5.5 The Expected Value of a Function of Random Variables

The material in this section is combined with the material in the next section and given in the next section.

5.6 Special Theorems

- Expected value of a function of random variables, $E[g(Y_1, Y_2, \ldots, Y_k)]$, equals

$$= \begin{cases} \sum_{\text{all } y_1} \sum_{\text{all } y_2} \cdots \sum_{\text{all } y_k} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k) & \text{if discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) f(y_1, y_2, \ldots, y_k)\, dy_1 dy_2 \cdots dy_k & \text{if continuous.} \end{cases}$$

- Also, the *variance* of a function of random variables, $V[g(Y_1, Y_2, \ldots, Y_k)]$, with expected value, $E[g(Y_1, Y_2, \ldots, Y_k)]$, is defined

$$\begin{aligned} V[g(Y_1, Y_2, \ldots, Y_k)] &= E\left[(g(Y_1, Y_2, \ldots, Y_k) - E[g(Y_1, Y_2, \ldots, Y_k)])^2\right] \\ &= E\left[g(Y_1, Y_2, \ldots, Y_k)^2\right] - \left[E\left[g(Y_1, Y_2, \ldots, Y_k)\right]\right]^2. \end{aligned}$$

- Related to this,

$$\begin{aligned} E[Y_1 + Y_2 + \cdots + Y_k] &= E[Y_1] + \cdots + E[Y_k], \\ E[g_1(Y_1, Y_2) + \cdots + g_k(Y_1, Y_2)] &= E[g_1(Y_1, Y_2)] + \cdots + E[g_k(Y_1, Y_2)], \\ E(c) &= c, \quad \text{if } c \text{ is a constant,} \\ E[cg(Y_1, Y_2)] &= cE[g(Y_1, Y_2)], \quad \text{if } c \text{ is a constant,} \\ E[g(Y_1)h(Y_2)] &= E[g(Y_1)]E[h(Y_2)], \quad \text{if } g(Y_1),\ h(Y_2) \text{ independent.} \end{aligned}$$

Exercise 5.6 (Special Theorems)

1. *Discrete Expected Value Calculations: Waiting Times To Catch Fish.* The joint density, $p(y_1, y_2)$, of the number of minutes waiting to catch the *first* fish, y_1, *and* the number of minutes waiting to catch the *second* fish, y_2, is given below.

$y_2 \downarrow y_1 \rightarrow$	1	2	3	total
1	0.01	0.01	0.07	0.09
2	0.02	0.02	0.08	0.12
3	0.08	0.08	0.63	0.79
total	0.11	0.11	0.78	1.00

(a) The *expected* average waiting time, $g(y_1, y_2) = \frac{y_1+y_2}{2}$, over two trips is

$$\begin{aligned}
E\left[\frac{Y_1+Y_2}{2}\right] &= \sum_{y_1=1}^{3}\sum_{y_2=1}^{3}\left(\frac{y_1+y_2}{2}\right)p(y_1,y_2) \\
&= \left(\frac{1+1}{2}\right)(0.01)+\left(\frac{1+2}{2}\right)(0.02)+\left(\frac{1+3}{2}\right)(0.08) \\
&\quad + \left(\frac{2+1}{2}\right)(0.01)+\left(\frac{2+2}{2}\right)(0.02)+\left(\frac{2+3}{2}\right)(0.08) \\
&\quad + \left(\frac{3+1}{2}\right)(0.07)+\left(\frac{3+2}{2}\right)(0.08)+\left(\frac{3+3}{2}\right)(0.63) =
\end{aligned}$$

(choose one) (i) **2.385** (ii) **2.685** (iii) **2.785**.

(Hint: Type $1, 1, 1, 2, 2, 2, 3, 3, 3$ in L_1 and $1, 2, 3, 1, 2, 3, 1, 2, 3$ in L_2, define $L_3 = \frac{L_1+L_2}{2}$, type $0.01, 0.02, 0.08, 0.01, 0.02, 0.08, 0.07, 0.07, 0.63$ in L_4, define $L_5 = L_3 \times L_4$, sum L_5 using 1-Var Stats L_5, read $\sum x = 2.685$.)

(b) The expected *total* waiting time, $g(y_1, y_2) = y_1 + y_2$, over two trips is

$$\begin{aligned}
E[Y_1+Y_2] &= \sum_{y_1=1}^{3}\sum_{y_2=1}^{3}(y_1+y_2)\,p(y_1,y_2) \\
&= (1+1)(0.01)+(1+2)(0.02)+(1+3)(0.08) \\
&\quad + (2+1)(0.01)+(2+2)(0.02)+(2+3)(0.08) \\
&\quad + (3+1)(0.07)+(3+2)(0.08)+(3+3)(0.63) =
\end{aligned}$$

(choose one) (i) **5.37** (ii) **6.37** (iii) **7.37**.

(Hint: Define $L_3 = L_1 + L_2$, define $L_5 = L_3 \times L_4$, sum L_5 using 1-Var Stats L_5, read $\sum x = 5.37$.)

(c) Marginal probability function for Y_1 is given by column totals; in this case:

y_1	1	2	3
$p_1(y_1)$	0.11	0.11	0.78

so expected waiting time for *first* trip is

$$E[Y_1] = \sum_{y_1=1}^{3} y_1 p_1(y_1) = (1)(0.11) + (2)(0.11) + (3)(0.78) =$$

(choose one) (i) **1.67** (ii) **2.67** (iii) **3.67**.

(d) Marginal probability function for Y_2 is given by row totals; in this case:

y_2	1	2	3
$p_2(y_2)$	0.09	0.12	0.79

so expected waiting time for *second* trip is

$$E[Y_2] = \sum_{y_2=1}^{3} y_2 p_2(y_2) = (1)(0.09) + (2)(0.12) + (3)(0.79) =$$

(choose one) (i) **1.7** (ii) **2.7** (iii) **3.7**.

