Appendix: Selected Answers

Chapter 1

Section 1.2

1.2.1 (a) 5/6. (b) 1. (c) 1/2.

1.2.4 No.

1.2.8 No.

1.2.10 Yes.

Section 1.3

1.3.1 (a) 0.9. (b) 0.1 is the smallest possible value of $P(\{1\})$.

1.3.4 The maximum is 25% and the minimum is 15%.

Section 1.4

1.4.1 (a) 1/1679616. (b) 1/279936. (c) 1/209952.

1.4.6 $P(\text{sum } \geq 4) = 1 - 1/221 - 8/663 = 652/663$.

1.4.10 5/12.

Section 1.5

1.5.1 (a) 0.75. (b) The conditional probability equals 0.762.

1.5.4 $P(\text{five Spades} \,|\, \text{at least 4 Spades}) \doteq 0.044$.

1.5.9 (a) A and B are not independent. (b) A and C are independent. (c) A and D are independent. (d) C and D are independent. (e) A and C and D are not all independent.

1.5.10 We have from the Exercise 1.4.11 solution that $P(\text{all red}) = 5/4488$, while $P(\text{all blue}) = 35/816$. Hence, $P(\text{all red} \,|\, \text{all same color}) = P(\text{all red}) \,/\, P(\text{all same color}) = (5/4488)/[(5/4488) + (35/816)] = 2/79 \doteq 0.025$.

Section 1.6

1.6.2 $\lim_{n\to\infty} P([1/4,\ 1 - e^{-n}]) = 3/4$.

1.6.3 $\lim_{n\to\infty} P(A_n) = P(A) = P(S) = 1$.

Chapter 2

Section 2.1

2.1.1(a) $\min_{s\in S} X(s) = X(1) = 1$. (b) $\max_{s\in S} X(s)$ does not exist. (c) $\min_{s\in S} Y(s)$ does not exist. (d) $\max_{s\in S} Y(s) = Y(1) = 1$.

2.1.4 $Z(1) = 3$, $Z(2) = 20$, $Z(3) = 87$, $Z(4) = 1032$, $Z(5) = 15,635$, $Z(6) = 279,948$.

2.1.7 No.

Section 2.2

2.2.1 X must equal 0, 1, or 2, and $P(X = 0) = 1/4$, $P(X = 1) = 1/2$, $P(X = x) = 0$ for $x \neq 0, 1, 2$.

2.2.5 (a) $P(X = 1) = .3, P(X = 2) = .2, P(X = 3) = .5$, and $P(X = x) = 0$ for all $x \notin \{1, 2, 3\}$. (b) $P(Y = 1) = .3, P(Y = 2) = .2, P(Y = 3) = .5$, and $P(Y = y) = 0$ for all $y \notin \{1, 2, 3\}$. (c) $P(W = 2) = 0.09$, $P(W = 3) = 0.12$, $P(W = 4) = 0.34$, $P(W = 5) = 0.2$, $P(W = 6) = 0.25$, and $P(W = w) = 0$ for all other choices of w.

2.2.8 Note that each number $w \in \{0, 1, \ldots, 99\}$ can occur and $P(W = w) = 1/100$.

Section 2.3

2.3.1 $p_Y(2) = 1/36$, $p_Y(3) = 2/36$, $p_Y(4) = 3/36$, $p_Y(5) = 4/36$, $p_Y(6) = 5/36$, $p_Y(7) = 6/36$, $p_Y(8) = 5/36$, $p_Y(9) = 4/36$, $p_Y(10) = 3/36$, $p_Y(11) = 2/36$, $p_Y(12) = 1/36$, and $p_Y(y) = 0$ otherwise.

2.3.5 $p_W(1) = 1/36$, $p_W(2) = 2/36$, $p_W(3) = 2/36$, $p_W(4) = 2/36 + 1/36 = 3/36$, $p_W(5) = 2/36$, $p_W(6) = 2/36 + 2/36 = 4/36$, $p_W(8) = 2/36$, $p_W(9) = 1/36$, $p_W(10) = 2/36$, $p_W(12) = 2/36 + 2/36 = 4/36$, $p_W(15) = 2/36$, $p_W(16) = 1/36$, $p_W(18) = 2/36$, $p_W(20) = 2/36$, $p_W(24) = 2/36$, $p_W(25) = 1/36$, $p_W(30) = 2/36$, $p_W(36) = 1/36$, and $p_W(w) = 0$ otherwise.

2.3.10 $P(X^2 \leq 15) = 369/625$.

2.3.15 (a) $\binom{10}{3}(.35)^3(.65)^7$. (b) $(.35)(.65)^9$. (c) $\binom{9}{1}(.35)^2(.65)^8$.

2.3.24 $Z \sim \text{Binomial}(n_1 + n_2, p)$.

2.3.28 $Z \sim \text{Negative Binomial}(r + s, \theta)$.

Section 2.4

2.4.1 (a) $P(U \leq 0) = 0$. (b) $P(U = 1/2) = 0$. (c) $P(U < -1/3) = 0$.
(d) $P(U \leq 2/3) = 2/3$. (f) $P(U < 1) = 1$. (g) $P(U \leq 17) = 1$.

2.4.4 (a) $c = 2$. (b) $c = n + 1$. (c) $c = 3/\left(\sqrt{2}4\right)$. (d) $c = 1$.

2.4.11 We have that $f(x) \geq 0$ for every x, and putting $u = x^\alpha, du = \alpha x^{\alpha-1}\,dx$, we have $\int_0^\infty \alpha x^{\alpha-1} e^{-x^\alpha}\,dx = \int_0^\infty e^{-u}\,du = -e^{-u}\big|_0^\infty = 1$.

Section 2.5

2.5.2 $F_X(x) = 0$ when $x < 1, F_X(x) = 1/6$ when $1 \leq x < 4, F_X(x) = 2/6$ when $4 \leq x < 9, F_X(x) = 3/6$ when $9 \leq x < 16, F_X(x) = 4/6$ when $16 \leq x < 25, F_X(x) = 5/6$ when $25 \leq x < 36$, and$F_X(x) = 1$ when $36 \leq x$.

2.5.3(a) No. (c) Yes. (g) No.

2.5.7 $\lim_{n\to\infty} |F(2n) - F(n)| = 0$.

2.5.11 $F(x) = (1 + e^{-x})^{-1}$.

2.5.15 $F(x) = \frac{1}{2}e^{x}$ when $x \leq 0$ and $F(x) = \frac{1}{2} + \frac{1}{2}(1 - e^{-x})$ when $x > 0$.

Section 2.6

2.6.3 Let $h(x) = cx + d$. Then $Y = h(X)$ and h is strictly increasing, so $f_Y(y) = f_X(h^{-1}(y)) \,/\, |h'(h^{-1}(y))| = e^{-[y-d-c\mu]^2/2c^2\sigma^2}/c\sigma\sqrt{2\pi}$.

2.6.13 $Y \sim$ Weibull(α/β).

2.6.15 $Y \sim$ Exponential(1).

Section 2.7

2.7.1 $F_{X,Y}(x, y) = 0$ when $\min[x,\ (y + 2)/4] < 0, F_{X,Y}(x, y) = 1/3$ when $0 \leq \min[x,\ (y + 2)/4] < 1$, and $F_{X,Y}(x, y) = 1$ when $\min[x,\ (y + 2)/4] \geq 1$.

2.7.3 (a) $p_X(2) = p_X(3) = p_X(-3) = p_X(-2) = p_X(17) = 1/5$, with $p_X(x) = 0$ otherwise (d) $P(Y = X) = 0$ since this never occurs.

2.7.4 (a) $C = 4$, and $P(X \leq 0.8,\ Y \leq 0.6) \doteq 0.0863$.

2.7.6 $F_{X,Y}(x, y) = 1 - e^{-\lambda \min(x,\ y^{1/3})}$ for $x, y > 0$, otherwise equals 0.

