

# Probability

<b>1</b>	<b>Axioms of Probability</b>	<b>2</b>
1.1	Sample Space and Events . . . . .	2
1.2	Axioms of Probability . . . . .	4
1.3	Sample Spaces with Equally Likely Outcomes . . . . .	6
<b>2</b>	<b>Conditional Probability and Independence</b>	<b>7</b>
2.1	Conditional Probability . . . . .	7
2.2	Bayes' Formula . . . . .	8
2.3	Independence . . . . .	9
<b>3</b>	<b>Discrete Random Variables</b>	<b>10</b>
3.1	Discrete Random Variables . . . . .	10
3.2	Expectation and Variance . . . . .	11
3.3	Bernoulli and Binomial Random Variables . . . . .	12
3.4	Poisson Random Variables . . . . .	14
3.5	Geometric and Negative Binomial Random Variables . . . . .	16
3.6	Hypergeometric Random Variables . . . . .	17
<b>4</b>	<b>Continuous Random Variables</b>	<b>18</b>
4.1	Continuous Random Variables . . . . .	18
4.2	Expectation and Variance . . . . .	19
4.3	Uniform Random Variables . . . . .	19
4.4	Normal Random Variables . . . . .	20
<b>5</b>	<b>Jointly Distributed Random Variables</b>	<b>21</b>
5.1	Jointly Distributed Random Variables . . . . .	21
5.2	Independent Random Variables . . . . .	23
5.3	Sums of Independent Random Variables . . . . .	24
<b>6</b>	<b>Properties of Expectation</b>	<b>25</b>
6.1	Covariance and Correlation . . . . .	25
6.2	Moment Generating Functions . . . . .	26

# Chapter 1

## Axioms of Probability

### 1.1 Sample Space and Events

**Definition 1.1.** The set of all possible outcomes of an experiment is called the **sample space** of the experiment and is denoted by  $\Omega$ .

**Example.** If the experiment consists of tossing two dice, then the sample space is

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

**Definition 1.2.** Let  $\Omega$  be a sample space of an experiment. A family  $\Sigma$  of subsets of  $\Omega$  is called a  **$\sigma$ -algebra** on  $\Omega$  if the following conditions hold.

- (a)  $\Omega \in \Sigma$ .
- (b) For all  $E \in \Sigma$ ,  $\Omega \setminus E \in \Sigma$ .
- (c) If  $E_1, E_2, \dots$  is a sequence of elements in  $\Sigma$ , then

$$\bigcup_{i=1}^{\infty} E_i \in \Sigma.$$

**Definition 1.3.** If  $\Sigma$  is a  $\sigma$ -algebra on  $\Omega$ , then  $(\Omega, \Sigma)$  is called a **measurable space**, and a subset of  $\Omega$  that belongs to  $\Sigma$  is called an **event**.

**Theorem 1.4.** Let  $I$  be an index set such that for each  $i \in I$ ,  $\Sigma_i$  is a  $\sigma$ -algebra on  $\Omega$ . Then

$$\Sigma^* = \bigcap_{i \in I} \Sigma_i$$

is also a  $\sigma$ -algebra on  $\Omega$ .

*Proof.*

- (a) Since  $\Omega \in \Sigma_i$  for each  $i \in I$ , it follows that  $\Omega \in \Sigma$ .
- (b) We have

$$\begin{aligned} E \in \Sigma &\Rightarrow E \in \Sigma_i \text{ for each } i \in I \\ &\Rightarrow \Omega \setminus E \in \Sigma_i \text{ for each } i \in I \\ &\Rightarrow \Omega \setminus E \in \Sigma. \end{aligned}$$

(c) We have

$$\begin{aligned} E_1, E_2, \dots \in \Sigma &\Rightarrow E_1, \dots, E_2 \in \Sigma_i \text{ for each } i \in I \\ &\Rightarrow \bigcup_{j=1}^{\infty} E_j \in \Sigma_i \text{ for each } i \in I \\ &\Rightarrow \bigcup_{j=1}^{\infty} E_j \in \Sigma. \quad \square \end{aligned}$$

**Definition 1.5.** Let  $\Phi$  be a family of subsets of  $\Omega$ . Then the  **$\sigma$ -algebra generated by  $\Phi$** , denoted by  $\sigma(\Phi)$ , is the intersection of all  $\sigma$ -algebras that contains  $\Phi$ .