(e) Notice

$$E[Y_1 + Y_2] = 5.37 = E[Y_1] + E[Y_2] = 2.67 + 2.7$$

(choose one) (i) **True** (ii) **False**

(f) Since

$$E[Y_2^2] = \sum_{y_2=1}^{3} y_2^2 p_2(y_2) = (1^2)(0.09) + (2^2)(0.12) + (3^2)(0.79) = 7.68,$$

then *variance* in *second* waiting time is

$$V[Y_2] = E\left[Y_2^2\right] - \left[E[Y_2]\right]^2 = 7.68 - 2.7^2 =$$

(choose one) (i) **0.29** (ii) **0.39** (iii) **0.49**.

2. *Discrete Expectation Value Calculations: Marbles In An Urn.* Marbles chosen at random with*out* replacement from an urn consisting of 8 blue and 6 black marbles. For ith marble chosen, $Y_i = 0$ if blue and $Y_i = 1$ if black.

$y_2 \downarrow y_1 \rightarrow$	blue, 0	black, 1	$p_2(y_2)$
blue, 0	$\frac{8\cdot7}{14\cdot13}$	$\frac{6\cdot8}{14\cdot13}$	$\frac{8\cdot7+6\cdot8}{14\cdot13}$
black, 1	$\frac{8\cdot6}{14\cdot13}$	$\frac{6\cdot5}{14\cdot13}$	$\frac{8\cdot6+6\cdot5}{14\cdot13}$
$p_1(y_1)$	$\frac{8\cdot7+8\cdot6}{14\cdot13}$	$\frac{6\cdot8+6\cdot5}{14\cdot13}$	1

(a) The expected value of $g(y_1, y_2) = y_1 y_2$ is

$$\begin{aligned} E\left[Y_1 Y_2\right] &= \sum_{y_1=0}^{1} \sum_{y_2=0}^{1} (y_1 y_2)\, p(y_1, y_2) \\ &= (0 \times 0)\left(\frac{8 \cdot 7}{14 \cdot 13}\right) + (0 \times 1)\left(\frac{8 \cdot 6}{14 \cdot 13}\right) \\ &\quad + (1 \times 0)\left(\frac{8 \cdot 8}{14 \cdot 13}\right) + (1 \times 1)\left(\frac{6 \cdot 5}{14 \cdot 13}\right) = \end{aligned}$$

(choose one) (i) $\mathbf{\frac{8\cdot6}{14\cdot13}}$ (ii) $\mathbf{\frac{8\cdot7}{14\cdot13}}$ (iii) $\mathbf{\frac{6\cdot5}{14\cdot13}}$.

(b) Expected value of Y_1 is

$$E[Y_1] = \sum_{y_1=0}^{1} y_1 p_1(y_1) = (0)\left(\frac{8\cdot 7 + 8\cdot 6}{14\cdot 13}\right) + (1)\left(\frac{6\cdot 8 + 6\cdot 5}{14\cdot 13}\right) =$$

(choose one) (i) $\mathbf{\frac{8\cdot 7+8\cdot 6}{14\cdot 13}}$ (ii) $\mathbf{\frac{8\cdot 7+2\times 8\cdot 6}{14\cdot 13}}$ (iii) $\mathbf{\frac{6\cdot 8+6\cdot 5}{14\cdot 13}}$.

(c) Expected value of Y_2 is

$$E[Y_2] = \sum_{y_2=0}^{1} y_2 p_2(y_2) = (0)\left(\frac{8\cdot 7 + 6\cdot 8}{14\cdot 13}\right) + (1)\left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right) =$$

(choose one) (i) $\mathbf{\frac{8\cdot 7+8\cdot 6}{14\cdot 13}}$ (ii) $\mathbf{\frac{8\cdot 7+2\times 8\cdot 6}{14\cdot 13}}$ (iii) $\mathbf{\frac{8\cdot 6+6\cdot 5}{14\cdot 13}}$.
(Notice, in this case, $E[Y_1] = E[Y_2]$.)

(d) In this case,

$$E[Y_1Y_2] = \frac{6\cdot 5}{14\cdot 13} \neq E[Y_1]E[Y_2] = \left(\frac{6\cdot 8 + 6\cdot 5}{14\cdot 13}\right)\left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right)$$

because Y_1 and Y_2 are *dependent*.
(choose one) (i) **True** (ii) **False**

(e) Using the properties of expectation above,

$$E[3Y_1 - 2Y_2] = 3E[Y_1] - 2E[Y_2] = 3\left(\frac{6\cdot 8 + 6\cdot 5}{14\cdot 13}\right) - 2\left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right) =$$

(i) $\mathbf{\frac{8\cdot 7+8\cdot 6}{14\cdot 13}}$ (ii) $\mathbf{\frac{8\cdot 7+2\times 8\cdot 6}{14\cdot 13}}$ (iii) $\mathbf{\frac{8\cdot 6+6\cdot 5}{14\cdot 13}}$.

(f) Since

$$E[Y_2^2] = \sum_{y_2=0}^{0} y_2^2 p_2(y_2) = (0^2)\left(\frac{8\cdot 7 + 6\cdot 8}{14\cdot 13}\right) + (1^2)\left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right) = \frac{8\cdot 6 + 6\cdot 5}{14\cdot 13},$$

then the variance is

$$V[Y_2] = E\left[Y_2^2\right] - \left[E[Y_2]\right]^2 = \frac{8\cdot 6 + 6\cdot 5}{14\cdot 13} - \left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right)^2 \approx$$

(choose one) (i) **0.245** (ii) **0.345** (iii) **0.445**.

