \sqrt{n} = \frac{1}{\Gamma(n)}(x\sqrt{n} + n)^{n-1}e^{-(x\sqrt{n}+n)}\sqrt{n}. \end{aligned}$$

Therefore, $(1/\Gamma(n))(x\sqrt{n} + n)^{n-1}e^{-(x\sqrt{n}+n)}\sqrt{n} \approx \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ as $n \to \infty$. Substituting $x = 0$ yields $n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}$.

5.37 a. For the exact calculations, use the fact that V_n is itself distributed negative binomial$(10r, p)$. The results are summarized in the following table. Note that the recursion relation of problem 3.48 can be used to simplify calculations.

	$P(V_n = v)$		
	(a)	(b)	(c)
v	Exact	Normal App.	Normal w/cont.
0	.0008	.0071	.0056
1	.0048	.0083	.0113
2	.0151	.0147	.0201
3	.0332	.0258	.0263
4	.0572	.0392	.0549
5	.0824	.0588	.0664
6	.1030	.0788	.0882
7	.1148	.0937	.1007
8	.1162	.1100	.1137
9	.1085	.1114	.1144
10	.0944	.1113	.1024

b. Using the normal approximation, we have $\mu_v = r(1-p)/p = 20(.3)/.7 = 8.57$ and

$$\sigma_v = \sqrt{r(1-p)/p^2} = \sqrt{(20)(.3)/.49} = 3.5.$$

Then,

$$P(V_n = 0) = 1 - P(V_n \geq 1) = 1 - P\left(\frac{V_n - 8.57}{3.5} \geq \frac{1 - 8.57}{3.5}\right) = 1 - P(Z \geq -2.16) = .0154.$$

Another way to approximate this probability is

$$P(V_n = 0) = P(V_n \leq 0) = P\left(\frac{V - 8.57}{3.5} \leq \frac{0 - 8.57}{3.5}\right) = P(Z \leq -2.45) = .0071.$$

Continuing in this way we have $P(V = 1) = P(V \leq 1) - P(V \leq 0) = .0154 - .0071 = .0083$, etc.

c. With the continuity correction, compute $P(V = k)$ by $P\left(\frac{(k-.5)-8.57}{3.5} \leq Z \leq \frac{(k+.5)-8.57}{3.5}\right)$, so $P(V = 0) = P(-9.07/3.5 \leq Z \leq -8.07/3.5) = .0104 - .0048 = .0056$, etc. Notice that the continuity correction gives some improvement over the uncorrected normal approximation.

5.39 a. If h is continuous given $\epsilon > 0$ there exits δ such that $|h(x_n) - h(x)| < \epsilon$ for $|x_n - x| < \delta$. Since $X_1, \ldots, X_n$ converges in probability to the random variable X, then $\lim_{n\to\infty} P(|X_n - X| < \delta) = 1$. Thus $\lim_{n\to\infty} P(|h(X_n) - h(X)| < \epsilon) = 1$.

b. Define the subsequence $X_j(s) = s + I_{[a,b]}(s)$ such that in $I_{[a,b]}$, a is always 0, i.e, the subsequence $X_1, X_2, X_4, X_7, \ldots$. For this subsequence

$$X_j(s) \to \begin{cases} s & \text{if } s > 0 \\ s+1 & \text{if } s = 0. \end{cases}$$

5.41 a. Let $\epsilon = |x - \mu|$.

(i) For $x - \mu \geq 0$

$$\begin{aligned} P(|X_n - \mu| > \epsilon) &= P(|X_n - \mu| > x - \mu) \\ &= P(X_n - \mu < -(x - \mu)) + P(X_n - \mu > x - \mu) \\ &\geq P(X_n - \mu > x - \mu) \\ &= P(X_n > x) \quad = \quad 1 - P(X_n \leq x). \end{aligned}$$

Therefore, $0 = \lim_{n\to\infty} P(|X_n - \mu| > \epsilon) \geq \lim_{n\to\infty} 1 - P(X_n \leq x)$. Thus $\lim_{n\to\infty} P(X_n \leq x) \geq 1$.

(ii) For $x - \mu < 0$

$$\begin{aligned} P(|X_n - \mu| > \epsilon) &= P(|X_n - \mu| > -(x - \mu)) \\ &= P(X_n - \mu < x - \mu) + P(X_n - \mu > -(x - \mu)) \\ &\geq P(X_n - \mu < x - \mu) \\ &= P(X_n \leq x). \end{aligned}$$

Therefore, $0 = \lim_{n\to\infty} P(|X_n - \mu| > \epsilon) \geq \lim_{n\to\infty} P(X_n \leq x)$.

By (i) and (ii) the results follows.

b. For every $\epsilon > 0$,

$$\begin{aligned} P(|X_n - \mu| > \epsilon) &\leq P(X_n - \mu < -\epsilon) + P(X_n - \mu > \epsilon) \\ &= P(X_n < \mu - \epsilon) + 1 - P(X_n \leq \mu + \epsilon) \quad \to \quad 0 \ \text{ as } n \to \infty. \end{aligned}$$

5.43 a. $P(|Y_n - \theta| < \epsilon) = P\left(\left|\sqrt{(n)}(Y_n - \theta)\right| < \sqrt{(n)}\epsilon\right)$. Therefore,

$$\lim_{n\to\infty} P(|Y_n - \theta| < \epsilon) = \lim_{n\to\infty} P\left(\left|\sqrt{(n)}(Y_n - \theta)\right| < \sqrt{(n)}\epsilon\right) = P(|Z| < \infty) = 1,$$

where $Z \sim \mathrm{n}(0, \sigma^2)$. Thus $Y_n \to \theta$ in probability.

b. By Slutsky's Theorem (a), $g'(\theta)\sqrt{n}(Y_n - \theta) \to g'(\theta)X$ where $X \sim \mathrm{n}(0, \sigma^2)$. Therefore $\sqrt{n}[g(Y_n) - g(\theta)] = g'(\theta)\sqrt{n}(Y_n - \theta) \to \mathrm{n}(0, \sigma^2[g'(\theta)]^2)$.

5.45 We do part (a), the other parts are similar. Using Mathematica, the exact calculation is

```
In[120]:=
f1[x_]=PDF[GammaDistribution[4,25],x]
p1=Integrate[f1[x],{x,100,\[Infinity]}]//N
1-CDF[BinomialDistribution[300,p1],149]

Out[120]=
e^(-x/25) x^3/2343750

Out[121]=
0.43347

Out[122]=
0.0119389.
```

The answer can also be simulated in Mathematica or in R. Here is the R code for simulating the same probability

```
p1<-mean(rgamma(10000,4,scale=25)>100)
mean(rbinom(10000, 300, p1)>149)
```

In each case 10,000 random variables were simulated. We obtained $p1 = 0.438$ and a binomial probability of 0.0108.

5.47 a. $-2\log(U_j) \sim \text{exponential}(2) \sim \chi^2_2$. Thus Y is the sum of ν independent χ^2_2 random variables. By Lemma 5.3.2(b), $Y \sim \chi^2_{2\nu}$.

b. $\beta\log(U_j) \sim \text{exponential}(2) \sim \text{gamma}(1, \beta)$. Thus Y is the sum of independent gamma random variables. By Example 4.6.8, $Y \sim \text{gamma}(a, \beta)$

c. Let $V = \sum_{j=1}^{a}\log(U_j) \sim \text{gamma}(a, 1)$. Similarly $W = \sum_{j=1}^{b}\log(U_j) \sim \text{gamma}(b, 1)$. By Exercise 4.24, $\frac{V}{V+W} \sim \text{beta}(a, b)$.

5.49 a. See Example 2.1.4.

b. $X = g(U) = -\log\frac{1-U}{U}$. Then $g^{-1}(x) = \frac{1}{1+e^{-y}}$. Thus

$$f_X(x) = 1 \times \left|\frac{e^{-y}}{(1+e^{-y})^2}\right| = \frac{e^{-y}}{(1+e^{-y})^2} \quad -\infty < y < \infty,$$

which is the density of a logistic$(0, 1)$ random variable.

c. Let $Y \sim \text{logistic}(\mu, \beta)$ then $f_Y(y) = \frac{1}{\beta}f_Z(\frac{-(y-\mu)}{\beta})$ where f_Z is the density of a logistic$(0, 1)$. Then $Y = \beta Z + \mu$. To generate a logistic(μ, β) random variable generate (i) generate $U \sim$ uniform$(0, 1)$, (ii) Set $Y = \beta\log\frac{U}{1-U} + \mu$.

5.51 a. For $U_i \sim \text{uniform}(0, 1)$, $\mathrm{E}U_i = 1/2$, $\mathrm{Var}U_i = 1/12$. Then

$$X = \sum_{i=1}^{12} U_i - 6 = 12\bar{U} - 6 = \sqrt{12}\left(\frac{\bar{U}-1/2}{1/\sqrt{12}}\right)$$

is in the form $\sqrt{n}\left((\bar{U}-\mathrm{E}U)/\sigma\right)$ with $n=12$, so X is approximately $\mathrm{n}(0,1)$ by the Central Limit Theorem.

b. The approximation does not have the same range as $Z \sim \mathrm{n}(0,1)$ where $-\infty < Z < +\infty$, since $-6 < X < 6$.

c.

$$\mathrm{E}X = \mathrm{E}\left(\sum_{i=1}^{12} U_i - 6\right) = \sum_{i=1}^{12} \mathrm{E}U_i - 6 = \left(\sum_{i=1}^{12} \frac{1}{2}\right) - 6 = 6 - 6 = 0.$$

$$\mathrm{Var}X = \mathrm{Var}\left(\sum_{i=1}^{12} U_i - 6\right) = \mathrm{Var}\sum_{i=1}^{12} U_i = 12\mathrm{Var}U_1 = 1$$

$\mathrm{E}X^3 = 0$ since X is symmetric about 0. (In fact, all odd moments of X are 0.) Thus, the first three moments of X all agree with the first three moments of a $\mathrm{n}(0,1)$. The fourth moment is not easy to get, one way to do it is to get the mgf of X. Since $\mathrm{E}e^{tU} = (e^t - 1)/t$,

$$\mathrm{E}\left[e^{t\left(\sum_{i=1}^{12} U_i - 6\right)}\right] = e^{-6t}\left(\frac{e^t - 1}{t}\right)^{12} = \left(\frac{e^{t/2} - e^{-t/2}}{t}\right)^{12}.$$

Computing the fourth derivative and evaluating it at $t = 0$ gives us $\mathrm{E}X^4$. This is a lengthy calculation. The answer is $\mathrm{E}X^4 = 29/10$, slightly smaller than $\mathrm{E}Z^4 = 3$, where $Z \sim \mathrm{n}(0,1)$.

5.53 The R code is the following:

a.
```
obs <- rbinom(1000,8,2/3)
meanobs <- mean(obs)
variance <- var(obs)
hist(obs)
Output:
> meanobs
[1] 5.231
> variance
[1] 1.707346
```

b.
```
obs<- rhyper(1000,8,2,4)
meanobs <- mean(obs)
variance <- var(obs)
hist(obs)
Output:
> meanobs
[1] 3.169
> variance
[1] 0.4488879
```

c.
```
obs <- rnbinom(1000,5,1/3)
meanobs <- mean(obs)
variance <- var(obs)
hist(obs)
Output:
> meanobs
[1] 10.308
> variance
[1] 29.51665
```

5.55 Let X denote the number of comparisons. Then

$$\begin{aligned}
\mathrm{E}X &= \sum_{k=0}^{\infty} P(X>k) \quad = \quad 1+\sum_{k=1}^{\infty} P(U > F_y(y_{k-1})) \\
&= 1+\sum_{k=1}^{\infty}(1-F_y(y_{k-1})) \quad = \quad 1+\sum_{k=0}^{\infty}(1-F_y(y_i)) \quad = \quad 1+\mathrm{E}Y
\end{aligned}$$

5.57 a. $\mathrm{Cov}(Y_1,Y_2) = \mathrm{Cov}(X_1+X_3, X_2+X_3) = \mathrm{Cov}(X_3,X_3) = \lambda_3$ since X_1, X_2 and X_3 are independent.

b.

$$Z_i = \begin{cases} 1 & \text{if } X_i = X_3 = 0 \\ 0 & \text{otherwise} \end{cases}$$

$p_i = P(Z_i = 0) = P(Y_i = 0) = P(X_i = 0, X_3 = 0) = e^{-(\lambda_i+\lambda_3)}$. Therefore Z_i are Bernoulli(p_i) with $\mathrm{E}[Z_i] = p_i$, $\mathrm{Var}(Z_i) = p_i(1-p_i)$ and

$$\begin{aligned}
\mathrm{E}[Z_1Z_2] &= P(Z_1=1,Z_2=1) \quad = \quad P(Y_1=0,Y_2=0) \\
&= P(X_1+X_3=0, X_2+X_3=0) \quad = \quad P(X_1=0)P(X_2=0)P(X_3=0) \\
&= e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Cov}(Z_1,Z_2) &= \mathrm{E}[Z_1Z_2] - \mathrm{E}[Z_1]\mathrm{E}[Z_2] \\
&= e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3} - e^{-(\lambda_i+\lambda_3)}e^{-(\lambda_2+\lambda_3)} \quad = \quad e^{-(\lambda_i+\lambda_3)}e^{-(\lambda_2+\lambda_3)}(e^{\lambda_3}-1) \\
&= p_1p_2(e^{\lambda_3}-1).
\end{aligned}$$

Thus $\mathrm{Corr}(Z_1,Z_2) = \frac{p_1p_2(e^{\lambda_3}-1)}{\sqrt{p_1(1-p_1)}\sqrt{p_2(1-p_2)}}$.

c. $\mathrm{E}[Z_1Z_2] \le p_i$, therefore

$$\begin{aligned}
\mathrm{Cov}(Z_1,Z_2) &= \mathrm{E}[Z_1Z_2] - \mathrm{E}[Z_1]\mathrm{E}[Z_2] \le p_1 - p_1p_2 \quad = \quad p_1(1-p_2), \quad \text{and} \\
\mathrm{Cov}(Z_1,Z_2) &\le p_2(1-p_1).
\end{aligned}$$

Therefore,

$$\mathrm{Corr}(Z_1,Z_2) \le \frac{p_1(1-p_2)}{\sqrt{p_1(1-p_1)}\sqrt{p_2(1-p_2)}} = \frac{\sqrt{p_1(1-p_2)}}{\sqrt{p_2(1-p_1)}}$$

and

$$\mathrm{Corr}(Z_1,Z_2) \le \frac{p_2(1-p_1)}{\sqrt{p_1(1-p_1)}\sqrt{p_2(1-p_2)}} = \frac{\sqrt{p_2(1-p_1)}}{\sqrt{p_1(1-p_2)}}$$

which implies the result.

5.59

$$\begin{aligned}
P(Y \le y) &= P(V \le y | U < \frac{1}{c}f_Y(V)) \quad = \quad \frac{P(V \le y, U < \frac{1}{c}f_Y(V))}{P(U < \frac{1}{c}f_Y(V))} \\
&= \frac{\int_0^y \int_0^{\frac{1}{c}f_Y(v)} dudv}{\frac{1}{c}} \quad = \quad \frac{\frac{1}{c}\int_0^y f_Y(v)dv}{\frac{1}{c}} \quad = \quad \int_0^y f_Y(v)dv
\end{aligned}$$

5.61 a. $M = \sup_y \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}}{\frac{\Gamma([a]+[b])}{\Gamma([a])\Gamma([b])}y^{[a]-1}(1-y)^{[b]-1}} < \infty$, since $a-[a]>0$ and $b-[b]>0$ and $y \in (0,1)$.

b. $M = \sup_y \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}}{\frac{\Gamma([a]+b)}{\Gamma([a])\Gamma(b)} y^{[a]-1}(1-y)^{b-1}} < \infty$, since $a-[a] > 0$ and $y \in (0,1)$.

c. $M = \sup_y \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}}{\frac{\Gamma([a]+1+\beta)}{\Gamma([a]+1)\Gamma(b')} y^{[a]+1-1}(1-y)^{b'-1}} < \infty$, since $a-[a]-1 < 0$ and $y \in (0,1)$. $b-b' > 0$ when $b' = [b]$ and will be equal to zero when $b' = b$, thus it does not affect the result.

d. Let $f(y) = y^\alpha(1-y)^\beta$. Then

$$\frac{df(y)}{dy} = \alpha y^{\alpha-1}(1-y)^\beta - y^\alpha\beta(1-y)^{\beta-1} = y^{\alpha-1}(1-y)^{\beta-1}[\alpha(1-y)+\beta y]$$

which is maximize at $y = \frac{\alpha}{\alpha+\beta}$. Therefore for, $\alpha = a-a'$ and $\beta = b-b'$

$$M = \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}{\frac{\Gamma(a'+b')}{\Gamma(a')\Gamma(b')}} \left(\frac{a-a'}{a-a'+b-b'}\right)^{a-a'} \left(\frac{b-b'}{a-a'+b-b'}\right)^{b-b'}.$$

We need to minimize M in a' and b'. First consider $\left(\frac{a-a'}{a-a'+b-b'}\right)^{a-a'} \left(\frac{b-b'}{a-a'+b-b'}\right)^{b-b'}$. Let $c = \alpha+\beta$, then this term becomes $\left(\frac{\alpha}{c}\right)^\alpha \left(\frac{c-\alpha}{c}\right)^{c-\alpha}$. This term is maximize at $\frac{\alpha}{c} = \frac{1}{2}$, this is at $\alpha = \frac{1}{2}c$. Then $M = (\frac{1}{2})^{(a-a'+b-b')} \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}{\frac{\Gamma(a'+b')}{\Gamma(a')\Gamma(b')}}$. Note that the minimum that M could be is one, which it is attain when $a = a'$ and $b = b'$. Otherwise the minimum will occur when $a-a'$ and $b-b'$ are minimum but greater or equal than zero, this is when $a' = [a]$ and $b' = [b]$ or $a' = a$ and $b' = [b]$ or $a' = [a]$ and $b' = b$.

5.63 $M = \sup_y \frac{\frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}}{\frac{1}{2\lambda} e^{\frac{-|y|}{\lambda}}}$. Let $f(y) = \frac{-y^2}{2} + \frac{|y|}{\lambda}$. Then $f(y)$ is maximize at $y = \frac{1}{\lambda}$ when $y \geq 0$ and at $y = \frac{-1}{\lambda}$ when $y < 0$. Therefore in both cases $M = \frac{\frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2\lambda^2}}}{\frac{1}{2\lambda} e^{\frac{-1}{\lambda^2}}}$. To minimize M let $M' = \lambda e^{\frac{1}{2\lambda^2}}$. Then $\frac{d \log M'}{d\lambda} = \frac{1}{\lambda} - \frac{1}{\lambda^3}$, therefore M is minimize at $\lambda = 1$ or $\lambda = -1$. Thus the value of λ that will optimize the algorithm is $\lambda = 1$.

5.65

$$\begin{aligned} P(X^* \leq x) &= \sum_{i=1}^m P(X^* \leq x|q_i)q_i = \sum_{i=1}^m I(Y_i \leq x)q_i = \frac{\frac{1}{m}\sum_{i=1}^m \frac{f(Y_i)}{g(Y_i)} I(Y_i \leq x)}{\frac{1}{m}\sum_{i=1}^m \frac{f(Y_i)}{g(Y_i)}} \\ &\underset{m\to\infty}{\longrightarrow} \frac{E_g \frac{f(Y)}{g(Y)} I(Y \leq x)}{E_g \frac{f(Y)}{g(Y)}} = \frac{\int_{-\infty}^x \frac{f(y)}{g(y)} g(y)dy}{\int_{-\infty}^\infty \frac{f(y)}{g(y)} g(y)dy} = \int_{-\infty}^x f(y)dy. \end{aligned}$$

5.67 An R code to generate the sample of size 100 from the specified distribution is shown for part c). The Metropolis Algorithm is used to generate 2000 variables. Among other options one can choose the 100 variables in positions 1001 to 1100 or the ones in positions $1010, 1020, ..., 2000$.

a. We want to generate $X = \sigma Z + \mu$ where $Z \sim$ Student's t with ν degrees of freedom. Therefore we first can generate a sample of size 100 from a Student's t distribution with ν degrees of freedom and then make the transformation to obtain the X's. Thus $f_Z(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\left(\frac{z^2}{\nu}\right)\right)^{(v+1)/2}}$. Let $V \sim \mathrm{n}(0, \frac{\nu}{\nu-2})$ since given ν we can set

$$\mathrm{E}V = \mathrm{E}Z = 0, \quad \text{and} \quad \mathrm{Var}(V) = \mathrm{Var}(Z) = \frac{\nu}{\nu-2}.$$

Now, follow the algorithm on page 254 and generate the sample $Z_1, Z_2 \ldots, Z_{100}$ and then calculate $X_i = \sigma Z_i + \mu$.

b. $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/2\sigma^2}}{x}$. Let $V \sim \text{gamma}(\alpha, \beta)$ where

$$\alpha = \frac{(e^{\mu+(\sigma^2/2)})^2}{e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}}, \quad \text{and} \quad \beta = \frac{e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}}{e^{\mu+(\sigma^2/2)}},$$

since given μ and σ^2 we can set

$$\mathrm{E}V = \alpha\beta = e^{\mu+(\sigma^2/2)} = \mathrm{E}X$$

and

$$\mathrm{Var}(V) = \alpha\beta^2 = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2} = \mathrm{Var}(X).$$

Now, follow the algorithm on page 254.

c. $f_X(x) = \frac{\alpha}{\beta}e^{\frac{-x^\alpha}{\beta}}x^{\alpha-1}$. Let $V \sim \text{exponential}(\beta)$. Now, follow the algorithm on page 254 where

$$\rho_i = \min\left\{\frac{V_i^{\alpha-1}}{Z_{i-1}^{\alpha-1}}e^{\frac{-V_i^\alpha+V_i-Z_{i-1}+Z_{i-1}^\alpha}{\beta}}, 1\right\}$$

An R code to generate a sample size of 100 from a Weibull(3,2) is:

```
#initialize a and b
b <- 2
a <- 3
Z <- rexp(1,1/b)
ranvars <- matrix(c(Z),byrow=T,ncol=1)
for( i in seq(2000))
{
U <- runif(1,min=0,max=1)
V <- rexp(1,1/b)
p <- pmin((V/Z)^(a-1)*exp((-V^a+V-Z+Z^a)/b),1)
if (U <= p)
  Z <- V
ranvars <- cbind(ranvars,Z)
}
#One option: choose elements in position 1001,1002,...,1100
to be the sample
vector.1 <- ranvars[1001:1100]
mean(vector.1)
var(vector.1)
#Another option: choose elements in position 1010,1020,...,2000
to be the sample
vector.2 <- ranvars[seq(1010,2000,10)]
mean(vector.2)
var(vector.2)
Output:
[1] 1.048035
[1] 0.1758335
[1] 1.130649
[1] 0.1778724
```

5.69 Let $w(v,z) = \frac{f_Y(v)f_V(z)}{f_V(v)f_Y(z)}$, and then $\rho(v,z) = \min\{w(v,z), 1\}$. We will show that

$$Z_i \sim f_Y \Rightarrow P(Z_{i+1} \leq a) = P(Y \leq a).$$

Write

$$P(Z_{i+1} \le a) = P(V_{i+1} \le a \text{ and } U_{i+1} \le \rho_{i+1}) + P(Z_i \le a \text{ and } U_{i+1} > \rho_{i+1}).$$

Since $Z_i \sim f_Y$, suppressing the unnecessary subscripts we can write

$$P(Z_{i+1} \le a) = P(V \le a \text{ and } U \le \rho(V,Y)) + P(Y \le a \text{ and } U > \rho(V,Y)).$$

Add and subtract $P(Y \le a \text{ and } U \le \rho(V,Y))$ to get

$$\begin{aligned} P(Z_{i+1} \le a) &= P(Y \le a) + P(V \le a \text{ and } U \le \rho(V,Y)) \\ &\quad - P(Y \le a \text{ and } U \le \rho(V,Y)). \end{aligned}$$

Thus we need to show that

$$P(V \le a \text{ and } U \le \rho(V,Y)) = P(Y \le a \text{ and } U \le \rho(V,Y)).$$

Write out the probability as

$$\begin{aligned} &P(V \le a \text{ and } U \le \rho(V,Y)) \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} \rho(v,y) f_Y(y) f_V(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \le 1)\left(\frac{f_Y(v)f_V(y)}{f_V(v)f_Y(y)}\right) f_Y(y)f_V(v)\,dy\,dv \\ &\quad + \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \ge 1) f_Y(y)f_V(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \le 1) f_Y(v)f_V(y)\,dy\,dv \\ &\quad + \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \ge 1) f_Y(y)f_V(v)\,dy\,dv. \end{aligned}$$

Now, notice that $w(v,y) = 1/w(y,v)$, and thus first term above can be written

$$\begin{aligned} &\int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \le 1) f_Y(v)f_V(y)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(y,v) > 1) f_Y(v)f_V(y)\,dy\,dv \\ &= P(Y \le a, \rho(V,Y) = 1, U \le \rho(V,Y)). \end{aligned}$$

The second term is

$$\begin{aligned} &\int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(v,y) \ge 1) f_Y(y)f_V(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(y,v) \le 1) f_Y(y)f_V(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(y,v) \le 1)\left(\frac{f_V(y)f_Y(v)}{f_V(y)f_Y(v)}\right) f_Y(y)f_V(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(y,v) \le 1)\left(\frac{f_Y(y)f_V(v)}{f_V(y)f_Y(v)}\right) f_V(y)f_Y(v)\,dy\,dv \\ &= \int_{-\infty}^{a}\int_{-\infty}^{\infty} I(w(y,v) \le 1) w(y,v) f_V(y)f_Y(v)\,dy\,dv \\ &= P(Y \le a, U \le \rho(V,Y), \rho(V,Y) \le 1). \end{aligned}$$

Putting it all together we have

$$\begin{aligned}
P(V \le a \text{ and } U \le \rho(V,Y)) &= P(Y \le a, \rho(V,Y)=1, U \le \rho(V,Y)) \\
&\quad + P(Y \le a, U \le \rho(V,Y), \rho(V,Y) \le 1) \\
&= P(Y \le a \text{ and } U \le \rho(V,Y)),
\end{aligned}$$

and hence

$$P(Z_{i+1} \le a) = P(Y \le a),$$

so f_Y is the stationary density.

Chapter 6

Principles of Data Reduction

6.1 By the Factorization Theorem, $|X|$ is sufficient because the pdf of X is

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-|x|^2/2\sigma^2} = g(\,|x|\,|\,\sigma^2) \cdot \underbrace{1}_{h(x)}.$$

6.2 By the Factorization Theorem, $T(X) = \min_i(X_i/i)$ is sufficient because the joint pdf is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} e^{i\theta - x_i} I_{(i\theta,+\infty)}(x_i) = \underbrace{e^{in\theta} I_{(\theta,+\infty)}(T(\mathbf{x}))}_{g(T(\mathbf{x})|\theta)} \cdot \underbrace{e^{-\Sigma_i x_i}}_{h(\mathbf{x})}.$$

Notice, we use the fact that $i > 0$, and the fact that all x_is $> i\theta$ if and only if $\min_i(x_i/i) > \theta$.

6.3 Let $x_{(1)} = \min_i x_i$. Then the joint pdf is

$$f(x_1, \ldots, x_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} e^{-(x_i-\mu)/\sigma} I_{(\mu,\infty)}(x_i) = \underbrace{\left(\frac{e^{\mu/\sigma}}{\sigma}\right)^n e^{-\Sigma_i x_i/\sigma} I_{(\mu,\infty)}(x_{(1)})}_{g\left(x_{(1)},\Sigma_i x_i \middle| \mu,\sigma\right)} \cdot \underbrace{1}_{h(\mathbf{x})}.$$

Thus, by the Factorization Theorem, $\left(X_{(1)}, \sum_i X_i\right)$ is a sufficient statistic for (μ, σ).

6.4 The joint pdf is

$$\prod_{j=1}^{n} \left\{ h(x_j) c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x_j)\right)\right\} = \underbrace{c(\boldsymbol{\theta})^n \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) \sum_{j=1}^{n} t_i(x_j)\right)}_{g(T(\mathbf{x})|\theta)} \cdot \underbrace{\prod_{j=1}^{n} h(x_j)}_{h(\mathbf{x})}.$$

By the Factorization Theorem, $\left(\sum_{j=1}^n t_1(X_j), \ldots, \sum_{j=1}^n t_k(X_j)\right)$ is a sufficient statistic for $\boldsymbol{\theta}$.

6.5 The sample density is given by

$$\begin{aligned}
\prod_{i=1}^{n} f(x_i|\theta) &= \prod_{i=1}^{n} \frac{1}{2i\theta} I\left(-i(\theta-1) \le x_i \le i(\theta+1)\right) \\
&= \left(\frac{1}{2\theta}\right)^n \left(\prod_{i=1}^{n} \frac{1}{i}\right) I\left(\min \frac{x_i}{i} \ge -(\theta-1)\right) I\left(\max \frac{x_i}{i} \le \theta+1\right).
\end{aligned}$$

Thus $(\min X_i/i, \max X_i/i)$ is sufficient for θ.

6.6 The joint pdf is given by

$$f(x_1,\dots,x_n|\alpha,\beta)=\prod_{i=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha}{x_i}^{\alpha-1}e^{-x_i/\beta}=\left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n\left(\prod_{i=1}^n x_i\right)^{\alpha-1}e^{-\Sigma_i x_i/\beta}.$$

By the Factorization Theorem, $(\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ is sufficient for (α,β).

6.7 Let $x_{(1)}=\min_i\{x_1,\dots,x_n\}$, $x_{(n)}=\max_i\{x_1,\dots,x_n\}$, $y_{(1)}=\min_i\{y_1,\dots,y_n\}$ and $y_{(n)}=\max_i\{y_1,\dots,y_n\}$. Then the joint pdf is

$$\begin{aligned} & f(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) \\ &= \prod_{i=1}^n \frac{1}{(\theta_3-\theta_1)(\theta_4-\theta_2)} I_{(\theta_1,\theta_3)}(x_i) I_{(\theta_2,\theta_4)}(y_i) \\ &= \underbrace{\left(\frac{1}{(\theta_3-\theta_1)(\theta_4-\theta_2)}\right)^n I_{(\theta_1,\infty)}(x_{(1)}) I_{(-\infty,\theta_3)}(x_{(n)}) I_{(\theta_2,\infty)}(y_{(1)}) I_{(-\infty,\theta_4)}(y_{(n)})}_{g(T(\mathbf{x})|\theta)} \cdot \underbrace{1}_{h(\mathbf{x})}. \end{aligned}$$

By the Factorization Theorem, $\left(X_{(1)},X_{(n)},Y_{(1)},Y_{(n)}\right)$ is sufficient for $(\theta_1,\theta_2,\theta_3,\theta_4)$.

6.9 Use Theorem 6.2.13.

a.

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}=\frac{(2\pi)^{-n/2}e^{-\Sigma_i(x_i-\theta)^2/2}}{(2\pi)^{-n/2}e^{-\Sigma_i(y_i-\theta)^2/2}}=\exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^n x_i^2-\sum_{i=1}^n y_i^2\right)+2\theta n(\bar y-\bar x)\right]\right\}.$$

This is constant as a function of θ if and only if $\bar y=\bar x$; therefore $\bar X$ is a minimal sufficient statistic for θ.

b. Note, for $X\sim$ location exponential(θ), the range depends on the parameter. Now

$$\begin{aligned}\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{\prod_{i=1}^n\left(e^{-(x_i-\theta)}I_{(\theta,\infty)}(x_i)\right)}{\prod_{i=1}^n\left(e^{-(y_i-\theta)}I_{(\theta,\infty)}(y_i)\right)} \\ &= \frac{e^{n\theta}e^{-\Sigma_i x_i}\prod_{i=1}^n I_{(\theta,\infty)}(x_i)}{e^{n\theta}e^{-\Sigma_i y_i}\prod_{i=1}^n I_{(\theta,\infty)}(y_i)} = \frac{e^{-\Sigma_i x_i}I_{(\theta,\infty)}(\min x_i)}{e^{-\Sigma_i y_i}I_{(\theta,\infty)}(\min y_i)}.\end{aligned}$$

To make the ratio independent of θ we need the ratio of indicator functions independent of θ. This will be the case if and only if $\min\{x_1,\dots,x_n\}=\min\{y_1,\dots,y_n\}$. So $T(\mathbf{X})=\min\{X_1,\dots,X_n\}$ is a minimal sufficient statistic.

c.

$$\begin{aligned}\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{e^{-\Sigma_i(x_i-\theta)}}{\prod_{i=1}^n\left(1+e^{-(x_i-\theta)}\right)^2}\,\frac{\prod_{i=1}^n\left(1+e^{-(y_i-\theta)}\right)^2}{e^{-\Sigma_i(y_i-\theta)}} \\ &= e^{-\Sigma_i(y_i-x_i)}\left(\frac{\prod_{i=1}^n\left(1+e^{-(y_i-\theta)}\right)}{\prod_{i=1}^n\left(1+e^{-(x_i-\theta)}\right)}\right)^2.\end{aligned}$$

This is constant as a function of θ if and only if $\mathbf{x}$ and $\mathbf{y}$ have the same order statistics. Therefore, the order statistics are minimal sufficient for θ.

d. This is a difficult problem. The order statistics are a minimal sufficient statistic.

e. Fix sample points $\mathbf{x}$ and $\mathbf{y}$. Define $A(\theta) = \{i : x_i \leq \theta\}$, $B(\theta) = \{i : y_i \leq \theta\}$, $a(\theta) =$ the number of elements in $A(\theta)$ and $b(\theta) =$ the number of elements in $B(\theta)$. Then the function $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ depends on θ only through the function

$$\begin{aligned}
&\sum_{i=1}^{n} | x_i - \theta | - \sum_{i=1}^{n} | y_i - \theta | \\
&= \sum_{i \in A(\theta)} (\theta - x_i) + \sum_{i \in A(\theta)^c} (x_i - \theta) - \sum_{i \in B(\theta)} (\theta - y_i) - \sum_{i \in B(\theta)^c} (y_i - \theta) \\
&= (a(\theta) - [n - a(\theta)] - b(\theta) + [n - b(\theta)])\theta \\
&\quad + \left(- \sum_{i \in A(\theta)} x_i + \sum_{i \in A(\theta)^c} x_i + \sum_{i \in B(\theta)} y_i - \sum_{i \in B(\theta)^c} y_i \right) \\
&= 2(a(\theta) - b(\theta))\theta + \left(- \sum_{i \in A(\theta)} x_i + \sum_{i \in A(\theta)^c} x_i + \sum_{i \in B(\theta)} y_i - \sum_{i \in B(\theta)^c} y_i \right).
\end{aligned}$$

Consider an interval of θs that does not contain any x_is or y_is. The second term is constant on such an interval. The first term will be constant, on the interval if and only if $a(\theta) = b(\theta)$. This will be true for all such intervals if and only if the order statistics for $\mathbf{x}$ are the same as the order statistics for $\mathbf{y}$. Therefore, the order statistics are a minimal sufficient statistic.

6.10 To prove $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ is not complete, we want to find $g[T(\mathbf{X})]$ such that $\mathrm{E}\, g[T(\mathbf{X})] = 0$ for all θ, but $g[T(\mathbf{X})] \not\equiv 0$. A natural candidate is $R = X_{(n)} - X_{(1)}$, the range of $\mathbf{X}$, because by Example 6.2.17 its distribution does not depend on θ. From Example 6.2.17, $R \sim \text{beta}(n-1, 2)$. Thus $\mathrm{E}\, R = (n-1)/(n+1)$ does not depend on θ, and $\mathrm{E}(R - \mathrm{E}\, R) = 0$ for all θ. Thus $g[X_{(n)}, X_{(1)}] = X_{(n)} - X_{(1)} - (n-1)/(n+1) = R - \mathrm{E}\, R$ is a nonzero function whose expected value is always 0. So, $(X_{(1)}, X_{(n)})$ is not complete. This problem can be generalized to show that if a function of a sufficient statistic is ancillary, then the sufficient statistic is not complete, because the expectation of that function does not depend on θ. That provides the opportunity to construct an unbiased, nonzero estimator of zero.

6.11 a. These are all location families. Let $Z_{(1)}, \ldots, Z_{(n)}$ be the order statistics from a random sample of size n from the standard pdf $f(z|0)$. Then $(Z_{(1)} + \theta, \ldots, Z_{(n)} + \theta)$ has the same joint distribution as $(X_{(1)}, \ldots, X_{(n)})$, and $(Y_{(1)}, \ldots, Y_{(n-1)})$ has the same joint distribution as $(Z_{(n)} + \theta - (Z_{(1)} + \theta), \ldots, Z_{(n)} + \theta - (Z_{(n-1)} + \theta)) = (Z_{(n)} - Z_{(1)}, \ldots, Z_{(n)} - Z_{(n-1)})$. The last vector depends only on $(Z_1, \ldots, Z_n)$ whose distribution does not depend on θ. So, $(Y_{(1)}, \ldots, Y_{(n-1)})$ is ancillary.

b. For a), Basu's lemma shows that $(Y_1, \ldots, Y_{n-1})$ is independent of the complete sufficient statistic. For c), d), and e) the order statistics are sufficient, so $(Y_1, \ldots, Y_{n-1})$ is not independent of the sufficient statistic. For b), $X_{(1)}$ is sufficient. Define $Y_n = X_{(1)}$. Then the joint pdf of $(Y_1, \ldots, Y_n)$ is

$$f(y_1, \ldots, y_n) = n! e^{-n(y_1 - \theta)} e^{-(n-1)y_n} \prod_{i=2}^{n-1} e^{y_i}, \qquad \begin{array}{c} 0 < y_{n-1} < y_{n-2} < \cdots < y_1 \\ 0 < y_n < \infty. \end{array}$$

Thus, $Y_n = X_{(1)}$ is independent of $(Y_1, \ldots, Y_{n-1})$.

6.12 a. Use Theorem 6.2.13 and write

$$\begin{aligned}
\frac{f(x, n|\theta)}{f(y, n'|\theta)} &= \frac{f(x|\theta, N = n) P(N = n)}{f(y|\theta, N = n') P(N = n')} \\
&= \frac{\binom{n}{x} \theta^x (1-\theta)^{n-x} p_n}{\binom{n'}{y} \theta^y (1-\theta)^{n'-y} p_{n'}} = \theta^{x-y} (1-\theta)^{n-n'-x+y} \frac{\binom{n}{x} p_n}{\binom{n'}{y} p_{n'}}.
\end{aligned}$$

The last ratio does not depend on θ. The other terms are constant as a function of θ if and only if $n = n'$ and $x = y$. So (X, N) is minimal sufficient for θ. Because $P(N = n) = p_n$ does not depend on θ, N is ancillary for θ. The point is that although N is independent of θ, the minimal sufficient statistic contains N in this case. A minimal sufficient statistic may contain an ancillary statistic.

b.

$$\begin{aligned}
\mathrm{E}\left(\frac{X}{N}\right) &= \mathrm{E}\left(\mathrm{E}\left(\frac{X}{N}\middle| N\right)\right) = \mathrm{E}\left(\frac{1}{N}\mathrm{E}\left(X \mid N\right)\right) = \mathrm{E}\left(\frac{1}{N}N\theta\right) = \mathrm{E}(\theta) = \theta. \\
\mathrm{Var}\left(\frac{X}{N}\right) &= \mathrm{Var}\left(\mathrm{E}\left(\frac{X}{N}\middle| N\right)\right) + \mathrm{E}\left(\mathrm{Var}\left(\frac{X}{N}\middle| N\right)\right) = \mathrm{Var}(\theta) + \mathrm{E}\left(\frac{1}{N^2}\mathrm{Var}\left(X \mid N\right)\right) \\
&= 0 + \mathrm{E}\left(\frac{N\theta(1-\theta)}{N^2}\right) = \theta(1-\theta)\mathrm{E}\left(\frac{1}{N}\right).
\end{aligned}$$

We used the fact that $X|N \sim \text{binomial}(N, \theta)$.

6.13 Let $Y_1 = \log X_1$ and $Y_2 = \log X_2$. Then Y_1 and Y_2 are iid and, by Theorem 2.1.5, the pdf of each is

$$f(y|\alpha) = \alpha \exp\left\{\alpha y - e^{\alpha y}\right\} = \frac{1}{1/\alpha}\exp\left\{\frac{y}{1/\alpha} - e^{y/(1/\alpha)}\right\}, \quad -\infty < y < \infty.$$

We see that the family of distributions of Y_i is a scale family with scale parameter $1/\alpha$. Thus, by Theorem 3.5.6, we can write $Y_i = \frac{1}{\alpha}Z_i$, where Z_1 and Z_2 are a random sample from $f(z|1)$. Then

$$\frac{\log X_1}{\log X_2} = \frac{Y_1}{Y_2} = \frac{(1/\alpha)Z_1}{(1/\alpha)Z_2} = \frac{Z_1}{Z_2}.$$

Because the distribution of Z_1/Z_2 does not depend on α, $(\log X_1)/(\log X_2)$ is an ancillary statistic.

6.14 Because $X_1, \ldots, X_n$ is from a location family, by Theorem 3.5.6, we can write $X_i = Z_i + \mu$, where $Z_1, \ldots, Z_n$ is a random sample from the standard pdf, $f(z)$, and μ is the location parameter. Let $M(\mathbf{X})$ denote the median calculated from $X_1, \ldots, X_n$. Then $M(\mathbf{X}) = M(\mathbf{Z}) + \mu$ and $\bar{X} = \bar{Z} + \mu$. Thus, $M(\mathbf{X}) - \bar{X} = (M(\mathbf{Z}) + \mu) - (\bar{Z} + \mu) = M(\mathbf{Z}) - \bar{Z}$. Because $M(\mathbf{X}) - \bar{X}$ is a function of only $Z_1, \ldots, Z_n$, the distribution of $M(\mathbf{X}) - \bar{X}$ does not depend on μ; that is, $M(\mathbf{X}) - \bar{X}$ is an ancillary statistic.

6.15 a. The parameter space consists only of the points (θ, ν) on the graph of the function $\nu = a\theta^2$. This quadratic graph is a line and does not contain a two-dimensional open set.

b. Use the same factorization as in Example 6.2.9 to show $(\bar{X}, S^2)$ is sufficient. $\mathrm{E}(S^2) = a\theta^2$ and $\mathrm{E}(\bar{X}^2) = \mathrm{Var}\bar{X} + (\mathrm{E}\bar{X})^2 = a\theta^2/n + \theta^2 = (a+n)\theta^2/n$. Therefore,

$$\mathrm{E}\left(\frac{n}{a+n}\bar{X}^2 - \frac{S^2}{a}\right) = \left(\frac{n}{a+n}\right)\left(\frac{a+n}{n}\theta^2\right) - \frac{1}{a}a\theta^2 = 0, \text{ for all } \theta.$$

Thus $g(\bar{X}, S^2) = \frac{n}{a+n}\bar{X}^2 - \frac{S^2}{a}$ has zero expectation so $(\bar{X}, S^2)$ not complete.

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).

6.18 The distribution of $Y = \sum_i X_i$ is Poisson$(n\lambda)$. Now

$$\mathrm{E}g(Y) = \sum_{y=0}^{\infty} g(y)\frac{(n\lambda)^y e^{-n\lambda}}{y!}.$$

If the expectation exists, this is an analytic function which cannot be identically zero.

6.19 To check if the family of distributions of X is complete, we check if $\mathrm{E}_p\, g(X) = 0$ for all p, implies that $g(X) \equiv 0$. For Distribution 1,

$$\mathrm{E}_p\, g(X) = \sum_{x=0}^{2} g(x) P(X = x) = pg(0) + 3pg(1) + (1 - 4p)g(2).$$

Note that if $g(0) = -3g(1)$ and $g(2) = 0$, then the expectation is zero for all p, but $g(x)$ need not be identically zero. Hence the family is not complete. For Distribution 2 calculate

$$\mathrm{E}_p\, g(X) = g(0)p + g(1)p^2 + g(2)(1 - p - p^2) = [g(1) - g(2)]p^2 + [g(0) - g(2)]p + g(2).$$

This is a polynomial of degree 2 in p. To make it zero for all p each coefficient must be zero. Thus, $g(0) = g(1) = g(2) = 0$, so the family of distributions is complete.

6.20 The pdfs in b), c), and e) are exponential families, so they have complete sufficient statistics from Theorem 6.2.25. For a), $Y = \max\{X_i\}$ is sufficient and

$$f(y) = \frac{2n}{\theta^{2n}} y^{2n-1}, \quad 0 < y < \theta.$$

For a function $g(y)$,

$$\mathrm{E}\, g(Y) = \int_0^\theta g(y) \frac{2n}{\theta^{2n}} y^{2n-1}\, dy = 0 \text{ for all } \theta \text{ implies } g(\theta) \frac{2n\theta^{2n-1}}{\theta^{2n}} = 0 \text{ for all } \theta$$

by taking derivatives. This can only be zero if $g(\theta) = 0$ for all θ, so $Y = \max\{X_i\}$ is complete. For d), the order statistics are minimal sufficient. This is a location family. Thus, by Example 6.2.18 the range $R = X_{(n)} - X_{(1)}$ is ancillary, and its expectation does not depend on θ. So this sufficient statistic is not complete.

6.21 a. X is sufficient because it is the data. To check completeness, calculate

$$\mathrm{E}g(X) = \frac{\theta}{2} g(-1) + (1 - \theta)g(0) + \frac{\theta}{2} g(1).$$

If $g(-1) = g(1)$ and $g(0) = 0$, then $\mathrm{E}g(X) = 0$ for all θ, but $g(x)$ need not be identically 0. So the family is not complete.

b. $|X|$ is sufficient by Theorem 6.2.6, because $f(x|\theta)$ depends on x only through the value of $|x|$. The distribution of $|X|$ is Bernoulli, because $P(|X| = 0) = 1 - \theta$ and $P(|X| = 1) = \theta$. By Example 6.2.22, a binomial family (Bernoulli is a special case) is complete.

c. Yes, $f(x|\theta) = (1 - \theta)(\theta/(2(1 - \theta))^{|x|} = (1 - \theta)e^{|x| \log[\theta/(2(1-\theta)]}$, the form of an exponential family.

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

6.23 Use Theorem 6.2.13. The ratio

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{\theta^{-n} I_{(x_{(n)}/2, x_{(1)})}(\theta)}{\theta^{-n} I_{(y_{(n)}/2, y_{(1)})}(\theta)}$$

is constant (in fact, one) if and only if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$. So $(X_{(1)}, X_{(n)})$ is a minimal sufficient statistic for θ. From Exercise 6.10, we know that if a function of the sufficient statistics is ancillary, then the sufficient statistic is not complete. The uniform$(\theta, 2\theta)$ family is a scale family, with standard pdf $f(z) \sim$ uniform$(1, 2)$. So if $Z_1, \ldots, Z_n$ is a random sample from a uniform$(1,2)$ population, then $X_1 = \theta Z_1, \ldots, X_n = \theta Z_n$ is a random sample from a uniform$(\theta, 2\theta)$ population, and $X_{(1)} = \theta Z_{(1)}$ and $X_{(n)} = \theta Z_{(n)}$. So $X_{(1)}/X_{(n)} = Z_{(1)}/Z_{(n)}$, a statistic whose distribution does not depend on θ. Thus, as in Exercise 6.10, $(X_{(1)}, X_{(n)})$ is not complete.

6.24 If $\lambda = 0$, $\mathrm{E}h(X) = h(0)$. If $\lambda = 1$,

$$\mathrm{E}h(X) = e^{-1}h(0) + e^{-1}\sum_{x=1}^{\infty}\frac{h(x)}{x!}.$$

Let $h(0) = 0$ and $\sum_{x=1}^{\infty}\frac{h(x)}{x!} = 0$, so $\mathrm{E}h(X) = 0$ but $h(x) \not\equiv 0$. (For example, take $h(0) = 0$, $h(1) = 1$, $h(2) = -2$, $h(x) = 0$ for $x \geq 3$.)

6.25 Using the fact that $(n-1)s_{\mathbf{x}}^2 = \sum_i x_i^2 - n\bar{x}^2$, for any (μ, σ^2) the ratio in Example 6.2.14 can be written as

$$\frac{f(\mathbf{x}|\mu,\sigma^2)}{f(\mathbf{y}|\mu,\sigma^2)} = \exp\left[\frac{\mu}{\sigma^2}\left(\sum_i x_i - \sum_i y_i\right) - \frac{1}{2\sigma^2}\left(\sum_i x_i^2 - \sum_i y_i^2\right)\right].$$

a. Do part b) first showing that $\sum_i X_i^2$ is a minimal sufficient statistic. Because $\left(\sum_i X_i, \sum_i X_i^2\right)$ is not a function of $\sum_i X_i^2$, by Definition 6.2.11 $\left(\sum_i X_i, \sum_i X_i^2\right)$ is not minimal.

b. Substituting $\sigma^2 = \mu$ in the above expression yields

$$\frac{f(\mathbf{x}|\mu,\mu)}{f(\mathbf{y}|\mu,\mu)} = \exp\left[\sum_i x_i - \sum_i y_i\right]\exp\left[-\frac{1}{2\mu}\left(\sum_i x_i^2 - \sum_i y_i^2\right)\right].$$

This is constant as a function of μ if and only if $\sum_i x_i^2 = \sum_i y_i^2$. Thus, $\sum_i X_i^2$ is a minimal sufficient statistic.

c. Substituting $\sigma^2 = \mu^2$ in the first expression yields

$$\frac{f(\mathbf{x}|\mu,\mu^2)}{f(\mathbf{y}|\mu,\mu^2)} = \exp\left[\frac{1}{\mu}\left(\sum_i x_i - \sum_i y_i\right) - \frac{1}{2\mu^2}\left(\sum_i x_i^2 - \sum_i y_i^2\right)\right].$$

This is constant as a function of μ if and only if $\sum_i x_i = \sum_i y_i$ and $\sum_i x_i^2 = \sum_i y_i^2$. Thus, $\left(\sum_i X_i, \sum_i X_i^2\right)$ is a minimal sufficient statistic.

d. The first expression for the ratio is constant a function of μ and σ^2 if and only if $\sum_i x_i = \sum_i y_i$ and $\sum_i x_i^2 = \sum_i y_i^2$. Thus, $\left(\sum_i X_i, \sum_i X_i^2\right)$ is a minimal sufficient statistic.

6.27 a. This pdf can be written as

$$f(x|\mu,\lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2}\left(\frac{1}{x^3}\right)^{1/2}\exp\left(\frac{\lambda}{\mu}\right)\exp\left(-\frac{\lambda}{2\mu^2}x - \frac{\lambda}{2}\frac{1}{x}\right).$$

This is an exponential family with $t_1(x) = x$ and $t_2(x) = 1/x$. By Theorem 6.2.25, the statistic $(\sum_i X_i, \sum_i(1/X_i))$ is a complete sufficient statistic. $(\bar{X}, T)$ given in the problem is a one-to-one function of $(\sum_i X_i, \sum_i(1/X_i))$. Thus, $(\bar{X}, T)$ is also a complete sufficient statistic.

b. This can be accomplished using the methods from Section 4.3 by a straightforward but messy two-variable transformation $U = (X_1 + X_2)/2$ and $V = 2\lambda/T = \lambda[(1/X_1) + (1/X_2) - (2/[X_1 + X_2])]$. This is a two-to-one transformation.

6.29 Let $f_j = \text{logistic}(\alpha_j, \beta_j)$, $j = 0, 1, \ldots, k$. From Theorem 6.6.5, the statistic

$$T(\mathbf{x}) = \left(\frac{\prod_{i=1}^n f_1(x_i)}{\prod_{i=1}^n f_0(x_i)}, \ldots, \frac{\prod_{i=1}^n f_k(x_i)}{\prod_{i=1}^n f_0(x_i)} \right) = \left(\frac{\prod_{i=1}^n f_1(x_{(i)})}{\prod_{i=1}^n f_0(x_{(i)})}, \ldots, \frac{\prod_{i=1}^n f_k(x_{(i)})}{\prod_{i=1}^n f_0(x_{(i)})} \right)$$

is minimal sufficient for the family $\{f_0, f_1, \ldots, f_k\}$. As T is a $1-1$ function of the order statistics, the order statistics are also minimal sufficient for the family $\{f_0, f_1, \ldots, f_k\}$. If $\mathcal{F}$ is a nonparametric family, $f_j \in \mathcal{F}$, so part (b) of Theorem 6.6.5 can now be directly applied to show that the order statistics are minimal sufficient for $\mathcal{F}$.

6.30 a. From Exercise 6.9b, we have that $X_{(1)}$ is a minimal sufficient statistic. To check completeness compute $f_{Y_1}(y)$, where $Y_1 = X_{(1)}$. From Theorem 5.4.4 we have

$$f_{Y_1}(y) = f_X(y)\,(1-F_X(y))^{n-1}\,n = e^{-(y-\mu)}\left[e^{-(y-\mu)}\right]^{n-1} n = ne^{-n(y-\mu)}, \quad y > \mu.$$

Now, write $\mathrm{E}_\mu\, g(Y_1) = \int_\mu^\infty g(y) n e^{-n(y-\mu)}\, dy$. If this is zero for all μ, then $\int_\mu^\infty g(y) e^{-ny}\, dy = 0$ for all μ (because $ne^{n\mu} > 0$ for all μ and does not depend on y). Moreover,

$$0 = \frac{d}{d\mu}\left[\int_\mu^\infty g(y) e^{-ny}\, dy\right] = -g(\mu)e^{-n\mu}$$

for all μ. This implies $g(\mu) = 0$ for all μ, so $X_{(1)}$ is complete.

b. Basu's Theorem says that if $X_{(1)}$ is a complete sufficient statistic for μ, then $X_{(1)}$ is independent of any ancillary statistic. Therefore, we need to show only that S^2 has distribution independent of μ; that is, S^2 is ancillary. Recognize that $f(x|\mu)$ is a location family. So we can write $X_i = Z_i + \mu$, where $Z_1, \ldots, Z_n$ is a random sample from $f(x|0)$. Then

$$S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2 = \frac{1}{n-1}\sum ((Z_i + \mu) - (\bar{Z} + \mu))^2 = \frac{1}{n-1}\sum (Z_i - \bar{Z})^2.$$

Because S^2 is a function of only $Z_1, \ldots, Z_n$, the distribution of S^2 does not depend on μ; that is, S^2 is ancillary. Therefore, by Basu's theorem, S^2 is independent of $X_{(1)}$.

6.31 a. (i) By Exercise 3.28 this is a one-dimensional exponential family with $t(x) = x$. By Theorem 6.2.25, $\sum_i X_i$ is a complete sufficient statistic. $\bar{X}$ is a one-to-one function of $\sum_i X_i$, so $\bar{X}$ is also a complete sufficient statistic. From Theorem 5.3.1 we know that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1} = \text{gamma}((n-1)/2, 2)$. $S^2 = [\sigma^2/(n-1)][(n-1)S^2/\sigma^2]$, a simple scale transformation, has a $\text{gamma}((n-1)/2, 2\sigma^2/(n-1))$ distribution, which does not depend on μ; that is, S^2 is ancillary. By Basu's Theorem, $\bar{X}$ and S^2 are independent.

