4. Random vectors and their distribution

Sometimes a single random variable is not enough to describe the outcomes of a random experiments. For example, to record the height and weight of every person in a certain community, we need a pair (x, y), where the components respectively represents he height and weight of a particular individuals. In many cases it is necessary to consider the joint behavior of two or more random variables.

Definition 4.1 (n-dimensional random vector). Let $X_1, X_2, \ldots, X_n$ be n real random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The function $\mathbf{X} : \Omega \to \mathbb{R}^n$ defined by

$$\mathbf{X}(\omega) := (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

is called an n-dimensional random vector.

Let $\mathbf{X}$ be a n-dimensional random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the function $P_{\mathbf{X}}$ on $\mathcal{B}(\mathbb{R}^n)$ defined by

$$P_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B), \quad B \in \mathcal{B}(\mathbb{R}^n)$$

is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. This is called distribution of $\mathbf{X}$.

Definition 4.2 (Joint cumulative distribution function (joint cdf)). Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be an n-dimensional random vector. The function $F_{(X_1,X_2,\ldots,X_n)} : \mathbb{R}^n \to [0, 1]$ defined by

$$F_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

is called the joint cumulative distribution funmction (joint cdf) of the random variables $X_1, X_2, \ldots, X_n$.

Marginal cumulative distribution function (marginal cdf): In the following, we consider $n = 2$, and the same results will hold for $n > 2$. Let X and Y be two random variables with joint cdf $F_{(X,Y)}$. One can find the cdf of X and Y from the joint cdf $F_{(X,Y)}$. Indeed

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\big(\cup_y \{X \le x, Y \le y\}\big) = \lim_{y\to\infty} \mathbb{P}(X \le x, Y \le y) = \lim_{y\to\infty} F_{(X,Y)}(x, y)$$

Similarly, we also have

$$F_Y(y) = \lim_{x\to\infty} F_{(X,Y)}(x, y).$$

The distribution functions F_X and F_Y are sometimes referred to as **marginal cdf** of X and Y. One can easily show that joint cdf is nondecreasing and right continuous on each of its arguments. Moreover, for any $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ with $x_1 \le x_2$ and $y_1 \le y_2$, set

$$A := \{x \le x_2, y \le y_2\}, \quad B = \{x \le x_1, y \le y_2\}, \quad C = \{x \le x_2, y \le y_1\}, \quad D = \{x \le x_1, y \le y_1\}.$$

Observe that

$$\begin{aligned} K_1 &:= \{x_1 < x \le x_2, y \le y_1\} = C \setminus D \implies P_{\mathbf{X}}(K_1) = P_{\mathbf{X}}(C) - P_{\mathbf{X}}(D) \\ K_2 &:= \{x_1 < x \le x_2, y \le y_2\} = A \setminus B \implies P_{\mathbf{X}}(K_2) = P_{\mathbf{X}}(A) - P_{\mathbf{X}}(B). \end{aligned}$$

Since $K_2 \setminus K_1 = \{x_1 < x \le x_2, y_1 < y \le y_2\}$, we have

$$\begin{aligned} 0 &\le P_{\mathbf{X}}(\{x_1 < x \le x_2, y_1 < y \le y_2\}) = P_{\mathbf{X}}(K_2) - P_{\mathbf{X}}(K_1) \\ &= P_{\mathbf{X}}(A) - P_{\mathbf{X}}(B) - \big(P_{\mathbf{X}}(C) - P_{\mathbf{X}}(D)\big) \\ &= F_{(X,Y)}(x_2, y_2) + F_{(X,Y)}(x_1, y_1) - F_{(X,Y)}(x_1, y_2) - F_{(X,Y)}(x_2, y_1)\,. \end{aligned}$$

Theorem 4.1. *A function $F : \mathbb{R}^2 \to [0, 1]$ is a joint cdf of some two dimensional random vector if and only if it satisfies the following conditions:*

a) F is nondecreasing and right continuous with respect to each arguments.
b) $\lim_{y\to-\infty} F(x, y) = 0 = \lim_{x\to-\infty} F(x, y)$ and $\lim_{(x,y)\to(\infty,\infty)} F(x, y) = 1$.

c) For any $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ *with* $x_1 \le x_2$ *and* $y_1 \le y_2$,

$$F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \ge 0.$$

Example 4.1. *The function* $F : \mathbb{R}^2 \to [0, 1]$ *given by*

$$F(x, y) = \begin{cases} 0, & x < 0, \ \text{or } y < 0, \ \text{or } x + y < 1, \\ 1, & \text{otherwise} \end{cases}$$

is NOT a joint cdf of any two dimensional random vector. If so, then

$$0 \le \mathbb{P}(\frac{1}{3} < X \le 1, \frac{1}{3} < Y \le 1) = F(1,1) + F(\frac{1}{3}, \frac{1}{3}) - F(1, \frac{1}{3}) - F(\frac{1}{3}, 1) = 1 + 0 - 1 - 1 = -1 < 0.$$

Definition 4.3 (**Discrere random vector**)**.** A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is said to be discrete if the random variables $X_1, X_2, \ldots, X_n$ are all discrete i.e., there exists a countable set $E \subseteq \mathbb{R}^n$ such that $\mathbb{P}(\mathbf{X} \in E) = 1$.

Definition 4.4 (**Joint probability mass function**)**.** Let $\mathbf{X}$ be a discrete random vector. The function $p_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ defined by

$$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \mathbb{P}(\mathbf{X} = \mathbf{x}), & \text{if } \mathbf{x} \text{ belongs to the image of } \mathbf{X} \\ 0, & \text{otherwise} \end{cases}$$

is called joint probability mass function (joint mpf) of $\mathbf{X}$.

Marginal pmf: Let X and Y be two discrete random variable with joint pmf $p_{(X,Y)}$. Then we can compute pmf of X and Y in terms of $p_{(X,Y)}$ as follows:

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}\big(\cup_y \{X = x, Y = y\}\big) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y p_{(X,Y)}(x, y)$$

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}\big(\cup_x \{X = x, Y = y\}\big) = \sum_x \mathbb{P}(X = x, Y = y) = \sum_x p_{(X,Y)}(x, y).$$

p_X and p_Y sometimes are referred as **marginal pmf** of X and Y.

Example 4.2. *A fair coin is tossed three times. Let* X *be the number of heads in three tossing, and let* Y *denotes the difference between number of heads and number of tails in absolute value. Then* $X \in \{0, 1, 2, 3\}$ *and* $Y \in \{1, 3\}$*. In this case,* $\Omega = \{H, T\}^3$*. We define* $\mathbb{P}(A) = \frac{|A|}{8}$*. Thus, for example*

$$\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(\{HTT, THT, TTH\}) = \frac{3}{8}, \quad \mathbb{P}(X = 2, Y = 1) = \mathbb{P}(\{HHT, HTH, THH\}) = \frac{3}{8}.$$

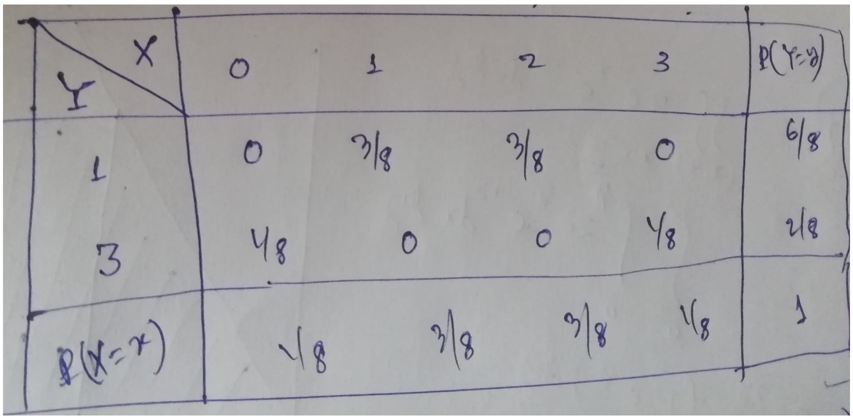
The joint pmf and the marginal pmf are given in the following table:

Like in one variable case, joint cdf can be determined in terms of joint pmf. Indeed, since image of (X, Y) is the countable set $E = \{(x_i, y_j) : i = 0, 1, \ldots, \ j = 0, 1, 2, \ldots\}$, we see that, for any $(x, y) \in \mathbb{R}^2$

$$F_{(X,Y)}(x, y) = \mathbb{P}(X \le x, Y \le y) = \sum_{x_i \le x, y_i \le y} \mathbb{P}(X = x_i, Y = y_i) = \sum_{x_i \le x, y_i \le y} p_{(X,Y)}(x_i, y_j).$$

Example 4.3. *A fair die is rolled and a fair coin is tossed independently. Let* X *be the face value of the die and let*

$$Y = \begin{cases} 0, & \text{if tail turns up} \\ 1, & \text{if head turns up} \end{cases}$$



Y \ X	0	1	2	3	P(Y=y)
1	0	3/8	3/8	0	6/8
3	1/8	0	0	1/8	2/8
P(X=x)	1/8	3/8	3/8	1/8	1

FIGURE 1. Joint pmf and Marginal pmf

where the joint pmf of X and Y are given by

$$p_{(X,Y)}(x,y) = \begin{cases} \frac{1}{12}, & \textit{if } (x,y) \textit{ is image of } (X,Y) \\ 0, & \textit{otherwise.} \end{cases}$$

Find the joint cdf of X and Y.

Solution: *Observe that $X \in \{1,2,3,4,5,6\}$ and $Y \in \{0,1\}$. By using the relation $F_{(X,Y)}(x,y) = \sum_{x_i \le x, y_i \le y} p_{(X,Y)}(x_i, y_j)$, we have*

$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x < 1, -\infty < y < \infty; \quad -\infty < x < \infty, y < 0 \\ \frac{1}{12}, & 1 \le x < 2, 0 \le y < 1 \\ \frac{1}{6}, & 2 \le x < 3, 0 \le y < 1; \quad 1 \le x < 2, y \ge 1 \\ \frac{1}{4}, & 3 \le x < 4, 0 \le y < 1 \\ \frac{1}{3}, & 4 \le x < 5, 0 \le y < 1; \quad 2 \le x < 3, y \ge 1 \\ \frac{5}{12}, & 5 \le x < 6, 0 \le y < 1 \\ \frac{1}{2}, & 6 \le x, 0 \le y < 1; \quad 3 \le x < 4, y \ge 1 \\ \frac{2}{3}, & 4 \le x < 5, y \ge 1 \\ 1, & x \ge 6, y \ge 1\,. \end{cases}$$

Definition 4.5. We say that X and Y are jointly continuous if there exists a nonnegative function $f_{(X,Y)}(\cdot,\cdot)$ defined for all real x and y, having the property that, for every Borel set $C \in \mathcal{B}(\mathbb{R}^2)$ such that

$$\mathbb{P}((X,Y) \in C) = \iint_{(x,y)\in C} f_{(X,Y)}(x,y)\,dx\,dy\,.$$

The function $f_{(X,Y)}(\cdot,\cdot)$ is called the joint probability density function (joint pdf) of X and Y.

Take $C = \{(x,y) : x \in A, y \in B\}$ where $A, B \in \mathcal{B}(\mathbb{R})$. Then we have

$$\mathbb{P}(X \in A, Y \in B) = \int_B \int_A f_{(X,Y)}(x,y)\,dx\,dy.$$

Thus, we have

$$F_{(X,Y)}(a,b) = \mathbb{P}(X \in (-\infty,a], Y \in (-\infty,b]) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{(X,Y)}(x,y)\,dx\,dy$$

$$\implies f_{(X,Y)}(a,b) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(a,b) = \frac{\partial^2}{\partial y \partial x} F_{(X,Y)}(a,b)\,.$$

Take

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dx.$$

Then $f_X \ge 0$ and $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. Same holds for $f_Y(\cdot)$. Moreover, for any $A, B \in \mathcal{B}(\mathbb{R})$, we have

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in (-\infty,\infty)) = \int_{-\infty}^{\infty} \int_A f_{(X,Y)}(x,y)\,dx\,dy = \int_A f_X(x)\,dx$$

$$\mathbb{P}(Y \in B) = \mathbb{P}(X \in (-\infty,\infty), Y \in B) = \int_B \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dx\,dy = \int_B f_Y(y)\,dy\,.$$

The functions $f_X(\cdot)$ and $f_Y(\cdot)$ are referred as **marginal probability density function (marginal pdf)** of X and Y.

Example 4.4. *The joint pdf of X and Y is given by*

$$f_{(X,Y)}(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & x > 0, y > 0 \\ 0, & otherwise \end{cases}$$

Then compute: i) $\mathbb{P}(X > 1, Y < 1)$, ii) $\mathbb{P}(X > Y)$, iii) $\mathbb{P}(X < a)$.

Solution: *Note that $\mathbb{P}(X > 1, Y < 1) = \mathbb{P}(X \in A, Y \in B)$ where $A = (1,\infty)$, $B = (-\infty, 1)$. Thus, we have*

$$\mathbb{P}(X > 1, Y < 1) = \int_{-\infty}^{1} \int_{1}^{\infty} f_{(X,Y)}(x,y)\,dx\,dy = \int_0^1 \int_1^{\infty} 2e^{-x}e^{-2y}\,dx\,dy = \int_0^1 2e^{-2y}\Big[-e^{-x}\Big]_1^{\infty}\,dy$$

$$= e^{-1} \int_0^1 2e^{-2y}\,dy = e^{-1}(1 - e^{-2})\,.$$

For ii)*, we proceed as follows:*

$$\mathbb{P}(X < Y) = \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy = 2\int_0^{\infty}\int_0^{y} e^{-x}e^{-2y}\,dx\,dy = \int_0^{\infty} 2e^{-2y}(1 - e^{-y})\,dy = \frac{1}{3}.$$

Observe that $\mathbb{P}(X < a) = \mathbb{P}(X \in A)$ where $A = (-\infty, a)$. Thus, we have

$$\mathbb{P}(X < a) = \int_{-\infty}^{a} f_X(x)\,dx = \int_{-\infty}^{a} \Big(\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy\Big)\,dx$$

$$= \int_0^a \Big(\int_0^{\infty} 2e^{-2y}e^{-x}\,dy\Big)\,dx = \int_0^a e^{-x}\,dx = 1 - e^{-a}.$$

4.1. **Independent random variables:** Recall that $E, F \in \mathcal{F}$ are independent if $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$. Let X and Y be two given random variables. For any $A, B \in \mathcal{B}(\mathbb{R})$, $E = \{X \in A\}$ and $F = \{Y \in B\}$ are elements of $\mathcal{F}$. We can examine whether E and F are independent events or not.