**Example.** Let  $\Phi$  be the collection of all open intervals on  $\mathbb{R}$ . Then the  $\sigma$ -algebra generated by  $\Phi$  is called the **Borel algebra** of  $\mathbb{R}$ , denoted by  $\mathcal{B}$ .

## 1.2 Axioms of Probability

**Definition 1.6.** Two events  $E$  and  $F$  are **mutually exclusive** if  $E \cap F = \emptyset$ .

**Definition 1.7.** Let  $(\Omega, \Sigma)$  be a measurable space. A function  $P : \Sigma \rightarrow \mathbb{R}$  is called a **probability function** and  $(\Omega, \Sigma, P)$  is a **probability space** if the following conditions hold.

- (a) For all  $E \in \Sigma$ ,  $P(E) \geq 0$ .
- (b)  $P(\Omega) = 1$ .
- (c) If  $E_1, E_2, \dots$  is a sequence of events that are pairwise mutually exclusive, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

**Theorem 1.8.** Let  $(\Omega, \Sigma, P)$  be a probability space. Let  $E, F \in \Sigma$ . Then  $P(F \setminus E) = P(F) - P(E \cap F)$ .

*Proof.* Since  $E \cap F$  and  $F \setminus E$  are mutually exclusive, we have

$$P(F) = P((E \cap F) \cup (F \setminus E)) = P(E \cap F) + P(F \setminus E).$$

Thus,  $P(F \setminus E) = P(F) - P(E \cap F)$ . □

**Corollary 1.9.**  $P(\Omega \setminus E) = 1 - P(E)$  holds for any event  $E$ , implying  $P(\emptyset) = 0$ .

**Corollary 1.10.** If  $E \subseteq F$ , then  $P(E) \leq P(F)$  because  $P(F) - P(E) = P(F \setminus E) \geq 0$ .

**Theorem 1.11 (Inclusive-exclusive Principle).** Let  $(\Omega, \Sigma, P)$  be a probability space. If  $E_1, \dots, E_n \in \Sigma$ , then

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P(E_{i_1} \cap \dots \cap E_{i_r}).$$

*Proof.* The proof is by induction on  $n$ . The theorem holds for  $n = 0$  and  $n = 1$  trivially. For  $n = 2$ , since  $E_1 \cap E_2$  and  $E_1 \setminus E_2$  are mutually exclusive, we have

$$P(E_1) = P((E_1 \cap E_2) \cup (E_1 \setminus E_2)) = P(E_1 \cap E_2) + P(E_1 \setminus E_2).$$

Thus, since  $E_1 \setminus E_2$  and  $E_2$  are mutually exclusive, we have

$$\begin{aligned} P(E_1 \cup E_2) &= P((E_1 \setminus E_2) \cup E_2) \\ &= P(E_1 \setminus E_2) + P(E_2) \\ &= P(E_1) - P(E_1 \cap E_2) + P(E_2). \end{aligned}$$

Now suppose that the theorem holds for some  $n \geq 2$ , and we prove that the theorem is true for  $n + 1$ . Since  $E_1 \cup \dots \cup E_n$  and  $E_{n+1}$  are mutually exclusive, we have

$$P(E_1 \cup \dots \cup E_n \cup E_{n+1}) = P(E_1 \cup \dots \cup E_n) + P(E_{n+1}) - P((E_1 \cup \dots \cup E_n) \cap E_{n+1}),$$

where the first term can be written as

$$P(E_1 \cup \dots \cup E_n) = \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P(E_{i_1} \cap \dots \cap E_{i_r}),$$

