Probability Theory

for Finance and Actuarial Science

An undergraduate course

Parin Chaipunya

King Mongkut's University of Technology Thonburi

parin.cha@kmutt.ac.th

How to use this text.

This text is designed for an undergraduate course MTH382 at King Mongkut's University of Technology Thonburi (KMUTT). Throughout the lecture note, we shall encounter several *examples*, *problems*, and *exercises*. *Examples* are meant to be solved together by the lecturer (me!) and students. *Problems* are meant to be solved by the students during the class. Lastly, *Exercises* are practice problems that the students should try to solve outside of class hours.

We adopt the following conventions throughout this lecture note:

Notations	**Meaning**
$\mathbb{R}$	The real line.
$\mathbb{N}$	The set of positive integers.
2^A	The power set of A.
$\#A$	The cardinality of A.
$A^{\complement}$	The complement of A.
$A \subset B$	A is a subset of B, with the possibility that $A = B$.
$A \uplus B$	The union of two disjoint sets A and B.
$\biguplus_{\lambda\in\Lambda} A_\lambda$	The union of a pairwise disjoint collection of sets $\{A_\lambda\}_{\lambda\in\Lambda}$.
$\blacksquare$	Completion of a proof (QED).
$\blacktriangle$	Completion of a definition, an example, or a remark.

Course topics

- Probability space.
- Conditional Probability.
- Probability distribution.
- Joint probability distribution.
- Random variable.
- Expected value and variance.
- Moment-generating function.
- Characteristic functions.
- Brancing process.
- Random sum.

- Central limit theorem.
- Basic Markov chain.
- Applications to finance and actuarial science.

Contents

Chapter 1. Soft landing on probability

Probability theory is a mathematical theory that helps dealing with uncertain nature of the world. The uncertainty here may range from coin flipping, dice rolling, casino games, stock prices, accidents, weather conditions, human behaviors, and much more. Being able to systematically and mathematically capture the randomness, probability theory is used in vast applications including finance, risk management, insurance, game theory, resource management, etc.

The subject can be studied with varying level of depths — the counting-based probability theory is studied in high schools, while the theory itself can be so deep that it involves measure and theory of integration, and functional analysis. This lecture note aims to provide a balance mix between the theoretical and practical perspectives of probability theory, with some examples and use cases from finance and insurance. The audience will require some background on combinatorics, calculus, set theory, and some basics of proofs in order to effectively follower the course.

This chapter's objective is twofold — firstly, to give a brief refreshment to probability with counting models, and secondly to shape the intuition for the formal probability spaces with an emphasis on the event space through examples.

1.1. Counting models

Before we actually start, let us fix some terminlogy. Each random phenomena (or experiment) is resulted with an *outcome*. It is agreed that *all* the possible outcomes of a random phenomena is known, but the *true* outcome is not known. The set Ω containing all possible outcomes is called the *sample space*. In this section (only), Ω will always be a finite set, and any subset A of Ω is an *event*. We shall see later that, in general, not every subset of Ω is an event. The *probability that an outcome belongs to an event A* can then be calculated by

$$P(A) = \frac{\#A}{\#\Omega}.$$

Example 1.1. A set of number cards consists of 10 cards with different faces $\mathtt{0}, \mathtt{1}, \ldots, \mathtt{9}$. A person is to randomly pick one card from this set. What is the probability that the drawn card has an odd number on it ?

The event that the drawn card has an odd face is described by

$$\mathsf{Odd} = \{1, 3, 5, 7, 9\}.$$

Thus we have

$$P(\mathsf{Odd}) = \frac{\#\mathsf{Odd}}{\#\Omega} = \frac{5}{10} = \frac{1}{2}. \qquad \blacktriangle$$

Example 1.2. Three cards are numbered with $\mathsf{1}, \mathsf{2}, \mathsf{3}$. A person randomly arrange them linearly.

- Find the number of all possible outcomes.
  Solution. We shall represent each outcome with $\omega = (\mathsf{a}, \mathsf{b}, \mathsf{c})$, where $\mathsf{a}, \mathsf{b}, \mathsf{c}$ denote respectively the faces of the first, second, and third cards. The sample space Ω consists of all permutations of $\mathsf{1}, \mathsf{2}, \mathsf{3}$, so that $\#\Omega = 3\,! = 6$.
- What is the probability that the middle card has even face ?
  Solution. Let us write Even to denote the event that the middle card is even, which gives $\mathsf{Even} = \{(1, 2, 3), (3, 2, 3)\}$. Hence

$$P(\mathsf{Even}) = \frac{\#\mathsf{Even}}{\#\Omega} = \frac{2}{3\,!} = \frac{1}{3}.$$

- What is the probability that the middle card has odd face ?
  Solution. Let us write Odd to denote the event that the middle card is odd, resulting with

$$\mathsf{Odd} = \{(1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2)\}.$$

  We then obtain

$$P(\mathsf{Odd}) = \frac{\#\mathsf{Odd}}{\#\Omega} = \frac{2}{3}. \qquad \blacktriangle$$

Example 1.3. A jar contains 3 black balls, 2 red balls, and 4 white balls. A person draws two balls from the jar, one after another without replacing.

- How many possible outcomes are there ?
  Solution. As usual, we denote with Ω the set of all possible outcomes. An outcome of this problem is represented by $\omega = (\mathsf{a}, \mathsf{b})$, where a and b denotes respectively the outcomes of the first and second balls. There are 9 ways to choose the first balls, and 8 ways to choose the second balls. Thus $\#\Omega = 9 \times 8 = 72$.
- What is the probability that the first ball is red ?
  Solution. We write $\{\mathsf{a} = \mathsf{red}\}$ to denote the event where the first draw is a red ball. We have $\#\{\mathsf{a} = \mathsf{red}\} = 3 \times 8 = 24$ and finally

$$P(\{\mathsf{a} = \mathsf{red}\}) = \frac{\#\{\mathsf{a} = \mathsf{red}\}}{\#\Omega} = \frac{24}{72} = \frac{1}{3}.$$

- What is the probability that the second ball is white ?