(ii) The independence of $\bar{X}$ and S^2 is determined by the joint distribution of $(\bar{X}, S^2)$ for each value of (μ, σ^2). By part (i), for each value of (μ, σ^2), $\bar{X}$ and S^2 are independent.

b. (i) μ is a location parameter. By Exercise 6.14, $M - \bar{X}$ is ancillary. As in part (a) $\bar{X}$ is a complete sufficient statistic. By Basu's Theorem, $\bar{X}$ and $M - \bar{X}$ are independent. Because they are independent, by Theorem 4.5.6 $\mathrm{Var}\, M = \mathrm{Var}(M - \bar{X} + \bar{X}) = \mathrm{Var}(M - \bar{X}) + \mathrm{Var}\, \bar{X}$.

(ii) If S^2 is a sample variance calculated from a normal sample of size N, $(N-1)S^2/\sigma^2 \sim \chi^2_{N-1}$. Hence, $(N-1)^2 \mathrm{Var}\, S^2/(\sigma^2)^2 = 2(N-1)$ and $\mathrm{Var}\, S^2 = 2(\sigma^2)^2/(N-1)$. Both M and $M - \bar{X}$ are asymptotically normal, so, $M_1, \ldots, M_N$ and $M_1 - \bar{X}_1, \ldots, M_N - \bar{X}_N$ are each approximately normal samples if n is reasonable large. Thus, using the above expression we get the two given expressions where in the straightforward case σ^2 refers to $\mathrm{Var}\, M$, and in the swindle case σ^2 refers to $\mathrm{Var}(M - \bar{X})$.

c. (i)

$$\mathrm{E}(X^k) = \mathrm{E}\left(\frac{X}{Y}Y\right)^k = \mathrm{E}\left[\left(\frac{X}{Y}\right)^k (Y^k)\right] \overset{\text{indep.}}{=} \mathrm{E}\left(\frac{X}{Y}\right)^k \mathrm{E}\left(Y^k\right).$$

Divide both sides by $\mathrm{E}\left(Y^k\right)$ to obtain the desired equality.

(ii) If α is fixed, $T = \sum_i X_i$ is a complete sufficient statistic for β by Theorem 6.2.25. Because β is a scale parameter, if $Z_1, \ldots, Z_n$ is a random sample from a gamma$(\alpha, 1)$ distribution, then $X_{(i)}/T$ has the same distribution as $(\beta Z_{(i)})/(\beta \sum_i Z_i) = Z_{(i)}/(\sum_i Z_i)$, and this distribution does not depend on β. Thus, $X_{(i)}/T$ is ancillary, and by Basu's Theorem, it is independent of T. We have

$$\mathrm{E}(X_{(i)}|T) = \mathrm{E}\left(\frac{X_{(i)}}{T}T\middle|T\right) = T\mathrm{E}\left(\frac{X_{(i)}}{T}\middle|T\right) \overset{\text{indep.}}{=} T\mathrm{E}\left(\frac{X_{(i)}}{T}\right) \overset{\text{part (i)}}{=} T\frac{\mathrm{E}(X_{(i)})}{\mathrm{E}T}.$$

Note, this expression is correct for each fixed value of (α, β), regardless whether α is "known" or not.

6.32 In the Formal Likelihood Principle, take $E_1 = E_2 = E$. Then the conclusion is $\mathrm{Ev}(E, x_1) = \mathrm{Ev}(E, x_2)$ if $L(\theta|x_1)/L(\theta|x_2) = c$. Thus evidence is equal whenever the likelihood functions are equal, and this follows from Formal Sufficiency and Conditionality.

6.33 a. For all sample points except $(2, \mathbf{x}_2^*)$ (but including $(1, \mathbf{x}_1^*)$), $T(j, \mathbf{x}_j) = (j, \mathbf{x}_j)$. Hence,

$$g(T(j, \mathbf{x}_j)|\theta)h(j, \mathbf{x}_j) = g((j, \mathbf{x}_j)|\theta)1 = f^*((j, \mathbf{x}_j)|\theta).$$

For $(2, \mathbf{x}_2^*)$ we also have

$$\begin{aligned} g(T(2, \mathbf{x}_2^*)|\theta)h(2, \mathbf{x}_2^*) &= g((1, \mathbf{x}_1^*)|\theta)C = f^*((1, \mathbf{x}_1^*)|\theta)C = C\frac{1}{2}f_1(\mathbf{x}_1^*|\theta) \\ &= C\frac{1}{2}L(\theta|\mathbf{x}_1^*) = \frac{1}{2}L(\theta|\mathbf{x}_2^*) = \frac{1}{2}f_2(\mathbf{x}_2^*|\theta) = f^*((2, \mathbf{x}_2^*)|\theta). \end{aligned}$$

By the Factorization Theorem, $T(J, \mathbf{X}_J)$ is sufficient.

b. Equations 6.3.4 and 6.3.5 follow immediately from the two Principles. Combining them we have $\mathrm{Ev}(E_1, \mathbf{x}_1^*) = \mathrm{Ev}(E_2, \mathbf{x}_2^*)$, the conclusion of the Formal Likelihood Principle.

c. To prove the Conditionality Principle. Let one experiment be the E^* experiment and the other E_j. Then

$$L(\theta|(j, \mathbf{x}_j)) = f^*((j, \mathbf{x}_j)|\theta) = \frac{1}{2}f_j(\mathbf{x}_j|\theta) = \frac{1}{2}L(\theta|\mathbf{x}_j).$$

Letting $(j, \mathbf{x}_j)$ and $\mathbf{x}_j$ play the roles of $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ in the Formal Likelihood Principle we can conclude $\mathrm{Ev}(E^*, (j, \mathbf{x}_j)) = \mathrm{Ev}(E_j, \mathbf{x}_j)$, the Conditionality Principle. Now consider the Formal Sufficiency Principle. If $T(\mathbf{X})$ is sufficient and $T(\mathbf{x}) = T(\mathbf{y})$, then $L(\theta|\mathbf{x}) = CL(\theta|\mathbf{y})$, where $C = h(\mathbf{x})/h(\mathbf{y})$ and h is the function from the Factorization Theorem. Hence, by the Formal Likelihood Principle, $\mathrm{Ev}(E, \mathbf{x}) = \mathrm{Ev}(E, \mathbf{y})$, the Formal Sufficiency Principle.

6.35 Let 1 = success and 0 = failure. The four sample points are $\{0, 10, 110, 111\}$. From the likelihood principle, inference about p is only through $L(p|\mathbf{x})$. The values of the likelihood are 1, p, p^2, and p^3, and the sample size does not directly influence the inference.

6.37 a. For one observation (X, Y) we have

$$I(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X, Y|\theta)\right) = -\mathrm{E}\left(-\frac{2Y}{\theta^3}\right) = \frac{2\mathrm{E}Y}{\theta^3}.$$

But, $Y \sim$ exponential(θ), and $\mathrm{E}Y = \theta$. Hence, $I(\theta) = 2/\theta^2$ for a sample of size one, and $I(\theta) = 2n/\theta^2$ for a sample of size n.

b. (i) The cdf of T is

$$P(T \le t) = P\left(\frac{\sum_i Y_i}{\sum_i X_i} \le t^2\right) = P\left(\frac{2\sum_i Y_i/\theta}{2\sum_i X_i\theta} \le t^2/\theta^2\right) = P(F_{2n,2n} \le t^2/\theta^2)$$

where $F_{2n,2n}$ is an F random variable with $2n$ degrees of freedom in the numerator and denominator. This follows since $2Y_i/\theta$ and $2X_i\theta$ are all independent exponential(1), or χ^2_2. Differentiating (in t) and simplifying gives the density of T as

$$f_T(t) = \frac{\Gamma(2n)}{\Gamma(n)^2}\frac{2}{t}\left(\frac{t^2}{t^2+\theta^2}\right)^n\left(\frac{\theta^2}{t^2+\theta^2}\right)^n,$$

and the second derivative (in θ) of the log density is

$$2n\frac{t^4+2t^2\theta^2-\theta^4}{\theta^2(t^2+\theta^2)^2} = \frac{2n}{\theta^2}\left(1-\frac{2}{(t^2/\theta^2+1)^2}\right),$$

and the information in T is

$$\frac{2n}{\theta^2}\left[1-2\mathrm{E}\left(\frac{1}{T^2/\theta^2+1}\right)^2\right] = \frac{2n}{\theta^2}\left[1-2\mathrm{E}\left(\frac{1}{F^2_{2n,2n}+1}\right)^2\right].$$

The expected value is

$$\mathrm{E}\left(\frac{1}{F^2_{2n,2n}+1}\right)^2 = \frac{\Gamma(2n)}{\Gamma(n)^2}\int_0^\infty \frac{1}{(1+w)^2}\frac{w^{n-1}}{(1+w)^{2n}} = \frac{\Gamma(2n)}{\Gamma(n)^2}\frac{\Gamma(n)\Gamma(n+2)}{\Gamma(2n+2)} = \frac{n+1}{2(2n+1)}.$$

Substituting this above gives the information in T as

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

which is not the answer reported by Joshi and Nabar.

(ii) Let $W=\sum_i X_i$ and $V=\sum_i Y_i$. In each pair, X_i and Y_i are independent, so W and V are independent. $X_i \sim$ exponential$(1/\theta)$; hence, $W \sim$ gamma$(n,1/\theta)$. $Y_i \sim$ exponential(θ); hence, $V \sim$ gamma(n,θ). Use this joint distribution of (W,V) to derive the joint pdf of (T,U) as

$$f(t,u|\theta) = \frac{2}{[\Gamma(n)]^2 t}u^{2n-1}\exp\left(-\frac{u\theta}{t}-\frac{ut}{\theta}\right), \quad u>0,\quad t>0.$$

Now, the information in (T,U) is

$$-\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(T,U|\theta)\right) = -\mathrm{E}\left(-\frac{2UT}{\theta^3}\right) = \mathrm{E}\left(\frac{2V}{\theta^3}\right) = \frac{2n\theta}{\theta^3} = \frac{2n}{\theta^2}.$$

(iii) The pdf of the sample is $f(\mathbf{x},\mathbf{y}) = \exp\left[-\theta\left(\sum_i x_i\right)-\left(\sum_i y_i\right)/\theta\right]$. Hence, (W,V) defined as in part (ii) is sufficient. (T,U) is a one-to-one function of (W,V), hence (T,U) is also sufficient. But, $\mathrm{E}\,U^2 = \mathrm{E}\,WV = (n/\theta)(n\theta) = n^2$ does not depend on θ. So $\mathrm{E}(U^2-n^2)=0$ for all θ, and (T,U) is not complete.

6.39 a. The transformation from Celsius to Fahrenheit is $y = 9x/5+32$. Hence,

$$\begin{aligned}\frac{5}{9}(T^*(y)-32) &= \frac{5}{9}((.5)(y)+(.5)(212)-32)\\ &= \frac{5}{9}((.5)(9x/5+32)+(.5)(212)-32) \;=\; (.5)x+50 \;=\; T(x).\end{aligned}$$

b. $T(x) = (.5)x+50 \neq (.5)x+106 = T^*(x)$. Thus, we do not have equivariance.

6.40 a. Because $X_1, \ldots, X_n$ is from a location scale family, by Theorem 3.5.6, we can write $X_i = \sigma Z_i + \mu$, where $Z_1, \ldots, Z_n$ is a random sample from the standard pdf $f(z)$. Then

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

Because T_1/T_2 is a function of only $Z_1, \ldots, Z_n$, the distribution of T_1/T_2 does not depend on μ or σ; that is, T_1/T_2 is an ancillary statistic.

b. $R(x_1, \ldots, x_n) = x_{(n)} - x_{(1)}$. Because $a > 0$, $\max\{ax_1 + b, \ldots, ax_n + b\} = ax_{(n)} + b$ and $\min\{ax_1+b, \ldots, ax_n+b\} = ax_{(1)}+b$. Thus, $R(ax_1+b, \ldots, ax_n+b) = (ax_{(n)}+b)-(ax_{(1)}+b) = a(x_{(n)} - x_{(1)}) = aR(x_1, \ldots, x_n)$. For the sample variance we have

$$\begin{aligned} S^2(ax_1+b,\ldots,ax_n+b) &= \frac{1}{n-1}\sum((ax_i+b)-(a\bar{x}+b))^2 \\ &= a^2\frac{1}{n-1}\sum(x_i-\bar{x})^2 \quad = \quad a^2S^2(x_1,\ldots,x_n). \end{aligned}$$

Thus, $S(ax_1 + b, \ldots, ax_n + b) = aS(x_1, \ldots, x_n)$. Therefore, R and S both satisfy the above condition, and R/S is ancillary by a).

6.41 a. Measurement equivariance requires that the estimate of μ based on $\mathbf{y}$ be the same as the estimate of μ based on $\mathbf{x}$; that is, $T^*(x_1 + a, \ldots, x_n + a) - a = T^*(\mathbf{y}) - a = T(\mathbf{x})$.

b. The formal structures for the problem involving $\mathbf{X}$ and the problem involving $\mathbf{Y}$ are the same. They both concern a random sample of size n from a normal population and estimation of the mean of the population. Thus, formal invariance requires that $T(\mathbf{x}) = T^*(\mathbf{x})$ for all $\mathbf{x}$. Combining this with part (a), the Equivariance Principle requires that $T(x_1+a, \ldots, x_n+a)-a = T^*(x_1+a, \ldots, x_n+a)-a = T(x_1, \ldots, x_n)$, i.e., $T(x_1+a, \ldots, x_n+a) = T(x_1, \ldots, x_n)+a$.

c. $W(x_1 + a, \ldots, x_n + a) = \sum_i(x_i + a)/n = \left(\sum_i x_i\right)/n + a = W(x_1, \ldots, x_n) + a$, so $W(\mathbf{x})$ is equivariant. The distribution of $(X_1, \ldots, X_n)$ is the same as the distribution of $(Z_1 + \theta, \ldots, Z_n + \theta)$, where $Z_1, \ldots, Z_n$ are a random sample from $f(x - 0)$ and $\mathrm{E}\,Z_i = 0$. Thus, $\mathrm{E}_\theta W = \mathrm{E}\sum_i(Z_i + \theta)/n = \theta$, for all θ.

6.43 a. For a location-scale family, if $X \sim f(x|\theta, \sigma^2)$, then $Y = g_{a,c}(X) \sim f(y|c\theta + a, c^2\sigma^2)$. So for estimating σ^2, $\bar{g}_{a,c}(\sigma^2) = c^2\sigma^2$. An estimator of σ^2 is invariant with respect to $\mathcal{G}_1$ if $W(cx_1 + a, \ldots, cx_n + a) = c^2W(x_1, \ldots, x_n)$. An estimator of the form kS^2 is invariant because

$$\begin{aligned} kS^2(cx_1+a,\ldots,cx_n+a) &= \frac{k}{n-1}\sum_{i=1}^n\left((cx_i+a)-\sum_{i=1}^n(cx_i+a)/n\right)^2 \\ &= \frac{k}{n-1}\sum_{i=1}^n((cx_i+a)-(c\bar{x}+a))^2 \\ &= c^2\frac{k}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2 \quad = \quad c^2kS^2(x_1,\ldots,x_n). \end{aligned}$$

To show invariance with respect to $\mathcal{G}_2$, use the above argument with $c = 1$. To show invariance with respect to $\mathcal{G}_3$, use the above argument with $a = 0$. ($\mathcal{G}_2$ and $\mathcal{G}_3$ are both subgroups of $\mathcal{G}_1$. So invariance with respect to $\mathcal{G}_1$ implies invariance with respect to $\mathcal{G}_2$ and $\mathcal{G}_3$.)

b. The transformations in $\mathcal{G}_2$ leave the scale parameter unchanged. Thus, $\bar{g}_a(\sigma^2) = \sigma^2$. An estimator of σ^2 is invariant with respect to this group if

$$W(x_1 + a, \ldots, x_n + a) = W(g_a(\mathbf{x})) = \bar{g}_a(W(\mathbf{x})) = W(x_1, \ldots, x_n).$$

An estimator of the given form is invariant if, for all a and $(x_1, \ldots, x_n)$,

$$W(x_1 + a, \ldots, x_n + a) = \phi\left(\frac{\bar{x}+a}{s}\right) s^2 = \phi\left(\frac{\bar{x}}{s}\right) s^2 = W(x_1, \ldots, x_n).$$

In particular, for a sample point with $s = 1$ and $\bar{x} = 0$, this implies we must have $\phi(a) = \phi(0)$, for all a; that is, ϕ must be constant. On the other hand, if ϕ is constant, then the estimators are invariant by part a). So we have invariance if and only if ϕ is constant. Invariance with respect to $\mathcal{G}_1$ also requires ϕ to be constant because $\mathcal{G}_2$ is a subgroup of $\mathcal{G}_1$. Finally, an estimator of σ^2 is invariant with respect to $\mathcal{G}_3$ if $W(cx_1, \ldots, cx_n) = c^2 W(x_1, \ldots, x_n)$. Estimators of the given form are invariant because

$$W(cx_1, \ldots, cx_n) = \phi\left(\frac{c\bar{x}}{cs}\right) c^2 s^2 = c^2 \phi\left(\frac{\bar{x}}{s}\right) s^2 = c^2 W(x_1, \ldots, x_n).$$

Chapter 7

Point Estimation

7.1 For each value of x, the MLE $\hat{\theta}$ is the value of θ that maximizes $f(x|\theta)$. These values are in the following table.

x	0	1	2	3	4
$\hat{\theta}$	1	1	2 or 3	3	3

At $x = 2$, $f(x|2) = f(x|3) = 1/4$ are both maxima, so both $\hat{\theta} = 2$ or $\hat{\theta} = 3$ are MLEs.

7.2 a.

$$\begin{aligned}
L(\beta|x) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_i^{\alpha-1} e^{-x_i/\beta} \quad = \quad \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/\beta} \\
\log L(\beta|x) &= -\log\Gamma(\alpha)^n - n\alpha\log\beta + (\alpha-1)\log\left[\prod_{i=1}^{n} x_i\right] - \frac{\sum_i x_i}{\beta} \\
\frac{\partial \log L}{\partial\beta} &= -\frac{n\alpha}{\beta} + \frac{\sum_i x_i}{\beta^2}
\end{aligned}$$

Set the partial derivative equal to 0 and solve for β to obtain $\hat{\beta} = \sum_i x_i/(n\alpha)$. To check that this is a maximum, calculate

$$\left.\frac{\partial^2 \log L}{\partial\beta^2}\right|_{\beta=\hat{\beta}} = \left.\frac{n\alpha}{\beta^2} - \frac{2\sum_i x_i}{\beta^3}\right|_{\beta=\hat{\beta}} = \frac{(n\alpha)^3}{\left(\sum_i x_i\right)^2} - \frac{2(n\alpha)^3}{\left(\sum_i x_i\right)^2} = -\frac{(n\alpha)^3}{\left(\sum_i x_i\right)^2} < 0.$$

Because $\hat{\beta}$ is the unique point where the derivative is 0 and it is a local maximum, it is a global maximum. That is, $\hat{\beta}$ is the MLE.

b. Now the likelihood function is

$$L(\alpha, \beta|x) = \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/\beta},$$

the same as in part (a) except α and β are both variables. There is no analytic form for the MLEs, The values $\hat{\alpha}$ and $\hat{\beta}$ that maximize L. One approach to finding $\hat{\alpha}$ and $\hat{\beta}$ would be to numerically maximize the function of two arguments. But it is usually best to do as much as possible analytically, first, and perhaps reduce the complexity of the numerical problem. From part (a), for each fixed value of α, the value of β that maximizes L is $\sum_i x_i/(n\alpha)$. Substitute this into L. Then we just need to maximize the function of the one variable α given by

$$\begin{aligned}
&\frac{1}{\Gamma(\alpha)^n \left(\sum_i x_i/(n\alpha)\right)^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\Sigma_i x_i/(\Sigma_i x_i/(n\alpha))} \\
&= \frac{1}{\Gamma(\alpha)^n \left(\sum_i x_i/(n\alpha)\right)^{n\alpha}} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-n\alpha}.
\end{aligned}$$

For the given data, $n = 14$ and $\sum_i x_i = 323.6$. Many computer programs can be used to maximize this function. From PROC NLIN in SAS we obtain $\hat{\alpha} = 514.219$ and, hence, $\hat{\beta} = \frac{323.6}{14(514.219)} = .0450$.

7.3 The log function is a strictly monotone increasing function. Therefore, $L(\theta|\mathbf{x}) > L(\theta'|\mathbf{x})$ if and only if $\log L(\theta|\mathbf{x}) > \log L(\theta'|\mathbf{x})$. So the value $\hat{\theta}$ that maximizes $\log L(\theta|\mathbf{x})$ is the same as the value that maximizes $L(\theta|\mathbf{x})$.

7.5 a. The value $\hat{z}$ solves the equation

$$(1-p)^n = \prod_i (1 - x_i z),$$

where $0 \le z \le (\max_i x_i)^{-1}$. Let $\hat{k}$ = greatest integer less than or equal to $1/\hat{z}$. Then from Example 7.2.9, $\hat{k}$ must satisfy

$$[k(1-p)]^n \ge \prod_i (k - x_i) \qquad \text{and} \qquad [(k+1)(1-p)]^n < \prod_i (k + 1 - x_i).$$

Because the right-hand side of the first equation is decreasing in $\hat{z}$, and because $\hat{k} \le 1/\hat{z}$ (so $\hat{z} \le 1/\hat{k}$) and $\hat{k} + 1 > 1/\hat{z}$, $\hat{k}$ must satisfy the two inequalities. Thus $\hat{k}$ is the MLE.

b. For $p = 1/2$, we must solve $\left(\frac{1}{2}\right)^4 = (1 - 20z)(1 - z)(1 - 19z)$, which can be reduced to the cubic equation $-380z^3 + 419z^2 - 40z + 15/16 = 0$. The roots are .9998, .0646, and .0381, leading to candidates of 1, 15, and 26 for $\hat{k}$. The first two are less than $\max_i x_i$. Thus $\hat{k} = 26$.

7.6 a. $f(\mathbf{x}|\theta) = \prod_i \theta x_i^{-2} I_{[\theta,\infty)}(x_i) = \left(\prod_i x_i^{-2}\right) \theta^n I_{[\theta,\infty)}(x_{(1)})$. Thus, $X_{(1)}$ is a sufficient statistic for θ by the Factorization Theorem.

b. $L(\theta|\mathbf{x}) = \theta^n \left(\prod_i x_i^{-2}\right) I_{[\theta,\infty)}(x_{(1)})$. θ^n is increasing in θ. The second term does not involve θ. So to maximize $L(\theta|\mathbf{x})$, we want to make θ as large as possible. But because of the indicator function, $L(\theta|\mathbf{x}) = 0$ if $\theta > x_{(1)}$. Thus, $\hat{\theta} = x_{(1)}$.

c. $\mathrm{E}\, X = \int_\theta^\infty \theta x^{-1}\, dx = \theta \log x|_\theta^\infty = \infty$. Thus the method of moments estimator of θ does not exist. (This is the Pareto distribution with $\alpha = \theta$, $\beta = 1$.)

7.7 $L(0|\mathbf{x}) = 1$, $0 < x_i < 1$, and $L(1|\mathbf{x}) = \prod_i 1/(2\sqrt{x_i})$, $0 < x_i < 1$. Thus, the MLE is 0 if $1 \ge \prod_i 1/(2\sqrt{x_i})$, and the MLE is 1 if $1 < \prod_i 1/(2\sqrt{x_i})$.

7.8 a. $\mathrm{E}\, X^2 = \mathrm{Var}\, X + \mu^2 = \sigma^2$. Therefore X^2 is an unbiased estimator of σ^2.

b.

$$\begin{aligned}
L(\sigma|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}. \quad \log L(\sigma|\mathbf{x}) = \log(2\pi)^{-1/2} - \log\sigma - x^2/(2\sigma^2). \\
\frac{\partial \log L}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{x^2}{\sigma^3} \stackrel{\text{set}}{=} 0 \Rightarrow \hat{\sigma} X^2 = \hat{\sigma}^3 \ \Rightarrow \hat{\sigma} = \sqrt{X^2} = |X|. \\
\frac{\partial^2 \log L}{\partial \sigma^2} &= \frac{-3x^2\sigma^2}{\sigma^6} + \frac{1}{\sigma^2}, \text{ which is negative at } \hat{\sigma} = |x|.
\end{aligned}$$

Thus, $\hat{\sigma} = |x|$ is a local maximum. Because it is the only place where the first derivative is zero, it is also a global maximum.

c. Because $\mathrm{E}\, X = 0$ is known, just equate $\mathrm{E}\, X^2 = \sigma^2 = \frac{1}{n}\sum_{i=1}^{1} X_i^2 = X^2 \Rightarrow \hat{\sigma} = |X|$.

7.9 This is a uniform$(0,\theta)$ model. So $\mathrm{E}\, X = (0+\theta)/2 = \theta/2$. The method of moments estimator is the solution to the equation $\tilde{\theta}/2 = \bar{X}$, that is, $\tilde{\theta} = 2\bar{X}$. Because $\tilde{\theta}$ is a simple function of the sample mean, its mean and variance are easy to calculate. We have

$$\mathrm{E}\, \tilde{\theta} = 2\mathrm{E}\, \bar{X} = 2\mathrm{E}\, X = 2\frac{\theta}{2} = \theta, \quad \text{and} \quad \mathrm{Var}\, \tilde{\theta} = 4\mathrm{Var}\, \bar{X} = 4\frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

The likelihood function is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}),$$

where $x_{(1)}$ and $x_{(n)}$ are the smallest and largest order statistics. For $\theta \geq x_{(n)}$, $L = 1/\theta^n$, a decreasing function. So for $\theta \geq x_{(n)}$, L is maximized at $\hat{\theta} = x_{(n)}$. $L = 0$ for $\theta < x_{(n)}$. So the overall maximum, the MLE, is $\hat{\theta} = X_{(n)}$. The pdf of $\hat{\theta} = X_{(n)}$ is nx^{n-1}/θ^n, $0 \leq x \leq \theta$. This can be used to calculate

$$\mathrm{E}\,\hat{\theta} = \frac{n}{n+1}\theta, \quad \mathrm{E}\,\hat{\theta}^2 = \frac{n}{n+2}\theta^2 \quad \text{and} \quad \mathrm{Var}\,\hat{\theta} = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

$\tilde{\theta}$ is an unbiased estimator of θ; $\hat{\theta}$ is a biased estimator. If n is large, the bias is not large because $n/(n+1)$ is close to one. But if n is small, the bias is quite large. On the other hand, $\mathrm{Var}\,\hat{\theta} < \mathrm{Var}\,\tilde{\theta}$ for all θ. So, if n is large, $\hat{\theta}$ is probably preferable to $\tilde{\theta}$.

7.10 a. $f(\mathbf{x}|\theta) = \prod_i \frac{\alpha}{\beta^\alpha} x_i^{\alpha-1} I_{[0,\beta]}(x_i) = \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_i x_i\right)^{\alpha-1} I_{(-\infty,\beta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}) = L(\alpha,\beta|\mathbf{x})$. By the Factorization Theorem, $(\prod_i X_i, X_{(n)})$ are sufficient.

b. For any fixed α, $L(\alpha,\beta|\mathbf{x}) = 0$ if $\beta < x_{(n)}$, and $L(\alpha,\beta|\mathbf{x})$ a decreasing function of β if $\beta \geq x_{(n)}$. Thus, $X_{(n)}$ is the MLE of β. For the MLE of α calculate

$$\frac{\partial}{\partial\alpha}\log L = \frac{\partial}{\partial\alpha}\left[n\log\alpha - n\alpha\log\beta + (\alpha-1)\log\prod_i x_i\right] = \frac{n}{\alpha} - n\log\beta + \log\prod_i x_i.$$

Set the derivative equal to zero and use $\hat{\beta} = X_{(n)}$ to obtain

$$\hat{\alpha} = \frac{n}{n\log X_{(n)} - \log\prod_i X_i} = \left[\frac{1}{n}\sum_i (\log X_{(n)} - \log X_i)\right]^{-1}.$$

The second derivative is $-n/\alpha^2 < 0$, so this is the MLE.

c. $X_{(n)} = 25.0$, $\log\prod_i X_i = \sum_i \log X_i = 43.95 \Rightarrow \hat{\beta} = 25.0$, $\hat{\alpha} = 12.59$.

7.11 a.

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_i \theta x_i^{\theta-1} = \theta^n \left(\prod_i x_i\right)^{\theta-1} = L(\theta|\mathbf{x}) \\ \frac{d}{d\theta}\log L &= \frac{d}{d\theta}\left[n\log\theta + (\theta-1)\log\prod_i x_i\right] = \frac{n}{\theta} + \sum_i \log x_i. \end{aligned}$$

Set the derivative equal to zero and solve for θ to obtain $\hat{\theta} = (-\frac{1}{n}\sum_i \log x_i)^{-1}$. The second derivative is $-n/\theta^2 < 0$, so this is the MLE. To calculate the variance of $\hat{\theta}$, note that $Y_i = -\log X_i \sim \text{exponential}(1/\theta)$, so $-\sum_i \log X_i \sim \text{gamma}(n, 1/\theta)$. Thus $\hat{\theta} = n/T$, where $T \sim \text{gamma}(n, 1/\theta)$. We can either calculate the first and second moments directly, or use the fact that $\hat{\theta}$ is inverted gamma (page 51). We have

$$\begin{aligned} \mathrm{E}\frac{1}{T} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}. \\ \mathrm{E}\frac{1}{T^2} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t^2} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}, \end{aligned}$$

and thus

$$\mathrm{E}\,\hat{\theta} = \frac{n}{n-1}\theta \quad \text{and} \quad \mathrm{Var}\,\hat{\theta} = \frac{n^2}{(n-1)^2(n-2)}\theta^2 \to 0 \text{ as } n \to \infty.$$

b. Because $X \sim \text{beta}(\theta, 1)$, $\mathrm{E}\,X = \theta/(\theta+1)$ and the method of moments estimator is the solution to

$$\frac{1}{n}\sum_i X_i = \frac{\theta}{\theta+1} \Rightarrow \tilde{\theta} = \frac{\sum_i X_i}{n-\sum_i X_i}.$$

7.12 $X_i \sim$ iid Bernoulli(θ), $0 \le \theta \le 1/2$.

a. method of moments:

$$\mathrm{E}\,X = \theta = \frac{1}{n}\sum_i X_i = \bar{X} \quad \Rightarrow \quad \tilde{\theta} = \bar{X}.$$

MLE: In Example 7.2.7, we showed that $L(\theta|\mathbf{x})$ is increasing for $\theta \le \bar{x}$ and is decreasing for $\theta \ge \bar{x}$. Remember that $0 \le \theta \le 1/2$ in this exercise. Therefore, when $\bar{X} \le 1/2$, $\bar{X}$ is the MLE of θ, because $\bar{X}$ is the overall maximum of $L(\theta|\mathbf{x})$. When $\bar{X} > 1/2$, $L(\theta|\mathbf{x})$ is an increasing function of θ on $[0, 1/2]$ and obtains its maximum at the upper bound of θ which is $1/2$. So the MLE is $\hat{\theta} = \min\left\{\bar{X}, 1/2\right\}$.

b. The MSE of $\tilde{\theta}$ is $\mathrm{MSE}(\tilde{\theta}) = \mathrm{Var}\,\tilde{\theta} + \mathrm{bias}(\tilde{\theta})^2 = (\theta(1-\theta)/n) + 0^2 = \theta(1-\theta)/n$. There is no simple formula for $\mathrm{MSE}(\hat{\theta})$, but an expression is

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= \mathrm{E}(\hat{\theta}-\theta)^2 = \sum_{y=0}^{n}(\hat{\theta}-\theta)^2\binom{n}{y}\theta^y(1-\theta)^{n-y} \\
&= \sum_{y=0}^{[n/2]}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n}\left(\frac{1}{2}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y},
\end{aligned}$$

where $Y = \sum_i X_i \sim \text{binomial}(n,\theta)$ and $[n/2] = n/2$, if n is even, and $[n/2] = (n-1)/2$, if n is odd.

c. Using the notation used in (b), we have

$$\mathrm{MSE}(\tilde{\theta}) = \mathrm{E}(\bar{X}-\theta)^2 = \sum_{y=0}^{n}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

Therefore,

$$\begin{aligned}
\mathrm{MSE}(\tilde{\theta}) - \mathrm{MSE}(\hat{\theta}) &= \sum_{y=[n/2]+1}^{n}\left[\left(\frac{y}{n}-\theta\right)^2 - \left(\frac{1}{2}-\theta\right)^2\right]\binom{n}{y}\theta^y(1-\theta)^{n-y} \\
&= \sum_{y=[n/2]+1}^{n}\left(\frac{y}{n}+\frac{1}{2}-2\theta\right)\left(\frac{y}{n}-\frac{1}{2}\right)\binom{n}{y}\theta^y(1-\theta)^{n-y}.
\end{aligned}$$

The facts that $y/n > 1/2$ in the sum and $\theta \le 1/2$ imply that every term in the sum is positive. Therefore $\mathrm{MSE}(\hat{\theta}) < \mathrm{MSE}(\tilde{\theta})$ for every θ in $0 < \theta \le 1/2$. (Note: $\mathrm{MSE}(\hat{\theta}) = \mathrm{MSE}(\tilde{\theta}) = 0$ at $\theta = 0$.)

7.13 $L(\theta|\mathbf{x}) = \prod_i \frac{1}{2}e^{-\frac{1}{2}|x_i-\theta|} = \frac{1}{2^n}e^{-\frac{1}{2}\Sigma_i|x_i-\theta|}$, so the MLE minimizes $\sum_i|x_i-\theta| = \sum_i|x_{(i)}-\theta|$, where $x_{(1)},\ldots,x_{(n)}$ are the order statistics. For $x_{(j)} \le \theta \le x_{(j+1)}$,

$$\sum_{i=1}^{n}|x_{(i)}-\theta| = \sum_{i=1}^{j}(\theta-x_{(i)}) + \sum_{i=j+1}^{n}(x_{(i)}-\theta) = (2j-n)\theta - \sum_{i=1}^{j}x_{(i)} + \sum_{i=j+1}^{n}x_{(i)}.$$

This is a linear function of θ that decreases for $j < n/2$ and increases for $j > n/2$. If n is even, $2j - n = 0$ if $j = n/2$. So the likelihood is constant between $x_{(n/2)}$ and $x_{((n/2)+1)}$, and any value in this interval is the MLE. Usually the midpoint of this interval is taken as the MLE. If n is odd, the likelihood is minimized at $\hat{\theta} = x_{((n+1)/2)}$.

7.15 a. The likelihood is

$$L(\mu, \lambda|\mathbf{x}) = \frac{\lambda^{n/2}}{(2\pi)^n \prod_i x_i} \exp\left\{ -\frac{\lambda}{2} \sum_i \frac{(x_i - \mu)^2}{\mu^2 x_i} \right\}.$$

For fixed λ, maximizing with respect to μ is equivalent to minimizing the sum in the exponential.

$$\frac{d}{d\mu} \sum_i \frac{(x_i - \mu)^2}{\mu^2 x_i} = \frac{d}{d\mu} \sum_i \frac{((x_i/\mu) - 1)^2}{x_i} = -\sum_i \frac{2\left((x_i/\mu) - 1\right)}{x_i} \frac{x_i}{\mu^2}.$$

Setting this equal to zero is equivalent to setting

$$\sum_i \left(\frac{x_i}{\mu} - 1 \right) = 0,$$

and solving for μ yields $\hat{\mu}_n = \bar{x}$. Plugging in this $\hat{\mu}_n$ and maximizing with respect to λ amounts to maximizing an expression of the form $\lambda^{n/2} e^{-\lambda b}$. Simple calculus yields

$$\hat{\lambda}_n = \frac{n}{2b} \quad \text{where} \quad b = \sum_i \frac{(x_i - \bar{x})^2}{2\bar{x}^2 x_i}.$$

Finally,

$$2b = \sum_i \frac{x_i}{\bar{x}^2} - 2\sum_i \frac{1}{\bar{x}} + \sum_i \frac{1}{x_i} = -\frac{n}{\bar{x}} + \sum_i \frac{1}{x_i} = \sum_i \left(\frac{1}{x_i} - \frac{1}{\bar{x}} \right).$$

b. This is the same as Exercise 6.27b.

c. This involved algebra can be found in Schwarz and Samanta (1991).

7.17 a. This is a special case of the computation in Exercise 7.2a.

b. Make the transformation

$$z = (x_2 - 1)/x_1, w = x_1 \quad \Rightarrow \quad x_1 = w, x_2 = wz + 1.$$

The Jacobean is $|w|$, and

$$f_Z(z) = \int f_{X_1}(w) f_{X_2}(wz + 1) w dw = \frac{1}{\theta^2} e^{-1/\theta} \int w e^{-w(1+z)/\theta} dw,$$

where the range of integration is $0 < w < -1/z$ if $z < 0$, $0 < w < \infty$ if $z > 0$. Thus,

$$f_Z(z) = \frac{1}{\theta^2} e^{-1/\theta} \begin{cases} \int_0^{-1/z} w e^{-w(1+z)/\theta} dw & \text{if } z < 0 \\ \int_0^{\infty} w e^{-w(1+z)/\theta} dw & \text{if } z \geq 0 \end{cases}$$

Using the fact that $\int w e^{-w/a} dw = -e^{-w/a}(aw + a^2)$, we have

$$f_Z(z) = e^{-1/\theta} \begin{cases} \frac{z\theta + e^{(1+z)/z\theta}(1+z-z\theta)}{\theta z(1+z)^2} & \text{if } z < 0 \\ \frac{1}{(1+z)^2} & \text{if } z \geq 0 \end{cases}$$

c. From part (a) we get $\hat\theta = 1$. From part (b), $X_2 = 1$ implies $Z = 0$ which, if we use the second density, gives us $\hat\theta = \infty$.

d. The posterior distributions are just the normalized likelihood times prior, so of course they are different.

7.18 a. The usual first two moment equations for X and Y are

$$\begin{aligned} \bar{x} &= \mathrm{E}\,X = \mu_X, & \frac{1}{n}\sum_i x_i^2 &= \mathrm{E}\,X^2 = \sigma_X^2 + \mu_X^2, \\ \bar{y} &= \mathrm{E}\,Y = \mu_Y, & \frac{1}{n}\sum_i y_i^2 &= \mathrm{E}\,Y^2 = \sigma_Y^2 + \mu_Y^2. \end{aligned}$$

We also need an equation involving ρ.

$$\frac{1}{n}\sum_i x_i y_i = \mathrm{E}\,XY = \mathrm{Cov}(X,Y) + (\mathrm{E}\,X)(\mathrm{E}\,Y) = \rho\sigma_X\sigma_Y + \mu_X\mu_Y.$$

Solving these five equations yields the estimators given. Facts such as

$$\frac{1}{n}\sum_i x_i^2 - \bar{x}^2 = \frac{\sum_i x_i^2 - \left(\sum_i x_i\right)^2/n}{n} = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

are used.

b. Two answers are provided. First, use the Miscellanea: For

$$L(\boldsymbol{\theta}|\mathbf{x}) = h(\mathbf{x})c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(\mathbf{x})\right),$$

the solutions to the k equations $\sum_{j=1}^n t_i(x_j) = \mathrm{E}_\theta\left(\sum_{j=1}^n t_i(X_j)\right) = n\mathrm{E}_\theta t_i(X_1)$, $i = 1,\ldots,k$, provide the unique MLE for $\boldsymbol{\theta}$. Multiplying out the exponent in the bivariate normal pdf shows it has this exponential family form with $k = 5$ and $t_1(x,y) = x$, $t_2(x,y) = y$, $t_3(x,y) = x^2$, $t_4(x,y) = y^2$ and $t_5(x,y) = xy$. Setting up the method of moment equations, we have

$$\begin{aligned} \sum_i x_i &= n\mu_X, & \sum_i x_i^2 &= n(\mu_X^2 + \sigma_X^2), \\ \sum_i y_i &= n\mu_Y, & \sum_i y_i^2 &= n(\mu_Y^2 + \sigma_Y^2), \\ \sum_i x_i y_i &= \sum_i [\mathrm{Cov}(X,Y) + \mu_X\mu_Y] &&= n(\rho\sigma_X\sigma_Y + \mu_X\mu_Y). \end{aligned}$$

These are the same equations as in part (a) if you divide each one by n. So the MLEs are the same as the method of moment estimators in part (a).

For the second answer, use the hint in the book to write

$$\begin{aligned} L(\theta|\mathbf{x},\mathbf{y}) &= L(\theta|\mathbf{x})L(\theta,\mathbf{x}|\mathbf{y}) \\ &= \underbrace{(2\pi\sigma_X^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_X^2}\sum_i (x_i - \mu_X)^2\right\}}_{A} \\ &\quad \times\underbrace{\left(2\pi\sigma_Y^2(1-\rho^2)\right)^{-\frac{n}{2}}\exp\left[\frac{-1}{2\sigma_Y^2(1-\rho^2)}\sum_i\left\{y_i - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x_i - \mu_X)\right)\right\}^2\right]}_{B} \end{aligned}$$

We know that $\bar{x}$ and $\hat{\sigma}_X^2 = \sum_i (x_i - \bar{x})^2/n$ maximizes A; the question is whether given σ_Y, μ_Y, and ρ, does $\bar{x}$, $\hat{\sigma}_X^2$ maximize B? Let us first fix σ_X^2 and look for $\hat{\mu}_X$, that maximizes B. We have

$$\begin{aligned} \frac{\partial \log B}{\partial \mu_X} &\propto -2\left(\sum_i \left[(y_i - \mu_Y) - \frac{\rho\sigma_Y}{\sigma_X}(x_i - \mu_X)\right]\right)\frac{\rho\sigma_Y}{\sigma_X} \stackrel{\text{set}}{=} 0 \\ \Rightarrow \sum_i (y_i - \mu_Y) &= \frac{\rho\sigma_Y}{\sigma_X}\Sigma(x_i - \hat{\mu}_X). \end{aligned}$$

Similarly do the same procedure for $L(\theta|\mathbf{y})L(\theta,\mathbf{y}|\mathbf{x})$ This implies $\sum_i (x_i - \mu_X) = \frac{\rho\sigma_X}{\sigma_Y}\sum_i (y_i - \hat{\mu}_Y)$. The solutions $\hat{\mu}_X$ and $\hat{\mu}_Y$ therefore must satisfy both equations. If $\sum_i (y_i - \hat{\mu}_Y) \neq 0$ or $\sum_i (x_i - \hat{\mu}_X) \neq 0$, we will get $\rho = 1/\rho$, so we need $\sum_i (y_i - \hat{\mu}_Y) = 0$ and $\sum_i (x_i - \hat{\mu}_X) = 0$. This implies $\hat{\mu}_X = \bar{x}$ and $\hat{\mu}_Y = \bar{y}$. ($\frac{\partial^2 \log B}{\partial \mu_X^2} < 0$. Therefore it is maximum). To get $\hat{\sigma}_X^2$ take

$$\begin{aligned} \frac{\partial \log B}{\partial \sigma_X^2} &\propto \sum_i \frac{\rho\sigma_Y}{\sigma_X^2}(x_i - \hat{\mu}_X)\left[(y_i - \mu_Y) - \frac{\rho\sigma_Y}{\sigma_X}(x_i - \mu_X)\right] \stackrel{\text{set}}{=} 0. \\ \Rightarrow \sum_i (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) &= \frac{\rho\sigma_Y}{\hat{\sigma}_X}\sum (x_i - \hat{\mu}_X)^2. \end{aligned}$$

Similarly, $\sum_i (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y) = \frac{\rho\sigma_X}{\hat{\sigma}_Y}\sum_i (y_i - \hat{\mu}_Y)^2$. Thus $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ must satisfy the above two equations with $\hat{\mu}_X = \bar{X}$, $\hat{\mu}_Y = \bar{Y}$. This implies

$$\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\sum_i (x_i - \bar{x})^2 = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y}\sum_i (y_i - \bar{y})^2 \Rightarrow \frac{\sum_i (x_i - \bar{x})^2}{\hat{\sigma}_X^2} = \frac{\sum_i (y_i - \bar{y})^2}{\hat{\sigma}_Y^2}.$$

Therefore, $\hat{\sigma}_X^2 = a\sum_i (x_i - \bar{x})^2$, $\hat{\sigma}_Y^2 = a\sum_i (y_i - \bar{y})^2$ where a is a constant. Combining the knowledge that $\left(\bar{x}, \frac{1}{n}\sum_i (x_i - \bar{x})^2\right) = (\hat{\mu}_X, \hat{\sigma}_X^2)$ maximizes A, we conclude that $a = 1/n$. Lastly, we find $\hat{\rho}$, the MLE of ρ. Write

$$\begin{aligned} &\log L(\bar{x}, \bar{y}, \hat{\sigma}_X^2, \hat{\sigma}_Y^2, \rho|\mathbf{x}, \mathbf{y}) \\ &= -\frac{n}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\sum_i \left[\frac{(x_i - \bar{x})^2}{\hat{\sigma}_X^2} - \frac{2\rho(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X, \hat{\sigma}_Y} + \frac{(y_i - \bar{y})^2}{\hat{\sigma}_Y^2}\right] \\ &= -\frac{n}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\left[2n - 2\rho\underbrace{\sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X\hat{\sigma}_Y}}_{A}\right] \end{aligned}$$

because $\hat{\sigma}_X^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$ and $\hat{\sigma}_Y^2 = \frac{1}{n}\sum_i (y_i - \bar{y})^2$. Now

$$\log L = -\frac{n}{2}\log(1-\rho^2) - \frac{n}{1-\rho^2} + \frac{\rho}{1-\rho^2}A$$

and

$$\frac{\partial \log L}{\partial \rho} = \frac{n}{1-\rho^2} - \frac{n\rho}{(1-\rho^2)^2} + \frac{A(1-\rho^2) + 2A\rho^2}{(1-\rho^2)^2} \stackrel{\text{set}}{=} 0.$$

This implies

$$\begin{aligned} \frac{A + A\rho^2 - n\hat{\rho} - n\hat{\rho}^3}{(1-\rho^2)^2} = 0 \quad &\Rightarrow \quad A(1 + \hat{\rho}^2) = n\hat{\rho}(1 + \hat{\rho}^2) \\ &\Rightarrow \quad \hat{\rho} = \frac{A}{n} = \frac{1}{n}\sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X\hat{\sigma}_Y}. \end{aligned}$$

7.19 a.

$$\begin{aligned}
L(\theta|\mathbf{y}) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\beta x_i)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2)\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\beta^2 \sum_i x_i^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i\right).
\end{aligned}$$

By Theorem 6.1.2, $(\sum_i Y_i^2, \sum_i x_i Y_i)$ is a sufficient statistic for (β, σ^2).

b.

$$\log L(\beta,\sigma^2|\mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i - \frac{\beta^2}{2\sigma^2}\sum_i x_i^2.$$

For a fixed value of σ^2,

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2}\sum_i x_i y_i - \frac{\beta}{\sigma^2}\sum_i x_i^2 \overset{\text{set}}{=} 0 \quad\Rightarrow\quad \hat\beta = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

Also,

$$\frac{\partial^2 \log L}{\partial \beta^2} = \frac{1}{\sigma^2}\sum_i x_i^2 < 0,$$

so it is a maximum. Because $\hat\beta$ does not depend on σ^2, it is the MLE. And $\hat\beta$ is unbiased because

$$\mathrm{E}\,\hat\beta = \frac{\sum_i x_i \mathrm{E}\,Y_i}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot \beta x_i}{\sum_i x_i^2} = \beta.$$

c. $\hat\beta = \sum_i a_i Y_i$, where $a_i = x_i / \sum_j x_j^2$ are constants. By Corollary 4.6.10, $\hat\beta$ is normally distributed with mean β, and

$$\mathrm{Var}\,\hat\beta = \sum_i a_i^2 \mathrm{Var}\,Y_i = \sum_i \left(\frac{x_i}{\sum_j x_j^2}\right)^2 \sigma^2 = \frac{\sum_i x_i^2}{(\sum_j x_j^2)^2}\sigma^2 = \frac{\sigma^2}{\sum_i x_i^2}.$$

7.20 a.

$$\mathrm{E}\frac{\sum_i Y_i}{\sum_i x_i} = \frac{1}{\sum_i x_i}\sum_i \mathrm{E}\,Y_i = \frac{1}{\sum_i x_i}\sum_i \beta x_i = \beta.$$

b.

$$\mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right) = \frac{1}{(\sum_i x_i)^2}\sum_i \mathrm{Var}\,Y_i = \frac{\sum_i \sigma^2}{(\sum_i x_i)^2} = \frac{n\sigma^2}{n^2\bar x^2} = \frac{\sigma^2}{n\bar x^2}.$$

Because $\sum_i x_i^2 - n\bar x^2 = \sum_i (x_i - \bar x)^2 \ge 0$, $\sum_i x_i^2 \ge n\bar x^2$. Hence,

$$\mathrm{Var}\,\hat\beta = \frac{\sigma^2}{\sum_i x_i^2} \le \frac{\sigma^2}{n\bar x^2} = \mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right).$$

(In fact, $\hat\beta$ is BLUE (Best Linear Unbiased Estimator of β), as discussed in Section 11.3.2.)

7.21 a.

$$\mathrm{E}\frac{1}{n}\sum_i \frac{Y_i}{x_i} = \frac{1}{n}\sum_i \frac{\mathrm{E}\,Y_i}{x_i} = \frac{1}{n}\sum_i \frac{\beta x_i}{x_i} = \beta.$$

b.

$$\mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i} = \frac{1}{n^2}\sum_i \frac{\mathrm{Var}\,Y_i}{x_i^2} = \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2}.$$

Using Example 4.7.8 with $a_i = 1/x_i^2$ we obtain

$$\frac{1}{n}\sum_i \frac{1}{x_i^2} \geq \frac{n}{\sum_i x_i^2}.$$

Thus,

$$\mathrm{Var}\,\hat{\beta} = \frac{\sigma^2}{\sum_i x_i^2} \leq \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2} = \mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i}.$$

Because $g(u) = 1/u^2$ is convex, using Jensen's Inequality we have

$$\frac{1}{\bar{x}^2} \leq \frac{1}{n}\sum_i \frac{1}{x_i^2}.$$

Thus,

$$\mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right) = \frac{\sigma^2}{n\bar{x}^2} \leq \frac{\sigma^2}{n^2}\sum_i \frac{1}{x_i^2} = \mathrm{Var}\frac{1}{n}\sum_i \frac{Y_i}{x_i}.$$

7.22 a.

$$f(\bar{x},\theta) = f(\bar{x}|\theta)\pi(\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-n(\bar{x}-\theta)^2/(2\sigma^2)}\frac{1}{\sqrt{2\pi}\tau}e^{-(\theta-\mu)^2/2\tau^2}.$$

b. Factor the exponent in part (a) as

$$\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2 = -\frac{1}{2v^2}(\theta-\delta(\mathbf{x}))^2 - \frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2,$$

where $\delta(\mathbf{x}) = (\tau^2\bar{x} + (\sigma^2/n)\mu)/(\tau^2+\sigma^2/n)$ and $v = (\sigma^2\tau^2/n)\big/(\tau+\sigma^2/n)$. Let $\mathrm{n}(a,b)$ denote the pdf of a normal distribution with mean a and variance b. The above factorization shows that

$$f(\mathbf{x},\theta) = \mathrm{n}(\theta,\sigma^2/n)\times\mathrm{n}(\mu,\tau^2) = \mathrm{n}(\delta(\mathbf{x}),v^2)\times\mathrm{n}(\mu,\tau^2+\sigma^2/n),$$

where the marginal distribution of $\bar{X}$ is $\mathrm{n}(\mu,\tau^2+\sigma^2/n)$ and the posterior distribution of $\theta|\mathbf{x}$ is $\mathrm{n}(\delta(\mathbf{x}),v^2)$. This also completes part (c).

7.23 Let $t = s^2$ and $\theta = \sigma^2$. Because $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, we have

$$f(t|\theta) = \frac{1}{\Gamma\left((n-1)/2\right)2^{(n-1)/2}}\left(\frac{n-1}{\theta}t\right)^{[(n-1)/2]-1}e^{-(n-1)t/2\theta}\frac{n-1}{\theta}.$$

With $\pi(\theta)$ as given, we have (ignoring terms that do not depend on θ)

$$\begin{aligned}\pi(\theta|t) &\propto \left[\left(\frac{1}{\theta}\right)^{((n-1)/2)-1}e^{-(n-1)t/2\theta}\frac{1}{\theta}\right]\left[\frac{1}{\theta^{\alpha+1}}e^{-1/\beta\theta}\right]\\ &\propto \left(\frac{1}{\theta}\right)^{((n-1)/2)+\alpha+1}\exp\left\{-\frac{1}{\theta}\left[\frac{(n-1)t}{2}+\frac{1}{\beta}\right]\right\},\end{aligned}$$

which we recognize as the kernel of an inverted gamma pdf, IG(a, b), with

$$a = \frac{n-1}{2} + \alpha \qquad \text{and} \qquad b = \left[\frac{(n-1)t}{2} + \frac{1}{\beta}\right]^{-1}.$$

Direct calculation shows that the mean of an IG(a, b) is $1/((a-1)b)$, so

$$\mathrm{E}(\theta|t) = \frac{\frac{n-1}{2}t + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1} = \frac{\frac{n-1}{2}s^2 + \frac{1}{\beta}}{\frac{n-1}{2} + \alpha - 1}.$$

This is a Bayes estimator of σ^2.