2.7.11 (a) $C = 2$. (b) We have that $f_X(x) = 2e^{-2x}$so $X \sim$ Exponential(2) and $f_Y(y) = 2e^{-y}(1 - e^{-y})$ for $y > 0$.

Section 2.8

2.8.1 (a) $p_X(-2) = 1/4$, $p_X(9) = 1/4$, $p_X(13) = 1/2$ otherwise $p_X(x) = 0$.
(b) $p_Y(3) = 2/3$, $p_Y(5) = 1/3$, otherwise $p_Y(y) = 0$. (c) Yes.

2.8.5 (a) $P(Y = 4 \,|\, X = 9) = 1/6$. (c) $P(Y = 0 \,|\, X = -4) = 0$.

2.8.7 (a) $C = 4$ and X and Y are not independent since $f_{Y|X}(y \,|\, x) \neq f_Y(y)$.

2.8.12 Since X and Y are independent, $P(X = 1 \,|\, Y = 5) = P(X = 1) = 1/3$.

2.8.19 $P(X_2 = f_2 \,|\, X_1 = f_1) = \binom{n-f_1}{f_2}\left(\frac{\theta_2}{1-\theta_1}\right)^{f_2}\left(1 - \frac{\theta_2}{1-\theta_1}\right)^{n-f_1-f_2}$, so $X_2 \,|\, X_1 = f_1 \sim$ Binomial$(n - f_1, \theta_2/(1 - \theta_1))$.

2.8.22 The distribution function of $X_{(3)}$, for $0 < x < 1$, is given by $P\left(X_{(3)} \leq x\right)$ $= 10x^3(1 - x)^2 + 5x^4(1 - x) + x^5 = 10x^3 - 15x^4 + 6x^5$, so $f(x) = 30x^2 - 60x^3 + 30x^4 = 30x^2(x - 1)^2$. This is the Beta(3,3) density.

Section 2.9

2.9.2 (a) $f_{X,Y}(x, y) = e^{-x}$ for $x \geq 0$ and $1 \leq y \leq 4$, otherwise $f_{X,Y}(x, y) = 0$.
(b) $h(x, y) = (x + y,\ x - y)$. (c) $h^{-1}(z, w) = ((z + w)/2,\ (z - w)/2)$.
(d) Here $J(x, y) = \frac{\partial h_1}{\partial x}\frac{\partial h_2}{\partial y} - \frac{\partial h_2}{\partial x}\frac{\partial h_1}{\partial y} = |(1)(-1) - (1)(1)| = 2$, so $f_{Z,W}(z, w) = f_{X,Y}(h^{-1}(z, w)) \,/\, |J(h^{-1}(z, w))| = f_{X,Y}((z + w)/2,\ (z - w)/2) \,/\, 2$, which equals

$e^{-(z+w)/2}$ for $(z+w)/2 \geq 0$ and $1 \leq (z-w)/2 \leq 4$, i.e., for $z \geq 1$ and $\max(-z, z-8) \leq w \leq z-2$, otherwise $f_{Z,W}(z,w) = 0$.

2.9.6 (a) $p_{Z,W}(5,5) = 1/7$, $p_{Z,W}(8,2) = 1/7$, $p_{Z,W}(9,1) = 1/7$, $p_{Z,W}(8,0) = 3/7$, $p_{Z,W}(12,4) = 1/7$, and $p_{Z,W}(z,w) = 0$ otherwise.

Section 2.10

2.10.2 (a) $F^{-1}(t) = t$, so $X = U$ (b) $F^{-1}(t) = \sqrt{t}$, so $X = \sqrt{U}$.

2.10.6 $x = F^{-1}(u) = \ln(u/(1-u))$ for $0 \leq u \leq 1$.

2.10.10 $x = F^{-1}(u) = \ln(2u)$ and, for $1/2 \leq u \leq 1$, $x = F^{-1}(u) = -\ln 2(1-u)$.

Chapter 3

Section 3.1

3.1.1 (a) $E(X) = 8/7$. (b) $E(X) = 1$. (c) $E(X) = 8$.

3.1.3 (a) $E(X) \doteq -14.4$. (d)$E(Y^2) \doteq 123.3$.

3.1.7 $E(XY) = 30$.

3.1.9 $E(X) = 6$.

3.1.12 (b) $E(Y - X) = E(Y) - E(X) = -(1-\theta)^{101}(1/\theta + 100)$.

Section 3.2

3.2.1 (a) $C = 1/4$, $E(X) = 7$. (b) $C = 1/16$, $E(X) \doteq 7.04$. (c) $C = 5/3093$ $E(X) \doteq -4.19$.

3.2.4 (a) $E(X) = 57/70$. (b) $E(Y) = 157/280$.

3.2.8 $E(Y+Z) = 49/36$.

3.2.15 $E(X) = \int_0^1 x \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1}\,dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}$.

Section 3.3

3.3.1 (a) $\text{Cov}(X,Y) = 2/3$. (b) $\text{Var}(X) = 2$, $\text{Var}(Y) = 32/9$. (c) $\text{Corr}(X,Y) = 1/4$.

3.3.4 $E(X) = 63/80$, $E(Y) = 61/72$, $\text{Var}(X) = 191/6400$,
$\text{Var}(Y) = 3253/181440$, $\text{Cov}(X,Y) = -1/1920$.

3.3.10 $\text{Corr}(X,Y) = sgn(c)$, where $sgn(c) = 1$ for $c > 0$, $sgn(c) = 0$ for $c = 0$, and $sgn(c) = -1$ for $c < 0$, (a) $\lim_{c \searrow 0} \text{Cov}(X,Y) = \lim_{c \searrow 0} c = 0$, (b)$\lim_{c \nearrow 0} \text{Cov}(X,Y) = \lim_{c \nearrow 0} c = 0$.

3.2.17 $\text{Var}(X) = 2$.

Section 3.4

3.4.1 (a) $r_Z(t) = t/(2-t)$ (b) $E(Z) = 2$, $E(Z^2) = 6$.

3.4.4 $r_Y(t) = t^4 r_X(t^3)$.

3.4.11 $r_X(t) = \theta^r (1 - t(1-\theta))^{-r}$ provided $|t(1-\theta)| < 1$.

3.4.15 Write $Z = \mu + \sigma X$, where $X \sim \text{Normal}(0,1)$. Then $m_Z(s) = E(e^{sZ}) = E(e^{s(\mu+\sigma X)}) = e^{s\mu} + E(e^{s\sigma X}) = e^{s\mu} + m_X(\sigma s) = e^{s\mu} + e^{(\sigma s)^2/2} = e^{s\mu} + e^{\sigma^2 s^2/2}$.

3.4.25 $m_{S_N}(s) = r_N(m_{X_1}(s)) = \theta / \left(1 - \frac{\lambda}{\lambda - s}(1-\theta)\right), m'_{S_N}(0) = (1-\theta)/(\lambda\theta).$

Section 3.5

3.5.1 (a) $E(X \mid Y = 3) = 5/2$. (b) $E(Y \mid X = 3) = 22/3$. (c) $E(X \mid Y = 2) = 5/2$, $E(X \mid Y = 17) = 3$. (d) $E(Y \mid X = 2) = 5/2$.

3.5.5 $E(\text{earnings} \mid Y = \text{“takes course”}) = \$2700$, $E(X \mid Y = \text{“doesn't take course”}) = \$2100$, $E(\text{earnings}) = \$2340$

3.5.14 (a) $\text{Var}(X) \doteq 0.244967$. (b) $\text{Var}(E(X \mid Y)) \doteq 0.0000002196$. (c) $\text{Var}(X \mid Y) = (8/49)(49 + 8064Y^4 + 98304Y^8)/(3 + 256Y^4)^2$. (d) $E(\text{Var}(X \mid Y)) \doteq 0.244967$.

Section 3.6

3.6.1 $P(Z \geq 7) \leq 3/7$.