Solution. We write the event by $\{\mathsf{b} = \mathsf{white}\}$. To effectively compute its cardinality, we partition this set into

$$\{\mathsf{b} = \mathsf{white}\} = \underbrace{\{\mathsf{a} = \mathsf{white},\ \mathsf{b} = \mathsf{white}\}}_{\substack{\text{The first ball is white and}\\ \text{the second ball is white.}}} \cup \underbrace{\{\mathsf{a} \neq \mathsf{white},\ \mathsf{b} = \mathsf{white}\}}_{\substack{\text{The first ball is not white but}\\ \text{the second ball is white.}}}$$

and since both sets are disjoint, we get

$$\begin{aligned}\#\{\mathsf{b} = \mathsf{white}\} &= \underbrace{\#\{\mathsf{a} = \mathsf{white},\ \mathsf{b} = \mathsf{white}\}}_{=4\times 3} + \underbrace{\#\{\mathsf{a} \neq \mathsf{white},\ \mathsf{b} = \mathsf{white}\}}_{=5\times 4} \\ &= 4 \times 3 + 5 \times 4 = 32.\end{aligned}$$

Finally we obtain

$$P(\{\mathsf{b} = \mathsf{white}\}) = \frac{\#\{\mathsf{b} = \mathsf{white}\}}{\#\Omega} = \frac{4}{9}.$$

○ What is the probability that at least one of the balls is black ?
Solution. We first argue that the event $\{\mathsf{a} = \mathsf{black} \text{ or } \mathsf{b} = \mathsf{black}\}$ can be decomposed disjointly as

$$\begin{aligned}\{\mathsf{a} = \mathsf{black} \text{ or } \mathsf{b} = \mathsf{black}\} = \{&\mathsf{a} = \mathsf{black},\ \mathsf{b} \neq \mathsf{black}\} \\ &\cup \{\mathsf{a} \neq \mathsf{black},\ \mathsf{b} = \mathsf{black}\} \\ &\cup \{\mathsf{a} = \mathsf{black},\ \mathsf{b} = \mathsf{black}\}.\end{aligned}$$

This leads to

$$\begin{aligned}\#\{\mathsf{a} = \mathsf{black} \text{ or } \mathsf{b} = \mathsf{black}\} &= \underbrace{\#\{\mathsf{a} = \mathsf{black},\ \mathsf{b} \neq \mathsf{black}\}}_{=3\times 6} \\ &\quad + \underbrace{\#\{\mathsf{a} \neq \mathsf{black},\ \mathsf{b} = \mathsf{black}\}}_{=6\times 3} \\ &\quad + \#\underbrace{\{\mathsf{a} = \mathsf{black},\ \mathsf{b} = \mathsf{black}\}}_{=3\times 2} \\ &= 42.\end{aligned}$$

Finally we get

$$P(\{\mathsf{a} = \mathsf{black} \text{ or } \mathsf{b} = \mathsf{black}\}) = \frac{42}{72} = \frac{7}{12}.$$ ▲

Example 1.4. A jar contains 3 black balls, 2 red balls, and 4 white balls. A person draws two balls from the jar, one after another with the balls returned to the jar.

○ How many possible outcomes are there ?
Solution. Both balls can be drawn from the jar in 9 ways. Hence the sample space Ω consists of $\#\Omega = 9 \times 9 = 81$ elements.

- What is the probability that none of the balls is red ?
  Solution. Following the same notation of the previous examples, we may obtain

$$\#\{\mathsf{a} \neq \mathsf{red},\ \mathsf{b} \neq \mathsf{red}\} = 7 \times 7 = 49.$$

  Then we conclude that $P(\#\{\mathsf{a} \neq \mathsf{red},\ \mathsf{b} \neq \mathsf{red}\}) = \frac{49}{81}$.
- What is the probability that at least one of the balls is read ?
  Solution. It is noticeable that this event is the complement of $\#\{\mathsf{a} \neq \mathsf{red},\ \mathsf{b} \neq \mathsf{red}\}$, so that the corresponding probability is $1 - \frac{49}{81} = \frac{32}{81}$.
- What is the probability that after the two draws, one ball is black and the other one is red ?
  Solution. This event, denoted by A, can be partitioned into two disjoint cases. In Case 1, the first draw gives a black ball and the second one is red. This case covers a possibility of $3 \times 2 = 6$ outcomes. In Case 2, the first draw is red and the second one is black. This gives another $2 \times 3 = 6$ cases. Finally, the enumeration of this event is

$$\#A = \underbrace{6}_{\text{Case 1}} + \underbrace{6}_{\text{Case 2}} = 12,$$

  which imples that $P(A) = \frac{12}{81}$. ▲

Example 1.5. A jar contains 3 black balls, 2 red balls, and 4 white balls. A person draws two balls, at the same time, from the jar.

- How many possible outcomes are there ?
  Solution. Let Ω be the sample space, whose cardinality is computed with the binomial formula

$$\#\Omega = \mathrm{C}_{9,2} = \binom{9}{2} = \frac{9\,!}{2\,!\;7\,!} = 36.$$

- What is the probability that both balls are white ?
  Solution. We denote such an event with WW and we have $\#\mathsf{WW} = \binom{4}{2} = 6$. We thus have $P(\mathsf{WW}) = \frac{6}{36} = \frac{1}{6}$.
- What is the probability that one of the ball is black and the other one is red.
  Solution. We denote such an event with $A \subset \Omega$, and enumerate such an event by

$$\#A = \underbrace{\binom{3}{1}}_{\text{Black balls.}} \times \underbrace{\binom{2}{1}}_{\text{Red balls.}} = 6.$$

  Subsequently $P(A) = \frac{6}{36} = \frac{1}{6}$.
- What is the probability that exactly one of the two balls is white ?
  Solution. Let B denote such an event. Then B can be enumerated by

$$\#B = \underbrace{4}_{\text{White balls.}} \times \underbrace{5}_{\text{Non-white balls.}} = 20,$$

  which gives $P(B) = \frac{20}{36} = \frac{5}{9}$.

- What is the probability that at least one of the two balls is black ?
  Solution. Let C denote such an event. We focus instead of $C^{\complement}$ which represents the event where both balls are not black. Hence we get

$$\#C^{\complement} = \underbrace{\binom{6}{2}}_{\text{Non-black balls.}} = 15,$$

  implying that

$$P(C) = 1 - P(C^{\complement}) = 1 - \frac{15}{36} = \frac{7}{13}. \qquad \blacktriangle$$

Example 1.6. A jar contains 3 black balls, 2 red balls, and 4 white balls. A person draws four balls, at the same time, from the jar.