7.24 For n observations, $Y = \sum_i X_i \sim$ Poisson$(n\lambda)$.

a. The marginal pmf of Y is

$$\begin{aligned} m(y) &= \int_0^\infty \frac{(n\lambda)^y e^{-n\lambda}}{y!} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}\, d\lambda \\ &= \frac{n^y}{y!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{(y+\alpha)-1} e^{-\frac{\lambda}{\beta/(n\beta+1)}}\, d\lambda \;=\; \frac{n^y}{y!\Gamma(\alpha)\beta^\alpha} \Gamma(y+\alpha) \left(\frac{\beta}{n\beta+1}\right)^{y+\alpha}. \end{aligned}$$

Thus,

$$\pi(\lambda|y) = \frac{f(y|\lambda)\pi(\lambda)}{m(y)} = \frac{\lambda^{(y+\alpha)-1} e^{-\frac{\lambda}{\beta/(n\beta+1)}}}{\Gamma(y+\alpha)\left(\frac{\beta}{n\beta+1}\right)^{y+\alpha}} \sim \text{gamma}\left(y+\alpha, \frac{\beta}{n\beta+1}\right).$$

b.

$$\begin{aligned} \mathrm{E}(\lambda|y) &= (y+\alpha)\frac{\beta}{n\beta+1} \;=\; \frac{\beta}{n\beta+1} y + \frac{1}{n\beta+1}(\alpha\beta). \\ \mathrm{Var}(\lambda|y) &= (y+\alpha)\frac{\beta^2}{(n\beta+1)^2}. \end{aligned}$$

7.25 a. We will use the results and notation from part (b) to do this special case. From part (b), the X_is are independent and each X_i has marginal pdf

$$m(x|\mu, \sigma^2, \tau^2) = \int_{-\infty}^\infty f(x|\theta, \sigma^2)\pi(\theta|\mu, \tau^2)\, d\theta = \int_{-\infty}^\infty \frac{1}{2\pi\sigma\tau} e^{-(x-\theta)^2/2\sigma^2} e^{-(\theta-\mu)^2/2\tau^2}\, d\theta.$$

Complete the square in θ to write the sum of the two exponents as

$$-\frac{\left(\theta - \left[\frac{x\tau^2}{\sigma^2+\tau^2} + \frac{\mu\sigma^2}{\sigma^2+\tau^2}\right]\right)^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} - \frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}.$$

Only the first term involves θ; call it $-A(\theta)$. Also, $e^{-A(\theta)}$ is the kernel of a normal pdf. Thus,

$$\int_{-\infty}^\infty e^{-A(\theta)}\, d\theta = \sqrt{2\pi}\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}},$$

and the marginal pdf is

$$\begin{aligned} m(x|\mu, \sigma^2, \tau^2) &= \frac{1}{2\pi\sigma\tau}\sqrt{2\pi}\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}} \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+\tau^2}} \exp\left\{-\frac{(x-\mu)^2}{2(\sigma^2+\tau^2)}\right\}, \end{aligned}$$

a n$(\mu, \sigma^2+\tau^2)$ pdf.

b. For one observation of X and θ the joint pdf is

$$h(x,\theta|\tau) = f(x|\theta)\pi(\theta|\tau),$$

and the marginal pdf of X is

$$m(x|\tau) = \int_{-\infty}^{\infty} h(x,\theta|\tau)\,d\theta.$$

Thus, the joint pdf of $\mathbf{X} = (X_1,\ldots,X_n)$ and $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$ is

$$h(\mathbf{x},\boldsymbol{\theta}|\tau) = \prod_i h(x_i,\theta_i|\tau),$$

and the marginal pdf of $\mathbf{X}$ is

$$\begin{aligned} m(\mathbf{x}|\tau) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_i h(x_i,\theta_i|\tau)\,d\theta_1\ldots d\theta_n \\ &= \int_{-\infty}^{\infty}\cdots\left\{\int_{-\infty}^{\infty} h(x_1,\theta_1|\tau)\,d\theta_1\right\}\prod_{i=2}^{n} h(x_i,\theta_i|\tau)\,d\theta_2\ldots d\theta_n. \end{aligned}$$

The $d\theta_1$ integral is just $m(x_1|\tau)$, and this is not a function of $\theta_2,\ldots,\theta_n$. So, $m(x_1|\tau)$ can be pulled out of the integrals. Doing each integral in turn yields the marginal pdf

$$m(\mathbf{x}|\tau) = \prod_i m(x_i|\tau).$$

Because this marginal pdf factors, this shows that marginally $X_1,\ldots,X_n$ are independent, and they each have the same marginal distribution, $m(x|\tau)$.

7.26 First write

$$f(x_1,\ldots,x_n|\theta)\pi(\theta) \propto e^{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2-|\theta|/a}$$

where the exponent can be written

$$\frac{n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{|\theta|}{a} = \frac{n}{2\sigma^2}\left(\theta-\delta_{\pm}(\mathbf{x})\right) + \frac{n}{2\sigma^2}\left(\bar{x}^2-\delta_{\pm}^2(\mathbf{x})\right)$$

with $\delta_{\pm}(\mathbf{x}) = \bar{x} \pm \frac{\sigma^2}{na}$, where we use the "+" if $\theta > 0$ and the "−" if $\theta < 0$. Thus, the posterior mean is

$$\mathrm{E}(\theta|\mathbf{x}) = \frac{\int_{-\infty}^{\infty}\theta e^{-\frac{n}{2\sigma^2}(\theta-\delta_{\pm}(\mathbf{x}))^2}\,d\theta}{\int_{-\infty}^{\infty} e^{-\frac{n}{2\sigma^2}(\theta-\delta_{\pm}(\mathbf{x}))^2}\,d\theta}.$$

Now use the facts that for constants a and b,

$$\begin{aligned} \int_0^{\infty} e^{-\frac{a}{2}(t-b)^2}\,dt &= \int_{-\infty}^{0} e^{-\frac{a}{2}(t-b)^2}\,dt = \sqrt{\frac{\pi}{2a}}, \\ \int_0^{\infty} te^{-\frac{a}{2}(t-b)^2}\,dt &= \int_0^{\infty}(t-b)e^{-\frac{a}{2}(t-b)^2}\,dt + \int_0^{\infty} be^{-\frac{a}{2}(t-b)^2}\,dt = \frac{1}{a}e^{-\frac{a}{2}b^2} + b\sqrt{\frac{\pi}{2a}}, \\ \int_{-\infty}^{0} te^{-\frac{a}{2}(t-b)^2}\,dt &= -\frac{1}{a}e^{-\frac{a}{2}b^2} + b\sqrt{\frac{\pi}{2a}}, \end{aligned}$$

to get

$$\mathrm{E}(\theta|\mathbf{x}) = \frac{\sqrt{\frac{\pi\sigma^2}{2n}}\left(\delta_-(\mathbf{x})+\delta_+(\mathbf{x})\right) + \frac{\sigma^2}{n}\left(e^{-\frac{n}{2\sigma^2}\delta_+^2(\mathbf{x})} - e^{-\frac{n}{2\sigma^2}\delta_-^2(\mathbf{x})}\right)}{2\sqrt{\frac{\pi\sigma^2}{2n}}}.$$

7.27 a. The log likelihood is

$$\log L = \sum_{i=1}^{n} -\beta\tau_i + y_i \log(\beta\tau_i) - \tau_i + x_i \log(\tau_i) - \log y_i! - \log x_i!$$

and differentiation gives

$$\begin{aligned}
\frac{\partial}{\partial\beta}\log L &= \sum_{i=1}^{n} -\tau_i + \frac{y_i\tau_i}{\beta\tau_i} \quad \Rightarrow \quad \beta = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \tau_i} \\
\frac{\partial}{\partial\tau_j}\log L &= -\beta + \frac{y_j\beta}{\beta\tau_j} - i + \frac{x_j}{\tau_j} \quad \Rightarrow \quad \tau_j = \frac{x_j + y_j}{1+\beta} \\
&\Rightarrow \sum_{j=1}^{n}\tau_j = \frac{\sum_{j=1}^n x_j + \sum_{j=1}^n y_j}{1+\beta}.
\end{aligned}$$

Combining these expressions yields $\hat\beta = \sum_{j=1}^n y_j / \sum_{j=1}^n x_j$ and $\hat\tau_j = \frac{x_j+y_j}{1+\hat\beta}$.

b. The stationary point of the EM algorithm will satisfy

$$\begin{aligned}
\hat\beta &= \frac{\sum_{i=1}^n y_i}{\hat\tau_1 + \sum_{i=2}^n x_i} \\
\hat\tau_1 &= \frac{\hat\tau_1 + y_1}{\hat\beta + 1} \\
\hat\tau_j &= \frac{x_j + y_j}{\hat\beta + 1}.
\end{aligned}$$

The second equation yields $\tau_1 = y_1/\beta$, and substituting this into the first equation yields $\beta = \sum_{j=2}^n y_j / \sum_{j=2}^n x_j$. Summing over j in the third equation, and substituting $\beta = \sum_{j=2}^n y_j / \sum_{j=2}^n x_j$ shows us that $\sum_{j=2}^n \hat\tau_j = \sum_{j=2}^n x_j$, and plugging this into the first equation gives the desired expression for $\hat\beta$. The other two equations in (7.2.16) are obviously satisfied.

c. The expression for $\hat\beta$ was derived in part (b), as were the expressions for $\hat\tau_i$.

7.29 a. The joint density is the product of the individual densities.

b. The log likelihood is

$$\log L = \sum_{i=1}^{n} -m\beta\tau_i + y_i \log(m\beta\tau_i) + x_i \log(\tau_i) + \log m! - \log y_i! - \log x_i!$$

and

$$\begin{aligned}
\frac{\partial}{\partial\beta}\log L &= 0 \quad \Rightarrow \quad \beta = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m\tau_i} \\
\frac{\partial}{\partial\tau_j}\log L &= 0 \quad \Rightarrow \quad \tau_j = \frac{x_j + y_j}{m\beta}.
\end{aligned}$$

Since $\sum \tau_j = 1$, $\hat\beta = \sum_{i=1}^n y_i/m = \sum_{i=1}^n y_i / \sum_{i=1}^n x_i$. Also, $\sum_j \tau_j = \sum_j (y_j + x_j) = 1$, which implies that $m\beta = \sum_j (y_j + x_j)$ and $\hat\tau_j = (x_j + y_j)/\sum_i (y_i + x_i)$.

c. In the likelihood function we can ignore the factorial terms, and the expected complete-data likelihood is obtained by on the r^{th} iteration by replacing x_1 with $\mathrm{E}(X_1|\hat\tau_1^{(r)}) = m\hat\tau_1^{(r)}$. Substituting this into the MLEs of part (b) gives the EM sequence.

The MLEs from the full data set are $\hat{\beta} = 0.0008413892$ and

$$\begin{aligned}\hat{\tau} &= (0.06337310, 0.06374873, 0.06689681, 0.04981487, 0.04604075, 0.04883109,\\ &\quad 0.07072460, 0.01776164, 0.03416388, 0.01695673, 0.02098127, 0.01878119,\\ &\quad 0.05621836, 0.09818091, 0.09945087, 0.05267677, 0.08896918, 0.08642925).\end{aligned}$$

The MLEs for the incomplete data were computed using R, where we take $m = \sum x_i$. The R code is

```
#mles on the incomplete data#
xdatam<-c(3560,3739,2784,2571,2729,3952,993,1908,948,1172,
    1047,3138,5485,5554,2943,4969,4828)
ydata<-c(3,4,1,1,3,1,2,0,2,0,1,3,5,4,6,2,5,4)
xdata<-c(mean(xdatam),xdatam); for (j in 1:500) {
xdata<-c(sum(xdata)*tau[1],xdatam) beta<-sum(ydata)/sum(xdata)
tau<-c((xdata+ydata)/(sum(xdata)+sum(ydata))) } beta tau
```

The MLEs from the incomplete data set are $\hat{\beta} = 0.0008415534$ and

$$\begin{aligned}\hat{\tau} &= (0.06319044, 0.06376116, 0.06690986, 0.04982459, 0.04604973, 0.04884062,\\ &\quad 0.07073839, 0.01776510, 0.03417054, 0.01696004, 0.02098536, 0.01878485,\\ &\quad 0.05622933, 0.09820005, 0.09947027, 0.05268704, 0.08898653, 0.08644610).\end{aligned}$$

7.31 a. By direct substitution we can write

$$\log L(\theta|\mathbf{y}) = \mathrm{E}\left[\log L(\theta|\mathbf{y},\mathbf{X})|\,\hat{\theta}^{(r)},\mathbf{y}\right] - \mathrm{E}\left[\log k(\mathbf{X}|\theta,\mathbf{y})|\,\hat{\theta}^{(r)},\mathbf{y}\right].$$

The next iterate, $\hat{\theta}^{(r+1)}$ is obtained by maximizing the expected complete-data log likelihood, so for any θ, $\mathrm{E}\left[\log L(\hat{\theta}^{(r+1)}\mathbf{y},\mathbf{X})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right] \geq \mathrm{E}\left[\log L(\theta|\mathbf{y},\mathbf{X})|\,\hat{\theta}^{(r)},\mathbf{y}\right]$

b. Write

$$\mathrm{E}\left[\log k(\mathbf{X}|\theta,\mathbf{y})|\theta',\mathbf{y}\right] = \int \log k(\mathbf{x}|\theta,\mathbf{y}) \log k(\mathbf{x}|\theta',\mathbf{y})d\mathbf{x} \leq \int \log k(\mathbf{x}|\theta',\mathbf{y}) \log k(\mathbf{x}|\theta',\mathbf{y})d\mathbf{x},$$

from the hint. Hence $\mathrm{E}\left[\log k(\mathbf{X}|\hat{\theta}^{(r+1)},\mathbf{y})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right] \leq \mathrm{E}\left[\log k(\mathbf{X}|\hat{\theta}^{(r)},\mathbf{y})\middle|\,\hat{\theta}^{(r)},\mathbf{y}\right]$, and so the entire right hand side in part (a) is decreasing.

7.33 Substitute $\alpha = \beta = \sqrt{n/4}$ into $\mathrm{MSE}(\hat{p}_B) = \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2$ and simplify to obtain

$$\mathrm{MSE}(\hat{p}_B) = \frac{n}{4(\sqrt{n}+n)^2},$$

independent of p, as desired.

7.35 a.

$$\begin{aligned}\delta_p\left(g(\mathbf{x})\right) &= \delta_p(x_1+a,\ldots,x_n+a)\\ &= \frac{\int_{-\infty}^{\infty} t\prod_i f(x_i+a-t)\,dt}{\int_{-\infty}^{\infty}\prod_i f(x_i+a-t)\,dt} = \frac{\int_{-\infty}^{\infty}(y+a)\prod_i f(x_i-y)\,dy}{\int_{-\infty}^{\infty}\prod_i f(x_i-y)\,dy} \qquad (y=t-a)\\ &= a+\delta_p(\mathbf{x}) = \bar{g}\left(\delta_p(\mathbf{x})\right).\end{aligned}$$

b.

$$\prod_i f(x_i - t) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\Sigma_i(x_i-t)^2} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}n(\bar{x}-t)^2} e^{-\frac{1}{2}(n-1)s^2},$$

so

$$\delta_p(\mathbf{x}) = \frac{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} te^{-\frac{1}{2}n(\bar{x}-t)^2} dt}{(\sqrt{n}/\sqrt{2\pi}) \int_{-\infty}^{\infty} e^{-\frac{1}{2}n(\bar{x}-t)^2} dt} = \frac{\bar{x}}{1} = \bar{x}.$$

c.

$$\prod_i f(x_i - t) = \prod_i I\left(t - \frac{1}{2} \le x_i \le t + \frac{1}{2}\right) = I\left(x_{(n)} - \frac{1}{2} \le t \le x_{(1)} + \frac{1}{2}\right),$$

so

$$\delta_p(\mathbf{x}) = \frac{\int_{x_{(n)}+1/2}^{x_{(1)}+1/2} t\, dt}{\int_{x_{(n)}+1/2}^{x_{(1)}+1/2} 1\, dt} = \frac{x_{(1)} + x_{(n)}}{2}.$$

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 $\left(X_{(1)}, X_{(n)}\right)$ 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.)

7.38 Use Corollary 7.3.15.

a.

$$\begin{aligned} \frac{\partial}{\partial\theta} \log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial\theta} \log \prod_i \theta x_i^{\theta-1} = \frac{\partial}{\partial\theta} \sum_i [\log\theta + (\theta - 1) \log x_i] \\ &= \sum_i \left[\frac{1}{\theta} + \log x_i\right] = -n\left[-\sum_i \frac{\log x_i}{n} - \frac{1}{\theta}\right]. \end{aligned}$$

Thus, $-\sum_i \log X_i/n$ is the UMVUE of $1/\theta$ and attains the Cramér-Rao bound.

b.

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial\theta}\log\prod_i \frac{\log\theta}{\theta-1}\theta^{x_i} \quad = \quad \frac{\partial}{\partial\theta}\sum_i \left[\log\log\theta - \log(\theta-1) + x_i\log\theta\right] \\
&= \sum_i \left(\frac{1}{\theta\log\theta} - \frac{1}{\theta-1}\right) + \frac{1}{\theta}\sum_i x_i \quad = \quad \frac{n}{\theta\log\theta} - \frac{n}{\theta-1} + \frac{n\bar{x}}{\theta} \\
&= \frac{n}{\theta}\left[\bar{x} - \left(\frac{\theta}{\theta-1} - \frac{1}{\log\theta}\right)\right].
\end{aligned}$$

Thus, $\bar{X}$ is the UMVUE of $\frac{\theta}{\theta-1} - \frac{1}{\log\theta}$ and attains the Cramér-Rao lower bound.

Note: We claim that if $\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{X}) = a(\theta)[W(\mathbf{X}) - \tau(\theta)]$, then $\mathrm{E}\,W(\mathbf{X}) = \tau(\theta)$, because under the condition of the Cramér-Rao Theorem, $\mathrm{E}\,\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) = 0$. To be rigorous, we need to check the "interchange differentiation and integration" condition. Both (a) and (b) are exponential families, and this condition is satisfied for all exponential families.

7.39

$$\begin{aligned}
\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X}|\theta)\right] &= E_\theta\left[\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)\right] \\
&= \mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right)\right] \quad = \quad \mathrm{E}_\theta\left[\frac{\frac{\partial^2}{\partial\theta^2}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)} - \left(\frac{\frac{\partial}{\partial\theta}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right)^2\right].
\end{aligned}$$

Now consider the first term:

$$\begin{aligned}
\mathrm{E}_\theta\left[\frac{\frac{\partial^2}{\partial\theta^2}f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)}\right] &= \int\left[\frac{\partial^2}{\partial\theta^2}f(\mathbf{x}|\theta)\right]d\mathbf{x} \quad = \quad \frac{d}{d\theta}\int\frac{\partial}{\partial\theta}f(\mathbf{x}|\theta)\,d\mathbf{x} \qquad \text{(assumption)} \\
&= \frac{d}{d\theta}\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right] \quad = \quad 0, \qquad \text{(7.3.8)}
\end{aligned}$$

and the identity is proved.

7.40

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x}) &= \frac{\partial}{\partial p}\log\prod_i p^{x_i}(1-p)^{1-x_i} \quad = \quad \frac{\partial}{\partial p}\sum_i x_i\log p + (1-x_i)\log(1-p) \\
&= \sum_i\left[\frac{x_i}{p} - \frac{(1-x_i)}{1-p}\right] \quad = \quad \frac{n\bar{x}}{p} - \frac{n-n\bar{x}}{1-p} \quad = \quad \frac{n}{p(1-p)}[\bar{x}-p].
\end{aligned}$$

By Corollary 7.3.15, $\bar{X}$ is the UMVUE of p and attains the Cramér-Rao lower bound. Alternatively, we could calculate

$$\begin{aligned}
&-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right) \\
&= -n\mathrm{E}\left(\frac{\partial^2}{\partial p^2}\log\left[p^X(1-p)^{1-X}\right]\right) \quad = \quad -n\mathrm{E}\left(\frac{\partial^2}{\partial p^2}\left[X\log p + (1-X)\log(1-p)\right]\right) \\
&= -n\mathrm{E}\left(\frac{\partial}{\partial p}\left[\frac{X}{p} - \frac{(1-X)}{1-p}\right]\right) \quad = \quad -n\mathrm{E}\left(\frac{-X}{p^2} - \frac{1-X}{(1-p)^2}\right) \\
&= -n\left(-\frac{1}{p} - \frac{1}{1-p}\right) \quad = \quad \frac{n}{p(1-p)}.
\end{aligned}$$

Then using $\tau(\theta) = p$ and $\tau'(\theta) = 1$,

$$\frac{\tau'(\theta)}{-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} = \frac{1}{n/p(1-p)} = \frac{p(1-p)}{n} = \mathrm{Var}\bar{X}.$$

We know that $\mathrm{E}\bar{X} = p$. Thus, $\bar{X}$ attains the Cramér-Rao bound.

7.41 a. $\mathrm{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathrm{E}\, X_i = \sum_i a_i\mu = \mu\sum_i a_i = \mu$. Hence the estimator is unbiased.

b. $\mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathrm{Var}\, X_i = \sum_i a_i^2\sigma^2 = \sigma^2\sum_i a_i^2$. Therefore, we need to minimize $\sum_i a_i^2$, subject to the constraint $\sum_i a_i = 1$. Add and subtract the mean of the a_i, $1/n$, to get

$$\sum_i a_i^2 = \sum_i \left[\left(a_i - \frac{1}{n}\right) + \frac{1}{n}\right]^2 = \sum_i \left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n},$$

because the cross-term is zero. Hence, $\sum_i a_i^2$ is minimized by choosing $a_i = 1/n$ for all i. Thus, $\sum_i (1/n)X_i = \bar{X}$ has the minimum variance among all linear unbiased estimators.

7.43 a. This one is real hard - it was taken from an *American Statistician* article, but the proof is not there. A cryptic version of the proof is in Tukey (Approximate Weights, *Ann. Math. Statist.* 1948, 91-92); here is a more detailed version.

Let $q_i = q_i^*(1+\lambda t_i)$ with $0 \le \lambda \le 1$ and $|t_i| \le 1$. Recall that $q_i^* = (1/\sigma_i^2)/\sum_j(1/\sigma_j^2)$ and $\mathrm{Var}W^* = 1/\sum_j(1/\sigma_j^2)$. Then

$$\begin{aligned}
\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) &= \frac{1}{(\sum_j q_j)^2}\sum_i q_i\sigma_i^2 \\
&= \frac{1}{[\sum_j q_j^*(1+\lambda t_j)]^2}\sum_i q_i^{*2}(1+\lambda t_i)^2\sigma_i^2 \\
&= \frac{1}{[\sum_j q_j^*(1+\lambda t_j)]^2\sum_j(1/\sigma_j^2)}\sum_i q_i^*(1+\lambda t_i)^2,
\end{aligned}$$

using the definition of q_i^*. Now write

$$\sum_i q_i^*(1+\lambda t_i)^2 = 1 + 2\lambda\sum_j q_j t_j + \lambda^2\sum_j q_j t_j^2 = [1+\lambda\sum_j q_j t_j]^2 + \lambda^2[\sum_j q_j t_j^2 - (\sum_j q_j t_j)^2],$$

where we used the fact that $\sum_j q_j^* = 1$. Now since

$$[\sum_j q_j^*(1+\lambda t_j)]^2 = [1+\lambda\sum_j q_j t_j]^2,$$

$$\begin{aligned}
\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) &= \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[\sum_j q_j t_j^2 - (\sum_j q_j t_j)^2]}{[1+\lambda\sum_j q_j t_j]^2}\right] \\
&\le \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[1 - (\sum_j q_j t_j)^2]}{[1+\lambda\sum_j q_j t_j]^2}\right],
\end{aligned}$$

since $\sum_j q_j t_j^2 \le 1$. Now let $T = \sum_j q_j t_j$, and

$$\mathrm{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) \le \frac{1}{\sum_j(1/\sigma_j^2)}\left[1 + \frac{\lambda^2[1-T^2]}{[1+\lambda T]^2}\right],$$

and the right hand side is maximized at $T = -\lambda$, with maximizing value

$$\operatorname{Var}\left(\frac{q_i W_i}{\sum_j q_j}\right) \leq \frac{1}{\sum_j (1/\sigma_j^2)}\left[1+\frac{\lambda^2[1-\lambda^2]}{[1-\lambda^2]^2}\right] = \operatorname{Var} W^* \frac{1}{1-\lambda^2}.$$

Bloch and Moses (1988) define λ as the solution to

$$b_{\max}/b_{\min} = \frac{1+\lambda}{1-\lambda},$$

where b_i/b_j are the ratio of the normalized weights which, in the present notation, is

$$b_i/b_j = (1+\lambda t_i)/(1+\lambda t_j).$$

The right hand side is maximized by taking t_i as large as possible and t_j as small as possible, and setting $t_i = 1$ and $t_j = -1$ (the extremes) yields the Bloch and Moses (1988) solution.

b.

$$b_i = \frac{1/k}{(1/\sigma_i^2)\Big/\left(\sum_j 1/\sigma_j^2\right)} = \frac{\sigma_i^2}{k}\sum_j 1/\sigma_j^2.$$

Thus,

$$b_{\max} = \frac{\sigma_{\max}^2}{k}\sum_j 1/\sigma_j^2 \qquad \text{and} \qquad b_{\min} = \frac{\sigma_{\min}^2}{k}\sum_j 1/\sigma_j^2$$

and $B = b_{\max}/b_{\min} = \sigma_{\max}^2/\sigma_{\min}^2$. Solving $B = (1+\lambda)/(1-\lambda)$ yields $\lambda = (B-1)/(B+1)$. Substituting this into Tukey's inequality yields

$$\frac{\operatorname{Var} W}{\operatorname{Var} W^*} \leq \frac{(B+1)^2}{4B} = \frac{((\sigma_{\max}^2/\sigma_{\min}^2)+1)^2}{4(\sigma_{\max}^2/\sigma_{\min}^2)}.$$

7.44 $\sum_i X_i$ is a complete sufficient statistic for θ when $X_i \sim \mathrm{n}(\theta, 1)$. $\bar{X}^2 - 1/n$ is a function of $\sum_i X_i$. Therefore, by Theorem 7.3.23, $\bar{X}^2 - 1/n$ is the unique best unbiased estimator of its expectation.

$$\mathrm{E}\left(\bar{X}^2 - \frac{1}{n}\right) = \operatorname{Var}\bar{X} + (\mathrm{E}\,\bar{X})^2 - \frac{1}{n} = \frac{1}{n} + \theta^2 - \frac{1}{n} = \theta^2.$$

Therefore, $\bar{X}^2 - 1/n$ is the UMVUE of θ^2. We will calculate

$$\operatorname{Var}\left(\bar{X}^2 - 1/n\right) = \operatorname{Var}(\bar{X}^2) = \mathrm{E}(\bar{X}^4) - [\mathrm{E}(\bar{X}^2)]^2, \quad \text{where } \bar{X} \sim \mathrm{n}\left(\theta, 1/n\right),$$

but first we derive some general formulas that will also be useful in later exercises. Let $Y \sim \mathrm{n}(\theta, \sigma^2)$. Then here are formulas for $\mathrm{E}\,Y^4$ and $\operatorname{Var} Y^2$.

$$\begin{aligned}
\mathrm{E}\,Y^4 &= \mathrm{E}[Y^3(Y-\theta+\theta)] = \mathrm{E}\,Y^3(Y-\theta) + \mathrm{E}\,Y^3\theta = E\,Y^3(Y-\theta) + \theta\mathrm{E}\,Y^3. \\
\mathrm{E}\,Y^3(Y-\theta) &= \sigma^2\mathrm{E}(3Y^2) = \sigma^2 3\left(\sigma^2+\theta^2\right) = 3\sigma^4 + 3\theta^2\sigma^2. && \text{(Stein's Lemma)} \\
\theta\mathrm{E}\,Y^3 &= \theta\left(3\theta\sigma^2+\theta^3\right) = 3\theta^2\sigma^2 + \theta^4. && \text{(Example 3.6.6)} \\
\operatorname{Var} Y^2 &= 3\sigma^4 + 6\theta^2\sigma^2 + \theta^4 - (\sigma^2+\theta^2)^2 = 2\sigma^4 + 4\theta^2\sigma^2.
\end{aligned}$$

Thus,

$$\operatorname{Var}\left(\bar{X}^2 - \frac{1}{n}\right) = \operatorname{Var}\bar{X}^2 = 2\frac{1}{n^2} + 4\theta^2\frac{1}{n} > \frac{4\theta^2}{n}.$$

To calculate the Cramér-Rao lower bound, we have

$$\begin{aligned}
\mathrm{E}_\theta\left(\frac{\partial^2 \log f(X|\theta)}{\partial\theta^2}\right) &= \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log\frac{1}{\sqrt{2\pi}}e^{-(X-\theta)^2/2}\right) \\
&= \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\left[\log(2\pi)^{-1/2}-\frac{1}{2}(X-\theta)^2\right]\right) = \mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}(X-\theta)\right) = -1,
\end{aligned}$$

and $\tau(\theta) = \theta^2$, $[\tau'(\theta)]^2 = (2\theta)^2 = 4\theta^2$ so the Cramér-Rao Lower Bound for estimating θ^2 is

$$\frac{[\tau'(\theta)]^2}{-n\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} = \frac{4\theta^2}{n}.$$

Thus, the UMVUE of θ^2 does not attain the Cramér-Rao bound. (However, the ratio of the variance and the lower bound $\to 1$ as $n \to \infty$.)

7.45 a. Because $\mathrm{E}\, S^2 = \sigma^2$, $\mathrm{bias}(aS^2) = \mathrm{E}(aS^2) - \sigma^2 = (a-1)\sigma^2$. Hence,

$$\mathrm{MSE}(aS^2) = \mathrm{Var}(aS^2) + \mathrm{bias}(aS^2)^2 = a^2\mathrm{Var}(S^2) + (a-1)^2\sigma^4.$$

b. There were two typos in early printings; $\kappa = \mathrm{E}[X-\mu]^4/\sigma^4$ and

$$\mathrm{Var}(S^2) = \frac{1}{n}\left(\kappa - \frac{n-3}{n-1}\right)\sigma^4.$$

See Exercise 5.8b for the proof.

c. There was a typo in early printings; under normality $\kappa = 3$. Under normality we have

$$\kappa = \frac{\mathrm{E}[X-\mu]^4}{\sigma^4} = \mathrm{E}\left[\frac{X-\mu}{\sigma}\right]^4 = \mathrm{E}\, Z^4,$$

where $Z \sim \mathrm{n}(0,1)$. Now, using Lemma 3.6.5 with $g(z) = z^3$ we have

$$\kappa = \mathrm{E}\, Z^4 = \mathrm{E}\, g(Z)Z = 1\mathrm{E}(3Z^2) = 3\mathrm{E}\, Z^2 = 3.$$

To minimize $\mathrm{MSE}(S^2)$ in general, write $\mathrm{Var}(S^2) = B\sigma^4$. Then minimizing $\mathrm{MSE}(S^2)$ is equivalent to minimizing $a^2B + (a-1)^2$. Set the derivative of this equal to 0 (B is not a function of a) to obtain the minimizing value of a is $1/(B+1)$. Using the expression in part (b), under normality the minimizing value of a is

$$\frac{1}{B+1} = \frac{1}{\frac{1}{n}\left(3 - \frac{n-3}{n-1}\right)+1} = \frac{n-1}{n+1}.$$

d. There was a typo in early printings; the minimizing a is

$$a = \frac{n-1}{(n+1) + \frac{(\kappa-3)(n-1)}{n}}.$$

To obtain this simply calculate $1/(B+1)$ with (from part (b))

$$B = \frac{1}{n}\left(\kappa - \frac{n-3}{n-1}\right).$$

e. Using the expression for a in part (d), if $\kappa = 3$ the second term in the denominator is zero and $a = (n-1)/(n+1)$, the normal result from part (c). If $\kappa < 3$, the second term in the denominator is negative. Because we are dividing by a smaller value, we have $a > (n-1)/(n+1)$. Because $\mathrm{Var}(S^2) = B\sigma^4$, $B > 0$, and, hence, $a = 1/(B+1) < 1$. Similarly, if $\kappa > 3$, the second term in the denominator is positive. Because we are dividing by a larger value, we have $a < (n-1)/(n+1)$.

7.46 a. For the uniform$(\theta, 2\theta)$ distribution we have $\mathrm{E}\,X = (2\theta+\theta)/2 = 3\theta/2$. So we solve $3\theta/2 = \bar{X}$ for θ to obtain the method of moments estimator $\tilde{\theta} = 2\bar{X}/3$.

b. Let $x_{(1)}, \ldots, x_{(n)}$ denote the observed order statistics. Then, the likelihood function is

$$L(\theta|\mathbf{x}) = \frac{1}{\theta^n} I_{[x_{(n)}/2,\, x_{(1)}]}(\theta).$$

Because $1/\theta^n$ is decreasing, this is maximized at $\hat{\theta} = x_{(n)}/2$. So $\hat{\theta} = X_{(n)}/2$ is the MLE. Use the pdf of $X_{(n)}$ to calculate $\mathrm{E}\,X_{(n)} = \frac{2n+1}{n+1}\theta$. So $\mathrm{E}\,\hat{\theta} = \frac{2n+1}{2n+2}\theta$, and if $k = (2n+2)/(2n+1)$, $\mathrm{E}\,k\hat{\theta} = \theta$.

c. From Exercise 6.23, a minimal sufficient statistic for θ is $(X_{(1)}, X_{(n)})$. $\tilde{\theta}$ is not a function of this minimal sufficient statistic. So by the Rao-Blackwell Theorem, $\mathrm{E}(\tilde{\theta}|X_{(1)}, X_{(n)})$ is an unbiased estimator of θ ($\tilde{\theta}$ is unbiased) with smaller variance than $\tilde{\theta}$. The MLE is a function of $(X_{(1)}, X_{(n)})$, so it can not be improved with the Rao-Blackwell Theorem.

d. $\tilde{\theta} = 2(1.16)/3 = .7733$ and $\hat{\theta} = 1.33/2 = .6650$.

7.47 $X_i \sim \mathrm{n}(r, \sigma^2)$, so $\bar{X} \sim \mathrm{n}(r, \sigma^2/n)$ and $\mathrm{E}\,\bar{X}^2 = r^2 + \sigma^2/n$. Thus $\mathrm{E}\left[(\pi\bar{X}^2 - \pi\sigma^2/n)\right] = \pi r^2$ is best unbiased because $\bar{X}$ is a complete sufficient statistic. If σ^2 is unknown replace it with s^2 and the conclusion still holds.

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 $Y \sim \text{exponential}(\lambda/n)$ 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.

7.50 a. $\mathrm{E}(a\bar{X} + (1-a)cS) = a\mathrm{E}\,\bar{X} + (1-a)\mathrm{E}(cS) = a\theta + (1-a)\theta = \theta$. So $a\bar{X} + (1-a)cS$ is an unbiased estimator of θ.

b. Because $\bar{X}$ and S^2 are independent for this normal model, $\mathrm{Var}(a\bar{X}+(1-a)cS) = a^2V_1+(1-a)^2V_2$, where $V_1 = \mathrm{Var}\bar{X} = \theta^2/n$ and $V_2 = \mathrm{Var}(cS) = c^2\mathrm{E}\,S^2 - \theta^2 = c^2\theta^2 - \theta^2 = (c^2-1)\theta^2$. Use calculus to show that this quadratic function of a is minimized at

$$a = \frac{V_2}{V_1+V_2} = \frac{(c^2-1)\theta^2}{((1/n)+c^2-1)\theta^2} = \frac{(c^2-1)}{((1/n)+c^2-1)}.$$

c. Use the factorization in Example 6.2.9, with the special values $\mu = \theta$ and $\sigma^2 = \theta^2$, to show that $(\bar{X}, S^2)$ is sufficient. $\mathrm{E}(\bar{X} - cS) = \theta - \theta = 0$, for all θ. So $\bar{X} - cS$ is a nonzero function of $(\bar{X}, S^2)$ whose expected value is always zero. Thus $(\bar{X}, S^2)$ is not complete.

7.51 a. Straightforward calculation gives:

$$\mathrm{E}\left[\theta - (a_1\bar{X} + a_2cS)\right]^2 = a_1^2\mathrm{Var}\,\bar{X} + a_2^2c^2\mathrm{Var}\,S + \theta^2(a_1+a_2-1)^2.$$

Because $\mathrm{Var}\,\bar{X} = \theta^2/n$ and $\mathrm{Var}\,S = \mathrm{E}\,S^2 - (\mathrm{E}\,S)^2 = \theta^2\left(\frac{c^2-1}{c^2}\right)$, we have

$$\mathrm{E}\left[\theta - (a_1\bar{X} + a_2cS)\right]^2 = \theta^2\left[a_1^2\big/n + a_2^2(c^2-1) + (a_1+a_2-1)^2\right],$$

and we only need minimize the expression in square brackets, which is independent of θ. Differentiating yields $a_2 = \left[(n+1)c^2 - n\right]^{-1}$ and $a_1 = 1 - \left[(n+1)c^2 - n\right]^{-1}$.

b. The estimator T^* has minimum MSE over a class of estimators that contain those in Exercise 7.50.

c. Because $\theta > 0$, restricting $T^* \geq 0$ will improve the MSE.

d. No. It does not fit the definition of either one.

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 $\mathrm{Var}\,T'(X) > \mathrm{Var}\,T(X)$.

7.53 Let a be a constant and suppose $\mathrm{Cov}_{\theta_0}(W,U) > 0$. Then

$$\mathrm{Var}_{\theta_0}(W + aU) = \mathrm{Var}_{\theta_0} W + a^2 \mathrm{Var}_{\theta_0} U + 2a\mathrm{Cov}_{\theta_0}(W,U).$$

Choose $a \in \left(-2\mathrm{Cov}_{\theta_0}(W,U)\Big/\mathrm{Var}_{\theta_0}U, 0\right)$. Then $\mathrm{Var}_{\theta_0}(W + aU) < \mathrm{Var}_{\theta_0} W$, so W cannot be best unbiased.

7.55 All three parts can be solved by this general method. Suppose $X \sim f(x|\theta) = c(\theta)m(x)$, $a < x < \theta$. Then $1/c(\theta) = \int_a^\theta m(x)\,dx$, and the cdf of X is $F(x) = c(\theta)/c(x)$, $a < x < \theta$. Let $Y = X_{(n)}$ be the largest order statistic. Arguing as in Example 6.2.23 we see that Y is a complete sufficient statistic. Thus, any function $T(Y)$ that is an unbiased estimator of $h(\theta)$ is the best unbiased estimator of $h(\theta)$. By Theorem 5.4.4 the pdf of Y is $g(y|\theta) = nm(y)c(\theta)^n/c(y)^{n-1}$, $a < y < \theta$. Consider the equations

$$\int_a^\theta f(x|\theta)\,dx = 1 \qquad \text{and} \qquad \int_a^\theta T(y)g(y|\theta)\,dy = h(\theta),$$

which are equivalent to

$$\int_a^\theta m(x)\,dx = \frac{1}{c(\theta)} \qquad \text{and} \qquad \int_a^\theta \frac{T(y)nm(y)}{c(y)^{n-1}}\,dy = \frac{h(\theta)}{c(\theta)^n}.$$

Differentiating both sides of these two equations with respect to θ and using the Fundamental Theorem of Calculus yields

$$m(\theta) = -\frac{c'(\theta)}{c(\theta)^2} \qquad \text{and} \qquad \frac{T(\theta)nm(\theta)}{c(\theta)^{n-1}} = \frac{c(\theta)^n h'(\theta) - h(\theta)nc(\theta)^{n-1}c'(\theta)}{c(\theta)^{2n}}.$$

Change θs to ys and solve these two equations for $T(y)$ to get the best unbiased estimator of $h(\theta)$ is

$$T(y) = h(y) + \frac{h'(y)}{nm(y)c(y)}.$$

For $h(\theta) = \theta^r$, $h'(\theta) = r\theta^{r-1}$.

a. For this pdf, $m(x) = 1$ and $c(\theta) = 1/\theta$. Hence

$$T(y) = y^r + \frac{ry^{r-1}}{n(1/y)} = \frac{n+r}{n}y^r.$$

b. If θ is the lower endpoint of the support, the smallest order statistic $Y = X_{(1)}$ is a complete sufficient statistic. Arguing as above yields the best unbiased estimator of $h(\theta)$ is

$$T(y) = h(y) - \frac{h'(y)}{nm(y)c(y)}.$$

For this pdf, $m(x) = e^{-x}$ and $c(\theta) = e^\theta$. Hence

$$T(y) = y^r - \frac{ry^{r-1}}{ne^{-y}e^y} = y^r - \frac{ry^{r-1}}{n}.$$

c. For this pdf, $m(x) = e^{-x}$ and $c(\theta) = 1/(e^{-\theta} - e^{-b})$. Hence

$$T(y) = y^r - \frac{ry^{r-1}}{ne^{-y}}(e^{-y} - e^{-b}) = y^r - \frac{ry^{r-1}(1 - e^{-(b-y)})}{n}.$$

7.56 Because T is sufficient, $\phi(T) = \mathrm{E}[h(X_1,\ldots,X_n)|T]$ is a function only of T. That is, $\phi(T)$ is an estimator. If $\mathrm{E}\,h(X_1,\ldots,X_n) = \tau(\theta)$, then

$$\mathrm{E}\,h(X_1,\cdots,X_n) = \mathrm{E}\left[\mathrm{E}\left(h(X_1,\ldots,X_n)|\,T\right)\right] = \tau(\theta),$$

so $\phi(T)$ is an unbiased estimator of $\tau(\theta)$. By Theorem 7.3.23, $\phi(T)$ is the best unbiased estimator of $\tau(\theta)$.

7.57 a. T is a Bernoulli random variable. Hence,

$$\mathrm{E}_p T = P_p(T=1) = P_p\left(\sum_{i=1}^n X_i > X_{n+1}\right) = h(p).$$

b. $\sum_{i=1}^{n+1} X_i$ is a complete sufficient statistic for θ, so $\mathrm{E}\left(T \,\middle|\, \sum_{i=1}^{n+1} X_i\right)$ is the best unbiased estimator of $h(p)$. We have

$$\begin{aligned}\mathrm{E}\left(T \,\middle|\, \sum_{i=1}^{n+1} X_i = y\right) &= P\left(\sum_{i=1}^n X_i > X_{n+1} \,\middle|\, \sum_{i=1}^{n+1} X_i = y\right)\\ &= P\left(\sum_{i=1}^n X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y\right) \Big/ P\left(\sum_{i=1}^{n+1} X_i = y\right).\end{aligned}$$

The denominator equals $\binom{n+1}{y} p^y(1-p)^{n+1-y}$. If $y = 0$ the numerator is

$$P\left(\sum_{i=1}^n X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = 0\right) = 0.$$

If $y > 0$ the numerator is

$$P\left(\sum_{i=1}^n X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 0\right) + P\left(\sum_{i=1}^n X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 1\right)$$

which equals

$$P\left(\sum_{i=1}^n X_i > 0, \sum_{i=1}^n X_i = y\right) P(X_{n+1} = 0) + P\left(\sum_{i=1}^n X_i > 1, \sum_{i=1}^n X_i = y-1\right) P(X_{n+1} = 1).$$

For all $y > 0$,

$$P\left(\sum_{i=1}^n X_i > 0, \sum_{i=1}^n X_i = y\right) = P\left(\sum_{i=1}^n X_i = y\right) = \binom{n}{y} p^y (1-p)^{n-y}.$$

If $y = 1$ or 2, then

$$P\left(\sum_{i=1}^n X_i > 1, \sum_{i=1}^n X_i = y-1\right) = 0.$$

And if $y > 2$, then

$$P\left(\sum_{i=1}^n X_i > 1, \sum_{i=1}^n X_i = y-1\right) = P\left(\sum_{i=1}^n X_i = y-1\right) = \binom{n}{y-1} p^{y-1}(1-p)^{n-y+1}.$$

Therefore, the UMVUE is

$$\mathrm{E}\left(T \middle| \sum_{i=1}^{n+1} X_i = y\right) = \begin{cases} 0 & \text{if } y = 0 \\ \frac{\binom{n}{y}p^y(1-p)^{n-y}(1-p)}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}}{\binom{n+1}{y}} = \frac{1}{(n+1)(n+1-y)} & \text{if } y = 1 \text{ or } 2 \\ \frac{\left(\binom{n}{y}+\binom{n}{y-1}\right)p^y(1-p)^{n-y+1}}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}+\binom{n}{y-1}}{\binom{n+1}{y}} = 1 & \text{if } y > 2. \end{cases}$$

7.59 We know $T = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Then

$$\mathrm{E}\,T^{p/2} = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}} \int_0^\infty t^{\frac{p+n-1}{2}-1} e^{-\frac{t}{2}}\,dt = \frac{2^{\frac{p}{2}}\Gamma\left(\frac{p+n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} = C_{p,n}.$$

Thus

$$\mathrm{E}\left(\frac{(n-1)S^2}{\sigma^2}\right)^{p/2} = C_{p,n},$$

so $(n-1)^{p/2}S^p \Big/ C_{p,n}$ is an unbiased estimator of σ^p. From Theorem 6.2.25, $(\bar{X}, S^2)$ is a complete, sufficient statistic. The unbiased estimator $(n-1)^{p/2}S^p \Big/ C_{p,n}$ is a function of $(\bar{X}, S^2)$. Hence, it is the best unbiased estimator.

7.61 The pdf for $Y \sim \chi^2_\nu$ is

$$f(y) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} y^{\nu/2-1} e^{-y/2}.$$

Thus the pdf for $S^2 = \sigma^2 Y/\nu$ is

$$g(s^2) = \frac{\nu}{\sigma^2} \frac{1}{\Gamma(\nu/2)2^{\nu/2}} \left(\frac{s^2\nu}{\sigma^2}\right)^{\nu/2-1} e^{-s^2\nu/(2\sigma^2)}.$$

Thus, the log-likelihood has the form (gathering together constants that do not depend on s^2 or σ^2)

$$\log L(\sigma^2|s^2) = \log\left(\frac{1}{\sigma^2}\right) + K\log\left(\frac{s^2}{\sigma^2}\right) - K'\frac{s^2}{\sigma^2} + K'',$$

where $K > 0$ and $K' > 0$.

The loss function in Example 7.3.27 is

$$L(\sigma^2, a) = \frac{a}{\sigma^2} - \log\left(\frac{a}{\sigma^2}\right) - 1,$$

so the loss of an estimator is the negative of its likelihood.

7.63 Let $a = \tau^2/(\tau^2+1)$, so the Bayes estimator is $\delta^\pi(x) = ax$. Then $R(\mu, \delta^\pi) = (a-1)^2\mu^2 + a^2$. As τ^2 increases, $R(\mu, \delta^\pi)$ becomes flatter.

7.65 a. Figure omitted.

b. The posterior expected loss is $\mathrm{E}\left(L(\theta, a)|x\right) = e^{ca}\mathrm{E}\,e^{-c\theta} - c\mathrm{E}(a-\theta) - 1$, where the expectation is with respect to $\pi(\theta|x)$. Then

$$\frac{d}{da}\mathrm{E}\left(L(\theta, a)|x\right) = ce^{ca}\mathrm{E}\,e^{-c\theta} - c \stackrel{\text{set}}{=} 0,$$

and $a = -\frac{1}{c}\log \mathrm{E}\,e^{-c\theta}$ is the solution. The second derivative is positive, so this is the minimum.

c. $\pi(\theta|x) = \mathrm{n}(\bar{x}, \sigma^2/n)$. So, substituting into the formula for a normal mgf, we find $\mathrm{E}\, e^{-c\theta} = e^{-c\bar{x}+\sigma^2 c^2/2n}$, and the LINEX posterior loss is

$$\mathrm{E}\left(L(\theta,a)|x\right) = e^{c(a-\bar{x})+\sigma^2c^2/2n} - c(a-\bar{x}) - 1.$$

Substitute $\mathrm{E}\, e^{-c\theta} = e^{-c\bar{x}+\sigma^2c^2/2n}$ into the formula in part (b) to find the Bayes rule is $\bar{x} - c\sigma^2/2n$.

d. For an estimator $\bar{X}+b$, the LINEX posterior loss (from part (c)) is

$$\mathrm{E}\left(L(\theta,\bar{x}+b)|x\right) = e^{cb}e^{c^2\sigma^2/2n} - cb - 1.$$

For $\bar{X}$ the expected loss is $e^{c^2\sigma^2/2n} - 1$, and for the Bayes estimator ($b = -c\sigma^2/2n$) the expected loss is $c^2\sigma^2/2n$. The marginal distribution of $\bar{X}$ is $m(\bar{x}) = 1$, so the Bayes risk is infinite for any estimator of the form $\bar{X}+b$.

e. For $\bar{X}+b$, the squared error risk is $\mathrm{E}\left[(\bar{X}+b)-\theta\right]^2 = \sigma^2/n + b^2$, so $\bar{X}$ is better than the Bayes estimator. The Bayes risk is infinite for both estimators.

7.66 Let $S = \sum_i X_i \sim \text{binomial}(n,\theta)$.

a. $\mathrm{E}\,\hat{\theta}^2 = \mathrm{E}\frac{S^2}{n^2} = \frac{1}{n^2}\mathrm{E}\,S^2 = \frac{1}{n^2}(n\theta(1-\theta)+(n\theta)^2) = \frac{\theta}{n} + \frac{n-1}{n}\theta^2$.

b. $T_n^{(i)} = \left(\sum_{j\neq i} X_j\right)^2 \Big/ (n-1)^2$. For S values of i, $T_n^{(i)} = (S-1)^2/(n-1)^2$ because the X_i that is dropped out equals 1. For the other $n-S$ values of i, $T_n^{(i)} = S^2/(n-1)^2$ because the X_i that is dropped out equals 0. Thus we can write the estimator as

$$\mathrm{JK}(T_n) = n\frac{S^2}{n^2} - \frac{n-1}{n}\left(S\frac{(S-1)^2}{(n-1)^2} + (n-S)\frac{S^2}{(n-1)^2}\right) = \frac{S^2-S}{n(n-1)}.$$

c. $\mathrm{E}\,\mathrm{JK}(T_n) = \frac{1}{n(n-1)}(n\theta(1-\theta)+(n\theta)^2 - n\theta) = \frac{n^2\theta^2 - n\theta^2}{n(n-1)} = \theta^2$.

d. For this binomial model, S is a complete sufficient statistic. Because $\mathrm{JK}(T_n)$ is a function of S that is an unbiased estimator of θ^2, it is the best unbiased estimator of θ^2.

Chapter 8

Hypothesis Testing

8.1 Let $X = \#$ of heads out of 1000. If the coin is fair, then $X \sim \text{binomial}(1000, 1/2)$. So

$$P(X \geq 560) = \sum_{x=560}^{1000} \binom{1000}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \approx .0000825,$$

where a computer was used to do the calculation. For this binomial, $\mathrm{E}\,X = 1000p = 500$ and $\mathrm{Var}\,X = 1000p(1-p) = 250$. A normal approximation is also very good for this calculation.

$$P\{X \geq 560\} = P\left\{\frac{X-500}{\sqrt{250}} \geq \frac{559.5-500}{\sqrt{250}}\right\} \approx P\{Z \geq 3.763\} \approx .0000839.$$

Thus, if the coin is fair, the probability of observing 560 or more heads out of 1000 is very small. We might tend to believe that the coin is not fair, and $p > 1/2$.

8.2 Let $X \sim \text{Poisson}(\lambda)$, and we observed $X = 10$. To assess if the accident rate has dropped, we could calculate

$$P(\,X \leq 10|\,\lambda = 15) = \sum_{i=0}^{10} \frac{e^{-15}\,15^i}{i!} = e^{-15}\,\left[1+15+\frac{15^2}{2!}+\cdots+\frac{15^{10}}{10!}\right] \approx .11846.$$

This is a fairly large value, not overwhelming evidence that the accident rate has dropped. (A normal approximation with continuity correction gives a value of .12264.)

8.3 The LRT statistic is

$$\lambda(y) = \frac{\sup_{\theta \leq \theta_0} L(\theta|y_1, \ldots, y_m)}{\sup_{\Theta} L(\theta|y_1, \ldots, y_m)}.$$

Let $y = \sum_{i=1}^m y_i$, and note that the MLE in the numerator is $\min\{y/m,\theta_0\}$ (see Exercise 7.12) while the denominator has y/m as the MLE (see Example 7.2.7). Thus

$$\lambda(y) = \begin{cases} 1 & \text{if } y/m \leq \theta_0 \\ \frac{(\theta_0)^y(1-\theta_0)^{m-y}}{(y/m)^y(1-y/m)^{m-y}} & \text{if } y/m > \theta_0, \end{cases}$$

and we reject H_0 if

$$\frac{(\theta_0)^y(1-\theta_0)^{m-y}}{(y/m)^y(1-y/m)^{m-y}} < c.$$

To show that this is equivalent to rejecting if $y > b$, we could show $\lambda(y)$ is decreasing in y so that $\lambda(y) < c$ occurs for $y > b > m\theta_0$. It is easier to work with $\log\lambda(y)$, and we have

$$\log\lambda(y) = y\,\log\,\theta_0 + (m-y)\,\log\,(1-\theta_0) - y\,\log\,\left(\frac{y}{m}\right) - (m-y)\,\log\,\left(\frac{m-y}{m}\right),$$

and

$$\begin{aligned}\frac{d}{dy}\log\lambda(y) &= \log\theta_0 - \log(1-\theta_0) - \log\left(\frac{y}{m}\right) - y\frac{1}{y} + \log\left(\frac{m-y}{m}\right) + (m-y)\frac{1}{m-y} \\ &= \log\left(\frac{\theta_0}{y/m}\frac{\left(\frac{m-y}{m}\right)}{1-\theta_0}\right).\end{aligned}$$

For $y/m > \theta_0$, $1 - y/m = (m-y)/m < 1 - \theta_0$, so each fraction above is less than 1, and the log is less than 0. Thus $\frac{d}{dy}\log\lambda < 0$ which shows that λ is decreasing in y and $\lambda(y) < c$ if and only if $y > b$.

8.4 For discrete random variables, $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = P(\mathbf{X} = \mathbf{x}|\theta)$. So the numerator and denominator of $\lambda(\mathbf{x})$ are the supremum of this probability over the indicated sets.

8.5 a. The log-likelihood is

$$\log L(\theta,\nu|\mathbf{x}) = n\log\theta + n\theta\log\nu - (\theta+1)\log\left(\prod_i x_i\right), \quad \nu \le x_{(1)},$$

where $x_{(1)} = \min_i x_i$. For any value of θ, this is an increasing function of ν for $\nu \le x_{(1)}$. So both the restricted and unrestricted MLEs of ν are $\hat{\nu} = x_{(1)}$. To find the MLE of θ, set

$$\frac{\partial}{\partial\theta}\log L(\theta, x_{(1)}|\mathbf{x}) = \frac{n}{\theta} + n\log x_{(1)} - \log\left(\prod_i x_i\right) = 0,$$

and solve for θ yielding

$$\hat{\theta} = \frac{n}{\log(\prod_i x_i/x_{(1)}^n)} = \frac{n}{T}.$$

$(\partial^2/\partial\theta^2)\log L(\theta, x_{(1)}|\mathbf{x}) = -n/\theta^2 < 0$, for all θ. So $\hat{\theta}$ is a maximum.

b. Under H_0, the MLE of θ is $\hat{\theta}_0 = 1$, and the MLE of ν is still $\hat{\nu} = x_{(1)}$. So the likelihood ratio statistic is

$$\lambda(\mathbf{x}) = \frac{x_{(1)}^n/\left(\prod_i x_i\right)^2}{(n/T)^n x_{(1)}^{n^2/T}\Big/\left(\prod_i x_i\right)^{n/T+1}} = \left(\frac{T}{n}\right)^n \frac{e^{-T}}{(e^{-T})^{n/T}} = \left(\frac{T}{n}\right)^n e^{-T+n}.$$

$(\partial/\partial T)\log\lambda(\mathbf{x}) = (n/T) - 1$. Hence, $\lambda(\mathbf{x})$ is increasing if $T \le n$ and decreasing if $T \ge n$. Thus, $T \le c$ is equivalent to $T \le c_1$ or $T \ge c_2$, for appropriately chosen constants c_1 and c_2.

c. We will not use the hint, although the problem can be solved that way. Instead, make the following three transformations. First, let $Y_i = \log X_i$, $i = 1, \ldots, n$. Next, make the n-to-1 transformation that sets $Z_1 = \min_i Y_i$ and sets $Z_2, \ldots, Z_n$ equal to the remaining Y_is, with their order unchanged. Finally, let $W_1 = Z_1$ and $W_i = Z_i - Z_1$, $i = 2, \ldots, n$. Then you find that the W_is are independent with $W_1 \sim f_{W_1}(w) = n\nu^n e^{-nw}$, $w > \log\nu$, and $W_i \sim$ exponential(1), $i = 2, \ldots, n$. Now $T = \sum_{i=2}^n W_i \sim$ gamma$(n-1, 1)$, and, hence, $2T \sim$ gamma$(n-1, 2) = \chi^2_{2(n-1)}$.

