Multivariate Distributions

1

> **2.1.6.** Let $f(x, y) = e^{-x-y}$, $0 < x < \infty$, $0 < y < \infty$, zero elsewhere, be the pdf of X and Y. Then if $Z = X + Y$, compute $P(Z \leq 0)$, $P(Z \leq 6)$, and, more generally, $P(Z \leq z)$, for $0 < z < \infty$. What is the pdf of Z?

2.1.7. Let X and Y have the pdf $f(x, y) = 1$, $0 < x < 1$, $0 < y < 1$, zero elsewhere. Find the cdf and pdf of the product $Z = XY$.

2.1.6

$$\begin{aligned} G(z) &= P(X + Y \leq z) = \int_0^z \int_0^{z-x} e^{-x-y}\, dydx \\ &= \int_0^z [1 - e^{-(z-x)}]e^{-x}\, dx = 1 - e^{-z} - ze^{-z}. \\ g(z) &= G'(z) = \begin{cases} ze^{-z} & 0 < z < \infty \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

2.1.7

$$\begin{aligned} G(z) &= P(XY \leq z) = 1 - \int_z^1 \int_{z/x}^1 dydx \\ &= 1 - \int_z^1 \left(1 - \frac{z}{x}\right) dx = z - z \log z \\ g(z) &= G'(z) = \begin{cases} -\log z & 0 < z < 1 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

Why is $-\log z > 0$?

2

2.2.1. If $p(x_1, x_2) = (\frac{2}{3})^{x_1+x_2}(\frac{1}{3})^{2-x_1-x_2}$, $(x_1, x_2) = (0,0), (0,1), (1,0), (1,1)$, zero elsewhere, is the joint pmf of X_1 and X_2, find the joint pmf of $Y_1 = X_1 - X_2$ and $Y_2 = X_1 + X_2$.

2.2.2. Let X_1 and X_2 have the joint pmf $p(x_1, x_2) = x_1x_2/36$, $x_1 = 1, 2, 3$ and $x_2 = 1, 2, 3$, zero elsewhere. Find first the joint pmf of $Y_1 = X_1X_2$ and $Y_2 = X_2$, and then find the marginal pmf of Y_1.

2.2.1

$$p(y_1, y_2) = \begin{cases} \left(\frac{2}{3}\right)^{y_2} \left(\frac{1}{3}\right)^{2-y_2} & (y_1, y_2) = (0,0), (-1,1), (1,1), (0,2) \\ 0 & \text{elsewhere.} \end{cases}$$

2.2.2

$$p(y_1, y_2) = \begin{cases} y_1/36 & y_1 = y_2, 2y_2, 3y_2; y_2 = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases}$$

y_1	1	2	3	4	6	9
$p(y_1)$	1/36	4/36	6/36	4/36	12/36	9/36

3

2.3.2. Let $f_{1|2}(x_1|x_2) = c_1 x_1/x_2^2$, $0 < x_1 < x_2$, $0 < x_2 < 1$, zero elsewhere, and $f_2(x_2) = c_2 x_2^4$, $0 < x_2 < 1$, zero elsewhere, denote, respectively, the conditional pdf of X_1, given $X_2 = x_2$, and the marginal pdf of X_2. Determine:

(a) The constants c_1 and c_2.

(b) The joint pdf of X_1 and X_2.

(c) $P(\frac{1}{4} < X_1 < \frac{1}{2} | X_2 = \frac{5}{8})$.

(d) $P(\frac{1}{4} < X_1 < \frac{1}{2})$.

2.3.3. Let $f(x_1, x_2) = 21x_1^2 x_2^3$, $0 < x_1 < x_2 < 1$, zero elsewhere, be the joint pdf of X_1 and X_2.

(a) Find the conditional mean and variance of X_1, given $X_2 = x_2$, $0 < x_2 < 1$.

(b) Find the distribution of $Y = E(X_1|X_2)$.

(c) Determine $E(Y)$ and $\text{Var}(Y)$ and compare these to $E(X_1)$ and $\text{Var}(X_1)$, respectively.

2.3.2

$$
\begin{array}{ll}
(a) & c_1 \int_0^{x_2} x_1/x_2^2 \, dx_1 = \frac{c_1}{2} = 1 \Rightarrow c_1 = 2 \text{ and } c_2 = 5. \\
(b) & 10x_1x_2^2, 0 < x_1 < x_2 < 1; \text{zero elsewhere} \\
(c) & \int_{1/4}^{1/2} 2x_1/(5/8)^2 \, dx = \frac{64}{25}\left(\frac{1}{4} - \frac{1}{16}\right) = \frac{12}{25}. \\
(d) & \int_{1/4}^{1/2} \int_{x_1}^{1} 10x_1x_2^2 \, dx_2 dx_1 = \int_{1/4}^{1/2} \frac{10}{3} x_1(1 - x_1^3) \, dx_1 = \frac{135}{512}.
\end{array}
$$

2.3.3

$$
\begin{aligned}
f_2(x_2) &= \int_0^{x_2} 21x_1^2x_2^3 \, dx_1 = 7x_2^6, \quad 0 < x_2 < 1. \\
f_{1|2}(x_1|x_2) &= 21x_1^2x_2^3/7x_2^6 = 3x_1^2/x_2^3, \quad 0 < x_1 < x_2. \\
E(X_1|x_2) &= \int_0^{x_2} x_1(3x_1^2/x_2^3) \, dx_1 = \frac{3}{4}x_2. \\
G(y) &= P\left(\frac{3}{4}X_2 \le y\right) = \int_0^{4y/3} 7x_2^6 \, dx_2 = \left(\frac{4y}{3}\right)^7, \quad 0 < y < \frac{3}{4} \\
g(y) &= \begin{cases} 7\left(\frac{4}{3}\right)^7 y^6 & 0 < y < \frac{3}{4} \\ 0 & \text{elsewhere.} \end{cases} \\
E(Y) &= \frac{7}{8}\frac{3}{4} = \frac{21}{32}. \\
\operatorname{Var}(Y) &= \frac{7}{1024}. \\
E(X_1) &= \frac{21}{32}. \\
\operatorname{Var}(X_1) &= \frac{553}{15360} > \frac{7}{1024}.
\end{aligned}
$$

2.3.8. Let X and Y have the joint pdf $f(x, y) = 2\exp\{-(x + y)\}$, $0 < x < y < \infty$, zero elsewhere. Find the conditional mean $E(Y|x)$ of Y, given $X = x$.