- How many possible outcomes are there ?
  Solution. Let Ω be the sample space, whose cardinality is computed with the binomial formula

$$\#\Omega = \binom{9}{4} = \frac{9\,!}{4\,!\;5\,!} = 126.$$

- What is the probability that two of the balls are black and the other two are white ?
  Solution. Let A denote such an event. We then have

$$\#A = \underbrace{\binom{3}{2}}_{\text{Black balls.}} \times \underbrace{\binom{4}{2}}_{\text{White balls}} = 18,$$

  which yields $P(A) = \frac{18}{126} = \frac{1}{7}$.
- What is the probability that at least two of the balls have the same color ?
  Solution. We denote such an event with B. We get

$$\#B = \underbrace{\binom{3}{2}\binom{6}{2}}_{\text{2 black balls.}} + \underbrace{\binom{3}{3}\binom{6}{1}}_{\text{3 black balls.}} + \underbrace{\binom{2}{2}\binom{7}{2}}_{\text{2 red balls.}} + \underbrace{\binom{4}{2}\binom{5}{2}}_{\text{2 white balls.}} + \underbrace{\binom{4}{3}\binom{5}{1}}_{\text{3 white balls.}} + \underbrace{\binom{4}{4}}_{\text{4 white balls.}}$$

  and the probability $P(B)$ can be calculated using a similar formula. $\blacktriangle$

1.2. Describing a probability space

We have seen in the previous section that an event is a subset of the sample space. However, not every subset is considered as an event. In this section we shall motivate ourselves, through examples, to get closer to the formal definition of a probabilistic space. The rigorous analysis of this definition, however, will be postponed until the Chapter 1.

To describe a *probability space*, three ingredients are required:

- a *sample space* Ω containing all possible *outcomes*,
- a *event space* $\mathcal{F} \subset 2^{\Omega}$ containing all observable *events*,

- a *probability measure* $P : \mathcal{F} \to [0,1]$, assigning to each event a numerical value between 0 and 1 representing its probability.

Again, we insist that it might be *impossible* to assign a probability to every subset of Ω. This behavior is typical in measure theory, but is also quintessential in probability theory. One could see from the following example that a probability could not be practically assigned to some events.

Example 1.7. Consider tossing two identical coins simultaneously.

- The sample space is $\Omega = \{(\mathsf{H},\mathsf{H}),(\mathsf{H},\mathsf{T})(\mathsf{T},\mathsf{H}),(\mathsf{H},\mathsf{H})\}$, where the first (second) coordinate corresponds to the first (second) coin. Here H and T denote respectively the Head and Tail of each coin.
- The base events are

$$\underbrace{\{(\mathsf{H},\mathsf{H})\}}_{\text{Both coins show heads.}} \quad , \quad \underbrace{\{(\mathsf{T},\mathsf{T})\}}_{\text{Both coins show tails.}} \quad , \quad \underbrace{\{(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H})\}}_{\text{Two coins show different sides.}}$$

  but we could also observe the secondary events including their (countable) complements, unions and intersections. Indeed, the event space in this case consists of such base events and their secondary events.
- Based on counting principles, we obtain that

$$P(\{(\mathsf{H},\mathsf{H})\}) = \frac{1}{4} \quad P(\{(\mathsf{T},\mathsf{T})\}) = \frac{1}{4} \quad P(\{(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H})\}) = \frac{1}{2}.$$

  The probability of the secondary events could be computed likewise.

Make a note that it is impossible, and is not even practical, to assign a probability to the set $\{(\mathsf{H},\mathsf{T})\}$. The fact that the two coins are tossed simultaneously makes it impossible to distinguish between the two. Hence it is not possible to observe the difference between the outcomes (H,T) and (T,H). It is, therefore, does not make sense to assign a probability to an *unobservable* event. ▲

Example 1.8. Mr.Ali tosses two fair coins simultaneously and Mr.Charlie observes the outcome and inform to Mr.Ali the *parity* of the toss, i.e. whether the two coins show the same face. Mr.Ali has the following in his probability space.

- The sample space is $\Omega = \{(\mathsf{H},\mathsf{H}),(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H}),(\mathsf{T},\mathsf{T})\}$, where H and T stand for Head and Tail, respectively, and the ordered pair represents the outcomes of the first and second coin in their respected orders.
- The event space $\mathcal{F}$ includes the two base events as well as their unions and intersections, so that

$$\mathcal{F} = \{\mathsf{Similar}, \mathsf{Different}, \varnothing, \Omega\}.$$

- The probability assigned to each event is defined by

$$P(\mathsf{Similar}) = P(\mathsf{Different}) = \frac{1}{2}, \quad P(\varnothing) = 0, \quad P(\Omega) = 1.$$ ▲

Example 1.9. Consider the tossing two fair coins in sequence. This time we focus on observing the outcomes of each toss.

- The sample space is $\Omega = \{(\mathsf{H}, \mathsf{H}), (\mathsf{H}, \mathsf{T}), (\mathsf{T}, \mathsf{H}), (\mathsf{T}, \mathsf{T})\}$.
- Let us write the outcomes in the form $(\mathsf{X}_1, \mathsf{X}_2)$. The base events in questions are

$$\underbrace{\{\mathsf{X}_1 = \mathsf{H}\} = \{(\mathsf{H}, \mathsf{H}), (\mathsf{H}, \mathsf{T})\}}_{\text{The first toss is } \mathsf{H}.}, \qquad \underbrace{\{\mathsf{X}_1 = \mathsf{T}\} = \{(\mathsf{T}, \mathsf{H}), (\mathsf{T}, \mathsf{T})\}}_{\text{The first toss is } \mathsf{T}.},$$
$$\underbrace{\{\mathsf{X}_2 = \mathsf{H}\} = \{(\mathsf{H}, \mathsf{H}), (\mathsf{T}, \mathsf{H})\}}_{\text{The second toss is } \mathsf{H}.}, \qquad \underbrace{\{\mathsf{X}_2 = \mathsf{T}\} = \{(\mathsf{H}, \mathsf{T}), (\mathsf{T}, \mathsf{T})\}}_{\text{The second toss is } \mathsf{T}.}.$$

- The full event space is formed by all unions and intersections of these base events.
- The probability assigned to each base event is defined by

$$P(\{\mathsf{X}_1 = \mathsf{H}\}) = P(\{\mathsf{X}_1 = \mathsf{T}\}) = P(\{\mathsf{X}_2 = \mathsf{H}\}) = P(\{\mathsf{X}_2 = \mathsf{T}\}) = \frac{1}{4}. \qquad \blacktriangle$$

Example 1.10. We consider dice rolling and focus on the "Odd – Even" scenarios.