8.6 a.

$$\begin{aligned}\lambda(\mathbf{x},\mathbf{y}) &= \frac{\sup_{\Theta_0} L(\theta|\mathbf{x},\mathbf{y})}{\sup_{\Theta} L(\theta|\mathbf{x},\mathbf{y})} = \frac{\sup_\theta \prod_{i=1}^n \frac{1}{\theta}e^{-x_i/\theta}\prod_{j=1}^m \frac{1}{\theta}e^{-y_j/\theta}}{\sup_{\theta,\mu}\prod_{i=1}^n \frac{1}{\theta}e^{-x_i/\theta}\prod_{j=1}^m \frac{1}{\mu}e^{-y_j/\mu}} \\ &= \frac{\sup_\theta \frac{1}{\theta^{m+n}}\exp\left\{-\left(\sum_{i=1}^n x_i + \sum_{j=1}^m y_j\right)\Big/\theta\right\}}{\sup_{\theta,\mu}\frac{1}{\theta^n}\exp\left\{-\sum_{i=1}^n x_i/\theta\right\}\frac{1}{\mu^m}\exp\left\{-\sum_{j=1}^m y_j/\mu\right\}}.\end{aligned}$$

Differentiation will show that in the numerator $\hat\theta_0 = (\sum_i x_i + \sum_j y_j)/(n+m)$, while in the denominator $\hat\theta = \bar x$ and $\hat\mu = \bar y$. Therefore,

$$\begin{aligned}\lambda(\mathbf{x},\mathbf{y}) &= \frac{\left(\frac{n+m}{\sum_i x_i+\sum_j y_j}\right)^{n+m}\exp\left\{-\left(\frac{n+m}{\sum_i x_i+\sum_j y_j}\right)\left(\sum_i x_i+\sum_j y_j\right)\right\}}{\left(\frac{n}{\sum_i x_i}\right)^{n}\exp\left\{-\left(\frac{n}{\sum_i x_i}\right)\sum_i x_i\right\}\left(\frac{m}{\sum_j y_j}\right)^{m}\exp\left\{-\left(\frac{m}{\sum_j y_j}\right)\sum_j y_j\right\}} \\ &= \frac{(n+m)^{n+m}}{n^n m^m}\frac{\left(\sum_i x_i\right)^n\left(\sum_j y_j\right)^m}{\left(\sum_i x_i+\sum_j y_j\right)^{n+m}}.\end{aligned}$$

And the LRT is to reject H_0 if $\lambda(\mathbf{x},\mathbf{y}) \le c$.

b.

$$\lambda = \frac{(n+m)^{n+m}}{n^n m^m}\left(\frac{\sum_i x_i}{\sum_i x_i+\sum_j y_j}\right)^n\left(\frac{\sum_j y_j}{\sum_i x_i+\sum_j y_j}\right)^m = \frac{(n+m)^{n+m}}{n^n m^m}T^n(1-T)^m.$$

Therefore λ is a function of T. λ is a unimodal function of T which is maximized when $T = \frac{n}{m+n}$. Rejection for $\lambda \le c$ is equivalent to rejection for $T \le a$ or $T \ge b$, where a and b are constants that satisfy $a^n(1-a)^m = b^n(1-b)^m$.

c. When H_0 is true, $\sum_i X_i \sim \text{gamma}(n,\theta)$ and $\sum_j Y_j \sim \text{gamma}(m,\theta)$ and they are independent. So by an extension of Exercise 4.19b, $T \sim \text{beta}(n,m)$.

8.7 a.

$$L(\theta,\lambda|\mathbf{x}) = \prod_{i=1}^n \frac{1}{\lambda}e^{-(x_i-\theta)/\lambda}I_{[\theta,\infty)}(x_i) = \left(\frac{1}{\lambda}\right)^n e^{-(\Sigma_i x_i - n\theta)/\lambda}I_{[\theta,\infty)}(x_{(1)}),$$

which is increasing in θ if $x_{(1)} \ge \theta$ (regardless of λ). So the MLE of θ is $\hat\theta = x_{(1)}$. Then

$$\frac{\partial \log L}{\partial\lambda} = -\frac{n}{\lambda} + \frac{\sum_i x_i - n\hat\theta}{\lambda^2} \stackrel{\text{set}}{=} 0 \quad\Rightarrow\quad n\hat\lambda = \sum_i x_i - n\hat\theta \quad\Rightarrow\quad \hat\lambda = \bar x - x_{(1)}.$$

Because

$$\frac{\partial^2 \log L}{\partial\lambda^2} = \frac{n}{\lambda^2} - 2\left.\frac{\sum_i x_i - n\hat\theta}{\lambda^3}\right|_{\bar x - x_{(1)}} = \frac{n}{(\bar x - x_{(1)})^2} - \frac{2n(\bar x - x_{(1)})}{(\bar x - x_{(1)})^3} = \frac{-n}{(\bar x - x_{(1)})^2} < 0,$$

we have $\hat\theta = x_{(1)}$ and $\hat\lambda = \bar x - x_{(1)}$ as the unrestricted MLEs of θ and λ. Under the restriction $\theta \le 0$, the MLE of θ (regardless of λ) is

$$\hat\theta_0 = \begin{cases} 0 & \text{if } x_{(1)} > 0 \\ x_{(1)} & \text{if } x_{(1)} \le 0. \end{cases}$$

For $x_{(1)} > 0$, substituting $\hat\theta_0 = 0$ and maximizing with respect to λ, as above, yields $\hat\lambda_0 = \bar x$. Therefore,

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta,\lambda \mid \mathbf{x})}{\sup_{\Theta} L(\theta,\lambda \mid \mathbf{x})} = \frac{\sup_{\{(\lambda,\theta):\theta\le 0\}} L(\lambda,\theta \mid \mathbf{x})}{L(\hat\theta,\hat\lambda \mid \mathbf{x})} = \begin{cases} 1 & \text{if } x_{(1)} \le 0 \\ \frac{L(\bar x, 0|\mathbf{x})}{L(\hat\lambda,\hat\theta|\mathbf{x})} & \text{if } x_{(1)} > 0, \end{cases}$$

where

$$\frac{L(\bar x, 0 \mid \mathbf{x})}{L(\hat\lambda,\hat\theta \mid \mathbf{x})} = \frac{(1/\bar x)^n e^{-n\bar x/\bar x}}{\left(1/\hat\lambda\right)^n e^{-n(\bar x - x_{(1)})/(\bar x - x_{(1)})}} = \left(\frac{\hat\lambda}{\bar x}\right)^n = \left(\frac{\bar x - x_{(1)}}{\bar x}\right)^n = \left(1 - \frac{x_{(1)}}{\bar x}\right)^n.$$

So rejecting if $\lambda(\mathbf{x}) \le c$ is equivalent to rejecting if $x_{(1)}/\bar x \ge c^*$, where c^* is some constant.

b. The LRT statistic is

$$\lambda(\mathbf{x}) = \frac{\sup_\beta (1/\beta^n) e^{-\Sigma_i x_i/\beta}}{\sup_{\beta,\gamma} (\gamma^n/\beta^n)(\prod_i x_i)^{\gamma-1} e^{-\Sigma_i x_i^\gamma/\beta}}.$$

The numerator is maximized at $\hat{\beta}_0 = \bar{x}$. For fixed γ, the denominator is maximized at $\hat{\beta}_\gamma = \sum_i x_i^\gamma/n$. Thus

$$\lambda(\mathbf{x}) = \frac{\bar{x}^{-n} e^{-n}}{\sup_\gamma (\gamma^n/\hat{\beta}_\gamma^n)(\prod_i x_i)^{\gamma-1} e^{-\Sigma_i x_i^\gamma/\hat{\beta}_\gamma}} = \frac{\bar{x}^{-n}}{\sup_\gamma (\gamma^n/\hat{\beta}_\gamma^n)(\prod_i x_i)^{\gamma-1}}.$$

The denominator cannot be maximized in closed form. Numeric maximization could be used to compute the statistic for observed data $\mathbf{x}$.

8.8 a. We will first find the MLEs of a and θ. We have

$$\begin{aligned} L(a,\theta \mid \mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi a\theta}} e^{-(x_i-\theta)^2/(2a\theta)}, \\ \log L(a,\theta \mid \mathbf{x}) &= \sum_{i=1}^n -\frac{1}{2}\log(2\pi a\theta) - \frac{1}{2a\theta}(x_i-\theta)^2. \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial \log L}{\partial a} &= \sum_{i=1}^n \left(-\frac{1}{2a} + \frac{1}{2\theta a^2}(x_i-\theta)^2\right) = -\frac{n}{2a} + \frac{1}{2\theta a^2}\sum_{i=1}^n (x_i-\theta)^2 \stackrel{\text{set}}{=} 0 \\ \frac{\partial \log L}{\partial \theta} &= \sum_{i=1}^n \left[-\frac{1}{2\theta} + \frac{1}{2a\theta^2}(x_i-\theta)^2 + \frac{1}{a\theta}(x_i-\theta)\right] \\ &= -\frac{n}{2\theta} + \frac{1}{2a\theta^2}\sum_{i=1}^n (x_i-\theta)^2 + \frac{n\bar{x}-n\theta}{a\theta} \stackrel{\text{set}}{=} 0. \end{aligned}$$

We have to solve these two equations simultaneously to get MLEs of a and θ, say $\hat{a}$ and $\hat{\theta}$. Solve the first equation for a in terms of θ to get

$$a = \frac{1}{n\theta}\sum_{i=1}^n (x_i-\theta)^2.$$

Substitute this into the second equation to get

$$-\frac{n}{2\theta} + \frac{n}{2\theta} + \frac{n(\bar{x}-\theta)}{a\theta} = 0.$$

So we get $\hat{\theta} = \bar{x}$, and

$$\hat{a} = \frac{1}{n\bar{x}}\sum_{i=1}^n (x_i-\bar{x})^2 = \frac{\hat{\sigma}^2}{\bar{x}},$$

the ratio of the usual MLEs of the mean and variance. (Verification that this is a maximum is lengthy. We omit it.) For $a = 1$, we just solve the second equation, which gives a quadratic in θ that leads to the restricted MLE

$$\hat{\theta}_R = \frac{-1+\sqrt{1+4(\hat{\sigma}^2+\bar{x}^2)}}{2}.$$

Noting that $\hat{a}\hat{\theta} = \hat{\sigma}^2$, we obtain

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{L(\hat{\theta}_R \mid \mathbf{x})}{L(\hat{a}, \hat{\theta} \mid \mathbf{x})} = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\theta}_R}} e^{-(x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R)}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{a}\hat{\theta}}} e^{-(x_i-\hat{\theta})^2/(2\hat{a}\hat{\theta})}} \\
&= \frac{\left(1/(2\pi\hat{\theta}_R)\right)^{n/2} e^{-\Sigma_i (x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R)}}{(1/(2\pi\hat{\sigma}^2))^{n/2} e^{-\Sigma_i (x_i-\bar{x})^2/(2\hat{\sigma}^2)}} \\
&= \left(\hat{\sigma}^2/\hat{\theta}_R\right)^{n/2} e^{(n/2)-\Sigma_i (x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R)}.
\end{aligned}$$

b. In this case we have

$$\log L(a, \theta \mid \mathbf{x}) = \sum_{i=1}^n \left[-\frac{1}{2}\log(2\pi a\theta^2) - \frac{1}{2a\theta^2}(x_i-\theta)^2 \right].$$

Thus

$$\begin{aligned}
\frac{\partial \log L}{\partial a} &= \sum_{i=1}^n \left(-\frac{1}{2a} + \frac{1}{2a^2\theta^2}(x_i-\theta)^2 \right) = -\frac{n}{2a} + \frac{1}{2a^2\theta^2}\sum_{i=1}^n (x_i-\theta)^2 \stackrel{\text{set}}{=} 0. \\
\frac{\partial \log L}{\partial \theta} &= \sum_{i=1}^n \left[-\frac{1}{\theta} + \frac{1}{a\theta^3}(x_i-\theta)^2 + \frac{1}{a\theta^2}(x_i-\theta) \right] \\
&= -\frac{n}{\theta} + \frac{1}{a\theta^3}\sum_{i=1}^n (x_i-\theta)^2 + \frac{1}{a\theta^2}\sum_{i=1}^n (x_i-\theta) \stackrel{\text{set}}{=} 0.
\end{aligned}$$

Solving the first equation for a in terms of θ yields

$$a = \frac{1}{n\theta^2}\sum_{i=1}^n (x_i-\theta)^2.$$

Substituting this into the second equation, we get

$$-\frac{n}{\theta} + \frac{n}{\theta} + n\frac{\sum_i (x_i-\theta)}{\sum_i (x_i-\theta)^2} = 0.$$

So again, $\hat{\theta} = \bar{x}$ and

$$\hat{a} = \frac{1}{n\bar{x}^2}\sum_{i=1}^n (x_i-\bar{x})^2 = \frac{\hat{\sigma}^2}{\bar{x}^2}$$

in the unrestricted case. In the restricted case, set $a = 1$ in the second equation to obtain

$$\frac{\partial \log L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^n (x_i-\theta)^2 + \frac{1}{\theta^2}\sum_{i=1}^n (x_i-\theta) \stackrel{\text{set}}{=} 0.$$

Multiply through by θ^3/n to get

$$-\theta^2 + \frac{1}{n}\sum_{i=1}^n (x_i-\theta)^2 - \frac{\theta}{n}\sum_{i=1}^n (x_i-\theta) = 0.$$

Add $\pm\bar{x}$ inside the square and complete all sums to get the equation

$$-\theta^2 + \hat{\sigma}^2 + (\bar{x}-\theta)^2 + \theta(\bar{x}-\theta) = 0.$$

This is a quadratic in θ with solution for the MLE

$$\hat{\theta}_R = \bar{x} + \sqrt{\bar{x}+4(\hat{\sigma}^2+\bar{x}^2)}\Big/ 2.$$

which yields the LRT statistic

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_R \mid \mathbf{x})}{L(\hat{a},\hat{\theta} \mid \mathbf{x})} = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\theta}_R^2}} e^{-(x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R^2)}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{a}\hat{\theta}^2}} e^{-(x_i-\hat{\theta})^2/(2\hat{a}\hat{\theta}^2)}} = \left(\frac{\hat{\sigma}}{\hat{\theta}_R}\right)^n e^{(n/2)-\Sigma_i(x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R)}.$$

8.9 a. The MLE of λ under H_0 is $\hat{\lambda}_0 = \left(\bar{Y}\right)^{-1}$, and the MLE of λ_i under H_1 is $\hat{\lambda}_i = Y_i^{-1}$. The LRT statistic is bounded above by 1 and is given by

$$1 \geq \frac{\left(\bar{Y}\right)^{-n} e^{-n}}{\left(\prod_i Y_i\right)^{-1} e^{-n}}.$$

Rearrangement of this inequality yields $\bar{Y} \geq \left(\prod_i Y_i\right)^{1/n}$, the arithmetic-geometric mean inequality.

b. The pdf of X_i is $f(x_i|\lambda_i) = (\lambda_i/x_i^2)e^{-\lambda_i/x_i}$, $x_i > 0$. The MLE of λ under H_0 is $\hat{\lambda}_0 = n/\left[\sum_i(1/X_i)\right]$, and the MLE of λ_i under H_1 is $\hat{\lambda}_i = X_i$. Now, the argument proceeds as in part (a).

8.10 Let $Y = \sum_i X_i$. The posterior distribution of $\lambda|y$ is gamma $(y+\alpha, \beta/(\beta+1))$.

a.

$$P(\lambda \leq \lambda_0|y) = \frac{(\beta+1)^{y+\alpha}}{\Gamma(y+\alpha)\beta^{y+\alpha}} \int_0^{\lambda_0} t^{y+\alpha-1}e^{-t(\beta+1)/\beta}\,dt.$$

$P(\lambda > \lambda_0|y) = 1 - P(\lambda \leq \lambda_0|y)$.

b. Because $\beta/(\beta+1)$ is a scale parameter in the posterior distribution, $(2(\beta+1)\lambda/\beta)|y$ has a gamma$(y+\alpha,2)$ distribution. If 2α is an integer, this is a $\chi^2_{2y+2\alpha}$ distribution. So, for $\alpha = 5/2$ and $\beta = 2$,

$$P(\lambda \leq \lambda_0|y) = P\left(\frac{2(\beta+1)\lambda}{\beta} \leq \frac{2(\beta+1)\lambda_0}{\beta}\,\middle|\, y\right) = P(\chi^2_{2y+5} \leq 3\lambda_0).$$

8.11 a. From Exercise 7.23, the posterior distribution of σ^2 given S^2 is IG(γ,δ), where $\gamma = \alpha+(n-1)/2$ and $\delta = [(n-1)S^2/2+1/\beta]^{-1}$. Let $Y = 2/(\sigma^2\delta)$. Then $Y|S^2 \sim$ gamma$(\gamma,2)$. (Note: If 2α is an integer, this is a $\chi^2_{2\gamma}$ distribution.) Let M denote the median of a gamma$(\gamma,2)$ distribution. Note that M depends on only α and n, not on S^2 or β. Then we have $P(Y \geq 2/\delta|S^2) = P(\sigma^2 \leq 1|S^2) > 1/2$ if and only if

$$M > \frac{2}{\delta} = (n-1)S^2 + \frac{2}{\beta}, \quad \text{that is,} \quad S^2 < \frac{M-2/\beta}{n-1}.$$

b. From Example 7.2.11, the unrestricted MLEs are $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = (n-1)S^2/n$. Under H_0, $\hat{\mu}$ is still $\bar{X}$, because this was the maximizing value of μ, regardless of σ^2. Then because $L(\bar{x},\sigma^2|\mathbf{x})$ is a unimodal function of σ^2, the restricted MLE of σ^2 is $\hat{\sigma}^2$, if $\hat{\sigma}^2 \leq 1$, and is 1, if $\hat{\sigma}^2 > 1$. So the LRT statistic is

$$\lambda(\mathbf{x}) = \begin{cases} 1 & \text{if } \hat{\sigma}^2 \leq 1 \\ (\hat{\sigma}^2)^{n/2}e^{-n(\hat{\sigma}^2-1)/2} & \text{if } \hat{\sigma}^2 > 1. \end{cases}$$

We have that, for $\hat\sigma^2 > 1$,

$$\frac{\partial}{\partial(\hat\sigma^2)} \log\lambda(\mathbf{x}) = \frac{n}{2}\left(\frac{1}{\hat\sigma^2} - 1\right) < 0.$$

So $\lambda(\mathbf{x})$ is decreasing in $\hat\sigma^2$, and rejecting H_0 for small values of $\lambda(\mathbf{x})$ is equivalent to rejecting for large values of $\hat\sigma^2$, that is, large values of S^2. The LRT accepts H_0 if and only if $S^2 < k$, where k is a constant. We can pick the prior parameters so that the acceptance regions match in this way. First, pick α large enough that $M/(n-1) > k$. Then, as β varies between 0 and ∞, $(M-2/\beta)/(n-1)$ varies between $-\infty$ and $M/(n-1)$. So, for some choice of β, $(M-2/\beta)/(n-1) = k$ and the acceptance regions match.

8.12 a. For $H_0: \mu \le 0$ vs. $H_1: \mu > 0$ the LRT is to reject H_0 if $\bar x > c\sigma/\sqrt{n}$ (Example 8.3.3). For $\alpha = .05$ take $c = 1.645$. The power function is

$$\beta(\mu) = P\left(\frac{\bar X - \mu}{\sigma/\sqrt{n}} > 1.645 - \frac{\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > 1.645 - \frac{\sqrt{n}\mu}{\sigma}\right).$$

Note that the power will equal .5 when $\mu = 1.645\sigma/\sqrt{n}$.

b. For $H_0: \mu = 0$ vs. $H_A: \mu \ne 0$ the LRT is to reject H_0 if $|\bar x| > c\sigma/\sqrt{n}$ (Example 8.2.2). For $\alpha = .05$ take $c = 1.96$. The power function is

$$\beta(\mu) = P\left(-1.96 - \sqrt{n}\mu/\sigma \le Z \le 1.96 + \sqrt{n}\mu/\sigma\right).$$

In this case, $\mu = \pm 1.96\sigma/\sqrt{n}$ gives power of approximately .5.

8.13 a. The size of ϕ_1 is $\alpha_1 = P(X_1 > .95|\theta = 0) = .05$. The size of ϕ_2 is $\alpha_2 = P(X_1 + X_2 > C|\theta = 0)$. If $1 \le C \le 2$, this is

$$\alpha_2 = P(X_1 + X_2 > C|\theta = 0) = \int_{1-C}^{1}\int_{C-x_1}^{1} 1\, dx_2\, dx_1 = \frac{(2-C)^2}{2}.$$

Setting this equal to α and solving for C gives $C = 2 - \sqrt{2\alpha}$, and for $\alpha = .05$, we get $C = 2 - \sqrt{.1} \approx 1.68$.

b. For the first test we have the power function

$$\beta_1(\theta) = P_\theta(X_1 > .95) = \begin{cases} 0 & \text{if } \theta \le -.05 \\ \theta + .05 & \text{if } -.05 < \theta \le .95 \\ 1 & \text{if } .95 < \theta. \end{cases}$$

Using the distribution of $Y = X_1 + X_2$, given by

$$f_Y(y|\theta) = \begin{cases} y - 2\theta & \text{if } 2\theta \le y < 2\theta + 1 \\ 2\theta + 2 - y & \text{if } 2\theta + 1 \le y < 2\theta + 2 \\ 0 & \text{otherwise,} \end{cases}$$

we obtain the power function for the second test as

$$\beta_2(\theta) = P_\theta(Y > C) = \begin{cases} 0 & \text{if } \theta \le (C/2) - 1 \\ (2\theta + 2 - C)^2/2 & \text{if } (C/2) - 1 < \theta \le (C-1)/2 \\ 1 - (C - 2\theta)^2/2 & \text{if } (C-1)/2 < \theta \le C/2 \\ 1 & \text{if } C/2 < \theta. \end{cases}$$

c. From the graph it is clear that ϕ_1 is more powerful for θ near 0, but ϕ_2 is more powerful for larger θs. ϕ_2 is not uniformly more powerful than ϕ_1.

d. If either $X_1 \geq 1$ or $X_2 \geq 1$, we should reject H_0, because if $\theta = 0$, $P(X_i < 1) = 1$. Thus, consider the rejection region given by

$$\{(x_1, x_2)\colon x_1 + x_2 > C\} \bigcup \{(x_1, x_2)\colon x_1 > 1\} \bigcup \{(x_1, x_2)\colon x_2 > 1\}.$$

The first set is the rejection region for ϕ_2. The test with this rejection region has the same size as ϕ_2 because the last two sets both have probability 0 if $\theta = 0$. But for $0 < \theta < C - 1$, The power function of this test is strictly larger than $\beta_2(\theta)$. If $C - 1 \leq \theta$, this test and ϕ_2 have the same power.

8.14 The CLT tells us that $Z = (\sum_i X_i - np)/\sqrt{np(1-p)}$ is approximately $\mathrm{n}(0, 1)$. For a test that rejects H_0 when $\sum_i X_i > c$, we need to find c and n to satisfy

$$P\left(Z > \frac{c - n(.49)}{\sqrt{n(.49)(.51)}}\right) = .01 \quad \text{and} \quad P\left(Z > \frac{c - n(.51)}{\sqrt{n(.51)(.49)}}\right) = .99.$$

We thus want

$$\frac{c - n(.49)}{\sqrt{n(.49)(.51)}} = 2.33 \quad \text{and} \quad \frac{c - n(.51)}{\sqrt{n(.51)(.49)}} = -2.33.$$

Solving these equations gives $n = 13{,}567$ and $c = 6{,}783.5$.

8.15 From the Neyman-Pearson lemma the UMP test rejects H_0 if

$$\frac{f(x \mid \sigma_1)}{f(x \mid \sigma_0)} = \frac{(2\pi\sigma_1^2)^{-n/2} e^{-\Sigma_i x_i^2/(2\sigma_1^2)}}{(2\pi\sigma_0^2)^{-n/2} e^{-\Sigma_i x_i^2/(2\sigma_0^2)}} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left\{\frac{1}{2}\sum_i x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right\} > k$$

for some $k \geq 0$. After some algebra, this is equivalent to rejecting if

$$\sum_i x_i^2 > \frac{2\log\left(k\left(\sigma_1/\sigma_0\right)^n\right)}{\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)} = c \qquad \left(\text{because } \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} > 0\right).$$

This is the UMP test of size α, where $\alpha = P_{\sigma_0}(\sum_i X_i^2 > c)$. To determine c to obtain a specified α, use the fact that $\sum_i X_i^2/\sigma_0^2 \sim \chi_n^2$. Thus

$$\alpha = P_{\sigma_0}\left(\sum_i X_i^2/\sigma_0^2 > c/\sigma_0^2\right) = P\left(\chi_n^2 > c/\sigma_0^2\right),$$

so we must have $c/\sigma_0^2 = \chi_{n,\alpha}^2$, which means $c = \sigma_0^2 \chi_{n,\alpha}^2$.

8.16 a.

$$\begin{aligned} \text{Size} &= P(\text{reject } H_0 \mid H_0 \text{ is true}) = 1 \Rightarrow \text{Type I error} = 1. \\ \text{Power} &= P(\text{reject } H_0 \mid H_A \text{ is true}) = 1 \Rightarrow \text{Type II error} = 0. \end{aligned}$$

b.

$$\begin{aligned} \text{Size} &= P(\text{reject } H_0 \mid H_0 \text{ is true}) = 0 \Rightarrow \text{Type I error} = 0. \\ \text{Power} &= P(\text{reject } H_0 \mid H_A \text{ is true}) = 0 \Rightarrow \text{Type II error} = 1. \end{aligned}$$

8.17 a. The likelihood function is

$$L(\mu, \theta \mid \mathbf{x}, \mathbf{y}) = \mu^n \left(\prod_i x_i\right)^{\mu - 1} \theta^n \left(\prod_j y_j\right)^{\theta - 1}.$$

Maximizing, by differentiating the log-likelihood, yields the MLEs

$$\hat{\mu} = -\frac{n}{\sum_i \log x_i} \quad \text{and} \quad \hat{\theta} = -\frac{m}{\sum_j \log y_j}.$$

Under H_0, the likelihood is

$$L(\theta|\mathbf{x},\mathbf{y}) = \theta^{n+m}\left(\prod_i x_i \prod_j y_j\right)^{\theta-1},$$

and maximizing as above yields the restricted MLE,

$$\hat{\theta}_0 = -\frac{n+m}{\sum_i \log x_i + \sum_j \log y_j}.$$

The LRT statistic is

$$\lambda(\mathbf{x},\mathbf{y}) = \frac{\hat{\theta}_0^{m+n}}{\hat{\mu}^n\hat{\theta}^m}\left(\prod_i x_i\right)^{\hat{\theta}_0-\hat{\mu}}\left(\prod_j y_j\right)^{\hat{\theta}_0-\hat{\theta}}.$$

b. Substituting in the formulas for $\hat{\theta}$, $\hat{\mu}$ and $\hat{\theta}_0$ yields $\left(\prod_i x_i\right)^{\hat{\theta}_0-\hat{\mu}}\left(\prod_j y_j\right)^{\hat{\theta}_0-\hat{\theta}} = 1$ and

$$\lambda(\mathbf{x},\mathbf{y}) = \frac{\hat{\theta}_0^{m+n}}{\hat{\mu}^n\hat{\theta}^m} = \frac{\hat{\theta}_0^n}{\hat{\mu}^n}\frac{\hat{\theta}_0^m}{\hat{\theta}^m} = \left(\frac{m+n}{m}\right)^m\left(\frac{m+n}{n}\right)^n(1-T)^m T^n.$$

This is a unimodal function of T. So rejecting if $\lambda(\mathbf{x},\mathbf{y}) \le c$ is equivalent to rejecting if $T \le c_1$ or $T \ge c_2$, where c_1 and c_2 are appropriately chosen constants.

c. Simple transformations yield $-\log X_i \sim \text{exponential}(1/\mu)$ and $-\log Y_i \sim \text{exponential}(1/\theta)$. Therefore, $T = W/(W+V)$ where W and V are independent, $W \sim \text{gamma}(n, 1/\mu)$ and $V \sim \text{gamma}(m, 1/\theta)$. Under H_0, the scale parameters of W and V are equal. Then, a simple generalization of Exercise 4.19b yields $T \sim \text{beta}(n,m)$. The constants c_1 and c_2 are determined by the two equations

$$P(T \le c_1) + P(T \ge c_2) = \alpha \quad \text{and} \quad (1-c_1)^m c_1^n = (1-c_2)^m c_2^n.$$

8.18 a.

$$\begin{aligned}
\beta(\theta) &= P_\theta\left(\frac{|\bar{X}-\theta_0|}{\sigma/\sqrt{n}} > c\right) = 1 - P_\theta\left(\frac{|\bar{X}-\theta_0|}{\sigma/\sqrt{n}} \le c\right) \\
&= 1 - P_\theta\left(-\frac{c\sigma}{\sqrt{n}} \le \bar{X}-\theta_0 \le \frac{c\sigma}{\sqrt{n}}\right) \\
&= 1 - P_\theta\left(\frac{-c\sigma/\sqrt{n}+\theta_0-\theta}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \le \frac{c\sigma/\sqrt{n}+\theta_0-\theta}{\sigma/\sqrt{n}}\right) \\
&= 1 - P\left(-c+\frac{\theta_0-\theta}{\sigma/\sqrt{n}} \le Z \le c+\frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right) \\
&= 1 + \Phi\left(-c+\frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right) - \Phi\left(c+\frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right),
\end{aligned}$$

where $Z \sim \text{n}(0,1)$ and Φ is the standard normal cdf.

b. The size is $.05 = \beta(\theta_0) = 1 + \Phi(-c) - \Phi(c)$ which implies $c = 1.96$. The power $(1 -$ type II error) is

$$.75 \leq \beta(\theta_0 + \sigma) = 1 + \Phi(-c - \sqrt{n}) - \Phi(c - \sqrt{n}) = 1 + \underbrace{\Phi(-1.96 - \sqrt{n})}_{\approx 0} - \Phi(1.96 - \sqrt{n}).$$

$\Phi(-.675) \approx .25$ implies $1.96 - \sqrt{n} = -.675$ implies $n = 6.943 \approx 7$.

8.19 The pdf of Y is

$$f(y|\theta) = \frac{1}{\theta} y^{(1/\theta)-1} e^{-y^{1/\theta}}, \quad y > 0.$$

By the Neyman-Pearson Lemma, the UMP test will reject if

$$\frac{1}{2} y^{-1/2} e^{y - y^{1/2}} = \frac{f(y|2)}{f(y|1)} > k.$$

To see the form of this rejection region, we compute

$$\frac{d}{dy}\left(\frac{1}{2} y^{-1/2} e^{y-y^{1/2}}\right) = \frac{1}{2} y^{-3/2} e^{y-y^{1/2}} \left(y - \frac{y^{1/2}}{2} - \frac{1}{2}\right)$$

which is negative for $y < 1$ and positive for $y > 1$. Thus $f(y|2)/f(y|1)$ is decreasing for $y \leq 1$ and increasing for $y \geq 1$. Hence, rejecting for $f(y|2)/f(y|1) > k$ is equivalent to rejecting for $y \leq c_0$ or $y \geq c_1$. To obtain a size α test, the constants c_0 and c_1 must satisfy

$$\alpha = P(Y \leq c_0|\theta = 1) + P(Y \geq c_1|\theta = 1) = 1 - e^{-c_0} + e^{-c_1} \quad \text{and} \quad \frac{f(c_0|2)}{f(c_0|1)} = \frac{f(c_1|2)}{f(c_1|1)}.$$

Solving these two equations numerically, for $\alpha = .10$, yields $c_0 = .076546$ and $c_1 = 3.637798$. The Type II error probability is

$$P(c_0 < Y < c_1|\theta = 2) = \int_{c_0}^{c_1} \frac{1}{2} y^{-1/2} e^{-y^{1/2}}\, dy = -e^{-y^{1/2}}\Big|_{c_0}^{c_1} = .609824.$$

8.20 By the Neyman-Pearson Lemma, the UMP test rejects for large values of $f(x|H_1)/f(x|H_0)$. Computing this ratio we obtain

x	1	2	3	4	5	6	7
$\frac{f(x\mid H_1)}{f(x\mid H_0)}$	6	5	4	3	2	1	.84

The ratio is decreasing in x. So rejecting for large values of $f(x|H_1)/f(x|H_0)$ corresponds to rejecting for small values of x. To get a size α test, we need to choose c so that $P(X \leq c|H_0) = \alpha$. The value $c = 4$ gives the UMP size $\alpha = .04$ test. The Type II error probability is $P(X = 5, 6, 7|H_1) = .82$.

8.21 The proof is the same with integrals replaced by sums.

8.22 a. From Corollary 8.3.13 we can base the test on $\sum_i X_i$, the sufficient statistic. Let $Y = \sum_i X_i \sim \text{binomial}(10, p)$ and let $f(y|p)$ denote the pmf of Y. By Corollary 8.3.13, a test that rejects if $f(y|1/4)/f(y|1/2) > k$ is UMP of its size. By Exercise 8.25c, the ratio $f(y|1/2)/f(y|1/4)$ is increasing in y. So the ratio $f(y|1/4)/f(y|1/2)$ is decreasing in y, and rejecting for large value of the ratio is equivalent to rejecting for small values of y. To get $\alpha = .0547$, we must find c such that $P(Y \leq c|p = 1/2) = .0547$. Trying values $c = 0, 1, \ldots$, we find that for $c = 2$, $P(Y \leq 2|p = 1/2) = .0547$. So the test that rejects if $Y \leq 2$ is the UMP size $\alpha = .0547$ test. The power of the test is $P(Y \leq 2|p = 1/4) \approx .526$.

b. The size of the test is $P(Y \geq 6|p = 1/2) = \sum_{k=6}^{10} \binom{10}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10-k} \approx .377$. The power function is $\beta(\theta) = \sum_{k=6}^{10} \binom{10}{k} \theta^k (1-\theta)^{10-k}$

c. There is a nonrandomized UMP test for all α levels corresponding to the probabilities $P(Y \leq i|p = 1/2)$, where i is an integer. For $n = 10$, α can have any of the values 0, $\frac{1}{1024}$, $\frac{11}{1024}$, $\frac{56}{1024}$, $\frac{176}{1024}$, $\frac{386}{1024}$, $\frac{638}{1024}$, $\frac{848}{1024}$, $\frac{968}{1024}$, $\frac{1013}{1024}$, $\frac{1023}{1024}$, and 1.

8.23 a. The test is Reject H_0 if $X > 1/2$. So the power function is

$$\beta(\theta) = P_\theta(X > 1/2) = \int_{1/2}^{1} \frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)} x^{\theta-1}(1-x)^{1-1}\,dx = \theta\,\frac{1}{\theta}x^\theta\Big|_{1/2}^{1} = 1 - \frac{1}{2^\theta}.$$

The size is $\sup_{\theta \in H_0} \beta(\theta) = \sup_{\theta \leq 1}(1 - 1/2^\theta) = 1 - 1/2 = 1/2$.

b. By the Neyman-Pearson Lemma, the most powerful test of $H_0\colon \theta = 1$ vs. $H_1\colon \theta = 2$ is given by Reject H_0 if $f(x \mid 2)/f(x \mid 1) > k$ for some $k \geq 0$. Substituting the beta pdf gives

$$\frac{f(x|2)}{f(x|1)} = \frac{\frac{1}{\beta(2,1)} x^{2-1}(1-x)^{1-1}}{\frac{1}{\beta(1,1)} x^{1-1}(1-x)^{1-1}} = \frac{\Gamma(3)}{\Gamma(2)\Gamma(1)} x = 2x.$$

Thus, the MP test is Reject H_0 if $X > k/2$. We now use the α level to determine k. We have

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \beta(1) = \int_{k/2}^{1} f_X(x|1)\,dx = \int_{k/2}^{1} \frac{1}{\beta(1,1)} x^{1-1}(1-x)^{1-1}\,dx = 1 - \frac{k}{2}.$$

Thus $1 - k/2 = \alpha$, so the most powerful α level test is reject H_0 if $X > 1 - \alpha$.

c. For $\theta_2 > \theta_1$, $f(x|\theta_2)/f(x|\theta_1) = (\theta_2/\theta_1)x^{\theta_2-\theta_1}$, an increasing function of x because $\theta_2 > \theta_1$. So this family has MLR. By the Karlin-Rubin Theorem, the test that rejects H_0 if $X > t$ is the UMP test of its size. By the argument in part (b), use $t = 1 - \alpha$ to get size α.

8.24 For $H_0\colon \theta = \theta_0$ vs. $H_1\colon \theta = \theta_1$, the LRT statistic is

$$\lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{\max\{L(\theta_0|\mathbf{x}), L(\theta_1|\mathbf{x})\}} = \begin{cases} 1 & \text{if } L(\theta_0|\mathbf{x}) \geq L(\theta_1|\mathbf{x}) \\ L(\theta_0|\mathbf{x})/L(\theta_1|\mathbf{x}) & \text{if } L(\theta_0|\mathbf{x}) < L(\theta_1|\mathbf{x}). \end{cases}$$

The LRT rejects H_0 if $\lambda(\mathbf{x}) < c$. The Neyman-Pearson test rejects H_0 if $f(\mathbf{x}|\theta_1)/f(\mathbf{x}|\theta_0) = L(\theta_1|\mathbf{x})/L(\theta_0|\mathbf{x}) > k$. If $k = 1/c > 1$, this is equivalent to $L(\theta_0|\mathbf{x})/L(\theta_1|\mathbf{x}) < c$, the LRT. But if $c \geq 1$ or $k \leq 1$, the tests will not be the same. Because c is usually chosen to be small (k large) to get a small size α, in practice the two tests are often the same.

8.25 a. For $\theta_2 > \theta_1$,

$$\frac{g(x \mid \theta_2)}{g(x \mid \theta_1)} = \frac{e^{-(x-\theta_2)^2/2\sigma^2}}{e^{-(x-\theta_1)^2/2\sigma^2}} = e^{x(\theta_2-\theta_1)/\sigma^2} e^{(\theta_1^2-\theta_2^2)/2\sigma^2}.$$

Because $\theta_2 - \theta_1 > 0$, the ratio is increasing in x. So the families of $\mathrm{n}(\theta, \sigma^2)$ have MLR.

b. For $\theta_2 > \theta_1$,

$$\frac{g(x \mid \theta_2)}{g(x \mid \theta_1)} = \frac{e^{-\theta_2}\theta_2^x/x!}{e^{-\theta_1}\theta_1^x/x!} = \left(\frac{\theta_2}{\theta_1}\right)^x e^{\theta_1-\theta_2},$$

which is increasing in x because $\theta_2/\theta_1 > 1$. Thus the Poisson(θ) family has an MLR.

c. For $\theta_2 > \theta_1$,

$$\frac{g(x \mid \theta_2)}{g(x \mid \theta_1)} = \frac{\binom{n}{x}\theta_2^x(1-\theta_2)^{n-x}}{\binom{n}{x}\theta_1^x(1-\theta_1)^{n-x}} = \left(\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right)^x \left(\frac{1-\theta_2}{1-\theta_1}\right)^n.$$

Both $\theta_2/\theta_1 > 1$ and $(1-\theta_1)/(1-\theta_2) > 1$. Thus the ratio is increasing in x, and the family has MLR.

(Note: You can also use the fact that an exponential family $h(x)c(\theta)\exp(w(\theta)x)$ has MLR if $w(\theta)$ is increasing in θ (Exercise 8.27). For example, the Poisson(θ) pmf is $e^{-\theta}\exp(x \log \theta)/x!$, and the family has MLR because $\log\theta$ is increasing in θ.)

8.26 a. We will prove the result for continuous distributions. But it is also true for discrete MLR families. For $\theta_1 > \theta_2$, we must show $F(x|\theta_1) \le F(x|\theta_2)$. Now

$$\frac{d}{dx}\left[F(x|\theta_1) - F(x|\theta_2)\right] = f(x|\theta_1) - f(x|\theta_2) = f(x|\theta_2)\left(\frac{f(x|\theta_1)}{f(x|\theta_2)} - 1\right).$$

Because f has MLR, the ratio on the right-hand side is increasing, so the derivative can only change sign from negative to positive showing that any interior extremum is a minimum. Thus the function in square brackets is maximized by its value at ∞ or $-\infty$, which is zero.

b. From Exercise 3.42, location families are stochastically increasing in their location parameter, so the location Cauchy family with pdf $f(x|\theta) = (\pi[1+(x-\theta)^2])^{-1}$ is stochastically increasing. The family does not have MLR.

8.27 For $\theta_2 > \theta_1$,

$$\frac{g(t|\theta_2)}{g(t|\theta_1)} = \frac{c(\theta_2)}{c(\theta_1)} e^{[w(\theta_2)-w(\theta_1)]t}$$

which is increasing in t because $w(\theta_2) - w(\theta_1) > 0$. Examples include $\mathrm{n}(\theta, 1)$, $\mathrm{beta}(\theta, 1)$, and $\mathrm{Bernoulli}(\theta)$.

8.28 a. For $\theta_2 > \theta_1$, the likelihood ratio is

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = e^{\theta_1-\theta_2}\left[\frac{1+e^{x-\theta_1}}{1+e^{x-\theta_2}}\right]^2.$$

The derivative of the quantity in brackets is

$$\frac{d}{dx}\frac{1+e^{x-\theta_1}}{1+e^{x-\theta_2}} = \frac{e^{x-\theta_1} - e^{x-\theta_2}}{(1+e^{x-\theta_2})^2}.$$

Because $\theta_2 > \theta_1$, $e^{x-\theta_1} > e^{x-\theta_2}$, and, hence, the ratio is increasing. This family has MLR.

b. The best test is to reject H_0 if $f(x|1)/f(x|0) > k$. From part (a), this ratio is increasing in x. Thus this inequality is equivalent to rejecting if $x > k'$. The cdf of this logistic is $F(x|\theta) = e^{x-\theta}\Big/(1 + e^{x-\theta})$. Thus

$$\alpha = 1 - F(k'|0) = \frac{1}{1+e^{k'}} \quad \text{and} \quad \beta = F(k'|1) = \frac{e^{k'-1}}{1+e^{k'-1}}.$$

For a specified α, $k' = \log(1-\alpha)/\alpha$. So for $\alpha = .2$, $k' \approx 1.386$ and $\beta \approx .595$.

c. The Karlin-Rubin Theorem is satisfied, so the test is UMP of its size.

8.29 a. Let $\theta_2 > \theta_1$. Then

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{1+(x-\theta_1)^2}{1+(x-\theta_2)^2} = \frac{1 + (1+\theta_1)^2/x^2 - 2\theta_1/x}{1 + (1+\theta_2)^2/x^2 - 2\theta_2/x}.$$

The limit of this ratio as $x \to \infty$ or as $x \to -\infty$ is 1. So the ratio cannot be monotone increasing (or decreasing) between $-\infty$ and ∞. Thus, the family does not have MLR.

b. By the Neyman-Pearson Lemma, a test will be UMP if it rejects when $f(x|1)/f(x|0) > k$, for some constant k. Examination of the derivative shows that $f(x|1)/f(x|0)$ is decreasing for $x \le (1-\sqrt{5})/2 = -.618$, is increasing for $(1-\sqrt{5})/2 \le x \le (1+\sqrt{5})/2 = 1.618$, and is decreasing for $(1+\sqrt{5})/2 \le x$. Furthermore, $f(1|1)/f(1|0) = f(3|1)/f(3|0) = 2$. So rejecting if $f(x|1)/f(x|0) > 2$ is equivalent to rejecting if $1 < x < 3$. Thus, the given test is UMP of its size. The size of the test is

$$P(1 < X < 3|\theta = 0) = \int_1^3 \frac{1}{\pi}\frac{1}{1+x^2}\,dx = \frac{1}{\pi}\arctan x\Big|_1^3 \approx .1476.$$

The Type II error probability is

$$1 - P(1 < X < 3|\theta = 1) = 1 - \int_1^3 \frac{1}{\pi} \frac{1}{1+(x-1)^2}\, dx = 1 - \frac{1}{\pi}\arctan(x-1)\bigg|_1^3 \approx .6476.$$

c. We will not have $f(1|\theta)/f(1|0) = f(3|\theta)/f(3|0)$ for any other value of $\theta \neq 1$. Try $\theta = 2$, for example. So the rejection region $1 < x < 3$ will not be most powerful at any other value of θ. The test is not UMP for testing $H_0\colon \theta \leq 0$ versus $H_1\colon \theta > 0$.

8.30 a. For $\theta_2 > \theta_1 > 0$, the likelihood ratio and its derivative are

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{\theta_2}{\theta_1}\frac{\theta_1^2+x^2}{\theta_2^2+x^2} \quad \text{and} \quad \frac{d}{dx}\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{\theta_2}{\theta_1}\frac{\theta_2^2-\theta_1^2}{(\theta_2^2+x^2)^2}x.$$

The sign of the derivative is the same as the sign of x (recall, $\theta_2^2 - \theta_1^2 > 0$), which changes sign. Hence the ratio is not monotone.

b. Because $f(x|\theta) = (\theta/\pi)(\theta^2 + |x|^2)^{-1}$, $Y = |X|$ is sufficient. Its pdf is

$$f(y|\theta) = \frac{2\theta}{\pi}\frac{1}{\theta^2+y^2}, \quad y > 0.$$

Differentiating as above, the sign of the derivative is the same as the sign of y, which is positive. Hence the family has MLR.

8.31 a. By the Karlin-Rubin Theorem, the UMP test is to reject H_0 if $\sum_i X_i > k$, because $\sum_i X_i$ is sufficient and $\sum_i X_i \sim \text{Poisson}(n\lambda)$ which has MLR. Choose the constant k to satisfy $P\left(\sum_i X_i > k|\lambda = \lambda_0\right) = \alpha$.

b.

$$\begin{aligned} P\left(\sum_i X_i > k \,\middle|\, \lambda = 1\right) &\approx P\left(Z > (k-n)/\sqrt{n}\right) \stackrel{\text{set}}{=} .05, \\ P\left(\sum_i X_i > k \,\middle|\, \lambda = 2\right) &\approx P\left(Z > (k-2n)/\sqrt{2n}\right) \stackrel{\text{set}}{=} .90. \end{aligned}$$

Thus, solve for k and n in

$$\frac{k-n}{\sqrt{n}} = 1.645 \quad \text{and} \quad \frac{k-2n}{\sqrt{2n}} = -1.28,$$

yielding $n = 12$ and $k = 17.70$.

8.32 a. This is Example 8.3.15.

b. This is Example 8.3.19.

8.33 a. From Theorems 5.4.4 and 5.4.6, the marginal pdf of Y_1 and the joint pdf of (Y_1, Y_n) are

$$\begin{aligned} f(y_1|\theta) &= n(1-(y_1-\theta))^{n-1}, \quad \theta < y_1 < \theta + 1, \\ f(y_1, y_n|\theta) &= n(n-1)(y_n - y_1)^{n-2}, \quad \theta < y_1 < y_n < \theta + 1. \end{aligned}$$

Under H_0, $P(Y_n \geq 1) = 0$. So

$$\alpha = P(Y_1 \geq k|0) = \int_k^1 n(1-y_1)^{n-1}\, dy_1 = (1-k)^n.$$

Thus, use $k = 1 - \alpha^{1/n}$ to have a size α test.

b. For $\theta \le k-1$, $\beta(\theta) = 0$. For $k-1 < \theta \le 0$,

$$\beta(\theta) = \int_k^{\theta+1} n(1-(y_1-\theta))^{n-1}\,dy_1 = (1-k+\theta)^n.$$

For $0 < \theta \le k$,

$$\begin{aligned}\beta(\theta) &= \int_k^{\theta+1} n(1-(y_1-\theta))^{n-1}\,dy_1 + \int_\theta^k \int_1^{\theta+1} n(n-1)(y_n-y_1)^{n-2}\,dy_n\,dy_1 \\ &= \alpha + 1 - (1-\theta)^n.\end{aligned}$$

And for $k < \theta$, $\beta(\theta) = 1$.

c. (Y_1, Y_n) are sufficient statistics. So we can attempt to find a UMP test using Corollary 8.3.13 and the joint pdf $f(y_1, y_n|\theta)$ in part (a). For $0 < \theta < 1$, the ratio of pdfs is

$$\frac{f(y_1,y_n|\theta)}{f(y_1,y_n|0)} = \begin{cases} 0 & \text{if } 0 < y_1 \le \theta,\ y_1 < y_n < 1 \\ 1 & \text{if } \theta < y_1 < y_n < 1 \\ \infty & \text{if } 1 \le y_n < \theta+1, \theta < y_1 < y_n. \end{cases}$$

For $1 \le \theta$, the ratio of pdfs is

$$\frac{f(y_1,y_n|\theta)}{f(y_1,y_n|0)} = \begin{cases} 0 & \text{if } y_1 < y_n < 1 \\ \infty & \text{if } \theta < y_1 < y_n < \theta+1. \end{cases}$$

For $0 < \theta < k$, use $k' = 1$. The given test always rejects if $f(y_1,y_n|\theta)/f(y_1,y_n|0) > 1$ and always accepts if $f(y_1,y_n|\theta)/f(y_1,y_n|0) < 1$. For $\theta \ge k$, use $k' = 0$. The given test always rejects if $f(y_1,y_n|\theta)/f(y_1,y_n|0) > 0$ and always accepts if $f(y_1,y_n|\theta)/f(y_1,y_n|0) < 0$. Thus the given test is UMP by Corollary 8.3.13.

d. According to the power function in part (b), $\beta(\theta) = 1$ for all $\theta \ge k = 1 - \alpha^{1/n}$. So these conditions are satisfied for any n.

8.34 a. This is Exercise 3.42a.

b. This is Exercise 8.26a.

8.35 a. We will use the equality in Exercise 3.17 which remains true so long as $\nu > -\alpha$. Recall that $Y \sim \chi^2_\nu = \text{gamma}(\nu/2, 2)$. Thus, using the independence of X and Y we have

$$\mathrm{E}\,T' = \mathrm{E}\frac{X}{\sqrt{Y/\nu}} = (\mathrm{E}\,X)\sqrt{\nu}\mathrm{E}\,Y^{-1/2} = \mu\sqrt{\nu}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)\sqrt{2}}$$

if $\nu > 1$. To calculate the variance, compute

$$\mathrm{E}(T')^2 = \mathrm{E}\frac{X^2}{Y/\nu} = (\mathrm{E}\,X^2)\nu\mathrm{E}\,Y^{-1} = (\mu^2+1)\nu\frac{\Gamma((\nu-2)/2)}{\Gamma(\nu/2)2} = \frac{(\mu^2+1)\nu}{\nu-2}$$

if $\nu > 2$. Thus, if $\nu > 2$,

$$\mathrm{Var}\,T' = \frac{(\mu^2+1)\nu}{\nu-2} - \left(\mu\sqrt{\nu}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)\sqrt{2}}\right)^2$$

b. If $\delta = 0$, all the terms in the sum for $k = 1, 2, \ldots$ are zero because of the δ^k term. The expression with just the $k = 0$ term and $\delta = 0$ simplifies to the central t pdf.

c. The argument that the noncentral t has an MLR is fairly involved. It may be found in Lehmann (1986, p. 295).

8.37 a. $P(\bar{X} > \theta_0 + z_\alpha \sigma/\sqrt{n}|\theta_0) = P\left((\bar{X}-\theta_0)/(\sigma/\sqrt{n}) > z_\alpha|\theta_0\right) = P(Z > z_\alpha) = \alpha$, where $Z \sim$ n$(0,1)$. Because $\bar{x}$ is the unrestricted MLE, and the restricted MLE is θ_0 if $\bar{x} > \theta_0$, the LRT statistic is, for $\bar{x} \geq \theta_0$

$$\lambda(\mathbf{x}) = \frac{(2\pi\sigma^2)^{-n/2} e^{-\Sigma_i (x_i-\theta_0)^2/2\sigma^2}}{(2\pi\sigma^2)^{-n/2} e^{-\Sigma_i (x_i-\bar{x})^2/2\sigma^2}} = \frac{e^{-\left[n(\bar{x}-\theta_0)^2+(n-1)s^2\right]\Big/2\sigma^2}}{e^{-(n-1)s^2/2\sigma^2}} = e^{-n(\bar{x}-\theta_0)^2/2\sigma^2}.$$

and the LRT statistic is 1 for $\bar{x} < \theta_0$. Thus, rejecting if $\lambda < c$ is equivalent to rejecting if $(\bar{x}-\theta_0)/(\sigma/\sqrt{n}) > c'$ (as long as $c < 1$ – see Exercise 8.24).

b. The test is UMP by the Karlin-Rubin Theorem.

c. $P(\bar{X} > \theta_0 + t_{n-1,\alpha} S/\sqrt{n}|\theta = \theta_0) = P(T_{n-1} > t_{n-1,\alpha}) = \alpha$, when T_{n-1} is a Student's t random variable with $n-1$ degrees of freedom. If we define $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$ and $\hat{\sigma}_0^2 = \frac{1}{n}\sum(x_i-\theta_0)^2$, then for $\bar{x} \geq \theta_0$ the LRT statistic is $\lambda = (\hat{\sigma}^2/\hat{\sigma}_0^2)^{n/2}$, and for $\bar{x} < \theta_0$ the LRT statistic is $\lambda = 1$. Writing $\hat{\sigma}^2 = \frac{n-1}{n}s^2$ and $\hat{\sigma}_0^2 = (\bar{x}-\theta_0)^2 + \frac{n-1}{n}s^2$, it is clear that the LRT is equivalent to the t-test because $\lambda < c$ when

$$\frac{\frac{n-1}{n}s^2}{(\bar{x}-\theta_0)^2 + \frac{n-1}{n}s^2} = \frac{(n-1)/n}{(\bar{x}-\theta_0)^2/s^2 + (n-1)/n} < c' \quad \text{and} \quad \bar{x} \geq \theta_0,$$

which is the same as rejecting when $(\bar{x}-\theta_0)/(s/\sqrt{n})$ is large.

d. The proof that the one-sided t test is UMP unbiased is rather involved, using the bounded completeness of the normal distribution and other facts. See Chapter 5 of Lehmann (1986) for a complete treatment.