and the last term can be written as

$$\begin{aligned} & P((E_1 \cup \dots \cup E_n) \cap E_{n+1}) \\ &= P((E_1 \cap E_{n+1}) \cup \dots \cup (E_n \cap E_{n+1})) \\ &= \sum_{s=1}^n (-1)^{s+1} \sum_{1 \leq i_1 < \dots < i_s \leq n} P((E_{i_1} \cap E_{n+1}) \cap \dots \cap (E_{i_s} \cap E_{n+1})) \\ &= \sum_{s=1}^n (-1)^{s+1} \sum_{1 \leq i_1 < \dots < i_s \leq n} P(E_{i_1} \cap \dots \cap E_{i_s} \cap E_{n+1}) \\ &= - \sum_{r=2}^{n+1} (-1)^{r+1} \sum_{\substack{1 \leq i_1 < \dots < i_{r-1} \leq n \\ i_r = n+1}} P(E_{i_1} \cap \dots \cap E_{i_{r-1}} \cap E_{i_r}). \end{aligned}$$

Now we consider  $r$ , which is the number of sets in each intersection. The second term is actually the case with  $r = 1$ , and the last term consists of the cases with  $r \geq 2$ . Thus,

$$P(E_{n+1}) - P((E_1 \cup \dots \cup E_n) \cap E_{n+1}) = \sum_{r=1}^n (-1)^{r+1} \sum_{\substack{1 \leq i_1 < \dots < i_r \leq n \\ i_r = n+1}} P(E_{i_1} \cap \dots \cap E_{i_r}).$$

Furthermore, note that the first term consists of the case where  $E_{n+1}$  does not appear in the intersection, while the difference above consists of the case where  $E_{n+1}$  appears in the intersection. Thus, by summing up all terms, we have

$$P(E_1 \cup \dots \cup E_n \cup E_{n+1}) = \sum_{r=1}^{n+1} (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n+1} P(E_{i_1} \cap \dots \cap E_{i_r}),$$

which completes the proof.  $\square$

**Example.** For any three events  $E_1, E_2, E_3$ , we have  $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_3 \cap E_1) + P(E_1 \cap E_2 \cap E_3)$ .

### 1.3 Sample Spaces with Equally Likely Outcomes

**Theorem 1.12.** Let  $\Omega = \{\omega_1, \dots, \omega_n\}$  be a finite sample space and let  $P$  be a probability function such that  $P(\{\omega_i\}) = P(\{\omega_j\})$  for  $i, j \in \{1, \dots, n\}$ . Then for each event  $E \subseteq \Omega$  with  $|E| = k$ , we have

$$P(E) = \frac{k}{n}.$$

*Proof.* Let  $p$  denote the probability of each elementary event  $\{\omega_i\}$  for all  $i \in \{1, \dots, n\}$ . Then we have

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \{\omega_i\}\right) = \sum_{i=1}^n P(\{\omega_i\}) = np.$$

Thus,

$$p = \frac{1}{n}.$$

Let  $E = \{\omega_{i_1}, \dots, \omega_{i_k}\}$ . Then

$$P(E) = P\left(\bigcup_{r=1}^k \{\omega_{i_r}\}\right) = \sum_{r=1}^k P(\{\omega_{i_r}\}) = \frac{k}{n}. \quad \square$$

# Chapter 2

## Conditional Probability and Independence

### 2.1 Conditional Probability

**Definition 2.1.** Let  $(\Omega, \Sigma, P)$  be a probability space. If  $E$  is an event with  $P(E) > 0$ , then define

$$P(F | E) = \frac{P(E \cap F)}{P(E)}$$

for any event  $F$ .

**Theorem 2.2.** Let  $(\Omega, \Sigma, P)$  be a probability space. If  $E$  is an event with  $P(E) > 0$ , then the function  $P_E : \Sigma \rightarrow \mathbb{R}$  is a probability function if

$$P_E(F) = P(F | E)$$

for any event  $F$ .

*Proof.* For events  $E$  and  $F$ ,

$$P_E(F) = \frac{P(E \cap F)}{P(E)} \geq 0.$$

Moreover,

$$P_E(\Omega) = \frac{P(E \cap \Omega)}{P(E)} = \frac{P(E)}{P(E)} = 1.$$

If  $F_1, F_2, \dots$  is a sequence of events that are piecewise mutually exclusive, then

$$P_E\left(\bigcup_{i=1}^{\infty} F_i\right) = \frac{P\left(E \cap \bigcup_{i=1}^{\infty} F_i\right)}{P(E)} = \frac{P\left(\bigcup_{i=1}^{\infty} (E \cap F_i)\right)}{P(E)} = \sum_{i=1}^{\infty} \frac{P(E \cap F_i)}{P(E)} = \sum_{i=1}^{\infty} P_E(F_i).$$

Thus,  $P_E$  is a probability function. □

## 2.2 Bayes' Formula

**Definition 2.3.** A **partition** of  $\Omega$  is a family of nonempty events such that each element in  $\Omega$  is in exactly one of these events.