- The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$, which contains all the possible outcomes of a dice roll.
- We consider two *base events*, namely $\mathsf{Odd} = \{1, 3, 5\}$ and $\mathsf{Even} = \{2, 4, 6\}$.
- The *events* are Odd and Even themselves, as well as

$$\varnothing = \underbrace{\mathsf{Odd} \cap \mathsf{Even}}_{\text{Odd and even at the same time.}} \quad \text{and} \quad \Omega = \underbrace{\mathsf{Odd} \cup \mathsf{Even}}_{\text{Odd or even.}}.$$

  This means the full event space is

$$\mathcal{F} = \{\mathsf{Odd}, \mathsf{Even}, \varnothing, \Omega\}.$$

- The *probability* is then assigned to each event as

$$P(\mathsf{Odd}) = P(\mathsf{Even}) = \frac{3}{6}, \quad P(\varnothing) = 0, \quad P(\Omega) = 1. \qquad \blacktriangle$$

Example 1.11. In this example, we consider the "Odd – Even – High – Low" scenarios for dice rolling.

- The sample space is again $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- We consider two *base events*, namely $\mathsf{Odd} = \{1, 3, 5\}$ and $\mathsf{Even} = \{2, 4, 6\}$.
- The *events* are Odd, Even, High and Low themselves, as well as $\varnothing$, Ω and additionally their unions and complements.
- The *probability* is then assigned to each event as

$$P(\mathsf{Odd}) = P(\mathsf{Even}) = P(\mathsf{High}) = P(\mathsf{Low}) = \frac{3}{6}, \quad P(\varnothing) = 0, \quad P(\Omega) = 1$$

  and

$$P(\mathsf{Odd} \cap \mathsf{High}) = P(\mathsf{Even} \cap \mathsf{Low}) = \frac{1}{6}, \quad P(\mathsf{Odd} \cap \mathsf{Low}) = P(\mathsf{Even} \cap \mathsf{High}) = \frac{1}{2},$$

$$P(\mathsf{Odd} \cup \mathsf{High}) = P(\mathsf{Even} \cup \mathsf{Low}) = \frac{5}{6}, \quad P(\mathsf{Odd} \cup \mathsf{Low}) = P(\mathsf{Even} \cup \mathsf{High}) = \frac{4}{6}. \qquad \blacktriangle$$

Example 1.12. Don Juan has two girlfriends, Amy and Betty, and he is now considering marrying one of them. Being a womanizer, we could not decide who to marry so he plans to make a phone call to each of them. Here are the rules of this marriage:

a. Don Juan makes the first call to one of the two girlfriends at random based on a coin toss.
b. Each girlfriend does not know if she got the first or second call.
c. Each girlfriend either accept or reject the proposal at random based on a coin toss.
d. If the girlfriend of the first call accepted the proposal, they get married and live happily everafter.
e. If the girlfriend of the first call rejected the proposal, then Don Juan makes the second call to the other girlfriend.

Let Y and N denote respectively the acceptance and rejection of either Amy or Betty. The sample space is given by

$$\Omega = \left\{ \begin{array}{cc} \underbrace{(\mathsf{Amy}, \mathsf{Y})}_{\substack{\text{Amy is called first} \\ \text{and she accepted.}}}, & \underbrace{(\mathsf{Betty}, \mathsf{Y})}_{\substack{\text{Betty is called first} \\ \text{and she accepted.}}}, \\ \underbrace{(\mathsf{Amy}, \mathsf{N}, \mathsf{Betty}, \mathsf{Y})}_{\substack{\text{Amy is called first, she rejected.} \\ \text{Betty is then called and she accpeted.}}}, & \underbrace{(\mathsf{Amy}, \mathsf{N}, \mathsf{Betty}, \mathsf{N})}_{\substack{\text{Amy is called first, she rejected.} \\ \text{Betty is then called and she rejected.}}}, \\ \underbrace{(\mathsf{Betty}, \mathsf{N}, \mathsf{Amy}, \mathsf{Y})}_{\substack{\text{Betty is called first, she rejected.} \\ \text{Amy is then called and she accpeted.}}}, & \underbrace{(\mathsf{Betty}, \mathsf{N}, \mathsf{Amy}, \mathsf{N})}_{\substack{\text{Betty is called first, she rejected.} \\ \text{Amy is then called and she rejected.}}} \end{array} \right\}.$$

Let us observe the event spaces from the viewpoints Don Juan, Amy, and Betty.

- Don Juan knows exactly the call sequence as well as the answers of each call. Therefore the event space that represents his information is the whole power set, that is

$$\mathcal{F}_{\mathrm{DJ}} = 2^{\Omega}.$$

- From the angle of Amy, she only knows whether Don Juan called and whether she accepted or rejected. Hence the event space that represents her information is

$$\mathcal{F}_{\mathrm{Amy}} = \sigma \left\{ \begin{array}{c} \underbrace{\{(\mathsf{Amy}, \mathsf{Y}), (\mathsf{Betty}, \mathsf{N}, \mathsf{Amy}, \mathsf{Y})\}}_{\text{Amy got a call and said yes.}}, \ \underbrace{\{(\mathsf{Amy}, \mathsf{N}), (\mathsf{Betty}, \mathsf{N}, \mathsf{Amy}, \mathsf{N})\}}_{\text{Amy got a call and said no.}}, \\ \underbrace{\{(\mathsf{Betty}, \mathsf{Y})\}}_{\text{Amy did not get a call.}} \end{array} \right\}.$$

  It should be noted that Amy could not tell if she was called first or second.

- Similar to Amy, Betty only knows whether Don Juan called and also her answer to the call. This means her event space that represents her information is

$$\mathcal{F}_{\mathrm{Betty}} = \sigma \left\{ \begin{array}{c} \underbrace{\{(\mathsf{Betty}, \mathsf{Y}), (\mathsf{Amy}, \mathsf{N}, \mathsf{Betty}, \mathsf{Y})\}}_{\text{Betty got a call and said yes.}}, \ \underbrace{\{(\mathsf{Betty}, \mathsf{N}), (\mathsf{Amy}, \mathsf{N}, \mathsf{Betty}, \mathsf{N})\}}_{\text{Betty got a call and said no.}}, \\ \underbrace{\{(\mathsf{Amy}, \mathsf{Y})\}}_{\text{Betty did not get a call.}} \end{array} \right\}.$$

  Again, Betty could not tell if she was the first or the second choice. ▲

Chapter 2. The formal introduction to probability

In this section, we move further to introduce the formal definition of a probability space and prove some of its fundmental properties.