Definition 4.6 (Independent random variables). Let X and Y be two real-valued random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X and Y are independent if for any $A, B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\,. \tag{4.1}$$

Taking $A = (-\infty, x]$ and $B = (-\infty, y]$, we see that condition of independence (4.1) is equivalent to

$$\mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y) \text{ i.e., } F_{(X,Y)}(x,y) = F_X(x)F_Y(y) \quad \forall\, x, y\,.$$

When X and Y are discrete random variables, the condition (4.1) is equivalent to

$$p_{(X,Y)}(x,y) = p_X(x)p_Y(y) \quad \forall\ x,y\,.$$

Indeed, taking $A = \{x\}$ and $B = \{y\}$, we have

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)) \text{ i.e., } p_{(X,Y)}(x,y) = p_X(x)p_Y(y).$$

Conversely, let $p_{(X,Y)}(x,y) = p_X(x)p_Y(y)$. For any $A, B \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{y\in B}\sum_{x\in A} p_{(X,Y)}(x,y) = \sum_{y\in B}\sum_{x\in A} p_X(x)p_Y(y)\\ &= \sum_{x\in A} p_X(x)\sum_{y\in B} p_Y(y) = \mathbb{P}(X\in A)\mathbb{P}(Y\in B)\,.\end{aligned}$$

In the jointly continuous case, the condition of independence is equivalent to

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y) \quad \forall\ x,y.$$

Theorem 4.2. *Let X and Y be independent random variables. Then for any Borel measurable functions f and g, the random variables $f(X)$ and $g(Y)$ are independent.*

Proof. We need to show that, for any $(x,y) \in \mathbb{R}^2$,

$$\mathbb{P}(f(X) \le x, g(Y) \le y) = \mathbb{P}(f(X) \le x)\mathbb{P}(g(Y) \le y).$$

Indeed, we have

$$\begin{aligned}&\mathbb{P}(f(X) \le x, g(Y) \le y) = \mathbb{P}\big(X \in f^{-1}(-\infty,x], Y \in g^{-1}(-\infty,y]\big)\\ &= \mathbb{P}\big(X \in f^{-1}(-\infty,x]\big)\mathbb{P}\big(Y \in g^{-1}(-\infty,y]\big) = \mathbb{P}(f(X)\le x)\mathbb{P}(g(Y)\le y)\,,\end{aligned}$$

where in the second equality, we have used that X and Y are independent. This completes the proof. □

A necessary and sufficient condition for the random variables X and Y to be independent is given in the following theorem.

Theorem 4.3. *The continuous (discrete) random variables X and Y are independent if and only iff their joint pdf (pmf) can be expressed as*

$$f_{(X,Y)}(x,y) = h(x)g(y), \quad -\infty < x < \infty,\ -\infty < y < \infty.$$

The concept of independence can be generalised for n random variables as follows: $X_1, X_2, \ldots, X_n$ are said to be independent if for any $A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}\big(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\big) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

which is equivalent to

$$F_{(X_1,X_2,\ldots,X_n)}(x_1,x_2,\ldots,x_n) = \prod_{i=1}^n F_{X_i}(x_i).$$

We say that an infinite collection of random variables is independent if every finite sub-collection is independent.

Definition 4.7. We say that two random variables X and Y are independent and identically distributed, in short **i.i.d**, random variables if

a) X and Y are independent
b) X and Y have same distribution i.e., $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for any $A \in \mathcal{B}(\mathbb{R})$.

If $X_1, X_2, \dots, X_n$ are n independent random variables, then $X_1, X_2, \dots, X_k$, $k \le n$ are also independent. Indeed,

$$\begin{aligned}\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_k \in A_k) &= \mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_k \in A_k, X_{k+1} \in \mathbb{R}, \dots, X_n \in R)\\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \quad \text{where } A_i = \mathbb{R} \text{ for } i = k+1, k+2, \dots, n\\ &= \prod_{i=1}^{k} \mathbb{P}(X_i \in A_i).\end{aligned}$$

Moreover, we have the following theorem.

Theorem 4.4. *Let $X_1, X_2, \dots, X_n$ are n independent random variables. Let Y be a random variable defined in terms of $X_1, X_2, \dots, X_k$ and Z be another random variable defined in terms of $X_{k+1}, X_{k+2}, \dots, X_n$ where $1 \le k \le n$. Then Y and Z are independent.*

Example 4.5. *Let X_1, X_2, X_3 and X_4 are independent random variables. Then $Y = X_1X_2 + X_3$ and $Z = e^{X_4}$ are independent, and hence*

$$\mathbb{E}[e^{X_4}(X_1X_2 + X_3)] = \mathbb{E}[e^{X_4}]\mathbb{E}[X_3] + \mathbb{E}[e^{X_4}]\mathbb{E}[X_1]\mathbb{E}[X_2].$$

Example 4.6. *Let X and Y be independent random variables with pdf given by*

$$f_X(x) = \frac{x^2}{9}\mathbf{1}_{(0,3)}(x), \quad f_Y(y) = \frac{1}{y^2}\mathbf{1}_{(1,\infty)}(y).$$

Find $\mathbb{P}(XY > 1)$.
Solution: *Since X and Y are independent, their joint pdf is given by*

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{x^2}{9y^2}, & 0 < x < 3,\ y > 1\\ 0, & \textit{otherwise.}\end{cases}$$

Hence, we obtain

$$\begin{aligned}\mathbb{P}(XY > 1) &= \int_1^\infty \int_{\frac{1}{y}}^3 \frac{x^2}{9y^2}\,dx\,dy = \int_1^\infty \frac{1}{27\,y^2}[27 - \frac{1}{y^3}]\,dy\\ &= \int_1^\infty \frac{1}{y^2}\,dy - \frac{1}{27}\int_1^\infty \frac{1}{y^5}\,dy = 1 - \frac{1}{108} \approx 0.99074.\end{aligned}$$

Sum of independent random variables: Let X and Y be two independent random variables. Let $Z = X + Y$. Then

$$\begin{aligned}F_Z(a) = \mathbb{P}(X + Y \le a) &= \iint_{x+y\le a} f_{(X,Y)}(x,y)\,dx\,dy = \iint_{x+y\le a} f_X(x)f_Y(y)\,dx\,dy\\ &= \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{a-y} f_X(x)\,dx\Big)f_Y(y)\,dy = \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,dy\\ \implies f_Z(a) &= \frac{d}{da}F_Z(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy = f_X * f_Y(a)\end{aligned}$$

where $*$ is the convolution operation.

Example 4.7. *Let X and Y be two independent random variables with pdf f_X and f_Y given by*

$$f_X(a) = f_Y(a) = \begin{cases} 1, & 0 < a < 1\\ 0, & \textit{otherwise.}\end{cases}$$

Then find the pdf of $X + Y$.

Solution: *We have* $f_{X+Y}(x) = \int_0^1 f_X(x-y)\,dy$. *We evaluate this integral for different values of* x. *Observe that, if* $x < 0$ *or* $x \geq 2$, $f_X(x-y) = 0$ *for all* $y \in (0,1)$ *and hence* $f_{X+Y}(x) = 0$.
Case 1: $0 \leq x \leq 1$**:** *Then* $f_{X+Y}(x) = \int_0^x dy = x$.
Case 2: $1 \leq x \leq 2$**:** *Then* $f_{X+Y}(x) = \int_{x-1}^1 dy = 2 - x$.
Thus, we obtain

$$f_{X+Y}(x) = \begin{cases} 0, & 0 < x; \quad x \geq 2 \\ x, & 0 \leq x \leq 1 \\ 2-x, & 1 < x < 2. \end{cases}$$

Example 4.8. *Let* X *and* Y *be two independent Poisson random variables with parameter* λ_1 *and* λ_2. *Then find the distribution of* $X + Y$.
Solution: *Note that the event* $\{X + Y = n\}$ *may be written as union of disjoint events:* $\{X = k, Y = n-k\}$, $0 \leq k \leq n$. *Thus, we have*

$$\begin{aligned}
\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n} \mathbb{P}(X=k, Y=n-k) = \sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k) = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!} e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}.
\end{aligned}$$

Thus, $X + Y$ *has a Poisson distribution with parameter* $(\lambda_1 + \lambda_2)$. *We can generalise it to any finite sum of independent Poisson random variables.*

Lemma 4.5. *If* X *and* Y *are independent then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Proof. Suppose X and Y are jointly continuous with joint pdf $f_{(X,Y)}$. Then

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{(X,Y)}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dx\,dy \\
&= \Big(\int_{-\infty}^{\infty} x f_X(x)\,dx\Big)\Big(\int_{-\infty}^{\infty} y f_Y(y)\,dy\Big) = \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

□

Converse of this lemma does not hold in general. To see this, let X be continuous random variable with pdf

$$f_X(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Take $Y = X^2$. Then clearly X and Y are NOT independent, but observe that

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0, \quad \mathbb{E}[X] = 0, \quad \mathbb{E}[Y] = \mathbb{E}[X^2] = \frac{1}{3} \implies \mathbb{E}[XY] = 0 = \mathbb{E}[X]\mathbb{E}[Y].$$