8.38 a.

$$\begin{aligned} \text{Size} &= P_{\theta_0}\left\{|\ \bar{X}-\theta_0\ | > t_{n-1,\alpha/2}\sqrt{S^2/n}\right\} \\ &= 1 - P_{\theta_0}\left\{-t_{n-1,\alpha/2}\sqrt{S^2/n} \leq \bar{X}-\theta_0 \leq t_{n-1,\alpha/2}\sqrt{S^2/n}\right\} \\ &= 1 - P_{\theta_0}\left\{-t_{n-1,\alpha/2} \leq \frac{\bar{X}-\theta_0}{\sqrt{S^2/n}} \leq t_{n-1,\alpha/2}\right\} \qquad \left(\frac{\bar{X}-\theta_0}{\sqrt{S^2/n}} \sim t_{n-1} \text{ under } H_0\right) \\ &= 1 - (1-\alpha) \quad = \quad \alpha. \end{aligned}$$

b. The unrestricted MLEs are $\hat{\theta} = \bar{X}$ and $\hat{\sigma}^2 = \sum_i (X_i - \bar{X})^2/n$. The restricted MLEs are $\hat{\theta}_0 = \theta_0$ and $\hat{\sigma}_0^2 = \sum_i (X_i-\theta_0)^2/n$. So the LRT statistic is

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{(2\pi\hat{\sigma}_0)^{-n/2}\exp\{-n\hat{\sigma}_0^2/(2\hat{\sigma}_0^2)\}}{(2\pi\hat{\sigma})^{-n/2}\exp\{-n\hat{\sigma}^2/(2\hat{\sigma}^2)\}} \\ &= \left[\frac{\sum_i (x_i-\bar{x})^2}{\sum_i (x_i-\theta_0)^2}\right]^{n/2} = \left[\frac{\sum_i (x_i-\bar{x})^2}{\sum_i (x_i-\bar{x})^2 + n(\bar{x}-\theta_0)^2}\right]^{n/2}. \end{aligned}$$

For a constant c, the LRT is

$$\text{reject } H_0 \text{ if } \left[\frac{\sum_i (x_i-\bar{x})^2}{\sum_i (x_i-\bar{x})^2 + n(\bar{x}-\theta_0)^2}\right] = \frac{1}{1 + n(\bar{x}-\theta_0)^2/\sum_i (x_i-\bar{x})^2} < c^{2/n}.$$

After some algebra we can write the test as

$$\text{reject } H_0 \text{ if } |\bar{x}-\theta_0| > \left[\left(c^{-2/n}-1\right)(n-1)\frac{s^2}{n}\right]^{1/2}.$$

We now choose the constant c to achieve size α, and we

$$\text{reject if } |\bar{x}-\theta_0| > t_{n-1,\alpha/2}\sqrt{s^2/n}.$$

c. Again, see Chapter 5 of Lehmann (1986).

8.39 a. From Exercise 4.45c, $W_i = X_i - Y_i \sim \mathrm{n}(\mu_W, \sigma_W^2)$, where $\mu_X - \mu_Y = \mu_W$ and $\sigma_X^2 + \sigma_Y^2 - \rho\sigma_X\sigma_Y = \sigma_W^2$. The W_is are independent because the pairs (X_i, Y_i) are.

b. The hypotheses are equivalent to $H_0: \mu_W = 0$ vs $H_1: \mu_W \neq 0$, and, from Exercise 8.38, if we reject H_0 when $|\bar{W}| > t_{n-1,\alpha/2}\sqrt{S_W^2/n}$, this is the LRT (based on $W_1, \ldots, W_n$) of size α. (Note that if $\rho > 0$, $\operatorname{Var} W_i$ can be small and the test will have good power.)

8.41 a.

$$\lambda(\mathbf{x},\mathbf{y}) = \frac{\sup_{H_0} L(\mu_X, \mu_Y, \sigma^2 \mid \mathbf{x},\mathbf{y})}{\sup L(\mu_X, \mu_Y, \sigma^2 \mid \mathbf{x},\mathbf{y})} = \frac{L(\hat{\mu}, \hat{\sigma}_0^2 \mid \mathbf{x},\mathbf{y})}{L(\hat{\mu}_X, \hat{\mu}_Y, \hat{\sigma}_1^2 \mid \mathbf{x},\mathbf{y})}.$$

Under H_0, the X_is and Y_is are one sample of size $m+n$ from a $\mathrm{n}(\mu, \sigma^2)$ population, where $\mu = \mu_X = \mu_Y$. So the restricted MLEs are

$$\hat{\mu} = \frac{\sum_i X_i + \sum_i Y_i}{n+m} = \frac{n\bar{x}+n\bar{y}}{n+m} \quad \text{and} \quad \hat{\sigma}_0^2 = \frac{\sum_i (X_i-\hat{\mu})^2 + \sum_i (Y_i-\hat{\mu})^2}{n+m}.$$

To obtain the unrestricted MLEs, $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}^2$, use

$$L(\mu_X, \mu_Y, \sigma^2|x,y) = (2\pi\sigma^2)^{-(n+m)/2} e^{-[\Sigma_i (x_i-\mu_X)^2 + \Sigma_i (y_i-\mu_Y)^2]/2\sigma^2}.$$

Firstly, note that $\hat{\mu}_X = \bar{x}$ and $\hat{\mu}_Y = \bar{y}$, because maximizing over μ_X does not involve μ_Y and vice versa. Then

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n+m}{2}\frac{1}{\sigma^2} + \frac{1}{2}\left[\sum_i (x_i-\hat{\mu}_X)^2 + \sum_i (y_i-\hat{\mu}_Y)^2\right]\frac{1}{(\sigma^2)^2} \overset{\text{set}}{=} 0$$

implies

$$\hat{\sigma}^2 = \left[\sum_{i=1}^n (x_i-\bar{x})^2 + \sum_{i=1}^m (y_i-\bar{y})^2\right]\frac{1}{n+m}.$$

To check that this is a maximum,

$$\begin{aligned}
\left.\frac{\partial^2 \log L}{\partial (\sigma^2)^2}\right|_{\hat{\sigma}^2} &= \left.\frac{n+m}{2}\frac{1}{(\sigma^2)^2} - \left[\sum_i (x_i-\hat{\mu}_X)^2 + \sum_i (y_i-\hat{\mu}_Y)^2\right]\frac{1}{(\sigma^2)^3}\right|_{\hat{\sigma}^2} \\
&= \frac{n+m}{2}\frac{1}{(\hat{\sigma}^2)^2} - (n+m)\frac{1}{(\hat{\sigma}^2)^2} = -\frac{n+m}{2}\frac{1}{(\hat{\sigma}^2)^2} < 0.
\end{aligned}$$

Thus, it is a maximum. We then have

$$\lambda(\mathbf{x},\mathbf{y}) = \frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n+m}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}_0^2}\left[\sum_{i=1}^n (x_i-\hat{\mu})^2 + \sum_{i=1}^m (y_i-\hat{\mu})^2\right]\right\}}{(2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}^2}\left[\sum_{i=1}^n (x_i-\bar{x})^2 + \sum_{i=1}^m (y_i-\bar{y})^2\right]\right\}} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{-\frac{n+m}{2}}$$

and the LRT is rejects H_0 if $\hat{\sigma}_0^2/\hat{\sigma}^2 > k$. In the numerator, first substitute $\hat{\mu} = (n\bar{x}+m\bar{y})/(n+m)$ and write

$$\sum_{i=1}^n \left(x_i - \frac{n\bar{x}+m\bar{y}}{n+m}\right)^2 = \sum_{i=1}^n \left((x_i-\bar{x}) + \left(\bar{x} - \frac{n\bar{x}+m\bar{y}}{n+m}\right)\right)^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + \frac{nm^2}{(n+m)^2}(\bar{x}-\bar{y})^2,$$

because the cross term is zero. Performing a similar operation on the Y sum yields

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 + \frac{nm}{n+m}(\bar{x}-\bar{y})^2}{\hat{\sigma}^2} = n + m + \frac{nm}{n+m}\frac{(\bar{x}-\bar{y})^2}{\hat{\sigma}^2}.$$

Because $\hat{\sigma}^2 = \frac{n+m-2}{n+m}S_p^2$, large values of $\hat{\sigma}_0^2 \big/ \hat{\sigma}^2$ are equivalent to large values of $(\bar{x}-\bar{y})^2 \big/ S_p^2$ and large values of $|T|$. Hence, the LRT is the two-sample t-test.

b.

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2(1/n + 1/m)}} = \frac{(\bar{X} - \bar{Y}) \Big/ \sqrt{\sigma^2(1/n + 1/m)}}{\sqrt{[(n+m-2)S_p^2/\sigma^2]/(n+m-2)}}.$$

Under H_0, $(\bar{X}-\bar{Y}) \sim \mathrm{n}(0, \sigma^2(1/n+1/m))$. Under the model, $(n-1)S_X^2/\sigma^2$ and $(m-1)S_Y^2/\sigma^2$ are independent χ^2 random variables with $(n-1)$ and $(m-1)$ degrees of freedom. Thus, $(n+m-2)S_p^2/\sigma^2 = (n-1)S_X^2/\sigma^2 + (m-1)S_Y^2/\sigma^2 \sim \chi^2_{n+m-2}$. Furthermore, $\bar{X} - \bar{Y}$ is independent of S_X^2 and S_Y^2, and, hence, S_p^2. So $T \sim t_{n+m-2}$.

c. The two-sample t test is UMP unbiased, but the proof is rather involved. See Chapter 5 of Lehmann (1986).

d. For these data we have $n = 14$, $\bar{X} = 1249.86$, $S_X^2 = 591.36$, $m = 9$, $\bar{Y} = 1261.33$, $S_Y^2 = 176.00$ and $S_p^2 = 433.13$. Therefore, $T = -1.29$ and comparing this to a t_{21} distribution gives a p-value of .21. So there is no evidence that the mean age differs between the core and periphery.

8.42 a. The Satterthwaite approximation states that if $Y_i \sim \chi^2_{r_i}$, where the Y_i's are independent, then

$$\sum_i a_i Y_i \overset{\text{approx}}{\sim} \frac{\chi^2_{\hat{\nu}}}{\hat{\nu}} \quad \text{where} \quad \hat{\nu} = \frac{\left(\sum_i a_i Y_i\right)^2}{\sum_i a_i^2 Y_i^2 / r_i}.$$

We have $Y_1 = (n-1)S_X^2/\sigma_X^2 \sim \chi^2_{n-1}$ and $Y_2 = (m-1)S_Y^2/\sigma_Y^2 \sim \chi^2_{m-1}$. Now define

$$a_1 = \frac{\sigma_X^2}{n(n-1)\left[(\sigma_X^2/n) + (\sigma_Y^2/m)\right]} \quad \text{and} \quad a_2 = \frac{\sigma_Y^2}{m(m-1)\left[(\sigma_X^2/n) + (\sigma_Y^2/m)\right]}.$$

Then,

$$\begin{aligned} \sum a_i Y_i &= \frac{\sigma_X^2}{n(n-1)\left[(\sigma_X^2/n) + (\sigma_Y^2/m)\right]}\frac{(n-1)S_X^2}{\sigma_X^2} \\ &\quad + \frac{\sigma_Y^2}{m(m-1)\left[(\sigma_X^2/n) + (\sigma_Y^2/m)\right]}\frac{(m-1)S_Y^2}{\sigma_Y^2} \\ &= \frac{S_X^2/n + S_Y^2/m}{\sigma_X^2/n + \sigma_Y^2/m} \sim \frac{\chi^2_{\hat{\nu}}}{\hat{\nu}} \end{aligned}$$

where

$$\hat{\nu} = \frac{\left(\frac{S_X^2/n + S_Y^2/m}{\sigma_X^2/n + \sigma_Y^2/m}\right)^2}{\frac{1}{(n-1)}\frac{S_X^4}{n^2\left(\sigma_X^2/n + \sigma_Y^2/m\right)^2} + \frac{1}{(m-1)}\frac{S_Y^4}{m^2\left(\sigma_X^2/n + \sigma_Y^2/m\right)^2}} = \frac{\left(S_X^2/n + S_Y^2/m\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}.$$

b. Because $\bar{X} - \bar{Y} \sim \mathrm{n}\left(\mu_X - \mu_Y, \sigma_X^2/n + \sigma_Y^2/m\right)$ and $\frac{S_X^2/n + S_Y^2/m}{\sigma_X^2/n + \sigma_Y^2/m} \overset{\text{approx}}{\sim} \chi^2_{\hat{\nu}}/\hat{\nu}$, under $H_0: \mu_X - \mu_Y = 0$ we have

$$T' = \frac{\bar{X} - \bar{Y}}{\sqrt{S_X^2/n + S_Y^2/m}} = \frac{(\bar{X} - \bar{Y}) \Big/ \sqrt{\sigma_X^2/n + \sigma_Y^2/m}}{\sqrt{\frac{\left(S_X^2/n + S_Y^2/m\right)}{\left(\sigma_X^2/n + \sigma_Y^2/m\right)}}} \overset{\text{approx}}{\sim} t_{\hat{\nu}}.$$

c. Using the values in Exercise 8.41d, we obtain $T' = -1.46$ and $\hat{\nu} = 20.64$. So the p-value is .16. There is no evidence that the mean age differs between the core and periphery.

d. $F = S_X^2/S_Y^2 = 3.36$. Comparing this with an $F_{13,8}$ distribution yields a p-value of $2P(F \geq 3.36) = .09$. So there is some slight evidence that the variance differs between the core and periphery.

8.43 There were typos in early printings. The t statistic should be

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n_1}+\frac{\rho^2}{n_2}}\sqrt{\frac{(n_1-1)s_X^2+(n_2-1)s_Y^2/\rho^2}{n_1+n_2-2}}},$$

and the F statistic should be $s_Y^2/(\rho^2 s_X^2)$. Multiply and divide the denominator of the t statistic by σ to express it as

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\rho^2\sigma^2}{n_2}}}$$

divided by

$$\sqrt{\frac{(n_1-1)s_X^2/\sigma^2+(n_2-1)s_Y^2/(\rho^2\sigma^2)}{n_1+n_2-2}}.$$

The numerator has a $\mathrm{n}(0,1)$ distribution. In the denominator, $(n_1-1)s_X^2/\sigma^2 \sim \chi^2_{n_1-1}$ and $(n_2-1)s_Y^2/(\rho^2\sigma^2) \sim \chi^2_{n_2-1}$ and they are independent, so their sum has a $\chi^2_{n_1+n_2-2}$ distribution. Thus, the statistic has the form of $\mathrm{n}(0,1)/\sqrt{\chi^2_\nu/\nu}$ where $\nu = n_1+n_2-2$, and the numerator and denominator are independent because of the independence of sample means and variances in normal sampling. Thus the statistic has a $t_{n_1+n_2-2}$ distribution. The F statistic can be written as

$$\frac{s_Y^2}{\rho^2 s_X^2} = \frac{s_Y^2/(\rho^2\sigma^2)}{s_X^2/\sigma^2} = \frac{[(n_2-1)s_Y^2/(\rho^2\sigma^2)]/(n_2-1)}{[(n_1-1)s_X^2/(\sigma^2)]/(n_1-1)}$$

which has the form $[\chi^2_{n_2-1}/(n_2-1)]/[\chi^2_{n_1-1}/(n_1-1)]$ which has an F_{n_2-1,n_1-1} distribution. (Note, early printings had a typo with the numerator and denominator degrees of freedom switched.)

8.44 Test 3 rejects $H_0\colon \theta=\theta_0$ in favor of $H_1\colon \theta \neq \theta_0$ if $\bar{X} > \theta_0 + z_{\alpha/2}\sigma/\sqrt{n}$ or $\bar{X} < \theta_0 - z_{\alpha/2}\sigma/\sqrt{n}$. Let Φ and ϕ denote the standard normal cdf and pdf, respectively. Because $\bar{X} \sim \mathrm{n}(\theta, \sigma^2/n)$, the power function of Test 3 is

$$\begin{aligned}\beta(\theta) &= P_\theta(\bar{X} < \theta_0 - z_{\alpha/2}\sigma/\sqrt{n}) + P_\theta(\bar{X} > \theta_0 + z_{\alpha/2}\sigma/\sqrt{n}) \\ &= \Phi\left(\frac{\theta_0-\theta}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) + 1 - \Phi\left(\frac{\theta_0-\theta}{\sigma/\sqrt{n}} + z_{\alpha/2}\right),\end{aligned}$$

and its derivative is

$$\frac{d\beta(\theta)}{d\theta} = -\frac{\sqrt{n}}{\sigma}\phi\left(\frac{\theta_0-\theta}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) + \frac{\sqrt{n}}{\sigma}\phi\left(\frac{\theta_0-\theta}{\sigma/\sqrt{n}} + z_{\alpha/2}\right).$$

Because ϕ is symmetric and unimodal about zero, this derivative will be zero only if

$$-\left(\frac{\theta_0-\theta}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) = \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + z_{\alpha/2},$$

that is, only if $\theta=\theta_0$. So, $\theta=\theta_0$ is the only possible local maximum or minimum of the power function. $\beta(\theta_0)=\alpha$ and $\lim_{\theta\to\pm\infty}\beta(\theta)=1$. Thus, $\theta=\theta_0$ is the global minimum of $\beta(\theta)$, and, for any $\theta' \neq \theta_0$, $\beta(\theta') > \beta(\theta_0)$. That is, Test 3 is unbiased.

8.45 The verification of size α is the same computation as in Exercise 8.37a. Example 8.3.3 shows that the power function $\beta_m(\theta)$ for each of these tests is an increasing function. So for $\theta > \theta_0$, $\beta_m(\theta) > \beta_m(\theta_0) = \alpha$. Hence, the tests are all unbiased.

8.47 a. This is very similar to the argument for Exercise 8.41.

b. By an argument similar to part (a), this LRT rejects H_0^+ if

$$T^+ = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq -t_{n+m-2,\alpha}.$$

c. Because H_0 is the union of H_0^+ and H_0^-, by the IUT method of Theorem 8.3.23 the test that rejects H_0 if the tests in parts (a) and (b) both reject is a level α test of H_0. That is, the test rejects H_0 if $T^+ \leq -t_{n+m-2,\alpha}$ and $T^- \geq t_{n+m-2,\alpha}$.

d. Use Theorem 8.3.24. Consider parameter points with $\mu_X - \mu_Y = \delta$ and $\sigma \to 0$. For any σ, $P(T^+ \leq -t_{n+m-2,\alpha}) = \alpha$. The power of the T^- test is computed from the noncentral t distribution with noncentrality parameter $|\mu_x - \mu_Y - (-\delta)|/[\sigma(1/n + 1/m)] = 2\delta/[\sigma(1/n + 1/m)]$ which converges to ∞ as $\sigma \to 0$. Thus, $P(T^- \geq t_{n+m-2,\alpha}) \to 1$ as $\sigma \to 0$. By Theorem 8.3.24, this IUT is a size α test of H_0.

8.49 a. The p-value is

$$\begin{aligned} & P\left\{\left.\begin{pmatrix}\text{7 or more successes}\\ \text{out of 10 Bernoulli trials}\end{pmatrix}\right|\theta=\frac{1}{2}\right\} \\ = & \binom{10}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 \\ = & .171875. \end{aligned}$$

b.

$$\begin{aligned} \text{P-value} &= P\{X \geq 3 \mid \lambda = 1\} = 1 - P(X < 3 \mid \lambda = 1) \\ &= 1 - \left[\frac{e^{-1}1^2}{2!} + \frac{e^{-1}1^1}{1!} + \frac{e^{-1}1^0}{0!}\right] \quad \approx \quad .0803. \end{aligned}$$

c.

$$\begin{aligned} \text{P-value} &= P\{\sum_i X_i \geq 9 \mid 3\lambda = 3\} \quad = \quad 1 - P(Y < 9 \mid 3\lambda = 3) \\ &= 1 - e^{-3}\left[\frac{3^8}{8!} + \frac{3^7}{7!} + \frac{3^6}{6!} + \frac{3^5}{5!} + \cdots + \frac{3^1}{1!} + \frac{3^0}{0!}\right] \quad \approx \quad .0038, \end{aligned}$$

where $Y = \sum_{i=1}^3 X_i \sim \text{Poisson}(3\lambda)$.

8.50 From Exercise 7.26,

$$\pi(\theta|\mathbf{x}) = \sqrt{\frac{n}{2\pi\sigma^2}} e^{-n(\theta - \delta_\pm(\mathbf{x}))^2/(2\sigma^2)},$$

where $\delta_\pm(\mathbf{x}) = \bar{x} \pm \frac{\sigma^2}{na}$ and we use the "+" if $\theta > 0$ and the "−" if $\theta < 0$.

a. For $K > 0$,

$$P(\theta > K|\mathbf{x}, a) = \sqrt{\frac{n}{2\pi\sigma^2}} \int_K^\infty e^{-n(\theta - \delta_+(\mathbf{x}))^2/(2\sigma^2)}\, d\theta = P\left(Z > \frac{\sqrt{n}}{\sigma}[K - \delta_+(\mathbf{x})]\right),$$

where $Z \sim \mathrm{n}(0,1)$.

b. As $a \to \infty$, $\delta_+(\mathbf{x}) \to \bar{x}$ so $P(\theta > K) \to P\left(Z > \frac{\sqrt{n}}{\sigma}(K-\bar{x})\right)$.

c. For $K = 0$, the answer in part (b) is $1 -$ (p-value) for $H_0\colon \theta \le 0$.

8.51 If $\alpha < p(\mathbf{x})$,

$$\sup_{\theta\in\Theta_0} P(W(\mathbf{X}) \ge c_\alpha) = \alpha < p(\mathbf{x}) = \sup_{\theta\in\Theta_0} P(W(\mathbf{X}) \ge W(\mathbf{x})).$$

Thus $W(\mathbf{x}) < c_\alpha$ and we could not reject H_0 at level α having observed $\mathbf{x}$. On the other hand, if $\alpha \ge p(\mathbf{x})$,

$$\sup_{\theta\in\Theta_0} P(W(\mathbf{X}) \ge c_\alpha) = \alpha \ge p(\mathbf{x}) = \sup_{\theta\in\Theta_0} P(W(\mathbf{X}) \ge W(\mathbf{x})).$$

Either $W(\mathbf{x}) \ge c_\alpha$ in which case we could reject H_0 at level α having observed $\mathbf{x}$ or $W(\mathbf{x}) < c_\alpha$. But, in the latter case we could use $c'_\alpha = W(\mathbf{x})$ and have $\{\mathbf{x}' : W(\mathbf{x}') \ge c'_\alpha\}$ define a size α rejection region. Then we could reject H_0 at level α having observed $\mathbf{x}$.

8.53 a.

$$P(-\infty < \theta < \infty) = \frac{1}{2} + \frac{1}{2}\frac{1}{\sqrt{2\pi\tau^2}}\int_{-\infty}^{\infty} e^{-\theta^2/(2\tau^2)}\, d\theta = \frac{1}{2} + \frac{1}{2} = 1.$$

b. First calculate the posterior density. Because

$$f(\bar{x}|\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-n(\bar{x}-\theta)^2/(2\sigma^2)},$$

we can calculate the marginal density as

$$\begin{aligned} m_\pi(\bar{x}) &= \frac{1}{2}f(\bar{x}|0) + \frac{1}{2}\int_{-\infty}^{\infty} f(\bar{x}|\theta)\frac{1}{\sqrt{2\pi}\tau}e^{-\theta^2/(2\tau^2)}d\theta \\ &= \frac{1}{2}\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-n\bar{x}^2/(2\sigma^2)} + \frac{1}{2}\frac{1}{\sqrt{2\pi}\sqrt{(\sigma^2/n)+\tau^2}}e^{-\bar{x}^2/[2((\sigma^2/n)+\tau^2)]} \end{aligned}$$

(see Exercise 7.22). Then $P(\theta = 0|\bar{x}) = \frac{1}{2}f(\bar{x}|0)/m_\pi(\bar{x})$.

c.

$$\begin{aligned} P\left(|\bar{X}| > \bar{x}\middle|\, \theta = 0\right) &= 1 - P\left(|\bar{X}| \le \bar{x}\middle|\, \theta = 0\right) \\ &= 1 - P\left(-\bar{x} \le \bar{X} \le \bar{x}\middle|\, \theta = 0\right) = 2\left[1 - \Phi\left(\bar{x}/(\sigma/\sqrt{n})\right)\right], \end{aligned}$$

where Φ is the standard normal cdf.

d. For $\sigma^2 = \tau^2 = 1$ and $n = 9$ we have a p-value of $2\left(1 - \Phi(3\bar{x})\right)$ and

$$P\left(\theta = 0|\, \bar{x}\right) = \left(1 + \sqrt{\frac{1}{10}}e^{81\bar{x}^2/20}\right)^{-1}.$$

The p-value of $\bar{x}$ is usually smaller than the Bayes posterior probability except when $\bar{x}$ is very close to the θ value specified by H_0. The following table illustrates this.

Some p-values and posterior probabilities ($n = 9$)

	$\bar{x}$								
	0	$\pm.1$	$\pm.15$	$\pm.2$	$\pm.5$	$\pm.6533$	$\pm.7$	±1	±2
p-value of $\bar{x}$	1	.7642	.6528	.5486	.1336	.05	.0358	.0026	≈ 0
posterior $P(\theta = 0\|\bar{x})$	.7597	.7523	.7427	.7290	.5347	.3595	.3030	.0522	≈ 0

8.54 a. From Exercise 7.22, the posterior distribution of $\theta|\mathbf{x}$ is normal with mean $[\tau^2/(\tau^2+\sigma^2/n)]\bar{x}$ and variance $\tau^2/(1+n\tau^2/\sigma^2)$. So

$$\begin{aligned} P(\theta \le 0|\mathbf{x}) &= P\left(Z \le \frac{0-[\tau^2/(\tau^2+\sigma^2/n)]\bar{x}}{\sqrt{\tau^2/(1+n\tau^2/\sigma^2)}}\right) \\ &= P\left(Z \le -\frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2+\sigma^2/n)}}\bar{x}\right) \;=\; P\left(Z \ge \frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2+\sigma^2/n)}}\bar{x}\right). \end{aligned}$$

b. Using the fact that if $\theta = 0$, $\bar{X} \sim \mathrm{n}(0, \sigma^2/n)$, the p-value is

$$P(\bar{X} \ge \bar{x}) = P\left(Z \ge \frac{\bar{x}-0}{\sigma/\sqrt{n}}\right) = P\left(Z \ge \frac{1}{\sigma/\sqrt{n}}\bar{x}\right)$$

c. For $\sigma^2 = \tau^2 = 1$,

$$P(\theta \le 0|x) = P\left(Z \ge \frac{1}{\sqrt{(1/n)(1+1/n)}}\bar{x}\right) \quad \text{and} \quad P(\bar{X} \ge \bar{x}) = P\left(Z \ge \frac{1}{\sqrt{1/n}}\bar{x}\right).$$

Because

$$\frac{1}{\sqrt{(1/n)(1+1/n)}} < \frac{1}{\sqrt{1/n}},$$

the Bayes probability is larger than the p-value if $\bar{x} \ge 0$. (Note: The inequality is in the opposite direction for $\bar{x} < 0$, but the primary interest would be in large values of $\bar{x}$.)

d. As $\tau^2 \to \infty$, the constant in the Bayes probability,

$$\frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2+\sigma^2/n)}} = \frac{1}{\sqrt{(\sigma^2/n)(1+\sigma^2/(\tau^2 n))}} \to \frac{1}{\sigma/\sqrt{n}},$$

the constant in the p-value. So the indicated equality is true.

8.55 The formulas for the risk functions are obtained from (8.3.14) using the power function $\beta(\theta) = \Phi(-z_\alpha + \theta_0 - \theta)$, where Φ is the standard normal cdf.

8.57 For 0–1 loss by (8.3.12) the risk function for any test is the power function $\beta(\mu)$ for $\mu \le 0$ and $1-\beta(\mu)$ for $\mu > 0$. Let $\alpha = P(1 < Z < 2)$, the size of test δ. By the Karlin-Rubin Theorem, the test δ_{z_α} that rejects if $X > z_\alpha$ is also size α and is uniformly more powerful than δ, that is, $\beta_{\delta_{z_\alpha}}(\mu) > \beta_\delta(\mu)$ for all $\mu > 0$. Hence,

$$R(\mu, \delta_{z_\alpha}) = 1 - \beta_{\delta_{z_\alpha}}(\mu) < 1 - \beta_\delta(\mu) = R(\mu, \delta), \quad \text{for all } \mu > 0.$$

Now reverse the roles of H_0 and H_1 and consider testing $H_0^*\colon \mu > 0$ versus $H_1^*\colon \mu \le 0$. Consider the test δ^* that rejects H_0^* if $X \le 1$ or $X \ge 2$, and the test $\delta^*_{z_\alpha}$ that rejects H_0^* if $X \le z_\alpha$. It is easily verified that for 0–1 loss δ and δ^* have the same risk functions, and $\delta^*_{z_\alpha}$ and δ_{z_α} have the same risk functions. Furthermore, using the Karlin-Rubin Theorem as before, we can conclude that $\delta^*_{z_\alpha}$ is uniformly more powerful than δ^*. Thus we have

$$R(\mu, \delta) = R(\mu, \delta^*) \ge R(\mu, \delta^*_{z_\alpha}) = R(\mu, \delta_{z_\alpha}), \quad \text{for all } \mu \le 0,$$

with strict inequality if $\mu < 0$. Thus, δ_{z_α} is better than δ.

Chapter 9

Interval Estimation

9.1 Denote $A = \{x\colon L(x) \le \theta\}$ and $B = \{x\colon U(x) \ge \theta\}$. Then $A \cap B = \{x\colon L(x) \le \theta \le U(x)\}$ and $1 \ge P\{A \cup B\} = P\{L(X) \le \theta \text{ or } \theta \le U(X)\} \ge P\{L(X) \le \theta \text{ or } \theta \le L(X)\} = 1$, since $L(x) \le U(x)$. Therefore, $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 - \alpha_1 + 1 - \alpha_2 - 1 = 1 - \alpha_1 - \alpha_2$.

9.3 a. The MLE of β is $X_{(n)} = \max X_i$. Since β is a scale parameter, $X_{(n)}/\beta$ is a pivot, and

$$.05 = P_\beta(X_{(n)}/\beta \le c) = P_\beta(\text{all } X_i \le c\beta) = \left(\frac{c\beta}{\beta}\right)^{\alpha_0 n} = c^{\alpha_0 n}$$

implies $c = .05^{1/\alpha_0 n}$. Thus, $.95 = P_\beta(X_{(n)}/\beta > c) = P_\beta(X_{(n)}/c > \beta)$, and $\{\beta : \beta < X_{(n)}/(.05^{1/\alpha_0 n})\}$ is a 95% upper confidence limit for β.

b. From 7.10, $\hat\alpha = 12.59$ and $X_{(n)} = 25$. So the confidence interval is $(0, 25/[.05^{1/(12.59 \cdot 14)}]) = (0, 25.43)$.

9.4 a.

$$\lambda(x,y) = \frac{\sup_{\lambda=\lambda_0} L\left(\sigma_X^2, \sigma_Y^2 \middle| x, y\right)}{\sup_{\lambda \in (0,+\infty)} L\left(\sigma_X^2, \sigma_Y^2 \middle| x, y\right)}$$

The unrestricted MLEs of σ_X^2 and σ_Y^2 are $\hat\sigma_X^2 = \frac{\Sigma X_i^2}{n}$ and $\hat\sigma_Y^2 = \frac{\Sigma Y_i^2}{m}$, as usual. Under the restriction, $\lambda = \lambda_0$, $\sigma_Y^2 = \lambda_0 \sigma_X^2$, and

$$\begin{aligned} L\left(\sigma_X^2, \lambda_0\sigma_X^2 \middle| x, y\right) &= \left(2\pi\sigma_X^2\right)^{-n/2} \left(2\pi\lambda_0\sigma_X^2\right)^{-m/2} e^{-\Sigma x_i^2/(2\sigma_X^2)} \cdot e^{-\Sigma y_i^2/(2\lambda_0\sigma_X^2)} \\ &= \left(2\pi\sigma_X^2\right)^{-(m+n)/2} \lambda_0^{-m/2} e^{-(\lambda_0\Sigma x_i^2 + \Sigma y_i^2)/(2\lambda_0\sigma_X^2)} \end{aligned}$$

Differentiating the log likelihood gives

$$\begin{aligned} \frac{d \log L}{d\left(\sigma_X^2\right)^2} &= \frac{d}{d\sigma_X^2}\left[-\frac{m+n}{2}\log\sigma_X^2 - \frac{m+n}{2}\log(2\pi) - \frac{m}{2}\log\lambda_0 - \frac{\lambda_0\Sigma x_i^2 + \Sigma y_i^2}{2\lambda_0\sigma_X^2}\right] \\ &= -\frac{m+n}{2}\left(\sigma_X^2\right)^{-1} + \frac{\lambda_0\Sigma x_i^2 + \Sigma y_i^2}{2\lambda_0}\left(\sigma_X^2\right)^{-2} \stackrel{\text{set}}{=} 0 \end{aligned}$$

which implies

$$\hat\sigma_0^2 = \frac{\lambda_0\Sigma x_i^2 + \Sigma y_i^2}{\lambda_0(m+n)}.$$

To see this is a maximum, check the second derivative:

$$\begin{aligned} \frac{d^2 \log L}{d\left(\sigma_X^2\right)^2} &= \frac{m+n}{2}\left(\sigma_X^2\right)^{-2} - \frac{1}{\lambda_0}\left(\lambda_0\Sigma x_i^2 + \Sigma y_i^2\right)\left(\sigma_X^2\right)^{-3}\Big|_{\sigma_X^2 = \hat\sigma_0^2} \\ &= -\frac{m+n}{2}(\hat\sigma_0^2)^{-2} < 0, \end{aligned}$$

therefore $\hat\sigma_0^2$ is the MLE. The LRT statistic is

$$\frac{\left(\hat\sigma_X^2\right)^{n/2}\left(\hat\sigma_Y^2\right)^{m/2}}{\lambda_0^{m/2}\left(\hat\sigma_0^2\right)^{(m+n)/2}},$$

and the test is: Reject H_0 if $\lambda(x,y)<k$, where k is chosen to give the test size α.

b. Under H_0, $\sum Y_i^2/(\lambda_0\sigma_X^2)\sim\chi_m^2$ and $\sum X_i^2/\sigma_X^2\sim\chi_n^2$, independent. Also, we can write

$$\begin{aligned}\lambda(X,Y) &= \left(\frac{1}{\frac{n}{m+n}+\frac{(\Sigma Y_i^2/\lambda_0\sigma_X^2)/m}{(\Sigma X_i^2/\sigma_X^2)/n}\cdot\frac{m}{m+n}}\right)^{n/2}\left(\frac{1}{\frac{m}{m+n}+\frac{(\Sigma X_i^2/\sigma_X^2)/n}{(\Sigma Y_i^2/\lambda_0\sigma_X^2)/m}\cdot\frac{n}{m+n}}\right)^{m/2}\\ &= \left[\frac{1}{\frac{n}{n+m}+\frac{m}{m+n}F}\right]^{n/2}\left[\frac{1}{\frac{m}{m+n}+\frac{n}{m+n}F^{-1}}\right]^{m/2}\end{aligned}$$

where $F=\frac{\Sigma Y_i^2/\lambda_0 m}{\Sigma X_i^2/n}\sim F_{m,n}$ under H_0. The rejection region is

$$\left\{(x,y)\colon \frac{1}{\left[\frac{n}{n+m}+\frac{m}{m+n}F\right]^{n/2}}\cdot\frac{1}{\left[\frac{m}{m+n}+\frac{n}{m+n}F^{-1}\right]^{m/2}}<c_\alpha\right\}$$

where c_α is chosen to satisfy

$$P\left\{\left[\frac{n}{n+m}+\frac{m}{m+n}F\right]^{-n/2}\left[\frac{m}{n+m}+\frac{n}{m+n}F^{-1}\right]^{-m/2}<c_\alpha\right\}=\alpha.$$

c. To ease notation, let $a=m/(n+m)$ and $b=a\sum y_i^2/\sum x_i^2$. From the duality of hypothesis tests and confidence sets, the set

$$c(\lambda)=\left\{\lambda\colon\left(\frac{1}{a+b/\lambda}\right)^{n/2}\left(\frac{1}{(1-a)+\frac{a(1-a)}{b}\lambda}\right)^{m/2}\ge c_\alpha\right\}$$

is a $1-\alpha$ confidence set for λ. We now must establish that this set is indeed an interval. To do this, we establish that the function on the left hand side of the inequality has only an interior maximum. That is, it looks like an upside-down bowl. Furthermore, it is straightforward to establish that the function is zero at both $\lambda=0$ and $\lambda=\infty$. These facts imply that the set of λ values for which the function is greater than or equal to c_α must be an interval. We make some further simplifications. If we multiply both sides of the inequality by $[(1-a)/b]^{m/2}$, we need be concerned with only the behavior of the function

$$h(\lambda)=\left(\frac{1}{a+b/\lambda}\right)^{n/2}\left(\frac{1}{b+a\lambda}\right)^{m/2}.$$

Moreover, since we are most interested in the sign of the derivative of h, this is the same as the sign of the derivative of $\log h$, which is much easier to work with. We have

$$\begin{aligned}\frac{d}{d\lambda}\log h(\lambda) &= \frac{d}{d\lambda}\left[-\frac{n}{2}\log(a+b/\lambda)-\frac{m}{2}\log(b+a\lambda)\right]\\ &= \frac{n}{2}\frac{b/\lambda^2}{a+b/\lambda}-\frac{m}{2}\frac{a}{b+a\lambda}\\ &= \frac{1}{2\lambda^2(a+b/\lambda)(b+a\lambda)}\left[-a^2m\lambda^2+ab(n-m)\lambda+nb^2\right].\end{aligned}$$

The sign of the derivative is given by the expression in square brackets, a parabola. It is easy to see that for $\lambda \ge 0$, the parabola changes sign from positive to negative. Since this is the sign change of the derivative, the function must increase then decrease. Hence, the function is an upside-down bowl, and the set is an interval.

9.5 a. Analogous to Example 9.2.5, the test here will reject H_0 if $T < k(p_0)$. Thus the confidence set is $C = \{p\colon T \ge k(p)\}$. Since $k(p)$ is nondecreasing, this gives an upper bound on p.

b. $k(p)$ is the integer that simultaneously satisfies

$$\sum_{y=k(p)}^{n} \binom{n}{y} p^y (1-p)^{n-y} \ge 1-\alpha \quad \text{and} \quad \sum_{y=k(p)+1}^{n} \binom{n}{y} p^y (1-p)^{n-y} < 1-\alpha.$$

9.6 a. For $Y = \sum X_i \sim \text{binomial}(n,p)$, the LRT statistic is

$$\lambda(y) = \frac{\binom{n}{y} p_0^y (1-p_0)^{n-y}}{\binom{n}{y} \hat{p}^y (1-\hat{p})^{n-y}} = \left(\frac{p_0(1-\hat{p})}{\hat{p}(1-p_0)}\right)^y \left(\frac{1-p_0}{1-\hat{p}}\right)^n$$

where $\hat{p} = y/n$ is the MLE of p. The acceptance region is

$$A(p_0) = \left\{ y\colon \left(\frac{p_0}{\hat{p}}\right)^y \left(\frac{1-p_0}{1-\hat{p}}\right)^{n-y} \ge k^* \right\}$$

where k^* is chosen to satisfy $P_{p_0}(Y \in A(p_0)) = 1-\alpha$. Inverting the acceptance region to a confidence set, we have

$$C(y) = \left\{ p\colon \left(\frac{p}{\hat{p}}\right)^y \left(\frac{(1-p)}{1-\hat{p}}\right)^{n-y} \ge k^* \right\}.$$

b. For given n and observed y, write

$$C(y) = \left\{ p\colon (n/y)^y \,(n/(n-y))^{n-y}\, p^y (1-p)^{n-y} \ge k^* \right\}.$$

This is clearly a highest density region. The endpoints of $C(y)$ are roots of the n^{th} degree polynomial (in p), $(n/y)^y\,(n/(n-y))^{n-y}\,p^y(1-p)^{n-y} - k^*$. The interval of (10.4.4) is

$$\left\{ p\colon \left| \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \right| \le z_{\alpha/2} \right\}.$$

The endpoints of this interval are the roots of the second degree polynomial (in p), $(\hat{p}-p)^2 - z_{\alpha/2}^2 p(1-p)/n$. Typically, the second degree and n^{th} degree polynomials will not have the same roots. Therefore, the two intervals are different. (Note that when $n \to \infty$ and $y \to \infty$, the density becomes symmetric (CLT). Then the two intervals are the same.)

9.7 These densities have already appeared in Exercise 8.8, where LRT statistics were calculated for testing $H_0\colon a = 1$.

a. Using the result of Exercise 8.8(a), the restricted MLE of θ (when $a = a_0$) is

$$\hat{\theta}_0 = \frac{-a_0 + \sqrt{a_0^2 + 4\sum x_i^2/n}}{2},$$

and the unrestricted MLEs are

$$\hat{\theta} = \bar{x} \quad \text{and} \quad \hat{a} = \frac{\sum (x_i - \bar{x})^2}{n\bar{x}}.$$

The LRT statistic is

$$\lambda(x) = \frac{\left(\frac{\hat{a}\hat{\theta}}{a_0\hat{\theta}_0}\right)^{n/2} e^{-\frac{1}{2a_0\hat{\theta}_0}\Sigma(x_i-\hat{\theta}_0)^2}}{e^{-\frac{1}{2\hat{a}\hat{\theta}}\Sigma(x_i-\hat{\theta})^2}} = \left(\frac{1}{2\pi a_0\hat{\theta}_0}\right)^{n/2} e^{n/2} e^{-\frac{1}{2a_0\hat{\theta}_0}\Sigma(x_i-\hat{\theta}_0)^2}$$

The rejection region of a size α test is $\{x\colon \lambda(x) \le c_\alpha\}$, and a $1-\alpha$ confidence set is $\{a_0\colon \lambda(x) \ge c_\alpha\}$.

b. Using the results of Exercise 8.8b, the restricted MLE (for $a = a_0$) is found by solving

$$-a_0\theta^2 + [\hat{\sigma}^2 + (\bar{x}-\theta)^2] + \theta(\bar{x}-\theta) = 0,$$

yielding the MLE

$$\hat{\theta}_R = \bar{x} + \sqrt{\bar{x} + 4a_0(\hat{\sigma}^2 + \bar{x}^2)}/2a_0.$$

The unrestricted MLEs are

$$\hat{\theta} = \bar{x} \quad \text{and} \quad \hat{a} = \frac{1}{n\bar{x}^2}\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\hat{\sigma}^2}{\bar{x}^2},$$

yielding the LRT statistic

$$\lambda(x) = \left(\hat{\sigma}/\hat{\theta}_R\right)^n e^{(n/2)-\Sigma(x_i-\hat{\theta}_R)^2/(2\hat{\theta}_R)}.$$

The rejection region of a size α test is $\{x\colon \lambda(x) \le c_\alpha\}$, and a $1-\alpha$ confidence set is $\{a_0\colon \lambda(x) \ge c_\alpha\}$.

9.9 Let $Z_1, \ldots, Z_n$ be iid with pdf $f(z)$.

a. For $X_i \sim f(x-\mu)$, $(X_1,\ldots,X_n) \sim (Z_1+\mu,\ldots,Z_n+\mu)$, and $\bar{X}-\mu \sim \overline{Z+\mu}-\mu = \bar{Z}$. The distribution of $\bar{Z}$ does not depend on μ.

b. For $X_i \sim f(x/\sigma)/\sigma$, $(X_1,\ldots,X_n) \sim (\sigma Z_1,\ldots,\sigma Z_n)$, and $\bar{X}/\sigma \sim \overline{\sigma Z}/\sigma = \bar{Z}$. The distribution of $\bar{Z}$ does not depend on σ.

c. For $X_i \sim f((x-\mu)/\sigma)/\sigma$, $(X_1,\ldots,X_n) \sim (\sigma Z_1+\mu,\ldots,\sigma Z_n+\mu)$, and $(\bar{X}-\mu)/S_X \sim (\overline{\sigma Z+\mu}-\mu)/S_{\sigma Z+\mu} = \sigma\bar{Z}/(\sigma S_Z) = \bar{Z}/S_Z$. The distribution of $\bar{Z}/S_Z$ does not depend on μ or σ.

9.11 Recall that if θ is the true parameter, then $F_T(T|\theta) \sim \text{uniform}(0,1)$. Thus,

$$P_{\theta_0}(\{T\colon \alpha_1 \le F_T(T|\theta_0) \le 1-\alpha_2\}) = P(\alpha_1 \le U \le 1-\alpha_2) = 1-\alpha_2-\alpha_1,$$

where $U \sim \text{uniform}(0,1)$. Since

$$t \in \{t\colon \alpha_1 \le F_T(t|\theta) \le 1-\alpha_2\} \quad \Leftrightarrow \quad \theta \in \{\theta : \alpha_1 \le F_T(t|\theta) \le 1-\alpha_2\}$$

the same calculation shows that the interval has confidence $1-\alpha_2-\alpha_1$.

9.12 If $X_1,\ldots,X_n \sim$ iid $\text{n}(\theta,\theta)$, then $\sqrt{n}(\bar{X}-\theta)/\sqrt{\theta} \sim \text{n}(0,1)$ and a $1-\alpha$ confidence interval is $\{\theta\colon |\sqrt{n}(\bar{x}-\theta)/\sqrt{\theta}| \le z_{\alpha/2}\}$. Solving for θ, we get

$$\left\{\theta\colon n\theta^2 - \theta\left(2n\bar{x} + z^2_{\alpha/2}\right) + n\bar{x}^2 \le 0\right\} = \left\{\theta\colon \theta \in \left(2n\bar{x} + z^2_{\alpha/2} \pm \sqrt{4n\bar{x}z^2_{\alpha/2} + z^4_{\alpha/2}}\right)/2n\right\}.$$

Simpler answers can be obtained using the t pivot, $(\bar{X}-\theta)/(S/\sqrt{n})$, or the χ^2 pivot, $(n-1)S^2/\theta^2$. (Tom Werhley of Texas A&M university notes the following: The largest probability of getting a negative discriminant (hence empty confidence interval) occurs when $\sqrt{n\theta} = \frac{1}{2}z_{\alpha/2}$, and the probability is equal to $\alpha/2$. The behavior of the intervals for negative values of $\bar{x}$ is also interesting. When $\bar{x} = 0$ the lefthand endpoint is also equal to 0, but when $\bar{x} < 0$, the lefthand endpoint is positive. Thus, the interval based on $\bar{x} = 0$ contains smaller values of θ than that based on $\bar{x} < 0$. The intervals get smaller as $\bar{x}$ decreases, finally becoming empty.)

9.13 a. For $Y = -(\log X)^{-1}$, the pdf of Y is $f_Y(y) = \frac{\theta}{y^2}e^{-\theta/y}$, $0 < y < \infty$, and

$$P(Y/2 \le \theta \le Y) = \int_\theta^{2\theta} \frac{\theta}{y^2} e^{-\theta/y} dy = e^{-\theta/y}\Big|_\theta^{2\theta} = e^{-1/2} - e^{-1} = .239.$$

b. Since $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$, $T = X^\theta$ is a good guess at a pivot, and it is since $f_T(t) = 1$, $0 < t < 1$. Thus a pivotal interval is formed from $P(a < X^\theta < b) = b - a$ and is

$$\left\{\theta\colon \frac{\log b}{\log x} \le \theta \le \frac{\log a}{\log x}\right\}.$$

Since $X^\theta \sim \text{uniform}(0,1)$, the interval will have confidence .239 as long as $b - a = .239$.

c. The interval in part a) is a special case of the one in part b). To find the best interval, we minimize $\log b - \log a$ subject to $b - a = 1 - \alpha$, or $b = 1 - \alpha + a$. Thus we want to minimize $\log(1 - \alpha + a) - \log a = \log\left(1+\frac{1-\alpha}{a}\right)$, which is minimized by taking a as big as possible. Thus, take $b = 1$ and $a = \alpha$, and the best $1 - \alpha$ pivotal interval is $\left\{\theta\colon 0 \le \theta \le \frac{\log\alpha}{\log x}\right\}$. Thus the interval in part a) is nonoptimal. A shorter interval with confidence coefficient .239 is $\{\theta\colon 0 \le \theta \le \log(1 - .239)/\log(x)\}$.

9.14 a. Recall the Bonferroni Inequality (1.2.9), $P(A_1 \cap A_2) \ge P(A_1) + P(A_2) - 1$. Let $A_1 = P(\text{interval covers } \mu)$ and $A_2 = P(\text{interval covers } \sigma^2)$. Use the interval (9.2.14), with $t_{n-1,\alpha/4}$ to get a $1 - \alpha/2$ confidence interval for μ. Use the interval after (9.2.14) with $b = \chi^2_{n-1,\alpha/4}$ and $a = \chi^2_{n-1,1-\alpha/4}$ to get a $1-\alpha/2$ confidence interval for σ. Then the natural simultaneous set is

$$\begin{aligned} C_a(x) &= \Bigg\{(\mu,\sigma^2)\colon \bar{x} - t_{n-1,\alpha/4}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,\alpha/4}\frac{s}{\sqrt{n}} \\ &\qquad \text{and } \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/4}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/4}}\Bigg\} \end{aligned}$$

and $P\left(C_a(X) \text{ covers } (\mu,\sigma^2)\right) = P(A_1 \cap A_2) \ge P(A_1) + P(A_2) - 1 = 2(1-\alpha/2) - 1 = 1-\alpha$.

b. If we replace the μ interval in a) by $\left\{\mu\colon \bar{x} - \frac{k\sigma}{\sqrt{n}} \le \mu \le \bar{x} + \frac{k\sigma}{\sqrt{n}}\right\}$ then $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \text{n}(0,1)$, so we use $z_{\alpha/4}$ and

$$C_b(x) = \left\{(\mu,\sigma^2)\colon \bar{x} - z_{\alpha/4}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/4}\frac{\sigma}{\sqrt{n}} \text{ and } \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/4}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/4}}\right\}$$

and $P\left(C_b(X) \text{ covers } (\mu,\sigma^2)\right) \ge 2(1-\alpha/2) - 1 = 1 - \alpha$.

c. The sets can be compared graphically in the (μ,σ) plane: C_a is a rectangle, since μ and σ^2 are treated independently, while C_b is a trapezoid, with larger σ^2 giving a longer interval. Their areas can also be calculated

$$\begin{aligned} \text{Area of } C_a &= \left[2t_{n-1,\alpha/4}\frac{s}{\sqrt{n}}\right]\left\{\sqrt{(n-1)s^2}\left(\frac{1}{\chi^2_{n-1,1-\alpha/4}} - \frac{1}{\chi^2_{n-1,\alpha/4}}\right)\right\} \\ \text{Area of } C_b &= \left[z_{\alpha/4}\frac{s}{\sqrt{n}}\left(\sqrt{\frac{n-1}{\chi^2_{n-1,1-\alpha/4}}} + \sqrt{\frac{n-1}{\chi^2_{n-1,\alpha/4}}}\right)\right] \\ &\quad \times \left\{\sqrt{(n-1)s^2}\left(\frac{1}{\chi^2_{n-1,1-\alpha/4}} - \frac{1}{\chi^2_{n-1,\alpha/4}}\right)\right\} \end{aligned}$$

and compared numerically.

9.15 Fieller's Theorem says that a $1-\alpha$ confidence set for $\theta = \mu_Y/\mu_X$ is

$$\left\{\theta\colon \left(\bar{x}^2 - \frac{t^2_{n-1,\alpha/2}}{n-1}s^2_X\right)\theta^2 - 2\left(\bar{x}\bar{y} - \frac{t^2_{n-1,\alpha/2}}{n-1}s_{YX}\right)\theta + \left(\bar{y}^2 - \frac{t^2_{n-1,\alpha/2}}{n-1}s^2_Y\right) \le 0\right\}.$$

a. Define $a = \bar{x}^2 - ts^2_X$, $b = \bar{x}\bar{y} - ts_{YX}$, $c = \bar{y}^2 - ts^2_Y$, where $t = \frac{t^2_{n-1,\alpha/2}}{n-1}$. Then the parabola opens upward if $a > 0$. Furthermore, if $a > 0$, then there always exists at least one real root. This follows from the fact that at $\theta = \bar{y}/\bar{x}$, the value of the function is negative. For $\bar{\theta} = \bar{y}/\bar{x}$ we have

$$\begin{aligned}
&\left(\bar{x}^2 - ts^2_X\right)\left(\frac{\bar{y}}{\bar{x}}\right)^2 - 2\left(\bar{x}\bar{y} - ts_{XY}\right)\left(\frac{\bar{y}}{\bar{x}}\right) + \left(\bar{y}^2 - as^2_Y\right) \\
&= -t\left[\frac{\bar{y}^2}{\bar{x}^2}s^2_X - 2\frac{\bar{y}}{\bar{x}}s_{XY} + s^2_Y\right] \\
&= -t\left[\sum_{i=1}^n\left(\frac{\bar{y}^2}{\bar{x}^2}(x_i-\bar{x})^2 - 2\frac{\bar{y}}{\bar{x}}(x_i-\bar{x})(y_i-\bar{y}) + (y_i-\bar{y})^2\right)\right] \\
&= -t\left[\sum_{i=1}^n\left(\frac{\bar{y}}{\bar{x}}(x_i-\bar{x}) - (y_i-\bar{y})\right)^2\right]
\end{aligned}$$

which is negative.

b. The parabola opens downward if $a < 0$, that is, if $\bar{x}^2 < ts^2_X$. This will happen if the test of $H_0\colon \mu_X = 0$ accepts H_0 at level α.

c. The parabola has no real roots if $b^2 < ac$. This can only occur if $a < 0$.

9.16 a. The LRT (see Example 8.2.1) has rejection region $\{x : |\bar{x} - \theta_0| > z_{\alpha/2}\sigma/\sqrt{n}\}$, acceptance region $A(\theta_0) = \{x\colon -z_{\alpha/2}\sigma/\sqrt{n} \le \bar{x} - \theta_0 \le z_{\alpha/2}\sigma/\sqrt{n}\}$, and $1-\alpha$ confidence interval $C(\theta) = \{\theta\colon \bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \theta \le \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\}$.

b. We have a UMP test with rejection region $\{x : \bar{x} - \theta_0 < -z_\alpha\sigma/\sqrt{n}\}$, acceptance region $A(\theta_0) = \{x\colon \bar{x} - \theta_0 \ge -z_\alpha\sigma/\sqrt{n}\}$, and $1-\alpha$ confidence interval $C(\theta) = \{\theta\colon \bar{x} + z_\alpha\sigma/\sqrt{n} \ge \theta\}$.

c. Similar to b), the UMP test has rejection region $\{x\colon \bar{x} - \theta_0 > z_\alpha\sigma/\sqrt{n}\}$, acceptance region $A(\theta_0) = \{x\colon \bar{x} - \theta_0 \le z_\alpha\sigma/\sqrt{n}\}$, and $1-\alpha$ confidence interval $C(\theta) = \{\theta\colon \bar{x} - z_\alpha\sigma/\sqrt{n} \le \theta\}$.