(g) The definition of variance is

$$\begin{aligned}
V[Y_2] &= E\left[(Y_2 - E[Y_2])])^2\right] \\
&= \sum_{y_2=0}^{1} (Y_2 - E[Y_2])^2 p_2(y_2) \\
&= \left(0 - \frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right)^2 \left(\frac{8\cdot 7 + 6\cdot 8}{14\cdot 13}\right) + \left(1 - \frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right)^2 \left(\frac{8\cdot 6 + 6\cdot 5}{14\cdot 13}\right) \approx
\end{aligned}$$

(choose one) (i) **0.245** (ii) **0.345** (iii) **0.445**.

3. *Continuous Expectation Value Calculations: Potato Chips.* Although each bag should weigh 50 grams each and contain 5 milligrams of salt, in fact, because of differing machines, weight and amount of salt placed in each bag varies according to two probability functions below.

   (a) *Machine A.* Bivariate density function for machine A is

$$f(y_1, y_2) = \begin{cases} \frac{1}{12}, & 49 \leq y_1 \leq 51, 2 \leq y_2 \leq 8 \\ 0 & \text{elsewhere} \end{cases}$$

   i. The expected value of $g(y_1, y_2) = y_1 y_2$ is

$$\begin{aligned} E\left[Y_1 Y_2\right] &= \int_2^8 \int_{49}^{51} (y_1 y_2)\, f(y_1, y_2)\, dy_1 dy_2 \\ &= \int_2^8 \int_{49}^{51} (y_1 y_2)\, \frac{1}{12}\, dy_1 dy_2 \\ &= \int_2^8 \frac{1}{12} y_2 \left(\frac{1}{2} y_1^2\right)_{y_1=49}^{y_1=51} dy_2 \\ &= \left(\frac{100}{24} y_2^2\right)_{y_1=2}^{y_1=8} = \end{aligned}$$

   (choose one) (i) **200** (ii) **250** (iii) **300**.

   ii. Since marginal,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 = \int_2^8 \frac{1}{12}\, dy_2 = \frac{1}{12} (y_2)_{y_2=2}^{y_2=8} = \frac{1}{2},$$

   $49 \leq y_1 \leq 51$, expected value of Y_1 is

$$E[Y_1] = \int_{49}^{51} y_1 f_1(y_1)\, dy_1 = \int_{49}^{51} y_1 \frac{1}{2}\, dy_1 = \left(\frac{1}{4} y_1^2\right)_{y_2=49}^{y_2=51} =$$

   (choose one) (i) **50** (ii) **100** (iii) **150**.

   iii. Since marginal,

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 = \int_{49}^{51} \frac{1}{12}\, dy_1 = \frac{1}{12} (y_1)_{y_1=49}^{y_1=51} = \frac{1}{6},$$

   $2 \leq y_2 \leq 8$, expected value of Y_2 is

$$E[Y_2] = \int_2^8 y_2 f_2(y_2)\, dy_2 = \int_2^8 y_2 \frac{1}{6}\, dy_2 = \left(\frac{1}{12} y_2^2\right)_{y_1=2}^{y_2=8} =$$

   (choose one) (i) **5** (ii) **10** (iii) **15**.

iv. Since

$$E[Y_1Y_2] = 250 = E[Y_1]E[Y_2] = (50)\,(5)$$

Y_1 and Y_2 are (choose one) (i) **dependent** (ii) **independent**.

v. Using the properties of expectation above,

$$E[3Y_1 - 2Y_2] = 3E[Y_1] - 2E[Y_2] = 3\,(50) - 2\,(5) =$$

(i) **130** (ii) **140** (iii) **150**.

vi. Since

$$E[Y_2^2] = \int_2^8 y_2^2 f_2(y_2)\,dy_2 = \int_2^8 y_2^2 \frac{1}{6}\,dy_2 = \left(\frac{1}{18}y_2^3\right)_{y_2=2}^{y_2=8} = 28,$$

then variance is

$$V[Y_2] = E\left[Y_2^2\right] - \left[E[Y_2]\right]^2 = 28 - (5)^2 \approx$$

(choose one) (i) **1** (ii) **2** (iii) **3**.

(b) *Machine B.* Bivariate density function for machine B is

$$f(y_1, y_2) = \begin{cases} \frac{1}{4}, & 49 \leq y_1 \leq 51, 4 \leq y_2 \leq 6 \\ 0 & \text{elsewhere} \end{cases}$$

i. The expected value of $g(y_1, y_2) = y_1y_2$ is

$$\begin{aligned} E\left[Y_1Y_2\right] &= \int_4^6 \int_{49}^{51} (y_1y_2)\, f(y_1, y_2)\, dy_1 dy_2 \\ &= \int_4^6 \int_{49}^{51} (y_1y_2)\, \frac{1}{4}\, dy_1 dy_2 \\ &= \int_4^6 \frac{1}{4}y_2 \left(\frac{1}{2}y_1^2\right)_{y_1=49}^{y_1=51} dy_2 \\ &= \left(\frac{100}{8}y_2^2\right)_{y_1=4}^{y_1=6} = \end{aligned}$$

(choose one) (i) $\mathbf{\frac{250}{3}}$ (ii) $\mathbf{\frac{500}{3}}$ (iii) $\mathbf{\frac{750}{3}}$.

ii. Since marginal,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 = \int_4^6 \frac{1}{4}\, dy_2 = \frac{1}{4}(y_2)_{y_2=4}^{y_2=6} = \frac{1}{2},$$

$49 \leq y_1 \leq 51$, expected value of Y_1 is

$$E[Y_1] = \int_{49}^{51} y_1 f_1(y_1)\, dy_1 = \int_{49}^{51} y_1 \frac{1}{2}\, dy_1 = \left(\frac{1}{4}y_1^2\right)_{y_2=49}^{y_2=51} =$$

(choose one) (i) **50** (ii) **100** (iii) **150**.