2.1. Outcomes and events

The first two concepts that we need for a probability space are the *outcomes* and *events*. One may think of an outcome as a result of a random phenomena or a random experiment. The set containing all possible outcomes ω is called the *sample space* Ω. An *event*, on the other hand, is a subset of Ω that represents a situation in which a *probability* will be assigned. We say that an outcome ω *realizes* the event A if $\omega \in A$.

In practice, we may not be able to assign probability values to all subsets $A \subset \Omega$ (see e.g. Example 1.12). Even in theory, doing so might lead to some strange unwanted situations as well. Hence we make a requirement that the interested events constitute a nicely behaved collection called a σ-field.

Definition 2.1. Let Ω be a nonempty set. A family $\mathcal{F} \subset 2^{\Omega}$ is called a *σ-field* (also called a *σ-algebra*) over Ω if the following conditions are satisfied:
(σ1) $\Omega \in \mathcal{F}$.
(σ2) If $A \in \mathcal{F}$, then $A^{\complement} := \Omega \setminus A \in \mathcal{F}$.
(σ3) If $A_i \in \mathcal{F}$ for each $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
The pair $(\Omega, \mathcal{F})$, in the language of *measure theory*, is known as a *measurable space*. ▲

Immediately from the definition, we have the following proposition.

Proposition 2.2. *Let $(\Omega, \mathcal{F})$ be a measurable space. Then the following conditions hold:*
(a) $\varnothing \in \mathcal{F}$.
(b) *If $A_i \in \mathcal{F}$ for each $i \in \mathbb{N}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.*
(c) *If $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$.*
(d) *If $A, B \in \mathcal{F}$, then $A \setminus B \in \mathcal{F}$.*

Proof. (a) Since $\Omega \in \mathcal{F}$, (σ2) implies that $\varnothing = \Omega^{\complement} \in \mathcal{F}$.

(b) For each $i \in \mathbb{N}$, define

$$B_i = A_i^{\complement} \in \mathcal{F}.$$

Using $(\sigma 2)$ and $(\sigma 3)$, we obtain

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} B_i^{\complement} = \left(\bigcup_{i=1}^{\infty} B_i\right)^{\complement} \in \mathcal{F}.$$

(c) Let $A, B \in \mathcal{F}$. Then the result is obtained from $(\sigma 3)$ and (b) by setting $A_1 := A$, $A_i = B$ for $i = 2, 3, \ldots$.

(d) This follows since $B^{\complement} \in \mathcal{F}$ and that $A \setminus B = A \cap B^{\complement}$. ∎

Let us provide some first examples.

Example 2.3. Let Ω be a nonempty set.

- $\mathcal{F}_0 = \{\varnothing, \Omega\}$ is the smallest σ-field over Ω, known as the *trivial σ-field*.
- $\mathcal{F}_\infty = 2^\Omega$ is the largest σ-field over Ω, known as the *gross σ-field*.

Here, $\mathcal{F}_0$ and $\mathcal{F}_\infty$ are the *weakest* and the *strongest* σ-fields, respectively, in the sense that

$$\mathcal{F}_0 \subset \mathcal{F} \subset \mathcal{F}_\infty$$

for any σ-field $\mathcal{F}$ over Ω.

- Let $\Omega = \{a, b, c, d\}$. Then $\mathcal{F} = \{\varnothing, \{a, b\}, \{c, d\}, \Omega\}$ is a σ-field over Ω.
- For any Ω and $A \subset \Omega$, the set $\{\varnothing, A, A^{\complement}, \Omega\}$ is a σ-field.
- For any Ω and a nonempty subset $A \subset \Omega$, the family $\{\varnothing, A, \Omega\}$ is *not* a σ-field.
- Consider an interval $\Omega = (0, 1]$. Then

$$\mathcal{F} = \{\varnothing, (0, 0.25], (0.25, 0.5], (0, 0.5], (0.25, 1], (0.5, 1], (0, 0.25] \cup (0.5, 1], (0, 1]\}$$

is a σ-field over Ω. ▲

The following propositions serve as some more sophistcated examples of a σ-field.

Proposition 2.4. *Let Ω be a nonempty set and let*

$$\mathcal{A} = \{A \subset \Omega \mid \text{either } A \text{ or } A^{\complement} \text{ is countable}\}.$$

Then $\mathcal{A}$ is a σ-field over Ω.

Proof. It is clear that $\Omega \in \mathcal{A}$ since its complement $\Omega^{\complement} = \varnothing$ is countable. It is also obvious that, if $A \in \mathcal{A}$, then $A^{\complement} \in \mathcal{A}$. Now, let $A_i \in \mathcal{A}$ for each $i \in \mathbb{N}$. Then either

Case I. all A_i's are countable, or

Case II. there exists $i_0 \in \mathbb{N}$ such that A_{i_0} is uncountable.

In the first case, the union $\bigcup_{i \in \mathbb{N}} A_i$ itself is also countable. In the second case, the complement $A_{i_0}^{\complement}$ must be countable and so

$$\left(\bigcup_{i \in \mathbb{N}} A_i\right)^{\complement} = \bigcap_{i \in \mathbb{N}} A_i^{\complement} \subset A_{i_0}^{\complement}.$$

This means $\bigcup_{i \in \mathbb{N}} A_i$ is countable. Hence the union $\bigcup_{i \in \mathbb{N}} A_i$ belongs to $\mathcal{A}$, and the proof is completed. ∎

Proposition 2.5 (Trace σ-field)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $E \subset \Omega$ *is a given nonempty set. Then*

$$\mathcal{F}_E := \{E \cap A \mid A \in \mathcal{F}\}$$

is a σ*-field over* Ω.

Proof. The proof is left as an Exercise 2.1.. ∎

Proposition 2.6 (Pre-image σ-field)**.** *Let* $(\Omega', \mathcal{F}')$ *be a measurable space,* Ω *a nonempty set, and* $f : \Omega \to \Omega'$ *a given map. Then*

$$\mathcal{F}_f := \{f^{-1}(A') \mid A' \in \mathcal{F}'\}$$

is a σ*-field. Then is a* σ*-field over* Ω

Proof. The proof is left as an Exercise 2.6. ∎

In probability theory, the properties (σ1) – (σ3) are interpretted in terms of events as follows.