**Theorem 2.4.** Let  $E_1, \dots, E_n$  form a partition of  $\Omega$  such that  $P(E_i) > 0$  for each  $i \in \{1, \dots, n\}$ . Then for any event  $F$ ,

$$P(F) = \sum_{i=1}^n P(F \mid E_i)P(E_i).$$

*Proof.* Since

$$F = F \cap \Omega = F \cap \bigcup_{i=1}^n E_i = \bigcup_{i=1}^n (F \cap E_i),$$

it follows that

$$P(F) = P\left(\bigcup_{i=1}^n (F \cap E_i)\right) = \sum_{i=1}^n P(F \cap E_i) = \sum_{i=1}^n P(F \mid E_i)P(E_i). \quad \square$$

**Theorem 2.5 (Bayes' Formula).** Let  $E_1, \dots, E_n$  form a partition of  $\Omega$  such that  $P(E_j) > 0$  for each  $j \in \{1, \dots, n\}$ . Then for any event  $F$  with  $P(F) > 0$ , for any  $i \in \{1, \dots, n\}$ , we have

$$P(E_i \mid F) = \frac{P(F \mid E_i)P(E_i)}{\sum_{j=1}^n P(F \mid E_j)P(E_j)}.$$

*Proof.* By Theorem 2.4, we have

$$P(E_i \mid F) = \frac{P(E_i \cap F)}{P(F)} = \frac{P(F \mid E_i)P(E_i)}{\sum_{j=1}^n P(F \mid E_j)P(E_j)}. \quad \square$$



## 2.3 Independence

**Definition 2.6.** Let  $E_1, \dots, E_n$  be events in a probability space  $(\Omega, \Sigma, P)$ .

- $E_1, \dots, E_n$  are **independent** if

$$P\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} P(E_i)$$

holds for any nonempty subset  $I$  of  $\{1, \dots, n\}$ .

- $E_1, \dots, E_n$  are **dependent** if they are not independent.

**Definition 2.7.** Let  $E_1, \dots, E_n$  and  $F$  be events in a probability space  $(\Omega, \Sigma, P)$ , where  $P(F) > 0$ . Then  $E_1, \dots, E_n$  are **independent given event  $F$**  if

$$P\left(\bigcap_{i \in I} E_i \mid F\right) = \prod_{i \in I} P(E_i \mid F)$$

holds for any nonempty subset  $I$  of  $\{1, \dots, n\}$ .

# Chapter 3

## Discrete Random Variables

### 3.1 Discrete Random Variables

**Definition 3.1.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a function in a probability space  $(\Omega, \Sigma, P)$ . Then  $X$  is called a **random variable** if  $X^{-1}(S) \in \Sigma$  for all  $S \in \mathcal{B}$ , where

$$X^{-1}(S) = \{\omega \in \Omega : X(\omega) \in S\}.$$

**Remark.** Since each  $S \in \mathcal{B}$  is mapped to a event  $X^{-1}(S) \in \Sigma$ , we will use conditions related to random variables to denote events. For example,

$$P(-1 \leq X \leq 1) = P(\{\omega \in \Omega : -1 \leq X(\omega) \leq 1\}).$$

**Definition 3.2.** A random variable  $X$  is a discrete random variable if there is a countable set  $S \subseteq \mathbb{R}$  such that  $P(X \in S) = 1$ .

**Definition 3.3.** Let  $X$  be a random variable in a probability space  $(\Omega, \Sigma, P)$ . The **probability mass function**  $p_X : \mathbb{R} \rightarrow \mathbb{R}$  of  $X$  is defined as

$$p_X(x) = P(X = x)$$

for each  $x \in \mathbb{R}$ .