4.2. **Moment generating function of sum of independent random variables:** Let X and Y be two independent random variables with moment generating function $m_X(t)$ amd $m_Y(t)$. Then the moment generating function of $X + Y$ can be calculated as follows:

$$m_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}e^{tY}].$$

Since X and Y are independent, the random variables e^{tX} and e^{tY} and independent and hence

$$m_{X+Y}(t) = \mathbb{E}[e^{tX}e^{tY}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] = m_X(t)m_Y(t).$$

Example 4.9. *Let X and Y be two independent binomial random variables with parameters (n,p) and (m,p) respectively. Find the distribution of $X+Y$.*
Solution: *We have seen that moment generating function of a binomial (n,p) random variable is $m_X(t)=(pe^t+1-p)^n$. The mgf of $X+Y$ is*

$$m_{X+Y}(t)=m_X(t)m_Y(t)=(pe^t+1-p)^n(pe^t+1-p)^m=(pe^t+1-p)^{n+m}.$$

This shows that $X+Y$ is again a binomial random variable with parameters $(m+n)$ and p. Similarly, if X_i are independent binomial random variables with parameters (n_i,p) for $i=1,2,\dots N$, then $\sum_{i=1}^N X_i$ is binomial random variable with parameters $(\sum_{i=1}^N n_i,p)$.

4.3. **Characteristic function of sum of independent random variables:** Let X and Y be two independent random variables with characteristic $\phi_X(t)$ amd $\phi_Y(t)$ respectively. Then the characteristic function of $X+Y$ is

$$\begin{aligned}\phi_{X+Y}&=\mathbb{E}[e^{it(X+Y)}]=\mathbb{E}[e^{itX}e^{itY}]=\mathbb{E}\big[\{\cos(tX)+i\sin(tX)\}\cdot\{\cos(tY)+i\sin(tY)\}\big]\\&=\mathbb{E}[\cos(tX)]\mathbb{E}[\cos(tY)]-\mathbb{E}[\sin(tX)]\mathbb{E}[\sin(tY)]\\&\qquad+i\Big\{\mathbb{E}[\sin(tX)]\mathbb{E}[\cos(tY)]+\mathbb{E}[\cos(tX)]\mathbb{E}[\sin(tY)]\Big\}\\&=\big\{\mathbb{E}[\cos(tX)]+i\,\mathbb{E}[\sin(tX)]\big\}\cdot\big\{\mathbb{E}[\cos(tY)]+i\,\mathbb{E}[\sin(tY)]\big\}=\phi_X(t)\phi_Y(t).\end{aligned}$$

Example 4.10. *Suppose $X\sim\mathcal{N}(\mu,\sigma^2)$ and $Y\sim\mathcal{N}(\nu,\tau^2)$ such that X and Y are independent. Then $X+Y\sim\mathcal{N}(\mu+\nu,\sigma^2+\tau^2)$. Indeed,*

$$\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)=e^{i\mu t-\frac{t^2\sigma^2}{2}}e^{i\nu t-\frac{t^2\tau^2}{2}}=e^{it(\mu+\nu)-\frac{t^2(\sigma^2+\tau^2)}{2}}$$

which is the characteristic function of a normal $\mathcal{N}(\mu+\nu,\sigma^2+\tau^2)$. Similarly, if $X_i\sim\mathcal{N}(\mu_i,\sigma_i^2)$ for $i=1,2,\dots N$ and they are independent, then $\sum_{i=1}^N X_i\sim\mathcal{N}(\sum_{i=1}^N\mu_i,\sum_{i=1}^N\sigma_i^2)$.

4.4. **Joint probability distribution of functions of random variables:** Let $X_1,X_2,\dots,X_n$ be random variables and $g:\mathbb{R}^n\to\mathbb{R}$ be a measurable function such that $Y=g(X_1,X_2,\dots,X_n)$ is a real-valued random variable. Then

$$\begin{aligned}\mathbb{P}(Y\le y)&=\mathbb{P}(g(X_1,X_2,\dots,X_n)\le y)\\&=\begin{cases}\displaystyle\sum_{\{(x_1,x_2,\dots,x_n):g(x_1,x_2,\dots,x_n)\le y\}}\mathbb{P}(X_1=x_1,X_2=x_2,\dots,X_n=x_n) & \text{discrete case}\\ \displaystyle\iint_{\{(x_1,x_2,\dots,x_n):g(x_1,x_2,\dots,x_n)\le y\}} f_{(X_1,X_2,\dots,X_n)}(x_1,x_2,\dots,x_n)\,dx_1\,dx_2\dots dx_n & \text{continuous case}\end{cases}\end{aligned}$$

where in the continuous case, $f_{(X_1,X_2,\dots,X_n)}(\cdot)$ is the joint density function of the n random variables $X_1,X_2,\dots,X_n$.

Example 4.11. *Let (X_1,X_2) be a two dimensional random vector with joint pdf given by*

$$f_{(X_1,X_2)}(x_1,x_2)=\begin{cases}2, & 0\le x_1\le x_2\le 1\\ 0, & otherwise\ .\end{cases}$$

What is the pdf of X_1+X_2?
Solution: *Let $Y=X_1+X_2$. Observe that for $y<0$, $\mathbb{P}(Y\le y)=0$ and for $y>2$, $\mathbb{P}(Y\le y)=1$. Let $0\le y\le 1$. Then*

$$\mathbb{P}(Y\le y)=\int_{x_1=0}^{\frac{y}{2}}\int_{x_2=x_1}^{y-x_1}2dx_2\,dx_1=2\int_0^{\frac{y}{2}}(y-2x_1)\,dx_1=\frac{y^2}{2}\,.$$

Let $1<y\le 2$. Then

$$\mathbb{P}(Y\le y)=\int_{x_1=0}^{y-1}\int_{x_2=x_1}^{2}2dx_2\,dx_1+\int_{x_1=y-1}^{\frac{y}{2}}\int_{x_2=x_1}^{y-x_1}2dx_2\,dx_1=-\frac{y^2}{2}+2y-1$$

Hence, we have

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 \le y \le 1 \\ -\frac{y^2}{2} + 2y - 1 & 1 < y \le 2 \\ 0 & y > 2\,. \end{cases}$$

Hence the density function of Y is given by

$$f_Y(y) = \begin{cases} y & 0 \le y \le 1 \\ 2 - y & 1 < y \le 2 \\ 0 & \textit{otherwise.} \end{cases}$$

Let $f_{(X_1,X_2,\ldots,X_n)}(\cdot)$ be a given joint density function of the n random variables $X_1, X_2, \ldots, X_n$. We want to find joint density function of $Y_1, Y_2, \ldots Y_n$, where $Y_i = g_i(X_1, X_2, \ldots, X_n)$, $i \le i \le n$. Assume that the functions g_i have continuous partial derivatives and that the Jacobian determinant

$$J(x_1, x_2, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0$$

at all points $(x_1, x_2, \ldots, x_n)$. Further we assume that the equations $y_i = g_i(x_1, x_2, \ldots, x_n)$ have a unique solution, say $x_i = h_i(y_1, y_2, \ldots, y_n)$ for $1 \le i \le n$. Then the joint density function of $Y_1, Y_2, \ldots Y_n$, is given by

$$f_{(Y_1,Y_2,\ldots,Y_n)}(y_1, y_2, \ldots, y_n) = f_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n)|J(x_1, x_2, \ldots, x_n)|^{-1}$$

where $x_i = h_i(y_1, y_2, \ldots, y_n)$ for $i = 1, 2, \ldots, n$.