2.3.8 The marginal pdf of X is

$$f_X(x) = 2\int_x^\infty e^{-x}e^{-y} \, dy = 2e^{-2x}, \quad 0 < x < \infty.$$

Hence, the conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{2e^{-x}e^{-y}}{2e^{-2x}} = e^{-(y-x)}, \quad 0 < x < y < \infty,$$

with conditional mean

$$E(Y|X = x) = \int_x^\infty ye^{-(y-x)} \, dy = x + 1, \quad x > 0.$$

4

2.4.7. If the correlation coefficient ρ of X and Y exists, show that $-1 \le \rho \le 1$. *Hint:* Consider the discriminant of the nonnegative quadratic function

$$h(v) = E\{[(X - \mu_1) + v(Y - \mu_2)]^2\},$$

where v is real and is not a function of X nor of Y.

2.4.7
$$h(v) = \text{var}(X) + 2v\text{cov}(X,Y) + v^2\text{var}(Y) \geq 0,$$

for all v. Hence, the discriminant of this quadratic must satisfy $b^2 - 4ac \leq 0$ which yields
$$[2\text{cov}(X,Y)]^2 - 4\text{var}(X)\text{var}(Y) \leq 0.$$

Equivalently,
$$\rho^2 = [\text{cov}(X,Y)]^2/\text{var}(X)\text{var}(Y) \leq 1.$$

2.4.11. Let $\sigma_1^2 = \sigma_2^2 = \sigma^2$ be the common variance of X_1 and X_2 and let ρ be the correlation coefficient of X_1 and X_2. Show for $k > 0$ that

$$P[|(X_1 - \mu_1) + (X_2 - \mu_2)| \geq k\sigma] \leq \frac{2(1+\rho)}{k^2}.$$

2.4.11 Let $Y = (X_1 - \mu_1) + (X_2 - \mu_2)$. Then the mean of Y is 0 and its variance is

$$\text{Var}(Y) = \text{Var}(X_1 + X_2) = \sigma^2 + \sigma^2 + 2\rho\sigma^2 = 2\sigma^2(1-\rho).$$

Use Chebyshev's inequality to obtain the result.

5

2.5.4. Find $P(0 < X_1 < \frac{1}{3}, 0 < X_2 < \frac{1}{3})$ if the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2) = 4x_1(1-x_2)$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere.

2.5.4 Because X_1 and X_2 are independent, the probability equals

$$\left[\int_0^{1/3} 2x_1\,dx_1\right]\left[\int_0^{1/3} 2(1-x_2)\,dx_2\right] = (1/3)^2[1-(2/3)^2] = 5/81.$$

2.5.9. Suppose that a man leaves for work between 8:00 a.m. and 8:30 a.m. and takes between 40 and 50 minutes to get to the office. Let X denote the time of departure and let Y denote the time of travel. If we assume that these random variables are independent and uniformly distributed, find the probability that he arrives at the office before 9:00 a.m.

2.5.9

$$\begin{aligned} P(X+Y \leq 60) &= P(X \leq 10) + \int_{10}^{20}\int_{40}^{60-x} \frac{1}{300}\,dy\,dx \\ &= \frac{1}{3} + \int_{10}^{20} (20-x)/300\,dx = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

6

2.6.3. Let X_1, X_2, X_3, and X_4 be four independent random variables, each with pdf $f(x) = 3(1-x)^2$, $0 < x < 1$, zero elsewhere. If Y is the minimum of these four variables, find the cdf and the pdf of Y.
Hint: $P(Y > y) = P(X_i > y\ , i = 1, \ldots, 4)$.

2.6.3

$$\begin{aligned} G(y) &= 1 - P(y < X_i, i = 1,2,3,4) = 1 - [(1-y)^3]^4 = 1 - (1-y)^{12} \\ g(y) &= G'(y) = \begin{cases} 12(1-y)^{11} & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

2.6.9. Let X_1, X_2, X_3 be iid with common pdf $f(x) = \exp(-x)$, $0 < x < \infty$, zero elsewhere. Evaluate:

(a) $P(X_1 < X_2 | X_1 < 2X_2)$.

2.6.9

$$\begin{aligned} (a) \qquad & \int_0^\infty \int_{x_1}^\infty e^{-x_1-x_2}\, dx_2\, dx_1 \Big/ \int_0^\infty \int_{x_1/2}^\infty e^{-x_1-x_2}\, dx_2\, dx_1 \\ & + \int_0^\infty e^{-2x_1}\, dx_1 \Big/ \int_0^\infty e^{-3x_1/2}\, dx_1 = \frac{1}{2}\frac{2}{3} = \frac{3}{4}. \end{aligned}$$

7

2.7.1. Let X_1, X_2, X_3 be iid, each with the distribution having pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \quad Y_3 = X_1 + X_2 + X_3$$

are mutually independent.

2.7.2. If $f(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere, is the pdf of the random variable X, find the pdf of $Y = X^2$.

2.7.1

$$x_1 = y_1 y_2 y_3,\ x_2 = y_2 y_3 - y_1 y_2 y_3,\ x_3 = y_3 - y_2 y_3.$$

with $J = y_2 y_3^2$, and $0 < y_1 < 1, 0 < y_2 < 1, 0 < y_3 < \infty$. This yields

$$g(y_1, y_2, y_3) = y_2 y_3^2 e^{-y_3} = (1)(2y_2)(y_3^2 e^{-y_3}/2) = g_1(y_1) g_2(y_2) g_3(y_3).$$

2.7.2

$$x_1 = \sqrt{y},\ x_2 = -\sqrt{y} \text{ and } J_i = \tfrac{1}{2\sqrt{y}},\ i = 1.2.$$

This yields

$$g(y) = \frac{1}{2}\left(\frac{1}{2\sqrt{y}}\right) + \frac{1}{2}\left(\frac{1}{2\sqrt{y}}\right) = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1.$$

8

2.8.2. Let X_1, X_2, X_3, X_4 be four iid random variables having the same pdf $f(x) = 2x$, $0 < x < 1$, zero elsewhere. Find the mean and variance of the sum Y of these four random variables.