9.17 a. Since $X - \theta \sim$ uniform$(-1/2, 1/2)$, $P(a \le X - \theta \le b) = b - a$. Any a and b satisfying $b = a + 1 - \alpha$ will do. One choice is $a = -\frac{1}{2} + \frac{\alpha}{2}$, $b = \frac{1}{2} - \frac{\alpha}{2}$.

b. Since $T = X/\theta$ has pdf $f(t) = 2t$, $0 \le t \le 1$,

$$P(a \le X/\theta \le b) = \int_a^b 2t\,dt = b^2 - a^2.$$

Any a and b satisfying $b^2 = a^2 + 1 - \alpha$ will do. One choice is $a = \sqrt{\alpha/2}$, $b = \sqrt{1-\alpha/2}$.

9.18 a. $P_p(X = 1) = \binom{3}{1}p^1(1-p)^{3-1} = 3p(1-p)^2$, maximum at $p = 1/3$.
$P_p(X = 2) = \binom{3}{2}p^2(1-p)^{3-2} = 3p^2(1-p)$, maximum at $p = 2/3$.

b. $P(X = 0) = \binom{3}{0}p^0(1-p)^{3-0} = (1-p)^3$, and this is greater than $P(X = 2)$ if $(1-p)^2 > 3p^2$, or $2p^2 + 2p - 1 < 0$. At $p = 1/3$, $2p^2 + 2p - 1 = -1/9$.

c. To show that this is a $1 - \alpha = .442$ interval, compare with the interval in Example 9.2.11. There are only two discrepancies. For example,

$$P(p \in \text{interval} \mid .362 < p < .634) = P(X = 1 \text{ or } X = 2) > .442$$

by comparison with Sterne's procedure, which is given by

x	interval
0	[.000,.305)
1	[.305,.634)
2	[.362,.762)
3	[.695,1].

9.19 For $F_T(t|\theta)$ increasing in θ, there are unique values $\theta_U(t)$ and $\theta_L(t)$ such that $F_T(t|\theta) < 1 - \frac{\alpha}{2}$ if and only if $\theta < \theta_U(t)$ and $F_T(t|\theta) > \frac{\alpha}{2}$ if and only if $\theta > \theta_L(t)$. Hence,

$$\begin{aligned} P\left(\theta_L(T) \le \theta \le \theta_U(T)\right) &= P\left(\theta \le \theta_U(T)\right) - P\left(\theta \le \theta_L(T)\right) \\ &= P\left(F_T(T) \le 1 - \frac{\alpha}{2}\right) - P\left(F_T(T) \le \frac{\alpha}{2}\right) \\ &= 1 - \alpha. \end{aligned}$$

9.21 To construct a $1-\alpha$ confidence interval for p of the form $\{p\colon \ell \le p \le u\}$ with $P(\ell \le p \le u) = 1-\alpha$, we use the method of Theorem 9.2.12. We must solve for ℓ and u in the equations

$$(1)\quad \frac{\alpha}{2} = \sum_{k=0}^{x} \binom{n}{k} u^k (1-u)^{n-k} \quad \text{and} \quad (2)\quad \frac{\alpha}{2} = \sum_{k=x}^{n} \binom{n}{k} \ell^k (1-\ell)^{n-k}.$$

In equation (1) $\alpha/2 = P(K \le x) = P(Y \le 1-u)$, where $Y \sim \text{beta}(n-x, x+1)$ and $K \sim \text{binomial}(n,u)$. This is Exercise 2.40. Let $Z \sim F_{2(n-x),2(x+1)}$ and $c = (n-x)/(x+1)$. By Theorem 5.3.8c, $cZ/(1+cZ) \sim \text{beta}(n-x, x+1) \sim Y$. So we want

$$\alpha/2 = P\left(\frac{cZ}{(1+cZ)} \le 1-u\right) = P\left(\frac{1}{Z} \ge \frac{cu}{1-u}\right).$$

From Theorem 5.3.8a, $1/Z \sim F_{2(x+1),2(n-x)}$. So we need $cu/(1-u) = F_{2(x+1),2(n-x),\alpha/2}$. Solving for u yields

$$u = \frac{\frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}.$$

A similar manipulation on equation (2) yields the value for ℓ.

9.23 a. The LRT statistic for $H_0\colon \lambda = \lambda_0$ versus $H_1\colon \lambda \ne \lambda_0$ is

$$g(y) = e^{-n\lambda_0}(n\lambda_0)^y / e^{-n\hat\lambda}(n\hat\lambda)^y,$$

where $Y = \sum X_i \sim \text{Poisson}(n\lambda)$ and $\hat\lambda = y/n$. The acceptance region for this test is $A(\lambda_0) = \{y\colon g(y) > c(\lambda_0))$ where $c(\lambda_0)$ is chosen so that $P(Y \in A(\lambda_0)) \ge 1-\alpha$. $g(y)$ is a unimodal function of y so $A(\lambda_0)$ is an interval of y values. Consider constructing $A(\lambda_0)$ for each $\lambda_0 > 0$. Then, for a fixed y, there will be a smallest λ_0, call it $a(y)$, and a largest λ_0, call it $b(y)$, such that $y \in A(\lambda_0)$. The confidence interval for λ is then $C(y) = (a(y), b(y))$. The values $a(y)$ and $b(y)$ are not expressible in closed form. They can be determined by a numerical search, constructing $A(\lambda_0)$ for different values of λ_0 and determining those values for which $y \in A(\lambda_0)$. (Jay Beder of the University of Wisconsin, Milwaukee, reminds us that since c is a function of λ, the resulting confidence set need not be a highest density region of a likelihood function. This is an example of the effect of the imposition of one type of inference (frequentist) on another theory (likelihood).)

b. The procedure in part a) was carried out for $y = 558$ and the confidence interval was found to be $(57.78, 66.45)$. For the confidence interval in Example 9.2.15, we need the values $\chi^2_{1116,.95} = 1039.444$ and $\chi^2_{1118,.05} = 1196.899$. This confidence interval is $(1039.444/18, 1196.899/18) = (57.75, 66.49)$. The two confidence intervals are virtually the same.

9.25 The confidence interval derived by the method of Section 9.2.3 is

$$C(y)=\left\{\mu\colon y+\frac{1}{n}\log\left(\frac{\alpha}{2}\right)\le\mu\le y+\frac{1}{n}\log\left(1-\frac{\alpha}{2}\right)\right\}$$

where $y=\min_i x_i$. The LRT method derives its interval from the test of $H_0:\mu=\mu_0$ versus $H_1\colon\mu\ne\mu_0$. Since Y is sufficient for μ, we can use $f_Y(y\mid\mu)$. We have

$$\begin{aligned}\lambda(y) &= \frac{\sup_{\mu=\mu_0}L(\mu|y)}{\sup_{\mu\in(-\infty,\infty)}L(\mu|y)} = \frac{ne^{-n}(y-\mu_0)I_{[\mu_0,\infty)(y)}}{ne^{-(y-y)}I_{[\mu,\infty)(y)}}\\ &= e^{-n(y-\mu_0)}I_{[\mu_0,\infty)}(y) = \begin{cases}0 & \text{if } y<\mu_0\\ e^{-n(y-\mu_0)} & \text{if } y\ge\mu_0.\end{cases}\end{aligned}$$

We reject H_0 if $\lambda(y)=e^{-n(y-\mu_0)}<c_\alpha$, where $0\le c_\alpha\le 1$ is chosen to give the test level α. To determine c_α, set

$$\begin{aligned}\alpha &= P\{\text{reject } H_0|\,\mu=\mu_0\} = P\left\{Y>\mu_0-\frac{\log c_\alpha}{n}\text{ or } Y<\mu_0\,\middle|\,\mu=\mu_0\right\}\\ &= P\left\{Y>\mu_0-\frac{\log c_\alpha}{n}\,\middle|\,\mu=\mu_0\right\} = \int_{\mu_0-\frac{\log c_\alpha}{n}}^{\infty} ne^{-n(y-\mu_0)}\,dy\\ &= -e^{-n(y-\mu_0)}\Big|_{\mu_0-\frac{\log c_\alpha}{n}}^{\infty} = e^{\log c_\alpha} = c_\alpha.\end{aligned}$$

Therefore, $c_\alpha=\alpha$ and the $1-\alpha$ confidence interval is

$$C(y)=\left\{\mu\colon\mu\le y\le\mu-\frac{\log\alpha}{n}\right\}=\left\{\mu\colon y+\frac{1}{n}\ \log\alpha\le\mu\le y\right\}.$$

To use the pivotal method, note that since μ is a location parameter, a natural pivotal quantity is $Z=Y-\mu$. Then, $f_Z(z)=ne^{-nz}I_{(0,\infty)}(z)$. Let $P\{a\le Z\le b\}=1-\alpha$, where a and b satisfy

$$\begin{aligned}\frac{\alpha}{2}=\int_0^a ne^{-nz}\,dz=-e^{-nz}\Big|_0^a=1-e^{-na} &\quad\Rightarrow\quad e^{-na}=1-\frac{\alpha}{2}\\ &\quad\Rightarrow\quad a=\frac{-\log\left(1-\frac{\alpha}{2}\right)}{n}\\ \frac{\alpha}{2}=\int_b^\infty ne^{-nz}\,dz=-e^{-nz}\Big|_b^\infty=e^{-nb} &\quad\Rightarrow\quad -nb=\log\frac{\alpha}{2}\\ &\quad\Rightarrow\quad b=-\frac{1}{n}\log\left(\frac{\alpha}{2}\right)\end{aligned}$$

Thus, the pivotal interval is $Y+\log(\alpha/2)/n\le\mu\le Y+\log(1-\alpha/2)$, the same interval as from Example 9.2.13. To compare the intervals we compare their lengths. We have

$$\begin{aligned}\text{Length of LRT interval} &= y-(y+\frac{1}{n}\log\alpha) = -\frac{1}{n}\log\alpha\\ \text{Length of Pivotal interval} &= \left(y+\frac{1}{n}\log(1-\alpha/2)\right)-(y+\frac{1}{n}\log\alpha/2) = \frac{1}{n}\log\frac{1-\alpha/2}{\alpha/2}\end{aligned}$$

Thus, the LRT interval is shorter if $-\log\alpha<\log[(1-\alpha/2)/(\alpha/2)]$, but this is always satisfied.

9.27 a. $Y=\sum X_i\sim\text{gamma}(n,\lambda)$, and the posterior distribution of λ is

$$\pi(\lambda|y)=\frac{(y+\frac{1}{b})^{n+a}}{\Gamma(n+a)}\frac{1}{\lambda^{n+a+1}}e^{-\frac{1}{\lambda}(y+\frac{1}{b})},$$

an IG $\left(n+a,(y+\frac{1}{b})^{-1}\right)$. The Bayes HPD region is of the form $\{\lambda\colon \pi(\lambda|y)\geq k\}$, which is an interval since $\pi(\lambda|y)$ is unimodal. It thus has the form $\{\lambda\colon a_1(y)\leq\lambda\leq a_2(y)\}$, where a_1 and a_2 satisfy

$$\frac{1}{{a_1}^{n+a+1}}e^{-\frac{1}{a_1}(y+\frac{1}{b})}=\frac{1}{{a_2}^{n+a+1}}e^{-\frac{1}{a_2}(y+\frac{1}{b})}.$$

b. The posterior distribution is IG$(((n-1)/2)+a,(((n-1)s^2/2)+1/b)^{-1})$. So the Bayes HPD region is as in part a) with these parameters replacing $n+a$ and $y+1/b$.

c. As $a\to 0$ and $b\to\infty$, the condition on a_1 and a_2 becomes

$$\frac{1}{{a_1}^{((n-1)/2)+1}}e^{-\frac{1}{a_1}\frac{(n-1)s^2}{2}}=\frac{1}{{a_2}^{((n-1)/2)+1}}e^{-\frac{1}{a_2}\frac{(n-1)s^2}{2}}.$$

9.29 a. We know from Example 7.2.14 that if $\pi(p)\sim\text{beta}(a,b)$, the posterior is $\pi(p|y)\sim\text{beta}(y+a,n-y+b)$ for $y=\sum x_i$. So a $1-\alpha$ credible set for p is:

$$\{p\colon \beta_{y+a,n-y+b,1-\alpha/2}\leq p\leq\beta_{y+a,n-y+b,\alpha/2}\}.$$

b. Converting to an F distribution, $\beta_{c,d}=\frac{(c/d)F_{2c,2d}}{1+(c/d)F_{2c,2d}}$, the interval is

$$\frac{\frac{y+a}{n-y+b}F_{2(y+a),2(n-y+b),1-\alpha/2}}{1+\frac{y+a}{n-y+b}F_{2(y+a),2(n-y+b),1-\alpha/2}}\leq p\leq\frac{\frac{y+a}{n-y+b}F_{2(y+a),2(n-y+b),\alpha/2}}{1+\frac{y+a}{n-y+b}F_{2(y+a),2(n-y+b),\alpha/2}}$$

or, using the fact that $F_{m,n}=F^{-1}_{n,m}$,

$$\frac{1}{1+\frac{n-y+b}{y+a}F_{2(n-y+b),2(y+a),\alpha/2}}\leq p\leq\frac{\frac{y+a}{n-y+b}F_{2(y+a),2(n+b),\alpha/2}}{1+\frac{y+a}{n-y+b}F_{2(y+a),2(n-y+b),\alpha/2}}.$$

For this to match the interval of Exercise 9.21, we need $x=y$ and

$$\begin{aligned}\text{Lower limit: }\ n-y+b=n-x+1\quad&\Rightarrow\quad b=1\\ y+a=x\quad&\Rightarrow\quad a=0\\ \text{Upper limit: }\ y+a=x+1\quad&\Rightarrow\quad a=1\\ n-y+b=n-x\quad&\Rightarrow\quad b=0.\end{aligned}$$

So no values of a and b will make the intervals match.

9.31 a. We continually use the fact that given $Y=y$, χ^2_{2y} is a central χ^2 random variable with $2y$ degrees of freedom. Hence

$$\begin{aligned}\mathrm{E}\chi^2_{2Y}&=\mathrm{E}[\mathrm{E}(\chi^2_{2Y}|Y)]\ =\ \mathrm{E}2Y\ =\ 2\lambda\\ \mathrm{Var}\chi^2_{2Y}&=\mathrm{E}[\mathrm{Var}(\chi^2_{2Y}|Y)]+\mathrm{Var}[\mathrm{E}(\chi^2_{2Y}|Y)]\\ &=\mathrm{E}[4Y]+\mathrm{Var}[2Y]\ =\ 4\lambda+4\lambda\ =\ 8\lambda\\ \text{mgf}&=\mathrm{E}e^{t\chi^2_{2Y}}\ =\ \mathrm{E}[\mathrm{E}(e^{t\chi^2_{2Y}}|Y)]\ =\ \mathrm{E}\left(\frac{1}{1-2t}\right)^Y\\ &=\sum_{y=0}^{\infty}\frac{e^{-\lambda}\left(\frac{\lambda}{1-2t}\right)^y}{y!}\ =\ e^{-\lambda+\frac{\lambda}{1-2t}}.\end{aligned}$$

From Theorem 2.3.15, the mgf of $(\chi^2_{2Y}-2\lambda)/\sqrt{8\lambda}$ is

$$e^{-t\sqrt{\lambda/2}}\left[e^{-\lambda+\frac{\lambda}{1-t/\sqrt{2\lambda}}}\right].$$

The log of this is

$$-\sqrt{\lambda/2}t - \lambda + \frac{\lambda}{1 - t/\sqrt{2\lambda}} = \frac{t^2\sqrt{\lambda}}{-t\sqrt{2} + 2\sqrt{\lambda}} = \frac{t^2}{-(t\sqrt{2}/\sqrt{\lambda}) + 2} \to t^2/2 \text{ as } \lambda \to \infty,$$

so the mgf converges to $e^{t^2/2}$, the mgf of a standard normal.

b. Since $P(\chi^2_{2Y} \le \chi^2_{2Y,\alpha}) = \alpha$ for all λ,

$$\frac{\chi^2_{2Y,\alpha} - 2\lambda}{\sqrt{8\lambda}} \to z_\alpha \text{ as } \lambda \to \infty.$$

In standardizing (9.2.22), the upper bound is

$$\frac{\frac{nb}{nb+1}\chi^2_{2(Y+a),\alpha/2} - 2\lambda}{\sqrt{8\lambda}} = \sqrt{\frac{8(\lambda+a)}{8\lambda}}\left[\frac{\frac{nb}{nb+1}[\chi^2_{2(Y+a),\alpha/2} - 2(\lambda+a)]}{\sqrt{8(\lambda+a)}} + \frac{\frac{nb}{nb+1}2(\lambda+a) - 2\lambda}{\sqrt{8(\lambda+a)}}\right].$$

While the first quantity in square brackets $\to z_{\alpha/2}$, the second one has limit

$$\lim_{\lambda\to\infty} \frac{-2\frac{1}{nb+1}\lambda + a\frac{nb}{nb+1}}{\sqrt{8(\lambda+a)}} \to -\infty,$$

so the coverage probability goes to zero.

9.33 a. Since $0 \in C_a(x)$ for every x, $P(0 \in C_a(X)|\, \mu = 0) = 1$. If $\mu > 0$,

$$\begin{aligned} P(\mu \in C_a(X)) &= P(\mu \le \max\{0, X+a\}) = P(\mu \le X + a) && (\text{since } \mu > 0) \\ &= P(Z \ge -a) && (Z \sim \mathrm{n}(0,1)) \\ &= .95 && (a = 1.645.) \end{aligned}$$

A similar calculation holds for $\mu < 0$.

b. The credible probability is

$$\begin{aligned} \int_{\min(0,x-a)}^{\max(0,x+a)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu-x)^2}\, d\mu &= \int_{\min(-x,-a)}^{\max(-x,a)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt \\ &= P\left(\min(-x,-a) \le Z \le \max(-x,a)\right). \end{aligned}$$

To evaluate this probability we have two cases:

$$\begin{array}{llll} \text{(i)} & |x| \le a & \Rightarrow & \text{credible probability} = P(|Z| \le a) \\ \text{(ii)} & |x| > a & \Rightarrow & \text{credible probability} = P(-a \le Z \le |x|) \end{array}$$

Thus we see that for $a = 1.645$, the credible probability is equal to .90 if $|x| \le 1.645$ and increases to .95 as $|x| \to \infty$.

9.34 a. A $1-\alpha$ confidence interval for μ is $\{\mu : \bar{x} - 1.96\sigma/\sqrt{n} \le \mu \le \bar{x} + 1.96\sigma/\sqrt{n}\}$. We need $2(1.96)\sigma/\sqrt{n} \le \sigma/4$ or $\sqrt{n} \ge 4(2)(1.96)$. Thus we need $n \ge 64(1.96)^2 \approx 245.9$. So $n = 246$ suffices.

b. The length of a 95% confidence interval is $2t_{n-1,.025}S/\sqrt{n}$. Thus we need

$$\begin{aligned} P\left(2t_{n-1,.025}\frac{S}{\sqrt{n}} \le \frac{\sigma}{4}\right) \ge .9 \quad &\Rightarrow \quad P\left(4t^2_{n-1,.025}\frac{S^2}{n} \le \frac{\sigma^2}{16}\right) \ge .9 \\ &\Rightarrow \quad P\left(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\sim\chi^2_{n-1}} \le \frac{(n-1)n}{t^2_{n-1,.025}\cdot 64}\right) \ge .9. \end{aligned}$$

We need to solve this numerically for the smallest n that satisfies the inequality

$$\frac{(n-1)n}{t^2_{n-1,.025}\cdot 64} \geq \chi^2_{n-1,.1}.$$

Trying different values of n we find that the smallest such n is $n = 276$ for which

$$\frac{(n-1)n}{t^2_{n-1,.025}\cdot 64} = 306.0 \geq 305.5 = \chi^2_{n-1,.1}.$$

As to be expected, this is somewhat larger than the value found in a).

9.35 length $= 2z_{\alpha/2}\sigma/\sqrt{n}$, and if it is unknown, E(length) $= 2t_{\alpha/2,n-1}c\sigma/\sqrt{n}$, where

$$c = \frac{\sqrt{n-1}\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(n/2)}$$

and E$cS = \sigma$ (Exercise 7.50). Thus the difference in lengths is $(2\sigma/\sqrt{n})(z_{\alpha/2} - ct_{\alpha/2})$. A little work will show that, as $n \to \infty$, $c \to$ constant. (This can be done using Stirling's formula along with Lemma 2.3.14. In fact, some careful algebra will show that $c \to 1$ as $n \to \infty$.) Also, we know that, as $n \to \infty$, $t_{\alpha/2,n-1} \to z_{\alpha/2}$. Thus, the difference in lengths $(2\sigma/\sqrt{n})(z_{\alpha/2} - ct_{\alpha/2}) \to 0$ as $n \to \infty$.

9.36 The sample pdf is

$$f(x_1,\ldots,x_n|\theta) = \prod_{i=1}^n e^{i\theta - x_i} I_{(i\theta,\infty)}(x_i) = e^{\Sigma(i\theta - x_i)} I_{(\theta,\infty)}[\min(x_i/i)].$$

Thus $T = \min(X_i/i)$ is sufficient by the Factorization Theorem, and

$$P(T > t) = \prod_{i=1}^n P(X_i > it) = \prod_{i=1}^n \int_{it}^{\infty} e^{i\theta - x}\,dx = \prod_{i=1}^n e^{i(\theta - t)} = e^{-\frac{n(n+1)}{2}(t-\theta)},$$

and

$$f_T(t) = \frac{n(n+1)}{2} e^{-\frac{n(n+1)}{2}(t-\theta)}, \quad t \geq \theta.$$

Clearly, θ is a location parameter and $Y = T - \theta$ is a pivot. To find the shortest confidence interval of the form $[T + a, T + b]$, we must minimize $b - a$ subject to the constraint $P(-b \leq Y \leq -a) = 1 - \alpha$. Now the pdf of Y is strictly decreasing, so the interval length is shortest if $-b = 0$ and a satisfies

$$P(0 \leq Y \leq -a) = e^{-\frac{n(n+1)}{2}a} = 1 - \alpha.$$

So $a = 2\log(1-\alpha)/(n(n+1))$.

9.37 a. The density of $Y = X_{(n)}$ is $f_Y(y) = ny^{n-1}/\theta^n$, $0 < y < \theta$. So θ is a scale parameter, and $T = Y/\theta$ is a pivotal quantity. The pdf of T is $f_T(t) = nt^{n-1}$, $0 \leq t \leq 1$.

b. A pivotal interval is formed from the set

$$\{\theta\colon a \leq t \leq b\} = \left\{\theta\colon a \leq \frac{y}{\theta} \leq b\right\} = \left\{\theta\colon \frac{y}{b} \leq \theta \leq \frac{y}{a}\right\},$$

and has length $Y(1/a - 1/b) = Y(b-a)/ab$. Since $f_T(t)$ is increasing, $b - a$ is minimized and ab is maximized if $b = 1$. Thus shortest interval will have $b = 1$ and a satisfying $\alpha = \int_0^a nt^{n-1}dt = a^n \;\Rightarrow\; a = \alpha^{1/n}$. So the shortest $1-\alpha$ confidence interval is $\{\theta\colon y \leq \theta \leq y/\alpha^{1/n}\}$.

9.39 Let a be such that $\int_{-\infty}^{a} f(x)\,dx = \alpha/2$. This value is unique for a unimodal pdf if $\alpha > 0$. Let μ be the point of symmetry and let $b = 2\mu - a$. Then $f(b) = f(a)$ and $\int_b^\infty f(x)\,dx = \alpha/2$. $a \le \mu$ since $\int_{-\infty}^{a} f(x)\,dx = \alpha/2 \le 1/2 = \int_{-\infty}^{\mu} f(x)\,dx$. Similarly, $b \ge \mu$. And, $f(b) = f(a) > 0$ since $f(a) \ge f(x)$ for all $x \le a$ and $\int_{-\infty}^{a} f(x)\,dx = \alpha/2 > 0 \Rightarrow f(x) > 0$ for some $x < a \Rightarrow f(a) > 0$. So the conditions of Theorem 9.3.2 are satisfied.

9.41 a. We show that for any interval $[a, b]$ and $\epsilon > 0$, the probability content of $[a - \epsilon, b - \epsilon]$ is greater (as long as $b - \epsilon > a$). Write

$$\begin{aligned}\int_b^a f(x)\,dx - \int_{a-\epsilon}^{b-\epsilon} f(x)\,dx &= \int_{b-\epsilon}^{b} f(x)\,dx - \int_{a-\epsilon}^{a} f(x)\,dx \\ &\le f(b-\epsilon)[b-(b-\epsilon)] - f(a)[a-(a-\epsilon)] \\ &\le \epsilon[f(b-\epsilon) - f(a)] \ \le\ 0,\end{aligned}$$

where all of the inequalities follow because $f(x)$ is decreasing. So moving the interval toward zero increases the probability, and it is therefore maximized by moving a all the way to zero.

b. $T = Y - \mu$ is a pivot with decreasing pdf $f_T(t) = ne^{-nt} I_{[0,\infty]}(t)$. The shortest $1 - \alpha$ interval on T is $[0, -\frac{1}{n}\log\alpha]$, since

$$\int_0^b ne^{-nt}\,dt = 1 - \alpha \ \Rightarrow\ b = -\frac{1}{n}\log\alpha.$$

Since $a \le T \le b$ implies $Y - b \le \mu \le Y - a$, the best $1 - \alpha$ interval on μ is $Y + \frac{1}{n}\log\alpha \le \mu \le Y$.

9.43 a. Using Theorem 8.3.12, identify $g(t)$ with $f(x|\theta_1)$ and $f(t)$ with $f(x|\theta_0)$. Define $\phi(t) = 1$ if $t \in C$ and 0 otherwise, and let ϕ' be the indicator of any other set C' satisfying $\int_{C'} f(t)\,dt \ge 1 - \alpha$. Then $(\phi(t) - \phi'(t))(g(t) - \lambda f(t)) \le 0$ and

$$0 \ge \int (\phi - \phi')(g - \lambda f) = \int_C g - \int_{C'} g - \lambda\left[\int_C f - \int_{C'} f\right] \ge \int_C g - \int_{C'} g,$$

showing that C is the best set.

b. For Exercise 9.37, the pivot $T = Y/\theta$ has density nt^{n-1}, and the pivotal interval $a \le T \le b$ results in the θ interval $Y/b \le \theta \le Y/a$. The length is proportional to $1/a - 1/b$, and thus $g(t) = 1/t^2$. The best set is $\{t\colon 1/t^2 \le \lambda n t^{n-1}\}$, which is a set of the form $\{t\colon a \le t \le 1\}$. This has probability content $1 - \alpha$ if $a = \alpha^{1/n}$. For Exercise 9.24 (or Example 9.3.4), the g function is the same and the density of the pivot is f_k, the density of a gamma$(k, 1)$. The set $\{t\colon 1/t^2 \le \lambda f_k(t)\} = \{t\colon f_{k+2}(t) \ge \lambda'\}$, so the best a and b satisfy $\int_a^b f_k(t)\,dt = 1 - \alpha$ and $f_{k+2}(a) = f_{k+2}(b)$.

9.45 a. Since $Y = \sum X_i \sim$ gamma(n, λ) has MLR, the Karlin-Rubin Theorem (Theorem 8.3.2) shows that the UMP test is to reject H_0 if $Y < k(\lambda_0)$, where $P\left(Y < k(\lambda_0)|\lambda = \lambda_0\right) = \alpha$.

b. $T = 2Y/\lambda \sim \chi^2_{2n}$ so choose $k(\lambda_0) = \frac{1}{2}\lambda_0\chi^2_{2n,\alpha}$ and

$$\{\lambda\colon Y \ge k(\lambda)\} = \left\{\lambda\colon Y \ge \frac{1}{2}\lambda\chi^2_{2n,\alpha}\right\} = \left\{\lambda\colon 0 < \lambda \le 2Y/\chi^2_{2n,\alpha}\right\}$$

is the UMA confidence set.

c. The expected length is $\mathrm{E}\frac{2Y}{\chi^2_{2n,\alpha}} = \frac{2n\lambda}{\chi^2_{2n,\alpha}}$.

d. $X_{(1)} \sim$ exponential(λ/n), so $\mathrm{E}X_{(1)} = \lambda/n$. Thus

$$\begin{aligned}\mathrm{E}(\text{length}(C^*)) &= \frac{2 \times 120}{251.046}\lambda \ =\ .956\lambda \\ \mathrm{E}(\text{length}(C^m)) &= \frac{-\lambda}{120 \times \log(.99)} \ =\ .829\lambda.\end{aligned}$$

9.46 The proof is similar to that of Theorem 9.3.5:

$$P_\theta\left(\theta' \in C^*(X)\right) = P_\theta\left(X \in A^*(\theta')\right) \le P_\theta\left(X \in A(\theta')\right) = P_\theta\left(\theta' \in C(X)\right),$$

where A and C are any competitors. The inequality follows directly from Definition 8.3.11.

9.47 Referring to (9.3.2), we want to show that for the upper confidence bound, $P_\theta(\theta' \in C) \le 1-\alpha$ if $\theta' \ge \theta$. We have

$$P_\theta(\theta' \in C) = P_\theta(\theta' \le \bar{X} + z_\alpha\sigma/\sqrt{n}).$$

Subtract θ from both sides and rearrange to get

$$P_\theta(\theta' \in C) = P_\theta\left(\frac{\theta'-\theta}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\theta}{\sigma/\sqrt{n}} + z_\alpha\right) = P\left(Z \ge \frac{\theta'-\theta}{\sigma/\sqrt{n}} - z_\alpha\right),$$

which is less than $1-\alpha$ as long as $\theta' \ge \theta$. The solution for the lower confidence interval is similar.

9.48 a. Start with the hypothesis test $H_0\colon \theta \ge \theta_0$ versus $H_1\colon \theta < \theta_0$. Arguing as in Example 8.2.4 and Exercise 8.47, we find that the LRT rejects H_0 if $(\bar{X}-\theta_0)/(S/\sqrt{n}) < -t_{n-1,\alpha}$. So the acceptance region is $\{x\colon (\bar{x}-\theta_0)/(s/\sqrt{n}) \ge -t_{n-1,\alpha}\}$ and the corresponding confidence set is $\{\theta\colon \bar{x} + t_{n-1,\alpha}s/\sqrt{n} \ge \theta\}$.

b. The test in part a) is the UMP unbiased test so the interval is the UMA unbiased interval.

9.49 a. Clearly, for each σ, the conditional probability $P_{\theta_0}(\bar{X} > \theta_0 + z_\alpha\sigma/\sqrt{n} \mid \sigma) = \alpha$, hence the test has unconditional size α. The confidence set is $\{(\theta,\sigma)\colon \theta \ge \bar{x} - z_\alpha\sigma/\sqrt{n}\}$, which has confidence coefficient $1-\alpha$ conditionally and, hence, unconditionally.

b. From the Karlin-Rubin Theorem, the UMP test is to reject H_0 if $X > c$. To make this size α,

$$\begin{aligned}
P_{\theta_0}(X > c) &= P_{\theta_0}\left(X > c \middle|\, \sigma = 10\right) P(\sigma = 10) + P\left(X > c \middle|\, \sigma = 1\right) P(\sigma = 1) \\
&= pP\left(\frac{X-\theta_0}{10} > \frac{c-\theta_0}{10}\right) + (1-p)P(X-\theta_0 > c-\theta_0) \\
&= pP\left(Z > \frac{c-\theta_0}{10}\right) + (1-p)P(Z > c-\theta_0),
\end{aligned}$$

where $Z \sim \mathrm{n}(0,1)$. Without loss of generality take $\theta_0 = 0$. For $c = z_{(\alpha-p)/(1-p)}$ we have for the proposed test

$$\begin{aligned}
P_{\theta_0}(\text{reject}) &= p + (1-p)P\left(Z > z_{(\alpha-p)/(1-p)}\right) \\
&= p + (1-p)\frac{(\alpha-p)}{(1-p)} \;=\; p + \alpha - p \;=\; \alpha.
\end{aligned}$$

This is not UMP, but more powerful than part a. To get UMP, solve for c in $pP(Z > c/10) + (1-p)P(Z > c) = \alpha$, and the UMP test is to reject if $X > c$. For $p = 1/2$, $\alpha = .05$, we get $c = 12.81$. If $\alpha = .1$ and $p = .05$, $c = 1.392$ and $z_{\frac{.1-.05}{.95}} =_{.0526}= 1.62$.

9.51

$$\begin{aligned}
P_\theta\left(\theta \in C(X_1,\ldots,X_n)\right) &= P_\theta\left(\bar{X} - k_1 \le \theta \le \bar{X} + k_2\right) \\
&= P_\theta\left(-k_2 \le \bar{X} - \theta \le k_1\right) \\
&= P_\theta\left(-k_2 \le \sum Z_i/n \le k_1\right),
\end{aligned}$$

where $Z_i = X_i - \theta$, $i = 1,\ldots,n$. Since this is a location family, for any θ, $Z_1,\ldots,Z_n$ are iid with pdf $f(z)$, i. e., the Z_is are pivots. So the last probability does not depend on θ.

9.52 a. The LRT of $H_0\colon \sigma=\sigma_0$ versus $H_1\colon \sigma\neq\sigma_0$ is based on the statistic

$$\lambda(x)=\frac{\sup_{\mu,\sigma=\sigma_0} L\left(\mu,\,\sigma_0|\,x\right)}{\sup_{\mu,\sigma\in(0,\infty)} L(\mu,\,\sigma^2|\,x)}.$$

In the denominator, $\hat{\sigma}^2=\sum(x_i-\bar{x})^2/n$ and $\hat{\mu}=\bar{x}$ are the MLEs, while in the numerator, σ_0^2 and $\hat{\mu}$ are the MLEs. Thus

$$\lambda(x)=\frac{\left(2\pi\sigma_0^2\right)^{-n/2}e^{-\frac{\Sigma(x_i-\bar{x})^2}{2\sigma_0^2}}}{\left(2\pi\hat{\sigma}^2\right)^{-n/2}e^{-\frac{\Sigma(x_i-\bar{x})^2}{2\sigma^2}}}=\left(\frac{\sigma_0^2}{\hat{\sigma}^2}\right)^{-n/2}\frac{e^{-\frac{\Sigma(x_i-\bar{x})^2}{2\sigma_0^2}}}{e^{-n/2}},$$

and, writing $\hat{\sigma}^2=[(n-1)/n]s^2$, the LRT rejects H_0 if

$$\left(\frac{\sigma_0^2}{\frac{n-1}{n}s^2}\right)^{-n/2}e^{-\frac{(n-1)s^2}{2\sigma_0^2}}<k_\alpha,$$

where k_α is chosen to give a size α test. If we denote $t=\frac{(n-1)s^2}{\sigma_0^2}$, then $T\sim\chi^2_{n-1}$ under H_0, and the test can be written: reject H_0 if $t^{n/2}e^{-t/2}<k'_\alpha$. Thus, a $1-\alpha$ confidence set is

$$\left\{\sigma^2\colon t^{n/2}e^{-t/2}\geq k'_\alpha\right\}=\left\{\sigma^2\colon \left(\frac{(n-1)s^2}{\sigma^2}\right)^{n/2}e^{-\frac{(n-1)s^2}{\sigma^2}/2}\geq k'_\alpha\right\}.$$

Note that the function $t^{n/2}e^{-t/2}$ is unimodal (it is the kernel of a gamma density) so it follows that the confidence set is of the form

$$\begin{aligned}\left\{\sigma^2\colon t^{n/2}e^{-t/2}\geq k'_\alpha\right\} &= \left\{\sigma^2\colon a\leq t\leq b\right\} = \left\{\sigma^2\colon a\leq\frac{(n-1)s^2}{\sigma^2}\leq b\right\}\\ &= \left\{\sigma^2\colon \frac{(n-1)s^2}{b}\leq\sigma^2\leq\frac{(n-1)s^2}{b}\right\},\end{aligned}$$

where a and b satisfy $a^{n/2}e^{-a/2}=b^{n/2}e^{-b/2}$ (since they are points on the curve $t^{n/2}e^{-t/2}$). Since $\frac{n}{2}=\frac{n+2}{2}-1$, a and b also satisfy

$$\frac{1}{\Gamma\left(\frac{n+2}{2}\right)2^{(n+2)/2}}a^{((n+2)/2)-1}e^{-a/2}=\frac{1}{\Gamma\left(\frac{n+2}{2}\right)2^{(n+2)/2}}b^{((n+2)/2)-1}e^{-b/2},$$

or, $f_{n+2}(a)=f_{n+2}(b)$.

b. The constants a and b must satisfy $f_{n-1}(b)b^2=f_{n-1}(a)a^2$. But since $b^{((n-1)/2)-1}b^2=b^{((n+3)/2)-1}$, after adjusting constants, this is equivalent to $f_{n+3}(b)=f_{n+3}(a)$. Thus, the values of a and b that give the minimum length interval must satisfy this along with the probability constraint. The confidence interval, say $I(s^2)$ will be unbiased if (Definition 9.3.7)

c.

$$P_{\sigma^2}\left(\sigma'^2\in I(S^2)\right)\leq P_{\sigma^2}\left(\sigma^2\in I(S^2)\right)=1-\alpha.$$

Some algebra will establish

$$\begin{aligned}P_{\sigma^2}\left(\sigma'^2\in I(S^2)\right) &= P_{\sigma^2}\left(\frac{(n-1)S^2}{b\sigma^2}\leq\frac{\sigma'^2}{\sigma^2}\leq\frac{(n-1)S^2}{a\sigma^2}\right)\\ &= P_{\sigma^2}\left(\frac{\chi^2_{n-1}}{b}\leq\frac{\sigma'^2}{\sigma^2}\leq\frac{\chi^2_{n-1}}{a}\right) = \int_{ac}^{bc}f_{n-1}(t)\,dt,\end{aligned}$$

where $c = \sigma'^2/\sigma^2$. The derivative (with respect to c) of this last expression is $bf_{n-1}(bc) - af_{n-1}(ac)$, and hence is equal to zero if both $c = 1$ (so the interval is unbiased) and $bf_{n-1}(b) = af_{n-1}(a)$. From the form of the chi squared pdf, this latter condition is equivalent to $f_{n+1}(b) = f_{n+1}(a)$.

d. By construction, the interval will be $1-\alpha$ equal-tailed.

9.53 a. $\mathrm{E}\left[b\mathrm{length}(C) - I_C(\mu)\right] = 2c\sigma b - P(|Z| \le c)$, where $Z \sim \mathrm{n}(0,1)$.

b. $\frac{d}{dc}\left[2c\sigma b - P(|Z| \le c)\right] = 2\sigma b - 2\left(\frac{1}{\sqrt{2\pi}} e^{-c^2/2}\right)$.

c. If $b\sigma > 1/\sqrt{2\pi}$ the derivative is always positive since $e^{-c^2/2} < 1$.

9.55

$$\begin{aligned}
\mathrm{E}[L((\mu,\sigma),C)] &= \mathrm{E}\left[L((\mu,\sigma),C)|S<K\right]P(S<K) + \mathrm{E}\left[L((\mu,\sigma),C)|S>K\right]P(S>K)\\
&= \mathrm{E}\left[L((\mu,\sigma),C')|S<K\right]P(S<K) + \mathrm{E}\left[L((\mu,\sigma),C)|S>K\right]P(S>K)\\
&= R\left[L((\mu,\sigma),C')\right] + \mathrm{E}\left[L((\mu,\sigma),C)|S>K\right]P(S>K),
\end{aligned}$$

where the last equality follows because $C' = \emptyset$ if $S > K$. The conditional expectation in the second term is bounded by

$$\begin{aligned}
\mathrm{E}\left[L((\mu,\sigma),C)|S>K\right] &= \mathrm{E}\left[b\mathrm{length}(C) - I_C(\mu)|S>K\right]\\
&= \mathrm{E}\left[2bcS - I_C(\mu)|S>K\right]\\
&> \mathrm{E}\left[2bcK - 1|S>K\right] \qquad \text{(since } S>K \text{ and } I_C \le 1)\\
&= 2bcK - 1,
\end{aligned}$$

which is positive if $K > 1/2bc$. For those values of K, C' dominates C.

9.57 a. The distribution of $X_{n+1} - \bar{X}$ is $\mathrm{n}[0, \sigma^2(1+1/n)]$, so

$$P\left(X_{n+1} \in \bar{X} \pm z_{\alpha/2}\sigma\sqrt{1+1/n}\right) = P(|Z| \le z_{\alpha/2}) = 1-\alpha.$$

b. p percent of the normal population is in the interval $\mu \pm z_{p/2}\sigma$, so $\bar{x} \pm k\sigma$ is a $1-\alpha$ tolerance interval if

$$P(\mu \pm z_{p/2} \subseteq \sigma\bar{X} \pm k\sigma) = P(\bar{X} - k\sigma \le \mu - z_{p/2}\sigma \text{ and } \bar{X} + k\sigma \ge \mu + z_{p/2}\sigma) \ge 1-\alpha.$$

This can be attained by requiring

$$P(\bar{X} - k\sigma \ge \mu - z_{p/2}\sigma) = \alpha/2 \quad \text{and} \quad P(\bar{X} + k\sigma \le \mu + z_{p/2}\sigma) = \alpha/2,$$

which is attained for $k = z_{p/2} + z_{\alpha/2}/\sqrt{n}$.

c. From part (a), $(X_{n+1} - \bar{X})/(S\sqrt{1+1/n}) \sim t_{n-1}$, so a $1-\alpha$ prediction interval is $\bar{X} \pm t_{n-1,\alpha/2}S\sqrt{1+1/n}$.

Chapter 10

Asymptotic Evaluations

10.1 First calculate some moments for this distribution.

$$\mathrm{E}X = \theta/3, \quad \mathrm{E}\,X^2 = 1/3, \quad \mathrm{Var}X = \frac{1}{3} - \frac{\theta^2}{9}.$$

So $3\bar{X}_n$ is an unbiased estimator of θ with variance

$$\mathrm{Var}(3\bar{X}_n) = 9(\mathrm{Var}X)/n = (3-\theta^2)/n \to 0 \text{ as } n \to \infty.$$

So by Theorem 10.1.3, $3\bar{X}_n$ is a consistent estimator of θ.

10.3 a. The log likelihood is

$$-\frac{n}{2}\log(2\pi\theta) - \frac{1}{2}\sum(x_i-\theta)/\theta.$$

Differentiate and set equal to zero, and a little algebra will show that the MLE is the root of $\theta^2+\theta-W=0$. The roots of this equation are $(-1\pm\sqrt{1+4W})/2$, and the MLE is the root with the plus sign, as it has to be nonnegative.

b. The second derivative of the log likelihood is $(-2\sum x_i^2+n\theta)/(2\theta^3)$, yielding an expected Fisher information of

$$I(\theta) = -\mathrm{E}_\theta\frac{-2\sum X_i^2+n\theta}{2\theta^3} = \frac{2n\theta+n}{2\theta^2},$$

and by Theorem 10.1.12 the variance of the MLE is $1/I(\theta)$.

10.4 a. Write

$$\frac{\sum X_iY_i}{\sum X_i^2} = \frac{\sum X_i(X_i+\epsilon_i)}{\sum X_i^2} = 1 + \frac{\sum X_i\epsilon_i}{\sum X_i^2}.$$

From normality and independence

$$\mathrm{E}X_i\epsilon_i = 0, \quad \mathrm{Var}X_i\epsilon_i = \sigma^2(\mu^2+\tau^2), \quad \mathrm{E}X_i^2 = \mu^2+\tau^2, \quad \mathrm{Var}X_i^2 = 2\tau^2(2\mu^2+\tau^2),$$

and $\mathrm{Cov}(X_i, X_i\epsilon_i) = 0$. Applying the formulas of Example 5.5.27, the asymptotic mean and variance are

$$\mathrm{E}\left(\frac{\sum X_iY_i}{\sum X_i^2}\right) \approx 1 \text{ and } \mathrm{Var}\left(\frac{\sum X_iY_i}{\sum X_i^2}\right) \approx \frac{n\sigma^2(\mu^2+\tau^2)}{[n(\mu^2+\tau^2)]^2} = \frac{\sigma^2}{n(\mu^2+\tau^2)}$$

b.

$$\frac{\sum Y_i}{\sum X_i} = \beta + \frac{\sum \epsilon_i}{\sum X_i}$$

with approximate mean β and variance $\sigma^2/(n\mu^2)$.

c.
$$\frac{1}{n}\sum\frac{Y_i}{X_i}=\beta+\frac{1}{n}\sum\frac{\epsilon_i}{X_i}$$
with approximate mean β and variance $\sigma^2/(n\mu^2)$.

10.5 a. The integral of $\mathrm{E}T_n^2$ is unbounded near zero. We have
$$\mathrm{E}T_n^2>\sqrt{\frac{n}{2\pi\sigma^2}}\int_0^1\frac{1}{x^2}e^{-(x-\mu)^2/2\sigma^2}dx>\sqrt{\frac{n}{2\pi\sigma^2}}K\int_0^1\frac{1}{x^2}dx=\infty,$$
where $K=\max_{0\le x\le 1}e^{-(x-\mu)^2/2\sigma^2}$

b. If we delete the interval $(-\delta,\delta)$, then the integrand is bounded, that is, over the range of integration $1/x^2<1/\delta^2$.

c. Assume $\mu>0$. A similar argument works for $\mu<0$. Then
$$P(-\delta<X<\delta)=P[\sqrt{n}(-\delta-\mu)<\sqrt{n}(X-\mu)<\sqrt{n}(\delta-\mu)]<P[Z<\sqrt{n}(\delta-\mu)],$$
where $Z\sim \mathrm{n}(0,1)$. For $\delta<\mu$, the probability goes to 0 as $n\to\infty$.

10.7 We need to assume that $\tau(\theta)$ is differentiable at $\theta=\theta_0$, the true value of the parameter. Then we apply Theorem 5.5.24 to Theorem 10.1.12.

10.9 We will do a more general problem that includes *a*) and *b*) as special cases. Suppose we want to estimate $\lambda^te^{-\lambda}/t!=P(X=t)$. Let
$$T=T(X_1,\dots,X_n)=\begin{cases}1 & \text{if } X_1=t\\ 0 & \text{if } X_1\ne t.\end{cases}$$
Then $\mathrm{E}T=P(T=1)=P(X_1=t)$, so T is an unbiased estimator. Since $\sum X_i$ is a complete sufficient statistic for λ, $\mathrm{E}(T|\sum X_i)$ is UMVUE. The UMVUE is 0 for $y=\sum X_i<t$, and for $y\ge t$,
$$\begin{aligned}\mathrm{E}(T|y) &= P(X_1=t|\textstyle\sum X_i=y)\\ &= \frac{P(X_1=t,\sum X_i=y)}{P(\sum X_i=y)}\\ &= \frac{P(X_1=t)P(\sum_{i=2}^n X_i=y-t)}{P(\sum X_i=y)}\\ &= \frac{\{\lambda^te^{-\lambda}/t!\}\{[(n-1)\lambda]^{y-t}e^{-(n-1)\lambda}/(y-t)!\}}{(n\lambda)^ye^{-n\lambda}/y!}\\ &= \binom{y}{t}\frac{(n-1)^{y-t}}{n^y}.\end{aligned}$$

a. The best unbiased estimator of $e^{-\lambda}$ is $((n-1)/n)^y$.

b. The best unbiased estimator of $\lambda e^{-\lambda}$ is $(y/n)[(n-1)/n]^{y-1}$

c. Use the fact that for constants a and b,
$$\frac{d}{d\lambda}\lambda^ab^\lambda=b^\lambda\lambda^{a-1}(a+\lambda\log b),$$
to calculate the asymptotic variances of the UMVUEs. We have for $t=0$,
$$\mathrm{ARE}\left(\left(\frac{n-1}{n}\right)^{n\hat\lambda},e^{-\lambda}\right)=\left[\frac{e^{-\lambda}}{\left(\frac{n-1}{n}\right)^{n\lambda}\log\left(\frac{n-1}{n}\right)^n}\right]^2,$$

and for $t = 1$

$$\mathrm{ARE}\left(\frac{n}{n-1}\hat{\lambda}\left(\frac{n-1}{n}\right)^{n\hat{\lambda}}, \hat{\lambda}e^{-\lambda}\right) = \left[\frac{(\lambda-1)e^{-\lambda}}{\frac{n}{n-1}\left(\frac{n-1}{n}\right)^{n\lambda}\left[1+\log\left(\frac{n-1}{n}\right)^{n}\right]}\right]^2.$$

Since $[(n-1)/n]^n \to e^{-1}$ as $n \to \infty$, both of these AREs are equal to 1 in the limit.

d. For these data, $n = 15$, $\sum X_i = y = 104$ and the MLE of λ is $\hat{\lambda} = \bar{X} = 6.9333$. The estimates are

	MLE	UMVUE
$P(X=0)$	.000975	.000765
$P(X=1)$	.006758	.005684

10.11 a. It is easiest to use the Mathematica code in Example A.0.7. The second derivative of the log likelihood is

$$\frac{\partial^2}{\partial\mu^2}\log\left(\frac{1}{\Gamma[\mu/\beta]\beta^{\mu/\beta}}x^{-1+\mu/\beta}e^{-x/\beta}\right) = \frac{1}{\beta^2}\psi'(\mu/\beta),$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

b. Estimation of β does not affect the calculation.

c. For $\mu = \alpha\beta$ known, the MOM estimate of β is $\bar{x}/\alpha$. The MLE comes from differentiating the log likelihood

$$\frac{d}{d\beta}\left(-\alpha n\log\beta - \sum_i x_i/\beta\right) \stackrel{\text{set}}{=} 0 \Rightarrow \beta = \bar{x}/\alpha.$$

d. The MOM estimate of β comes from solving

$$\frac{1}{n}\sum_i x_i = \mu \text{ and } \frac{1}{n}\sum_i x_i^2 = \mu^2 + \mu\beta,$$

which yields $\tilde{\beta} = \hat{\sigma}^2/\bar{x}$. The approximate variance is quite a pain to calculate. Start from

$$\mathrm{E}\bar{X} = \mu, \quad \mathrm{Var}\bar{X} = \frac{1}{n}\mu\beta, \quad \mathrm{E}\hat{\sigma}^2 \approx \mu\beta, \quad \mathrm{Var}\hat{\sigma}^2 \approx \frac{2}{n}\mu\beta^3,$$

where we used Exercise 5.8(b) for the variance of $\hat{\sigma}^2$. Now using Example 5.5.27 and (and assuming the covariance is zero), we have $\mathrm{Var}\tilde{\beta} \approx \frac{3\beta^3}{n\mu}$. The ARE is then

$$\mathrm{ARE}(\hat{\beta}, \tilde{\beta}) = \left[3\beta^3/\mu\right]\left[\mathrm{E}\left(-\frac{d^2}{d\beta^2}l(\mu,\beta|\mathbf{X}\right)\right].$$

Here is a small table of AREs. There are some entries that are less than one - this is due to using an approximation for the MOM variance.

	μ			
β	1	3	6	10
1	1.878	0.547	0.262	0.154
2	4.238	1.179	0.547	0.317
3	6.816	1.878	0.853	0.488
4	9.509	2.629	1.179	0.667
5	12.27	3.419	1.521	0.853
6	15.075	4.238	1.878	1.046
7	17.913	5.08	2.248	1.246
8	20.774	5.941	2.629	1.451
9	23.653	6.816	3.02	1.662
10	26.546	7.704	3.419	1.878

10.13 Here are the 35 distinct samples from $\{2, 4, 9, 12\}$ and their weights.

$\{12, 12, 12, 12\}, 1/256$	$\{9, 12, 12, 12\}, 1/64$	$\{9, 9, 12, 12\}, 3/128$
$\{9, 9, 9, 12\}, 1/64$	$\{9, 9, 9, 9\}, 1/256$	$\{4, 12, 12, 12\}, 1/64$
$\{4, 9, 12, 12\}, 3/64$	$\{4, 9, 9, 12\}, 3/64$	$\{4, 9, 9, 9\}, 1/64$
$\{4, 4, 12, 12\}, 3/128$	$\{4, 4, 9, 12\}, 3/64$	$\{4, 4, 9, 9\}, 3/128$
$\{4, 4, 4, 12\}, 1/64$	$\{4, 4, 4, 9\}, 1/64$	$\{4, 4, 4, 4\}, 1/256$
$\{2, 12, 12, 12\}, 1/64$	$\{2, 9, 12, 12\}, 3/64$	$\{2, 9, 9, 12\}, 3/64$
$\{2, 9, 9, 9\}, 1/64$	$\{2, 4, 12, 12\}, 3/64$	$\{2, 4, 9, 12\}, 3/32$
$\{2, 4, 9, 9\}, 3/64$	$\{2, 4, 4, 12\}, 3/64$	$\{2, 4, 4, 9\}, 3/64$
$\{2, 4, 4, 4\}, 1/64$	$\{2, 2, 12, 12\}, 3/128$	$\{2, 2, 9, 12\}, 3/64$
$\{2, 2, 9, 9\}, 3/128$	$\{2, 2, 4, 12\}, 3/64$	$\{2, 2, 4, 9\}, 3/64$
$\{2, 2, 4, 4\}, 3/128$	$\{2, 2, 2, 12\}, 1/64$	$\{2, 2, 2, 9\}, 1/64$
$\{2, 2, 2, 4\}, 1/64$	$\{2, 2, 2, 2\}, 1/256$	

The verifications of parts $(a) - (d)$ can be done with this table, or the table of means in Example $A.0.1$ can be used. For part (e),verifying the bootstrap identities can involve much painful algebra, but it can be made easier if we understand what the bootstrap sample space (the space of all n^n bootstrap samples) looks like. Given a sample $x_1, x_2, \ldots, x_n$, the bootstrap sample space can be thought of as a data array with n^n rows (one for each bootstrap sample) and n columns, so each row of the data array is one bootstrap sample. For example, if the sample size is $n = 3$, the bootstrap sample space is

$$
\begin{array}{ccc}
x_1 & x_1 & x_1 \\
x_1 & x_1 & x_2 \\
x_1 & x_1 & x_3 \\
x_1 & x_2 & x_1 \\
x_1 & x_2 & x_2 \\
x_1 & x_2 & x_3 \\
x_1 & x_3 & x_1 \\
x_1 & x_3 & x_2 \\
x_1 & x_3 & x_3 \\
\hline
x_2 & x_1 & x_1 \\
x_2 & x_1 & x_2 \\
x_2 & x_1 & x_3 \\
x_2 & x_2 & x_1 \\
x_2 & x_2 & x_2 \\
x_2 & x_2 & x_3 \\
x_2 & x_3 & x_1 \\
x_2 & x_3 & x_2 \\
x_2 & x_3 & x_3 \\
\hline
x_3 & x_1 & x_1 \\
x_3 & x_1 & x_2 \\
x_3 & x_1 & x_3 \\
x_3 & x_2 & x_1 \\
x_3 & x_2 & x_2 \\
x_3 & x_2 & x_3 \\
x_3 & x_3 & x_1 \\
x_3 & x_3 & x_2 \\
x_3 & x_3 & x_3
\end{array}
$$

Note the pattern. The first column is 9 x_1s followed by 9 x_2s followed by 9 x_3s, the second column is 3 x_1s followed by 3 x_2s followed by 3 x_3s, then repeated, etc. In general, for the entire bootstrap sample,

- The first column is n^{n-1} x_1s followed by n^{n-1} x_2s followed by, …, followed by n^{n-1} x_ns
- The second column is n^{n-2} x_1s followed by n^{n-2} x_2s followed by, …, followed by n^{n-2} x_ns, repeated n times
- The third column is n^{n-3} x_1s followed by n^{n-3} x_2s followed by, …, followed by n^{n-3} x_ns, repeated n^2 times

  $\vdots$
- The n^{th} column is 1 x_1 followed by 1 x_2 followed by, …, followed by 1 x_n, repeated n^{n-1} times

So now it is easy to see that each column in the data array has mean $\bar{x}$, hence the entire bootstrap data set has mean $\bar{x}$. Appealing to the $3^3 \times 3$ data array, we can write the numerator of the variance of the bootstrap means as

$$\begin{aligned}
&\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[\frac{1}{3}(x_i+x_j+x_k)-\bar{x}\right]^2 \\
&= \frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})+(x_j-\bar{x})+(x_k-\bar{x})\right]^2 \\
&= \frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})^2+(x_j-\bar{x})^2+(x_k-\bar{x})^2\right],
\end{aligned}$$

because all of the cross terms are zero (since they are the sum of deviations from the mean). Summing up and collecting terms shows that

$$\frac{1}{3^2}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left[(x_i-\bar{x})^2+(x_j-\bar{x})^2+(x_k-\bar{x})^2\right] = 3\sum_{i=1}^{3}(x_i-\bar{x})^2,$$

and thus the average of the variance of the bootstrap means is

$$\frac{3\sum_{i=1}^{3}(x_i-\bar{x})^2}{3^3}$$

which is the usual estimate of the variance of $\bar{X}$ if we divide by n instead of $n-1$. The general result should now be clear. The variance of the bootstrap means is

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n}\left[\frac{1}{n}(x_{i_1}+x_{i_2}+\cdots+x_{i_n})-\bar{x}\right]^2 \\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n}\left[(x_{i_1}-\bar{x})^2+(x_{i_2}-\bar{x})^2+\cdots+(x_{i_n}-\bar{x})^2\right],
\end{aligned}$$

since all of the cross terms are zero. Summing and collecting terms shows that the sum is $n^{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^2$, and the variance of the bootstrap means is $n^{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^2/n^n = \sum_{i=1}^{n}(x_i-\bar{x})^2/n^2$.