3.6.6 Largest value is $\text{Cov}(Y, Z) \doteq 38.73$ and smallest is $\text{Cov}(Y, Z) \doteq -38.73$. Therefore, r_{XY} is as stated.

3.6.11 $P(X \notin (\bar{x} - 2\hat{s}, \bar{x} + 2\hat{s})) \leq \frac{\hat{s}_X^2}{(2\hat{s}_X)^2} = \frac{1}{4}$**,** so the largest possible proportion is 1/4.

Section 3.7

3.7.1 $E(X_1) = 3$, $E(X_2) = 0$, and $E(Y) = (1/5)E(X_1) + (4/5)E(X_2) = 3/5$.

Chapter 4

Section 4.1

4.1.2 If Z is the sample mean, then $P(Z = 1) = 1/36$, $P(Z = 1.5) = 2/36$, $P(Z = 2) = 3/36$, $P(Z = 2.5) = 4/36$, $P(Z = 3) = 5/36$, $P(Z = 3.5) = 6/36$, $P(Z = 4) = 5/36$, $P(Z = 4.5) = 4/36$, $P(Z = 5) = 3/36$, $P(Z = 5.5) = 2/36$, and $P(Z = 6) = 1/36$.

4.1.4 If Z is the sample mean, then $P(Z = 0) = \frac{N}{N+M}\frac{N-1}{N+M-1}$, $P(Z = 0.5) = 2\frac{N}{N+M}\frac{M}{N+M-1}$, and $P(Z = 1) = \frac{M}{N+M}\frac{M-1}{N+M-1}$.

4.1.8 $Y \sim \text{Poisson}(n\lambda)$.

Section 4.2

4.2.2 For any $\epsilon > 0$, $P(|X_n - 0| \geq \epsilon) = P(Y^n \geq \epsilon) = P(Y \geq \epsilon^{1/n}) = 1 - \epsilon^{1/n} \to 0$ as $n \to \infty$, so $X_n \to 0$ in probability.

Section 4.3

4.3.1 Note that $Z_n = Z$ unless $7 \leq U < 7 + 1/n^2$. Hence, if $U < 7$, then $Z_n = Z$ for all n, so of course $Z_n \to Z$. Also, if $U > 7$, then $Z_n = Z$ whenever $1/n^2 < 7 - U$, i.e., $n > 1/\sqrt{7 - U}$, so again $Z_n \to Z$. Hence, $P(Z_n \to Z) \geq P(U \neq 7) = 1 - P(U = 7) = 1 - 0 = 1$, i.e., $Z_n \to Z$ with probability 1.

4.3.5 $P(X_n \to X \text{ and } Y_n \to Y) \geq 1 - P(X_n \not\to X) - P(Y_n \not\to Y)$.

4.3.12 This is false.

Section 4.4

4.4.1 Here $\lim_{n\to\infty} P(X_n = i) = 1/3 = P(X = i)$ for $i = 1, 2, 3$, so $\lim_{n\to\infty} P(X_n \le x) = P(X \le x)$ for all x, so $X_n \to X$ in distribution.

4.4.5$P(S \le 540) \approx 0.6915$.

4.4.16 $P(M_n \le m) = P(\frac{\sqrt{n}(M_n-0)}{\sqrt{15}} \le \frac{\sqrt{n}(m-0)}{\sqrt{15}}) \approx P(Z \le \frac{\sqrt{n}(m-0)}{\sqrt{15}})$, where $Z \sim N(0,1)$.

Section 4.5

4.5.1 This integral equals $\sqrt{2\pi}E(\cos^2(Z))$, where $Z \sim N(0,1)$.

4.5.4 $n \ge 9\,(10)\,/(.1)^2 = 9000.0$.

4.5.11 $n \ge 9/\left(4\delta^2\right)$.

Section 4.6

4.6.1 (a) $U \sim N(44,\, 629)$. $V \sim N(-18 - 8C,\, 144 + 25C^2)$. (b) U and V are independent if and only if $C = -24/125$.

4.6.3 $C_1 = 1/\sqrt{5}$. $C_2 = -3$. $C_3 = 1/\sqrt{2}$. $C_4 = 7$. $C_5 = 2$.

4.6.11 We see that $f_Z(-z) = \Gamma((n+1)/2)(1+(-z)^2/n)^{-(n+1)/2}/\Gamma(n/2)\sqrt{\pi n} = \Gamma((n+1)/2)(1+z^2/n)^{-(n+1)/2}/\Gamma(n/2)\sqrt{\pi n} = f_Z(z)$. Then using the substitution $s = -t$, we have $P(Z < -x) = \int_{-\infty}^{-x} f_Z(t)\,dt = -\int_x^{\infty} f_Z(-s)\,(-ds) = \int_x^{\infty} f_Z(s)\,ds = P(Z > x)$.

Chapter 5

Section 5.1

5.1.1 The mean survival times for the control group and the treatment group are 93.2 days and 356.2 days, respectively. As we can see, there is a big difference between the two means, which might suggest that the treatment is indeed effective, but we can't base our conclusions about the effectiveness of the treatment based only on these numbers. We have to consider sampling variability as well.

5.1.3 For those who are still alive their survival times will be longer than the recorded values, so these data values are incomplete.

Section 5.2

5.2.1 In Example 5.2.1 the mode is given by 0. In Example 5.2.2 the mode of this density is 1. In both cases the mode is at the extreme left end of the distribution and so doesn't seem like a very good predictor.

5.2.3 The density is given by $.5\,(2\pi)^{-1/2}\exp\{-\left(x+4\right)^2/2\}+$ $.5\,(2\pi)^{-1/2}\exp\{-\left(x-4\right)^2/2\}$ for $-\infty < x < \infty$.

5.2.8 Suppose that $X \sim \text{Beta}(a,b)$. We have that $E\left(X\right) = a/\left(a+b\right)$ with $E((X - a/\left(a+b\right))^2) = ab/\left(a+b+1\right)\left(a+b\right)^2$. The mode is given by $(a-1)\,/(a+b-2)$ and $E((X-(a-1)\,/\,(a+b-2))^2) = ab/\left(a+b+1\right)\left(a+b\right)^2 + (a/(a+b) - (a-1)/(a+b-2))^2 \ge E((X - a/\left(a+b\right))^2)$. Therefore, the mean is a better predictor.

Section 5.3

5.3.1 The statistical model for a single response consists of three probability functions $\{f_1, f_2, f_3\}$, where f_1 is the probability function for the Bernoulli(1/2) distribution, f_1 is the probability function for the Bernoulli(1/3) distribution, and f_3 is the probability function for the Bernoulli(2/3) distribution. Then $(x_1, x_2, \ldots, x_5)$ is a sample from one of these Bernoulli(θ) distributions.

5.3.6 The first quartile of the Uniform$[0, \beta]$ distribution is $c = 0.25\beta$. Since c is a 1-1 transformation of β, we can parameterize this model by the first quartile.

5.3.9 $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^n ((x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - \mu) + (\mu - \bar{x})^2) = \sum_{i=1}^n (x_i - \bar{x})^2 - 2(\mu - \bar{x}) \sum_{i=1}^n (x_i - \bar{x}) + \sum_{i=1}^n (\mu - \bar{x})^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\mu - \bar{x})^2$ since $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = 0$.

5.3.11 The first quartile of a $N(\mu, \sigma^2)$ distribution is $c = \mu + \sigma z_{.25}$, where $z_{.25}$ is the first quartile of the $N(0, 1)$ distribution, i.e., $\Phi(z_{.25}) = .25$. But we see from this that several different values of (μ, σ^2) can give the same first quartile, e.g., $(\mu, \sigma^2) = (0, 1)$ and $(\mu, \sigma^2) = (z_{.25}/2, 1/4)$ both give rise to normal distributions whose first quartile equals $z_{.25}$. Therefore, we cannot parameterize this model by the first quartile.