2.8.2 Note that

$$\begin{aligned} \mu_1 &= E(X_i) = \int_0^1 2x^2\,dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3} \\ E(X_i^2) &= \int_0^1 2x^3\,dx = \frac{2}{4}x^4\Big|_0^1 = \frac{1}{2} \end{aligned}$$

So

$$\sigma^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

Hence,

$$\begin{aligned} E(Y) &= \sum_{i=1}^4 E(X_i) = \frac{8}{3} \\ V(Y) &= \sum_{i=1}^4 V(X_i) = \frac{4}{18}, \end{aligned}$$

where we used the independence of $X_1, \ldots, X_4$ to establish the variance of Y.

2.8.4. If the independent variables X_1 and X_2 have means μ_1, μ_2 and variances σ_1^2, σ_2^2, respectively, show that the mean and variance of the product $Y = X_1X_2$ are $\mu_1\mu_2$ and $\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$, respectively.

2.8.4 By independence

$$\begin{aligned} E(X_1X_2) &= E(X_1)E(X_2) = \mu_1\mu_2 \\ E(X_1^2X_2^2) &= E(X_1^2)E(X_2^2) = (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2). \end{aligned}$$

So,

$$V(X_1X_2) = (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2) - \mu_1^2\mu_2^2,$$

which simplifies to the answer.

Some Special Distributions

1

3.1.2. The mgf of a random variable X is $(\frac{2}{3} + \frac{1}{3}e^t)^9$. Show that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \sum_{x=1}^{5} \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}.$$

3.1.3. If X is $b(n, p)$, show that

$$E\left(\frac{X}{n}\right) = p \quad \text{and} \quad E\left[\left(\frac{X}{n} - p\right)^2\right] = \frac{p(1-p)}{n}.$$

3.1.11. Let X be $b(2, p)$ and let Y be $b(4, p)$. If $P(X \geq 1) = \frac{5}{9}$, find $P(Y \geq 1)$.

3.1.12. If $x = r$ is the unique mode of a distribution that is $b(n, p)$, show that

$$(n+1)p - 1 < r < (n+1)p.$$

Hint: Determine the values of x for which the ratio $p(x+1)/p(x) > 1$.

3.1.11

$$\begin{aligned} P(X \geq 1) &= 1-(1-p)^2 = 5/9 \Rightarrow (1-p)^2 = 4/9 \\ P(Y \geq 1) &= 1-(1-p)^4 = 1-(4/9)^2 = 65/81. \end{aligned}$$

3.1.12 Let $f(x)$ denote the pmf which is $b(n,p)$. Show, for $x \geq 1$, that $f(x)/f(x-1) = 1 + [(n+1)p - x]/x(1-p)$. Then $f(x) > f(x-1)$ if $(n+1)p > x$ and $f(x) < f(x-1)$ if $(n+1)p < x$. Thus the mode is the greatest integer less than $(n+1)p$. If $(n+1)p$ is an integer, there is no unique mode but $f[(n+1)p] = f[(n+1)p-1]$ is the maximum of $f(x)$.

3.1.18. If a fair coin is tossed at random five independent times, find the conditional probability of five heads given that there are at least four heads.

3.1.19. Let an unbiased die be cast at random seven independent times. Compute the conditional probability that each side appears at least once given that side 1 appears exactly twice.

3.1.18

$$\binom{5}{5}\left(\frac{1}{2}\right)^5 \Big/ \left[\binom{5}{4}\left(\frac{1}{2}\right)^5 + \binom{5}{5}\left(\frac{1}{2}\right)^5\right] = \frac{1}{6},$$

which is much different than 1/2 that some might have arrived at by letting 4 coins be heads and tossing the fifth coin.

3.1.19

$$\left[\frac{7!}{2!1!\cdots 1!}\left(\frac{1}{6}\right)^7\right] \Big/ \left[\frac{7!}{2!5!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^5\right] = \frac{5!}{1!\cdots 1!}\left(\frac{1}{5}\right)^5.$$

2

3.2.1. If the random variable X has a Poisson distribution such that $P(X=1) = P(X=2)$, find $P(X=4)$.

3.2.1

$$\frac{e^{-\mu}\mu}{1!} = \frac{e^{-\mu}\mu^2}{2!} \Rightarrow \mu = 2 \text{ and } P(X=4) = \tfrac{e^{-2}2^4}{4!}.$$

3.2.4. Let the pmf $p(x)$ be positive on and only on the nonnegative integers. Given that $p(x) = (4/x)p(x-1)$, $x = 1, 2, 3, \ldots$, find the formula for $p(x)$.
Hint: Note that $p(1) = 4p(0)$, $p(2) = (4^2/2!)p(0)$, and so on. That is, find each $p(x)$ in terms of $p(0)$ and then determine $p(0)$ from

$$1 = p(0) + p(1) + p(2) + \cdots.$$

3.2.4 Given $p(x) = 4p(x-1)/x$, $x = 1, 2, 3, \ldots$. Thus $p(1) = 4p(0)$, $p(2) = 4^2p(0)/2!$, $p(3) = 4^3p(0)/3!$. Use induction to show that $p(x) = 4^x p(0)/x!$. Then

$$1 = \sum_{x=0}^{\infty} p(x) = p(0)\sum_{x=0}^{\infty} 4^x/x! = p(0)e^4 \text{ and } p(x) = 4^x e^{-4}/x!, x = 0, 1, 2, \ldots.$$

3.2.12. Let X have a Poisson distribution with mean 1. Compute, if it exists, the expected value $E(X!)$.

3.2.12

$$E(X!) = \textstyle\sum_{x=0}^{\infty} x!\frac{e^{-1}}{x!} = \sum_{x=0}^{\infty} e^{-1} \text{ does not exist.}$$

3

3.3.9. Let X have a gamma distribution with parameters α and β. Show that $P(X \geq 2\alpha\beta) \leq (2/e)^\alpha$.
Hint: Use the result of Exercise 1.10.4.

3.3.9

$$P(X \geq 2\alpha\beta) \leq e^{-2\alpha\beta t}(1-\beta t)^{-\alpha},$$

for all $t < 1/\beta$. The minimum of the right side, say $K(t)$, can be found by

$$K'(t) = e^{-2\alpha\beta t}(\alpha\beta)(1-\beta t)^{-\alpha-1} + e^{-2\alpha\beta t}(-2\alpha\beta)(1-\beta t)^{-\alpha} = 0$$

which implies that

$$(1-\beta t)^{-1} - 2 = 0 \text{ and } t = 1/2\beta.$$

That minimum is

$$K(1/2\beta) = e^{-\alpha}(1-(1/2))^{-\alpha} = (2/e)^\alpha.$$

3.3.15. Let X have a Poisson distribution with parameter m. If m is an experimental value of a random variable having a gamma distribution with $\alpha = 2$ and $\beta = 1$, compute $P(X = 0, 1, 2)$.
Hint: Find an expression that represents the joint distribution of X and m. Then integrate out m to find the marginal distribution of X.

