# Summary for Elementary Probability

SEUNGWOO HAN

# Contents

# Chapter 1

# Basic Concepts

## 1.1 Events and Probability

**Definition 1.1.1: Probability Space**

A *probability space* contains of a triple $(\Omega, \mathcal{F}, P)$ where

- $\Omega$ is the sample space,
- $\mathcal{F} \subseteq 2^{\Omega}$ (each $A \in \mathcal{F}$ is called an *event*), and
- $P : \mathcal{F} \to [0,1]$ maps each event $A \in \mathcal{F}$ to the *probability* of $A$

which satisfies the following conditions:

**Axioms Relative to the Events** The family $\mathcal{F}$ of events must be a $\sigma$-field on $\Omega$:

(1) $\Omega \in \mathcal{F}$;

(2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (where $A^c$ is the complement of $A$);

(3) If $\langle A_n \rangle_{n \in \mathbb{Z}_+}$ is a sequence on $\mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

**Axioms Relative to the Probability** The function $P$ must satisfy the following conditions:

(1) $P(\Omega) = 1$;

(2) $\sigma$-additivity holds: if $\langle A_n \rangle_{n \in \mathbb{Z}_+}$ is a sequence of pairwise disjoint events, then

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

**Note**

Here are immediate properties of probability:

- $P(A^c) = 1 - P(A)$;
- $\varnothing = \Omega^c \in \mathcal{F}$ and $P(\varnothing) = 0$;
- If $\langle A_n \rangle_{n \in \mathbb{Z}_+}$ is a sequence of events, then $\bigcap_{n=1}^{\infty} A_n$ is also an event;
- $A, B \in \mathcal{F}$ and $A \subseteq B$ implies $P(A) \leq P(B)$.

**Lemma 1.1.2** sub-$\sigma$-additivity

If $\langle A_n \rangle_{n\in\mathbb{Z}_+}$ is a sequence of events, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} P(A_n).$$

***Proof.*** Let $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$ for each $n \ge 1$ and use $\sigma$-additivity. □

**Lemma 1.1.3** Inclusion-Exclusion Principle

If $A_1, \cdots, A_n$ are events, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\varnothing \neq I \subseteq [n]} (-1)^{|I|-1} P\left(\bigcap_{i\in I} A_i\right).$$

***Proof.*** Classic. □

**Theorem 1.1.4** Sequential Continuity of Probability

(1) Let $\langle B_n \rangle_{n\in\mathbb{Z}_+}$ be a sequence of events such that $B_n \subseteq B_{n+1}$ for all $n \ge 1$. Then,

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n).$$

(2) Let $\langle C_n \rangle_{n\in\mathbb{Z}_+}$ be a sequence of events such that $C_n \supseteq C_{n+1}$ for all $n \ge 1$. Then,

$$P\left(\bigcap_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} P(C_n).$$

***Proof.***

(1) Let $B_n' := B_n \setminus B_{n-1}$ for each $n \ge 2$ and $B_1' := B_1$. so that $B_m = \bigcup_{n=1}^{m} B_n'$ and $B_i'$'s are pairwise disjoint. Hence, by $\sigma$-additivity, we have

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n'\right) = \sum_{n=1}^{\infty} P(B_n') = P(B_1) + \sum_{n=1}^{\infty} \big(P(B_n) - P(B_{n-1})\big) = \lim_{n\to\infty} P(B_n).$$

(2) Let $C_n' := C_n{}^{\mathsf{c}}$ for each $n \ge 1$ so that $C_n' \subseteq C_{n+1}'$ for all $n$. Hence, by (1), we have $P\left(\bigcup_{n=1}^{\infty} C_n'\right) = \lim_{n\to\infty} P(C_n')$. The result follows from the fact that $\bigcup_{n=1}^{\infty} C_n' = \Omega \setminus \bigcap_{n=1}^{\infty} C_n$. □

## 1.2 Random Variables and Their Distributions

**Definition 1.2.1: Random Variable**

A *random variable* on $(\Omega, \mathcal{F})$ is any mapping $X : \Omega \to \overline{\mathbb{R}}$ such that for all $a \in \mathbb{R}$, $\{X \le a\} \triangleq \{\omega \in \Omega \mid X(\omega) \le a\} \in \mathcal{F}$. Here, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

- If $X$ only takes finite values, $X$ is called a *real random variable*.
- If $X$ only takes only a countable set of values $\{a_n\}_{n\in\mathbb{Z}_{\ge 0}}$, $X$ is called a *discrete random variable*.

