MULTIVARIATE PROBABILITY DISTRIBUTIONS

1. Preliminaries

1.1. **Example.** Consider an experiment that consists of tossing a die and a coin at the same time. We can consider a number of random variables defined on this sample space. We will assign an indicator random variable to the result of tossing the coin. If it comes up heads we will assign a value of one, and if it comes up zero we will assign a value of zero. Consider the following random variables.

X_1**:** The number of dots appearing on the die.
X_2**:** The sum of the number of dots on the die and the indicator for the coin.
X_3**:** The value of the indicator for tossing the coin.
X_4**:** The product of the number of dots on the die and the indicator for the coin.

There are twelve sample points associated with this experiment where the first element of the pair is the number on the die and the second is whether the coin comes up heads or tails.

$$\begin{array}{cccccc} E_1 : 1H & E_2 : 2H & E_3 : 3H & E_4 : 4H & E_5 : 5H & E_6 : 6H \\ E_7 : 1T & E_8 : 2T & E_9 : 3T & E_{10} : 4T & E_{11} : 5T & E_{12} : 6T \end{array}$$

Random variable X_1 has six possible outcomes, each with probability $\frac{1}{6}$. Random variable X_3 has two possible outcomes, each with probability $\frac{1}{2}$. Consider the values of X_2 for each of the sample points. The possible outcomes and the probabilities for X_2 are as follows:

Table 1. Probability of X_2

Value of Random Variable	Probability
1	1/12
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
7	1/12

The possible outcomes and the probabilities for X_4 are as follows:

Date: January 23, 2006.

TABLE 2. Probability of X_4

Value of Random Variable	Probability
0	1/2
1	1/12
2	1/12
3	1/12
4	1/12
5	1/12
6	1/12

1.2. **Bivariate Random Variables.** Now consider the intersection of $X_1 = 3$ and $X_2 = 3$. We call this intersection a bivariate random variable. For a general bivariate case we write this as $P(X_1 = x_1, X_2 = x_2)$. We can write the probability distribution in the form of a table as follows for the above example.

TABLE 3. Joint Probability of X_1 and X_2

		X_2						
		1	2	3	4	5	6	7
	1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	0
	2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0
X_1	3	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0
	4	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0
	5	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0
	6	0	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$

For the example, $P(X_1 = 3,\, X_2 = 3) = \frac{1}{12}$, which is the probability of sample point E_9.

2. PROBABILITY DISTRIBUTIONS FOR DISCRETE MULTIVARIATE RANDOM VARIABLES

2.1. **Definition.** If X_1 and X_2 be discrete random variables, the function given by

$$p(x_1,\, x_2) = P(X_1 = x_1,\, X_2 = x_2)$$

for each pair of values of $(x_1,\, x_2)$ within the range of X_1 and X_2 is called the joint (or bivariate) probability distribution for X_1 and X_2. Specifically we write

$$p(x_1,\, x_2) = P(X_1 = x_1,\, X_2 = x_2), \qquad -\infty < x_1 < \infty,\, -\infty < x_2 < \infty. \tag{1}$$

In the single-variable case, the probability function for a discrete random variable X assigns non-zero probabilities to a countable number of distinct values of X in such a way that the sum of the probabilities is equal to 1. Similarly, in the bivariate case the joint probability function $p(x_1,\, x_2)$ assigns non-zero probabilities to only a countable number of pairs of values $(x_1,\, x_2)$. Further, the non-zero probabilities must sum to 1.

2.2. **Properties of the Joint Probability (or Density) Function.**

Theorem 1. *If X_1 and X_2 are discrete random variables with joint probability function $p(x_1,\, x_2)$, then*

(i) $p(x_1,\, x_2) \geq 0$ *for all* $x_1,\, x_2$.
(ii) $\sum\limits_{x_1,\, x_2} p(x_1,\, x_2) = 1$ *, where the sum is over all values* $(x_1,\, x_2)$ *that are assigned non-zero probabilities.*

Once the joint probability function has been determined for discrete random variables X_1 and X_2, calculating joint probabilities involving X_1 and X_2 is straightforward.

2.3. **Example 1.** Roll a red die and a green die. Let

$$X_1 = \text{number of dots on the red die}$$
$$X_2 = \text{number of dots on the green die}$$

There are 36 points in the sample space.

TABLE 4. Possible Outcomes of Rolling a Red Die and a Green Die. (First number in pair is number on red die.)

Green Red	1	2	3	4	5	6
1	1 1	1 2	1 3	1 4	1 5	1 6
2	2 1	2 2	2 3	2 4	2 5	2 6
3	3 1	3 2	3 3	3 4	3 5	3 6
4	4 1	4 2	4 3	4 4	4 5	4 6
5	5 1	5 2	5 3	5 4	5 5	5 6
6	6 1	6 2	6 3	6 4	6 5	6 6

The probability of (1, 1) is $\frac{1}{36}$. The probability of (6, 3) is also $\frac{1}{36}$.

Now consider $P(2 \leq X_1 \leq 3,\, 1 \leq X_2 \leq 2)$. This is given as

$$\begin{aligned} P(2 \leq X_1 \leq 3,\, 1 \leq X_2 \leq 2) &= p(2,\, 1) + p(2,\, 2) + p(3,\, 1) + p(3,\, 2) \\ &= \frac{4}{36} = \frac{1}{9} \end{aligned}$$

2.4. **Example 2.** Consider the example of tossing a coin and rolling a die from section 1. Now consider $P(2 \leq X_1 \leq 3,\, 1 \leq X_2 \leq 2)$. This is given as

$$\begin{aligned} P(2 \leq X_1 \leq 4,\, 3 \leq X_2 \leq 5) &= p(2,\,3) + p(2,\,4) + p(2,\,5) \\ &+ p(3,\,3) + p(3,\,4) + p(3,\,5) \\ &+ p(4,\,3) + p(4,\,4) + p(4,\,5) \\ &= \frac{5}{36} \end{aligned}$$

2.5. **Example 3.** Two caplets are selected at random from a bottle containing three aspirin, two sedative, and four cold caplets. If X and Y are, respectively, the numbers of aspirin and sedative caplets included among the two caplets drawn from the bottle, find the probabilities associated with all possible pairs of values of X and Y?

The possible pairs are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0). To find the probability associated with (1, 0), for example, observe that we are concerned with the event of getting one of the three aspirin caplets, none of the two sedative caplets, and hence, one of the four cold caplets. The number of ways in which this can be done is

$$\binom{3}{1}\binom{2}{0}\binom{4}{1} = 12$$

and the total number of ways in which two of the nine caplets can be selected is

$$\binom{9}{2} = 36.$$

Since those possibilities are all equally likely by virtue of the assumption that the selection is random, it follows that the probability associated with (1, 0) is $\frac{12}{36} = \frac{1}{3}$. Similarly, the probability associated with (1, 1) is

$$\frac{\binom{3}{1}\binom{2}{1}\binom{4}{0}}{36} = \frac{6}{36} = \frac{1}{6}$$

and, continuing this way, we obtain the values shown in the following table:

TABLE 5. Joint Probability of Drawing Aspirin (X_1) and Sedative Caplets (Y).

		x		
		0	1	2
	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
y	1	$\frac{2}{9}$	$\frac{1}{6}$	0
	2	$\frac{1}{36}$	0	0

We can also represent this joint probability distribution as a formula

$$p(x,\,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}}{36},\; x = 0,\,1,\,2; \quad y = 0,\,1,\,2; \quad 0 \le (x+y) \le 2$$

3. Distribution Functions for Discrete Multivariate Random Variables

3.1. **Definition of the Distribution Function.** If X_1 and X_2 are discrete random variables, the function given by

$$F(x_1,\,x_2) = P\left[X_1 \le x_1,\, X_2 \le x_2\right] = \sum_{u_1 \le x_1} \sum_{u_2 \le x_2} p(u_1,\,u_2) \qquad \begin{array}{c} -\infty < x_1 < \infty \\ -\infty < x_2 < \infty \end{array} \tag{2}$$

where $p(u_1,\,u_2)$ is the value of the joint probability function of X_1 and X_2 at $(u_1,\,u_2)$ is called the joint distribution function, or the joint cumulative distribution of X_1 and X_2.

3.2. **Examples.**

3.2.1. *Example 1.* Consider the experiment of tossing a red and green die where X_1 is the number of the red die and X_2 is the number on the green die.

Now find $F(2,\,3) = P(X_1 \le 2,\, X_2 \le 3)$. This is given by summing as in the definition (equation 2).

$$\begin{aligned} F(2,\,3) = P\left[X_1 \le 2,\, X_2 \le 3\right] &= \sum_{u_1 \le 2} \sum_{u_2 \le 3} p(u_1,\,u_2) \\ &= p(1,\,1) + p(1,\,2) + p(1,\,3) + p(2,\,1) + p(2,\,2) + p(2,\,3) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\ &= \frac{6}{36} = \frac{1}{6} \end{aligned}$$

3.2.2. *Example 2.* Consider Example 3 from Section 2. The joint probability distribution is given in Table 5 which is repeated here for convenience.

Table 5. Joint Probability of Drawing Aspirin (X_1) and Sedative Caplets (Y).

		x		
		0	1	2
	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
y	1	$\frac{2}{9}$	$\frac{1}{6}$	0
	2	$\frac{1}{36}$	0	0

The joint probability distribution is

$$p(x,\, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}}{36},\; x = 0,\, 1,\, 2;\; y = 0,\, 1,\, 2;\; 0 \le (x+y) \le 2$$

For this problem find $F(1\,,2) = P(X \le 1,\, Y \le 2)$. This is given by

$$\begin{aligned} F(1,\, 2) = P\,[X \le 1,\, Y \le 2] &= \sum_{u_1 \le 1} \sum_{u_2 \le 2} p(u_1,\, u_2) \\ &= p(0,\, 0) + p(0,\, 1) + p(0,\, 2) + p(1,\, 0) + p(1,\, 1) + p(1,\, 2) \\ &= \frac{1}{6} + \frac{2}{9} + \frac{1}{36} + \frac{1}{3} + \frac{1}{6} + 0 \\ &= \frac{6}{36} + \frac{8}{36} + \frac{1}{36} + \frac{12}{36} + \frac{6}{36} \\ &= \frac{33}{36} \end{aligned}$$

4. Probability Distributions for Continuous Bivariate Random Variables

4.1. **Definition of a Joint Probability Density Function.** A bivariate function with values $f(x_1,\, x_2)$ defined over the x_1x_2-plane is called a joint probability density function of the continuous random variables X_1 and X_2 if, and only if,

$$P\,[(X_1,\, X_2) \in A] = \int_A \int f(x_1,\, x_2)\, dx_1\, dx_2 \quad \text{for any region } A \in \text{the } x_1x_2 \text{ -plane} \tag{3}$$

4.2. **Properties of the Joint Probability (or Density) Function in the Continuous Case.**

Theorem 2. *A bivariate function can serve as a joint probability density function of a pair of continuous random variables X_1 and X_2 if its values, $f(x_1,\, x_2)$, satisfy the conditions*

(i) $f(x_1,\, x_2) \ge 0$ *for* $-\infty < x_1 < \infty,\, \infty < x_2 < \infty$

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1,\, x_2) dx_1\, dx_2 = 1$

4.3. **Example of a Joint Probability Density Function.** Given the joint probability density function

$$f(x_1,\, x_2) = \begin{cases} 6x_1^2 x_2 & 0 < x_1 < 1,\, 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

of the two random variables, X_1 and X_2, find $P\,[(X_1,\, X_2) \in A]$, where A is the region $\{(x_1,\, x_2) \mid 0 < x_1 < \frac{3}{4},\, \frac{1}{3} < x_2 < 2\}$.