3.3.15 The joint pdf of X and the parameter is

$$\begin{aligned} f(x|m)g(m) &= \frac{e^{-m}m^x}{x!}me^{-m}, \quad x = 0,1,2,\ldots, \quad 0 < m < \infty \\ P(X = 0,1,2) &= \sum_{x=0}^{2}\int_0^\infty \frac{m^{x+1}e^{-2m}}{x!}\,dm = \sum_{x=0}^{2}\frac{\Gamma(x+2)(1/2)^{x+2}}{x!} \\ &= \sum_{x=0}^{2}(x+1)(1/2)^{x+2} = \frac{1}{4} + \frac{2}{8} + \frac{3}{16} = \frac{11}{16}. \end{aligned}$$

3.3.16. Let X have the uniform distribution with pdf $f(x) = 1$, $0 < x < 1$, zero elsewhere. Find the cdf of $Y = -2\log X$. What is the pdf of Y?

3.3.16

$$\begin{aligned} G(y) &= P(Y \leq y) = P(-2\log X \leq y) = P(X \geq \exp\{-y/2\}) \\ &= \int_{\exp\{-y/2\}}^{1}(1)\,dx = 1 - \exp\{-y/2\}, \quad 0 < y < \infty \\ g(y) &= G'(y) = (1/2)\exp\{-y/2\}, \quad 0 < y < \infty; \end{aligned}$$

so Y is $\chi^2(2)$.

4

3.4.1. If
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw,$$
show that $\Phi(-z) = 1 - \Phi(z)$.

3.4.1 In the integral for $\Phi(-z)$, let $w = -v$ and it follows that $\Phi(-z) = 1 - \Phi(z)$.

3.4.4. Let X be $N(\mu, \sigma^2)$ so that $P(X < 89) = 0.90$ and $P(X < 94) = 0.95$. Find μ and σ^2.

3.4.5. Show that the constant c can be selected so that $f(x) = c2^{-x^2}$, $-\infty < x < \infty$, satisfies the conditions of a normal pdf.
Hint: Write $2 = e^{\log 2}$.

3.4.4
$$\begin{aligned} P\left(\frac{X-\mu}{\sigma} < \frac{89-\mu}{\sigma}\right) &= 0.90 \\ P\left(\frac{X-\mu}{\sigma} < \frac{94-\mu}{\sigma}\right) &= 0.95. \end{aligned}$$
Thus $\frac{89-\mu}{\sigma} = 1.282$ and $\frac{94-\mu}{\sigma} = 1.645$. Solve for μ and σ.

3.4.5
$$c2^{-x^2} = ce^{-x^2 \log\, 2} = c \exp\left\{-\frac{(2\log\, 2)x^2}{2}\right\}.$$
Thus if $c = 1/[\sqrt{2\pi}\sqrt{1/(2\log\, 2)}]$, we would have a $N(0, 1/(2\log\, 2))$ distribution.

3.4.12. Let X be $N(5, 10)$. Find $P[0.04 < (X-5)^2 < 38.4]$.

3.4.13. If X is $N(1, 4)$, compute the probability $P(1 < X^2 < 9)$.

3.4.12
$$P\left[0.0004 < \frac{(X-5)^2}{10} < 3.84\right] \text{ and } \frac{(X-5)^2}{10} \text{ is } \chi^2(1),$$
so, the answer is $0.95 - 0.05 = 0.90$.

3.4.13
$$\begin{aligned} P(1 < X^2 < 9) &= p(-3 < X < -1) + P(1 < X < 3) \\ &= \left[\Phi\left(\frac{-1-1}{2}\right) - \Phi\left(\frac{-3-1}{2}\right)\right] + \left[\Phi\left(\frac{3-1}{2}\right) - \Phi\,(0)\right]. \end{aligned}$$

3.4.22. Let $f(x)$ and $F(x)$ be the pdf and the cdf, respectively, of a distribution of the continuous type such that $f'(x)$ exists for all x. Let the mean of the truncated distribution that has pdf $g(y) = f(y)/F(b)$, $-\infty < y < b$, zero elsewhere, be equal to $-f(b)/F(b)$ for all real b. Prove that $f(x)$ is a pdf of a standard normal distribution.

3.4.22

$$\int_{-\infty}^{b} yf(y)/F(b)\,dy = -f(b)/F(b).$$

Multiply both sides by $F(b)$ then differentiate both sides with respect to b. This yields,

$$bf(b) = f'(b) \text{ and } -(b^2/2) + c = \log\, f(b).$$

Thus

$$f(b) = c_1 e^{-b^2/2},$$

which is the pdf of a $N(0,1)$ distribution.

5

3.5.5. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 5$, $\mu_2 = 10$, $\sigma_1^2 = 1$, $\sigma_2^2 = 25$, and $\rho > 0$. If $P(4 < Y < 16|X = 5) = 0.954$, determine ρ.

3.5.5 Because $E(Y|x = 5) = 10 + \rho(5/1)(5-5) = 10$, this probability requires that

$$\tfrac{16-10}{5\sqrt{1-\rho^2}} = 2,\ \tfrac{9}{25} = 1 - \rho^2, \text{ and } \rho = \tfrac{4}{5}.$$

3.5.8. Let

$$f(x,y) = (1/2\pi)\exp\left[-\frac{1}{2}(x^2+y^2)\right]\left\{1 + xy\exp\left[-\frac{1}{2}(x^2+y^2-2)\right]\right\},$$

where $-\infty < x < \infty$, $-\infty < y < \infty$. If $f(x,y)$ is a joint pdf, it is not a normal bivariate pdf. Show that $f(x,y)$ actually is a joint pdf and that each marginal pdf is normal. Thus the fact that each marginal pdf is normal does not imply that the joint pdf is bivariate normal.