- (σ1) says that we can always observe the *certain event*.
- (σ2) says that if we could observe the occurrence of an event A, then we could also observe the non-occurrence of A.
- (σ3) and its implication (c) from Proposition 2.2 says that if we could observe the events A and B, then we could also observe whether at least one or both of them occurs.

Example 2.7. Consider $\Omega = \{(\mathsf{H},\mathsf{H}), (\mathsf{H},\mathsf{T}), (\mathsf{T},\mathsf{H}), (\mathsf{T},\mathsf{T})\}$, $\Omega' = \{0,1\}$ and $\mathcal{F}' = \{\varnothing, \{0\}, \{1\}, \Omega'\}$. This makes the pair $(\Omega', \mathcal{F}')$ a measurable space. Define a map $f : \Omega \to \Omega'$ by

$$f((\mathsf{H},\mathsf{T})) = f((\mathsf{T},\mathsf{H})) = 0, \quad f((\mathsf{H},\mathsf{H})) = f((\mathsf{T},\mathsf{T})) = 1.$$

Then we get

$$\begin{aligned}\mathcal{F}_f &= \{f^{-1}(\varnothing), f^{-1}(\{0\}), f^{-1}(\{1\}), f^{-1}(\Omega')\}\\ &= \{\varnothing, \{(\mathsf{H},\mathsf{T}), (\mathsf{T},\mathsf{H})\}, \{(\mathsf{H},\mathsf{H}), (\mathsf{T},\mathsf{T})\}, \Omega\}.\end{aligned}$$ ▲

In practice, we usually start from few *base events* (or *primitive events*) rather than the whole σ-field of events. We may, for example, recall from Example 1.11 that the primitive events Odd, Even, High, and Low are of our initial focus — other events are *generated* later. We take the same approach here: starting with few primitive events and generate a complete σ-field from there. To accomplish this idea, we first consider the following results.

Proposition 2.8. *Let* Ω *be a nonempty set and* $\{\mathcal{F}_\lambda\}_{\lambda \in \Lambda}$ *be a collection of* σ*-fields over* Ω*. Then* $\mathcal{F} := \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$ *is also a* σ*-field over* Ω.

Proof. Follows directly from the definition of a σ-field. ∎

Proposition 2.9. *Suppose that Ω is a nonempty set and $\mathcal{E} \subset 2^\Omega$ be a nonempty family of subsets of Ω. Then there exists a unique smallest σ-field over Ω containing $\mathcal{E}$, defined by*

$$\sigma(\mathcal{E}) := \bigcap \{\mathcal{F} \subset 2^\Omega \,|\, \mathcal{F} \supset \mathcal{E} \text{ is a } \sigma\text{-field over } \Omega.\} \tag{2.1}$$

Proof. Since the powerset 2^Ω is a σ-field containing $\mathcal{E}$, hence the intersection in (2.1) is nonempty. In view of Proposition 2.8, the family $\sigma(\mathcal{E})$ is a σ-field. Moreover, this intersection is a subset of all the σ-fields containing $\mathcal{E}$. This shows that $\sigma(\mathcal{E})$ is the smallest of such σ-fields. ∎

Definition 2.10. Suppose that Ω is nonempty and let $\mathcal{E} \subset 2^\Omega$ be nonempty family containing *base events*. Then the collection $\sigma(\mathcal{E})$, as defined in (2.1), is called the *σ-field generated by* $\mathcal{E}$. On the other hand, if $\mathcal{F}$ is a σ-field over Ω such that $\mathcal{F} = \sigma(\mathcal{E})$ for some collection of sets $\mathcal{E} \subset 2^\Omega$, then $\mathcal{F}$ is said to be *generated by* $\mathcal{E}$. ▲

It is unfortunate that $\sigma(\mathcal{E})$ has no explicit or constructive formula. Therefore it is usually not possible to make a descriptive list of its elements. One should have seen from the earlier examples that even in the case where Ω is finite, the enumeration of $\sigma(\mathcal{E})$ is a tough job. One might think that $\sigma(\mathcal{E})$ could be constructed, for any $\mathcal{E}$, by adding to the family $\mathcal{E}$ all possible countable unions of its members and complements. But this it not true, even when we repeat such a process uncountably (see a discussion in Schilling [1, pp.20–21]).

Let us consider some of the simpler examples.

Example 2.11. Let Ω be a finite set, and $\mathcal{E} = \{\{\omega\} \mid \omega \in \Omega\}$. Then $\sigma(\mathcal{E}) = 2^\Omega$. ▲

Example 2.12. Let $\Omega = \{a, b, c, d\}$ and $\mathcal{E} = \{\{a, b\}, \{b, c\}\}$. Then $\sigma(\mathcal{E}) = 2^\Omega$. ▲

More generally, one could come up with the following result.

Proposition 2.13. *Let Ω be given, and Ω is partitioned into $A_1, A_2, \ldots, A_N$ (which means $\Omega = A_1 \uplus \cdots \uplus A_N$). Then $\#\sigma(\{A_1, \ldots, A_N\}) = 2^N$.*

Proof. The proof is left as an Exercise 2.4. ∎

2.2. Probability measures

We split this section into two subsections. The first subsection focuses on the definition of a probability measure itself as well as some examples, without the need to use additional properties. The second subsection focuses on deriving further properties and their applications in finance and actuarial examples.

2.2.1. The definition

Let us begin directly with the definition of a probability measure and a probability space.

Definition 2.14. Let $(\Omega, \mathcal{F})$ be a measurable space. A *probability measure* (or simply a *probability*) on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that the following conditions are satisfied:

(P1) $P(\Omega) = 1$.

(P2) For any sequence of events $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{E}$ such that $A_i \cap A_j = \varnothing$ for all $i, j \in \mathbb{N}$ with $i \neq j$, it holds
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Whenever P is a probability measure on $(\Omega, \mathcal{F})$, the tripple $(\Omega, \mathcal{F}, P)$ is called a *probability space.* ▲

The defining properties (P1) and (P2) of a probability measure could be utilized, as seen in the following examples.

Example 2.15 (SOA Exam P Sample Question)**.** You are given that $P(A \cup B) = 0.7$ and $P(A \cup B^{\complement}) = 0.9$. Calculate $P(A)$.