Section 5.4

5.4.1 $F_X(x) = 0$ when $x < 1$, $F_X(x) = 4/10$ when $1 \le x < 2$, $F_X(x) = 7/10$ when $2 \le x < 3$, $F_X(x) = 9/10$ when $3 \le x < 4$, $F_X(x) = 1$ when $4 \le x$, $f_X(1) = 4/10, f_X(2) = 3/10, f_X(3) = 2/10, f_X(4) = 1/10$, and $\mu_X = 2, \sigma_X^2 = 1$.

5.4.4 (a) $f_X(0) = a/N, f_X(1) = (N - a)/N$. This is a Bernoulli($(N - a)/N$) distribution. (b) $P\left(\hat{f}_X(0) = f_X(0)\right) = P\left(n\hat{f}_X(0) = nf_X(0)\right)$ $= \binom{a}{n - nf_X(0)}\binom{N-a}{nf_X(0)}/\binom{N}{n}$ since $n\hat{f}_X(0) \sim$ Hypergeometric(N, a, n). (c) We have that $n\hat{f}_X(0) \sim$ Binomial$(n, a/N)$ so $P\left(\hat{f}_X(0) = f_X(0)\right)$ $= P\left(n\hat{f}_X(0) = nf_X(0)\right) = P$(number of 0's in the sample equals $nf_X(0)$) $= \binom{n}{nf_X(0)} (a/N)^{nf_X(0)} (1 - a/N)^{n - nf_X(0)}$.

5.4.9 (a) $f_X(0) = a/N, f_X(1) = b/N, f_X(2) = (N - a - b)/N$. (b) Assuming f_1, f_2, f_3 are nonnegative integers summing to n (otherwise probability is 0), the probability is $\binom{a}{f_0}\binom{b}{f_1}\binom{N-a-b}{f_2}/\binom{N}{n}$. (c) The probability that $\hat{f}_X(0) = f_0, \hat{f}_X(1) = f_1$ and $\hat{f}_X(2) = f_2$ is $\binom{a}{f_0\ f_1\ f_2} (a/N)^{f_0} (b/N)^{f_1} ((N - a - b)/N)^{f_2}$.

5.4.12 When $f_X(0) = a/N$ is unknown we estimate it by $\hat{f}_X(0)$. Now $N = a/f_X(0)$, so we can estimate N by setting $\hat{N} = a/\hat{f}_X(0)$, provided $\hat{f}_X(0) \neq 0$.

Section 5.5

5.5.1 (a) $\hat{f}_X(0) = .2667, \hat{f}_X(1) = .2, \hat{f}_X(2) = .2667$, and $\hat{f}_X(3) = \hat{f}_X(4) = .1333$. (b) $\hat{F}_X(0) = .2667, \hat{F}_X(1) = .4667, \hat{F}_X(2) = .7333, \hat{F}_X(3) = .8667$, and $\hat{F}_X(4) = 1.000$. (d) The mean $\bar{x} = 15$ and the variance $s^2 = 1.952$. (e) The median is 2 and the $IQR = 3$. According to the 1.5 IQR rule, there are no outliers.

5.5.6 The distribution is skewed to the right, so we choose the median as a measure of location and the IQR as a measure of spread.

5.5.7 $\psi(\mu) = x_{0.25} = \mu + \sigma_0 z_{0.25}$, where $z_{0.25}$ satisfies $\Phi(z_{0.25}) = .25$.

5.5.13 $\psi(\theta) = 2\theta(1-\theta)$.

Chapter 6

Section 6.1

6.1.1 The appropriate statistical model is the Binomial(n, θ), where $\theta \in \Omega = [0,1]$ is the probability of having this antibody in the blood (we can also think of θ as the unknown proportion of the population who have this antibody in their blood). The likelihood function is given by $L(\theta \,|\, s) = \binom{n}{s}\theta^s(1-\theta)^{n-s}$, where s is the number of people whose result was positive. The likelihood function for $n = 10$ people and $s = 3$ is given by $L(\theta \,|\, 3) = \binom{10}{3}\theta^3(1-\theta)^7$.

6.1.3 The likelihood function is given by $L(\theta \,|\, x_1, ..., x_{20}) = \theta^{20}\exp(-(20\bar{x})\,\theta)$. By the factorization theorem (Theorem 6.1.1) $\bar{x}$ is a sufficient statistic, so we only need to observe its value to obtain a representative likelihood. The likelihood function when $\bar{x} = 5.2$ is given by $L(\theta \,|\, x_1, ..., x_{20}) = \theta^{20}\exp(-20\,(5.2)\,\theta)$.

6.1.7 The likelihood function is given by $L(\theta \,|\, x_1, ..., x_n) = \theta^{n\bar{x}}e^{-n\theta}/\prod_{i=1}^{n} x_i!$. By the factorization theorem $\bar{x}$ is a sufficient statistic. If we differentiate $\ln L(\theta \,|\, x_1, ..., x_n) = -\ln\prod x_i! + n\bar{x}\ln\theta - n\theta$, we get $(\ln L(\theta \,|\, x_1, ..., x_n))' = n\bar{x}/\theta - n$ and setting this equal to 0 gives the solution $\theta = \bar{x}$. Therefore, we can obtain $\bar{x}$ from the likelihood and we conclude that it is a minimal sufficient statistic.

6.1.13 The likelihood function is given by $L(\theta \,|\, x_1, ..., x_n) = \theta^{-n} I_{[x_{(n)},\infty)}(\theta)$ when $\theta > 0$. By the factorization theorem $x_{(n)}$ is a sufficient statistic. Now notice that the likelihood function is 0 to the left of $x_{(n)}$ and positive to the right. So given the likelihood, we can determine $x_{(n)}$ and it is minimal sufficient.

Section 6.2

6.2.1 The MLEs are $\hat{\theta}(1) = a, \hat{\theta}(2) = b,\ \hat{\theta}(3) = b$, and $\hat{\theta}(4) = a$.

6.2.4 The likelihood function is $L(\theta \,|\, x_1, ..., x_n) = e^{-n\theta}\theta^{n\bar{x}}$, the log-likelihood function is $l(\theta \,|\, x_1, ..., x_n) = -n\theta + n\bar{x}\ln\theta$ and the score function is given by $S(\theta \,|\, x_1, ..., x_n) = -n + n\bar{x}/\theta$. Solving the score equation gives $\hat{\theta}(x_1, ..., x_n) = \bar{x}$. Note that since $\bar{x} \geq 0$, we have $\left.\frac{\partial S(\theta \,|\, x_1,...,x_n)}{\partial\theta}\right|_{\theta=\bar{x}} = -\left.\frac{n\bar{x}}{\theta^2}\right|_{\theta=\bar{x}} = -\frac{n}{\bar{x}} < 0$. So $\bar{x}$ is the MLE.

6.2.9 The likelihood function is $L(\alpha \,|\, x_1, ..., x_n) = \alpha^n\left(\prod_{i=1}^{n}(1+x_i)\right)^{-(\alpha+1)}$, the log-likelihood function is $l(\alpha \,|\, x_1, ..., x_n) = n\ln\alpha - (\alpha+1)\sum_{i=1}^{n}\ln(1+x_i)$, and the score function is $S(\alpha \,|\, x_1, ..., x_n) = \frac{n}{\alpha} - \sum_{i=1}^{n}\ln(1+x_i)$. Solving the score equation gives $\hat{\alpha}(x_1, ..., x_n) = n/\sum_{i=1}^{n}\ln(1+x_i)$. Note also that $\frac{\partial}{\partial\alpha}S(\alpha \,|\, x_1, ..., x_n) = -\frac{n}{\alpha^2} < 0$ for every α, so $\hat{\alpha}$ is the MLE.

6.2.13 The log-likelihood function is given by $l(\mu \mid x_1, ..., x_n) = -n(\bar{x} - \mu)^2/2$, and as a function of μ, its graph is a concave parabola and its maximum value occurs at $\bar{x}$. So if $\bar{x} \geq 0$, this is the MLE. If $\bar{x} < 0$, however, the maximum occurs at 0 and this is the MLE.