10.15 a. As $B \to \infty$ $\mathrm{Var}^*_B(\hat{\theta}) = \mathrm{Var}^*(\hat{\theta})$.

b. Each $\mathrm{Var}^*_{B_i}(\hat{\theta})$ is a sample variance, and they are independent so the LLN applies and

$$\frac{1}{m}\sum_{i=1}^{m}\mathrm{Var}^*_{B_i}(\hat{\theta}) \overset{m\to\infty}{\to} \mathrm{E}\mathrm{Var}^*_B(\hat{\theta}) = \mathrm{Var}^*(\hat{\theta}),$$

where the last equality follows from Theorem 5.2.6(c).

10.17 a. The correlation is .7781

b. Here is R code (R is available free at http://cran.r-project.org/) to bootstrap the data, calculate the standard deviation, and produce the histogram:

```
cor(law)
n <- 15
theta <- function(x,law){ cor(law[x,1],law[x,2]) }
results <- bootstrap(1:n,1000,theta,law,func=sd)
results[2]
hist(results[[1]])
```

The data "law" is in two columns of length 15, "results[2]" contains the standard deviation. The vector "results[[1]]" is the bootstrap sample. The output is

```
          V1        V2
V1 1.0000000 0.7781716
V2 0.7781716 1.0000000
$func.thetastar
[1] 0.1322881
```

showing a correlation of .7781 and a bootstrap standard deviation of .1323.

c. The R code for the parametric bootstrap is

```
mx<-600.6;my<-3.09
sdx<-sqrt(1791.83);sdy<-sqrt(.059)
rho<-.7782;b<-rho*sdx/sdy;sdxy<-sqrt(1-rho^2)*sdx
rhodata<-rho
for (j in 1:1000)  {
y<-rnorm(15,mean=my,sd=sdy)
x<-rnorm(15,mean=mx+b*(y-my),sd=sdxy)
rhodata<-c(rhodata,cor(x,y))
}
sd(rhodata)
hist(rhodata)
```

where we generate the bivariate normal by first generating the marginal then the condidional, as R does not have a bivariate normal generator. The bootstrap standard deviation is 0.1159, smaller than the nonparametric estimate. The histogram looks similar to the nonparametric bootstrap histogram, displaying a skewness left.

d. The Delta Method approximation is

$$r \sim \mathrm{n}(\rho, (1-\rho^2)^2/n),$$

and the "plug-in" estimate of standard error is $\sqrt{(1-.7782^2)^2/15} = .1018$, the smallest so far. Also, the approximate pdf of r will be normal, hence symmetric.

e. By the change of variables

$$t = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right), \qquad dt = \frac{1}{1-r^2},$$

the density of r is

$$\frac{1}{\sqrt{2\pi}(1-r^2)}\exp\left(-\frac{n}{2}\left[\frac{1}{2}\log\left(\frac{1+r}{1-r}\right)-\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)\right]^2\right), \quad -1 \le r \le 1.$$

More formally, we could start with the random variable T, normal with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and variance $1/n$, and make the transformation to $R = \frac{e^{2T}+1}{e^{2T}-1}$ and get the same answer.

10.19 a. The variance of $\bar{X}$ is

$$\begin{aligned}
\mathrm{Var}\bar{X} = \mathrm{E}(\bar{X}-\mu)^2 &= \mathrm{E}\left(\frac{1}{n}\sum_i X_i - \mu\right)^2 \\
&= \frac{1}{n^2}\mathrm{E}\left(\sum_i (X_i-\mu)^2 + 2\sum_{i>j}(X_i-\mu)(X_j-\mu)\right) \\
&= \frac{1}{n^2}\left(n\sigma^2 + 2\frac{n(n-1)}{2}\rho\sigma^2\right) \\
&= \frac{\sigma^2}{n} + \frac{n-1}{n}\rho\sigma^2
\end{aligned}$$

b. In this case we have

$$\mathrm{E}\left[\sum_{i>j}(X_i-\mu)(X_j-\mu)\right] = \sigma^2\sum_{i=2}^{n}\sum_{j=1}^{i-1}\rho^{i-j}.$$

In the double sum ρ appears $n-1$ times, ρ^2 appears $n-2$ times, etc.. so

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\rho^{i-j} = \sum_{i=1}^{n-1}(n-i)\rho^i = \frac{\rho}{1-\rho}\left(n - \frac{1-\rho^n}{1-\rho}\right),$$

where the series can be summed using (1.5.4), the partial sum of the geometric series, or using Mathematica.

c. The mean and variance of X_i are

$$\mathrm{E}X_i = \mathrm{E}[\mathrm{E}(X_i|X_{i-1})] = \mathrm{E}\rho X_{i-1} = \cdots = \rho^{i-1}\mathrm{E}X_1$$

and

$$\mathrm{Var}X_i = \mathrm{Var}\mathrm{E}(X_i|X_{i-1}) + \mathrm{E}\mathrm{Var}(X_i|X_{i-1}) = \rho^2\sigma^2 + 1 = \sigma^2$$

for $\sigma^2 = 1/(1-\rho^2)$. Also, by iterating the expectation

$$\mathrm{E}X_1X_i = \mathrm{E}[\mathrm{E}(X_1X_i|X_{i-1})] = \mathrm{E}[\mathrm{E}(X_1|X_{i-1})\mathrm{E}(X_i|X_{i-1})] = \rho\mathrm{E}[X_1X_{i-1}],$$

where we used the facts that X_1 and X_i are independent conditional on X_{i-1}. Continuing with the argument we get that $\mathrm{E}X_1X_i = \rho^{i-1}\mathrm{E}X_1^2$. Thus,

$$\mathrm{Corr}(X_1,X_i) = \frac{\rho^{i-1}\mathrm{E}X_1^2 - \rho^{i-1}(\mathrm{E}X_1)^2}{\sqrt{\mathrm{Var}X_1\mathrm{Var}X_i}} = \frac{\rho^{i-1}\sigma^2}{\sqrt{\sigma^2\sigma^2}} = \rho^{i-1}.$$

10.21 a. If any $x_i \to \infty$, $s^2 \to \infty$, so it has breakdown value 0. To see this, suppose that $x_1 \to \infty$. Write

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left([(1-\frac{1}{n})x_1 - \bar{x}_{-1}]^2 + \sum_{i=2}^{n}(x_i-\bar{x})^2\right),$$

where $\bar{x}_{-1} = (x_2 + \ldots + x_n)/n$. It is easy to see that as $x_1 \to \infty$, each term in the sum $\to \infty$.

b. If less than 50% of the sample $\to \infty$, the median remains the same, and the median of $|x_i - M|$ remains the same. If more than 50% of the sample $\to \infty$, $M \to \infty$ and so does the MAD.

10.23 a. The ARE is $[2\sigma f(\mu)]^2$. We have

Distribution	Parameters	variance	$f(\mu)$	ARE
normal	$\mu = 0, \sigma = 1$	1	.3989	.64
logistic	$\mu = 0, \beta = 1$	$\pi^2/3$	.25	.82
double exp.	$\mu = 0, \sigma = 1$	2	.5	2

b. If $X_1, X_2, \ldots, X_n$ are iid f_X with $\mathrm{E}X_1 = \mu$ and $\mathrm{Var}X_1 = \sigma^2$, the ARE is $\sigma^2[2 * f_X(\mu)]^2$. If we transform to $Y_i = (X_i - \mu)/\sigma$, the pdf of Y_i is $f_Y(y) = \sigma f_X(\sigma y + \mu)$ with ARE $[2 * f_Y(0)]^2 = \sigma^2[2 * f_X(\mu)]^2$

c. The median is more efficient for smaller ν, the distributions with heavier tails.

ν	$\mathrm{Var}X$	$f(0)$	ARE
3	3	.367	1.62
5	5/3	.379	.960
10	5/4	.389	.757
25	25/23	.395	.678
50	25/24	.397	.657
∞	1	.399	.637

d. Again the heavier tails favor the median.

δ	σ	ARE
.01	2	.649
.1	2	.747
.5	2	.895
.01	5	.777
.1	5	1.83
.5	5	2.98

10.25 By transforming $y = x - \theta$,

$$\int_{-\infty}^{\infty} \psi(x-\theta) f(x-\theta) dx = \int_{-\infty}^{\infty} \psi(y) f(y) dy.$$

Since ψ is an odd function, $\psi(y) = -\psi(-y)$, and

$$\begin{aligned}\int_{-\infty}^{\infty} \psi(y) f(y) dy &= \int_{-\infty}^{0} \psi(y) f(y) dy + \int_{0}^{\infty} \psi(y) f(y) dy \\ &= \int_{-\infty}^{0} -\psi(-y) f(y) dy + \int_{0}^{\infty} \psi(y) f(y) dy \\ &= -\int_{0}^{\infty} \psi(y) f(y) dy + \int_{0}^{\infty} \psi(y) f(y) dy = 0,\end{aligned}$$

where in the last line we made the transformation $y \to -y$ and used the fact the f is symmetric, so $f(y) = f(-y)$. From the discussion preceding Example 10.2.6, $\hat{\theta}_M$ is asymptotically normal with mean equal to the true θ.

10.27 a.

$$\lim_{\delta \to 0} \frac{1}{\delta}[(1-\delta)\mu + \delta x - \mu] = \lim_{\delta \to 0} \frac{\delta(x-\mu)}{\delta} = x - \mu.$$

b.

$$P(X \le a) = P(X \le a | X \sim F)(1-\delta) + P(x \le a | X = x)\delta = (1-\delta)F(a) + \delta I(x \le a)$$

and

$$(1-\delta)F(a) = \frac{1}{2} \quad \Rightarrow \quad a = F^{-1}\left(\frac{1}{2(1-\delta)}\right)$$
$$(1-\delta)F(a) + \delta = \frac{1}{2} \quad \Rightarrow \quad a = F^{-1}\left(\frac{\frac{1}{2}-\delta}{2(1-\delta)}\right)$$

c. The limit is
$$\lim_{\delta\to 0} \frac{a_\delta - a_0}{\delta} = a'_\delta|_{\delta=0}$$
by the definition of derivative. Since $F(a_\delta) = \frac{1}{2(1-\delta)}$,
$$\frac{d}{d\delta}F(a_\delta) = \frac{d}{d\delta}\frac{1}{2(1-\delta)}$$
or
$$f(a_\delta)a'_\delta = \frac{1}{2(1-\delta)^2} \Rightarrow a'_\delta = \frac{1}{2(1-\delta)^2 f(a_\delta)}.$$
Since $a_0 = m$, the result follows. The other limit can be calculated in a similar manner.

10.29 a. Substituting cl' for ψ makes the ARE equal to 1.

b. For each distribution is the case that the given ψ function is equal to cl', hence the resulting M-estimator is asymptotically efficient by (10.2.9).

10.31 a. By the CLT,
$$\sqrt{n_1}\frac{\hat{p}_1 - p_1}{\sqrt{p_1(1-p_1)}} \to \mathrm{n}(0,1) \quad \text{and} \quad \sqrt{n_2}\frac{\hat{p}_2 - p_2}{\sqrt{p_2(1-p_2)}} \to \mathrm{n}(0,1),$$
so if $\hat{p}_1$ and $\hat{p}_2$ are independent, under $H_0: p_1 = p_2 = p$,
$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \to \mathrm{n}(0,1)$$
where we use Slutsky's Theorem and the fact that $\hat{p} = (S_1+S_2)/(n_1+n_2)$ is the MLE of p under H_0 and converges to p in probability. Therefore, $T \to \chi^2_1$.

b. Substitute $\hat{p}_i$s for S_i and F_is to get
$$\begin{aligned} T^* &= \frac{n_1^2(\hat{p}_1-\hat{p})^2}{n_1\hat{p}} + \frac{n_2^2(\hat{p}_2-\hat{p})^2}{n_2\hat{p}} \\ &\quad + \frac{n_1^2\left[(1-\hat{p}_1)-(1-\hat{p})\right]^2}{n_1(1-\hat{p})} + \frac{n_2^2\left[(1-\hat{p}_2)-(1-\hat{p})\right]^2}{n_2\hat{p}} \\ &= \frac{n_1(\hat{p}_1-\hat{p})^2}{\hat{p}(1-\hat{p})} + \frac{n_2(\hat{p}_2-\hat{p})^2}{\hat{p}(1-\hat{p})} \end{aligned}$$
Write $\hat{p} = (n_1\hat{p}_1 + n_2\hat{p}_2)/(n_1+n_2)$. Substitute this into the numerator, and some algebra will get
$$n_1(\hat{p}_1-\hat{p})^2 + n_2(\hat{p}_2-\hat{p})^2 = \frac{(\hat{p}_1-\hat{p}_2)^2}{\frac{1}{n_1}+\frac{1}{n_2}},$$
so $T^* = T$.

c. Under H_0,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1-p)}} \to \mathrm{n}(0,1)$$

and both $\hat{p}_1$ and $\hat{p}_2$ are consistent, so $\hat{p}_1(1-\hat{p}_1) \to p(1-p)$ and $\hat{p}_2(1-\hat{p}_2) \to p(1-p)$ in probability. Therefore, by Slutsky's Theorem,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \to \mathrm{n}(0,1),$$

and $(T^{**})^2 \to \chi^2_1$. It is easy to see that $T^{**} \neq T$ in general.

d. The estimator $(1/n_1 + 1/n_2)\hat{p}(1-\hat{p})$ is the MLE of $\mathrm{Var}(\hat{p}_1 - \hat{p}_2)$ under H_0, while the estimator $\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_1$ is the MLE of $\mathrm{Var}(\hat{p}_1 - \hat{p}_2)$ under H_1. One might argue that in hypothesis testing, the first one should be used, since under H_0, it provides a better estimator of variance. If interest is in finding the confidence interval, however, we are making inference under both H_0 and H_1, and the second one is preferred.

e. We have $\hat{p}_1 = 34/40$, $\hat{p}_2 = 19/35$, $\hat{p} = (34+19)/(40+35) = 53/75$, and $T = 8.495$. Since $\chi^2_{1,.05} = 3.84$, we can reject H_0 at $\alpha = .05$.

10.32 a. First calculate the MLEs under $p_1 = p_2 = p$. We have

$$L(p|x) = p^{x_1} p^{x_2} p^{x_3} \cdots p_{n-1}^{x_{n-1}} \left(1 - 2p - \sum_{i=3}^{n-1} p_i\right)^{m - x_1 - x_2 - \cdots - x_{n-1}}.$$

Taking logs and differentiating yield the following equations for the MLEs:

$$\frac{\partial \log L}{\partial p} = \frac{x_1 + x_2}{p} - \frac{2\left(m - \sum_{i=1}^{n-1} x_i\right)}{1 - 2p - \sum_{i=3}^{n-1} p_i} = 0$$

$$\frac{\partial \log L}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_n}{1 - 2p - \sum_{i=3}^{n-1} p_i} = 0, \quad i = 3, \ldots, n-1,$$

with solutions $\hat{p} = \frac{x_1 + x_2}{2m}$, $\hat{p}_i = \frac{x_i}{m}, i = 3, \ldots, n-1$, and $\hat{p}_n = \left(m - \sum_{i=1}^{n-1} x_i\right)/m$. Except for the first and second cells, we have expected = observed, since both are equal to x_i. For the first two terms, expected $= m\hat{p} = (x_1 + x_2)/2$ and we get

$$\sum \frac{(\textit{observed} - \textit{expected})^2}{\textit{expected}} = \frac{\left(x_1 - \frac{x_1 + x_2}{2}\right)^2}{\frac{x_1 + x_2}{2}} + \frac{\left(x_2 - \frac{x_1 + x_2}{2}\right)^2}{\frac{x_1 + x_2}{2}} = \frac{(x_1 - x_2)^2}{x_1 + x_2}.$$

b. Now the hypothesis is about conditional probabilities is given by H_0: P(change—initial agree)=P(change—initial disagree) or, in terms of the parameters $H_0 : \frac{p_1}{p_1 + p_3} = \frac{p_2}{p_2 + p_4}$. This is the same as $p_1 p_4 = p_2 p_3$, which is not the same as $p_1 = p_2$.

10.33 Theorem 10.1.12 and Slutsky's Theorem imply that

$$\frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{n} I_n(\hat{\theta})}} \to \mathrm{n}(0,1)$$

and the result follows.

10.35 a. Since $\sigma/\sqrt{n}$ is the estimated standard deviation of $\bar{X}$ in this case, the statistic is a Wald statistic

b. The MLE of σ^2 is $\hat{\sigma}^2_\mu = \sum_i (x_i - \mu)^2/n$. The information number is

$$-\frac{d^2}{d(\sigma^2)^2}\left(-\frac{n}{2}\log\sigma^2 - \frac{1}{2}\frac{\hat{\sigma}^2_\mu}{\sigma^2}\right)\bigg|_{\sigma^2=\hat{\sigma}^2_\mu} = \frac{n}{2\hat{\sigma}^2_\mu}.$$

Using the Delta method, the variance of $\hat{\sigma}_\mu = \sqrt{\hat{\sigma}^2_\mu}$ is $\hat{\sigma}^2_\mu/8n$, and a Wald statistic is

$$\frac{\hat{\sigma}_\mu - \sigma_0}{\sqrt{\sigma^2_\mu/8n}}.$$

10.37 a. The log likelihood is

$$\log L = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_i (x_i - \mu)^2/\sigma^2$$

with

$$\begin{aligned}\frac{d}{d\mu} &= \frac{1}{\sigma^2}\sum_i (x_i - \mu) = \frac{n}{\sigma^2}(\bar{x} - \mu)\\ \frac{d^2}{d\mu^2} &= -\frac{n}{\sigma^2},\end{aligned}$$

so the test statistic for the score test is

$$\frac{\frac{n}{\sigma^2}(\bar{x} - \mu)}{\sqrt{\sigma^2/n}} = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}$$

b. We test the equivalent hypothesis $H_0 : \sigma^2 = \sigma^2_0$. The likelihood is the same as Exercise 10.35(b), with first derivative

$$-\frac{d}{d\sigma^2} = \frac{n(\hat{\sigma}^2_\mu - \sigma^2)}{2\sigma^4}$$

and expected information number

$$\mathrm{E}\left(\frac{n(2\hat{\sigma}^2_\mu - \sigma^2)}{2\sigma^6}\right) = \frac{n(2\sigma^2 - \sigma^2)}{2\sigma^6} = \frac{n}{2\sigma^4}.$$

The score test statistic is

$$\sqrt{\frac{n}{2}}\frac{\hat{\sigma}^2_\mu - \sigma^2_0}{\sigma^2_0}$$

10.39 We summarize the results for (a) – (c) in the following table. We assume that the underlying distribution is normal, and use that for all score calculations. The actual data is generated from normal, logistic, and double exponential. The sample size is 15, we use 1000 simulations and draw 20 bootstrap samples. Here $\theta_0 = 0$, and the power is tabulated for a nominal $\alpha = .1$ test.

Underlying pdf	Test	θ_0	$\theta_0+.25\sigma$	$\theta_0+.5\sigma$	$\theta_0+.75\sigma$	$\theta_0+1\sigma$	$\theta_0+2\sigma$
Laplace	Naive	0.101	0.366	0.774	0.957	0.993	1.
	Boot	0.097	0.364	0.749	0.932	0.986	1.
	Median	0.065	0.245	0.706	0.962	0.995	1.
Logistic	Naive	0.137	0.341	0.683	0.896	0.97	1.
	Boot	0.133	0.312	0.641	0.871	0.967	1.
	Median	0.297	0.448	0.772	0.944	0.993	1.
Normal	Naive	0.168	0.316	0.628	0.878	0.967	1.
	Boot	0.148	0.306	0.58	0.836	0.957	1.
	Median	0.096	0.191	0.479	0.761	0.935	1.

Here is Mathematica code:

This program calculates size and power for Exercise 10.39, Second Edition

We do our calculations assuming normality, but simulate power and size under other distributions. We test $H_0 : \theta = 0$.

```
theta_0=0;
Needs["Statistics‘Master‘"]
Clear[x]
f1[x_]=PDF[NormalDistribution[0,1],x];
F1[x_]=CDF[NormalDistribution[0,1],x];
f2[x_]=PDF[LogisticDistribution[0,1],x];
f3[x_]=PDF[LaplaceDistribution[0,1],x];
v1=Variance[NormalDistribution[0,1]];
v2=Variance[LogisticDistribution[0,1]];
v3=Variance[LaplaceDistribution[0,1]];
```

Calculate m-estimate

```
Clear[k,k1,k2,t,x,y,d,n,nsim,a,w1]
ind[x_,k_]:=If[Abs[x]<k,1,0]
rho[y_,k_]:=ind[y,k]*y^2 + (1-ind[y,k])*(k*Abs[y]-k^2)
alow[d_]:=Min[Mean[d],Median[d]]
aup[d_]:=Max[Mean[d],Median[d]]
sol[k_,d_]:=FindMinimum[Sum[rho[d[[i]]-a,k],{i,1,n}],{a,{alow[d],aup[d]}}]
mest[k_,d_]:=sol[k,d][[2]]
```

generate data - to change underlying distributions change the sd and the distribution in the Random statement.

```
n = 15; nsim = 1000; sd = Sqrt[v1];
theta = {theta_0, theta_0 +.25*sd, theta_0 +.5*sd,
        theta_0 +.75*sd, theta_0 + 1*sd, theta_0 +2*sd}
ntheta = Length[theta]
data = Table[Table[Random[NormalDistribution[0, 1]],
       {i, 1, n}],{j, 1,nsim}];
m1 = Table[Table[a /. mest[k1, data[[j]] - theta[[i]]],
       {j, 1, nsim}], {i, 1, n\theta}];
```

Calculation of naive variance and test statistic

```
Psi[x_, k_] =  x*If[Abs[x]<= k, 1, 0]-  k*If[x < -k, 1, 0] +
```

```
                k*If[x > k, 1, 0];
Psi1[x_, k_] = If[Abs[x]  <=  k, 1, 0];
num =Table[Psi[w1[[j]][[i]], k1], {j, 1, nsim}, {i, 1,n}];
den =Table[Psi1[w1[[j]][[i]], k1], {j, 1, nsim}, {i, 1,n}];
varnaive = Map[Mean, num^2]/Map[Mean, den]^2;
naivestat = Table[Table[m1[[i]][[j]] -theta_0/Sqrt[varnaive[[j]]/n],
            {j, 1, nsim}],{i, 1, ntheta}];
absnaive = Map[Abs, naivestat];
N[Table[Mean[Table[If[absnaive[[i]][[j]] > 1.645, 1, 0],
           {j, 1, nsim}]], {i, 1, n\theta}]]
```

Calculation of bootstrap variance and test statistic

```
nboot=20;
u:=Random[DiscreteUniformDistribution[n]]
databoot=Table[data[[jj]][[u]],{jj,1,nsim},{j,1,nboot},{i,1,n}];
m1boot=Table[Table[a/.mest[k1,databoot[[j]][[jj]]],
          {jj,1,nboot}],{j,1,nsim}];
varboot = Map[Variance, m1boot];
bootstat = Table[Table[m1[[i]][[j]] -theta_0/Sqrt[varboot[[j]]],
          {j, 1, nsim}], {i, 1, ntheta}];
absboot = Map[Abs, bootstat];
N[Table[Mean[Table[If[absboot[[i]][[j]] > 1.645, 1,0],
          {j, 1, nsim}]], {i, 1, ntheta}]]\)
```

Calculation of median test - use the score variance at the root density (normal)

```
med = Map[Median, data];
medsd = 1/(n*2*f1[theta_0]);
medstat = Table[Table[med[[j]] + \theta[[i]] - theta_0/medsd,
          {j, 1, nsim}], {i, 1, ntheta}];
absmed = Map[Abs, medstat];
N[Table[Mean[Table[If[\(absmed[[i]][[j]] > 1.645, 1, 0],
          {j, 1, nsim}]], {i, 1, ntheta}]]
```

10.41 a. The log likelihood is

$$\log L = nr\log p + n\bar{x}\log(1-p)$$

with

$$\frac{d}{dp}\log L = \frac{nr}{p} - \frac{n\bar{x}}{1-p} \quad \text{and} \quad \frac{d^2}{dp^2}\log L = -\frac{nr}{p^2} - \frac{n\bar{x}}{(1-p)^2},$$

expected information $\frac{nr}{p^2(1-p)}$ and (Wilks) score test statistic

$$\sqrt{n}\frac{\left(\frac{r}{p} - \frac{n\bar{x}}{1-p}\right)}{\sqrt{\frac{r}{p^2(1-p)}}} = \sqrt{\frac{n}{r}}\left(\frac{(1-p)r + p\bar{x}}{\sqrt{1-p}}\right).$$

Since this is approximately $\mathrm{n}(0,1)$, a $1-\alpha$ confidence set is

$$\left\{p : \left|\sqrt{\frac{n}{r}}\left(\frac{(1-p)r - p\bar{x}}{\sqrt{1-p}}\right)\right| \le z_{\alpha/2}\right\}.$$

b. The mean is $\mu = r(1-p)/p$, and a little algebra will verify that the variance, $r(1-p)/p^2$ can be written $r(1-p)/p^2 = \mu + \mu^2/r$. Thus

$$\sqrt{\frac{n}{r}}\left(\frac{(1-p)r - p\bar{x}}{\sqrt{1-p}}\right) = \sqrt{n}\frac{\mu - \bar{x}}{\sqrt{\mu + \mu^2/r}}.$$

The confidence interval is found by setting this equal to $z_{\alpha/2}$, squaring both sides, and solving the quadratic for μ. The endpoints of the interval are

$$\frac{r(8\bar{x} + z^2_{\alpha/2}) \pm \sqrt{rz^2_{\alpha/2}}\sqrt{16r\bar{x} + 16\bar{x}^2 + rz^2_{\alpha/2}}}{8r - 2z^2_{\alpha/2}}.$$

For the continuity correction, replace $\bar{x}$ with $\bar{x}+1/(2n)$ when solving for the upper endpoint, and with $\bar{x} - 1/(2n)$ when solving for the lower endpoint.

c. We table the endpoints for $\alpha = .1$ and a range of values of r. Note that $r = \infty$ is the Poisson, and smaller values of r give a wider tail to the negative binomial distribution.

r	lower bound	upper bound
1	22.1796	364.42
5	36.2315	107.99
10	38.4565	95.28
50	40.6807	85.71
100	41.0015	84.53
1000	41.3008	83.46
∞	41.3348	83.34

10.43 a. Since

$$P\left(\sum_i X_i = 0\right) = (1-p)^n = \alpha/2 \Rightarrow p = 1 - \alpha^{1/n}$$

and

$$P\left(\sum_i X_i = n\right) = p^n = \alpha/2 \Rightarrow p = \alpha^{1/n},$$

these endpoints are exact, and are the shortest possible.

b. Since $p \in [0,1]$, any value outside has zero probability, so truncating the interval shortens it at no cost.

10.45 The continuity corrected roots are

$$\frac{2\hat{p} + z^2_{\alpha/2}/n \pm \frac{1}{n} \pm \sqrt{\frac{z^2_{\alpha/2}}{n^3}[\pm 2n(1-2\hat{p}) - 1] + (2\hat{p} + z^2_{\alpha/2}/n)^2 - 4\hat{p}^2(1 + z^2_{\alpha/2}/n)}}{2(1 + z^2_{\alpha/2}/n)}$$

where we use the upper sign for the upper root and the lower sign for the lower root. Note that the only differences between the continuity-corrected intervals and the ordinary score intervals are the terms with $\pm$ in front. But it is still difficult to analytically compare lengths with the non-corrected interval - we will do a numerical comparison. For $n = 10$ and $\alpha = .1$ we have the following table of length ratios, with the continuity-corrected length in the denominator

n	0	1	2	3	4	5	6	7	8	9	10
Ratio	0.79	0.82	0.84	0.85	0.86	0.86	0.86	0.85	0.84	0.82	0.79

The coverage probabilities are

p	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
score	.99	.93	.97	.92	.90	.89	.90	.92	.97	.93	.99
cc	.99	.99	.97	.92	.98	.98	.98	.92	.97	.99	.99

Mathematica code to do the calculations is:

```
Needs["Statistics`Master`"]
Clear[p, x]
pbino[p_, x_] = PDF[BinomialDistribution[n, p], x];
cut = 1.645^2;
n = 10;
```

The quadratic score interval with and without continuity correction

```
slowcc[x_] := p /. FindRoot[(x/n - 1/(2*n) - p)^2 ==
              cut*(p*((1 - p))/n, {p,  .001}]
supcc[x_] :=  p /. FindRoot[(x/n + 1/(2*n) - p)^2 ==
              cut*(p*((1 - p)/n, {p,  .999}]
slow[x_] :=   p /. FindRoot[(x/n - p))^2 ==
              cut*(p*(1 - p))/n, {p,  .001}]
sup[x_] :=    p /. FindRoot[(x/n - p)^2 ==
              cut*(p*(1 - p)/n, {p,  .999}]
scoreintcc=Partition[Flatten[{{0,sup[0]},Table[{slowcc[i],supcc[i]},
              {i,1,n-1}],{slowcc[n],1}},2],2];
scoreint=Partition[Flatten[{{0,sup[0]},Table[{slow[i],sup[i]},
              {i,1,n-1}],{slowcc[n],1}},2],2];
```

Length Comparison

```
Table[(sup[i] - slow[i])/(supcc[i] - slowcc[i]), {i, 0, n}]
```

Now we'll calculate coverage probabilities

```
scoreindcc[p_,x_]:=If[scoreintcc[[x+1]][[1]]<=p<=scoreintcc[[x+1]][[2]],1,0]
scorecovcc[p_]:=scorecovcc[p]=Sum[pbino[p,x]*scoreindcc[p,x],{x,0,n}]
scoreind[p_,x_]:=If[scoreint[[x+1]][[1]]<=p<=scoreint[[x+1]][[2]],1,0]
scorecov[p_]:=scorecov[p]=Sum[pbino[p,x]*scoreind[p,x],{x,0,n}]
{scorecovcc[.0001],Table[scorecovcc[i/10],{i,1,9}],scorecovcc[.9999]}//N
{scorecov[.0001],Table[scorecov[i/10],{i,1,9}],scorecov[.9999]}//N
```

10.47 a. Since $2pY \sim \chi^2_{nr}$ (approximately)

$$P(\chi^2_{nr,1-\alpha/2} \leq 2pY \leq \chi^2_{nr,\alpha/2}) = 1-\alpha,$$

and rearrangment gives the interval.

b. The interval is of the form $P(a/2Y \leq p \leq b/2Y)$, so the length is proportional to $b-a$. This must be minimized subject to the constraint $\int_a^b f(y)dy = 1-\alpha$, where $f(y)$ is the pdf of a χ^2_{nr}. Treating b as a function of a, differentiating gives

$$b' - 1 = 0 \quad \text{and} \quad f(b)b' - f(a) = 0$$

which implies that we need $f(b) = f(a)$.

Chapter 11

Analysis of Variance and Regression

11.1 a. The first order Taylor's series approximation is

$$\mathrm{Var}[g(Y)] \approx [g'(\theta)]^2 \cdot \mathrm{Var}Y = [g'(\theta)]^2 \cdot v(\theta).$$

b. If we choose $g(y) = g^*(y) = \int_a^y \frac{1}{\sqrt{v(x)}}dx$, then

$$\frac{dg^*(\theta)}{d\theta} = \frac{d}{d\theta}\int_a^\theta \frac{1}{\sqrt{v(x)}}dx = \frac{1}{\sqrt{v(\theta)}},$$

by the Fundamental Theorem of Calculus. Then, for any θ,

$$\mathrm{Var}[g^*(Y)] \approx \left(\frac{1}{\sqrt{v(\theta)}}\right)^2 v(\theta) = 1.$$

11.2 a. $v(\lambda) = \lambda$, $g^*(y) = \sqrt{y}$, $\frac{dg^*(\lambda)}{d\lambda} = \frac{1}{2\sqrt{\lambda}}$, $\mathrm{Var}g^*(Y) \approx \left(\frac{dg^*(\lambda)}{d\lambda}\right)^2 \cdot v(\lambda) = 1/4$, independent of λ.

b. To use the Taylor's series approximation, we need to express everything in terms of $\theta = \mathrm{E}Y = np$. Then $v(\theta) = \theta(1-\theta/n)$ and

$$\left(\frac{dg^*(\theta)}{d\theta}\right)^2 = \left(\frac{1}{\sqrt{1-\frac{\theta}{n}}} \cdot \frac{1}{2\sqrt{\frac{\theta}{n}}} \cdot \frac{1}{n}\right)^2 = \frac{1}{4n\theta(1-\theta/n)}.$$

Therefore

$$\mathrm{Var}[g^*(Y)] \approx \left(\frac{dg^*(\theta)}{d\theta}\right)^2 v(\theta) = \frac{1}{4n},$$

independent of θ, that is, independent of p.

c. $v(\theta) = K\theta^2$, $\frac{dg^*(\theta)}{d\theta} = \frac{1}{\theta}$ and $\mathrm{Var}[g^*(Y)] \approx \left(\frac{1}{\theta}\right)^2 \cdot K\theta^2 = K$, independent of θ.

11.3 a. $g^*_\lambda(y)$ is clearly continuous with the possible exception of $\lambda = 0$. For that value use l'Hôpital's rule to get

$$\lim_{\lambda\to 0}\frac{y^\lambda - 1}{\lambda} = \lim_{\lambda\to 0}\frac{(\log y)y^\lambda}{1} = \log y.$$

b. From Exercise 11.1, we want to find $v(\lambda)$ that satisfies

$$\frac{y^\lambda - 1}{\lambda} = \int_a^y \frac{1}{\sqrt{v(x)}}dx.$$

Taking derivatives

$$\frac{d}{dy}\left(\frac{y^\lambda - 1}{\lambda}\right) = y^{\lambda-1} = \frac{d}{dy}\int_a^y \frac{1}{\sqrt{v(x)}}dx = \frac{1}{\sqrt{v(y)}}.$$

Thus $v(y) = y^{-2(\lambda-1)}$. From Exercise 11.1,

$$\operatorname{Var}\left(\frac{y^\lambda - 1}{\lambda}\right) \approx \left(\frac{d}{dy}\frac{\theta^\lambda - 1}{\lambda}\right)^2 v(\theta) = \theta^{2(\lambda-1)}\theta^{-2(\lambda-1)} = 1.$$

Note: If $\lambda = 1/2$, $v(\theta) = \theta$, which agrees with Exercise 11.2(a). If $\lambda = 1$ then $v(\theta) = \theta^2$, which agrees with Exercise 11.2(c).

11.5 For the model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i,$$

take $k = 2$. The two parameter configurations

$$\begin{aligned}(\mu, \tau_1, \tau_2) &= (10, 5, 2)\\ (\mu, \tau_1, \tau_2) &= (7, 8, 5),\end{aligned}$$

have the same values for $\mu + \tau_1$ and $\mu + \tau_2$, so they give the same distributions for Y_1 and Y_2.

11.6 a. Under the ANOVA assumptions $Y_{ij} = \theta_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim$ independent $\mathrm{n}(0, \sigma^2)$, so $Y_{ij} \sim$ independent $\mathrm{n}(\theta_i, \sigma^2)$. Therefore the sample pdf is

$$\begin{aligned}\prod_{i=1}^{k}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-1/2}e^{-\frac{(y_{ij}-\theta_i)^2}{2\sigma^2}} &= (2\pi\sigma^2)^{-\Sigma n_i/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2\right\}\\ &= (2\pi\sigma^2)^{-\Sigma n_i/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}n_i\theta_i^2\right\}\\ &\quad\times\exp\left\{-\frac{1}{2\sigma^2}\sum_i\sum_j y_{ij}^2 + \frac{2}{2\sigma^2}\sum_{i=1}^{k}\theta_i n_i \bar{Y}_{i\cdot}\right\}.\end{aligned}$$

Therefore, by the Factorization Theorem,

$$\left(\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{k\cdot}, \sum_i\sum_j Y_{ij}^2\right)$$

is jointly sufficient for $\left(\theta_1, \ldots, \theta_k, \sigma^2\right)$. Since $(\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{k\cdot}, S_p^2)$ is a 1-to-1 function of this vector, $(\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{k\cdot}, S_p^2)$ is also jointly sufficient.

b. We can write

$$\begin{aligned}&(2\pi\sigma^2)^{-\Sigma n_i/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2\right\}\\ &= (2\pi\sigma^2)^{-\Sigma n_i/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}([y_{ij}-\bar{y}_{i\cdot}] + [\bar{y}_{i\cdot}-\theta_i])^2\right\}\\ &= (2\pi\sigma^2)^{-\Sigma n_i/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[y_{ij}-\bar{y}_{i\cdot}]^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}n_i[\bar{y}_{i\cdot}-\theta_i]^2\right\},\end{aligned}$$

so, by the Factorization Theorem, $\bar{Y}_{i\cdot}$, $i = 1, \ldots, n$, is independent of $Y_{ij} - \bar{Y}_{i\cdot}$, $j = 1, \ldots, n_i$, so S_p^2 is independent of each $\bar{Y}_{i\cdot}$.

c. Just identify $n_i\bar{Y}_{i\cdot}$ with X_i and redefine θ_i as $n_i\theta_i$.

11.7 Let $U_i = \bar{Y}_{i\cdot} - \theta_i$. Then

$$\sum_{i=1}^{k} n_i[(\bar{Y}_{i\cdot} - \bar{\bar{Y}}) - (\theta_i - \bar{\theta})]^2 = \sum_{i=1}^{k} n_i(U_i - \bar{U})^2.$$

The U_i are clearly $\mathrm{n}(0, \sigma^2/n_i)$. For $K = 2$ we have

$$\begin{aligned}
S_2^2 &= n_1(U_1 - \bar{U})^2 + n_2(U_2 - \bar{U})^2 \\
&= n_1\left(U_1 - \frac{n_1\bar{U}_1 + n_2\bar{U}_2}{n_1 + n_2}\right)^2 + n_2\left(U_2 - \frac{n_1\bar{U}_1 + n_2\bar{U}_2}{n_1 + n_2}\right)^2 \\
&= (U_1 - U_2)^2\left[n_1\left(\frac{n_2}{n_1 + n_2}\right)^2 + n_2\left(\frac{n_1}{n_1 + n_2}\right)^2\right] \\
&= \frac{(U_1 - U_2)^2}{\frac{1}{n_1} + \frac{1}{n_2}}.
\end{aligned}$$

Since $U_1 - U_2 \sim \mathrm{n}(0, \sigma^2(1/n_1 + 1/n_2))$, $S_2^2/\sigma^2 \sim \chi_1^2$. Let $\bar{U}_k$ be the weighted mean of k U_is, and note that

$$\bar{U}_{k+1} = \bar{U}_k + \frac{n_{k+1}}{N_{k+1}}(U_{k+1} - \bar{U}_k),$$

where $N_k = \sum_{j=1}^{k} n_j$. Then

$$\begin{aligned}
S_{k+1}^2 &= \sum_{i=1}^{k+1} n_i(U_i - \bar{U}_{k+1})^2 \quad = \quad \sum_{i=1}^{k+1} n_i\left[(U_i - \bar{U}_k) - \frac{n_{k+1}}{N_{k+1}}(U_{k+1} - \bar{U}_k)\right]^2 \\
&= S_k^2 + \frac{n_{k+1}N_k}{N_{k+1}}(U_{k+1} - \bar{U}_k)^2,
\end{aligned}$$

where we have expanded the square, noted that the cross-term (summed up to k) is zero, and did a boat-load of algebra. Now since

$$U_{k+1} - \bar{U}_k \sim \mathrm{n}(0, \sigma^2(1/n_{k+1} + 1/N_k)) = \mathrm{n}(0, \sigma^2(N_{k+1}/n_{k+1}N_k)),$$

independent of S_k^2, the rest of the argument is the same as in the proof of Theorem 5.3.1(c).

11.8 Under the oneway ANOVA assumptions, $Y_{ij} \sim$ independent $\mathrm{n}(\theta_i, \sigma^2)$. Therefore

$$\begin{aligned}
\bar{Y}_{i\cdot} &\sim \mathrm{n}\left(\theta_i, \sigma^2/n_i\right) \quad (Y_{ij}\text{'s are independent with common } \sigma^2.) \\
a_i\bar{Y}_{i\cdot} &\sim \mathrm{n}\left(a_i\theta_i, a_i^2\sigma^2/n_i\right) \\
\sum_{i=1}^{k} a_i\bar{Y}_{i\cdot} &\sim \mathrm{n}\left(\sum a_i\theta_i, \sigma^2\sum_{i=1}^{k} a_i^2/n_i\right).
\end{aligned}$$

All these distributions follow from Corollary 4.6.10.

11.9 a. From Exercise 11.8,

$$T = \sum a_i\bar{Y}_i \sim \mathrm{n}\left(\sum a_i\theta_i, \sigma^2\sum a_i^2\right),$$

and under H_0, $\mathrm{E}T = \delta$. Thus, under H_0,

$$\frac{\sum a_i\bar{Y}_i - \delta}{\sqrt{S_p^2\sum a_i^2}} \sim t_{N-k},$$

where $N = \sum n_i$. Therefore, the test is to reject H_0 if

$$\frac{\left|\sum a_i \bar{Y}_i - \delta\right|}{\sqrt{S_p^2 \sum a_i^2/n_i}} > t_{N-k,\frac{\alpha}{2}}.$$

b. Similarly for $H_0\colon \sum a_i\theta_i \le \delta$ vs. $H_1\colon \sum a_i\theta_i > \delta$, we reject H_0 if

$$\frac{\sum a_i \bar{Y}_i - \delta}{\sqrt{S_p^2 \sum a_i^2/n_i}} > t_{N-k,\alpha}.$$

11.10 a. Let H_0^i, $i = 1, \ldots, 4$ denote the null hypothesis using contrast a_i, of the form

$$H_0^i\colon \sum_j a_{ij}\theta_j \ge 0.$$

If H_0^1 is rejected, it indicates that the average of θ_2, θ_3, θ_4, and θ_5 is bigger than θ_1 which is the control mean. If all H_0^i's are rejected, it indicates that $\theta_5 > \theta_i$ for $i = 1, 2, 3, 4$. To see this, suppose H_0^4 and H_0^5 are rejected. This means $\theta_5 > \frac{\theta_5+\theta_4}{2} > \theta_3$; the first inequality is implied by the rejection of H_0^5 and the second inequality is the rejection of H_0^4. A similar argument implies $\theta_5 > \theta_2$ and $\theta_5 > \theta_1$. But, for example, it does not mean that $\theta_4 > \theta_3$ or $\theta_3 > \theta_2$. It also indicates that

$$\frac{1}{2}(\theta_5 + \theta_4) > \theta_3, \quad \frac{1}{3}(\theta_5 + \theta_4 + \theta_3) > \theta_2, \quad \frac{1}{4}(\theta_5 + \theta_4 + \theta_3 + \theta_2) > \theta_1.$$

b. In part a) all of the contrasts are orthogonal. For example,

$$\sum_{i=1}^{5} a_{2i}a_{3i} = \left(0, 1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right)\begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = -\frac{1}{3} + \frac{1}{6} + \frac{1}{6} = 0,$$

and this holds for all pairs of contrasts. Now, from Lemma 5.4.2,

$$\mathrm{Cov}\left(\sum_i a_{ji}\bar{Y}_{i\cdot}, \sum_i a_{j'i}\bar{Y}_{i\cdot}\right) = \frac{\sigma^2}{n}\sum_i a_{ji}a_{j'i},$$

which is zero because the contrasts are orthogonal. Note that the equal number of observations per treatment is important, since if $n_i \ne n_{i'}$ for some i, i', then

$$\mathrm{Cov}\left(\sum_{i=1}^{k} a_{ji}\bar{Y}_i, \sum_{i=1}^{k} a_{j'i}\bar{Y}_i\right) = \sum_{i=1}^{k} a_{ji}a_{j'i}\frac{\sigma^2}{n_i} = \sigma^2\sum_{i=1}^{k}\frac{a_{ji}a_{j'i}}{n_i} \ne 0.$$

c. This is not a set of orthogonal contrasts because, for example, $a_1 \times a_2 = -1$. However, each contrast can be interpreted meaningfully in the context of the experiment. For example, a_1 tests the effect of potassium alone, while a_5 looks at the effect of adding zinc to potassium.

11.11 This is a direct consequence of Lemma 5.3.3.

11.12 a. This is a special case of (11.2.6) and (11.2.7).

b. From Exercise 5.8(a) We know that

$$s^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_{i\cdot} - \bar{\bar{y}})^2 = \frac{1}{2k(k-1)}\sum_{i,i'}(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot})^2 .$$

Then

$$\begin{aligned}\frac{1}{k(k-1)}\sum_{i,i'} t_{ii'}^2 &= \frac{1}{2k(k-1)}\sum_{i,i'}\frac{(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot})^2}{s_p^2/n} = \sum_{i=1}^{k}\frac{(\bar{y}_{i\cdot} - \bar{\bar{y}})^2}{(k-1)s_p^2/n} \\ &= \frac{\sum_i n(\bar{y}_{i\cdot} - \bar{\bar{y}})^2/(k-1)}{s_p^2},\end{aligned}$$

which is distributed as $F_{k-1,N-k}$ under $H_0\colon \theta_1 = \cdots = \theta_k$. Note that

$$\sum_{i,i'} t_{ii'}^2 = \sum_{i=1}^{k}\sum_{i'=1}^{k} t_{ii'}^2,$$

therefore $t_{ii'}^2$ and $t_{i'i}^2$ are both included, which is why the divisor is $k(k-1)$, not $\frac{k(k-1)}{2} = \binom{k}{2}$. Also, to use the result of Example 5.9(a), we treated each mean $\bar{Y}_{i\cdot}$ as an observation, with overall mean $\bar{\bar{Y}}$. This is true for equal sample sizes.

11.13 a.

$$L(\theta|y) = \left(\frac{1}{2\pi\sigma^2}\right)^{Nk/2} e^{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2/\sigma^2}.$$

Note that

$$\begin{aligned}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2 &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2 + \sum_{i=1}^{k} n_i(\bar{y}_{i\cdot}-\theta_i)^2 \\ &= SSW + \sum_{i=1}^{k} n_i(\bar{y}_{i\cdot}-\theta_i)^2,\end{aligned}$$

and the LRT statistic is

$$\lambda = (\hat{\tau}^2/\hat{\tau}_0^2)^{Nk/2}$$

where

$$\hat{\tau}^2 = SSW \quad \text{and} \quad \hat{\tau}_0^2 = SSW + \sum_i n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = SSW + SSB.$$

Thus $\lambda < k$ if and only if SSB/SSW is large, which is equivalent to the F test.

b. The error probabilities of the test are a function of the θ_is only through $\eta = \sum \theta_i^2$. The distribution of F is that of a ratio of chi squared random variables, with the numerator being noncentral (dependent on η). Thus the Type II error is given by

$$P(F > k|\eta) = P\left(\frac{\chi_{k-1}^2(\eta)/(k-1)}{\chi_{N-k}^2/(N-k)} > k\right) \ge P\left(\frac{\chi_{k-1}^2(0)/(k-1)}{\chi_{N-k}^2/(N-k)} > k\right) = \alpha,$$

where the inequality follows from the fact that the noncentral chi squared is stochastically increasing in the noncentrality parameter.

11.14 Let $X_i \sim \mathrm{n}(\theta_i, \sigma^2)$. Then from Exercise 11.11

$$\mathrm{Cov}\left(\sum_i \tfrac{a_i}{\sqrt{c_i}} X_i, \sum_i \sqrt{c_i} v_i X_i\right) = \sigma^2 \sum a_i v_i$$
$$\mathrm{Var}\left(\sum_i \tfrac{a_i}{\sqrt{c_i}} X_i\right) = \sigma^2 \sum \tfrac{a_i^2}{c_i}, \qquad \mathrm{Var}\left(\sum_i \sqrt{c_i} v_i X_i\right) = \sigma^2 \sum c_i v_i^2,$$

and the Cauchy-Schwarz inequality gives

$$\left(\sum a_i v_i\right) \Big/ \left(\sum a_i^2/c_i\right) \le \sum c_i v_i^2.$$

If $a_i = c_i v_i$ this is an equality, hence the LHS is maximized. The simultaneous statement is equivalent to

$$\frac{\left(\sum_{i=1}^k a_i(\bar{y}_{i\cdot} - \theta_i)\right)^2}{\left(s_p^2 \sum_{i=1}^k a_i^2/n\right)} \le M \text{ for all } a_1, \ldots, a_k,$$

and the LHS is maximized by $a_i = n_i(\bar{y}_{i\cdot} - \theta_i)$. This produces the F statistic.

11.15 a. Since $t_\nu^2 = F_{1,\nu}$, it follows from Exercise 5.19(b) that for $k \ge 2$

$$P[(k-1)F_{k-1,\nu} \ge a] \ge P(t_\nu^2 \ge a).$$

So if $a = t^2_{\nu,\alpha/2}$, the F probability is greater than α, and thus the α-level cutoff for the F must be greater than $t^2_{\nu,\alpha/2}$.

b. The only difference in the intervals is the cutoff point, so the Scheffé intervals are wider.

c. Both sets of intervals have nominal level $1-\alpha$, but since the Scheffé intervals are wider, tests based on them have a smaller rejection region. In fact, the rejection region is contained in the t rejection region. So the t is more powerful.

11.16 a. If $\theta_i = \theta_j$ for all i, j, then $\theta_i - \theta_j = 0$ for all i, j, and the converse is also true.

b. $H_0\colon \boldsymbol{\theta} \in \cap_{ij}\Theta_{ij}$ and $H_1\colon \boldsymbol{\theta} \in \cup_{ij}(\Theta_{ij})^c$.

11.17 a. If all of the means are equal, the Scheffé test will only reject α of the time, so the t tests will be done only α of the time. The experimentwise error rate is preserved.

b. This follows from the fact that the t tests use a smaller cutoff point, so there can be rejection using the t test but no rejection using Scheffé. Since Scheffé has experimentwise level α, the t test has experimentwise error greater than α.

c. The pooled standard deviation is 2.358, and the means and t statistics are

Mean			t statistic		
Low	Medium	High	Med-Low	High-Med	High-Low
3.51.	9.27	24.93	3.86	10.49	14.36

The t statistics all have 12 degrees of freedom and, for example, $t_{12,.01} = 2.68$, so all of the tests reject and we conclude that the means are all significantly different.

11.18 a.

$$\begin{aligned} P(Y > a \mid Y > b) &= P(Y > a, Y > b)/P(Y > b) \\ &= P(Y > a)/P(Y > b) && (a > b) \\ &> P(Y > a). && (P(Y > b) < 1) \end{aligned}$$

b. If a is a cutoff point then we would declare significance if $Y > a$. But if we only check if Y is significant because we see a big Y $(Y > b)$, the proper significance level is $P(Y > a \mid Y > b)$, which will show less significance than $P(Y > a)$.