We find the probability by integrating the double integral over the relevant region, i.e.,

$$
\begin{aligned}
P\left(0 < X_1 < \tfrac{3}{4},\ \tfrac{1}{3} < X_2 < 2\right) &= \int_{\frac{1}{3}}^{2} \int_{0}^{\frac{3}{4}} f(x_1,\ x_2)\, dx_1\, dx_2 \\
&= \int_{\frac{1}{3}}^{1} \int_{0}^{\frac{3}{4}} 6x_1^2 x_2\, dx_1\, dx_2 + \int_{\frac{1}{3}}^{1} \int_{0}^{\frac{3}{4}} 0\, dx_1\, dx_2 \\
&= \int_{\frac{1}{3}}^{1} \int_{0}^{\frac{3}{4}} 6x_1^2 x_2\, dx_1\, dx_2
\end{aligned}
$$

Integrate the inner integral first.

$$
\begin{aligned}
P\left(0 < X_1 < \tfrac{3}{4},\ \tfrac{1}{3} < X_2 < 2\right) &= \int_{\frac{1}{3}}^{1} \int_{0}^{\frac{3}{4}} 6x_1^2 x_2\, dx_1\, dx_2 \\
&= \int_{\frac{1}{3}}^{1} \left(2x_1^3 x_2 \,\Big|_{0}^{\frac{3}{4}} \right)\ dx_2 \\
&= \int_{\frac{1}{3}}^{1} \left((2) \left(\frac{3}{4} \right)^3 x_2 - 0 \right) dx_2 \\
&= \int_{\frac{1}{3}}^{1} \left((2) \left(\frac{27}{64} \right) x_2 \right) dx_2 \\
&= \int_{\frac{1}{3}}^{1} \frac{54}{64} x_2\, dx_2
\end{aligned}
$$

Now integrate the remaining integral

$$
\begin{aligned}
P\left(0 < X_1 < \tfrac{3}{4},\ \tfrac{1}{3} < X_2 < 2\right) &= \int_{\frac{1}{3}}^{1} \frac{54}{64} x_2\, dx_2 \\
&= \frac{54}{128} x_2^2 \,\Big|_{\frac{1}{3}}^{1} \\
&= \left(\frac{54}{128} \right) (1) - \left(\frac{54}{128} \right) \left(\frac{1}{9} \right) \\
&= \left(\frac{54}{128} \right) (1) - \left(\frac{6}{128} \right) \\
&= \frac{48}{128} = \frac{3}{8}
\end{aligned}
$$

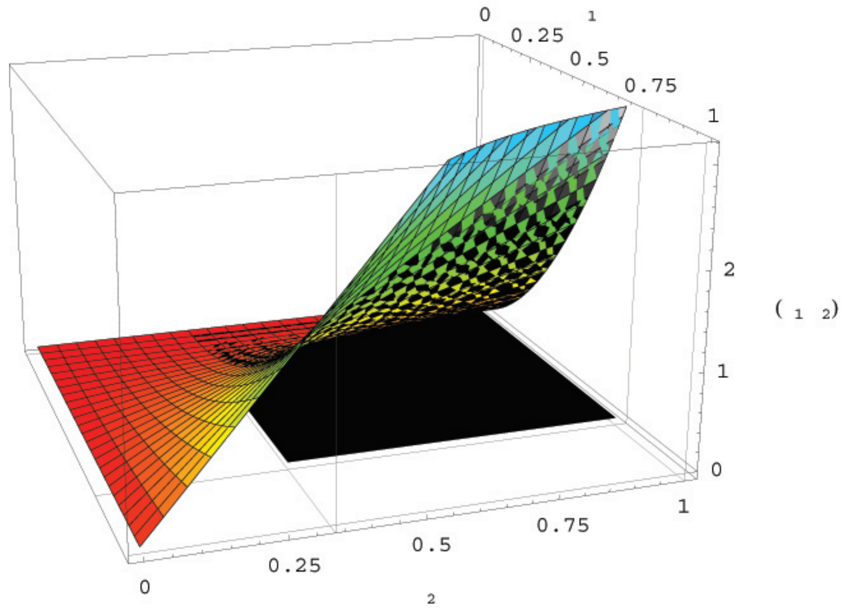
This probability is the volume under the surface $f(x_1, x_2) = 6x_1^2 x_2$ and above the rectangular set

$$\{(x_1, x_2) \mid 0 < x_1 < \tfrac{3}{4}, \quad \tfrac{1}{3} < x_2 < 1\}$$

in the x_1x_2-plane.

We can see this area in figure 1

FIGURE 1. Probability that $\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right)$



4.4. **Definition of a Joint Distribution Function.** If X_1 and X_2 are continuous random variables, the function given by

$$F(x_1, x_2) = P\left[X_1 \le x_1, X_2 \le x_2\right] = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(u_1, u_2)\, du_1\, du_2 \qquad \begin{matrix} -\infty < x_1 < \infty \\ -\infty < x_2 < \infty \end{matrix} \tag{4}$$

where $f(u_1, u_2)$ is the value of the joint probability function of X_1 and X_2 at (u_1, u_2) is called the joint distribution function, or the joint cumulative distribution of X_1 and X_2.

If the joint distribution function is continuous everywhere and partially differentiable with respect to x_1 and x_2 for all but a finite set of values then

$$f(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2) \tag{5}$$

wherever these partial derivatives exist.

4.5. **Properties of the Joint Distribution Function.**

Theorem 3. *If X_1 and X_2 are random variables with joint distribution function $F(x_1,\, x_2)$, then*

(i) $F(-\infty,\, -\infty) = F(-\infty,\, x_2) = F(x_1,\, -\infty) = 0$
(ii) $F(\infty,\, \infty) = 1$
(iii) *If $a < b$ and $c < d$, then $F(a,\, c) < F(b,\, d)$*
(iv) *If $a > x_1$ and $b > x_2$, then $F(a,\, b) - F(a,\, x_2) - F(x_1,\, b) + F\ (x_1,\, x_2) \geq 0$*

Part (iv) follows because

$$F(a,\, b) - F(a,\, x_2) - F(x_1,\, b) + F(x_1,\, x_2) = P\left[x_1 < X_1 \leq a,\, x_2 < X_2 \leq b\right] \geq 0$$

Note also that

$$F(\infty,\, \infty) \equiv \lim_{x_1 \to \infty} \lim_{x_2 \to \infty} F(x_1,\, x_2) = 1$$

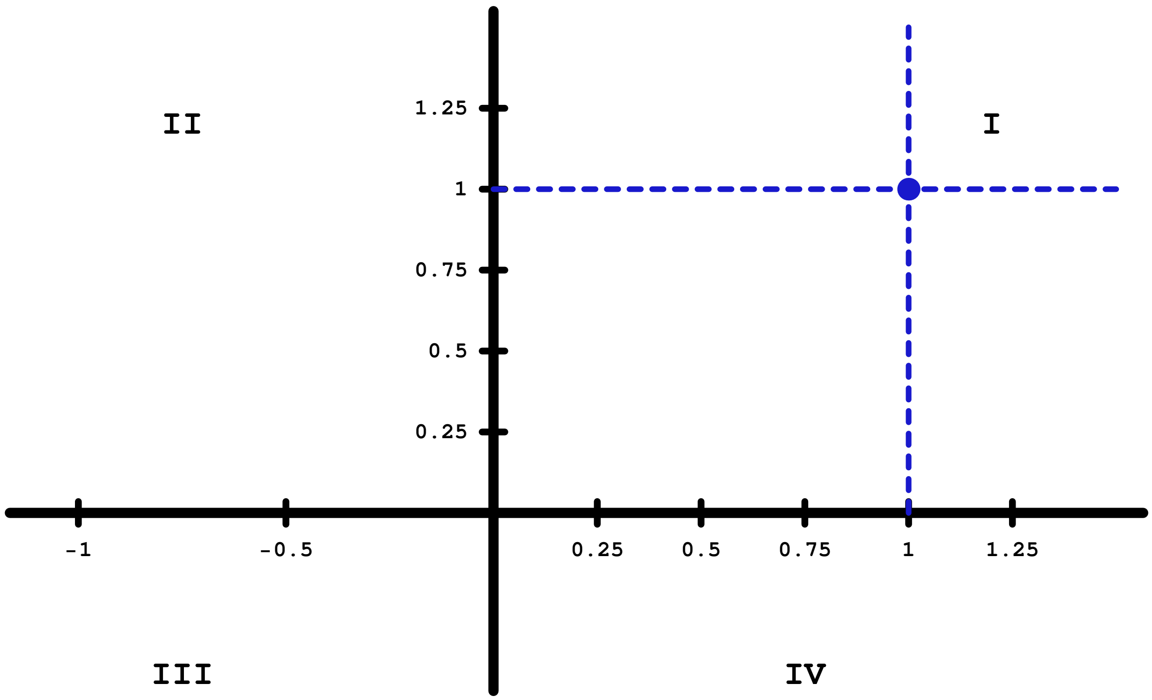
implies that the joint density function $f(x_1,\, x_2)$ must be such that the integral of $f(x_1,\, x_2)$ over all values of $(x_1,\, x_2)$ is 1.

4.6. **Examples of a Joint Distribution Function and Density Functions.**

4.6.1. *Deriving a Distribution Function from a Joint Density Function.* Consider a joint density function for X_1 and X_2 given by

$$f(x_1,\, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

This has a positive value in the square bounded by the horizontal and vertical axes and the vertical and horizontal lines at one. It is zero elsewhere. We will therefore need to find the value of the distribution function for five different regions: second, third and fourth quadrants, square defined by the vertical and horizontal lines at one, area between the vertical axis and a vertical line at one and above a horizontal line at one in the first quadrant, area between the horizontal axis and a horizontal line at one and to the right of a vertical line at one in the first quadrant, the area in the first quadrant not previously mentioned. This can be diagrammed as follows.



We find the distribution function by integrating the joint density function.
If either $x_1 < 0$ or $x_2 < 0$, it follows that

$$F(x_1,\, x_2) = 0$$

For $0 < x_1 < 1$ and $0 < x_2 < 1$, we get

$$F(x_1,\, x_2) = \int_0^{x_2} \int_0^{x_1} (s+t)\, ds\, dt = \frac{1}{2} x_1 x_2 (x_1 + x_2)$$

for $x_1 > 1$ and $0 < x_2 < 1$, we get

$$F(x_1,\, x_2) = \int_0^{x_2} \int_0^{1} (s+t) ds\, dt = \frac{1}{2} x_2 (x_2 + 1)$$

for $0 < x_1 < 1$ and $x_2 > 1$, we get

$$F(x_1,\, x_2) = \int_0^{1} \int_0^{x_1} (s+t) ds\, dt = \frac{1}{2} x_1 (x_1 + 1)$$

and for $x_1 > 1$ and $x_2 > 1$ we get

$$F(x_1,\, x_2) = \int_0^{1} \int_0^{1} (s+t)\, ds\, dt = 1$$

Because the joint distribution function is everywhere continuous, the boundaries between any two of these regions can be included in either one, and we can write

$$F(x_1,\,x_2) = \begin{cases} 0 & \text{for } x_1 \le 0 \text{ or } x_2 \le 0 \\ \frac{1}{2}x_1x_2(x_1+x_2) & \text{for } 0 < x_1 < 1,\, 0 < x_2 < 1 \\ \frac{1}{2}x_2(x_2+1) & \text{for } x_1 \ge 1,\, 0 < x_2 < 1 \\ \frac{1}{2}x_1(x_1+1) & \text{for } 0 < x_1 < 1,\, x_2 \ge 1 \\ 1 & \text{for } x_1 \ge 1,\, x_2 \ge 1 \end{cases}$$

4.7. **Deriving a Joint Density Function from a Distribution Function.** Consider two random variables X_1 and X_2 whose joint distribution function is given by

$$F(x_1,\,x_2) = \begin{cases} (1-\mathrm{e}^{-x_1})\,(1-\mathrm{e}^{-x_2}) & \text{for } x_1 > 0 \text{ and } x_2 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Partial differentiation yields

$$\frac{\partial^2}{\partial x_1 \partial x_2} F(x_1,\,x_2) = \mathrm{e}^{-(x_1+x_2)}$$

For $x_1 > 0$ and $x_2 > 0$ and 0 elsewhere we find that the joint probability density of X_1 and X_2 is given by

$$f(x_1,\,x_2) = \begin{cases} \mathrm{e}^{-(x_1+x_2)} & \text{for } x_1 > 0 \text{ and } x_2 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

5. Multivariate Distributions for Continuous Random Variables

5.1. **Joint Density of Several Random Variables.** The k-dimensional random variable $(X_1,\,X_2,\,\ldots,\,X_k)$ is said to be a k-dimensional random variable if there exists a function $f(\cdot,\,\cdot,\,\ldots\,,\,\cdot) \ge 0$ such that

$$F(x_1,\,x_2,\,\ldots\,,\,x_k) = \int_{-\infty}^{x_k}\int_{-\infty}^{x_{k-1}}\cdots\int_{-\infty}^{x_1} f(u_1,\,u_2,\,\ldots u_k)\,du_1\,\ldots\,du_k \tag{6}$$

for all $(x_1,\,x_2,\,\ldots,\,x_k)$ where

$$F(x_1,\,x_2,\,x_3,\,\ldots) = P\left[X_1 \le x_1,\,X_2 \le x_2,\,X_3 \le x_3,\,\ldots\right]$$

The function $f(\cdot)$ is defined to be a joint probability density function. It has the following properties:

$$\begin{gathered} f(x_1,\,x_2,\,\ldots\,,\,x_k) \ge 0 \\ \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\,x_2,\,\ldots\,,\,x_k)\,dx_1\ldots dx_k = 1 \end{gathered} \tag{7}$$

In order to make it clear the variables over which f is defined it is sometimes written

$$f(x_1,\,x_2,\,\ldots\,,\,x_k) = f_{X_1,\,X_2,\,\ldots\,,\,X_k}(x_1,\,x_2,\,\ldots\,,\,x_k) \tag{8}$$

6. Marginal Distributions

6.1. **Example Problem.** Consider the example of tossing a coin and rolling a die from section 1. The probability of any particular pair, (x_1, x_2) is given in the Table 6.