## 3.2 Expectation and Variance

**Definition 3.4.** Let  $X$  be a discrete random variable in a probability space  $(\Omega, \Sigma, P)$ . Then the **expectation** of  $X$ , denoted by  $E[X]$ , is defined as follows.

- If  $X$  is nonnegative, i.e,  $X(\omega) \geq 0$  for each  $\omega \in \Omega$ , then

$$E[X] = \sum_{x \in X(\Omega)} x \cdot p_X(x).$$

- Otherwise, we define  $E[X] = E[X^+] - E[X^-]$ , where  $X^+ = \max\{X, 0\}$  and  $X^- = \max\{-X, 0\}$ .

**Theorem 3.5.** Let  $X$  and  $Y$  be discrete random variables in a probability space  $(\Omega, \Sigma, P)$ . If both  $E[X]$  and  $E[Y]$  exist, then the following statements are true.

- (a)  $E[aX] = aE[X]$  for  $a \in \mathbb{R}$ .
- (b)  $E[X + Y] = E[X] + E[Y]$ .

*Proof.*

- (a) First, suppose that  $a \geq 0$ . If  $X$  is nonnegative, then so is  $aX$ . Thus, we have

$$E[aX] = \sum_{x \in X(\Omega)} ax \cdot p_X(x) = aE[X].$$

If  $X$  is not nonnegative, by the fact that  $(aX)^+ = aX^+$  and  $(aX)^- = aX^-$ , we have

$$E[aX] = E[aX^+] - E[aX^-] = aE[X^+] - aE[X^-] = aE[X].$$

since  $X^+$  and  $X^-$  are nonnegative. Thus the statement holds for  $a \geq 0$ .

For the case that  $a < 0$ , note that since  $(-X)^+ = X^-$  and  $(-X)^- = X^+$ , it follows that

$$E[-X] = E[X^-] - E[X^+] = -E[X].$$

Thus, we have

$$E[aX] = E[(-a)(-X)] = -aE[-X] = aE[X].$$

- (b) To be completed. □

**Definition 3.6.** Let  $X$  be a discrete random variable in a probability space  $(\Omega, \Sigma, P)$ . Then the **variance** of  $X$  is defined as

$$\text{Var}(X) = E[(X - E[X])^2].$$

**Theorem 3.7.** Let  $X$  be a discrete random variable. Then  $\text{Var}(X) = E[X^2] - (E[X])^2$ .

*Proof.* It is proved by

$$\begin{aligned} \text{Var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2X \cdot E[X] + (E[X])^2] \\ &= E[X^2] - 2E[X] \cdot E[X] + (E[X])^2 \\ &= E[X^2] - (E[X])^2. \end{aligned}$$

□

### 3.3 Bernoulli and Binomial Random Variables

**Definition 3.8.** Let  $0 \leq p \leq 1$ . A random variable  $X$  is called a **Bernoulli random variable** with parameter  $p$  if  $p_X(1) = p$  and  $p_X(0) = 1 - p$ .

**Theorem 3.9.** Let  $X$  be a Bernoulli random variable with parameter  $p$ .

(a)  $E[X] = p$ .

(b)  $\text{Var}(X) = p(1 - p)$ .

*Proof.* Since  $p(0) + p(1) = 1$ , we have  $p(x) = 0$  for  $x \notin \{0, 1\}$ .

(a) We have

$$E[X] = \sum_{x: p_X(x) > 0} x \cdot p_X(x) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

(b) By (a), we have

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= (1^2 \cdot p + 0^2 \cdot (1 - p)) - p^2 \\ &= p - p^2 \\ &= p(1 - p). \end{aligned}$$

□

**Definition 3.10.** Let  $n$  be a nonnegative integer and  $0 \leq p \leq 1$ . A random variable  $X$  is called a **binomial random variable** with parameter  $(n, p)$ , if

$$p_X(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for each  $x \in \{0, \dots, n\}$ .

**Theorem 3.11.** Let  $X$  be a binomial random variable with parameter  $(n, p)$ .

(a)  $E[X] = np$ .

(b)  $\text{Var}(X) = np(1 - p)$ .