**Definition 1.2.2: Cumulative Distribution Function**

The *cumulative distribution function* (CDF) of a random variable $X$ is the function $F : \mathbb{R} \to [0,1]$ defined by

$$F(x) = P(X \leq x) \triangleq P(\{X \leq x\}).$$

**Lemma 1.2.3**

Let $F$ be a cumulative distribution function of a random variable $X$.

(1) $F$ is monotone increasing.

(2) $F$ is right-continuous.

(3) If we define $F(\infty) := \lim_{x\to\infty} F(x)$ and $F(-\infty) = \lim_{x\to-\infty} F(x)$, then $1 - F(\infty) = P(X = \infty)$ and $F(-\infty) = P(X = -\infty)$.

***Proof.***

(1) Take any $x, y \in \mathbb{R}$ with $x \leq y$. Then, $\{X \leq x\} \subseteq \{X \leq y\}$. Hence, $F(x) = P(X \leq x) \leq P(X \leq y) \leq F(y)$.

(2) Take any decreasing nonnegative sequence $\langle \varepsilon_n \rangle_{n\in\mathbb{Z}_+}$ of real numbers converging to zero and a real number $x$. Let $C_n := \{X \leq x + \varepsilon_n\}$ so that $\langle C_n \rangle_{n\in\mathbb{Z}_+}$ is a decreasing sequence of events. Note also that $\{X \leq x\} = \bigcap_{n=1}^{\infty} C_n$ Then, by Theorem 1.1.4 (2),

$$F(x) = P(X \leq x) = \lim_{n\to\infty} P(X \leq x + \varepsilon_n) = \lim_{n\to\infty} F(x + \varepsilon_n).$$

(3) Let $B_n := \{X \leq n\}$ for each $n \in \mathbb{Z}_+$ so that $\bigcup_{n=1}^{\infty} B_n = \{X < \infty\}$ and $\langle B_n \rangle_{n\in\mathbb{Z}_+}$ is an increasing sequence of events. By Theorem 1.1.4 (1),

$$1 - P(X = \infty) = P(X < \infty) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} F(n) = F(\infty).$$

The last equality is due to (1). □

**Definition 1.2.4: Probability Density**

If a real random variable $X$ admits a cumulative distribution function $F$ such that

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

for some nonnegative function $f$, then $X$ is said to admit the *probability density* $f$.

**Note**

Note that the probability density $f$ satisfies

$$\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y = 1.$$

## 1.3 Conditional Probability and Independence

**Definition 1.3.1: Conditional Probability**

Let $B$ be an event with $P(B) > 0$. For any event $A$, we define

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

and it is called the *probability of $A$ given $B$*.

**Definition 1.3.2: Independent Events**

(1) Two events $A$ and $B$ are said to be *indepenent* if $P(A \cap B) = P(A)P(B)$.

(2) Let $\mathcal{A}$ be a nonempty family of events. $\mathcal{A}$ is said to be a *family of independent events* if for any finite subfamily $\langle A_1, \cdots, A_n \rangle$ of $\mathcal{A}$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

**Note**

When $P(B) > 0$, $A$ and $B$ are indepenent if and only if $P(A \mid B) = P(A)$.

**Definition 1.3.3: Independent Random Variables**

Two random variables $X$ and $Y$ defined on $(\Omega, \mathcal{F}, P)$ are said to be *independent* if

$$\forall a, b \in \mathbb{R},\ P(X \le a, Y \le b) = P(X \le a)P(Y \le a).$$

A family $\mathcal{X}$ of random variables is said to be *independent* if, for any finite subfamily $\{X_1, \cdots, X_n\} \subseteq \mathcal{X}$, and for any $a_1, \cdots, a_n \in \mathbb{R}$, we have

$$P(X_1 \le a_1, \cdots, X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i).$$

**Note**

If $X$ and $Y$ takes values $\langle a_n \rangle_{n \in \mathbb{Z}_+}$ and $\langle b_n \rangle_{n \in \mathbb{Z}_+}$, respectively, then $X$ and $Y$ are independent if and only if

$$P(X = a_i, Y = b_j) = P(X = a_i)P(Y = b_j)$$

for all $i, j \in \mathbb{Z}_+$. It is analogous to family of discrete random variables.