TABLE 6. Joint and Marginal Probabilities of X_1 and X_2.

		X_2							
		1	2	3	4	5	6	7	
X_1	1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	$\frac{1}{6}$
	3	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	

Notice that we have summed the columns and the rows and placed these sums at the bottom and right hand side of the table. The sum in the first column is the probability that $X_2 = 1$. The sum in the sixth row is the probability that $X_1 = 6$. Specifically the column totals are the probabilities that X_2 will take on the values 1, 2, 3, ... ,7. They are the values

$$g(x_2) = \sum_{x_1=1}^{6} p(x_1, x_2) \quad \text{for } x_2 = 1,\, 2,\, 3,\, \dots\, ,\, 7$$

In the same way, the row totals are the probabilities that X_1 will take on the values in its space.

Because these numbers are computed in the margin of the table, they are called marginal probabilities.

6.2. **Marginal Distributions for Discrete Random Variables.** If X_1 and X_2 are discrete random variables and $p(x_1, x_2)$ is the value of their joint distribution function at (x_1, x_2), the function given by

$$g(x_1) = \sum_{x_2} p(x_1, x_2) \tag{9}$$

for each x_1 within the range of X_1 is called the marginal distribution of X_1. Correspondingly, the function given by

$$h(x_2) = \sum_{x_1} p(x_1,\, x_2) \tag{10}$$

for each x_2 within the range of X_2 is called the marginal distribution of X_2.

6.3. **Marginal Distributions for Continuous Random Variables.** If X and Y are jointly continuous random variables, then the functions $f_X(\cdot)$ and $f_Y(\cdot)$ are called the marginal probability density functions. The subscripts remind us that f_X is defined for the random variable X. Intuitively, the marginal density is the density that results when we ignore any information about the random outcome Y. The marginal densities are obtained by integration of the joint density

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)\, dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)\, dx \end{aligned} \tag{11}$$

In a similar fashion for a k-dimensional random variable X

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1,\, x_2,\, \dots)\, dx_2\, dx_3 \dots\, dx_k \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1,\, x_2,\, \dots)\, dx_1\, dx_3\, \dots\, dx_k \end{aligned} \tag{12}$$

6.4. **Example 1.** Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1,\, x_2) = \begin{cases} 2x_2 \mathrm{e}^{-x_1} & x_1 \ge 0, \quad 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What are the marginal densities of x_1 and x_2?
First find the marginal density for x_1.

$$\begin{aligned} f_1(x_1) &= \int_0^1 2x_2 \mathrm{e}^{-x_1}\, dx_2 \\ &= x_2^2 \mathrm{e}^{-x_1} \Big|_0^1 \\ &= \mathrm{e}^{-x_1} - 0 \\ &= \mathrm{e}^{-x_1} \end{aligned}$$

Now find the marginal density for x_2.

$$
\begin{aligned}
f_2(x_2) &= \int_0^\infty 2x_2 \mathrm{e}^{-x_1}\, dx_1 \\
&= -2x_2 \mathrm{e}^{-x_1} \Big|_0^\infty \\
&= 0 - \left(-2x_2 \mathrm{e}^0\right) \\
&= 2x_2 e^0 \\
&= 2x_2
\end{aligned}
$$

6.5. **Example 2.** Let the joint density of two random variables x and y be given by

$$
f(x,\, y) = \begin{cases} \frac{1}{6}(x + 4y) & 0 < x < 2, \quad 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}
$$

What are the marginal densities of x and y?
First find the marginal density for x.

$$
\begin{aligned}
f_X(x) &= \int_0^1 \frac{1}{6}(x + 4y)\, dy \\
&= \frac{1}{6}(xy + 2y^2) \Big|_0^1 \\
&= \frac{1}{6}(x + 2) - \frac{1}{6}(0) \\
&= \frac{1}{6}(x + 2)
\end{aligned}
$$

Now find the marginal density for y.

$$
\begin{aligned}
f_Y(y) &= \int_0^2 \frac{1}{6}(x + 4y)\, dx \\
&= \frac{1}{6}\left(\frac{x^2}{2} + 4xy\right) \Big|_0^2 \\
&= \frac{1}{6}\left(\frac{4}{2} + 8y\right) - \frac{1}{6}(0) \\
&= \frac{1}{6}(2 + 8y)
\end{aligned}
$$

7. Conditional Distributions

7.1. **Conditional Probability Functions for Discrete Distributions.** We have previously shown that the conditional probability of A given B can be obtained by dividing the probability of the intersection by the probability of B, specifically,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{13}$$

Now consider two random variables X and Y. We can write the probability that $X = x$ and $Y = y$ as

$$\begin{aligned} P(X = x \mid Y = y) &= \frac{P(X = x,\, Y = y)}{P(Y = y)} \\ &= \frac{p(x,\, y)}{h(y)} \end{aligned} \tag{14}$$

provided $P(Y = y) \neq 0$, where $p(x,\, y)$ is the value of joint probability distribution of X and Y at $(x,\, y)$ and $h(y)$ is the value of the marginal distribution of Y at y. We can then define a conditional distribution of X given $Y = y$ as follows.

If $p(x,\, y)$ is the value of the joint probability distribution of the discrete random variables X and Y at $(x,\, y)$ and $h(y)$ is the value for the marginal distribution of Y at y, then the function given by

$$p(x \mid y) = \frac{p(x,\, y)}{h(y)} \quad h(y) \neq 0 \tag{15}$$

for each x within the range of X, is called the conditional distribution of X given $Y = y$.

7.2. **Example for discrete distribution.** Consider the example of tossing a coin and rolling a die from section 1. The probability of any particular pair, $(x_1,\, x_2)$ is given in the following table where x_1 is the value on the die and x_2 is the sum of the number on the die and an indicator that is one if the coin is a head and zero otherwise. The data is in the Table 6 is repeated here for convenience.

Consider the probability that $x_1 = 3$ given that $x_2 = 4$. We compute this as follows.

For the example, $P(X_1 = 3,\ X_2 = 3) = \frac{1}{12}$, which is the probability of sample point E_9.

$$\begin{aligned} p(x_1 \mid x_2) = p(3 \mid 4) &= \frac{p(x_1,\, x_2)}{h(x_2)} = \frac{p(3,\, 4)}{h(4)} \\ &= \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{2} \end{aligned}$$

We can then make a table for the conditional probability function for x_1.

We do the same for X_2 given X_1 in table 8.

7.3. **Conditional Distribution Functions for Continuous Distributions.**

TABLE 6. Joint and Marginal Probabilities of X_1 and X_2.

	X_2								
		1	2	3	4	5	6	7	
X_1	1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	$\frac{1}{6}$
	3	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	

TABLE 7. Probability Function for X_1 given X_2.

	X_2								
		1	2	3	4	5	6	7	
X_1	1	1	$\frac{1}{2}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	$\frac{1}{6}$
	3	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{2}$	1	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	

7.3.1. *Discussion.* In the continuous case, the idea of a conditional distribution takes on a slightly different meaning than in the discrete case. If X_1 and X_2 are both continuous, $P(X_1 = x_1 \mid X_2 = x_2)$ is not defined because the probability of any one point is identically zero. It make sense however to define a conditional distribution function, i.e.,

$$P(X_1 \leq x_1 \mid X_2 = x_2)$$

TABLE 8. Probability Function for X_2 given X_1.

		X_2							
		1	2	3	4	5	6	7	
	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	$\frac{1}{6}$
X_1	3	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	

because the value of X_2 is known when we compute the value the probability that X_1 is less than some specific value.

7.3.2. *Definition of a Continuous Distribution Function.* If X_1 and X_2 are jointly continuous random variables with joint density function $f(x_1, x_2)$, then the conditional distribution function of X_1 given $X_2 = x_2$ is

$$F\,(x_1 \mid x_2) = P(X_1 \le x_1 \mid X_2 = x_2) \tag{16}$$

We can obtain the unconditional distribution function by integrating the conditional one over x_2. This is done as follows.

$$F\,(x_1) = \int_{-\infty}^{\infty} F(x_1 \mid x_2) f_{X_2}(x_2)\, dx_2 \tag{17}$$

We can also find the probability that X_1 is less than x_1 is the usual fashion as

$$F\,(x_1) = \int_{-\infty}^{x_1} f_{X_1}(t_1)\, dt_1 \tag{18}$$

But the marginal distribution inside the integral is obtained by integrating the joint density over the range of x_2. Specifically,

$$f_{X_1}\,(t_1) = \int_{-\infty}^{\infty} f_{X_1 X_2}(t_1,\, x_2)\, dx_2 \tag{19}$$

This implies then that

$$F\,(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_1} f_{X_1 X_2}(t_1,\, x_2)\, dt_1\, dx_2 \tag{20}$$

Now compare the integrand in equation 20 with that in equation 17 to conclude that

$$\begin{aligned} F\,(x_1 \mid x_2)\, f_{X_2}(x_2) &= \int_{-\infty}^{x_1} f_{X_1 X_2}(t_1,\, x_2)\, dt_1 \\ \Rightarrow F\,(x_1 \mid x_2) &= \int_{-\infty}^{x_1} \frac{f_{X_1 X_2}(t_1\,, x_2)\, dt_1}{f_{X_2}(x_2)} \end{aligned} \tag{21}$$

We call the integrand in the second line of (21) the conditional density function of X_1 given $X_2 = x_2$. We denote it by $f(x_1 \mid x_2)$ or$f_{X_1|X_2}(x_1 \mid x_2)$. Specifically

Let X_1 and X_2 be jointly continuous random variables with joint probability density $f_{X_1 X_2}(x_1,\, x_2)$ and marginal densities $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, respectively. For any x_2 such that $f_{X_2}(x_2) > 0$, the conditional probability density function of X_1 given $X_2 = x_2$, is defined to be

$$\begin{aligned} f_{X_1|X_2}(x_1 \mid x_2) &= \frac{f_{X_1 X_2}(x_1,\, x_2)}{f_{X_2}(x_2)} \\ &= \frac{f(x_1,\, x_2)}{f(x_2)} \end{aligned} \tag{22}$$

And similarly

$$\begin{aligned} f_{X_2|X_1}(x_2 \mid x_1) &= \frac{f_{X_1 X_2}(x_1,\, x_2)}{f_{X_1}(x_1)} \\ &= \frac{f(x_1,\, x_2)}{f(x_1)} \end{aligned} \tag{23}$$

7.4. **Example.** Let the joint density of two random variables x and y be given by

$$f(x,\, y) = \begin{cases} \frac{1}{6}(x + 4y) & 0 < x < 2,\ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density of x is $f_X(x) = \frac{1}{6}(x + 2)$ while the marginal density of y is $f_y(y) = \frac{1}{6}(2 + 8y)$.

Now find the conditional distribution of x given y. This is given by

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f(x,\, y)}{f(y)} = \frac{\frac{1}{6}(x + 4y)}{\frac{1}{6}(2 + 8y)} \\ &= \frac{(x + 4y)}{(8y + 2)} \end{aligned}$$

for $0 < x < 2$ and $0 < y < 1$. Now find the probability that $X \leq 1$ given that $y = \frac{1}{2}$. First determine the density function when $y = \frac{1}{2}$ as follows

$$\begin{aligned}\frac{f(x,\,y)}{f(y)} &= \frac{(x+4y)}{(8y+2)} \\ &= \frac{\left(x+4\left(\frac{1}{2}\right)\right)}{\left(8\left(\frac{1}{2}\right)+2\right)} \\ &= \frac{(x+2)}{(4+2)} = \frac{(x+2)}{6}\end{aligned}$$

Then

$$\begin{aligned}P\left(X \le 1 \mid Y = \tfrac{1}{2}\right) &= \int_0^1 \frac{1}{6}(x+2)\,dx \\ &= \frac{1}{6}\left(\frac{x^2}{2} + 2x\right)\Big|_0^1 \\ &= \frac{1}{6}\left(\frac{1}{2} + 2\right) - 0 \\ &= \frac{1}{12} + \frac{2}{6} = \frac{5}{12}\end{aligned}$$

8. Independent Random Variables

8.1. **Discussion.** We have previously shown that two events A and B are independent if the probability of their intersection is the product of their individual probabilities, i.e.

$$P\,(A \cap B) = P(A)P(B) \tag{24}$$

In terms of random variables, X and Y, consistency with this definition would imply that

$$P\,(a \le X \le b,\, c \le Y \le d) = P\,(a \le X \le b)\,P\,(c \le Y \le d) \tag{25}$$

That is, if X and Y are independent, the joint probability can be written as the product of the marginal probabilities. We then have the following definition.