iii. Since marginal,

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 = \int_{49}^{51} \frac{1}{12}\, dy_1 = \frac{1}{12}(y_1)_{y_1=49}^{y_1=51} = \frac{1}{6},$$

$4 \leq y_2 \leq 6$, expected value of Y_2 is

$$E[Y_2] = \int_4^6 y_2 f_2(y_2)\, dy_2 = \int_4^6 y_2 \frac{1}{6}\, dy_2 = \left(\frac{1}{12} y_2^2\right)_{y_1=4}^{y_2=6} =$$

(choose one) (i) $\mathbf{\frac{5}{3}}$ (ii) $\mathbf{\frac{10}{3}}$ (iii) $\mathbf{\frac{15}{3}}$.

iv. Since

$$E[Y_1 Y_2] = \frac{250}{3} = E[Y_1]E[Y_2] = \left(\frac{50}{3}\right)(5)$$

Y_1 and Y_2 are (choose one) (i) **dependent** (ii) **independent**.

v. Using the properties of expectation above,

$$E[3Y_1 + 3Y_2] = 3E[Y_1] + 3E[Y_2] = 3\left(\frac{50}{3}\right) + 3\,(5) =$$

(i) **55** (ii) **60** (iii) **65**.

vi. Since

$$E[Y_2^2] = \int_4^6 y_2^2 f_2(y_2)\, dy_2 = \int_4^6 y_2^2 \frac{1}{6}\, dy_2 = \left(\frac{1}{18} y_2^3\right)_{y_2=4}^{y_2=6} = \frac{76}{9},$$

then variance is

$$V[Y_2] = E\left[Y_2^2\right] - \left[E[Y_2]\right]^2 = \frac{76}{9} - \left(\frac{5}{3}\right)^2 \approx$$

(choose one) (i) $\mathbf{\frac{17}{3}}$ (ii) $\mathbf{\frac{18}{3}}$ (iii) $\mathbf{\frac{19}{3}}$.

vii. Let $U = Y_1 - Y_2$.

$$\begin{aligned}
V(U) &= E(U^2) - E(U)^2 \\
&= E\left[(Y_1 - Y_2)^2\right] - E\left((Y_1 - Y_2)\right)^2 \\
&= E\left[Y_1^2 - 2Y_1Y_2 + Y_2^2\right] - \left[E(Y_1) - E(Y_2)\right]^2 \\
&= E\left[Y_1^2\right] - 2E\left[Y_1\right]E\left[Y_2\right] + E\left[Y_2^2\right] - \left\{\left[E(Y_1)\right]^2 - 2E\left[Y_1\right]E\left[Y_2\right] + \left[E(Y_2)\right]^2\right\} \\
&= E\left[Y_1^2\right] - \left[E(Y_1)\right]^2 + \left\{E\left[Y_2^2\right] - \left[E(Y_2)\right]^2\right\} \\
&= V(Y_1) + V(Y_2)
\end{aligned}$$

(i) **True** (ii) **False**

(Notice: $E(2Y_1Y_2) = 2E(Y_1)E(Y_2)$ because Y_1 and Y_2 are independent.)

4. *More Continuous Expectation Value Calculations.*
Consider

$$f(y_1, y_2) = \begin{cases} \frac{6}{5}(2 - y_1)(1 - y_2) & 0 \leq y_1 \leq 2,\ 0 \leq y_2 \leq 1,\ y_1 + 2y_2 < 2 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Since $y_1 + 2y_2 \leq 2$, then $y_2 \leq 1 - \frac{1}{2}y_1$, and marginal of Y_1 is

$$\begin{aligned} \frac{6}{5}\int_0^{1-\frac{1}{2}y_1} (2 - y_1)(1 - y_2)\, dy_2 &= \frac{6}{5}\int_0^{1-\frac{1}{2}y_1} (2 - y_1 - 2y_2 + y_1 y_2)\, dy_2 \\ &= \frac{6}{5}\left(2y_2 - y_1 y_2 - y_2^2 + \frac{1}{2}y_1 y_2^2 \right)_{y_2=0}^{y_2=1-\frac{1}{2}y_1} \\ &= \frac{6}{5}\left(1 - \frac{1}{2}y_1 - \frac{1}{4}y_1^2 + \frac{1}{8}y_1^3\right) \end{aligned}$$

where $0 \leq y_1 \leq 2$. So, the expected value of Y_1 is

$$\begin{aligned} E[Y_1] &= \int_0^2 y_1 f_1(y_1)\, dy_1 = \frac{6}{5}\int_0^2 y_1 \left(1 - \frac{1}{2}y_1 - \frac{1}{4}y_1^2 + \frac{1}{8}y_1^3\right) dy_1 \\ &= \frac{6}{5}\left(\frac{1}{2}y_1^2 - \frac{1}{6}y_1^3 - \frac{1}{16}y_1^4 + \frac{1}{40}y_1^5\right)_{y_2=0}^{y_2=2} = \end{aligned}$$

(choose one) (i) $\mathbf{\frac{13}{25}}$ (ii) $\mathbf{\frac{14}{25}}$ (iii) $\mathbf{\frac{15}{25}}$.