Section 6.3

6.3.1 This is a two sided z-test with the z statistic equal to -0.54 and the P-value is equal to 0.592. So there is no evidence against H_0. A .95-confidence interval for the unknown μ is $(4.442, 5.318)$. Note that the confidence interval contains the value 5, which confirms our conclusion using the above test.

6.3.4 This is a two sided t-test with the t statistic equal to 9.12 and (using the Student(3) distribution) the P-value equals 0.452, so we don't have evidence against the null hypothesis.

6.3.8 The P-value equals 0.32, so we conclude that there is no evidence against H_0. A .90-confidence interval for θ is given by $(0.559832, 0.680168)$ which includes the value 0.65, so agrees with the result of the above test.

6.3.19 The form of the power function associated with the above hypothesis assessment procedure is given by $\beta(\mu) = 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} + z_{1-\alpha}\right)$.

Section 6.4

6.4.1 An approximate .95-confidence interval for μ_3 is given by $(26.027, 151.373)$.

6.4.4 Let $\psi(\mu) = \exp(\mu)$ then $\psi'(\mu) = \exp(\mu)$. By the delta theorem (6.4.1) an approximate γ-confidence interval for $\psi(\mu)$ is given by $(-5.6975, 42.046)$.

6.4.13 For a random variable with this distribution, we have that $E_{\hat{F}}(X^i) = \sum_{j=1}^{n} x_{(j)}^i (\hat{F}(x_{(j)}) - \hat{F}(x_{(j-1)}))$ where we take $x_{(0)} = -\infty$. Now $\hat{F}(x_{(j)}) - \hat{F}(x_{(j-1)}) = 1/n$ since all the $x_{(j)}$ are distinct.

Section 6.5

6.5.1 The Fisher information is $nI(\sigma^2) = n/2\sigma^4$.

6.5.3 The Fisher information is $nI(\alpha) = n/\alpha^2$.

6.5.5 An approximate .90-confidence interval is given by $\frac{2}{\bar{x}} \pm \frac{1}{\sqrt{2n}}\left(\frac{2}{\bar{x}}\right) z_{.95} = \left(9.5413 \times 10^{-4}, 1.5045 \times 10^{-3}\right)$.

Chapter 7

Section 7.1

7.1.1 $\pi(1 \mid s = 1) = 3/16, \pi(2 \mid s = 1) = 4/16, \pi(3 \mid s = 1) = 9/16, \pi(1 \mid s = 2) = 3/14, \pi(2 \mid s = 2) = 4/7$, and $\pi(3 \mid s = 2) = 3/14$.

7.1.4 The posterior is a Gamma$(n\bar{x} + a, n + \beta)$ distribution.

7.1.13 $m(x_1, ...x_n) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n\bar{x}+\alpha)\Gamma(n(1-\bar{x})+\beta)}{\Gamma(\alpha+\beta+n)}$ for $(x_1, ...x_n) \in \{0,1\}^n$.

Section 7.2

7.2.1 $E(\theta^m \mid x_1, ..., x_n) = \frac{\Gamma(\alpha+\beta+n)\Gamma(n\bar{x}+\alpha+m)}{\Gamma(n\bar{x}+\alpha)\Gamma(\alpha+\beta+n+m)}$.

7.2.3 $E\left(1/\sigma^2 \,|\, x_1, ..., x_n\right) = (\alpha_0 + n/2)/\beta_x$, where β_x is given by (7.1.8) and the posterior mode is $1/\hat{\sigma}^2 = (\alpha_0 + n/2 - 1)/\beta_x$.

7.2.12 (a) The posterior mode is $\hat{\mu} = 64.053$. A .95-credible interval for μ is given by $64.053 \pm \sqrt{9/19} z_{0.975} = (62.704, 65.402)$. This interval has length equal to 2.698 and the margin of error is less than 1.5 marks, which is quite small, so we conclude that the estimate is quite accurate. (b) Based on the .95-credible interval, we can't reject $H_0 : \mu = 65$ at the 5% level since 65 falls inside the interval. (c) The posterior probability of the null hypothesis above is .3885. (d) The Bayes factor in favor of $H_0 : \mu = 65$ is given by .6353.

Section 7.3

7.3.7 As we increase the Monte Carlo sample size N, the interval that contains the exact value of the posterior expectation with virtual certainty becomes shorter and shorter. But for a given sample size n for the data, the posterior expectation will not be equal to the true value of $1/(\alpha + 1)$, so this interval will inevitably exclude the true value.

7.3.9 $F_{Y|X}^{-1}(u) = \left(\left(1 - x^2\right) u\right)^{1/2}$ for $0 < u < 1$. Therefore, we can generate Y given $X = x$ by generating $U \sim$ Uniform$[0, 1]$ and putting $Y = \left(\left(1 - x^2\right) U\right)^{1/2}$, $F_{X|Y}^{-1}(u) = \left(y^2 u\right)^{1/2} = y u^{1/2}$ for $0 < u < 1$. Therefore, we can generate X given $Y = y$ by generating $U \sim$ Uniform$[0, 1]$ and putting $X = yU^{1/2}$. So we select x_0. Then we generate $Y \sim f_{Y|X}(\cdot \,|\, x_0)$, using the above algorithm, obtaining y_1. Next we generate $X \sim f_{X|Y}(\cdot| \, y_1)$, using the above algorithm, obtaining x_1. Then we generate $Y \sim f_{Y|X}(\cdot \,|\, x_1)$, using the above algorithm, obtaining y_2, etc. We can generate exactly from this distribution as follows: $F_Y^{-1}(u) = u^{1/4}$ for $0 < u < 1$, so we can generate $Y \sim F_Y$ by generating $U \sim$ Uniform$[0, 1]$ and putting $y = U^{1/4}$. Then we use the above algorithm to generate $X \sim f_{X|Y}(\cdot \,|\, y)$.

Section 7.4

7.4.1 The posterior density is $\propto \lambda^{n+\alpha-1} \exp\left[-\lambda\left(\ln\left(\prod(1 + x_i)\right) + \beta\right)\right]$ and we recognize this as being proportional to the Gamma$(n + \alpha, \ln\left(\prod(1 + x_i)\right) + \beta)$ density. Hence, this is a conjugate family.

7.4.3 (a) $m_1(1, 1, 3) = 59/1728, m_2(1, 1, 3) = 43/1296$ and the maximum value of the prior predictive is obtained when $\tau = 1$. (b) $\pi_1(a \,|\, 1, 1, 3) = 32/59, \pi_1(b \,|\, 1, 1, 3) = 27/59$.

7.4.7 Jeffreys' prior for this model is $\sqrt{n}\theta^{-1/2}(1 - \theta)^{-1/2}$. The posterior density of θ is then proportional to $\theta^{n\bar{x}-1/2}(1 - \theta)^{n(1-\bar{x})-1/2}$, which we recognize as the unnormalized density of a Beta$(n\bar{x} + 1/2, n(1 - \bar{x}) + 1/2)$ distribution.

7.4.11 Jeffreys' prior is given by $1/\sigma^2$.

Chapter 8

Section 8.1

8.1.1 We have that $L(1 \,|\, \cdot) = (3/2) L(2 \,|\, \cdot)$, so by Section 6.1.1 T is a sufficient statistic. Then, given $T = 1$, the conditional distributions of s are given by

$f_a(1 \mid T=1) = 2/3, f_a(2 \mid T=1) = 1/3, f_a(3 \mid T=1) = 0, f_a(4 \mid T=1) = 0$, and $f_b(1 \mid T=1) = 2/3, f_b(2 \mid T=1) = 1/3, f_b(3 \mid T=1) = 0, f_b(4 \mid T=1) = 0$, and we see that these are the same (i.e., independent of θ). When $T = 3$ the conditional distributions of s are degenerate at $s = 3$ and when $T = 4$ the conditional distributions are degenerate at $s = 4$. These are also independent of θ.