Let X have distribution function $F_X(x)$, Y have distribution function $F_Y(y)$, and X and Y gave joint distribution function $F(x,\,y)$. Then X and Y are said to be independent if, and only if,

$$F(x,\,y) = F_X(X)F_Y(y) \tag{26}$$

for every pair of real numbers $(x,\,y)$. If X and Y are not independent, they are said to be dependent.

8.2. **Independence Defined in Terms of Density Functions.**

8.2.1. *Discrete Random Variables.* If X and Y are discrete random variables with joint probability density function $p(x, y)$ and marginal density functions $p_X(x)$ and $p_Y(y)$, respectively, then X and Y are independent if, and only if

$$\begin{aligned} p_{X,Y}(x, y) &= p_X(x)p_Y(y) \\ &= p(x)p(y) \end{aligned} \tag{27}$$

for all pairs of real numbers (x, y).

8.2.2. *Continuous Bivariate Random Variables.* If X and Y are continuous random variables with joint probability density function $f(x, y)$ and marginal density functions $f_X(x)$ and $f_Y(y)$, respectively then X and Y are independent if and only if

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x)f_Y(y) \\ &= f(x)f(y) \end{aligned} \tag{28}$$

for all pairs of real numbers (x, y).

8.3. **Continuous Multivariate Random Variables.** In a more general context the variables $X_1, X_2, \ldots, X_k$ are independent if, and only if

$$\begin{aligned} f_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) &= \prod_{i=1}^{k} f_{X_i}(x_i) \\ &= f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_k}(x_k) \\ &= f(x_1)f(x_2)\ldots f(x_k) \end{aligned} \tag{29}$$

In other words two random variables are independent if the joint density is equal to the product of the marginal densities.

8.4. **Examples.**

8.4.1. *Example 1 — Rolling a Die and Tossing a Coin.* Consider the previous example where we rolled a die and tossed a coin. X_1 is the number on the die, X_2 is the number of the die plus the value of the indicator on the coin ($H = 1$). Table 8 is repeated here for convenience. For independence $p(x, y) = p(x)p(y)$ for all values of x_1 and x_2.

To show that the variables are not independent, we only need show that

$$p(x, y) \neq p(x)p(y)$$

Consider $p(1, 2) = \frac{1}{2}$. If we multiply the marginal probabilities we obtain

$$\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36} \neq \frac{1}{2}$$

TABLE 8. Probability Function for X_2 given X_1.

		X_2							
		1	2	3	4	5	6	7	
X_1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	$\frac{1}{6}$
	3	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	

8.4.2. *Example 2 — A Continuous Multiplicative Joint Density.* Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1 x_2) = \begin{cases} 2x_2 e^{-x_1} & x_1 \geq 0,\ 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density for x_1 is given by

$$\begin{aligned} f_1(x_1) &= \int_0^1 2x_2 e^{-x_1}\, dx_2 \\ &= x_2^2 e^{-x_1} \Big|_0^1 \\ &= e^{-x_1} - 0 \\ &= e^{-x_1} \end{aligned}$$

The marginal density for x_2 is given by

$$\begin{aligned} f_2(x_2) &= \int_0^\infty 2x_2 \mathrm{e}^{-x_1}\, dx_1 \\ &= -2x_2 \mathrm{e}^{-x_1} \Big|_0^\infty \\ &= 0 - \left(-2x_2 \mathrm{e}^0\right) \\ &= 2x_2 \mathrm{e}^0 \\ &= 2x_2 \end{aligned}$$

It is clear the joint density is the product of the marginal densities.

8.4.3. *Example 3.* Let the joint density of two random variables x and y be given by

$$f(x,\, y) = \begin{cases} \dfrac{-3x^2 \log[y]}{2\left(1 + \log[2] - 2\log[4]\right)} & 0 \le x \le 1,\ 2 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

First find the marginal density for x.

$$\begin{aligned} f_X(x) &= \int_0^1 \frac{-3x^2 \log[y]}{2\left(1 + \log[2] - 2\log[4]\right)}\, dy \\ &= \frac{-3x^2 y\left(\log[y] - 1\right)}{2\left(1 + \log[2] - 2\log[4]\right)} \Bigg|_2^4 \\ &= \frac{-3x^2 4\left(\log[4] - 1\right) + 3x^2 2\left(\log[2] - 1\right)}{2\left(1 + \log[2] - 2\log[4]\right)} \\ &= \frac{3x^2\Big(2\left(\log[2] - 1\right) - 4\left(\log[4] - 1\right)\Big)}{2\left(1 + \log[2] - 2\log[4]\right)} \\ &= \frac{3x^2\left(2\log[2] - 2 - 4\log[4] + 4\right)}{2\left(1 + \log[2] - 2\log[4]\right)} \\ &= \frac{3x^2\left(2\log[2] - 4\log[4] + 2\right)}{2\left(1 + \log[2] - 2\log[4]\right)} \\ &= \frac{3x^2\Big(2\left(1 + \log[2] - 2\log[4]\right)\Big)}{2\left(1 + \log[2] - 2\log[4]\right)} \\ &= 3x^2 \end{aligned}$$

Now find the marginal density for y.

$$\begin{aligned} f_Y(y) &= \int_0^1 \frac{-3x^2 \log[y]}{2\big(1+\log[2]-2\log[4]\big)}\, dx \\ &= \frac{-x^3 \log[y]}{2\big(1+\log[2]-2\log[4]\big)} \Big|_0^1 \\ &= \frac{-\log[y]+0}{2\big(1+\log[2]-2\log[4]\big)} \\ &= \frac{-\log[y]}{2\big(1+\log[2]-2\log[4]\big)} \end{aligned}$$

It is clear the joint density is the product of the marginal densities.

8.4.4. *Example 4.* Let the joint density of two random variables X and Y be given by

$$f(x,\, y) = \begin{cases} \frac{3}{5}x^2 + \frac{3}{10}y & 0 \le x \le 1,\ 0 \le y \le 2 \\ \\ 0 & \text{otherwise} \end{cases}$$

First find the marginal density for x.

$$\begin{aligned} f_X(x) &= \int_0^2 \left(\frac{3}{5}x^2 + \frac{3}{10}y\right) dy \\ &= \left(\frac{3}{5}x^2 y + \frac{3}{20}y^2\right) \Big|_0^2 \\ &= \left(\frac{6}{5}x^2 + \frac{12}{20}\right) - 0 \\ &= \frac{6}{5}x^2 + \frac{3}{5} \end{aligned}$$

Now find the marginal density for y.

$$\begin{aligned} f_Y(y) &= \int_0^1 \left(\frac{3}{5}x^2 + \frac{3}{10}y\right) dx \\ &= \left(\frac{3}{15}x^3 + \frac{3}{10}xy\right) \Big|_0^1 \\ &= \left(\frac{3}{15} + \frac{3}{10}y\right) - 0 \\ &= \left(\frac{1}{5} + \frac{3}{10}y\right) \end{aligned}$$

The product of the marginal densities is not the joint density.

8.4.5. *Example 5.* Let the joint density of two random variables X and Y be given by

$$f(x,\, y) = \begin{cases} 2\mathrm{e}^{-(x+y)} & 0 \le x \le y,\ 0 \le y \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal density of X.

$$\begin{aligned} f_X(x) &= \int_x^\infty 2\mathrm{e}^{-(x+y)}\, dy \\ &= -2\mathrm{e}^{-(x+y)} \Big|_x^\infty \\ &= -2\mathrm{e}^{-(x+\infty)} - \left(-2\mathrm{e}^{-(x+x)}\right) \\ &= 0 + 2\mathrm{e}^{-2x} \\ &= 2\mathrm{e}^{-2x} \end{aligned}$$

The marginal density of Y is obtained as follows.

$$\begin{aligned} f_Y(y) &= \int_0^y 2\mathrm{e}^{-(x+y)}\, dx \\ &= -2\mathrm{e}^{-(x+y)} \Big|_0^y \\ &= -2\mathrm{e}^{-(y+y)} - \left(-2\mathrm{e}^{-(0+y)}\right) \\ &= -2\mathrm{e}^{-2y} + 2\mathrm{e}^{-y} \\ &= 2\mathrm{e}^{-y}\left(1 - \mathrm{e}^{-y}\right) \end{aligned}$$

We can show that this is a proper density function by integrating it over the range of x and y.

$$\begin{aligned}
\int_0^\infty \int_x^\infty 2e^{-(x+y)}\,dy\,dx &= \int_0^\infty \left[-2e^{-(x+y)} \Big|_x^\infty \right] dx \\
&= \int_0^\infty 2e^{-2x}\,dx \\
&= -e^{-2x} \Big|_0^\infty \\
&= -e^{-\infty} - \left[-e^0 \right] \\
&= 0 + 1 = 1
\end{aligned}$$

Or in the other order as follows.

$$\begin{aligned}
\int_0^\infty \int_0^y 2e^{-(x+y)}\,dx\,dy &= \int_0^\infty \left[-2e^{-(x+y)} \Big|_0^y \right] dy \\
&= \int_0^\infty \left[2e^{-y} - 2e^{-2y} \right] dy \\
&= -2e^{-y} \Big|_0^\infty - \left[-e^{-y} \right] \Big|_0^\infty \\
&= \left[-2e^{-\infty} + 2 \right] - \left[-e^{-\infty} + 1 \right] \\
&= [0 + 2] - [0 + 1] \\
&= 2 - 1 = 1
\end{aligned}$$

8.5. **Separation of a Joint Density Function.**

8.5.1. *Theorem 4.*

Theorem 4. *Let X_1 and X_2 have a joint density function $f(x_1, x_2)$ that is positive if, and only if, $a \leq x_1 \leq b$ and $c \leq x_2 \leq d$, for constants a, b, c and d; and $f(x_1, x_2) = 0$ otherwise. Then X_1 and X_2 are independent random variables if, and only if*

$$f(x_1, x_2) = g(x_1)h(x_2)$$

where $g(x_1)$ is a non-negative function of x_1 alone and $h(x_2)$ is a non-negative function of x_2 alone.

Thus if we can separate the joint density into two multiplicative terms, one depending on x_1 alone and one on x_2 alone, we know the random variables are independent without showing that these functions are actually the marginal densities.

8.5.2. *Example.* Let the joint density of two random variables x and y be given by

$$f(x, y) = \begin{cases} 8x & 0 \leq x \leq \frac{1}{2},\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can write $f(x, y)$ as $g(x)h(y)$, where

$$g(x) = \begin{cases} x & 0 \leq x \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$h(y) = \begin{cases} 8 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

These functions are not density functions because they do not integrate to one.

$$\int_0^{1/2} x\,dx = \frac{1}{2}x^2 \Big|_0^{1/2} = \frac{1}{8} \neq 1$$

$$\int_0^1 8\,dy = 8y \Big|_0^1 = 8 \neq 1$$

The marginal densities as defined below do sum to one.

$$f_X(x) = \begin{cases} 8x & 0 \leq x \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

9. Expected Value of a Function of Random Variables

9.1. **Definition.**

9.1.1. *Discrete Case.* Let $X = (X_1, X_2, \ldots, X_k)$ be a k-dimensional discrete random variable with probability function $p(x_1, x_2, \ldots, x_k)$. Let $g(\cdot, \cdot, \ldots, \cdot)$ be a function of the k random variables $(X_1, X_2, \ldots, X_k)$. Then the expected value of $g(X_1, X_2, \ldots, X_k)$ is

$$E\big[g(X_1, X_2, \ldots, X_k)\big] = \sum_{x_k}\sum_{x_{k-1}}\cdots\sum_{x_2}\sum_{x_1} g(x_1, \ldots, x_k)p(x_1, x_2, \ldots, x_k) \tag{30}$$

9.1.2. *Continuous Case.* Let $X = (X_1, X_2, \ldots, X_k)$ be a k-dimensional random variable with density $f(x_1, x_2, \ldots, x_k)$. Let $g(\cdot, \cdot, \ldots, \cdot)$ be a function of the k random variables $(X_1, X_2, \ldots, X_k)$. Then the expected value of $g(X_1, X_2, \ldots, X_k)$ is

$$\begin{aligned} E\left[g(X_1, X_2 \ldots, X_k)\right] &= \int_{x_k}\int_{x_{k-1}}\cdots\int_{x_2}\int_{x_1} g(x_1, \ldots, x_k) f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) dx_1 \ldots dx_k \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, \ldots, x_k) f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)\, dx_1 \ldots dx_k \end{aligned} \tag{31}$$

if the integral is defined.