3.5.9. Let X, Y, and Z have the joint pdf

$$\left(\frac{1}{2\pi}\right)^{3/2}\exp\left(-\frac{x^2+y^2+z^2}{2}\right)\left[1 + xyz\exp\left(-\frac{x^2+y^2+z^2}{2}\right)\right],$$

where $-\infty < x < \infty$, $-\infty < y < \infty$, and $-\infty < z < \infty$. While X, Y, and Z are obviously dependent, show that X, Y, and Z are pairwise independent and that each pair has a bivariate normal distribution.

3.5.8 $f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = (1/\sqrt{2\pi})\exp\{-x^2/2\}$, because the first term of the integral is obviously equal to the latter expression and the second term integrates to zero as it is an odd function of y. Likewise

$$f_2(y) = \frac{1}{\sqrt{2\pi}}\exp\{-y^2/2\}.$$

Of course, each of these marginal standard normal densities integrates to one.

3.5.9 Similar to 3.5.8 as the second term of

$$\int_{-\infty}^{\infty} f(x,y,z)\,dx$$

equals zero because it is an integral of an odd function of x.

6

3.6.8. Let F have an F-distribution with parameters r_1 and r_2. Argue that $1/F$ has an F-distribution with parameters r_2 and r_1.

3.6.8 Since $F = \frac{U/r_1}{V/r_2}$, then $\frac{1}{F} = \frac{V/r_2}{U/r_1}$, which has an F-distribution with r_2 and r_1 degrees of freedom.

3.6.12. Show that

$$Y = \frac{1}{1 + (r_1/r_2)W},$$

where W has an F-distribution with parameters r_1 and r_2, has a beta distribution.

3.6.13. Let X_1, X_2 be iid with common distribution having the pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that $Z = X_1/X_2$ has an F-distribution.

3.6.12 The change-of-variable technique can be used. An alternative method is to observe that

$$Y = \frac{1}{1 + (U/V)} = \frac{V}{V + U},$$

where V and U are independent gamma variables with respective parameters $(r_2/2, 2)$ and $(r_1/2, 2)$. Hence, Y is beta with $\alpha = r_2/2$ and $\beta = r_1/2$.

3.6.13 Note that the distribution of X_i is $\Gamma(1, 1)$. It follows that the mgf of $Y_i = 2X_i$ is

$$M_{Y_i}(t) = (1 - 2t)^{-2/2}, \quad t < 1/2.$$

Hence $2X_i$ is distributed as $\chi^2(2)$. Since X_1 and X_2 are independent, we have that

$$\frac{X_1}{X_2} = \frac{2X_1/2}{2X_2/2}$$

has an F-distribution with $\nu_1 = 2$ and $\nu_2 = 2$ degrees of freedom.

7

3.7.3. Consider the mixture distribution, $(9/10)N(0, 1) + (1/10)N(0, 9)$. Show that its kurtosis is 8.34.

3.7.4. Let X have the conditional geometric pmf $\theta(1 - \theta)^{x-1}$, $x = 1, 2, \ldots$, where θ is a value of a random variable having a beta pdf with parameters α and β. Show that the marginal (unconditional) pmf of X is

$$\frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)\Gamma(\beta + x - 1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + x)}, \quad x = 1, 2, \ldots.$$

If $\alpha = 1$, we obtain

$$\frac{\beta}{(\beta + x)(\beta + x - 1)}, \quad x = 1, 2, \ldots,$$

which is one form of **Zipf's law**.

3.7.3 Recall from Section 3.4, that we can write the random variable of interest as

$$X = IZ + 3(1 - I)Z,$$

where Z has a $N(0, 1)$ distribution, I is 0 or 1 with probabilities 0.1 and 0.9, respectively, and I and Z are independent. Note that $E(X) = 0$ and the variance of X is given by expression (3.4.13); hence, for the kurtosis we only need the fourth moment. Because I is 0 or 1, $I^k = I$ for all positive integers k. Also $I(I - 1) = 0$. Using these facts, we see that

$$E(X^4) = .9E(Z^4) + 3^4(.1)E(Z^4) = E(Z^4)(.9 + (.i)3^4).$$

Use expression (1.9.1) to get $E(Z^4)$.

3.7.4 The joint pdf is

$$f_{X,\theta}(x, \theta) = \theta(1 - \theta)^{x-1}\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1 - \theta)^{\beta-1}.$$

Integrating out θ, we have

$$\begin{aligned} f_X(x) &= \int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+1-1}(1 - \theta)^{\beta+x-1-1}\, d\theta \\ &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)\Gamma(\beta + x - 1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + x)}. \end{aligned}$$

Some Elementary Statistical Inferences

1

4.1.5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous-type distribution.

(a) Find $P(X_1 \le X_2), P(X_1 \le X_2, X_1 \le X_3), \ldots, P(X_1 \le X_i, i = 2, 3, \ldots, n)$.

(b) Suppose the sampling continues until X_1 is no longer the smallest observation (i.e., $X_j < X_1 \le X_i, i = 2, 3, \ldots, j-1$). Let Y equal the number of trials, not including X_1, until X_1 is no longer the smallest observation (i.e., $Y = j - 1$). Show that the distribution of Y is

$$P(Y = y) = \frac{1}{y(y+1)}, \quad y = 1, 2, 3, \ldots.$$

4.1.5 Parts (a) and (b).

Part (a). Using conditional expectation we have

$$\begin{aligned} P(X_1 \le X_i, i = 2, 3, \ldots, j) &= E[P(X_1 \le X_i, i = 2, 3, \ldots, j|X_1)] \\ &= E[(1 - F(X_1))^{j-1}] \\ &= \int_0^1 u^{j-1}\, du = j^{-1}, \end{aligned}$$

where we used the fact that the random variable $F(X_1)$ has a uniform$(0, 1)$ distribution.

Part (b). In the same way, for $j = 2, 3, \ldots$

$$\begin{aligned} P(Y = j - 1) &= P(X_1 \le X_2, \ldots, X_1 \le X_{j-1}, X_j > X_1) \\ &= E[(1 - F(X_1))^{j-2}F(X_1)] = \int_0^1 u^{j-2}(1 - u)\, du \\ &= \frac{1}{j(j-1)}. \end{aligned}$$

4.1.8. Recall that for the parameter $\eta = g(\theta)$, the mle of η is $g(\widehat{\theta})$, where $\widehat{\theta}$ is the mle of θ. Assuming that the data in Example 4.1.6 were drawn from a Poisson distribution with mean λ, obtain the mle of λ and then use it to obtain the mle of the pmf. Compare the mle of the pmf to the nonparametric estimate. Note: For the domain value 6, obtain the mle of $P(X \geq 6)$.