8.1.4 $E(\bar{x} + \sigma_0 z_{.25}) = E(\bar{x}) + \sigma_0 z_{.25} = \mu + \sigma_0 z_{.25}$. Since $\bar{x}$ is complete, this implies that $\bar{x} + \sigma_0 z_{.25}$ is UMVU.

8.1.14 Suppose that c is a function such that $E_\theta(c(U)) = 0$ for every θ. Then $E_\theta(c(h(T))) = 0$ for every θ and the completeness of T implies that $P_\theta(\{s : c(h(T(s))) = 0\}) = 1$ for every θ. Now suppose u is such that $c(u) \neq 0$. Then $P_\theta(U = u) = P_\theta(h(T) = u) = P_\theta\left(T = h^{-1}(u)\right) =$
$P_\theta\left(\left\{s : T(s) = h^{-1}(u)\right\}\right) = 0$ since $c(h(T(s)) = c(u)$ for s in
$\left\{s : T(s) = h^{-1}(u)\right\}$. This implies that U is complete.

Section 8.2

8.2.1 The ratio $f_b(s)/f_a(s)$ has the following distribution when $\theta = a$:
$P_a(f_b(s)/f_a(s) = 3/2) = P_a(\{1,2\}) = 1/2, P_a(f_b(s)/f_a(s) = 2) = P_a(\{3\}) = 1/12$, and $P_a(f_b(s)/f_a(s) = 1/5) = P_a(\{4\}) = 5/12$. When $\alpha = .1$, using (8.2.4) and (8.2.5), we have that $c_0 = 3/2$ and $\gamma = ((1/10) - (1/12))/(1/2) = 1/30$. The power of the test is $P_b(\{3\}) + P_b(\{1,2\})/30 = 1/6 + (3/4)/30 = 23/120$. When $\alpha = .05$ we have that $c_0 = 2$ and $\gamma = ((1/20) - 0)/(1/12) = 3/5$. The power of the test is $P_b(\{3\})(3/5) = (1/6)(3/5) = 1/10$.

8.2.3 By (8.2.6) the optimal size .01 test is of the form (using $z_{.99} = 2.3263$) $\varphi_0(\bar{x}) = 1$ when $\bar{x} \geq 2.0404$ and is 0 otherwise.

8.2.7 With $x_\alpha(\beta_0)$ denoting the αth quantile of the Gamma$(n\alpha_0, \beta_0)$ distribution, the UMP size α test for $H_0 : \beta \leq \beta_0$ versus $H_a : \beta > \beta_0$ is to reject whenever $n\bar{x} \leq x_\alpha(\beta_0)$.

Section 8.3

8.3.1 The posterior distribution of θ is given by $\Pi(\theta = 1 \mid 2) = 2/5, \Pi(\theta = 2 \mid 2) = 3/5$, so $\Pi(\theta = 2 \mid 2) > \Pi(\theta = 1 \mid 2)$. We accept $H_0 : \theta = 2$.

8.3.3 The Bayes rule is given by the posterior mean $(1/\tau_0^2 + n/\sigma_0^2)^{-1}(\mu_0/\tau_0^2 + n\bar{x}/\sigma_0^2)$ and this converges to $\bar{x}$ as $\tau_0 \to \infty$.

8.3.6 The Bayes rule is given by the posterior mean of $1/\beta$, namely
$(n\bar{x} + \upsilon_0)/(n\alpha_0 + \tau_0 - 1)$, which converges with probability 1 to$(\alpha_0/\beta)/\alpha_0 = \beta^{-1}$ as $n \to \infty$.

Section 8.4

8.4.1 The model is given by the collection of probability functions
$\left\{\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}} : \theta \in [0,1]\right\}$ on the set of all sequences $(x_1, \ldots, x_n)$ of 0's and 1's. The action space is $\mathcal{A} = [0,1]$, the correct action function is $A(\theta) = \theta$, and the loss function is $L(\theta, a) = (\theta - a)^2$. The risk function for T is given by $R_T(\theta) = E_\theta\left((\theta - \bar{x})^2\right) = \mathrm{Var}_\theta(\bar{x}) = \theta(1-\theta)/n$.

8.4.4 The model is given by the collection of probability functions $\left\{\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}} : \theta \in [0,1]\right\}$ on the set of all sequences $(x_1, \ldots, x_n)$ of 0's and 1's. The action space is $\mathcal{A} = \{H_0, H_a\}$, where $H_0 : \theta = 1/2$, the correct action function is $A(\theta) = H_0$ when $\theta = 1/2$, and $A(\theta) = H_a$ when $\theta \neq 1/2$. The loss function is $L(\theta, a) = 0$ when $\theta = 1/2, a = H_0$ or $\theta \neq 1/2, a = H_a$ and $L(\theta, a) = 1$ when $\theta = 1/2, a = H_a$ or $\theta \neq 1/2, a = H_0$. The test function $\varphi(n\bar{x}) = 0$ when $n\bar{x} \notin \{0, 1, n-1, n\}$ and $\varphi(n\bar{x}) = 1$ when $n\bar{x} \in \{0, 1, n-1, n\}$ has risk function given by $R_\varphi(\theta) = \binom{n}{0}(1-\theta)^n + \binom{n}{1}\theta(1-\theta)^{n-1} + \binom{n}{n-1}\theta^{n-1}(1-\theta) + \binom{n}{n}\theta^n$.

8.4.8 Suppose we have that $\delta(s, \cdot)$ is degenerate at $d(s)$ for each s. Then, clearly, $d : S \to \mathcal{A}$. Now suppose we have $d : S \to \mathcal{A}$ and define $\delta(s, B) = 1$ when $d(s) \in B$ and $\delta(s, B) = 0$ otherwise. Then $\delta(s, \mathcal{A}) = 1$, and if $B_1, B_2, \ldots$ are mutually disjoint subsets of $\mathcal{A}$, then $d(s) \in B_i$ for one i (and only one) if and only if $d(s) \in \cup_{j=1}^{\infty} B_j$, so $\delta\left(s, \cup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \delta(s, B_j)$. Therefore, $\delta(s, \cdot)$ is a probability measure for each s, and δ is a decision function. Now, using the fact that $\delta(s, \cdot)$ is a discrete probability measure degenerate at $d(s)$, we have that $R_\delta(\theta) = E_\theta\left(E_{\delta(s,\cdot)}(L(\theta, a))\right) = E_\theta(\delta(s, \{d(s)\})(L(\theta, d(s)))) = E_\theta(L(\theta, d(s)))$ since $\delta(s, \{d(s)\}) = 1$.

Chapter 9

Section 9.1

9.1.1 $D(r) = \left(\sigma_0^2\right)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 22.761$. Now $D(R) \sim \chi^2(19)$ distribution, so the P-value is then given by $P(D(R) > 22.761) = .248$, which doesn't indicate any evidence against the model being correct.

9.1.5 The Chi-squared statistic is equal to 3.50 and the P-value is given by $(X^2 \sim \chi^2(4))$ $P\left(X^2 \geq 3.5\right) = 0.4779$. Therefore, we have no evidence against the Uniform model being correct.

9.1.11 We have $E(a_1Y_1 + \cdots + a_kY_k) = a_1\mu_1 + \cdots + a_k\mu_k$ and so $E(Y_i) = \mu_i$ by taking $a_i = 1$ and $a_j = 0$ whenever $j \neq i$. By Theorem 3.3.3 (b) we have $\text{Var}(a_1Y_1 + \cdots + a_kY_k) = a_1^2 \text{Var}(Y_1) + \cdots + a_k^2 \text{Var}(Y_k) + 2\sum_{i<j} a_ia_j \text{Cov}(Y_i, Y_j) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_ia_j\sigma_{ij}$. Putting $a_i = 1$ and $a_j = 0$ whenever $i \neq j$, we obtain $\text{Var}(Y_i) = \sigma_{ii}$ and this implies that $Y_i \sim N(\mu_i, \sigma_{ii})$. Putting $a_i = a_j = 1$ and $a_l = 0$ whenever $l \notin \{i, j\}$, we obtain $\text{Var}(Y_i + Y_j) = \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij} = \text{Var}(Y_i) + \text{Var}(Y_j) + 2\,\text{Cov}(Y_i, Y_j)$. This immediately implies that $\text{Cov}(Y_i, Y_i) = \sigma_{ij}$.