(b) Since $y_1 + 2y_2 \leq 2$, then $y_1 \leq 2 - 2y_2$, and marginal of Y_2 is

$$\begin{aligned} \frac{6}{5}\int_0^{2-2y_2} (2 - y_1)(1 - y_2)\, dy_1 &= \frac{6}{5}\int_0^{2-2y_2} (2 - y_1 - 2y_2 + y_1 y_2)\, dy_1 \\ &= \frac{6}{5}\left(2y_1 - \frac{1}{2}y_1^2 - 2y_1 y_2 + \frac{1}{2}y_1^2 y_2 \right)_{y_1=0}^{y_1=2-2y_2} \\ &= \frac{6}{5}\left(2 - 2y_2 - 2y_2^2 + 2y_2^3\right) \end{aligned}$$

where $0 \leq y_1 \leq 1$. So, the expected value of Y_1 is

$$\begin{aligned} E[Y_2] &= \int_0^1 y_2 f_2(y_2)\, dy_2 = \frac{6}{5}\int_0^1 y_2 \left(2 - 2y_2 - 2y_2^2 + 2y_2^3\right) dy_2 \\ &= \frac{6}{5}\left(y_2^2 - \frac{2}{3}y_2^3 - \frac{2}{4}y_2^4 + \frac{2}{5}y_2^5\right)_{y_2=0}^{y_2=1} = \end{aligned}$$

(choose one) (i) $\mathbf{\frac{7}{25}}$ (ii) $\mathbf{\frac{8}{25}}$ (iii) $\mathbf{\frac{9}{25}}$.

(c) Using the properties of expectation above,

$$E[25Y_1 + 50Y_2] = 25E[Y_1] + 50E[Y_2] = 25\left(\frac{17}{25}\right) + 50\left(\frac{7}{25}\right) =$$

(i) **29** (ii) **30** (iii) **31**.

(d) Since

$$\begin{aligned} E[Y_2^2] &= \int_0^1 y_2^2 f_2(y_2)\,dy_2 = \frac{6}{5}\int_0^1 y_2^2\left(2 - 2y_2 - 2y_2^2 + 2y_2^3\right)dy_2 \\ &= \frac{6}{5}\left(\frac{2}{3}y_2^3 - \frac{2}{4}y_4 - \frac{2}{5}y_2^5 + \frac{2}{6}y_2^6\right)_{y_2=0}^{y_2=1} = \frac{3}{25}, \end{aligned}$$

then variance is

$$V[Y_2] = E\left[Y_2^2\right] - [E[Y_2]]^2 = \frac{3}{25} - \left(\frac{7}{25}\right)^2 \approx$$

(choose one) (i) $\mathbf{\frac{24}{625}}$ (ii) $\mathbf{\frac{25}{625}}$ (iii) $\mathbf{\frac{26}{625}}$.

(e) According to Tchebysheff's theorem, the probability variable Y_2 is *within* $k = 3$ standard deviations of mean $E[Y_2]$ is *at least*

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = \frac{8}{9}.$$

That is, there is at least a $\frac{8}{9}$th chance the following interval contains Y_2:

$$E[Y_2] \pm 3\sqrt{V[Y_2]} = \frac{7}{25} \pm 3\sqrt{\frac{26}{625}} \approx$$

(i) $\mathbf{(-0.23, 0.79)}$ (ii) $\mathbf{(-0.33, 0.89)}$ (iii) $\mathbf{(-0.43, 0.99)}$.

5. *Expected Number of Matches.* Ten people throw ten tickets with their names on each ticket into a jar, then draw one ticket out of the jar at random (and put it back in the jar). Let X be the number of people who select their own ticket out of the jar. Let

$$X = Y_1 + Y_2 + \cdots + Y_{10}$$

where

$$Y_i = \begin{cases} 1 & \text{if } i\text{th person selects own ticket} \\ 0 & \text{if } i\text{th person does not select their own ticket} \end{cases}$$

(a) Each person chooses any of the ten tickets with equal chance, $E[Y_i] = 1 \times p(1) + 0 \times p(0) =$ (choose one) (i) $\mathbf{\frac{1}{10}}$ (ii) $\mathbf{\frac{2}{10}}$ (iii) $\mathbf{\frac{3}{10}}$.

(b) Expected number of ten individuals to choose their own ticket is $E(X) = E(Y_1) + \cdots + E(Y_{10}) = 10 \times \frac{1}{10} =$ (i) $\mathbf{\frac{8}{10}}$ (ii) $\mathbf{\frac{9}{10}}$ (iii) $\mathbf{\frac{10}{10}}$.
We would expect one of ten individuals to choose their own ticket.

(c) *If* n individuals played this game, then we would expect $E(X) = E(Y_1) + \cdots + E(Y_n) = n\left(\frac{1}{n}\right) =$ (i) $\boldsymbol{\frac{n-1}{n}}$ (ii) $\boldsymbol{\frac{n}{n}}$ (iii) $\boldsymbol{\frac{n+1}{n}}$.

5.7 The Covariance of Two Random Variables

We look at *covariance*, a measure of how two random variables are linearly related. Covariance is defined by

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - E(Y_2))(Y_2 - E(Y_2))] = E(Y_1Y_2) - E(Y_1)E(Y_2),$$

and has the following properties,

- $\text{Cov}(Y_1, Y_2) = \text{Cov}(Y_2, Y_1)$,
- $\text{Cov}(Y_1, Y_1) = V(Y_1)$,
- $\text{Cov}(aY_1, Y_2) = a\text{Cov}(Y_1, Y_2)$, where a is a constant,
- $\text{Cov}(Y_1, Y_2) = 0$ if Y_1, Y_2 are uncorrelated or independent.

Uncorrelated random variables are *not* necessarily independent random variables. Independent random variables *are* necessarily uncorrelated random variables. The *correlation* ρ, $-1 \leq \rho \leq 1$, is given by

$$\rho(y_1, y_2) = \rho = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1\sigma_2}.$$

Exercise 5.7 (The Covariance of Two Random Variables)

1. *Covariance and Correlation: Waiting Times To Catch Fish.* The joint density, $p(y_1, y_2)$, of the number of minutes waiting to catch the *first* fish, y_1, *and* the number of minutes waiting to catch the *second* fish, y_2, is given below.