**Lemma 1.3.4** Bayes' Retrodiction Formula

If $A$ and $B$ are events of positive probability, then

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}.$$

**Lemma 1.3.5** Bayes' Sequential Formula

Let $A_1, \cdots, A_n$ be events such that $P(A_1 \cap \cdots \cap A_n) > 0$. Then,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

***Proof.*** Mathematical induction. □

**Lemma 1.3.6** Law of Total Probability

Let $A$ be an event, and let $\langle B_n \rangle_{n \in \mathbb{Z}_{>0}}$ be an exaustive sequence of events. In other words, $\bigcup_{n=1}^{\infty} B_n = \Omega$ and $B_i \cap B_j = \varnothing$ for all $1 \leq i < j$. Then, we have

$$P(A) = \sum_{n=1}^{\infty} P(A \mid B_n)P(B_n)$$

where we agree to have $P(A \mid B_n)P(B_n) = 0$ when $P(B_n) = 0$. Moreover, for all $m \in \mathbb{Z}_{>0}$, we have

$$P(B_m \mid A) = \frac{P(A \mid B_m)P(B_m)}{\sum_{n=1}^{\infty} P(A \mid B_n)P(B_n)}$$

if $P(A) > 0$.

***Proof.*** $A = A \cap \Omega = A \cap \left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty}(A \cap B_n)$. Apply $\sigma$-additivity to obtain the result. Note that $P(A \cap B_n) = P(A \mid B_n)P(B_n)$ always according to our convention. □

## 1.4 Counting and Probability

If $\Omega$ is finite and we let $p(\omega) \coloneqq P(\{\omega\})$ with equal probabilities, then we must have $P(A) = (\operatorname{card} A)/(\operatorname{card} \Omega)$ for all $A \subseteq \Omega$. Hence, we should *count*.

**Example 1.4.1**

- The number of injections from $E$ to $F$ with $p = \operatorname{card}(E)$ and $n = \operatorname{card}(F)$ when $p \leq n$ is $A_p^n = \frac{n!}{(n-p)!}$.
- In particular, if $p = n$, we have $A_n^n$, the number of permutations of $n$ elements, which is $n!$.
- The number of subsets of $F$ with $p$ elements is $\binom{n}{p} = \frac{n!}{p!(n-p)!}$.
- (Binomial formula) $(x+y)^n = \sum_{p=0}^{n} x^p y^{n-p}$. $2^n = \sum_{p=0}^{n} \binom{n}{p}$.
- $\binom{n}{p} = \binom{n}{n-p}$.
- (Pascal's formula) $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$.

# Chapter 2

# Discrete Probability

## 2.1 Discrete Random Elements

**Definition 2.1.1: Discrete Random Element**

Let $E$ be a denumerable set and let $(\Omega, \mathcal{F}, P)$ be a probability space. Any function $X : \Omega \to E$ such that

$$\forall x \in E, \{ \omega \mid X(\omega) = x \} \in \mathcal{F}$$

is called a *discrete random element* of $E$. When $E \subseteq \mathbb{R}$, we refer to $X$ as a *discrete random variable*. This allows us to define

$$p(x) := P(X = x)$$

for $x \in E$. The collection $\{ p(x) \mid x \in E \}$ is the *distribution* of $X$. It satisfies

$$0 \leq p(x) \leq 1 \quad \text{and} \quad \sum_{x \in E} p(x) = 1.$$

**Note**

$E$ being denumerable enables us to define in such way. Note the difference from Definition 1.2.1.

## 2.2 Expectation

**Definition 2.2.1: Expectation of Discrete Random Variable**

Let $X$ be a random element taking its values in $E$, and let $f : E \to \mathbb{R}$ be a function such that

$$\sum_{x \in E} |f(x)| p(x) < \infty. \tag{2.1}$$

One then defines the *expectation* of $f(X)$, denoted $\mathbb{E}[f(X)]$, by

$$\mathbb{E}[f(X)] := \sum_{x \in E} f(x) p(x).$$

**Note**

If ⟨2.1⟩ is satisfied, $\mathbb{E}[f(X)]$ is well-defined and finite. If ⟨2.1⟩ is not satisfied and $f$ is nonnegative, then $\mathbb{E}[f(X)]$ is well-defined but can be infinite. Otherwise, $\mathbb{E}[f(X)]$ may not be well-defined.