Section 9.2

9.2.1 (a) The probability of obtaining $s = 2$ from f_1 is $1/3$, which is a reasonable value, so we have no evidence against the model $\{f_1, f_2\}$. (b) The prior predictive M distribution is given by $m(1) = 1/3, m(2) = 1/10, m(3) = 17/30$. So the probability of a data set occurring with probability as small or smaller than $m(2)$ is $1/10$, so the observation 2 is not very surprising. Accordingly, there is no evidence of a prior-data conflict. (c) The prior predictive M now is given by $m(1) = 1/3, m(2) = 1/300, m(3) = 199/300$. So the probability of a data set occurring with probability as small or smaller than $m(2)$ is $1/300$, so the observation 2 is surprising. Accordingly, there is some evidence of a prior-data conflict.

9.2.4 First, by Corollary 4.6.1 we have $\bar{X} \sim N(\mu, \sigma_0^2/n)$. Then we can write $\bar{X}$ as $\bar{X} = \mu + Z/\sqrt{n}$, where $Z \sim N(0, \sigma_0^2)$ is independent of $\mu \sim N(\mu_0, \tau_0^2)$. Hence, by Theorem 4.6.1 we have that the prior predictive distribution of $\bar{X}$ is the $N(\mu_0, \tau_0^2 + \sigma_0^2/n)$ distribution.

Chapter 10

Section 10.1

10.1.1 From the definitions we know that if the conditional distribution of Y given X does not change as we change X, then X and Y are unrelated. Then for any x_1, x_2 (that occur with positive probability), and y we have $P(Y = y \mid X = x_1) = P(Y = y \mid X = x_2)$. Hence, $P(X = x_1, Y = y) /$ $P(X = x_1) = P(X = x_2, Y = y) / P(X = x_2)$, so $P(X = x_1, Y = y) = P(X = x_2, Y = y) P(X = x_1) / P(X = x_2)$. Summing this over x_1 gives $P(Y = y) = P(X = x_2, Y = y) / P(X = x_2)$, so we must have $P(X = x_2, Y = y) = P(X = x_2) P(Y = y)$ and this implies that X and Y are statistically independent. Now on the other hand, if X and Y are statistically independent, then for all x and y we have $P(Y = y \mid X = x) = P(Y = y)$, so the conditional distribution of Y given X does not change as we change X and therefore X and Y are unrelated.

10.1.3 The conditional distribution of Y given $X = x$ does change as we change x, so we conclude that X and Y are related.

10.1.7 If the conditional distribution of life-length given various smoking habits changes, then we can conclude that these two variables are related. However, since we can't assign the value of smoking habit (perhaps different amount of smoking), and considering there may be many other confounding variables that should be take into account, e.g., exercise habits, eating habits, sleeping habits, etc., we can't conclude that this relationship is a cause-effect relationship.

10.1.13 (a) The response variable could be whether or not they have watched the particular program. Predictor variables might be the number of members in each household, whether or not they received the brochure, number of televisions in the house, number of children, etc. (b) We cannot claim for a cause-effect relationship since for many of the predictor variables it would be impossible to assign the different values it can take.

10.1.19 U and V are related.

Section 10.2

10.2.1 The Chi-squared statistic is equal to $X_0^2 = 5.7143$ and, with $X^2 \sim \chi^2(2)$, the P-value equals $P\left(X^2 > 5.7143\right) = .05743$. Therefore, we don't have evidence against the null hypothesis of no difference in the distributions of thunderstorms between the two years (at least at the .05 level).

10.2.5

(a) First, note that the predictor variable, X (gender), is not random. The Chi-squared statistic is equal to $X_0^2 = 10.4674$ and, with $X^2 \sim \chi^2(4)$, the P-value equals $P\left(X^2 > 10.4674\right) = .03325$. Therefore, we have some evidence against the null hypothesis of no relationship between hair color and gender. (b) The appropriate bar plots are the two conditional distributions. (c) The standardized residuals are given in the following table. They all look reasonable, so nothing stands out as an explanation of why the model of independence doesn't fit. Overall, it looks like a large sample size has detected a small difference.

	$Y =$ fair	$Y =$ red	$Y =$ medium	$Y =$ dark	$Y =$ jet black
$X =$ m	-1.07303	0.20785	1.05934	-0.63250	1.73407
$X =$ f	1.16452	-0.22557	-1.14966	0.68642	-1.88191

10.2.16 $E\left(X_1^{l_1}\cdots X_k^{l_k}\right) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\frac{\Gamma(\alpha_1+l_1)\cdots\Gamma(\alpha_k+l_k)}{\Gamma(\alpha_1+\cdots+\alpha_k+l_1+\cdots+l_k)}$.

Section 10.3

10.3.2 Since $\bar{x} \in [0, \theta] \subset [0, \infty)$ with probability 1, we have that $\bar{x}$ is the least squares estimate of the mean $\theta/2 \in [0, \infty)$.

10.3.4 (b) The least squares estimates of β_1 and β_2 are given by $b_2 = 2.1024$ and $b_1 = \bar{y} = -0.00091$, so the least-squares line is given by $y = -0.00091 + 2.1024x$. (e) Both graphs indicate that the normal simple linear regression model is reasonable. (f) A .95-confidence interval for the intercept is $(-1.0533, 1.0515)$ and a .95-confidence interval for the slope is $(1.7696, 2.4352)$. (g) The F statistic for testing $H_0 : \beta_2 = 0$ is given by $F = 204.28$ and, since $F \sim F(1, 9)$ under H_0, the P-value is given by $P(F > 204.28) = .000$, so we reject the null hypothesis of no effect between X and Y. (h) The proportion of the observed variation in the response that is being explained by changes in the predictor is given by the coefficient of determination $R^2 = .9578$. (i) The prediction is given by $y = -0.00091$. This is an interpolation because 0.0 is in the range of observed X values. The standard error of this prediction is 0.46515. (j) The prediction is given by $y = 12.613$. This is an extrapolation because 6 is not in the range of observed X values. The standard error of this prediction is 0.99763. (k) The prediction is given by $y = 42.047$. This is an extrapolation because 12 is not in the range of observed X values. The standard error of this prediction is 2.9784. The standard errors get larger as we move away from the observed X values.

10.3.12 Since $\sum_{i=1}^n (x_i - \bar{x})^2 = 0$, we must have $(x_i - \bar{x})^2 = 0$, so $x_i = \bar{x}$ for every i and all the x_i are equal to the same value, say x. Then we need to estimate the conditional mean of Y at $X = x$ based on a sample $(y_1, \ldots, y_n)$ from this distribution. The model says that this conditional mean is of the form $E(Y \mid X = x) = \beta_1 + \beta_2 x$, where $\beta_1, \beta_2 \in R^1$. Therefore, $E(Y \mid X = x)$ can be any value in R^1 and the least squares estimate is given by the sample average $\bar{y}$.