4.1.8 If $X_1, \ldots, X_n$ are iid with a Poisson distribution having mean λ, then the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Taking the partial of the log of this likelihood function leads to $\overline{x}$ as the mle of λ. Hence, the mle of the pmf at k is

$$\widehat{p(k)} = e^{-\overline{x}} \frac{\overline{x}^k}{k!}$$

and the mle of $P(X \geq 6)$ is

$$\widehat{P(X \geq 6)} = e^{-\overline{x}} \sum_{k=6}^{\infty} \frac{\overline{x}^k}{k!}.$$

For the data set of this problem, we obtain $\overline{x} = 2.1333$. Using R, the mle of $P(X \geq 6)$ is `1 - ppois(5, 2.1333) = 0.0219`. Note, for comparison, from the tabled data, that the nonparametric estimate of this probability is 0.033.

2

4.2.10. Let $X_1, X_2, \ldots, X_9$ be a random sample of size 9 from a distribution that is $N(\mu, \sigma^2)$.

(a) If σ is known, find the length of a 95% confidence interval for μ if this interval is based on the random variable $\sqrt{9}(\overline{X} - \mu)/\sigma$.

(b) If σ is unknown, find the expected value of the length of a 95% confidence interval for μ if this interval is based on the random variable $\sqrt{9}(\overline{X} - \mu)/S$. *Hint:* Write $E(S) = (\sigma/\sqrt{n-1})E[((n-1)S^2/\sigma^2)^{1/2}]$.

4.2.10 (a). $\overline{X} \pm 1.96\sigma/\sqrt{9}$, length $= (2)(1.96)\sigma/3 = 1.31\sigma$.

(b). $\overline{X} \pm 2.306S/\sqrt{8}$, length $= (2)(2.306)S/\sqrt{8}$. Since

$$\begin{aligned}
E(S) &= (\sigma/\sqrt{n}) \int_0^\infty w^{1/2} \frac{w^{4-1}e^{-w/2}}{\Gamma(4)2^4}\, dw \\
&= (\sigma/\sqrt{9}) \frac{\Gamma(9/2)2^{9/2}}{\Gamma(4)2^4} = \frac{\sigma(7/2)(5/2)(3/2)(1/2)\Gamma(1/2)\sqrt{2}}{3 \cdot 3 \cdot 2 \cdot 1} \\
&= \frac{35\sqrt{2\pi}\sigma}{(6)(16)} = (0.914)\sigma, \\
&= E(\text{length}) = \left[(2)(2.306)(0.914)/\sqrt{8}\right] \sigma = 1.49\sigma.
\end{aligned}$$

4.2.11. Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be a random sample of size $n+1$, $n > 1$, from a distribution that is $N(\mu, \sigma^2)$. Let $\overline{X} = \sum_1^n X_i/n$ and $S^2 = \sum_1^n (X_i - \overline{X})^2/(n-1)$. Find the constant c so that the statistic $c(\overline{X} - X_{n+1})/S$ has a t-distribution. If $n = 8$, determine k such that $P(\overline{X} - kS < X_9 < \overline{X} + kS) = 0.80$. The observed interval $(\overline{x} - ks,\ \overline{x} + ks)$ is often called an 80% **prediction interval** for X_9.

4.2.11 $\frac{(\overline{X}-X_{n+1})/\sqrt{\sigma^2/n+\sigma^2}}{\sqrt{(nS^2/\sigma^2)/(n-1)}} = \sqrt{\frac{n-1}{n+1}}\frac{\overline{X}-X_{n+1}}{S}$ is $T(n-1)$.

$P(-1.415 < \sqrt{\frac{7}{9}}\left(\frac{\overline{X}-X_{n+1}}{S}\right) < 1.415) = 0.80$, or equivalently,

$P(\overline{X} - 1.415\sqrt{9/7}S < X_{n+1} < \overline{X} + 1.415\sqrt{9/7}S) = 0.80$

4.2.19. Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution with known parameter $\alpha = 3$ and unknown $\beta > 0$. Discuss the construction of a confidence interval for β.
Hint: What is the distribution of $2\sum_1^n X_i/\beta$? Follow the procedure outlined in Exercise 4.2.18.

4.2.19 $E[\exp\{t(2X/\beta)\}] = [1 - \beta(2t/\beta)]^{-3} = (1-2t)^{-6/2}$.
Since $2X/\beta$ is $\chi^2(6), 2\sum X_i/\beta$ is $\chi^2(6n)$. Using tables for $\chi^2(6n)$, find a and b such that

$$P\left(a < 2\sum X_i/\beta < b\right) = 0.95$$

or, equivalently,

$$P\left(\frac{2\sum X_i}{b} < \beta < \frac{2\sum X_i}{a}\right) = 0.95.$$

4.2.26. Let $\overline{X}$ and $\overline{Y}$ be the means of two independent random samples, each of size n, from the respective distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, where the common variance is known. Find n such that

$$P(\overline{X} - \overline{Y} - \sigma/5 < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + \sigma/5) = 0.90.$$

4.2.26 The distribution of $\overline{X}$ is $N(\mu_1, \sigma^2/n)$ and the distribution of $\overline{Y}$ is $N(\mu_2, \sigma^2/n)$. Because the samples are independent the distribution of $\overline{X} - \overline{Y}$ is $N(\mu_1 - \mu_2, 2\sigma^2/n)$. After some algebra, the equation to solve for n can be written as

$$P\left[\left|\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma/\sqrt{n}}\right| < \frac{\sqrt{n}}{5}\right] = 0.90,$$

which is equivalent to

$$P\left[|Z| < \frac{\sqrt{n}}{5}\right] = 0.90,$$

where Z has a $N(0,1)$ distribution. Hence, $\sqrt{n}/5 = 1.645$ or $n = 67.65$, i.e., $n = 68$.

3

4.4.5. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size 4 from the distribution having pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Find $P(Y_4 \geq 3)$.

4.4.5 The cdf of the Y_4 is

$$P(Y_4 \leq t) = (1 - e^{-t})^4, \quad t > 0.$$

Hence, $P(Y_4 \geq 3) = 1 - (1 - e^{-3})^4 = 0.1848$.

4.4.11. Find the probability that the range of a random sample of size 4 from the uniform distribution having the pdf $f(x) = 1$, $0 < x < 1$, zero elsewhere, is less than $\frac{1}{2}$.