**Note**

Definition 2.2.1 easily extends to $f : E \to \mathbb{C}$ with the same condition. Writing $f = g + ih$, ⟨2.1⟩ is equivalent to

$$\sum_{x\in E} |g(x)|p(x) < \infty \quad \text{and} \quad \sum_{x\in E} |h(x)|p(x) < \infty.$$

**Note**

Some properties of expectation:

- *Linearity*. $\mathbb{E}[\lambda_1 f_1(X) + \lambda_2 f_2(X)] = \lambda_1 \mathbb{E}[f_1(X)] + \lambda_2 \mathbb{E}[f_2(X)]$.
- *Monotonicity*. If $\forall x \in E$, $f_1(x) \le f_2(x)$, then $\mathbb{E}[f_1(X)] \le \mathbb{E}[f_2(X)]$.
- $|\mathbb{E}[f(X)]| \le \mathbb{E}[|f(X)|]$.
- Let $C \subseteq E$ and let $I_C$ be the *indicator function* of $C$ defined by

$$I_C(x) := \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{otherwise.} \end{cases}$$

  Then, $\mathbb{E}[I_C(X)] = \sum_{x\in E} I_C(x)p(x) = \sum_{x\in C} p(x) = \sum_{x\in C} P(X = x) = P\big(\bigcup_{x\in C}\{X = x\}\big)$.
- Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$. Defining the indicator function $I_A : \Omega \to \{0, 1\}$ for $A$, $I_A$ is clearly a discrete random variable taking values on $\{0, 1\}$. We have $\mathbb{E}[I_A] = P(A)$.

**Theorem 2.2.2** Markov's Inequality

Let $f : E \to \mathbb{R}$ satisfy ⟨2.1⟩. Then, for $a > 0$, we have

$$P(|f(X)| \ge a) \le \frac{\mathbb{E}[|f(X)|]}{a}.$$

***Proof.*** Let $C := \{ x \in E \mid |f(x)| \ge a \} \subseteq E$. Then, $|f(x)| \ge |f(x)|I_C(x)$ and thus

$$\begin{aligned} \mathbb{E}[|f(X)|] &\ge \mathbb{E}[|f(x)|I_C(x)] \\ &\ge \mathbb{E}[aI_C(X)] \\ &= a\mathbb{E}[I_C(X)] = aP(|f(X)| \ge a). \end{aligned}$$

□

## 2.3 Independence

**Definition 2.3.1: Independence of Discrete Random Elements**

Let $X$ and $Y$ be two discrete random elements with values in the denumerable spaces $E$ and $F$, respectively. Now, one can define another random element $Z$ on $G := E \times F$ by $Z(\omega) = (X(\omega), Y(\omega))$. We say $X$ and $Y$ are *independent* if

$$P(X = x, Y = y) := P(Z = (x, y)) = P(X = x)P(Y = y)$$

for all $x \in E$ and $y \in F$. This can be ge

**Lemma 2.3.2** Product Formula

Let $X$ and $Y$ be two discrete random elements with values in the denumerable spaces $E$ and $F$, respectively. If $f : E \to \mathbb{R}$ and $g : F \to \mathbb{R}$ satisfy ⟨2.1⟩, and if $X$ and $Y$ are independent, then $\mathbb{E}[f(X)g(Y)]$ is well-defined and

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)].$$

**Note**

Definition 2.3.1 and Lemma 2.3.2 can readily be generalized to finite number of discrete random elements.

## 2.4 Mean and Variance

**Definition 2.4.1: Mean and Variance of Discrete Random Variable**

If $X$ is a discrete random variable, the quantities

$$m \triangleq \mathbb{E}[X] \quad \text{and} \quad \sigma^2 \triangleq \mathrm{Var}[X] \triangleq \mathbb{E}[(X - m)^2]$$

are called the *mean* and *variance* of $X$, respectively. The quantity $\sigma \triangleq \sqrt{\sigma^2}$ is called the *standard deviation* of $X$.

**Note**

Some properties of mean and variance:

- $\mathrm{Var}[aX] = a^2 \mathrm{Var}[X]$.
- $\sigma^2 = 0$ implies that $p(x) = 0$ for all $x \neq m$.
- If $X_1, \cdots, X_n$ are independent discrete random variables, then $\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathrm{Var}[X_i]$.

**Theorem 2.4.2** Chebyshev's Inequality

Let $X$ be a discrete random variable. Then, for any $\varepsilon > 0$, we have

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

***Proof.*** Apply Markov's Inequality to $X$ with $f(x) = (x-m)^2$ and $a = \varepsilon^2$ to get

$$\begin{aligned} P(|X-m| \geq \varepsilon) &= P((X-m)^2 \geq \varepsilon^2) \\ &\leq \frac{\mathbb{E}[|X-m|^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}. \end{aligned}$$

□

**Theorem 2.4.3** Weak Law of Large Numbers

Let $\langle X_n \rangle_{n \in \mathbb{Z}_{>0}}$ be a sequence of discrete random variables, identically distributed with common mean $m$ and common variance $\sigma^2$.

***End.***