Example 4.12. *Let X_1 and X_2 be two independent standard normal random variables. Find the joint density function of Y_1 and Y_2 where*

$$Y_1 = X_1 + X_2, \quad Y_2 = X_1 - X_2.$$

Solution: *Let $g_1(x_1, x_2) = x_1 + x_2$ and $g_2(x_1, x_2) = x_1 - x_2$. Then*

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \neq 0\,.$$

Moreover, the equations $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ have unique solution given by

$$x_1 = \frac{y_1 + y_2}{2}, \quad x_2 = \frac{y_1 - y_2}{2}.$$

Observe that, due to independence, the joint density function of X_1 and X_2 is given by

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}}.$$

Thus, the joint density function of Y_1 and Y_2 is

$$f_{(Y_1,Y_2)}(y_1, y_2) = \frac{f_{(X_1,X_2)}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right)}{2} = \frac{1}{\sqrt{4\pi}} e^{-\frac{y_1^2}{4}} \frac{1}{\sqrt{4\pi}} e^{-\frac{y_2^2}{4}}\,.$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable function and $X_1, X_2, \ldots, X_n$ are given random variables. If $(X_1, X_2, \ldots, X_n)$ is discrete type and

$$\sum_{x_1, x_2, \ldots, x_n} |g(x_1, x_2, \ldots, x_n)| \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) < +\infty$$

then the expected value of $Y = g(X_1, X_2, \dots, X_n)$ is given by

$$\mathbb{E}[g(X_1, X_2, \dots, X_n)] = \sum_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n) \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) .$$

If $(X_1, X_2, \dots, X_n)$ is of continuous type with joint density function $f_{(X_1, X_2, \dots, X_n)}(\cdot, \cdot, \dots, \cdot)$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, x_2, \dots, x_n)| f_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) \prod_{i=1}^{n} dx_i < +\infty$$

then the expected value of $Y = g(X_1, X_2, \dots, X_n)$ is given by

$$\mathbb{E}[g(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) \prod_{i=1}^{n} dx_i .$$

Example 4.13. *An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident, an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.*

Solution: *We need to compute $\mathbb{E}[|X - Y|]$. Let $g(x, y) = |x - y|$. Then g is continuous on $\mathbb{R}^2$ and hence $|X - Y|$ is a random variable. Due to independence, the joint density function of X and Y is given by*

$$f_{(X,Y)}(x, y) = \frac{1}{L^2}, \quad 0 < x < L, \; 0 < y < L .$$

Therefore,

$$\mathbb{E}[|X - Y|] = \frac{1}{L^2} \int_0^L \int_0^L |x - y| \, dy \, dx .$$

Observe that

$$\int_0^L |x - y| \, dy = \int_0^x (x - y) \, dy + \int_x^L (y - x) \, dy = \frac{L^2}{2} + x^2 - xL .$$

Hence, we have

$$\mathbb{E}[|X - Y|] = \frac{1}{L^2} \int_0^L \big(\frac{L^2}{2} + x^2 - xL \big) \, dx = \frac{L}{3} .$$

4.5. **Conditional expectation and its properties:** Recall that for any events E and F, $\mathbb{P}(E \big| F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$ provided $\mathbb{P}(F) > 0$. Let X and Y be jointly discrete random variables. Then we define the conditional probability mass function (conditional pmf) of X given that $Y = y$, denoted by $p_{X \big| Y}(x \big| y)$

$$p_{X \big| Y}(x \big| y) := \mathbb{P}(X = x \big| Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{(X,Y)}(x, y)}{p_Y(y)}$$

for all y such that $p_Y(y) > 0$. The conditional distribution function of X given that $Y = y$ is defined, for all y such that $p_Y(y) > 0$, by

$$F_{X \big| Y}(x \big| y) := \mathbb{P}(X \le x | Y = y) = \sum_{a \le x} p_{X \big| Y}(a \big| y).$$

Remark 4.1. If X is independent of Y, then $p_{X \big| Y}(x \big| y) = \mathbb{P}(X = x)$. Indeed

$$p_{X \big| Y}(x \big| y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(X = x) \mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} = \mathbb{P}(X = x) .$$

Example 4.14. *Let X and Y be two discrete random variables such that their joint pmf is given by*

$$p_{(X,Y)}(0,0) = 0.4, \quad p_{(X,Y)}(0,1) = 0.2, \quad p_{(X,Y)}(1,0) = 0.1, \quad p_{(X,Y)}(1,1) = 0.3.$$

Find $\mathbb{P}(X = i \big| Y = 1)$ and $\mathbb{P}(X = i \big| Y = 0)$.
Solution: *First note that*

$$p_Y(0) = \sum_{i=0}^{1} p_{(X,Y)}(i,0) = 0.4 + 0.1 = 0.5, \quad p_Y(1) = \sum_{i=0}^{1} p_{(X,Y)}(i,1) = 0.2 + 0.3 = 0.5.$$

Hence, we have

$$\mathbb{P}(X = i \big| Y = 1) = \frac{p_{(X,Y)}(i,1)}{p_Y(1)} = \begin{cases} \frac{2}{5}, & i = 0 \\ \frac{3}{5}, & i = 1. \end{cases}$$

$$\mathbb{P}(X = i \big| Y = 0) = \frac{p_{(X,Y)}(i,0)}{p_Y(0)} = \begin{cases} \frac{4}{5}, & i = 0 \\ \frac{1}{5}, & i = 1. \end{cases}$$

Example 4.15. *Let X and Y be two independent real-valued random variables such that $X \sim \mathrm{Poisson}(\lambda_1)$ and $Y \sim \mathrm{Poisson}(\lambda_2)$. Calculate the conditional distribution of X given that $X+Y = n$.*
Solution: *We first calculate the conditional pmf of X given that $X + Y = n$. We have*

$$p_{X\big|X+Y}(k\big|n) = \frac{\mathbb{P}(X = k, X + Y = n)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k)\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)}$$

Since X and Y are independent, $X + Y \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$. Hence we obtain

$$\begin{aligned} p_{X\big|X+Y}(k\big|n) &= e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \Big(e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!} \Big)^{-1} \\ &= \binom{n}{k} \Big(\frac{\lambda_1}{\lambda_1 + \lambda_2} \Big)^k \Big(\frac{\lambda_2}{\lambda_1 + \lambda_2} \Big)^{n-k}. \end{aligned}$$

In other words, the conditional distribution of X given that $X+Y = n$ is the binomial distribution with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

4.5.1. **Conditional distribution: continuous case:** If X and Y are jointly continuous, then $\mathbb{P}(X = x) = 0 = \mathbb{P}(Y = y)$, and hence $\mathbb{P}(X \in B \big| Y = y)$ is not defined similar to discrete case. Let $\varepsilon > 0$, and suppose that $\mathbb{P}(Y \in (y-\varepsilon, y+\varepsilon]) > 0$. For every x and every interval $(y-\varepsilon, y+\varepsilon]$, consider the conditional probability of events $X \le x$ given that $Y \in (y - \varepsilon, y + \varepsilon]$

$$\mathbb{P}(X \le x \big| y - \varepsilon < Y \le y + \varepsilon) = \frac{\mathbb{P}(X \le x, y - \varepsilon < Y \le y + \varepsilon)}{\mathbb{P}(y - \varepsilon < Y \le y + \varepsilon)}.$$

Definition 4.8. The conditional cumulative distribution function (conditional cdf) of X given $Y = y$, denoted by $F_{X|Y)}(x|y)$, is defined as the limit

$$\lim_{\varepsilon \to 0} \mathbb{P}(X \le x \big| y - \varepsilon < Y \le y + \varepsilon)$$

provided the limit exists. The conditional density function of X, given $Y = y$, denoted as $f_{X|Y}(x|y)$, as a non-negative function satisfying

$$F_{X|Y)}(x|y) = \int_{\infty}^{x} f_{X|Y)}(t|y)\, dt\,.$$

Let $f_{(X,Y)}(x,y)$ be the joint pdf of X and Y such that $f_{(X,Y)}(\cdot,\cdot)$ is continuous and the marginal pdf $f_Y(\cdot)$ is continuous. Let y be such that $f_Y(y)>0$. Then we have

$$F_{X|Y)}(x|y) = \lim_{\varepsilon\to 0} \frac{\int_{-\infty}^{x}\int_{y-\varepsilon}^{y+\varepsilon} f_{(X,Y)}(t,u)\,du\,dt}{\int_{y-\varepsilon}^{y+\varepsilon} f_Y(u)\,du} = \frac{\int_{-\infty}^{x} f_{(X,Y)}(t,y)\,dt}{f_Y(y)} = \int_{-\infty}^{x} \frac{f_{(X,Y)}(t,y)}{f_Y(y)}\,dt$$