*Proof.* We have  $p(x) = 0$  for  $x \notin \{0, \dots, n\}$  because

$$\sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{n-x} = (p + (1 - p))^n = 1.$$

Also, we have the fact that

$$\begin{aligned} E[X^k] &= \sum_{x=0}^n x^k \binom{n}{x} p^x (1 - p)^{n-x} \\ &= \sum_{x=1}^n x^k \binom{n}{x} p^x (1 - p)^{n-x} \\ &= np \sum_{x=1}^n x^{k-1} \binom{n-1}{x-1} p^{x-1} (1 - p)^{n-x} \\ &= np \sum_{y=0}^{n-1} (y+1)^{k-1} \binom{n-1}{y} p^y (1 - p)^{(n-1)-y} \end{aligned} \tag{*}$$

holds for positive integer  $k$ .

(a) By (\*), the expectation of  $X$  is given by

$$E[X] = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} = np.$$

(b) By (\*), we have

$$E[X^2] = np \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^y (1-p)^{(n-1)-y} = np((n-1)p + 1).$$

Thus, the variance of  $X$  is given by

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= np((n-1)p + 1) - (np)^2 \\ &= np(1-p). \end{aligned}$$

□

### 3.4 Poisson Random Variables

**Theorem 3.12.** Let  $\lambda > 0$ . For integer  $n \geq \lambda$ , let  $X_n$  be a binomial random variable with parameter  $(n, \lambda/n)$ . Then for nonnegative integer  $x$ , we have

$$\lim_{n \rightarrow \infty} p_{X_n}(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}.$$

*Proof.* For  $n \geq \lambda$ , we have

$$\begin{aligned} p_{X_n}(x) &= \frac{n!}{(n-x)! \cdot x!} \cdot \left(\frac{\lambda}{n}\right)^x \left(\frac{n-\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \cdot \frac{n!}{(n-x)! \cdot (n-\lambda)^x} \cdot \left(\frac{n-\lambda}{n}\right)^n \\ &= \frac{\lambda^x}{x!} \cdot \prod_{i=1}^x \frac{n+1-i}{n-\lambda} \cdot \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

Thus,

$$\lim_{n \rightarrow \infty} p_{X_n}(x) = \frac{\lambda^x}{x!} \cdot \prod_{i=1}^x \left( \lim_{n \rightarrow \infty} \frac{n+1-i}{n-\lambda} \right) \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^x}{x!} \cdot e^{-\lambda}. \quad \square$$

**Definition 3.13.** Let  $\lambda > 0$ . A random variable  $X$  is called a **Poisson random variable** with parameter  $\lambda$ , if

$$p_X(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

holds for any nonnegative integer  $x$ .

**Theorem 3.14.** Let  $X$  be a Poisson random variable with parameter  $\lambda$ .

- (a)  $E[X] = \lambda$ .
- (b)  $\text{Var}(X) = \lambda$ .

*Proof.* We have  $p_X(x) = 0$  for  $x \notin \{0, 1, 2, \dots\}$  because

$$\sum_{x=0}^{\infty} p_X(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

where the second equality follows from the fact that  $e^t = \sum_{k=0}^{\infty} t^k/k!$  for  $t \in \mathbb{R}$ .

- (a) We have

$$E[X] = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda.$$

(b) Since

$$\begin{aligned} E[X^2] &= \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} x^2 \cdot \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{y=0}^{\infty} (y+1) \cdot \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \lambda(\lambda+1), \end{aligned}$$

we have

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.$$

□

### 3.5 Geometric and Negative Binomial Random Variables

**Definition 3.15.** Let  $0 \leq p \leq 1$ . A random variable  $X$  is called a **geometric random variable** with parameter  $p$ , if

$$p_X(x) = p \cdot (1 - p)^{x-1}$$

holds for any positive integer  $x$ .

**Definition 3.16.** Let  $r$  be a nonnegative integer and  $0 \leq p \leq 1$ . A random variable  $X$  is called a **negative binomial random variable** with parameter  $(r, p)$ , if

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

holds for any integer  $x \geq r$ .

**Theorem 3.17.** Let  $X$  be a negative binomial random variable with parameter  $(r, p)$ .

- (a)  $E[X] = r/p$ .
- (b)  $\text{Var}(X) = r(1-p)/p^2$ .