11.19 a. The marginal distributions of the Y_i are somewhat straightforward to derive. As $X_{i+1} \sim$ gamma$(\lambda_{i+1}, 1)$ and, independently, $\sum_{j=1}^{i} X_j \sim$ gamma$(\sum_{j=1}^{i} \lambda_j, 1)$ (Example 4.6.8), we only need to derive the distribution of the ratio of two independent gammas. Let $X \sim$ gamma$(\lambda_1, 1)$ and Y $\sim$ gamma$(\lambda_2, 1)$. Make the transformation

$$u = x/y, \quad v = y \quad \Rightarrow \quad x = uv, \quad y = v,$$

with Jacobian v. The density of (U, V) is

$$f(u, v) = \frac{1}{\Gamma(\lambda_1)\Gamma(\lambda_2)}(uv)^{\lambda_1 - 1} v^{\lambda_2 - 1} v e^{-uv} e^{-v} = \frac{u^{\lambda_1 - 1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} v^{\lambda_1 + \lambda_2 - 1} e^{-v(1+u)}.$$

To get the density of U, integrate with respect to v. Note that we have the kernel of a gamma$(\lambda_1 + \lambda_2, 1/(1 + u))$, which yields

$$f(u) = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \frac{u^{\lambda_1 - 1}}{(1 + u)^{\lambda_1 + \lambda_2 - 1}}.$$

The joint distribution is a nightmare. We have to make a multivariate change of variable. This is made a bit more palatable if we do it in two steps. First transform

$$W_1 = X_1, \quad W_2 = X_1 + X_2, \quad W_3 = X_1 + X_2 + X_3, \quad \ldots, \quad W_n = X_1 + X_2 + \cdots + X_n,$$

with

$$X_1 = W_1, \quad X_2 = W_2 - W_1, \quad X_3 = W_3 - W_2, \quad \ldots \quad X_n = W_n - W_{n-1},$$

and Jacobian 1. The joint density of the W_i is

$$f(w_1, w_2, \ldots, w_n) = \prod_{i=1}^{n} \frac{1}{\Gamma(\lambda_i)} (w_i - w_{i-1})^{\lambda_i - 1} e^{-w_n}, \quad w_1 \le w_2 \le \cdots \le w_n,$$

where we set $w_0 = 0$ and note that the exponent telescopes. Next note that

$$y_1 = \frac{w_2 - w_1}{w_1}, \quad y_2 = \frac{w_3 - w_2}{w_2}, \quad \ldots \quad y_{n-1} = \frac{w_n - w_{n-1}}{w_{n-1}}, \quad y_n = w_n,$$

with

$$w_i = \frac{y_n}{\prod_{j=i}^{n-1}(1 + y_j)}, \quad i = 1, \ldots, n-1, \quad w_n = y_n.$$

Since each w_i only involves y_j with $j \ge i$, the Jacobian matrix is triangular and the determinant is the product of the diagonal elements. We have

$$\frac{dw_i}{dy_i} = -\frac{y_n}{(1 + y_i)\prod_{j=i}^{n-1}(1 + y_j)}, \quad i = 1, \ldots, n-1, \quad \frac{dw_n}{dy_n} = 1,$$

and

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= \frac{1}{\Gamma(\lambda_1)} \left(\frac{y_n}{\prod_{j=1}^{n-1}(1 + y_j)} \right)^{\lambda_1 - 1} \\ &\quad \times \prod_{i=2}^{n-1} \frac{1}{\Gamma(\lambda_i)} \left(\frac{y_n}{\prod_{j=i}^{n-1}(1 + y_j)} - \frac{y_n}{\prod_{j=i-1}^{n-1}(1 + y_j)} \right)^{\lambda_i - 1} e^{-y_n} \\ &\quad \times \prod_{i=1}^{n-1} \frac{y_n}{(1 + y_i)\prod_{j=i}^{n-1}(1 + y_j)}. \end{aligned}$$

Factor out the terms with y_n and do some algebra on the middle term to get

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= y_n^{\Sigma_i \lambda_i - 1} e^{-y_n} \frac{1}{\Gamma(\lambda_1)} \left(\frac{1}{\prod_{j=1}^{n-1}(1+y_j)} \right)^{\lambda_1 - 1} \\ &\quad \times \prod_{i=2}^{n-1} \frac{1}{\Gamma(\lambda_i)} \left(\frac{y_{i-1}}{1+y_{i-1}} \frac{1}{\prod_{j=i}^{n-1}(1+y_j)} \right)^{\lambda_i - 1} \\ &\quad \times \prod_{i=1}^{n-1} \frac{1}{(1+y_i)\prod_{j=i}^{n-1}(1+y_j)}. \end{aligned}$$

We see that Y_n is independent of the other Y_i (and has a gamma distribution), but there does not seem to be any other obvious conclusion to draw from this density.

b. The Y_i are related to the F distribution in the ANOVA. For example, as long as the sum of the λ_i are integers,

$$Y_i = \frac{X_{i+1}}{\sum_{j=1}^{i} X_j} = \frac{2X_{i+1}}{2\sum_{j=1}^{i} X_j} = \frac{\chi^2_{\lambda_{i+1}}}{\chi^2_{\sum_{j=1}^{i}\lambda_j}} \sim F_{\lambda_{i+1}, \sum_{j=1}^{i}\lambda_j}.$$

Note that the F density makes sense even if the λ_i are not integers.

11.21 a.

$$\begin{aligned} \text{Grand mean } \bar{y}_{\cdot\cdot} &= \frac{188.54}{15} = 12.57 \\ \text{Total sum of squares} &= \sum_{i=1}^{3}\sum_{j=1}^{5}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = 1295.01. \\ \text{Within SS} &= \sum_{1}^{3}\sum_{1}^{5}(y_{ij} - \bar{y}_{i\cdot})^2 \\ &= \sum_{1}^{5}(y_{1j} - 3.508)^2 + \sum_{1}^{5}(y_{2j} - 9.274)^2 + \sum_{1}^{5}(y_{3j} - 24.926)^2 \\ &= 1.089 + 2.189 + 63.459 = 66.74 \\ \text{Between SS} &= 5\left(\sum_{1}^{3}(y_{ij} - \bar{y}_{i\cdot})^2\right) \\ &= 5(82.120 + 10.864 + 152.671) = 245.65 \cdot 5 = 1228.25. \end{aligned}$$

ANOVA table:

Source	df	SS	MS	F
Treatment	2	1228.25	614.125	110.42
Within	12	66.74	5.562	
Total	14	1294.99		

Note that the total SS here is different from above – round off error is to blame. Also, $F_{2,12} = 110.42$ is highly significant.

b. Completing the proof of (11.2.4), we have

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}((y_{ij} - \bar{y}_{i\cdot}) + (\bar{y}_i - \bar{\bar{y}}))^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot}-\bar{\bar{y}})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})(\bar{y}_{i\cdot}-\bar{\bar{y}}),$$

where the cross term (the sum over j) is zero, so the sum of squares is partitioned as

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2+\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{\bar{y}})^2$$

c. From a), the F statistic for the ANOVA is 110.42. The individual two-sample t's, using $s_p^2 = \frac{1}{15-3}(66.74) = 5.5617$, are

$$\begin{aligned}
t_{12}^2 &= \frac{(3.508-9.274)^2}{(5.5617)(2/5)} = \frac{33.247}{2.2247} = 14.945,\\
t_{13}^2 &= \frac{(3.508-24.926)^2}{2.2247} = 206.201,\\
t_{23}^2 &= \frac{(9.274-24.926)^2}{2.2247} = 110.122,
\end{aligned}$$

and

$$\frac{2(14.945)+2(206.201)+(110.122)}{6} = 110.42 = F.$$

11.23 a.

$$\begin{aligned}
\mathrm{E}Y_{ij} &= \mathrm{E}(\mu+\tau_i+b_j+\epsilon_{ij}) = \mu+\tau_i+\mathrm{E}b_j+\mathrm{E}\epsilon_{ij} = \mu+\tau_i\\
\mathrm{Var}Y_{ij} &= \mathrm{Var}b_j+\mathrm{Var}\epsilon_{ij} = \sigma_B^2+\sigma^2,
\end{aligned}$$

by independence of b_j and ϵ_{ij}.

b.

$$\mathrm{Var}\left(\sum_{i=1}^{n}a_i\bar{Y}_{i\cdot}\right)=\sum_{i=1}^{n}a_i^2\mathrm{Var}\bar{Y}_{i\cdot}+2\sum_{i>i'}\mathrm{Cov}(a_iY_{i\cdot},a_{i'}Y_{i'\cdot}).$$

The first term is

$$\sum_{i=1}^{n}a_i^2\mathrm{Var}\bar{Y}_{i\cdot}=\sum_{i=1}^{n}a_i^2\mathrm{Var}\left(\frac{1}{r}\sum_{j=1}^{r}\mu+\tau_i+b_j+\epsilon_{ij}\right)=\frac{1}{r^2}\sum_{i=1}^{n}a_i^2(r\sigma_B^2+r\sigma^2)$$

from part (a). For the covariance

$$\mathrm{E}\bar{Y}_{i\cdot}=\mu+\tau_i,$$

and

$$\begin{aligned}
\mathrm{E}(\bar{Y}_{i\cdot}\bar{Y}_{i'\cdot}) &= \mathrm{E}\left(\left[\mu+\tau_i+\frac{1}{r}\sum_j(b_j+\epsilon_{ij})\right]\left[\mu+\tau_{i'}+\frac{1}{r}\sum_j(b_j+\epsilon_{i'j})\right]\right)\\
&= (\mu+\tau_i)(\mu+\tau_{i'})+\frac{1}{r^2}\mathrm{E}\left(\left[\sum_j(b_j+\epsilon_{ij})\right]\left[\sum_j(b_j+\epsilon_{i'j})\right]\right)
\end{aligned}$$

since the cross terms have expectation zero. Next, expanding the product in the second term again gives all zero cross terms, and we have

$$\mathrm{E}(\bar{Y}_{i\cdot}\bar{Y}_{i'\cdot}) = (\mu+\tau_i)(\mu+\tau_{i'}) + \frac{1}{r^2}(r\sigma_B^2),$$

and

$$\mathrm{Cov}(\bar{Y}_{i\cdot}, \bar{Y}_{i'\cdot}) = \sigma_B^2/r.$$

Finally, this gives

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n a_i\bar{Y}_{i\cdot}\right) &= \frac{1}{r^2}\sum_{i=1}^n a_i^2(r\sigma_B^2 + r\sigma^2) + 2\sum_{i>i'} a_i a_{i'}\sigma_B^2/r \\
&= \frac{1}{r}\left[\sum_{i=1}^n a_i^2\sigma^2 + \sigma_B^2(\sum_{i=1}^n a_i)^2\right] \\
&= \frac{1}{r}\sigma^2\sum_{i=1}^n a_i^2 \\
&= \frac{1}{r}(\sigma^2+\sigma_B^2)(1-\rho)\sum_{i=1}^n a_i^2,
\end{aligned}$$

where, in the third equality we used the fact that $\sum_i a_i = 0$.

11.25 Differentiation yields

a. $\frac{\partial}{\partial c}RSS = 2\sum\left[y_i - (c+dx_i)\right](-1) \stackrel{\text{set}}{=} 0 \quad\Rightarrow\quad nc + d\sum x_i = \sum y_i$

$\frac{\partial}{\partial d}RSS = 2\sum\left[y_i - (c_i+dx_i)\right](-x_i) \stackrel{\text{set}}{=} 0 \quad\Rightarrow\quad c\sum x_i + d\sum x_i^2 = \sum x_i y_i.$

b. Note that $nc + d\sum x_i = \sum y_i \Rightarrow c = \bar{y} - d\bar{x}$. Then

$$(\bar{y}-d\bar{x})\sum x_i + d\sum x_i^2 = \sum x_i y_i \quad\text{and}\quad d\left(\sum x_i^2 - n\bar{x}^2\right) = \sum x_i y_i - \sum x_i\bar{y}$$

which simplifies to $d = \sum x_i(y_i-\bar{y})/\sum(x_i-\bar{x})^2$. Thus c and d are the least squares estimates.

c. The second derivatives are

$$\frac{\partial^2}{\partial c^2}\mathrm{RSS} = n, \quad \frac{\partial^2}{\partial c\partial d}\mathrm{RSS} = \sum x_i, \quad \frac{\partial^2}{\partial d^2}\mathrm{RSS} = \sum x_i^2.$$

Thus the Jacobian of the second-order partials is

$$\begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = n\sum x_i^2 - \left(\sum x_i\right)^2 = n\sum(x_i-\bar{x})^2 > 0.$$

11.27 For the linear estimator $\sum_i a_i Y_i$ to be unbiased for α we have

$$\mathrm{E}\left(\sum_i a_i Y_i\right) = \sum_i a_i(\alpha+\beta x_i) = \alpha \Rightarrow \sum_i a_i = 1 \text{ and } \sum_i a_i x_i = 0.$$

Since $\mathrm{Var}\sum_i a_i Y_i = \sigma^2\sum_i a_i^2$, we need to solve:

$$\text{minimize } \sum_i a_i^2 \text{ subject to } \sum_i a_i = 1 \text{ and } \sum_i a_i x_i = 0.$$

A solution can be found with Lagrange multipliers, but verifying that it is a minimum is excruciating. So instead we note that

$$\sum_i a_i = 1 \Rightarrow a_i = \frac{1}{n} + k(b_i - \bar{b}),$$

for some constants $k, b_1, b_2, \ldots, b_n$, and

$$\sum_i a_i x_i = 0 \Rightarrow k = \frac{-\bar{x}}{\sum_i (b_i - \bar{b})(x_i - \bar{x})} \text{ and } a_i = \frac{1}{n} - \frac{\bar{x}(b_i - \bar{b})}{\sum_i (b_i - \bar{b})(x_i - \bar{x})}.$$

Now

$$\sum_i a_i^2 = \sum_i \left[\frac{1}{n} - \frac{\bar{x}(b_i - \bar{b})}{\sum_i (b_i - \bar{b})(x_i - \bar{x})}\right]^2 = \frac{1}{n} + \frac{\bar{x}^2 \sum_i (b_i - \bar{b})^2}{[\sum_i (b_i - \bar{b})(x_i - \bar{x})]^2},$$

since the cross term is zero. So we need to minimize the last term. From Cauchy-Schwarz we know that

$$\frac{\sum_i (b_i - \bar{b})^2}{[\sum_i (b_i - \bar{b})(x_i - \bar{x})]^2} \geq \frac{1}{\sum_i (x_i - \bar{x})]^2},$$

and the minimum is attained at $b_i = x_i$. Substituting back we get that the minimizing a_i is $\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$, which results in $\sum_i a_i Y_i = \bar{Y} - \hat{\beta}\bar{x}$, the least squares estimator.

11.28 To calculate

$$\max_{\sigma^2} L(\sigma^2 | y, \hat{\alpha}\hat{\beta}) \quad = \quad \max_{\sigma^2} \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2}\Sigma_i [y_i - (\hat{\alpha} + \hat{\beta}x_i)]^2 / \sigma^2}$$

take logs and differentiate with respect to σ^2 to get

$$\frac{d}{d\sigma^2} \log L(\sigma^2 | y, \hat{\alpha}, \hat{\beta}) = -\frac{n}{2\sigma^2} + \frac{1}{2} \frac{\sum_i [y_i - (\hat{\alpha} + \hat{\beta}x_i)]^2}{(\sigma^2)^2}.$$

Set this equal to zero and solve for σ^2. The solution is $\hat{\sigma}^2$.

11.29 a.

$$\mathrm{E}\hat{\epsilon}_i = \mathrm{E}(Y_i - \hat{\alpha} - \hat{\beta}x_i) = (\alpha + \beta x_i) - \alpha - \beta x_i = 0.$$

b.

$$\begin{aligned}
\mathrm{Var}\hat{\epsilon}_i &= \mathrm{E}[Y_i - \hat{\alpha} - \hat{\beta}x_i]^2 \\
&= \mathrm{E}[(Y_i - \alpha - \beta x_i) - (\hat{\alpha} - \alpha) - x_i(\hat{\beta} - \beta)]^2 \\
&= \mathrm{Var}Y_i + \mathrm{Var}\hat{\alpha} + x_i^2 \mathrm{Var}\hat{\beta} - 2\mathrm{Cov}(Y_i, \hat{\alpha}) - 2x_i \mathrm{Cov}(Y_i, \hat{\beta}) + 2x_i \mathrm{Cov}(\hat{\alpha}, \hat{\beta}).
\end{aligned}$$

11.30 a. Straightforward algebra shows

$$\begin{aligned}
\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\
&= \sum \frac{1}{n} y_i - \frac{\bar{x} \sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \\
&= \sum \left[\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}\right] y_i.
\end{aligned}$$

b. Note that for $c_i = \frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$, $\sum c_i = 1$ and $\sum c_i x_i = 0$. Then

$$\begin{aligned} \mathrm{E}\hat{\alpha} &= \mathrm{E}\sum c_i Y_i &= \sum c_i(\alpha+\beta x_i &= \alpha, \\ \mathrm{Var}\hat{\alpha} &= \sum c_i^2 \mathrm{Var} Y_i &= \sigma^2 \sum c_i^2, \end{aligned}$$

and

$$\begin{aligned} \sum c_i^2 &= \sum \left[\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right]^2 &= \sum \frac{1}{n^2} + \frac{\sum \bar{x}^2(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2} \qquad \text{(cross term = 0)} \\ &= \frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2} &= \frac{\sum x_i^2}{nS_{xx}}. \end{aligned}$$

c. Write $\hat{\beta} = \sum d_i y_i$, where

$$d_i = \frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}.$$

From Exercise 11.11,

$$\begin{aligned} \mathrm{Cov}(\hat{\alpha},\hat{\beta}) &= \mathrm{Cov}\left(\sum c_i Y_i, \sum d_i Y_i\right) &= \sigma^2 \sum c_i d_i \\ &= \sigma^2 \sum \left[\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right] \frac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} &= \frac{-\sigma^2\bar{x}}{\sum(x_i-\bar{x})^2}. \end{aligned}$$

11.31 The fact that

$$\hat{\epsilon}_i = \sum_i [\delta_{ij} - (c_j + d_j x_i)] Y_j$$

follows directly from (11.3.27) and the definition of c_j and d_j. Since $\hat{\alpha} = \sum_i c_i Y_i$, from Lemma 11.3.2

$$\begin{aligned} \mathrm{Cov}(\hat{\epsilon}_i, \hat{\alpha}) &= \sigma^2 \sum_j c_j[\delta_{ij} - (c_j + d_j x_i)] \\ &= \sigma^2 \left[c_i - \sum_j c_j(c_j + d_j x_i)\right] \\ &= \sigma^2 \left[c_i - \sum_j c_j^2 - x_i \sum_j c_j d_j\right]. \end{aligned}$$

Substituting for c_j and d_j gives

$$\begin{aligned} c_i &= \frac{1}{n} - \frac{(x_i-\bar{x})\bar{x}}{S_{xx}} \\ \sum_j c_j^2 &= \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \\ x_i \sum_j c_j d_j &= -\frac{x_i\bar{x}}{S_{xx}}, \end{aligned}$$

and substituting these values shows $\mathrm{Cov}(\hat{\epsilon}_i, \hat{\alpha}) = 0$. Similarly, for $\hat{\beta}$,

$$\mathrm{Cov}(\hat{\epsilon}_i, \hat{\beta}) = \sigma^2 \left[d_i - \sum_j c_j d_j - x_i \sum_j d_j^2\right]$$

with

$$\begin{aligned} d_i &= \frac{(x_i - \bar{x})}{S_{xx}} \\ \sum_j c_j d_j &= -\frac{\bar{x}}{S_{xx}} \\ x_i \sum_j d_j^2 &= \frac{1}{S_{xx}}, \end{aligned}$$

and substituting these values shows $\mathrm{Cov}(\hat{\epsilon}_i, \hat{\beta}) = 0$.

11.32 Write the models as

$$\begin{aligned} 3y_i &= \alpha + \beta x_i + \epsilon_i \\ y_i &= \alpha' + \beta'(x_i - \bar{x}) + \epsilon_i \\ &= \alpha' + \beta' z_i + \epsilon_i. \end{aligned}$$

a. Since $\bar{z} = 0$,

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum z_i(y_i - \bar{y})}{\sum z_i^2} = \hat{\beta}'.$$

b.

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}, \\ \hat{\alpha}' &= \bar{y} - \hat{\beta}'\bar{z} = \bar{y} \end{aligned}$$

since $\bar{z} = 0$.

$$\hat{\alpha}' \sim \mathrm{n}(\alpha + \beta\bar{z}, \sigma^2/n) = \mathrm{n}(\alpha, \sigma^2/n).$$

c. Write

$$\hat{\alpha}' = \sum \frac{1}{n} y_i \hat{\beta}' = \sum \left(\frac{z_i}{\sum z_i^2} \right) y_i.$$

Then

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \sum \frac{1}{n} \left(\frac{z_i}{\sum z_i^2} \right) = 0,$$

since $\sum z_i = 0$.

11.33 a. From (11.23.25), $\beta = \rho(\sigma_Y/\sigma_X)$, so $\beta = 0$ if and only if $\rho = 0$ (since we assume that the variances are positive).

b. Start from the display following (11.3.35). We have

$$\begin{aligned} \frac{\hat{\beta}^2}{S^2/S_{xx}} &= \frac{S_{xy}^2/S_{xx}}{RSS/(n-2)} \\ &= (n-2)\frac{S_{xy}^2}{\left(S_{yy} - S_{xy}^2/S_{xx}\right) S_{xx}} \\ &= (n-2)\frac{S_{xy}^2}{\left(S_{yy}S_{xx} - S_{xy}^2\right)}, \end{aligned}$$

and dividing top and bottom by $S_{yy}S_{xx}$ finishes the proof.

c. From (11.3.33) if $\rho = 0$ (equivalently $\beta = 0$), then $\hat{\beta}/(S/\sqrt{S_{xx}}) = \sqrt{n-2}\, r/\sqrt{1-r^2}$ has a t_{n-2} distribution.

11.34 a. ANOVA table for height data

Source	df	SS	MS	F
Regression	1	60.36	60.36	50.7
Residual	6	7.14	1.19	
Total	7	67.50		

The least squares line is $\hat{y} = 35.18 + .93x$.

b. Since $y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$, we just need to show that the cross term is zero.

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n}\left[y_i - (\hat{\alpha} + \hat{\beta}x_i)\right]\left[(\hat{\alpha} + \hat{\beta}x_i) - \bar{y}\right] \\
&= \sum_{i=1}^{n}\left[(\hat{y}_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})\right]\left[\hat{\beta}(x_i - \bar{x})\right] \qquad (\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}) \\
&= \hat{\beta}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0,
\end{aligned}$$

from the definition of $\hat{\beta}$.

c.

$$\sum(\hat{y}_i - \bar{y})^2 = \hat{\beta}^2\sum(x_i - \bar{x})^2 = \frac{S_{xy}^2}{S_{xx}}.$$

11.35 a. For the least squares estimate:

$$\frac{d}{d\theta}\sum_i(y_i - \theta x_i^2)^2 = 2\sum_i(y_i - \theta x_i^2)x_i^2 = 0$$

which implies

$$\hat{\theta} = \frac{\sum_i y_i x_i^2}{\sum_i x_i^4}.$$

b. The log likelihood is

$$\log L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i(y_i - \theta x_i^2)^2,$$

and maximizing this is the same as the minimization in part (a).

c. The derivatives of the log likelihood are

$$\begin{aligned}
\frac{d}{d\theta}\log L &= \frac{1}{\sigma^2}\sum_i(y_i - \theta x_i^2)x_i^2 \\
\frac{d^2}{d\theta^2}\log L &= \frac{-1}{\sigma^2}\sum_i x_i^4,
\end{aligned}$$

so the CRLB is $\sigma^2/\sum_i x_i^4$. The variance of $\hat{\theta}$ is

$$\mathrm{Var}\hat{\theta} = \mathrm{Var}\left(\frac{\sum_i y_i x_i^2}{\sum_i x_i^4}\right) = \sum_i\left(\frac{x_i^2}{\sum_j x_j^4}\right)\sigma^2 = \sigma^2/\sum_i x_i^4,$$

so $\hat{\theta}$ is the best unbiased estimator.

11.36 a.

$$\begin{aligned}
\mathrm{E}\hat{\alpha} &= \mathrm{E}(\bar{Y}-\hat{\beta}\bar{X}) = \mathrm{E}\left[\mathrm{E}(\bar{Y}-\hat{\beta}\bar{X}|\bar{X})\right] = \mathrm{E}\left[\alpha+\beta\bar{X}-\beta\bar{X}\right] = \mathrm{E}\alpha = \alpha. \\
\mathrm{E}\hat{\beta} &= \mathrm{E}[\mathrm{E}(\hat{\beta}|\bar{X})] = \mathrm{E}\beta = \beta.
\end{aligned}$$

b. Recall

$$\begin{aligned}
\mathrm{Var}Y &= \mathrm{Var}[\mathrm{E}(Y|X)] + \mathrm{E}[\mathrm{Var}(Y|X)] \\
\mathrm{Cov}(Y,Z) &= \mathrm{Cov}[\mathrm{E}(Y|X),\mathrm{E}(Z|X)] + \mathrm{E}[\mathrm{Cov}(Y,Z|X)].
\end{aligned}$$

Thus

$$\begin{aligned}
\mathrm{Var}\hat{\alpha} &= \mathrm{E}[\mathrm{Var}(\hat{\alpha}|X)] = \sigma^2\mathrm{E}\left[\sum X_i^2 \middle/ S_{XX}\right] \\
\mathrm{Var}\hat{\beta} &= \sigma^2\mathrm{E}[1/S_{XX}] \\
\mathrm{Cov}(\hat{\alpha},\hat{\beta}) &= \mathrm{E}[\mathrm{Cov}(\hat{\alpha},\hat{\beta}|\hat{X})] = -\sigma^2\mathrm{E}[\bar{X}/S_{XX}].
\end{aligned}$$

11.37 This is almost the same problem as Exercise 11.35. The log likelihood is

$$\log L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i-\beta x_i)^2.$$

The MLE is $\sum_i x_i y_i / \sum_i x_i^2$, with mean β and variance $\sigma^2/\sum_i x_i^2$, the CRLB.

11.38 a. The model is $y_i = \theta x_i + \epsilon_i$, so the least squares estimate of θ is $\sum x_i y_i / \sum x_i^2$ (regression through the origin).

$$\begin{aligned}
\mathrm{E}\left(\frac{\sum x_i Y_i}{\sum x_i^2}\right) &= \frac{\sum x_i(x_i\theta)}{\sum x_i^2} = \theta \\
\mathrm{Var}\left(\frac{\sum x_i Y_i}{\sum x_i^2}\right) &= \frac{\sum x_i^2(x_i\theta)}{\left(\sum x_i^2\right)^2} = \theta\frac{\sum x_i^3}{\left(\sum x_i^2\right)^2}.
\end{aligned}$$

The estimator is unbiased.

b. The likelihood function is

$$\begin{aligned}
L(\theta|\mathbf{x}) &= \prod_{i=1}^n \frac{e^{-\theta x_i}(\theta x_i)^{y_i}}{(y_i)!} = \frac{e^{-\theta\Sigma x_i}\prod(\theta x_i)^{y_i}}{\prod y_i!} \\
\frac{\partial}{\partial\theta}\log L &= \frac{\partial}{\partial\theta}\left[-\theta\sum x_i + \sum y_i\log(\theta x_i) - \log\prod y_i!\right] \\
&= -\sum x_i + \sum\frac{x_i y_i}{\theta x_i} \stackrel{\text{set}}{=} 0
\end{aligned}$$

which implies

$$\hat{\theta} = \frac{\sum y_i}{\sum x_i}$$

$$\mathrm{E}\hat{\theta} = \frac{\sum\theta x_i}{\sum x_i} = \theta \quad \text{and} \quad \mathrm{Var}\hat{\theta} = \mathrm{Var}\left(\frac{\sum y_i}{\sum x_i}\right) = \frac{\sum\theta x_i}{\left(\sum x_i\right)^2} = \frac{\theta}{\sum x_i}.$$

c.

$$\frac{\partial^2}{\partial\theta^2}\log L = \frac{\partial}{\partial\theta}\left[-\sum x_i + \frac{\sum y_i}{\theta}\right] = \frac{-\sum y_i}{\theta^2} \quad \text{and} \quad \mathrm{E} - \frac{\partial^2}{\partial\theta^2}\log L = \frac{\sum x_i}{\theta}.$$

Thus, the CRLB is $\theta/\sum x_i$, and the MLE is the best unbiased estimator.

11.39 Let A_i be the set

$$A_i = \left\{ \hat{\alpha}, \hat{\beta} : \left[(\hat{\alpha} + \hat{\beta} x_{0i}) - (\alpha + \beta x_{0i}) \right] \Big/ \left[S \sqrt{\frac{1}{n} + \frac{(x_{0i} - \bar{x})^2}{S_{xx}}} \right] \leq t_{n-2,\alpha/2m} \right\}.$$

Then $P\left(\cap_{i=1}^m A_i\right)$ is the probability of simultaneous coverage, and using the Bonferroni Inequality (1.2.10) we have

$$P\left(\cap_{i=1}^m A_i\right) \geq \sum_{i=1}^m P(A_i) - (m-1) = \sum_{i=1}^m \left(1 - \frac{\alpha}{m}\right) - (m-1) = 1 - \alpha.$$

11.41 Assume that we have observed data $(y_1, x_1), (y_2, x_2), \ldots, (y_{n-1}, x_{n-1})$ and we have x_n but not y_n. Let $\phi(y_i|x_i)$ denote the density of Y_i, a $\mathrm{n}(a + bx_i, \sigma^2)$.

a. The expected complete-data log likelihood is

$$\mathrm{E}\left(\sum_{i=1}^n \log \phi(Y_i|x_i)\right) = \sum_{i=1}^{n-1} \log \phi(y_i|x_i) + \mathrm{E} \log \phi(Y|x_n),$$

where the expectation is respect to the distribution $\phi(y|x_n)$ with the current values of the parameter estimates. Thus we need to evaluate

$$\mathrm{E} \log \phi(Y|x_n) = \mathrm{E}\left(-\frac{1}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}(Y - \mu_1)^2\right),$$

where $Y \sim \mathrm{n}(\mu_0, \sigma_0^2)$. We have

$$\mathrm{E}(Y - \mu_1)^2 = \mathrm{E}([Y - \mu_0] + [\mu_0 - \mu_1])^2 = \sigma_0^2 + [\mu_0 - \mu_1]^2,$$

since the cross term is zero. Putting this all together, the expected complete-data log likelihood is

$$\begin{aligned} -\frac{n}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n-1}[y_i - (a_1 + b_1 x_i)]^2 - \frac{\sigma_0^2 + [(a_0 + b_0 x_n) - (a_1 + b_1 x_n)]^2}{2\sigma_1^2} \\ = -\frac{n}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}[y_i - (a_1 + b_1 x_i)]^2 - \frac{\sigma_0^2}{2\sigma_1^2} \end{aligned}$$

if we define $y_n = a_0 + b_0 x_n$.

b. For fixed a_0 and b_0, maximizing this likelihood gives the least squares estimates, while the maximum with respect to σ_1^2 is

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^n [y_i - (a_1 + b_1 x_i)]^2 + \sigma_0^2}{n}.$$

So the EM algorithm is the following: At iteration t, we have estimates $\hat{a}^{(t)}$, $\hat{b}^{(t)}$, and $\hat{\sigma}^{2(t)}$. We then set $y_n^{(t)} = \hat{a}^{(t)} + \hat{b}^{(t)} x_n$ (which is essentially the E-step) and then the M-step is to calculate $\hat{a}^{(t+1)}$ and $\hat{b}^{(t+1)}$ as the least squares estimators using (y_1, x_1), (y_2, x_2), $\ldots$ (y_{n-1}, x_{n-1}), $(y_n^{(t)}, x_n)$, and

$$\hat{\sigma}_1^{2(t+1)} = \frac{\sum_{i=1}^n [y_i - (a^{(t+1)} + b^{(t+1)} x_i)]^2 + \sigma_0^{2(t)}}{n}.$$

c. The EM calculations are simple here. Since $y_n^{(t)} = \hat{a}^{(t)} + \hat{b}^{(t)} x_n$, the estimates of a and b must converge to the least squares estimates (since they minimize the sum of squares of the observed data, and the last term adds nothing. For $\hat{\sigma}^2$ we have (substituting the least squares estimates) the stationary point

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n [y_i - (\hat{a} + \hat{b}x_i)]^2 + \hat{\sigma}^2}{n} \quad \Rightarrow \quad \hat{\sigma}^2 = \sigma^2_{\text{obs}},$$

where σ^2_{obs} is the MLE from the $n-1$ observed data points. So the MLE s are the same as those without the extra x_n.

d. Now we use the bivariate normal density (see Definition 4.5.10 and Exercise 4.45). Denote the density by $\phi(x, y)$. Then the expected complete-data log likelihood is

$$\sum_{i=1}^{n-1} \log \phi(x_i, y_i) + \mathrm{E} \log \phi(X, y_n),$$

where after iteration t the missing data density is the conditional density of X given $Y = y_n$,

$$X|Y = y_n \sim \mathrm{n}\left(\mu_X^{(t)} + \rho^{(t)}(\sigma_X^{(t)}/\sigma_Y^{(t)})(y_n - \mu_Y^{(t)}), (1 - \rho^{2(t)})\sigma_X^{2(t)}\right).$$

Denoting the mean by μ_0 and the variance by σ_0^2, the expected value of the last piece in the likelihood is

$$\begin{aligned}
&\mathrm{E} \log \phi(X, y_n) \\
&= -\frac{1}{2}\log(2\pi\sigma_X^2\sigma_Y^2(1-\rho^2)) \\
&\quad -\frac{1}{2(1-\rho^2)}\left[\mathrm{E}\left(\frac{X-\mu_X}{\sigma_X}\right)^2 - 2\rho\mathrm{E}\left(\frac{(X-\mu_X)(y_n-\mu_Y)}{\sigma_X\sigma_Y}\right) + \left(\frac{y_n-\mu_Y}{\sigma_Y}\right)^2\right] \\
&= -\frac{1}{2}\log(2\pi\sigma_X^2\sigma_Y^2(1-\rho^2)) \\
&\quad -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_0^2}{\sigma_X^2} + \left(\frac{\mu_0-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{(\mu_0-\mu_X)(y_n-\mu_Y)}{\sigma_X\sigma_Y}\right) + \left(\frac{y_n-\mu_Y}{\sigma_Y}\right)^2\right].
\end{aligned}$$

So the expected complete-data log likelihood is

$$\sum_{i=1}^{n-1} \log \phi(x_i, y_i) + \log \phi(\mu_0, y_n) - \frac{\sigma_0^2}{2(1-\rho^2)\sigma_X^2}.$$

The EM algorithm is similar to the previous one. First note that the MLEs of μ_Y and σ_Y^2 are the usual ones, $\bar{y}$ and $\hat{\sigma}_Y^2$, and don't change with the iterations. We update the other estimates as follows. At iteration t, the E-step consists of replacing $x_n^{(t)}$ by

$$x_n^{(t+1)} = \hat{\mu}_X^{(t)} + \rho^{(t)}\frac{\sigma_X^{(t)}}{\sigma_Y^{(t)}}(y_n - \bar{y}).$$

Then $\mu_X^{(t+1)} = \bar{x}$ and we can write the likelihood as

$$-\frac{1}{2}\log(2\pi\sigma_X^2\hat{\sigma}_Y^2(1-\rho^2)) - \frac{1}{2(1-\rho^2)}\left[\frac{S_{xx}+\sigma_0^2}{\sigma_X^2} - 2\rho\frac{S_{xy}}{\sigma_X\hat{\sigma}_Y} + \frac{S_{yy}}{\hat{\sigma}_Y^2}\right].$$

which is the usual bivariate normal likelihood except that we replace S_{xx} with $S_{xx} + \sigma_0^2$. So the MLEs are the usual ones, and the EM iterations are

$$
\begin{aligned}
x_n^{(t+1)} &= \hat{\mu}_X^{(t)} + \rho^{(t)} \frac{\sigma_X^{(t)}}{\sigma_Y^{(t)}}(y_n - \bar{y}) \\
\hat{\mu}_X^{(t+1)} &= \bar{x}^{(t)} \\
\hat{\sigma}_X^{2(t+1)} &= \frac{S_{xx}^{(t)} + (1-\hat{\rho}^{2(t)})\hat{\sigma}_X^{2(t)}}{n} \\
\hat{\rho}^{(t+1)} &= \frac{S_{xy}^{(t)}}{\sqrt{(S_{xx}^{(t)} + (1-\hat{\rho}^{2(t)})\hat{\sigma}_X^{2(t+1)})S_{yy}}}.
\end{aligned}
$$

Here is R code for the EM algorithm:

```
nsim<-20;
xdata0<-c(20,19.6,19.6,19.4,18.4,19,19,18.3,18.2,18.6,19.2,18.2,
18.7,18.5,18,17.4,16.5,17.2,17.3,17.8,17.3,18.4,16.9)
ydata0<-(1,1.2,1.1,1.4,2.3,1.7,1.7,2.4,2.1,2.1,1.2,2.3,1.9,2.4,2.6,
2.9,4,3.3,3,3.4,2.9,1.9,3.9,4.2)
nx<-length(xdata0);
ny<-length(ydata0);
#initial values from mles on the observed data#
xmean<-18.24167;xvar<-0.9597797;ymean<-2.370833;yvar<- 0.8312327;
rho<- -0.9700159;
for (j in 1:nsim) {
#This is the augmented x (O2) data#
xdata<-c(xdata0,xmean+rho*(4.2-ymean)/(sqrt(xvar*yvar)))
xmean<-mean(xdata);
Sxx<-(ny-1)*var(xdata)+(1-rho^2)*xvar
xvar<-Sxx/ny
rho<-cor(xdata,ydata0)*sqrt((ny-1)*var(xdata)/Sxx)
}
```

The algorithm converges very quickly. The MLEs are

$$\hat{\mu}_X = 18.24, \quad \hat{\mu}_Y = 2.37, \quad \hat{\sigma}_X^2 = .969, \quad \hat{\sigma}_Y^2 = .831, \quad \hat{\rho} = -0.969.$$

Chapter 12

Regression Models

12.1 The point $(\hat{x}', \hat{y}')$ is the closest if it lies on the vertex of the right triangle with vertices (x', y') and $(x', a + bx')$. By the Pythagorean theorem, we must have

$$\left[(\hat{x}'-x')^2+\left(\hat{y}'-(a+bx')\right)^2\right]+\left[(\hat{x}'-x')^2+(\hat{y}'-y')^2\right]=(x'-x')^2+(y'-(a+bx'))^2 .$$

Substituting the values of $\hat{x}'$ and $\hat{y}'$ from (12.2.7) we obtain for the LHS above

$$\left[\left(\frac{b(y'-bx'-a)}{1+b^2}\right)^2+\left(\frac{b^2(y'-bx'-a)}{1+b^2}\right)^2\right]+\left[\left(\frac{b(y'-bx'-a)}{1+b^2}\right)^2+\left(\frac{y'-bx-a}{1+b^2}\right)^2\right]$$
$$= (y'-(a+bx'))^2\left[\frac{b^2+b^4+b^2+1}{(1+b^2)^2}\right] = (y'-(a+bx'))^2 .$$

12.3 a. Differentiation yields $\partial f/\partial\xi_i = -2(x_i-\xi_i)-2\lambda\beta\left[y_i-(\alpha+\beta\xi_i)\right] \stackrel{\text{set}}{=} 0 \Rightarrow \xi_i(1+\lambda\beta^2) = x_i-\lambda\beta(y_i-\alpha)$, which is the required solution. Also, $\partial^2 f/\partial\xi^2 = 2(1+\lambda\beta^2) > 0$, so this is a minimum.

b. Parts i), ii), and iii) are immediate. For iv) just note that D is Euclidean distance between $(x_1, \sqrt{\lambda}y_1)$ and $(x_2, \sqrt{\lambda}y_2)$, hence satisfies the triangle inequality.

12.5 Differentiate $\log L$, for L in (12.2.17), to get

$$\frac{\partial}{\partial\sigma_\delta^2}\log L = \frac{-n}{\sigma_\delta^2}+\frac{1}{2(\sigma_\delta^2)^2}\frac{\lambda}{1+\hat{\beta}^2}\sum_{i=1}^n\left[y_i-(\hat{\alpha}+\hat{\beta}x_i)\right]^2 .$$

Set this equal to zero and solve for σ_δ^2. The answer is (12.2.18).

12.7 a. Suppressing the subscript i and the minus sign, the exponent is

$$\frac{(x-\xi)^2}{\sigma_\delta^2}+\frac{[y-(\alpha+\beta\xi)]^2}{\sigma_\epsilon^2} = \left(\frac{\sigma_\epsilon^2+\beta^2\sigma_\delta^2}{\sigma_\epsilon^2\sigma_\delta^2}\right)(\xi-k)^2+\frac{[y-(\alpha+\beta x)]^2}{\sigma_\epsilon^2+\beta^2\sigma_\delta^2},$$

where $k = \frac{\sigma_\epsilon^2 x+\sigma_\delta^2\beta(y-\alpha)}{\sigma_\epsilon^2+\beta^2\sigma_\delta^2}$. Thus, integrating with respect to ξ eliminates the first term.

b. The resulting function must be the joint pdf of X and Y. The double integral is infinite, however.

12.9 a. From the last two equations in (12.2.19),

$$\hat{\sigma}_\delta^2 = \frac{1}{n}S_{xx}-\hat{\sigma}_\xi^2 = \frac{1}{n}S_{xx}-\frac{1}{n}\frac{S_{xy}}{\hat{\beta}},$$

which is positive only if $S_{xx} > S_{xy}/\hat{\beta}$. Similarly,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}S_{yy}-\hat{\beta}^2\hat{\sigma}_\xi^2 = \frac{1}{n}S_{yy}-\hat{\beta}^2\frac{1}{n}\frac{S_{xy}}{\hat{\beta}},$$

which is positive only if $S_{yy} > \hat{\beta}S_{xy}$.

b. We have from part a), $\hat{\sigma}_\delta^2 > 0 \Rightarrow S_{xx} > S_{xy}/\hat{\beta}$ and $\hat{\sigma}_\epsilon^2 > 0 \Rightarrow S_{yy} > \hat{\beta} S_{xy}$. Furthermore, $\hat{\sigma}_\xi^2 > 0$ implies that S_{xy} and $\hat{\beta}$ have the same sign. Thus $S_{xx} > |S_{xy}|/|\hat{\beta}|$ and $S_{yy} > |\hat{\beta}||S_{xy}|$. Combining yields

$$\frac{|S_{xy}|}{S_{xx}} < \left|\hat{\beta}\right| < \frac{S_{yy}}{|S_{xy}|}.$$

12.11 a.

$$\begin{aligned}
&\mathrm{Cov}(aY{+}bX, cY{+}dX) \\
&\quad = \mathrm{E}(aY+bX)(cY+dX) - \mathrm{E}(aY+bX)E(cY+dX) \\
&\quad = \mathrm{E}\left(acY^2{+}(bc+ad)XY{+}bdX^2\right) - \mathrm{E}(aY+bX)E(cY+dX) \\
&\quad = ac\mathrm{Var}Y + ac(\mathrm{E}Y)^2 + (bc+ad)\mathrm{Cov}(X,Y) \\
&\qquad +(bc+ad)\mathrm{E}X\mathrm{E}Y + bd\mathrm{Var}X + bd(\mathrm{E}X)^2 - \mathrm{E}(aY+bX)\mathrm{E}(cY+dX) \\
&\quad = ac\mathrm{Var}Y + (bc+ad)\mathrm{Cov}(X,Y) + bd\mathrm{Var}X.
\end{aligned}$$

b. Identify $a = \beta\lambda$, $b = 1$, $c = 1$, $d = -\beta$, and using (12.3.19)

$$\begin{aligned}
\mathrm{Cov}(\beta\lambda Y_i{+}X_i, Y_i{-}\beta X_i) &= \beta\lambda\mathrm{Var}Y + (1-\lambda\beta^2)\mathrm{Cov}(X,Y) - \beta\mathrm{Var}X \\
&= \beta\lambda\left(\sigma_\epsilon^2 + \beta^2\sigma_\xi^2\right) + (1-\lambda\beta^2)\beta\sigma_\xi^2 - \beta\left(\sigma_\delta^2 + \sigma_\xi^2\right) \\
&= \beta\lambda\sigma_\epsilon^2 - \beta\sigma_\delta^2 \quad = \quad 0
\end{aligned}$$

if $\lambda\sigma_\epsilon^2 = \sigma_\delta^2$. (Note that we did not need the normality assumption, just the moments.)

c. Let $W_i = \beta\lambda Y_i + X_i$, $V_i = Y_i + \beta X_i$. Exercise 11.33 shows that if $\mathrm{Cov}(W_i, V_i) = 0$, then $\sqrt{n-2}r/\sqrt{1-r^2}$ has a t_{n-2} distribution. Thus $\sqrt{n-2}r_\lambda(\beta)/\sqrt{1-r_\lambda^2(\beta)}$ has a t_{n-2} distribution for all values of β, by part (b). Also

$$P\left(\left\{\beta\colon \frac{(n-2)r^2_{\lambda(\beta)}}{1-r^2_{\lambda(\beta)}} \le F_{1,n-2,\alpha}\right\}\right) = P\left(\left\{(X,Y)\colon \frac{(n-2)r_\lambda^2(\beta)}{1-r_\lambda^2(\beta)} \le F_{1,n-2,\alpha}\right\}\right) = 1-\alpha.$$

12.13 a. Rewrite (12.2.22) to get

$$\left\{\beta\colon \hat{\beta} - \frac{t\hat{\sigma}_\beta}{\sqrt{n-2}} \le \beta \le \hat{\beta} + \frac{t\hat{\sigma}_\beta}{\sqrt{n-2}}\right\} = \left\{\beta\colon \frac{(\hat{\beta}{-}\beta)^2}{\sigma_\beta^2\big/(n-2)} \le F\right\}.$$

b. For $\hat{\beta}$ of (12.2.16), the numerator of $r_\lambda(\beta)$ in (12.2.22) can be written

$$\beta\lambda S_{yy}{+}(1{-}\beta^2\lambda)S_{xy}{-}\beta S_{xy} = \beta^2(\lambda S_{xy}) + \beta(S_{xx} - \lambda S_{yy}) + S_{xy} = \lambda S_{xy}(\beta - \hat{\beta})\left(\beta{+}\frac{1}{\lambda\hat{\beta}}\right).$$

Again from (12.2.22), we have

$$\begin{aligned}
&\frac{r_\lambda^2(\beta)}{1-r_\lambda^2(\beta)} \\
&\quad = \frac{\left(\beta\lambda S_{yy}{+}(1{-}\beta^2\lambda)S_{xy}{-}\beta S_{xy}\right)^2}{\left(\beta^2\lambda^2 S_{yy}{+}2\beta\lambda S_{xy}{+}S_{xx}\right)\left(S_{yy}{-}2\beta S_{xy}{+}\beta^2 S_{xx}\right) - \left(\beta\lambda S_{yy}{+}(1{-}\beta^2\lambda)S_{xy}{-}\beta S_{xx}\right)^2},
\end{aligned}$$

and a great deal of straightforward (but tedious) algebra will show that the denominator of this expression is equal to

$$(1+\lambda\beta^2)^2\left(S_{yy}S_{xx} - S_{xy}^2\right).$$

Thus

$$\begin{aligned}\frac{r_\lambda^2(\beta)}{1-r_\lambda^2(\beta)} &= y\frac{\lambda^2 S_{xy}^2\left(\beta-\hat{\beta}\right)^2\left(\beta+\frac{1}{\lambda\hat{\beta}}\right)^2}{(1-\lambda\beta^2)^2\left(S_{yy}S_x-S_{xy}^2\right)}\\ &= \frac{\left(\beta-\hat{\beta}\right)^2}{\hat{\sigma}_\beta^2}\left(\frac{1+\lambda\beta\hat{\beta}}{1+\lambda\beta^2}\right)^2\frac{(1+\lambda\hat{\beta}^2)^2S_{xy}^2}{\hat{\beta}^2\left[\left(S_{xx}-\lambda S_{yy}\right)^2+4\lambda S_{xy}^2\right]},\end{aligned}$$

after substituting $\hat{\sigma}_\beta^2$ from page 588. Now using the fact that $\hat{\beta}$ and $-1/\lambda\hat{\beta}$ are both roots of the same quadratic equation, we have

$$\frac{(1+\lambda\hat{\beta}^2)^2}{\hat{\beta}^2}=\left(\frac{1}{\hat{\beta}}+\lambda\hat{\beta}\right)^2=\frac{\left(S_{xx}-\lambda S_{yy}\right)^2+4\lambda S_{xy}^2}{S_{xy}^2}.$$

Thus the expression in square brackets is equal to 1.

12.15 a.

$$\pi(-\alpha/\beta)=\frac{e^{\alpha+\beta(-\alpha/\beta)}}{1+e^{\alpha+\beta(-\alpha/\beta)}}=\frac{e^0}{1+e^0}=\frac{1}{2}.$$

b.

$$\pi((-\alpha/\beta)+c)=\frac{e^{\alpha+\beta((-\alpha/\beta)+c)}}{1+e^{\alpha+\beta((-\alpha/\beta)+c)}}=\frac{e^{\beta c}}{1+e^{\beta c}},$$

and

$$1-\pi((-\alpha/\beta)-c)=1-\frac{e^{-\beta c}}{1+e^{-\beta c}}=\frac{e^{\beta c}}{1+e^{\beta c}}.$$

c.

$$\frac{d}{dx}\pi(x)=\beta\frac{e^{\alpha+\beta x}}{[1+e^{\alpha+\beta x}]^2}=\beta\pi(x)(1-\pi(x)).$$

d. Because

$$\frac{\pi(x)}{1-\pi(x)}=e^{\alpha+\beta x},$$

the result follows from direct substitution.

e. Follows directly from (d).

f. Follows directly from

$$\frac{\partial}{\partial\alpha}F(\alpha+\beta x)=f(\alpha+\beta x)\text{ and }\frac{\partial}{\partial\beta}F(\alpha+\beta x)=xf(\alpha+\beta x).$$

g. For $F(x)=e^x/(1+e^x)$, $f(x)=F(x)(1-F(x))$ and the result follows. For $F(x)=\pi(x)$ of (12.3.2), from part (c) if follows that $\frac{f}{F(1-F)}=\beta$.

12.17 a. The likelihood equations and solution are the same as in Example 12.3.1 with the exception that here $\pi(x_j)=\Phi(\alpha+\beta x_j)$, where Φ is the cdf of a standard normal.

b. If the $0-1$ failure response in denoted "oring" and the temperature data is "temp", the following R code will generate the logit and probit regression:

```
summary(glm(oring~temp, family=binomial(link=logit)))
summary(glm(oring~temp, family=binomial(link=probit)))
```

For the logit model we have

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	15.0429	7.3719	2.041	0.0413
temp	-0.2322	0.1081	-2.147	0.0318

and for the probit model we have

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	8.77084	3.86222	2.271	0.0232
temp	-0.13504	0.05632	-2.398	0.0165

Although the coefficients are different, the fit is qualitatively the same, and the probability of failure at 31°, using the probit model, is .9999.

12.19 a. Using the notation of Example 12.3.1, the likelihood (joint density) is

$$\prod_{j=1}^{J}\left[\frac{e^{\alpha+\beta x_j}}{1+e^{\alpha+\beta x_j}}\right]^{y_j^*}\left[\frac{1}{1+e^{\alpha+\beta x_j}}\right]^{n_j-y_j^*}=\prod_{j=1}^{J}\left[\frac{1}{1+e^{\alpha+\beta x_j}}\right]^{n_j} e^{\alpha\sum_j y_j^*+\beta\sum_j x_j y_j^*}.$$

By the Factorization Theorem, $\sum_j y_j^*$ and $\sum_j x_j y_j^*$ are sufficient.

b. Straightforward substitution.

12.21 Since $\frac{d}{d\pi}\log(\pi/(1-\pi)) = 1/(\pi(1-\pi))$,

$$\operatorname{Var}\log\left(\frac{\hat\pi}{1-\hat\pi}\right)\approx\left(\frac{1}{\pi(1-\pi)}\right)^2\frac{\pi(1-\pi)}{n}=\frac{1}{n\pi(1-\pi)}$$

12.23 a. If $\sum a_i = 0$,

$$\mathrm{E}\sum_i a_iY_i=\sum_i a_i[\alpha+\beta\boldsymbol{x}_i+\mu(1-\delta)]=\beta\sum_i a_ix_i=\beta$$

for $a_i = x_i - \bar{x}$.

b.

$$\mathrm{E}(\bar{Y}-\beta\bar{x})=\frac{1}{n}\sum_i[\alpha+\beta x_i+\mu(1-\delta)]-\beta\bar{x}=\alpha+\mu(1-\delta),$$

so the least squares estimate a is unbiased in the model $Y_i = \alpha' + \beta x_i + \epsilon_i$, where $\alpha' = \alpha + \mu(1-\delta)$.

12.25 a. The least absolute deviation line minimizes

$$|y_1-(c+dx_1)|+|y_2-(c+dx_1)|+|y_3-(c+dx_3)|.$$

Any line that lies between (x_1, y_1) and (x_1, y_2) has the same value for the sum of the first two terms, and this value is smaller than that of any line outside of (x_1, y_1) and (x_2, y_2). Of all the lines that lie inside, the ones that go through (x_3, y_3) minimize the entire sum.

b. For the least squares line, $a = -53.88$ and $b = .53$. Any line with b between $(17.9-14.4)/9 = .39$ and $(17.9-11.9)/9 = .67$ and $a = 17.9 - 136b$ is a least absolute deviation line.

12.27 In the terminology of M-estimators (see the argument on pages 485 – 486), $\hat\beta_L$ is consistent for the β_0 that satisfies $\mathrm{E}_{\beta_0}\sum_i \psi(Y_i - \beta_0 x_i) = 0$, so we must take the "true" β to be this value. We then see that

$$\sum_i \psi(Y_i-\hat\beta_L x_i)\to 0$$

as long as the derivative term is bounded, which we assume is so.

12.29 The argument for the median is a special case of Example 12.4.3, where we take $x_i = 1$ so $\sigma_x^2 = 1$. The asymptotic distribution is given in (12.4.5) which, for $\sigma_x^2 = 1$, agrees with Example 10.2.3.

12.31 The LAD estimates, from Example 12.4.2 are $\tilde{\alpha} = 18.59$ and $\tilde{\beta} = -.89$. Here is Mathematica code to bootstrap the standard deviations. (Mathematica is probably not the best choice here, as it is somewhat slow. Also, the minimization seemed a bit delicate, and worked better when done iteratively.) Sad is the sum of the absolute deviations, which is minimized iteratively in bmin and amin. The residuals are bootstrapped by generating random indices u from the discrete uniform distribution on the integers 1 to 23.

1. First enter data and initialize

```
Needs["Statistics`Master`"]
Clear[a,b,r,u]
a0=18.59;b0=-.89;aboot=a0;bboot=b0;
y0={1,1.2,1.1,1.4,2.3,1.7,1.7,2.4,2.1,2.1,1.2,2.3,1.9,2.4,
        2.6,2.9,4,3.3,3,3.4,2.9,1.9,3.9};
x0={20,19.6,19.6,19.4,18.4,19,19,18.3,18.2,18.6,19.2,18.2,
        18.7,18.5,18,17.4,16.5,17.2,17.3,17.8,17.3,18.4,16.9};
model=a0+b0*x0;
r=y0-model;
u:=Random[DiscreteUniformDistribution[23]]
Sad[a_,b_]:=Mean[Abs[model+rstar-(a+b*x0)]]
bmin[a_]:=FindMinimum[Sad[a,b],{b,{.5,1.5}}]
amin:=FindMinimum[Sad[a,b/.bmin[a][[2]]],{a,{16,19}}]
```

2. Here is the actual bootstrap. The vectors *aboot* and *bboot* contain the bootstrapped values.

```
B=500;
Do[
  rstar=Table[r[[u]],{i,1,23}];
  astar=a/.amin[[2]];
  bstar=b/.bmin[astar][[2]];
  aboot=Flatten[{aboot,astar}];
  bboot=Flatten[{bboot,bstar}],
  {i,1,B}]
```

3. Summary Statistics

```
Mean[aboot]
StandardDeviation[aboot]
Mean[bboot]
StandardDeviation[bboot]
```

4. The results are Intercept: Mean 18.66, SD .923 Slope: Mean $-.893$, SD .050.