Solution. One may notice that $(A \cup B^{\complement})^{\complement} = A^{\complement} \cap B = B \setminus A$. This allows us to calculate

$$(B \setminus A) = P((A \cup B^{\complement})^{\complement}) = 1 - P(A \cup B^{\complement}) = 1 - 0.9 = 0.1.$$

Since $A \cup B = A \uplus (B \setminus A)$, we get from (P2) that

$$0.7 = P(A \cup B) = P(A \uplus (B \setminus A)) = P(A) + P(B \setminus A) = P(A) + 0.1,$$

which yields $P(A) = 0.6$. ▲

Example 2.16 (SOA Exam P Sample Question)**.** In modeling the number of claims filed by an individual under an automobile policy during a three-year period, an actuary makes the simplifying assumption that for all integers $n \geq 0$, $p(n + 1) = 0.2p(n)$ where $p(n)$ represents the probability that the policyholder files n claims during the period. Calculate the probability that a policyholder files more than one claim.

Solution. In this example, $\Omega = \{0, 1, 2, \dots\} = \mathbb{N}_0$ and $\mathcal{F} = 2^{\Omega}$. Note that the events that a policyholder files n claims are disjoint for different integers $n \geq 0$. This means

$$1 = P(\Omega) = P(0) + P(1) + P(2) + \cdots = \sum_{n=0}^{\infty} P(n) = P(0) \sum_{n=0}^{\infty} (0.2)^n = \frac{P(0)}{1 - 0.2},$$

so that we obtain $P(0) = 0.8$. Finally, we calculate

$$P(2) + P(3) + \cdots = 1 - P(0) - P(1) = 1 - 0.8 - 0.2 \times 0.8 = 0.04.$$ ▲

Apart from the examples above and the ones already seen in Chapter 1, we take this opportunity to present a more sophisticated example of a probability space.

Example 2.17. In the unit interval $[0, 1]$, we put $\mathcal{I} := \{(a, b) \mid 0 \le a < b \le 1\}$ to be the set of all open subintervals of $[0, 1]$.

- Consider $\Omega = [0, 1]$. The Borel σ-field, denoted by $\mathcal{B}([0, 1])$, is defined by

$$\mathcal{B}([0, 1]) = \sigma(\mathcal{I}).$$

Then there exists a unique probability measure $P : \mathcal{B}([0, 1]) \to [0, 1]$ such that

$$P(I) = b - a, .$$

for any subinterval $I \subset [0, 1]$ with endpoints $a \le b$. This is a known result in measure theory (see *e.g.* Schilling [1]), but the proof is outside of the scope of this note.

- A similar result can be carried out in the n-dimensional case. Consider $\Omega = [0, 1]^n$. The *Borel σ-field* over $[0, 1]^n = \underbrace{[0, 1] \times \cdots \times [0, 1]}_{n \text{ times}}$, denoted $\mathcal{B}([0, 1]^n)$, is defined by

$$\mathcal{B}([0, 1]^n) = \sigma(\mathcal{I}^n),$$

where $\mathcal{I}^n := \left\{ \prod_{j=1}^n I_j \mid I_j\text{'s are subintervals of } [0, 1]. \right\}$ denotes the set of all rectangles in $[0, 1]^n$. It could be proved (refer again to Schilling [1]) that there exists a unique probability measure $P : \mathcal{B}([0, 1]^n) \to [0, 1]$ such that

$$P\left(\prod_{j=1}^n I_j\right) = (b_j - a_j) \times \cdots \times (b_1 - a_1),$$

for any intervals I_j's having endpoints $0 \le a \le b \le 1$. It is important to note that $P(A)$ coincides with the n-dimensional volume of any event $A \in \mathcal{B}([0, 1]^n)$. ▲

Example 2.18. Consider the Borel construction $([0, 1], \mathcal{B}, P)$.
Calculate the probability that a randomly selected point $\omega \in [0, 1]$ belongs to the set $A = [0, 0.1) \cup (0.4, 0.6) \cup (0.9, 1]$.
Solution. COMPLETE

Example 2.19. Consider the Borel construction $([0, 1]^2, \mathcal{B}, P)$, and let $A = [0, 0.5] \times [0.5, 1] \cup [0.5, 1] \times [0, 0.5]$.
Calculate the probability that a randomly selected point $\omega \in [0, 1]^2$ belongs to the set A.
Solution. COMPLETE

2.2.2. Properties of a probability measure

Now we turn to discuss some basic properties of a probability measure, followed along the way by the sample cases where the properties apply.

We shall begin with the probability of a complemented event which then leads to say that the probability of an impossible event is 0.

Proposition 2.20. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then*
(a) $P(A^{\complement}) = 1 - P(A)$ *for any* $A \in \mathcal{F}$.
(b) $P(\varnothing) = 0$.

Proof. Let $A \in \mathcal{F}$, then we have $\Omega = A \cup A^{\complement}$. Since A and $A^{\complement}$ are disjoint, (P1) and (P2) imply that $1 = P(\Omega) = P(A) + P(A^{\complement})$, proving (a). The item (b) follows from (a) with $A = \Omega$. ∎

Example 2.21. Find the probability of *getting at least one tail* in the trial of 10 coin tosses.

Example 2.22. A company has 120 employees that owns a car. Amont them, 80 of them either use a Toyota or a Honda. Find the probability that an employee does not drive neither Toyota nor Honda.

The next proposition simply says that a larger event implies a larger probability.

Proposition 2.23 (Monotonicity.)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$. Then*

$$A \subset B \implies P(A) \leq P(B).$$

Proof. The conclusion follows since

$$P(A) \leq P(A) + P(B \setminus A) = P(A \cup (B \setminus A)) = P(B).$$ ∎

Example 2.24. In drawing a card randomly from a standard deck, the probability of drawing a *club* is less than of drawing a *black card*. ▲

The following proposition complements the additivity property (P2) in the definition of a probability measure. Here, we do not ask for the sets to be pairwise disjoint.