*Proof.* We have  $p_X(x) = 0$  for  $x \notin \{r, r+1, r+2, \dots\}$  because

$$\begin{aligned} \sum_{x=r}^{\infty} p_X(x) &= \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= \sum_{x=r}^{\infty} \binom{-r}{x-r} p^r (-(1-p))^{x-r} \\ &= p^r \sum_{y=0}^{\infty} \binom{-r}{y} (-(1-p))^y \\ &= p^r (1 - (1-p))^{-r} \\ &= 1, \end{aligned}$$

where the second equality follows from

$$\binom{x-1}{r-1} = \binom{x-1}{x-r} = \binom{-r}{x-r} \cdot (-1)^{x-r}.$$

- (a) To be completed.
- (b) To be completed. □



### 3.6 Hypergeometric Random Variables

**Definition 3.18.** Let  $n, K, N$  be nonnegative integers with  $n \leq N$  and  $K \leq N$ . A random variable  $X$  is called a **hypergeometric random variable** with parameter  $(n, K, N)$  if

$$p_X(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$$

holds for any integer  $x \in \{0, 1, \dots, K\}$ .

**Remark.** A hypergeometric random variable with parameter  $(1, K, N)$  is a Bernoulli random variable with parameter  $K/N$ .

**Remark.** A hypergeometric random variable  $X$  with parameter  $(n, K, N)$  can be seen as the number of successes in  $n$  draws without replacement from a population of size  $N$  that contains  $K$  objects that represent success.

**Remark.** If  $N$  and  $K$  are large compared to  $n$ , then a hypergeometric random variable  $X$  with parameter  $(n, K, N)$  behaves like a binomial random variable with parameter  $(n, K/N)$ .

# Chapter 4

## Continuous Random Variables

### 4.1 Continuous Random Variables

**Definition 4.1.** A random variable  $X$  is a **continuous random variable** if there exists a nonnegative function  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$P(X \in S) = \int_S f_X(x) dx$$

holds for any  $S \in \mathcal{B}$ . The function  $f_X$  is called a **probability density function** of  $X$ .

**Theorem 4.2.** Let  $X$  be a continuous random variable in a probability space  $(\Omega, \Sigma, P)$ . Then  $p_X(a) = 0$  for  $a \in \mathbb{R}$ .

*Proof.* It is proved by

$$p_X(a) = P(X = a) = \int_a^a f_X(x) dx = 0. \quad \square$$

**Definition 4.3.** The **cumulative distribution function**  $F_X$  of a random variable  $X$  is defined by

$$F_X(x) = P(X \leq x)$$

for all  $x \in \mathbb{R}$ .

**Theorem 4.4.** Let  $X$  be a continuous random variable in a probability space  $(\Omega, \Sigma, P)$ . If  $f_X$  is a probability density function of  $X$ , then

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

holds for all  $x \in \mathbb{R}$ .

*Proof.* For all  $x \in \mathbb{R}$ , we have

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt. \quad \square$$

## 4.2 Expectation and Variance

## 4.3 Uniform Random Variables

**Definition 4.5.** Let  $a, b$  be real numbers with  $a < b$ . A continuous random variable  $X$  is called a **uniform random variable** with parameters  $(a, b)$  if the function  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  with

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

for  $x \in \mathbb{R}$  is a probability density function of  $X$ .

## 4.4 Normal Random Variables

**Definition 4.6.** Let  $\mu, \sigma$  be real numbers with  $\sigma \geq 0$ . A continuous random variable  $X$  is called a **normal random variable** with parameters  $(\mu, \sigma^2)$  if the function  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  with

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

for  $x \in \mathbb{R}$  is a probability density function of  $X$ .

# Chapter 5

## Jointly Distributed Random Variables

### 5.1 Jointly Distributed Random Variables

**Definition 5.1.** Let  $X$  and  $Y$  be random variables on a probability space  $(\Omega, \Sigma, P)$ .

- The function  $p_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$  with

$$p_{X,Y}(x, y) = P(X = x \text{ and } Y = y)$$

for  $x, y \in \mathbb{R}$  is the **joint probability mass function** of  $X$  and  $Y$ .