$y_2 \downarrow y_1 \rightarrow$	1	2	3	total
1	0.01	0.01	0.07	0.09
2	0.02	0.02	0.08	0.12
3	0.08	0.08	0.63	0.79
total	0.11	0.11	0.78	1.00

(a) If $g(y_1, y_2) = y_1y_2$, then

$$\begin{aligned} E[Y_1Y_2] &= \sum_{y_1=1}^{3}\sum_{y_2=1}^{3}(y_1y_2)\,p(y_1, y_2) \\ &= (1\times 1)(0.01) + (1\times 2)(0.02) + (1\times 3)(0.08) \\ &\quad + (2\times 1)(0.01) + (2\times 2)(0.02) + (2\times 3)(0.08) \\ &\quad + (3\times 1)(0.07) + (3\times 2)(0.08) + (3\times 3)(0.63) = \end{aligned}$$

(choose one) (i) **5.17** (ii) **6.17** (iii) **7.17**.
(Hint: Type $1, 1, 1, 2, 2, 2, 3, 3, 3$ in L_1 and $1, 2, 3, 1, 2, 3, 1, 2, 3$ in L_2, define $L_3 = L_1 \times L_2$, type $0.01, 0.02, 0.08, 0.01, 0.02, 0.08, 0.07, 0.07, 0.63$ in L_4, define $L_5 = L_3 \times L_4$, sum L_5 using 1-Var Stats L_5, read $\sum x = 7.17$.)

(b) Since

$$\begin{aligned} E[Y_1] &= \sum_{y_1=1}^{3} y_1 p_1(y_1) = (1)(0.11) + (2)(0.11) + (3)(0.78) = 2.67 \\ E[Y_2] &= \sum_{y_2=1}^{3} y_2 p_2(y_2) = (1)(0.09) + (2)(0.12) + (3)(0.79) = 2.7, \end{aligned}$$

then covariance is

$$\text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = 7.17 - (2.67)(2.7) \approx$$

(choose one) (i) **−0.039** (ii) **0.039** (iii) **0.139**.

(c) Since

$$\begin{aligned} E[Y_1^2] &= \sum_{y_1=1}^{3} y_1^2 p_1(y_1) = (1^2)(0.11) + (2^2)(0.11) + (3^2)(0.78) = 7.57, \\ V[Y_1] &= E\left[Y_1^2\right] - \left[E[Y_1]\right]^2 = 7.57 - 2.67^2 = 0.4411, \\ E[Y_2^2] &= \sum_{y_2=1}^{3} y_2^2 p_2(y_2) = (1^2)(0.09) + (2^2)(0.12) + (3^2)(0.79) = 2.7, \\ V[Y_2] &= E\left[Y_2^2\right] - \left[E[Y_2]\right]^2 = 7.68 - 2.7^2 = 0.39, \end{aligned}$$

then correlation is

$$\rho = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{-0.039}{\sqrt{0.4411 \times 0.39}} \approx$$

(choose one) (i) **−0.235** (ii) **−0.139** (iii) **−0.094**.
There is very little linear correlation between two waiting times.

(d) Let $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 - Y_2$. Then

$$\begin{aligned} \text{Cov}(U_1, U_2) &= E(U_1 U_2) - E(U_1)E(U_2) \\ &= E\left[(Y_1 + Y_2)(Y_1 - Y_2)\right] - E\left((Y_1 + Y_2)\right) E\left((Y_1 - Y_2)\right) \\ &= E\left[Y_1^2 - Y_2^2\right] - \left[E(Y_1) + E(Y_2)\right]\left[E(Y_1) - E(Y_2)\right] \\ &= E\left[Y_1^2\right] - E\left[Y_2^2\right] - \left\{[E(Y_1)]^2 - [E(Y_2)]^2\right\} \\ &= E\left[Y_1^2\right] - [E(Y_1)]^2 - \left\{E\left[Y_2^2\right] - [E(Y_2)]^2\right\} \\ &= V(Y_1) - V(Y_2) = 0.4411 - 0.39 \approx \end{aligned}$$

(choose one) (i) **0.0355** (ii) **0.0392** (iii) **0.0511**.

2. *Covariance and Correlation: Marbles In An Urn.* Marbles chosen at random with*out* replacement from an urn consisting of 8 blue and 6 black marbles. For ith marble chosen, $Y_i = 0$ if blue and $Y_i = 1$ if black.

$y_2 \downarrow y_1 \rightarrow$	blue, 0	black, 1	$p_2(y_2)$
blue, 0	$\frac{8\cdot 7}{14\cdot 13}$	$\frac{6\cdot 8}{14\cdot 13}$	$\frac{8\cdot 7+6\cdot 8}{14\cdot 13}$
black, 1	$\frac{8\cdot 6}{14\cdot 13}$	$\frac{6\cdot 5}{14\cdot 13}$	$\frac{8\cdot 6+6\cdot 5}{14\cdot 13}$
$p_1(y_1)$	$\frac{8\cdot 7+8\cdot 6}{14\cdot 13}$	$\frac{6\cdot 8+6\cdot 5}{14\cdot 13}$	1

(a) Since

$$\begin{aligned}
E\left[Y_1Y_2\right] &= \sum_{y_1=0}^{1}\sum_{y_2=0}^{1}(y_1y_2)\,p(y_1,y_2) \\
&= (0\times 0)\left(\frac{8\cdot 7}{14\cdot 13}\right) + (0\times 1)\left(\frac{8\cdot 6}{14\cdot 13}\right) \\
&\quad + (1\times 0)\left(\frac{8\cdot 8}{14\cdot 13}\right) + (1\times 1)\left(\frac{6\cdot 5}{14\cdot 13}\right) = \frac{6\cdot 5}{14\cdot 13}, \\
E[Y_1] &= \sum_{y_1=0}^{1} y_1p_1(y_1) \\
&= (0)\left(\frac{8\cdot 7+8\cdot 6}{14\cdot 13}\right) + (1)\left(\frac{6\cdot 8+6\cdot 5}{14\cdot 13}\right) = \frac{6\cdot 8+6\cdot 5}{14\cdot 13}, \\
E[Y_2] &= \sum_{y_2=0}^{1} y_2p_2(y_2) \\
&= (0)\left(\frac{8\cdot 7+6\cdot 8}{14\cdot 13}\right) + (1)\left(\frac{8\cdot 6+6\cdot 5}{14\cdot 13}\right) = \frac{8\cdot 6+6\cdot 5}{14\cdot 13},
\end{aligned}$$

then covariance is

$$\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) =$$

(choose one) (i) $-\frac{\mathbf{12}}{\mathbf{637}}$ (ii) $-\frac{\mathbf{13}}{\mathbf{637}}$ (iii) $-\frac{\mathbf{14}}{\mathbf{637}}$.