It follows that, there exists a conditional pdf of X, given $Y=y$, that is expressed by

$$f_{X|Y)}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \quad f_Y(y)>0\,.$$

The use of conditional densities allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable. That is, if X and Y are jointly continuous, then, for any set A,

$$\mathbb{P}(X\in A|Y=y) = \int_A f_{X|Y)}(x|y)\,dx\,.$$

Example 4.16. *Suppose that the joint density function of X and Y is given by*

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}}e^{-y}}{y}, & 0<x<\infty,\ 0<y<\infty \\ 0, & otherwise \end{cases}.$$

Find $\mathbb{P}(X>1|Y=y)$.
Solution: *We need to find out the conditional density $f_{X|Y)}(x|y)$. Observe that*

$$f_{X|Y)}(x|y) = \frac{f_{(X,Y)}(x,y)}{\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dx} = \frac{f_{(X,Y)}(x,y)}{e^{-y}} = \frac{1}{y}e^{-\frac{x}{y}}$$

and therefore

$$\mathbb{P}(X>1|Y=y) = \int_1^{\infty} \frac{1}{y}e^{-\frac{x}{y}}\,dx = e^{-\frac{1}{y}}\,.$$

4.5.2. **Conditional expectation:** Let X and Y be two given random variables. The conditional expectation of X, given $Y=y$, denoted as $\mathbb{E}[X|Y=y]$, is defined by

$$\mathbb{E}[X|Y=y] = \begin{cases} \sum_x x\mathbb{P}(X=x|Y=y), & \text{if } X \text{ and } Y \text{ are discrete and } p_Y(y)>0, \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx, & \text{if } X \text{ and } Y \text{ are continuous and } f_Y(y)>0. \end{cases}$$

Example 4.17. *If X and Y are independent binomial random variables with identical parameters n and p, calculate $\mathbb{E}[X|X+Y=m]$.*
Solution: *First we note that since X and Y are independent and $X,Y\sim\mathcal{B}(n,p)$, the sum of the random variable $X+Y$ is again binomial random variable with parameters $2n$ and p i.e., $X+Y\sim\mathcal{B}(2n,p)$. Now for $k\le\min\{m,n\}$*

$$\begin{aligned}\mathbb{P}(X=k|X+Y=m) &= \frac{\mathbb{P}(X=k,X+Y=m)}{\mathbb{P}(X+Y=m)} = \frac{\mathbb{P}(X=k,Y=m-k)}{\mathbb{P}(X+Y=m)} \\ &= \frac{\mathbb{P}(X=k)\mathbb{P}(Y=m-k)}{\mathbb{P}(X+Y=m)} = \frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}} \\ &= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}\end{aligned}$$

Hence,

$$\mathbb{E}[X|X+Y=m] = \sum_{k:k\le \min\{m,n\}} k\mathbb{P}(X=k|X+Y=m) = \sum_{k:k\le \min\{m,n\}} k\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}} = \frac{m}{2}.$$

Example 4.18. *Suppose that the joint density function of X and Y is given by*

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}}e^{-y}}{y}, & 0<x<\infty,\ 0<y<\infty \\ 0, & otherwise \end{cases}.$$

Find $\mathbb{E}(X|Y=y)$.

Solution: *We have already calculated conditional density $f_{X|Y}(x|y) = \frac{1}{y}e^{-\frac{x}{y}}$; see Example 4.16. Hence the conditional distribution of X, given that $Y=y$, is just the exponential distribution with mean y. Thus, $\mathbb{E}(X|Y=y)=y$.*

Remark 4.2. Let $g:\mathbb{R}\to\mathbb{R}$ be a Borel measurable function and X, Y, X_i $1\le i\le n$ are given random variables. Then the following formulas remain valid:

$$\mathbb{E}[g(X)|Y=y] = \begin{cases} \sum_x g(x)\mathbb{P}(X=x|Y=y), & \text{if } X \text{ and } Y \text{ are discrete,} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,dx, & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

$$\mathbb{E}\Big[\sum_{i=1}^n X_i|Y=y\Big] = \sum_{i=1}^n \mathbb{E}[X_i|Y=y]\,.$$

Definition 4.9. Let X and Y be real-valued random variables defined on $(\Omega,\mathcal{F},\mathbb{P})$ and $h:\mathbb{R}\to\mathbb{R}$ be a Borel measurable function. Then the random variable $\mathbb{E}[h(X)|Y]$ defined by

$$\begin{aligned} \mathbb{E}[h(X)|Y] :&\Omega \to \mathbb{R} \\ &\omega \mapsto \mathbb{E}[h(X)|Y=Y(\omega)] \end{aligned}$$

is called the conditional expectation of $h(X)$ given Y. That is $\mathbb{E}[h(X)|Y]$ is a random variable which takes value $\mathbb{E}[h(X)|Y=y]$.

An extremely important property of conditional expectations is given by the following proposition.

Proposition 4.6. *Let X, Y and Z be real-valued random variables defined on $(\Omega,\mathcal{F},\mathbb{P})$ and $h:\mathbb{R}\to\mathbb{R}$ be a Borel measurable function. If $\mathbb{E}[h(X)]$ exists then*

$$\mathbb{E}[h(X)] = \mathbb{E}\big[\mathbb{E}[h(X)|Y]\big].$$

Moreover, the followings hold:

i) $\mathbb{E}[X|Y]\ge 0$ *if* $\mathbb{P}(X\ge 0)=1$.
ii) If X and Y are independent, then $\mathbb{E}[X|Y]=\mathbb{E}[X]$.
iii) $\mathbb{E}[Xh(Y)|Y] = h(Y)\mathbb{E}[X|Y]$.
iv) $\mathbb{E}[\alpha X+\beta Y|Z] = \alpha\mathbb{E}[X|Z]+\beta\mathbb{E}[Y|Z]$.

Proof. Let X and Y are discrete. Then

$$\begin{aligned} \mathbb{E}\big[\mathbb{E}[h(X)|Y]\big] &= \sum_y \mathbb{E}[h(X)|Y=y]\mathbb{P}(Y=y) = \sum_y\sum_x h(x)\mathbb{P}(X=x|Y=y)\mathbb{P}(Y=y) \\ &= \sum_x h(x)\Big(\sum_y \mathbb{P}(X=x,Y=y)\Big) = \sum_x h(x)\mathbb{P}(X=x) = \mathbb{E}[h(X)]\,. \end{aligned}$$

If X and Y are continuous random variables with joint probability density function $f_{(X,Y)}$, then

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[h(X)|Y]\big] &= \int_{-\infty}^{\infty} \mathbb{E}[h(X)|Y=y] f_Y(y)\,dy = \int_{-\infty}^{\infty}\Big(\int_{\infty}^{\infty} h(x) f_{(X|Y)}(x|y)\,dx\Big) f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty}\int_{\infty}^{\infty} h(x) f_{(X,Y)}(x,y)\,dx\,dy = \int_{\infty}^{\infty} h(x)\Big(\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy\Big)\,dx \\ &= \int_{\infty}^{\infty} h(x) f_X(x)\,dx = \mathbb{E}[h(X)]\,.\end{aligned}$$

Proof of i)-iv) is left as exercise. □

Example 4.19. *Suppose that the number of people entering a department store on a given day is a random variable with mean 20. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of \$4. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day ?*

Solution: *Let N be the number of customers enter the store on a given day and ith customer spent X_i amount of money. It is given that $\mathbb{E}[N]=20$, $\mathbb{E}[X_i]=4$ and N and X_i's are independent. The total amount of money spent is $Y := \sum_{i=1}^N X_i$. We need to find $\mathbb{E}[Y]$. Conditioning on N, we have $\mathbb{E}[Y]=\mathbb{E}\big[\mathbb{E}[Y|N]\big]$. Now*

$$\mathbb{E}[Y|N=n] = \mathbb{E}\big[\sum_{i=1}^N X_i|N=n\big] = \sum_{i=1}^n \mathbb{E}[X_i|N=n] = \sum_{i=1}^n \mathbb{E}[X_i] = n\mathbb{E}[X]$$

where $\mathbb{E}[X]=\mathbb{E}[X_i]$. This implies that $\mathbb{E}\big[\sum_{i=1}^N X_i|N\big] = N\mathbb{E}[X]$. Therefore, we have

$$\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y|N]\big] = \mathbb{E}\big[N\mathbb{E}[X]\big] = \mathbb{E}[N]\mathbb{E}[X] = \$80\,.$$

The expected amount of money spent in the store on a given day is \$80.