Similarly, if $g(X)$ is a bounded real function on the interval $[a, b]$ then

$$E(g(X)) = \int_a^b g(x)\, dF(x) = \int_a^b g\, dF \tag{32}$$

where the integral is in the sense of Lebesque and can be loosely interpreted as $f(x)\, dx$.

Consider as an example $g(x_1, \ldots, x_k) = x_i$. Then

$$\begin{aligned} E\left[g(X_1, \ldots, X_k)\right] \equiv E[X_i] &\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k \\ &\equiv \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\, dx_i \end{aligned} \tag{33}$$

because integration over all the other variables gives the marginal density of x_i.

9.2. **Example.** Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1 x_2) = \begin{cases} 2x_2 e^{-x_1} & x_1 \geq 0,\ 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density for x_1 is given by

$$
\begin{aligned}
f_1(x_1) &= \int_0^1 2x_2 \mathrm{e}^{-x_1} \, dx_2 \\
&= x_2^2 \mathrm{e}^{-x_1} \Big|_0^1 \\
&= \mathrm{e}^{-x_1} - 0 \\
&= \mathrm{e}^{-x_1}
\end{aligned}
$$

The marginal density for x_2 is given by

$$
\begin{aligned}
f_2(x_2) &= \int_0^\infty 2x_2 \mathrm{e}^{-x_1} \, dx_1 \\
&= -2x_2 \mathrm{e}^{-x_1} \Big|_0^\infty \\
&= 0 - \left(-2x_2 \mathrm{e}^{0}\right) \\
&= 2x_2 \mathrm{e}^{0} = 2x_2
\end{aligned}
$$

We can find the expected value of X_1 by integrating the joint density or the marginal density. First with the joint density.

$$
E\left[X_1\right] = \int_0^1 \int_0^\infty 2x_1 x_2 \mathrm{e}^{-x_1} \, dx_1 \, dx_2
$$

Consider the inside integral first. We will need a $u\,dv$ substitution to evaluate the integral. Let

$$
u = 2x_1 x_2 \quad \text{and} \quad dv = \mathrm{e}^{-x_1} \, dx_1
$$

then

$$
du = 2x_2 \, dx_1 \quad \text{and} \quad v = -\mathrm{e}^{-x_1}
$$

Then

$$
\begin{aligned}
\int_0^\infty 2x_1 x_2 \mathrm{e}^{-x_1} \, dx_1 &= -2x_1 x_2 \mathrm{e}^{-x_1} \Big|_0^\infty - \int_0^\infty -2x_2 \mathrm{e}^{-x_1} \, dx_1 \\
&= -2x_1 x_2 \mathrm{e}^{-x_1} \Big|_0^\infty + \int_0^\infty 2x_2 \mathrm{e}^{-x_1} \, dx_1 \\
&= 0 + -2x_2 \mathrm{e}^{-x_1} \Big|_0^\infty \\
&= 2x_2
\end{aligned}
$$

Now integrate with respect to x_2.

$$E[X_1] = \int_0^1 2x_2 \, dx_2$$
$$= x_2^2 \Big|_0^1 = 1$$

Now find it using the marginal density of x_1. Integrate as follows:

$$E[X_1] = \int_0^\infty x_1 \mathrm{e}^{-x_1} \, dx_1$$

We will need to use a $u \, dv$ substitution to evaluate the integral. Let

$$u = x_1 \quad \text{and} \quad dv = \mathrm{e}^{-x_1} \, dx_1$$

then

$$du = dx_1 \quad \text{and} \quad v = -\mathrm{e}^{-x_1}$$

Then

$$\begin{aligned}
\int_0^\infty x_1 \mathrm{e}^{-x_1} \, dx_1 &= -x_1 \mathrm{e}^{-x_1} \Big|_0^\infty - \int_0^\infty -\mathrm{e}^{-x_1} \, dx_1 \\
&= -x_1 \mathrm{e}^{-x_1} \Big|_0^\infty + \int_0^\infty \mathrm{e}^{-x_1} \, dx_1 \\
&= 0 + -\mathrm{e}^{-x_1} \Big|_0^\infty \\
&= -\mathrm{e}^{-\infty} - -\mathrm{e}^0 \\
&= 0 + 1 = 1
\end{aligned}$$

We can likewise show that the expected value of x_2 is $\frac{2}{3}$. Now consider $E[x_1 x_2]$. We can obtain it as

$$E[X_1 X_2] = \int_0^1 \int_0^\infty 2x_1 x_2^2 \mathrm{e}^{-x_1} \, dx_1 \, dx_2$$

Consider the inside integral first. We will need a $u \, dv$ substitution to evaluate the integral. Let

$$u = 2x_1 x_2^2 \quad \text{and} \quad dv = \mathrm{e}^{-x_1} \, dx_1$$

then

$$du = 2x_2^2 \, dx_1 \quad \text{and} \quad v = -\mathrm{e}^{-x_1}$$

Then

$$\begin{aligned}
\int_0^\infty 2x_1x_2^2\mathrm{e}^{-x_1}\,dx_1 &= -2x_1x_2^2\mathrm{e}^{-x_1}\,\big|_0^\infty - \int_0^\infty -2x_2^2\mathrm{e}^{-x_1}\,dx_1 \\
&= -2x_1x_2^2\mathrm{e}^{-x_1}\,\big|_0^\infty + \int_0^\infty 2x_2^2\mathrm{e}^{-x_1}\,dx_1 \\
&= 0 + -2x_2^2\mathrm{e}^{-x_1}\,\big|_0^\infty \\
&= 2x_2^2
\end{aligned}$$

Now integrate with respect to x_2.

$$\begin{aligned}
E\left[X_1X_2\right] &= \int_0^1 2x_2^2\,dx_2 \\
&= \frac{2}{3}x_2^3\bigg|_0^1 = \frac{2}{3}
\end{aligned}$$

9.3. **Properties of Expectation.**

9.3.1. *Constants.*

Theorem 5. *Let c be a constant. Then*

$$\begin{aligned}
E[c] &\equiv \int_x\int_y cf(x,\,y)\,dy\,dx \\
&\equiv c\int_x\int_y f(x,\,y)\,dy\,dx \\
&\equiv c
\end{aligned} \tag{34}$$

9.3.2. *Theorem.*

Theorem 6. *Let $g(X_1,\,X_2)$ be a function of the random variables X_1 and X_2 and let a be a constant. Then*

$$\begin{aligned}
E[ag(X_1,\,X_2)] &\equiv \int_{x_1}\int_{x_2} ag(x_1,\,x_2)f(x_1,\,x_2)\,dx_2\,dx_1 \\
&\equiv a\int_{x_1}\int_{x_2} g(x_1,\,x_2)fx_1,\,x_2)\,dx_2\,dx_1 \\
&\equiv aE[g(X_1,\,X_2)]
\end{aligned} \tag{35}$$

9.3.3. *Theorem.*

Theorem 7. *Let X and Y denote two random variables defined on the same probability space and let $f(x, y)$ be their joint density. Then*

$$\begin{aligned} E[aX + bY] &= \int_y \int_x (ax + by) f(x,\, y)\, dx\, dy \\ &= a \int_y \int_x x f(x,\, y)\, dx\, dy \\ &+ b \int_y \int_x y f(x,\, y)\, dx\, dy \\ &= aE[X] + bE[Y] \end{aligned} \tag{36}$$

In matrix notation we can write this as

$$E[a_1\, a_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [a_1\, a_2] \begin{bmatrix} E(x_1) \\ E(x_2) \end{bmatrix} = a_1\mu_1 + a_2\mu_2 \tag{37}$$

9.3.4. *Theorem.*

Theorem 8. *Let X and Y denote two random variables defined on the same probability space and let $g_1(X,\, Y),\ g_2(X,\, Y),\ g_3(X,\, Y),\ \ldots\ ,\ g_k(X,\, Y)$ be functions of $(X,\, Y)$. Then*

$$\begin{aligned} &E[g_1(X,\, Y) + g_2(X,\, Y) + \cdots + g_k(X,\, Y)] \\ &\qquad = E[g_1(X,\, Y)] + E[g_2(X,\, Y)] + \cdots + E[g_k(X,\, Y)] \end{aligned} \tag{38}$$

9.3.5. *Independence.*

Theorem 9. *Let X_1 and X_2 be independent random variables and $g(X_1)$ and $h(X_2)$ be functions of X_1 and X_2, respectively. Then*

$$E\left[g(X_1)h(X_2)\right] = E\left[g(X_1)\right] E\left[h(X_2)\right] \tag{39}$$

provided that the expectations exist.

Proof: Let $f(x_1,\, x_2)$ be the joint density of X_1 and X_2. The product $g(X_1)h(X_2)$ is a function of X_1 and X_2. Therefore we have

$$
\begin{aligned}
E\big[g(X_1)h(X_2)\big] &\equiv \int_{x_1}\int_{x_2} g(x_1)h(x_2)f(x_1,\, x_2)\,dx_2\,dx_1 \\
&\equiv \int_{x_1}\int_{x_2} g(x_1)h(x_2)f_{X_1}(x_1)f_{X_2}(x_2)\,dx_2\,dx_1 \\
&\equiv \int_{x_1} g(x_1)f_{X_1}(x_1)\left[\int_{x_2} h(x_2)f_{X_2}(x_2)\,dx_2\right]\,dx_1 \\
&\equiv \int_{x_1} g(x_1)f_{X_1}(x_1)\big(E\left[h(X_2)\right]\big)\,dx_1 \\
&\equiv E\left[h(X_2)\right]\int_{x_1} g(x_1)f_{X_1}(x_1)\,dx_1 \\
&\equiv E\left[h(X_2)\right]E\left[g(X_1)\right]
\end{aligned} \tag{40}
$$

10. Variance, Covariance and Correlation

10.1. **Variance of a Single Random Variable.** The variance of a random variable X with mean μ is given by

$$
\begin{aligned}
\operatorname{var}(X) \equiv \sigma^2 &\equiv E\left[\big(X - E(X)\big)^2\right] \\
&\equiv E\left[(X-\mu)^2\right] \\
&\equiv \int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx \\
&\equiv \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left[\int_{-\infty}^{\infty} x f(x) dx\right]^2 \\
&\equiv E(x^2) - E^2(x)
\end{aligned} \tag{41}
$$

The variance is a measure of the dispersion of the random variable about the mean.

10.2. **Covariance.**

10.2.1. *Definition.* Let X and Y be any two random variables defined in the same probability space. The covariance of X and Y, denoted $\operatorname{cov}[X,\, Y]$ or $\sigma_{X,\,Y}$, is defined as

$$\begin{aligned}
\operatorname{cov}[X,\,Y] &\equiv E\left[(X-\mu_X)(Y-\mu_Y)\right] \\
&\equiv E[XY] - E[\mu_X Y] - E[X\mu_Y] + E[\mu_Y \mu_X] \\
&\equiv E[XY] - E[X]E[Y] \\
&\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,\,y)\,dx\,dy - \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,\,y)\,dx\,dy \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,\,y)\,dx\,dy\right] \\
&\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,\,y)\,dx\,dy - \left[\int_{-\infty}^{\infty} xf_X(x,\,y)\,dx \cdot \int_{-\infty}^{\infty} yf_Y(x,\,y)\,dy\right]
\end{aligned} \tag{42}$$

The covariance measures the interaction between two random variables, but its numerical value is not independent of the units of measurement of X and Y. Positive values of the covariance imply that X increases when Y increases; negative values indicate X decreases as Y decreases.