(b) Since

$$\begin{aligned}
E[Y_1^2] &= \sum_{y_1=1}^{3} y_1^2 p_1(y_1) = (0^2)\left(\frac{8\cdot 7+8\cdot 6}{14\cdot 13}\right) + (1^2)\left(\frac{6\cdot 8+6\cdot 5}{14\cdot 13}\right) = \frac{6\cdot 8+6\cdot 5}{14\cdot 13}, \\
V[Y_1] &= E\left[Y_1^2\right] - [E[Y_1]]^2 = \frac{6\cdot 8+6\cdot 5}{14\cdot 13} - \left(\frac{8\cdot 6+6\cdot 5}{14\cdot 13}\right)^2 \approx 0.245, \\
E[Y_2^2] &= \sum_{y_2=0}^{0} y_2^2 p_2(y_2) = (0^2)\left(\frac{8\cdot 7+6\cdot 8}{14\cdot 13}\right) + (1^2)\left(\frac{8\cdot 6+6\cdot 5}{14\cdot 13}\right) = \frac{8\cdot 6+6\cdot 5}{14\cdot 13}, \\
V[Y_2] &= E\left[Y_2^2\right] - [E[Y_2]]^2 = \frac{8\cdot 6+6\cdot 5}{14\cdot 13} - \left(\frac{8\cdot 6+6\cdot 5}{14\cdot 13}\right)^2 \approx 0.245,
\end{aligned}$$

then correlation is

$$\rho = \frac{\operatorname{Cov}(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{-\frac{12}{637}}{\sqrt{0.245 \times 0.245}} \approx$$

(choose one) (i) $\mathbf{-0.077}$ (ii) $\mathbf{-0.139}$ (iii) $\mathbf{-0.294}$.

(c) *Relationship between covariance and variance.*

$$\begin{aligned}
\operatorname{Cov}(Y_1, Y_1) &= E(Y_1Y_1) - E(Y_1)E(Y_1) = E(Y_1^2) - [E(Y_1)]^2 = V(V_1) \approx 0.245 \\
\operatorname{Cov}(Y_2, Y_2) &= E(Y_2Y_2) - E(Y_2)E(Y_2) = E(Y_2^2) - [E(Y_2)]^2 = V(V_2) \approx 0.245
\end{aligned}$$

(i) **True** (ii) **False**

3. *Covariance and Correlation: Weight and Amount of Salt in Potato Chips.*

(a) *Machine A.* Bivariate density function for machine A is

$$f(y_1, y_2) = \begin{cases} \frac{1}{12}, & 49 \leq y_1 \leq 51, 2 \leq y_2 \leq 8 \\ 0 & \text{elsewhere} \end{cases}$$

i. Since

$$\begin{aligned}
E\left[Y_1Y_2\right] &= \int_2^8 \int_{49}^{51} (y_1y_2)\, f(y_1,y_2)\, dy_1dy_2 \\
&= \int_2^8 \int_{49}^{51} (y_1y_2)\, \frac{1}{12}\, dy_1dy_2 = \int_2^8 \frac{1}{12} y_2 \left(\frac{1}{2}y_1^2\right)_{y_1=49}^{y_1=51} dy_2 \\
&= \left(\frac{100}{24}y_2^2\right)_{y_2=2}^{y_2=8} = 250, \\
f_1(y_1) &= \int_{-\infty}^{\infty} f(y_1,y_2)\, dy_2 \\
&= \int_2^8 \frac{1}{12}\, dy_2 = \frac{1}{12}(y_2)_{y_2=2}^{y_2=8} = \frac{1}{2}, \\
E[Y_1] &= \int_{49}^{51} y_1 f_1(y_1)\, dy_1 \\
&= \int_{49}^{51} y_1 \frac{1}{2}\, dy_1 = \left(\frac{1}{4}y_1^2\right)_{y_1=49}^{y_1=51} = 50, \\
f_2(y_2) &= \int_{-\infty}^{\infty} f(y_1,y_2)\, dy_1 \\
&= \int_{49}^{51} \frac{1}{12}\, dy_1 = \frac{1}{12}(y_1)_{y_1=49}^{y_1=51} = \frac{1}{6}, \\
E[Y_2] &= \int_2^8 y_2 f_2(y_2)\, dy_2 \\
&= \int_2^8 y_2 \frac{1}{6}\, dy_2 = \left(\frac{1}{12}y_2^2\right)_{y_2=2}^{y_2=8} = 5,
\end{aligned}$$

then covariance is

$$\operatorname{Cov}(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = 250 - (50)(5) =$$

(choose one) (i) **0** (ii) **1** (iii) **2**.

ii. Since Y_1 and Y_2 are uncorrelated, $\operatorname{Cov}(Y_1,Y_2) = 0$, then correlation is

$$\rho = \frac{\operatorname{Cov}(Y_1,Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{0}{\sqrt{0.245 \times 0.245}} =$$

(choose one) (i) **0** (ii) **1** (iii) **2**.