- If there exists a nonnegative function  $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that

$$\iint_S f_{X,Y}(x, y) dx dy = P((X, Y) \in S)$$

holds for all  $S \in \mathcal{B}^2$ , then  $f_{X,Y}$  is a **joint probability density function** of  $X$  and  $Y$ .

- The function  $F_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$  with

$$F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$$

for  $x, y \in \mathbb{R}$  is the **joint cumulative distribution function** of  $X$  and  $Y$ .

**Definition 5.2.** Let  $X_1, X_2, \dots, X_n$  be random variables on a probability space  $(\Omega, \Sigma, P)$ .

- The function  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  with

$$p(x_1, \dots, x_n) = P(X_i = x_i \text{ for } i \in \{1, \dots, n\})$$

for  $x_1, \dots, x_n \in \mathbb{R}$  is the **joint probability mass function** of  $X_1, \dots, X_n$ .

- If there exists a nonnegative function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$\int \cdots \int_S f(x_1, \dots, x_n) dx_1 \cdots dx_n = P((X_1, \dots, X_n) \in S)$$

holds for all  $S \in \mathcal{B}^n$ , then  $f$  is a **joint probability density function** of  $X_1, \dots, X_n$ .

- The function  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  with

$$F(x_1, \dots, x_n) = P(X_i \leq x_i \text{ for } i \in \{1, \dots, n\})$$

for  $x_1, \dots, x_n \in \mathbb{R}$  is the **joint cumulative distribution function** of  $X_1, \dots, X_n$ .

## 5.2 Independent Random Variables

**Definition 5.3.** Let  $X$  and  $Y$  be random variables on a probability space  $(\Omega, \Sigma, P)$ . Then  $X$  and  $Y$  are **independent** if the events  $X \in S_1$  and  $Y \in S_2$  are independent for any  $S_1, S_2 \in \mathcal{B}$ .

**Definition 5.4.** Let  $X_1, X_2, \dots, X_n$  be random variables on a probability space  $(\Omega, \Sigma, P)$ . Then  $X_1, X_2, \dots, X_n$  are **independent** if the events  $X_1 \in S_1, X_2 \in S_2, \dots, X_n \in S_n$  are independent for any  $S_1, S_2, \dots, S_n \in \mathcal{B}$ .

### 5.3 Sums of Independent Random Variables

**Theorem 5.5.** Let  $X$  and  $Y$  be independent continuous random variables and  $Z = X + Y$ . Then  $f_Z : \mathbb{R} \rightarrow \mathbb{R}$  with

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) dx$$

for  $z \in \mathbb{R}$  is a probability density function of  $Z$ , where  $f_X$  and  $f_Y$  are probability density functions of  $X$  and  $Y$ , respectively.

*Proof.* We have

$$\begin{aligned} F_Z(z) &= P(X + Y \leq z) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z f_{X,Y}(x, u - x) du dx && (u = x + y) \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} f_{X,Y}(x, u - x) dx du. \end{aligned}$$

Thus,

$$\frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x) dx = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) dx = f_Z(z),$$

implying  $f_Z$  is a probability density function of  $Z$ . □



# Chapter 6

## Properties of Expectation

### 6.1 Covariance and Correlation

**Definition 6.1.** Let  $X$  and  $Y$  be random variables on a probability space  $(\Omega, \Sigma, P)$ . The **covariance** of  $X$  and  $Y$  is defined by

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

**Definition 6.2.** Let  $X$  and  $Y$  be random variables on a probability space  $(\Omega, \Sigma, P)$ . The **correlation** of  $X$  and  $Y$  is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$

## 6.2 Moment Generating Functions

**Definition 6.3.** Let  $X$  be a random variable. Then the **moment generating function**  $M_X$  of  $X$  is defined by

$$M_X(t) = E[e^{tX}].$$

**Proposition 6.4.** Let  $X$  be a random variable. Then the following statements are true.

- (a)  $M_X(0) = 1$ .
- (b) For each positive integer  $k$ ,  $M_X^{(k)}(0) = E[X^k]$  if  $E[X^k]$  exists.