4.6. **Covariance, variance of sums, correlations and conditional variance:**

Definition 4.10. The covariance between two random variables X and Y, denoted by $\mathrm{Cov}(X,Y)$, is defined by

$$\mathrm{Cov}(X,Y) = \mathbb{E}\Big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\Big] = \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\,.$$

Proposition 4.7 (Properties of covariance)**.** *The followings hold:*

i) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$ *and* $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$.
ii) $\mathrm{Cov}(aX+b,Y)=a\mathrm{Cov}(X,Y)$ *for any* $a,b\in\mathbb{R}$.
iii) $\mathrm{Cov}\Big(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\Big) = \sum_{i=1}^n\sum_{j=1}^m \mathrm{Cov}(X_i,Y_j)$.
iv) $\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)$.

Proof. i) follows from the definition. To prove ii), we use linearity of expectation and have

$$\begin{aligned}\mathrm{Cov}(aX+b,Y) &= \mathbb{E}\Big[\big(aX+b-a\mathbb{E}[X]-b\big)(Y-\mathbb{E}[Y])\Big] \\ &= \mathbb{E}\Big[a(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\Big] = a\,\mathrm{Cov}(X,Y)\,.\end{aligned}$$

Proof of iii): Let $\mu_i=\mathbb{E}[X_i]$ and $\gamma_j=\mathbb{E}[Y_j]$. Then $\mathbb{E}\Big[\sum_{i=1}^n X_i\Big]=\sum_{i=1}^n\mu_i$ and $\mathbb{E}\Big[\sum_{j=1}^m Y_j\Big]=\sum_{j=1}^m\gamma_j$. Using these, we have

$$\mathrm{Cov}\Big(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\Big) = \mathbb{E}\Big[\sum_{i=1}^n (X_i-\mu_i)\sum_{j=1}^m (Y_j-\gamma_j)\Big] = \mathbb{E}\Big[\sum_{i=1}^n\sum_{j=1}^m (X_i-\mu_i)(Y_j-\gamma_j)\Big]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}\Big[(X_i-\mu_i)(Y_j-\gamma_j)\Big] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Cov}(X_i,Y_j)\,,$$

where we have used the fact that the expected value of a sum of random variables is equal to sum of the expected values.
Proof of iv): In view of i), ii) and iii), we have

$$\begin{aligned}\mathrm{Var}\Big(\sum_{i=1}^{n}X_i\Big) &= \mathrm{Cov}\Big(\sum_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i\Big) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,X_j)\\ &= \sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)\,.\end{aligned}$$

□

Remark 4.3. If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$. However, the converse is NOT true in general. For example, consider a random variable X satisfying

$$\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=-1)=\frac{1}{3}$$

and Y given by

$$Y=\begin{cases}0, & \text{if } X\neq 0\\ 1, & \text{otherwise.}\end{cases}$$

Observe that $XY=0$ and $\mathbb{E}[X]=0$. Therefore, $\mathrm{Cov}(X,Y)=0$. However X and Y are clearly not independent.

In view of iv), we see that, if $X_1,X_2,\ldots,X_n$ are pairwise independent, then

$$\mathrm{Var}\Big(\sum_{i=1}^{n}X_i\Big)=\sum_{i=1}^{n}\mathrm{Var}(X_i)\,.$$

Example 4.20. *Let $X_1,X_2,\ldots,X_n$ be* **i.i.d** *random variables with $\mu=\mathbb{E}[X_i]$ and $\sigma^2=\mathrm{Var}(X_i)$. Find $\mathrm{Var}(\bar{X})$ and $\mathbb{E}[S^2]$ where*

$$\bar{X}:=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2:=\frac{1}{n-1}\sum_{i=1}^{n}\big(X_i-\bar{X}\big)^2\,.$$

$\bar{X}$ is called sample mean and S^2 is called sample variance.
Solution: *Since the expected value of a sum of random variables is equal to sum of the expected values, we see that $\mathbb{E}[\bar{X}]=\mu$. By using property of variance and independence of X_i's, we get*

$$\mathrm{Var}(\bar{X})=\frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n}X_i\Big)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{\sigma^2}{n}\,.$$

To find $\mathbb{E}[S^2]$, we start with the following identity:

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}\big(X_i-\bar{X}\big)^2=\sum_{i=1}^{n}\Big\{(X_i-\mu)^2+(\bar{X}-\mu)^2-2(\bar{X}-\mu)(X_i-\mu)\Big\}\\ &= \sum_{i=1}^{n}(X_i-\mu)^2+n(\bar{X}-\mu)^2-2n\,(\bar{X}-\mu)\sum_{i=1}^{n}\frac{X_i-\mu}{n}\\ &= \sum_{i=1}^{n}(X_i-\mu)^2-n(\bar{X}-\mu)^2\,.\end{aligned}$$

Taking expectation, we have

$$(n-1)\mathbb{E}[S^2] = \mathbb{E}\Big[\sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2\Big] = \sum_{i=1}^n \mathrm{Var}(X_i) - n\mathrm{Var}(\bar{X}) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$
$$\implies \mathbb{E}[S^2] = \sigma^2 .$$

We now prove one of the important inequality in probability theory, called Cauchy-Schwartz inequality.

Theorem 4.8 (Cauchy-Schwartz inequality)**.** *For random variables X and Y with $\mathbb{E}[X^2] < +\infty, \mathbb{E}[Y^2] < +\infty$, one has*

$$\big|\mathbb{E}[XY]\big|^2 \le \mathbb{E}[X^2]\mathbb{E}[Y^2] .$$

The equality holds if and only if there are real constants a and b, not both simultaneously zero, such that

$$\mathbb{P}(aX+bY=0)=1.$$

Proof. Let $\alpha = \mathbb{E}[Y^2]$ and $\beta = -\mathbb{E}[XY]$. Then $\alpha \ge 0$. For $\alpha = 0$, given the result holds. Let $\alpha > 0$. Thanks to linearity of expectation, we have

$$\begin{aligned}0 &\le \mathbb{E}\Big[\big(\alpha X+\beta Y\big)^2\Big] = \alpha^2\mathbb{E}[X^2] + \beta^2\mathbb{E}[Y^2] + 2\alpha\beta\mathbb{E}[XY] \\ &= \alpha\mathbb{E}[Y^2]\mathbb{E}[X^2] + \alpha\big(-\mathbb{E}[XY]\big)^2 + 2\alpha(-\mathbb{E}[XY])\mathbb{E}[XY] = \alpha\Big(\mathbb{E}[Y^2]\mathbb{E}[X^2] - \big(\mathbb{E}[XY]\big)^2\Big) .\end{aligned}$$

Since $\alpha > 0$, the assertion follows from the above inequality.