4.4.11 The distribution of the range $Y_4 - Y_1$ could be found. An alternative method is

$$P(Y_4 - Y_1 < 1/2) = 1 - \int_0^{1/2} \int_{y_1+1/2}^{1} 12(y_4 - y_1)^2 \, dy_4 \, dy_1.$$

4.4.12. Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from a distribution having the pdf $f(x) = 2x$, $0 < x < 1$, zero elsewhere. Show that $Z_1 = Y_1/Y_2$, $Z_2 = Y_2/Y_3$, and $Z_3 = Y_3$ are mutually independent.

4.4.12 $y_1 = z_1 z_2 z_3$, $y_2 = z_2 z_3$, $y_3 = z_3$, with $J = z_2 z_3^2$, $0 < z_1 < 1$, $0 < z_2 < 1$, $0 < z_3 < 1$. Accordingly,

$$\begin{aligned} g(z_1, z_2, z_3) &= 3!\, 2(z_1 z_2 z_3)\, 2(z_2 z_3)\, 2(z_3) z_2 z_3^2 \\ &= (2z_1)(4z_2^3)(6z_3^5), \quad 0 < z_i < 1, \quad i = 1, 2, 3. \end{aligned}$$

4.4.17. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size $n = 4$ from a distribution with pdf $f(x) = 2x$, $0 < x < 1$, zero elsewhere.

(a) Find the joint pdf of Y_3 and Y_4.

(b) Find the conditional pdf of Y_3, given $Y_4 = y_4$.

(c) Evaluate $E(Y_3|y_4)$.

4.4.17

$$\begin{aligned} F(x) &= x^2, \ 0 \le x < 1. \\ g_{34}(y_3, y_4) &= \frac{4!}{2!}(y_3^2)^2(2y_3)(2y_4), \ 0 < y_3 < y_4 < 1. \\ g_4(y_4) &= 4(y_4^2)^3(2y_4) = 8y_4^7, \ 0 < y_4 < 1. \\ g_{3|4}(y_3|y_4) &= 6y_3^5/y_4^6, \ 0 < y_3 < y_4. \\ E(Y_3|y_4) &= (6/7)y_4. \end{aligned}$$

4.4.24. Let Y_n denote the nth order statistic of a random sample of size n from a distribution of the continuous type. Find the smallest value of n for which the inequality $P(\xi_{0.9} < Y_n) \ge 0.75$ is true.

4.4.24 Let $F(x)$ denote the common cdf of the sample. Then $\xi_{0.9} = F^{-1}(0.9)$. The solution to the desired inequality is

$$\begin{aligned} 1 - (F(\xi_{0.9}))^n &\ge 0.75 \\ 1 - F(F^{-1}(0.9)))^n &\ge \frac{3}{4} \\ 1 - 0.9^n &\ge \frac{3}{4} \\ n\log(0.9) &\le \frac{1}{4} \\ n &\ge -\frac{\log(4)}{\log(0.9)} = 13.14. \end{aligned}$$

Hence, take $n = 14$.

5

4.5.3. Let X have a pdf of the form $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta \in \{\theta : \theta = 1, 2\}$. To test the simple hypothesis $H_0 : \theta = 1$ against the alternative simple hypothesis $H_1 : \theta = 2$, use a random sample X_1, X_2 of size $n = 2$ and define the critical region to be $C = \{(x_1, x_2) : \frac{3}{4} \le x_1 x_2\}$. Find the power function of the test.

4.5.3 For a general θ the probability of rejecting H_0 is

$$\gamma(\theta) = \int_{3/4}^{1}\int_{3/4x_1}^{1} \theta^2 (x_1x_2)^{\theta-1}\, dx_2\, dx_1 = 1 - \left(\frac{3}{4}\right)^{\theta} + \theta\left(\frac{3}{4}\right)^{\theta} \log\left(\frac{3}{4}\right)$$

$\gamma(1)$ is the significance level and $\gamma(2)$ is the power when $\theta = 2$.

4.5.8. Let us say the life of a tire in miles, say X, is normally distributed with mean θ and standard deviation 5000. Past experience indicates that $\theta = 30{,}000$. The manufacturer claims that the tires made by a new process have mean $\theta > 30{,}000$. It is possible that $\theta = 35{,}000$. Check his claim by testing $H_0 : \theta = 30{,}000$ against $H_1 : \theta > 30{,}000$. We observe n independent values of X, say $x_1, \ldots, x_n$, and we reject H_0 (thus accept H_1) if and only if $\overline{x} \geq c$. Determine n and c so that the power function $\gamma(\theta)$ of the test has the values $\gamma(30{,}000) = 0.01$ and $\gamma(35{,}000) = 0.98$.

4.5.8

$$\begin{aligned}\gamma(\theta) &= P(\overline{X} \geq c; \theta) = P\left(\frac{\overline{X} - \theta}{5000/\sqrt{n}} \geq \frac{c - \theta}{5000/\sqrt{n}}; \theta\right) \\ &= 1 - \Phi\left(\frac{c - \theta}{5000/\sqrt{n}}\right).\end{aligned}$$

Thus, solve for n and c knowing that

$$\frac{c - 30000}{5000/\sqrt{n}} = 2.325 \text{ and } \frac{c - 35000}{5000/\sqrt{n}} = -2.05.$$

4.5.12. Let $X_1, X_2, \ldots, X_8$ be a random sample of size $n = 8$ from a Poisson distribution with mean μ. Reject the simple null hypothesis $H_0 : \mu = 0.5$ and accept $H_1 : \mu > 0.5$ if the observed sum $\sum_{i=1}^{8} x_i \geq 8$.

(a) Compute the significance level α of the test.

(b) Find the power function $\gamma(\mu)$ of the test as a sum of Poisson probabilities.

(c) Using Table I of Appendix C, determine $\gamma(0.75)$, $\gamma(1)$, and $\gamma(1.25)$.

4.5.12 Let $Y = \sum_{i=1}^{8} X_i$. Then Y has a Poisson(8μ) distribution.

Part (a). The significance level of the test is

$$\alpha = P_{H_0}[Y \geq 8] = P[\text{Poisson}(4) \geq 8] = 0.051.$$

Part (b). The power function is

$$\gamma(\mu) = P_\mu[Y \geq 8] = P[\text{Poisson}(8\mu) \geq 8].$$

Part (c). $\gamma(0.75) = 0.256$.