10.2.2. *Examples.*

(i) Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1x_2) = \begin{cases} 2x_2\mathrm{e}^{-x_1} & x_1 \geq 0,\ 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We showed in Example 9.2 that

$$E\left[X_1\right] = 1$$

$$E\left[X_2\right] = \frac{2}{3}$$

$$E\left[X_1X_2\right] = \frac{2}{3}$$

The covariance is then given by

$$\begin{aligned}
\operatorname{cov}[X,\,Y] &\equiv E[XY] - E[X]E[Y] \\
&\equiv \frac{2}{3} - (1)\left(\frac{2}{3}\right) = 0
\end{aligned}$$

(ii) Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1x_2) = \begin{cases} \frac{1}{6}x_1 & 0 \leq x_1 \leq 2,\ \ 0 \leq x_2 \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

First compute the expected value of X_1X_2 as follows.

$$
\begin{aligned}
E\left[X_1 X_2\right] &= \int_0^3 \int_0^2 \frac{1}{6} x_1^2 x_2 \, dx_1 \, dx_2 \\
&= \int_0^3 \left(\frac{1}{18} x_1^3 x_2 \Big|_0^2 \right) dx_2 \\
&= \int_0^3 \frac{8}{18} x_2 \, dx_2 \\
&= \int_0^3 \frac{4}{9} x_2 \, dx_2 \\
&= \frac{4}{18} x_2^2 \Big|_0^3 \\
&= \frac{36}{18} \\
&= 2
\end{aligned}
$$

Then compute expected value of X_1 as follows

$$
\begin{aligned}
E\left[X_1\right] &= \int_0^3 \int_0^2 \frac{1}{6} x_1^2 \, dx_1 \, dx_2 \\
&= \int_0^3 \left(\frac{1}{12} x_1^2 \Big|_0^2 \right) \, dx_2 \\
&= \int_0^3 \frac{4}{12} \, dx_2 \\
&= \int_0^3 \frac{1}{3} dx_2 \\
&= \frac{1}{3} x_2 \Big|_0^3 \\
&= \frac{3}{3} \\
&= 1
\end{aligned}
$$

Then compute the expected value of X_2 as follows.

$$\begin{aligned}
E\left[X_2\right] &= \int_0^3 \int_0^2 \frac{1}{6} x_1 x_2 \, dx_1 \, dx_2 \\
&= \int_0^3 \left(\frac{1}{12} x_1^2 x_2 \Big|_0^2 \right) dx_2 \\
&= \int_0^3 \frac{4}{12} x_2 \, dx_2 \\
&= \int_0^3 \frac{1}{3} x_2 \, dx_2 \\
&= \frac{1}{6} x_2^2 \Big|_0^3 \\
&= \frac{9}{6} \\
&= \frac{3}{2}
\end{aligned}$$

The covariance is then given by

$$\begin{aligned}
\operatorname{cov}[X, Y] &\equiv E[XY] - E[X]E[Y] \\
&\equiv 2 - \left(\frac{4}{3}\right)\left(\frac{3}{2}\right) \\
&= 2 - 2 = 0
\end{aligned}$$

(iii) Let the joint density of two random variables x_1 and x_2 be given by

$$f(x_1 \, x_2) = \begin{cases} \frac{3}{8} x_1 & 0 \leq x_2 \leq x_1 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

First compute the expected value of $X_1 X_2$ as follows.

$$\begin{aligned}
E\left[X_1\, X_2\right] &= \int_0^2 \int_{x_2}^2 \frac{3}{8} x_1^2 x_2 \, dx_1 \, dx_2 \\
&= \int_0^2 \left(\frac{3}{24} x_1^3 x_2 \Big|_{x_2}^{2} \right) dx_2 \\
&= \int_0^2 \left(\frac{24}{24} x_2 - \frac{3}{24} x_2^4 \right) dx_2 \\
&= \int_0^2 \left(x_2 - \frac{1}{8} x_2^4 \right) dx_2 \\
&= \left(\frac{x_2^2}{2} - \frac{1}{40} x_2^5 \right) \Big|_0^2 \\
&= \frac{4}{2} - \frac{32}{40} \\
&= 2 - \frac{4}{5} \\
&= \frac{6}{5}
\end{aligned}$$

Then compute expected value of X_1 as follows

$$\begin{aligned}
E\left[X_1\right] &= \int_0^2 \int_0^{x_1} \frac{3}{8} x_1^2 \, dx_2 \, dx_1 \\
&= \int_0^2 \left(\frac{3}{8} x_1^2 x_2 \Big|_0^{x_1} \right) dx_1 \\
&= \int_0^2 \frac{3}{8} x_1^3 \, dx_1 \\
&= \frac{3}{32} x_1^4 \Big|_0^2 \\
&= \frac{48}{32} \\
&= \frac{3}{2}
\end{aligned}$$

Then compute the expected value of X_2 as follows.

$$\begin{aligned}
E[X_2] &= \int_0^2 \int_{x_2}^2 \frac{3}{8} x_1 x_2 \, dx_1 \, dx_2 \\
&= \int_0^2 \left(\frac{3}{16} x_1^2 x_2 \Big|_{x_2}^2 \right) dx_2 \\
&= \int_0^2 \left(\frac{12}{16} x_2 - \frac{3}{16} x_2^3 \right) dx_2 \\
&= \int_0^2 \left(\frac{3}{4} x_2 - \frac{3}{16} x_2^3 \right) dx_2 \\
&= \left(\frac{3}{8} x_2^2 - \frac{3}{64} x_2^4 \right) \Big|_0^2 \\
&= \frac{12}{8} - \frac{48}{64} \\
&= \frac{96}{64} - \frac{48}{64} = \frac{48}{64} \\
&= \frac{3}{4}
\end{aligned}$$

The covariance is then given by

$$\begin{aligned}
\text{cov}[X, Y] &\equiv E[XY] - E[X]E[Y] \\
&\equiv \frac{6}{5} - \left(\frac{3}{2} \right) \left(\frac{3}{4} \right) \\
&\equiv \frac{6}{5} - \frac{9}{8} \\
&\equiv \frac{48}{40} - \frac{45}{40} \\
&= \frac{3}{40}
\end{aligned}$$

10.3. **Correlation.** The correlation coefficient, denoted by $\rho[X, Y]$, or $\rho_{X,Y}$ of random variables X and Y is defined to be

$$\rho_{X,Y} = \frac{\text{cov}[X, Y]}{\sigma_X \sigma_Y} \tag{43}$$

provided that $\text{cov}[X, Y]$, σ_X and σ_Y exist, and σ_X, σ_Y are positive. The correlation coefficient between two random variables is a measure of the interaction between them. It also has the property of being independent of the units of measurement and being bounded between negative one and one. The sign of the correlation coefficient is the same as the sign of the covariance. Thus $\rho > 0$ indicates that X_2 increases as X_1 increases and $\rho = 1$ indicates perfect correlation, with all the points falling on a straight line with positive slope. If $\rho = 0$, there is no correlation and the covariance is zero.

10.4. **Independence and Covariance.**

10.4.1. *Theorem.*

Theorem 10. *If X and Y are independent random variables, then*

$$\text{cov}[X, Y] = 0. \tag{44}$$

Proof:
We know from equation 42 that

$$\text{cov}[X, Y] = E[XY] - E[X]E[Y] \tag{45}$$

We also know from equation 39 that if X and Y are independent, then

$$E\left[g(X)h(Y)\right] = E\left[g(X)\right]E\left[h(Y)\right] \tag{46}$$

Let $g(X) = X$ and $h(Y) = Y$ to obtain

$$E\left[XY\right] = E\left[X\right]E\left[Y\right] \tag{47}$$

Substituting into equation 45 we obtain

$$\text{cov}[X, Y] = E[X]E[Y] - E[X]E[Y] = 0 \tag{48}$$

The converse of Theorem 10 is not true, i.e., $\text{cov}[X, Y] = 0$ does not imply X and Y are independent.

10.4.2. *Example.* Consider the following discrete probability distribution.

		x_1			
		-1	0	1	
x_2	-1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{5}{16}$
	0	$\frac{3}{16}$	0	$\frac{3}{16}$	$\frac{6}{16} = \frac{3}{8}$
	1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{5}{16}$
		$\frac{5}{16}$	$\frac{6}{16} = \frac{3}{8}$	$\frac{5}{16}$	1

These random variables are not independent because the joint probabilities are not the product of the marginal probabilities. For example

$$p_{X_1X_2}[-1,\,-1] = \frac{1}{16} \neq p_{X_1}(-1)p_{X_1}(-1) = \left(\frac{5}{16}\right)\left(\frac{5}{16}\right) = \frac{25}{256}$$

Now compute the covariance between X_1 and X_2. First find $E[X_1]$ as follows

$$E[X_1] = (-1)\left(\frac{5}{16}\right) + (0)\left(\frac{6}{16}\right) + (1)\left(\frac{5}{16}\right) = 0$$

Similarly for the expected value of X_2.

$$E[X_2] = (-1)\left(\frac{5}{16}\right) + (0)\left(\frac{6}{16}\right) + (1)\left(\frac{5}{16}\right) = 0$$

Now compute $E[X_1X_2]$ as follows

$$\begin{aligned}
E[X_1X_2] = (-1)(-1)\left(\frac{1}{16}\right) + (-1)(0)\left(\frac{3}{16}\right) + (-1)(1)\left(\frac{1}{16}\right) \\
+ (0)(-1)\left(\frac{3}{16}\right) + (0)(0)(0) + (0)(1)\left(\frac{3}{16}\right) \\
+ (1)(-1)\left(\frac{1}{16}\right) + (1)(0)\left(\frac{3}{16}\right) + (1)(1)\left(\frac{1}{16}\right) \\
= \frac{1}{16} - \frac{1}{16} - \frac{1}{16} + \frac{1}{16} = 0
\end{aligned}$$

The covariance is then

$$\begin{aligned}
\text{cov}[X,\,Y] &\equiv E[XY] - E[X]E[Y] \\
&\equiv 0 - (0)\,(0) = 0
\end{aligned}$$

In this case the covariance is zero, but the variables are not independent.

10.5. Sum of Variances — $\text{var}[a_1x_1 + a_2x_2]$.

$$\begin{aligned}
\text{var}[a_1x_1 + a_2x_2] &= a_1^2\,\text{var}(x_1) + a_2^2\,\text{var}(x_2) + 2a_1a_2\,\text{cov}(x_1,\,x_2) \\
&= a_1^2\sigma_1^2 + 2a_1a_2\sigma_{12} + a_2^2\sigma_2^2 \\
&= [a_1\ a_2]\begin{bmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2\end{bmatrix}\begin{bmatrix}a_1 \\ a_2\end{bmatrix} \\
&= \text{var}\left[\,[a_1\ a_2]\begin{bmatrix}x_1 \\ x_2\end{bmatrix}\right]
\end{aligned} \tag{49}$$

10.6. **The Expected Value and Variance of Linear Functions of Random Variables.**

10.6.1. *Theorem.*

Theorem 11. *Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables with $E[Y_i] = \mu_i$ and $E[X_j] = \xi_j$. Define*

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j \tag{50}$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following three results hold:

(i) $E[U_1] = \sum_{i=1}^{n} a_i \mu_i$

(ii) $\mathrm{var}[U_1] = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}[Y_i] + 2\sum\sum_{i<j} a_i a_j \,\mathrm{cov}[Y_i,\, Y_j]$
where the double sum is over all pairs $(i,\, j)$ with $i < j$.

(iii) $\mathrm{cov}[U_1,\, U_2] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \,\mathrm{cov}[Y_i,\, X_j]$.

Proof:

(i) We want to show that

$$E[U_1] = \sum_{i=1}^{n} a_i \mu_i$$

Write out the $E[U_1]$ as follows:

$$\begin{aligned} E[U_1] &= E\left[\sum_{i=1}^{n} a_i Y_i\right] \\ &= \sum_{i=1}^{n} E\left[a_i Y_i\right] \\ &= \sum_{i=1}^{n} a_i E\left[Y_i\right] \\ &= \sum_{i=1}^{n} a_i \mu_i \end{aligned} \tag{51}$$

using Theorems 6–8 as appropriate.

(ii) Write out the $\text{var}[U_1]$ as follows:

$$\begin{aligned}
\text{var}(U_1) = E[U_1 - E(U_1)]^2 &= E\left[\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right]^2 \\
&= E\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]^2 \\
&= E\left[\sum_{i=1}^{n} a_i^2 (Y_i - \mu_i)^2 + \sum\sum_{i \neq j} a_i a_j (Y_i - \mu i)(Y_j - \mu_j)\right] \\
&= \sum_{i=1}^{n} a_i^2 E(Y_i - \mu_i)^2 + \sum\sum_{i \neq j} a_i a_j E[(Y_i - \mu_i)(Y_j - \mu_j)]
\end{aligned} \tag{52}$$

By definitions of variance and covariance, we have

$$\text{var}(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + \sum\sum_{i \neq j} a_i a_j \,\text{cov}(Y_i,\, Y_j) \tag{53}$$

Because

$$\text{cov}(Y_i,\, Y_j) = \text{cov}(Y_j,\, Y_i)$$

we can write

$$\text{var}(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2\sum\sum_{i < j} a_i a_j \,\text{cov}(Y_i,\, Y_j) \tag{54}$$

Similar steps can be used to obtain (iii).