10.3.16

(a) Putting $b = \sum_{i=1}^n x_i y_i / \sum_{i=1}^n x_i^2$, we have that $\sum_{i=1}^n (y_i - \beta x_i)^2 =$
$\sum_{i=1}^n (y_i - b x_i)^2 + (b - \beta)^2 \sum_{i=1}^n x_i^2$ since $\sum_{i=1}^n (y_i - b x_i) x_i =$

$\sum_{i=1}^n x_i y_i - b\sum_{i=1}^n x_i^2 = 0$, and this is clearly minimized, as a function of β, by b. (b) $E(B \mid X_1 = x_1, \ldots, X_n = x_n) = \sum_{i=1}^n x_i(\beta x_i)/\sum_{i=1}^n x_i^2 = \beta$ and $\text{Var}(B \mid X_1 = x_1, \ldots, X_n = x_n) = \sigma^2/\sum_{i=1}^n x_i^2$. (c) $E(S^2 \mid X_1 = x_1, \ldots, X_n = x_n) = (n-1)^{-1}\sum_{i=1}^n E\left((Y_i - Bx_i)^2 \mid X_1 = x_1, \ldots, X_n = x_n\right)$ and
$E\left((Y_i - Bx_i)^2 \mid X_1 = x_1, \ldots, X_n = x_n\right) = \sigma^2 - \sigma^2 x_i^2/\sum_{i=1}^n x_i^2 =$
$\sigma^2(1 - x_i^2/\sum_{i=1}^n x_i^2)$, so $E\left(S^2 \mid X_1 = x_1, \ldots, X_n = x_n\right) = \sigma^2$. (d) We have that $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n (y_i - bx_i)^2 + b^2\sum_{i=1}^n x_i^2$. Here we have that $\sum_{i=1}^n (y_i - bx_i)^2$ is the error sum of squares and $b^2\sum_{i=1}^n x_i^2$ is the regression sum of squares. The coefficient of determination is then given by $R^2 = b^2\sum_{i=1}^n x_i^2/\sum_{i=1}^n y_i^2$ and this is the proportion of the total variation observed in Y (as measured by $\sum_{i=1}^n y_i^2$) due to changes in X. (e) Since B is a linear combination of independent normal variables, we have that B is normally distributed with mean given by (part (b)) β and variance (part (b)) given by $\sigma^2/\sum_{i=1}^n x_i^2$. (f) We have that $(B - \beta)/\sigma\left(\sum_{i=1}^n x_i^2\right)^{-1/2} \sim N(0,1)$ independent of $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, so $(B - \beta)/S\left(\sum_{i=1}^n x_i^2\right)^{-1/2} \sim t(n-1)$. Now there is no relationship between X and Y if and only if $\beta = 0$ so we test $H_0 : \beta = 0$ by computing the P-value $P(|T| > |b/s\left(\sum_{i=1}^n x_i^2\right)^{-1/2}|)$, where $T \sim t(n-1)$. (g) We have that $y_i = bx_i + (y_i - bx_i)$, and when the model is correct $y_i - bx_i$ is a value from a distribution with mean 0 and variance (see part (b)) $\sigma^2\left(1 - x_i^2/\sum_{i=1}^n x_i^2\right)$. Therefore, the ith standardized residual is given by $(y_i - bx_i)/s\left(1 - x_i^2/\sum_{i=1}^n x_i^2\right)^{1/2}$. We can plot these in residual plots and normal probability plots to see if they look like samples from the $N(0,1)$ distribution.

Section 10.4

10.4.1 (c) The F statistic for testing H_0 is given by $F = 2.18/2.09 = 1.0431$ and, since $F \sim F(2,9)$ under H_0, we have P-value $P(F > 1.0431) = .39135$. Therefore, we don't have evidence against the null hypothesis of no difference among the conditional means of Y given X. (d) Since we didn't find any relationship between Y and X, there is no need to calculate these confidence intervals.

10.4.5 (c) The F statistic for testing H_0 is given by $F = 6.414/0.589 = 10.89$ and, since $F \sim F(3,20)$ under H_0, the P-value equals $P(F > 10.89) = .00019$. Therefore, we have strong evidence against the null hypothesis of no difference among the conditional means of Y given the predictor.

10.4.11 If an interaction exists between the two factors, then the b response curves are not parallel, so cannot be horizontal, i.e., there must be effect due to both factors.

10.4.16 An individual error rate of .01 gives a family error rate of 0.0455.

10.4.17 (d) The F statistic for testing H_0 : no interaction between Cheese and Lot, is given by $F = 0.151/0.110 = 1.3727$ and, since $F \sim F(2,6)$ under H_0, the P-value equals $P(F > 1.3727) = .32293$. Therefore, we don't have evidence against the null hypothesis of no interaction effect. We can then proceed to calculate the P-value for testing H_0 : no effect due to Cheese. Since $F \sim F(1,6)$

under H_0, this is given by $P(F > 0.114/0.110 = 1.0364) = .34794$. Therefore, we don't have any evidence against the null hypothesis of no effect due to Cheese. The P-value for testing H_0 : no effect due to Lot, since $F \sim F(2,6)$ under H_0, is given by $P(F > 12.950/0.110 = 117.73) = .00002$. Therefore, we have strong evidence against the null hypothesis of no effect due to Lot.

Section 10.5

10.5.4 (b) The Chi-squared statistic for testing the validity of the model is then equal to 4.66204 with P-value given by $P\left(\chi^2(8) > 4.66204\right) = .79301$. Therefore, we have no evidence that the model is incorrect. (c) The P-value for testing $H_0 : \beta_3 = 0$ is 0.638, so we don't have any evidence against the null hypothesis.

Chapter 11

Section 11.1

11.1.1 (a) 0. (b) 0. (f) 4/9. (k) 0.00925.

11.1.5 (a) $P(\tau_c < \tau_0) \doteq 0.89819$. (b) Here $P(\tau_c < \tau_0) \doteq 0.881065$. (d) Here $P(\tau_c < \tau_0) \doteq 0.0183155$.

Section 11.2

11.2.1 (a) $P(X_0 = 1) = \mu_1 = 0.7$. (b) $P(X_0 = 2) = \mu_2 = 0.1$. (d) $P(X_1 = 2 \,|\, X_0 = 1) = p_{12} = 1/4$

11.2.5 (a) $P_2(X_1 = 1) = p_{21} = 1/2$. (b) $P_2(X_1 = 2) = p_{22} = 0$. (g) $P_2(X_3 = 3) = 37/100$.

11.2.10 (a) This chain is irreducible. (b) The chain is aperiodic. (c) $\pi_1 = 1/4$ and $\pi_2 = \pi_3 = 1/2$. (d) $\lim_{n\to\infty} P_1(X_n = 2) = \pi_2 = 1/2$. Hence, $P_1(X_{500} = 2) \approx 1/2$.

Section 11.3

11.3.1 First, choose any initial value X_0. Then, given $X_n = i$, let $Y_{n+1} = i + 1$ or $i - 1$ with probability 1/2 each. Let $j = Y_{n+1}$ and let $\alpha_{ij} = \min(1,\ \pi_j/\pi_i) = \min(1,\ e^{-(j-13)^4+(i-13)^4})$. Then let $X_{n+1} = j$ with probability α_{ij}, otherwise let $X_{n+1} = i$ with probability $1 - \alpha_{ij}$.

Section 11.4

11.4.1 $C = 12/5$.

11.4.4 $P(X_n = 14) = 1.2$.

11.4.7 (a) $\{X_n\}$ is a martingale. (b) T is a stopping time. (c) $E(X_T) = 27$. (d) $P(X_T = 1) = 27/40$.

Section 11.5

11.5.1 (a) $P(Y_1^{(1)} = 1) = 1/2$. (c) $P(Y_1^{(2)} = \sqrt{2}) = 1/4$.

11.5.5 $E(B_{13}B_8) = 8$.

11.5.10 (a) $P(X_8 > 500) \doteq 0.00004276$.

Section 11.6

11.6.1 (a) $N_2 \sim \text{Poisson}(14)$, so $P(N_2 = 13) \doteq 0.1060$. (b) $P(N_5 = 3) \doteq 4.5 \times 10^{-12}$. (f) $P(N_2 = 13,\ N_5 = 20) \doteq 2.9 \times 10^{-5}$.
11.6.6 (a) $P(N_6 = 5 \mid N_9 = 5) \doteq 0.1317$.