iii. Random variables Y_1 and Y_2 are uncorrelated, $\operatorname{Cov}(Y_1,Y_2) = 0$, because they are *independent*,

$$E(Y_1Y_2) = E(Y_1)E(Y_2).$$

(i) **True** (ii) **False**

(b) *Machine B.* Bivariate density function for machine B is

$$f(y_1, y_2) = \begin{cases} \frac{1}{4}, & 49 \leq y_1 \leq 51, 4 \leq y_2 \leq 6 \\ 0 & \text{elsewhere} \end{cases}$$

i. Since Y_1 and Y_2 are independent,
$\text{Cov}(Y_1, Y_2) =$ (choose one) (i) **0** (ii) **1** (iii) **2**.

ii. Since Y_1 and Y_2 are independent,
correlation $\rho = \frac{\text{Cov}(Y_1,Y_2)}{\sqrt{V(Y_1)V(Y_2)}} =$ (choose one) (i) **0** (ii) **0.5** (iii) **1**.

4. *More Covariance and Correlation.*
Consider

$$f(y_1, y_2) = \begin{cases} \frac{6}{5}(2 - y_1)(1 - y_2) & 0 \leq y_1 \leq 2,\ 0 \leq y_2 \leq 1,\ y_1 + 2y_2 < 2 \\ 0 & \text{elsewhere} \end{cases}$$

which has a non–rectangular range.

(a) Since

$$\begin{aligned}
E[Y_1Y_2] &= \frac{6}{5}\int_0^1 \int_0^2 (y_1y_2)(2 - y_1 - 2y_2 + y_1y_2)\, dy_1dy_2 \\
&= \frac{6}{5}\int_0^1 \int_0^2 \left(2y_1y_2 - y_1^2y_2 - 2y_1y_2^2 + y_1^2y_2^2\right) dy_1dy_2 \\
&= \frac{6}{5}\int_0^1 \left(y_1^2y_2 - \frac{1}{3}y_1^3y_2 - y_1^2y_2^2 + \frac{1}{3}y_1^2y_2^2\right)_{y_1=0}^{y_1=2} dy_2 \\
&= \frac{6}{5}\int_0^1 \left(4y_2 - \frac{8}{3}y_2 - 4y_2^2 + \frac{8}{3}y_2^2\right) dy_2 \\
&= \frac{6}{5}\int_0^1 \left(\frac{4}{3}y_2 - \frac{4}{3}y_2^2\right) dy_2 \\
&= \frac{6}{5}\left(\frac{4}{6}y_2^2 - \frac{4}{9}y_2^3\right)_{y_2=0}^{y_2=1} = \frac{4}{15}, \\
E[Y_1] &= \int_0^2 y_1f_1(y_1)\, dy_1 = \frac{6}{5}\int_0^2 y_1\left(1 - \frac{1}{2}y_1 - \frac{1}{4}y_1^2 + \frac{1}{8}y_1^3\right) dy_1 \\
&= \frac{6}{5}\left(\frac{1}{2}y_1^2 - \frac{1}{6}y_1^3 - \frac{1}{16}y_1^4 + \frac{1}{40}y_1^5\right)_{y_2=0}^{y_2=2} = \frac{14}{25}, \\
E[Y_2] &= \int_0^1 y_2f_2(y_2)\, dy_2 = \frac{6}{5}\int_0^1 y_2\left(2 - 2y_2 - 2y_2^2 + 2y_2^3\right) dy_2 \\
&= \frac{6}{5}\left(y_2^2 - \frac{2}{3}y_2^3 - \frac{2}{4}y_2^4 + \frac{2}{5}y_2^5\right)_{y_2=0}^{y_2=1} = \frac{7}{25},
\end{aligned}$$

then covariance is

$$\operatorname{Cov}(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = \frac{4}{15} - \left(\frac{14}{25}\right)\left(\frac{7}{25}\right) =$$

(choose one) (i) $\mathbf{\frac{205}{1875}}$ (ii) $\mathbf{\frac{206}{1875}}$ (iii) $\mathbf{\frac{207}{1875}}$.

(b) Since

$$\begin{aligned}
E[Y_1^2] &= \int_0^2 y_1^2 f_1(y_1)\,dy_1 = \frac{6}{5}\int_0^2 y_1^2\left(1 - \frac{1}{2}y_1 - \frac{1}{4}y_1^2 + \frac{1}{8}y_1^3\right)\,dy_1 \\
&= \frac{6}{5}\left(\frac{1}{3}y_1^3 - \frac{1}{8}y_1^4 - \frac{1}{20}y_1^5 + \frac{1}{48}y_1^6\right)\Bigg|_{y_2=0}^{y_2=2} = \frac{202}{25}, \\
V[Y_1] &= E\left[Y_1^2\right] - [E[Y_1]]^2 = \frac{202}{25} - \left(\frac{14}{25}\right)^2 = \frac{4854}{625}, \\
E[Y_2^2] &= \int_0^1 y_2^2 f_2(y_2)\,dy_2 = \frac{6}{5}\int_0^1 y_2^2\left(2 - 2y_2 - 2y_2^2 + 2y_2^3\right)\,dy_2 \\
&= \frac{6}{5}\left(\frac{2}{3}y_2^3 - \frac{2}{4}y_4 - \frac{2}{5}y_2^5 + \frac{2}{6}y_2^6\right)\Bigg|_{y_2=0}^{y_2=1} = \frac{3}{25}, \\
V[Y_2] &= E\left[Y_2^2\right] - [E[Y_2]]^2 = \frac{3}{25} - \left(\frac{7}{25}\right)^2 = \frac{26}{625},
\end{aligned}$$

then correlation is

$$\rho = \frac{\operatorname{Cov}(Y_1, Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{\frac{206}{1875}}{\sqrt{\frac{4854}{625} \times \frac{26}{625}}} \approx$$

(choose one) (i) **0.19** (ii) **0.23** (iii) **0.34**.