6

4.6.2. Consider the power function $\gamma(\mu)$ and its derivative $\gamma'(\mu)$ given by (4.6.5) and (4.6.6). Show that $\gamma'(\mu)$ is strictly negative for $\mu < \mu_0$ and strictly positive for $\mu > \mu_0$.

4.6.2 Suppose $\mu > \mu_0$. Then

$$\left|\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}+z_{\alpha/2}\right| < \left|\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}-z_{\alpha/2}\right|.$$

Hence,

$$\phi\left(\left|\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}+z_{\alpha/2}\right|\right) > \phi\left(\left|\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}-z_{\alpha/2}\right|\right).$$

Because $\phi(t)$ is symmetric about 0, $\phi(t) = \phi(|t|)$. This observation plus the last inequality shows that $\gamma'(\mu)$ is increasing, (for $\mu > \mu_0$). Likewise for $\mu < \mu_0$, $\gamma'(\mu)$ is decreasing.

4.6.5. Assume that the weight of cereal in a "10-ounce box" is $N(\mu, \sigma^2)$. To test $H_0 : \mu = 10.1$ against $H_1 : \mu > 10.1$, we take a random sample of size $n = 16$ and observe that $\overline{x} = 10.4$ and $s = 0.4$.

(a) Do we accept or reject H_0 at the 5% significance level?

(b) What is the approximate p-value of this test?

4.6.5 (a). The critical region is

$$t = \frac{\overline{x}-10.1}{s/\sqrt{15}} \geq 1.753.$$

The observed value of t,

$$t = \frac{10.4-10.1}{0.4/\sqrt{15}} = 2.90,$$

is greater than 1.753 so we reject H_0.

(b). Since $t_{0.005}(15) = 2.947$ (from other tables), the approximate p-value of this test is 0.005.

4.6.7. Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles in μg/m^3. Let X and Y equal the concentration of suspended particles in μg/m^3 in the city center (commercial district) for Melbourne and Houston, respectively. Using $n = 13$ observations of X and $m = 16$ observations of Y, we test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$.

(a) Define the test statistic and critical region, assuming that the unknown variances are equal. Let $\alpha = 0.05$.

(b) If $\overline{x} = 72.9$, $s_x = 25.6$, $\overline{y} = 81.7$, and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.

4.6.7 Assume that Y and Y are normally distributed. Then the t-statistic

$$t = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{(1/n_1) + (1/n_2)}}$$

has under H_0 a t-distribution with $n_1 + n_2 - 2$ degrees of freedom. A level α test for the alternative $H_A: \ \mu_1 < \mu_2$ is

Reject H_0 in favor of H_A, if $t < -t_{\alpha,n_1+n_2-2}$.

For Part (b), based on the data we have,

$$\begin{aligned} s_p^2 &= \frac{(13-1)25.6^2 + (16-1)28.3^2}{27} \\ s_p &= \sqrt{s_p^2} = 27.133 \\ t &= \frac{72.9 - 81.7}{27.133\sqrt{(1/13) + (1/16)}} = -0.8685. \end{aligned}$$

Since $t = -0.8685 \not< -t_{.05,27} = -1.703$, we fail to reject H_0 at level 0.05. The p-value is $P[t(27) < -0.8685] = 0.1964$.

7

4.7.1. A number is to be selected from the interval $\{x : 0 < x < 2\}$ by a random process. Let $A_i = \{x : (i-1)/2 < x \leq i/2\}$, $i = 1,2,3$, and let $A_4 = \{x : \frac{3}{2} < x < 2\}$. For $i = 1,2,3,4$, suppose a certain hypothesis assigns probabilities p_{i0} to these sets in accordance with $p_{i0} = \int_{A_i} (\frac{1}{2})(2-x)\,dx$, $i = 1,2,3,4$. This hypothesis (concerning the multinomial pdf with $k = 4$) is to be tested at the 5% level of significance by a chi-square test. If the observed frequencies of the sets A_i, $i = 1,2,3,4$, are respectively, 30, 30, 10, 10, would H_0 be accepted at the (approximate) 5% level of significance?

4.7.1 $p_{10} = \int_0^{1/2} \frac{2-x}{2}\,dx = \frac{1}{2} - \frac{1}{16} = \frac{7}{16}$.
Likewise $p_{20} = 5/16, p_{30} = 3/16, p_{40} = 1/16$.
$Q_3 = \frac{(30-35)^2}{35} + \frac{(30-25)^2}{25} + \frac{(10-15)^2}{15} + \frac{(10-15)^2}{5} = 8.38$.
However, $8.38 > 7.81$ so we reject H_0 at $\alpha = 0.05$.

4.7.3. A die was cast $n = 120$ independent times and the following data resulted:

Spots Up	1	2	3	4	5	6
Frequency	b	20	20	20	20	$40-b$

If we use a chi-square test, for what values of b would the hypothesis that the die is unbiased be rejected at the 0.025 significance level?

4.7.3 $Q_5 = \frac{(b-20)^2}{20} + \frac{(40-b-20)^2}{20} = \frac{(b-20)^2}{10} = 12.8$,
which is the 97.5 percentile of a $\chi^2(5)$ distribution. Thus $(b-20)^2 = 128$ and $b = 20 \pm 11.3$. Hence $b < 8.7$ or $b > 31.3$ would lead to rejection.

4.7.7. A certain genetic model suggests that the probabilities of a particular trinomial distribution are, respectively, $p_1 = p^2$, $p_2 = 2p(1-p)$, and $p_3 = (1-p)^2$, where $0 < p < 1$. If X_1, X_2, X_3 represent the respective frequencies in n independent trials, explain how we could check on the adequacy of the genetic model.

4.7.7 The maximum likelihood statistic for p is defined by that value of p which maximizes

$$\frac{n!}{x_1!x_2!x_3!}[p^2]^{x_1}[2p(1-p)]^{x_2}[(1-p)^2]^{x_3};$$

it is $\hat{p} = (2X_1 + X_2)/(2X_1 + 2X_2 + 2X_3)$. Thus if $\hat{p}_1 = \hat{p}^2$, $\hat{p}_2 = 2\hat{p}(1-\hat{p})$, and $\hat{p}_3 = (1-\hat{p})^2$, the random variable $\sum_1^3 (X_i - n\hat{p}_i)^2/n\hat{p}_i$ has an approximate chi-square distribution with $3 - 1 - 1 = 1$ degree of freedom.