(iii) We have

$$
\begin{aligned}
\operatorname{cov}(U_1,\, U_2) &= E\left\{[U_1 - E(U_1)]\,[U_2 - E(U_2)]\right\} \\
&= E\left[\left(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{m} b_j X_j - \sum_{j=1}^{m} b_j \xi_j\right)\right] \\
&= E\left\{\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]\left[\sum_{j=1}^{m} b_j (x_j - \xi_j)\right]\right\} \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_i (Y_i - \mu_i)(X_j - \xi_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_i E\left[(Y_i - \mu_i)(X_j - \xi_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_i \operatorname{cov}(Y_i,\, X_j)
\end{aligned}
\tag{55}
$$

11. Conditional Expectations

11.1. **Definition.** If X_1 and X_2 are any two random variables, the conditional expectation of $g(X_1)$, given that $X_2 = x_2$, is defined to be

$$
E\left[g(X_1) \mid X_2\right] = \int_{-\infty}^{\infty} g(x_1) f(x_1 \mid x_2)\, dx_1 \tag{56}
$$

if X_1 and X_2 are jointly continuous and

$$
E\left[g(X_1) \mid X_2\right] = \sum_{x_1} g(x_1) p(x_1 \mid x_2) \tag{57}
$$

if X_1 and X_2 are jointly discrete.

11.2. **Example.** Let the joint density of two random variables X and Y be given by

$$
f(x,\, y) = \begin{cases} 2 & x \geq 0, \quad y \geq 0, \quad x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}
$$

We can find the marginal density of y by integrating the joint density with respect to x as follows

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,\, y)\, dx \\ &= \int_0^{1-y} 2\, dx \\ &= 2x \,\Big|_0^{1-y} \\ &= 2(1-y), \quad 0 \leq y \leq 1 \end{aligned}$$

We find the conditional density of X given that $Y = y$ by forming the ratio

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f(x,\, y)}{f_Y(y)} \\ &= \frac{2}{2(1-y)} \\ &= \frac{1}{(1-y)}, \quad 0 \leq x \leq 1-y \end{aligned}$$

We then form the expected value by multiplying the density by x and then integrating over x.

$$\begin{aligned} E\left[X \mid Y\right] &= \int_0^{1-y} x \frac{1}{(1-y)}\, dx \\ &= \frac{1}{(1-y)} \int_0^{1-y} x\, dx \\ &= \frac{1}{(1-y)} \left(\frac{x^2}{2} \Big|_0^{1-y} \right) \\ &= \frac{1}{(1-y)} \left(\frac{(1-y)^2}{2} \right) \\ &= \frac{(1-y)}{2} \end{aligned}$$

We can find the unconditional expected value of X by multiplying the marginal density of y by this expected value and integrating over y as follows

$$
\begin{aligned}
E[X] &= E_Y\big[E[X \mid Y]\big] \\
&= \int_0^1 \frac{1-y}{2}\big(2(1-y)\big)\,dy \\
&= \int_0^1 (1-y)^2\,dy \\
&= \frac{-(1-y)^3}{3}\,\Bigg|_0^1 \\
&= -\frac{1}{3}\big[(1-1)^3 - (1-0)^3\big] \\
&= -\frac{1}{3}[0-1] \\
&= \frac{1}{3}
\end{aligned}
$$

We can show this directly by multiplying the joint density by x and then integrating over x and y.

$$
\begin{aligned}
E[X] &= \int_0^1 \int_0^{1-y} 2x\,dx\,dy \\
&= \int_0^1 \left(x^2\,\big|_0^{1-y}\right)\,dy \\
&= \int_0^1 (1-y)^2\,dy \\
&= \int_0^1 \frac{-(1-y)^3}{3}\,dy \\
&= \frac{-(1-y)^3}{3}\,\big|_0^1 \\
&= -\frac{1}{3}\,\big[(1-1)^3 - (1-0)^3\big] \\
&= -\frac{1}{3}\,[0-1] \\
&= \frac{1}{3}
\end{aligned}
$$

The fact that we can find the expected value of X using the conditional distribution of X given Y is due to the following theorem.

11.3. **Theorem.**

Theorem 12. *Let X and Y denote random variables. Then*

$$E[X] = E_Y\left[E_{X|Y}[X \mid Y]\right] \tag{58}$$

The inner expectation is with respect to the conditional distribution of X given Y and the outer expectation is with respect to the distribution of Y.

Proof: Suppose that X and Y are jointly continuous with joint density $F(X, Y)$ and marginal distributions $f_X(x)$ and $f_Y(y)$, respectively. Then

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,\, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx\right] f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} E\left[X \mid Y = y\right] f_Y(y)\, dy \\
&= E_Y\left[E_{X|Y}[X \mid Y]\right]
\end{aligned} \tag{59}$$

The proof is similar for the discrete case.

11.4. **Conditional Variance.**

11.4.1. *Definition.* Just as we can compute a conditional expected value, we can compute a conditional variance. The idea is that the variance of the random variable X may be different for different values of Y. We define the conditional variance as follows.

$$\begin{aligned}
\text{var}[X \mid Y = y] &= E\Big[(X - E[X \mid Y = y])^2 \mid Y = y\Big] \\
&= E\left[X^2 \mid Y = y\right] - \big[E[X \mid Y = y]\big]^2
\end{aligned} \tag{60}$$

We can write the variance of X as a function of the expected value of the conditional variance. This is sometimes useful for specific problems.

11.4.2. *Theorem.*

Theorem 13. *Let X and Y denote random variables. Then*

$$\text{var}[X] = E\Big[\text{var}[X \mid Y = y]\Big] + \text{var}\Big[E[X \mid Y = y]\Big] \tag{61}$$

Proof: First note the following three definitions

$$\text{var}[X \mid Y] = E[X^2 \mid Y] - \big[E[X \mid Y]\big]^2 \tag{62a}$$

$$E\big[\text{var}[X \mid Y]\big] = E\big[E[X^2 \mid Y]\big] - E\big\{\big[E[X \mid Y]\big]^2\big\} \tag{62b}$$

$$\text{var}\big[E[X \mid Y]\big] = E\big\{\big[E[X \mid Y]\big]^2\big\} - \big\{E\big[E[X \mid Y]\big]\big\}^2 \tag{62c}$$

The variance of X is given by

$$\text{var}[X] = E[X^2] - \big[E[X]\big]^2 \tag{63}$$

We can find the expected value of a variable by taking the expected value of the conditional expectation as in Theorem 12. For this problem we can write $E[X^2]$ as the expected value of the conditional expectation of X^2 given Y. Specifically,

$$E[X^2] = E_Y\left\{E_{X|Y}[X^2 \mid Y]\right\} \tag{64}$$

and

$$\big[E[X]\big]^2 = \big[E_Y\big\{E_{X|Y}[X \mid Y]\big\}\big]^2 \tag{65}$$

Write (63) substituting in (64) and (65) as follows

$$\begin{aligned} \text{var}[X] &= E[X^2] - \big[E[X]\big]^2 \\ &= E_Y\left\{E_{X|Y}[X^2 \mid Y]\right\} - \big[E_Y\big\{E_{X|Y}[X \mid Y]\big\}\big]^2 \end{aligned} \tag{66}$$

Now subtract and add $E\big\{\big[E(X \mid Y)\big]^2\big\}$ to the right hand side of equation 66 as follows

$$\begin{aligned} \text{var}[X] &= E_Y\big\{E_{X|Y}[X^2 \mid Y]\big\} - \big[E_Y\big\{E_{X|Y}[X \mid Y]\big\}\big]^2 \\ &= E_Y\big\{E_{X|Y}[X^2 \mid Y]\big\} - E\big\{\big[E(X \mid Y)\big]^2\big\} \\ &\qquad + E\big\{\big[E(X \mid Y)\big]^2\big\} - \big[E_Y\big\{E_{X|Y}[X \mid Y]\big\}\big]^2 \end{aligned} \tag{67}$$

Now notice that the first two terms in equation 67 are the same as the right hand side of equation 62b which is $\big(E\big[\text{var}[X \mid Y]\big]\big)$. Then notice that the second two terms in equation 67 are the same as the right hand side of equation 62c which is $\big(\text{var}\big[E[X \mid Y]\big]\big)$.
We can then write $\text{var}[X]$ as

$$\begin{aligned} \text{var}[X] &= E_Y\big\{E_{X|Y}[X^2 \mid Y]\big\} - \big[E_Y\big\{E_{X|Y}[X \mid Y]\big\}\big]^2 \\ &= E\big[\text{var}[X \mid Y]\big] + \text{var}\big[E[X \mid Y]\big] \end{aligned} \tag{68}$$

11.4.3. *Example.* Let the joint density of two random variables X and Y be given by

$$f(x,\,y) = \begin{cases} \frac{1}{4}(2x+y) & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

We can find the marginal density of x by integrating the joint density with respect to y as follows

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,\,y)\,dy \\ &= \int_0^2 \frac{1}{4}(2x+y)\,dy \\ &= \frac{1}{4}\left(2xy + \frac{y^2}{2}\right)\Big|_0^2 \\ &= \frac{1}{4}\left(4x + \frac{4}{2}\right) \\ &= \frac{1}{4}\left(4x+2\right), \quad 0 \le x \le 1 \end{aligned} \tag{69}$$

We can find the marginal density of y by integrating the joint density with respect to x as follows:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,\,y)\,dy \\ &= \int_0^1 \frac{1}{4}(2x+y)\,dx \\ &= \frac{1}{4}\left(x^2 + xy\right)\Big|_0^1 \\ &= \frac{1}{4}\left(1+y\right), \quad 0 \le y \le 2 \end{aligned} \tag{70}$$

We find the expected value of X by multiplying the conditional density by x and then integrating over x

$$\begin{aligned}
E[X] &= \int_0^1 \frac{1}{4} x\,(4x+2)\ dx \\
&= \int_0^1 \frac{1}{4}\left(4x^2+2x\right)\ dx \\
&= \frac{1}{4}\left(\frac{4}{3}x^3+x^2\right)\Big|_0^1 \\
&= \frac{1}{4}\left(\frac{4}{3}+1\right) = \frac{1}{4}\left(\frac{7}{3}\right) \\
&= \frac{7}{12}
\end{aligned} \tag{71}$$

To find the variance of X, we first need to find the $E[X^2]$. We do this as follows

$$\begin{aligned}
E[X^2] &= \int_0^1 \frac{1}{4} x^2\,(4x+2)\ dx \\
&= \int_0^1 \frac{1}{4}\left(4x^3+2x^2\right)\ dx \\
&= \frac{1}{4}\left(x^4+\frac{2}{3}x^3\right)\Big|_0^1 \\
&= \frac{1}{4}\left(1+\frac{2}{3}\right) = \frac{1}{4}\left(\frac{5}{3}\right) \\
&= \frac{5}{12}
\end{aligned} \tag{72}$$

The variance of X is then given by

$$\begin{aligned}
\operatorname{var}(X) &\equiv E\left[(X-E(X))^2\right] \\
&\equiv E(x^2)-E^2(x) \\
&= \frac{5}{12}-\left(\frac{7}{12}\right)^2 \\
&= \frac{5}{12}-\frac{49}{144} \\
&= \frac{60}{144}-\frac{49}{144} \\
&= \frac{11}{144}
\end{aligned} \tag{73}$$

We find the conditional density of X given that $Y = y$ by forming the ratio

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f(x,\, y)}{f_Y(y)} \\ &= \frac{\frac{1}{4}(2x+y)}{\frac{1}{4}(1+y)} \\ &= \frac{(2x+y)}{(1+y)} \end{aligned} \tag{74}$$

We then form the expected value of X given Y by multiplying the density by x and then integrating over x.

$$\begin{aligned} E\left[X \mid Y\right] &= \int_0^1 x \frac{(2x+y)}{(1+y)}\, dx \\ &= \frac{1}{1+y} \int_0^1 (2x^2 + xy)\, dx \\ &= \frac{1}{1+y} \left(\frac{2}{3}x^3 + \frac{1}{2}x^2 y \right) \Big|_0^1 \\ &= \frac{1}{(1+y)} \left(\frac{2}{3} + \frac{1}{2} y \right) \\ &= \frac{\left(\frac{2}{3} + \frac{y}{2}\right)}{(1+y)} \\ &= \frac{(4+3y)}{(6+6y)} \\ &= \left(\frac{1}{6}\right) \frac{(4+3y)}{(1+y)} \end{aligned} \tag{75}$$

We can find the unconditional expected value of X by multiplying the marginal density of y by this expected value and integrating over y as follows:

$$
\begin{aligned}
E[X] &= E_Y\left[E[X \mid Y]\right] \\
&= \int_0^2 \frac{(4+3y)}{(6+6y)} \frac{1}{4}(1+y)\, dy \\
&= \frac{1}{4} \int_0^2 \frac{(4+3y)(1+y)}{6(1+y)}\, dy \\
&= \frac{1}{24} \int_0^2 (4+3y)\, dy \\
&= \frac{1}{24} \left(4y + \frac{3}{2}y^2\right) \Big|_0^2 \\
&= \frac{1}{24}(8+6) \\
&= \frac{14}{24} = \frac{7}{12}
\end{aligned} \tag{76}
$$