Note that if $\mathbb{E}[Y^2]\mathbb{E}[X^2] = \big(\mathbb{E}[XY]\big)^2$, then $\mathbb{E}\big[(\alpha X+\beta Y)^2\big] = 0$ and therefore we have $\mathbb{P}(\alpha X+\beta Y=0)=1$. If $\alpha > 0$, take $a=\alpha$ and $\beta = b$. If $\alpha = 0$, take $a=0$ and $b=1$. Conversely, let $\mathbb{P}(aX+bY=0)=1$ with $a \ne 0$. Then with probability 1, $X = -\frac{b}{a}Y$. In this case, one can easily show the equality in the assertion. This completes the proof. □

Remark 4.4. Replacing X by $|X|$ and Y by $|Y|$ in the Cauchy-Schwartz inequality, we get

$$\mathbb{E}[|XY|] \le \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]} . \tag{4.2}$$

Inequality (4.2) is also known as Cauchy-Schwartz inequality. Again, replacing X by $X-\mathbb{E}[X]$ and Y by $Y-\mathbb{E}[Y]$ in (4.2), we get the bound of the covariance of X and Y.

$$\big|\mathrm{Cov}(X,Y)\big| \le \mathbb{E}\Big[\big|(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big|\Big] \le \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)} .$$

4.6.1. **Correlation coefficient.**

Definition 4.11 (Correlation)**.** The correlation(coefficient) of two random variables X and Y, denoted by $\rho(X,Y)$, is defined by

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

where σ_X and σ_Y are standard deviation of X and Y respectively.

If $\rho(X,Y)=0$, then X and Y are said to be **uncorrelated**.

Lemma 4.9. *The correlation coefficient of X and Y has the following bound:*

$$-1 \le \rho(X,Y) \le 1.$$

Proof. Let σ_X and σ_Y are standard deviation of X and Y respectively. We have

$$0 \le \mathrm{Var}\Big(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\Big) = \mathrm{Var}\Big(\frac{X}{\sigma_X}\Big) + \mathrm{Var}\Big(\frac{Y}{\sigma_Y}\Big) \pm 2\mathrm{Cov}\Big(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\Big)$$

$$= \frac{1}{\sigma_X^2}\mathrm{Var}(X) + \frac{1}{\sigma_Y^2}\mathrm{Var}(Y) \pm 2\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = 2\big(1 \pm \rho(X,Y)\big)$$
$$\implies -1 \le \rho(X,Y) \le 1\,.$$

□

Remark 4.5. Let $\rho(X,Y) = -1$. Then following the proof of Lemma 4.9, we see that $1 - \rho(X,Y) = \frac{1}{2}\mathrm{Var}\Big(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\Big)$ and therefore, in this case, $\mathrm{Var}\Big(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\Big) = 0$. Hence, with probability 1, $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$ is constant. Therefore, there exists $a \in \mathbb{R}$ such that $Y = bX + a$ where $b = \frac{\sigma_Y}{\sigma_X} > 0$. Similarly, if $\rho(X,Y) = 1$, then with probability 1, $Y = bX + a$ for some $a \in \mathbb{R}$ with $b = -\frac{\sigma_Y}{\sigma_X} < 0$.

Converse part is also true i.e., if $Y = bX + a$, then $\rho(X,Y)$ is either 1 or -1 depending on the sign of b.

The correlation coefficient is a measure of the degree of linearity between X and Y. A value of $\rho(X,Y)$ near 1 or -1 indicates a high degree of linearity between X and Y, whereas a value near 0 indicates that such linearity is absent. A positive value of $\rho(X,Y)$ indicates that Y tends to increase when X does, whereas a negative value indicates that Y tends to decrease when X increases.

Example 4.21. *Let X and Y be random variables with mean 0, variance 1 and correlation coefficient ρ. Show that*

$$\mathbb{E}\big[\max\{X^2, Y^2\}\big] \le 1 + \sqrt{1-\rho^2}.$$

Solution: *We know that for any $u, v \in \mathbb{R}$, $\max\{u,v\} = \frac{1}{2}\big(u + v + |u-v|\big)$. Thus, by using Cauchy-Schwartz inequality along with the given condition, we have*

$$\begin{aligned}\mathbb{E}\big[\max\{X^2,Y^2\}\big] &= \frac{1}{2}\Big(\mathbb{E}[X^2] + \mathbb{E}[Y^2] + \mathbb{E}\big[|(X-Y)(X+Y)|\big]\Big)\\ &\le 1 + \frac{1}{2}\sqrt{\mathbb{E}[(X-Y)^2]\mathbb{E}[(X-Y)^2]} = 1 + \frac{1}{2}\sqrt{(2+2\mathbb{E}[XY])(2-2\mathbb{E}[XY])}\\ &= 1 + \sqrt{1 - \big(\mathbb{E}[XY]\big)^2}\,.\end{aligned}$$

Note that, since mean is zero and variance is 1 for both X and Y, one has

$$\mathbb{E}[XY] = \mathrm{Cov}(X,Y) = \rho(X,Y) = \rho.$$

Thus the assertion follows.

4.6.2. **Conditional variance.**

Definition 4.12 (Conditional variance)**.** Let X and Y be real-valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[X^2]$ is finite. The conditional variance of X given Y, denoted as $\mathrm{Var}(X|Y)$, is defined by

$$\mathrm{Var}(X|Y) := \mathbb{E}\Big[\big(X - \mathbb{E}[X|Y]\big)^2|Y\Big].$$

Note that $\mathrm{Var}(X|Y)$ is a random variable which takes the value $\mathrm{Var}(X|Y=y)$. In view of the properties of conditional expectation, we see that

$$\mathrm{Var}(X|Y) = \mathbb{E}[X^2|Y] - \big(\mathbb{E}[X|Y]\big)^2\,.$$

There is a very useful relationship between $\mathrm{Var}(X)$, and $\mathrm{Var}(X|Y)$ which can often be applied to compute $\mathrm{Var}(X)$.

Lemma 4.10. *Let X and Y be real-valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[X^2]$ is finite. Then*

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}\big(\mathbb{E}[X|Y]\big).$$

Proof. Since $\mathrm{Var}(X|Y) = \mathbb{E}[X^2|Y] - \big(\mathbb{E}[X|Y]\big)^2$, one has, by using properties of conditional expectation

$$\mathbb{E}[\mathrm{Var}(X|Y)] = \mathbb{E}[X^2] - \mathbb{E}\Big[\big(\mathbb{E}[X|Y]\big)^2\Big]. \tag{4.3}$$

On the other hand,

$$\mathrm{Var}\big(\mathbb{E}[X|Y]\big) = \mathbb{E}\Big[\big(\mathbb{E}[X|Y]\big)^2\Big] - \Big(\mathbb{E}\big[\mathbb{E}[X|Y]\big]\Big)^2 = \mathbb{E}\Big[\big(\mathbb{E}[X|Y]\big)^2\Big] - \big(\mathbb{E}[X]\big)^2. \tag{4.4}$$

Hence the assertion follows from (4.3) and (4.4). □

Example 4.22. *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and let N be a nonnegative integer-valued random variable that is independent of the sequence $X_i,\ i \ge 1$. Compute $\mathrm{Var}\Big(\sum_{i=1}^N X_i\Big)$.*

Solution: *We have already seen that $\mathbb{E}\Big[\sum_{i=1}^N X_i|N\Big] = N\mathbb{E}[X]$ where $\mathbb{E}[X_i] = \mathbb{E}[X]$. For given N, $\sum_{i=1}^N X_i$ is just the sum of fixed number of independent random variables and hence $\mathrm{Var}\Big(\sum_{i=1}^N X_i|N\Big) = N\mathrm{Var}(X)$, where $\mathrm{Var}(X) = \mathrm{Var}(X_i)$. Hence, from the conditional variance formula, we have*

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^N X_i\Big) &= \mathbb{E}\Big[\mathrm{Var}\Big(\sum_{i=1}^N X_i|N\Big)\Big] + \mathrm{Var}\Big(\mathbb{E}\Big[\sum_{i=1}^N X_i|N\Big]\Big) \\
&= \mathbb{E}[N\mathrm{Var}(X)] + \mathrm{Var}\big(N\mathbb{E}[X]\big) \\
&= \mathbb{E}[N]\mathrm{Var}(X) + \big(\mathbb{E}[X]\big)^2\mathrm{Var}(N).
\end{aligned}$$