We find the conditional variance by finding the expected value of X^2 given Y and then subtracting the square of $E[X \mid Y]$.

$$
\begin{aligned}
E[X^2 \mid Y] &= \int_0^1 x^2 \frac{(2x+y)}{(1+y)}\, dx \\
&= \frac{1}{1+y} \int_0^1 (2x^3 + x^2 y)\, dx \\
&= \frac{1}{1+y} \left(\frac{1}{2}x^4 + \frac{1}{3}x^3 y\right) \Big|_0^1 \\
&= \frac{1}{1+y} \left(\frac{1}{2} + \frac{1}{3}y\right) \\
&= \frac{\left(\frac{1}{2} + \frac{y}{3}\right)}{1+y} \\
&= \left(\frac{1}{6}\right) \left(\frac{3+2y}{1+y}\right)
\end{aligned} \tag{77}
$$

Now square $E[X \mid Y]$.

$$
\begin{aligned}
E^2\left[X \mid Y\right] &= \left(\left(\frac{1}{6}\right) \frac{(4+3y)}{(1+y)}\right)^2 \\
&= \frac{1}{36} \frac{(4+3y)^2}{(1+y)^2}
\end{aligned} \tag{78}
$$

Now subtract equation 78 from equation 77

$$
\begin{aligned}
\operatorname{var}[X \mid Y] &= \left(\frac{1}{6}\right)\left(\frac{3+2y}{1+y}\right) - \frac{1}{36}\frac{(4+3y)^2}{(1+y)^2} \\
&= \left(\frac{1}{36}\right)\left(\frac{(18+12y)(1+y) - (4+3y)^2}{(1+y)^2}\right) \\
&= \left(\frac{12y^2+30y+18-(16+24y+9y^2)}{36(1+y)^2}\right) \\
&= \frac{3y^2+6y+2}{36(1+y)^2}
\end{aligned} \tag{79}
$$

For example, if $y = 1$, we obtain

$$
\begin{aligned}
\operatorname{var}[X \mid Y = 1] &= \left.\frac{3y^2+6y+2}{36(1+y)^2}\right|_{y=1} \\
&= \frac{11}{144}
\end{aligned} \tag{80}
$$

To find the expected value of this variance we need to multiply the expression in equation 80 by the marginal density of Y and then integrate over the range of Y.

$$
\begin{aligned}
E\big[\operatorname{var}[X \mid Y]\big] &= \int_0^2 \frac{3y^2+6y+2}{36(1+y)^2}\frac{1}{4}(1+y)\,dy \\
&= \frac{1}{144}\int_0^2 \frac{3y^2+6y+2}{(1+y)}\,dy
\end{aligned} \tag{81}
$$

Consider first the indefinite integral.

$$
z = \int \frac{3y^2+6y+2}{(1+y)}\,dy \tag{82}
$$

This integral would be easier to solve if $(1+y)$ in the denominator could be eliminated. This would be the case if it could be factored out of the numerator. One way to do this is carry out the specified division.

$$\begin{array}{r|rrrrr} & 3y & + & 3 & & \\ \hline 1 + y & 3y^2 & + & 6y & + & 2 \\ & 3y^2 & + & 3y & & \\ \hline & & & 3y & + & 2 \\ & & & 3y & + & 3 \\ \hline & & & & & -1 \end{array} \tag{83}$$

$$\Rightarrow \frac{3y^2 + 6y + 2}{1 + y} = (3y + 3) - \frac{1}{1 + y}$$

Now substitute equation 83 into equation 82 as follows

$$\begin{aligned} z &= \int \frac{3y^2 + 6y + 2}{(1 + y)}\, dy \\ &= \int \left[(3y + 3) - \frac{1}{1 + y} \right] dy \\ &= \frac{3y^2}{2} + 3y - \log[1 + y] \end{aligned} \tag{84}$$

Now compute the **expected value of the variance** as

$$\begin{aligned} E\big[\text{var}[X \mid Y]\big] &= \frac{1}{144} \int_0^2 \frac{3y^2 + 6y + 2}{(1 + y)}\, dy \\ &= \frac{1}{144} \left[\frac{3y^2}{2} + 3y - \log[1 + y] \,\Big|_0^2 \right] \\ &= \frac{1}{144} \left[\left(\frac{12}{2} + 6 - \log[3] \right) - \log[1] \right] \\ &= \frac{1}{144} \big[12 - \log[3] \big] \end{aligned} \tag{85}$$

To compute the variance of $E[X \mid Y]$ we need to find $E_Y\left[\left(E[X \mid Y]\right)^2\right]$ and then subtract $\left(E_Y\left[E[X \mid Y]\right]\right)^2$.

First find the second term. The expected value of X given Y comes from equation 75.

$$E[X \mid Y] = \left(\frac{1}{6}\right) \frac{(4 + 3y)}{(1 + y)} \tag{86}$$

We found the expected value of $E[X \mid Y]$ in equation 76. We repeat the derivation here by multiplying $E[X \mid Y]$ by the marginal density of Y and then integrating over the range of Y.

$$
\begin{aligned}
E_Y\big(E[X \mid Y]\big) &= \int_0^2 \left(\frac{1}{6}\right) \frac{(4+3y)}{(1+y)} \frac{1}{4}(y+1)\,dy \\
&= \frac{1}{24}\int_0^2 (4+3y)\,dy \\
&= \frac{1}{24}\left(4y + \frac{3}{2}y^2\right)\Big|_0^2 \\
&= \frac{1}{24}\left(8 + \frac{12}{2}\right) \\
&= \frac{1}{24}(14) \\
&= \frac{7}{12}
\end{aligned} \tag{87}
$$

Now find the first term

$$
\begin{aligned}
E_Y\Big(\big(E[X \mid Y]\big)^2\Big) &= \int_0^2 \left(\frac{1}{36}\right) \frac{(4+3y)^2}{(1+y)^2} \frac{1}{4}(y+1)\,dy \\
&= \frac{1}{144}\int_0^2 \frac{(4+3y)^2}{1+y}\,dy \\
&= \frac{1}{144}\int_0^2 \frac{9y^2+24y+16}{1+y}\,dy
\end{aligned} \tag{88}
$$

Now find the indefinite integral by first simplifying the integrand using long division.

$$
\frac{9y^2+24y+16}{1+y} = 1+y\,\overline{\big)\,9y^2 \quad + \quad 24y \quad + \quad 16} \tag{89}
$$

Now carry out the division

$$
\begin{array}{rrrrrrr}
 & & & 9y & + & 15 & \\
\cline{4-7}
1 & + & y\big) & 9y^2 & + & 24y & + \; 16 \\
 & & & 9y^2 & + & 9y & \\
\cline{4-7}
 & & & & & 15y & + \; 16 \\
 & & & & & 15y & + \; 15 \\
\cline{6-7}
 & & & & & & 1
\end{array} \tag{90}
$$

$$
\Rightarrow \frac{9y^2+24y+16}{1+y} = (9y+15) + \frac{1}{1+y}
$$

Now substitute in equation 90 into equation 88 as follows

$$
\begin{aligned}
E_Y\Big(\big(E[X \mid Y]\big)^2\Big) &= \frac{1}{144}\int_0^2 \frac{9y^2+24y+16}{1+y}\,dy \\
&= \frac{1}{144}\int_0^2 9y+15+\frac{1}{1+y}\,dy \\
&= \frac{1}{144}\Big[\frac{9y^2}{2}+15y+\log[y+1]\Big]\Big|_0^2 \\
&= \frac{1}{144}\Big[\frac{36}{2}+30+\log[3]\Big] \\
&= \frac{1}{144}\big[48+\log[3]\big]
\end{aligned} \tag{91}
$$

The variance is obtained by subtracting the square of (87) from (91)

$$
\begin{aligned}
\operatorname{var}\big[E[X \mid Y]\big] &= E_Y\Big(\big(E[X \mid Y]\big)^2\Big) + \Big(E_Y\big(E[X \mid Y]\big)\Big)^2 \\
&= \frac{1}{144}\big[48+\log[3]\big] - \left(\frac{7}{12}\right)^2 \\
&= \frac{1}{144}\big[48+\log[3]\big] - \frac{49}{144} \\
&= \frac{1}{144}\big[\log[3]-1\big]
\end{aligned} \tag{92}
$$

We can show that the sum of (85) and (92) is equal to the $\operatorname{var}[X_1]$ as in Theorem 13:

$$
\begin{aligned}
\operatorname{var}[X] &= E\big[\operatorname{var}[X \mid Y=y]\big] + \operatorname{var}\big[E[X \mid Y=y]\big] \\
&= \frac{1}{144}\big[\log[3]-1\big] + \frac{1}{144}\big[12-\log[3]\big] \\
&= \frac{\log[3]-1+12-\log[3]}{144} \\
&= \frac{11}{144}
\end{aligned} \tag{93}
$$

which is the same as in equation 73.

12. CAUCHY-SCHWARZ INEQUALITY

12.1. **Statement of Inequality.** For any functions $g(x)$ and $h(x)$ and cumulative distribution function $F(x)$,the following holds:

$$\int g(x)h(x)\,dF(x) \leq \left(\int g(x)^2\,dF(x)\right)^{\frac{1}{2}} \left(\int h(x)^2\,dF(x)\right)^{\frac{1}{2}} \tag{94}$$

where x is a vector random variable.

12.2. **Proof.** Form a linear combination of $g(x)$ and $h(x)$, square it and then integrate as follows:

$$\int \big(tg(x) + h(x)\big)^2\,dF(x) \geq 0 \tag{95}$$

The inequality holds because of the square and $dF(x) > 0$. Now expand the integrand in (95) to obtain

$$t^2 \int \big(g(x)\big)^2\,dF(x) + 2t \int g(x)h(x)\,dF(x) + \int \big(h(x)\big)^2\,dF(x) \geq 0 \tag{96}$$

This is a quadratic equation in t which holds for all t. Now define t as follows:

$$t = \frac{-\int g(x)h(x)\,dF(x)}{\int \big(g(x)\big)^2\,dF(x)} \tag{97}$$

and substitute in (96)

$$\begin{aligned}
&\frac{\left(\int g(x)h(x)\,dFx)\right)^2}{\int \big(g(x)\big)^2\,dF(x)} - 2\frac{\left(\int g(x)h(x)\,dF(x)\right)^2}{\int \big(g(x)\big)^2\,dF(x)} + \int \big(h(x)\big)^2\,dF(x) \geq 0 \\
&\Rightarrow -\frac{\left(\int g(x)h(x)\,dF(x)\right)^2}{\int \big(g(x)\big)^2\,dF(x)} \geq -\int \big(h(x)\big)^2\,dF(x) \\
&\Rightarrow \left(\int g(x)h(x)\,dF(x)\right)^2 \leq \int \big(h(x)\big)^2\,dF(x) \int \big(g(x)\big)^2\,dF(x) \\
&\Rightarrow \left|\int g(x)h(x)\,dF(x)\right| \leq \left(\int \big(h(x)\big)^2\,dF(x)\right)^{\frac{1}{2}} \left(\int \big(gx)\big)^2\,dF(x)\right)^2
\end{aligned} \tag{98}$$

12.3. **Corollary 1.** Consider two random variables X_1 and X_2 and the expectation of their product. Using (98) we obtain

$$\begin{aligned}
&\big(E(X_1X_2)\big)^2 \leq E(X_1^2)E(X_2^2) \\
&\left|\,E\,(X_1X_2)\,\right| \leq \big(E(X_1^2)\big)^{\frac{1}{2}} \big(E(X_2^2)\big)^{\frac{1}{2}}
\end{aligned} \tag{99}$$

12.4. **Corollary 2.**

$$|\operatorname{cov}(X_1 X_2)| < \big(\operatorname{var}(X_1)\big)^{\frac{1}{2}} \big(\operatorname{var}(X_2)\big)^{\frac{1}{2}} \tag{100}$$

Proof: Apply (98) to the centered random variables $g(X) = X_1 - \mu_1$ and $h(X) = X_2 - \mu_2$ where $\mu_i = E(X_i)$.

REFERENCES

[1] Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.
[2] Bickel, P. J. and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Vol 1).* 2nd edition. Upper Saddle River, NJ: Prentice Hall, 2001.
[3] Billingsley, P. *Probability and Measure.* 3rd edition. New York: Wiley, 1995.
[4] Casella, G. And R. L. Berger. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2002.
[5] Cramer, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.
[6] Goldberger, A. S. *Econometric Theory.* New York: Wiley, 1964.
[7] Lindgren, B. W. *Statistical Theory* 3rd edition. New York: Macmillan Publishing Company, 1976.
[8] Rao, C. R. *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley, 1973.