Proposition 2.25 (Sub-σ-additivity.)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_i\}_{i\in\mathbb{N}}$ be a countable collection of events in $\mathcal{F}$. Then*

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

Proof. Let $B_1 := A_1$ and define B_i by

$$B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$$

for $i \in \mathbb{N}$ with $i \geq 2$. It is clear that $\{B_i\}_{i\in\mathbb{N}}$ is a collection of pairwise disjoint events with

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

This gives

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{\infty} P\left(A_i \setminus \bigcup_{j=1}^{i-1} A_j\right) \leq \sum_{i=1}^{\infty} P(A_i),$$

where the last inequality follows from Proposition 2.23. ∎

Example 2.26. A health dataset collected from 100 patients shows that 10 of them have diabetes and 30 of them have high cholesterol. We may conclude that *no more than* 40 *of them have at least one of the two problems.* ▲

The next two results show the monotonicity of a probability measure when applied respectively to a nondecreasing and a nonincreasing sequence of events.

Proposition 2.27. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\{A_i\}_{i\in\mathbb{N}}$ *be a nondecreasing sequence of events in* $\mathcal{F}$*, i.e.* $A_i \subset A_{i+1}$ *for all* $i \in \mathbb{N}$*. Then*

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

Proof. Since $\{A_i\}_{i\in\mathbb{N}}$ is nondecreasing, we have

$$A_i = A_1 \uplus (A_2 \setminus A_1) \uplus \cdots \uplus (A_i \setminus A_{i-1})$$

for each $i \in \mathbb{N}$. It follows that

$$\bigcup_{i=1}^{\infty} A_i = A_1 \uplus \biguplus_{j=2}^{\infty} (A_j \setminus A_{j-1}).$$

Finally, the σ-additivity condition (P2) implies

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P(A_1) + \sum_{j=2}^{\infty} P(A_i \setminus A_{i-1}) \\ &= \lim_{i\to\infty} \left[P(A_1) + \sum_{j=2}^{i} P(A_i \setminus A_{i-1})\right] \\ &= \lim_{i\to\infty} [P(A_i)] \end{aligned}$$

∎

Corollary 2.28. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\{A_i\}_{i\in\mathbb{N}}$ *be a nonincreasing sequence of events in* $\mathcal{F}$*, i.e.* $A_{i+1} \subset A_i$ *for all* $i \in \mathbb{N}$*. Then*

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

Proof. The proof is left as an Exercise 2.5. ∎

The remaining of this section discusses the Poincaré formula, which allows one to make a computation of those secondary events (which includes those add/or events). We present the formula first for the case of 2 sets, then extend it to the one for 3 sets and finally the general formula for n sets. This formula is, perhaps, one of the most common topics in the SOA's Exam P.

Proposition 2.29 (Poincaré formula)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$ be any two events. Then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Proof. Recall that $A = (A \cap B) \cup (A \cap B^{\complement})$ and the sets $(A \cap B)$ and $(A \cap B^{\complement})$ are disjoint. Then $P(A) = P(A \cap B) + P(A \cap B^{\complement})$. Similarly, we obtain that $P(B) = P(A \cap B) + P(B \cap A^{\complement})$. Combining the two formulae, we get

$$P(A) + P(B) = P(A \cap B) + [P(A \cap B) + P(A \cap B^{\complement}) + P(B \cap A^{\complement})].$$

Since $(A \cap B)$, $(A \cap B^{\complement})$, and $(B \cap A^{\complement})$ are pairwise disjoint and their union is $A \cup B$, we have arrived at

$$P(A) + P(B) = P(A \cap B) + P(A \cup B),$$

which completes the proof. ∎

Proposition 2.30 (Poincaré formula for 3 sets)**.** *Let A_1, A_2, A_3 be three events. Then*

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) =& P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Proof. Follows from Proposition 2.29. ∎

Let us see the use of the Poincaré formulae (both Propositions 2.29 and 2.30) in the following examples.

Example 2.31 (SOA Exam P Sample Questions)**.** The probability that a visit to a primary care physician's (PCP) office results in neither labwork nor referral to a specialist is 35%. Of those coming to a PCP's office, 30% are referred to specialists and 40% require lab work. Calculate the probability that a visit to a PCP's office results in both lab work and referral to a specialist.

<u>Solution</u>. Let Ω be the set of visitors of PCP's office and $A, B \subset \Omega$ refers, respectively, to the set of those who require lab work and those who are referred to specialists. The σ-field here is probably $\mathcal{F} = \sigma(\{A, B\})$ (but this is not of interests for the question). We aim to calculate $P(A \cap B)$. From the given information, we get

$$P(A) = 0.4, \quad P(B) = 0.3, \quad P((A \cup B)^{\complement}) = 0.35.$$

It is implies that $P(A \cup B) = 1 - 0.35 = 0.65$. The Poincaré formula (Proposition 2.29), it follows that

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.65 = 0.05. \qquad \blacktriangle$$

Example 2.32 (SOA Exam P Sample Questions)**.** A survey of a group's viewing habits over the last year revealed the following information:

(i) 28% watched gynmastics
(ii) 29% watched baseball
(iii) 19% watched soccer
(iv) 14% watched gynmastics and baseball
(v) 12% watched baseball and soccer
(vi) 10% watched gynmastics and soccer
(vii) 8% watched all three sports.

Calculate the percentage of the group that watched none of the three sports during the last year.

Solution. Take Ω to be the set of all people in the group. Set A, B, and C to be subsets of Ω consisting of people in the group that watched gymnastics, baseball, and soccer, respectively. The σ-field here is probably $\mathcal{F} = \sigma(\{A, B, C\})$ (but it is not of interests here). The question aims to calculate the probability of $(A \cup B \cup C)^{\complement}$. From the information of the question, we deduce that

$$\begin{aligned} &P(A) = 0.28, && P(B) = 0.29, && P(C) = 0.19, \\ &P(A \cap B) = 0.14, && P(B \cap C) = 0.12, && P(A \cap C) = 0.1, && P(A \cap B \cap C) = 0.08. \end{aligned}$$

Using the Poincaré formula (Proposition 2.30), we obtain $P(A \cup B \cup C) = 0.48$ and finally we get

$$P((A \cup B \cup C)^{\complement}) = 1 - P(A \cup B \cup C) = 0.52. \qquad \blacktriangle$$

The following is the general Poncaré formula for n events. The proof could be established using mathematical induction, which we shall omit in this lecture note.

Theorem 2.33. *Let $A_1, \ldots, A_n$ be events. Then*

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

2.3. Zero-probability events and almost sure events

In the following, we make a counter-intuitive remark on events with zero probability.

Definition 2.34. Consider a probability space $(\Omega, \mathcal{F}, P)$. An event $A \in \mathcal{F}$ that has zero probability, *i.e.* $P(A) = 0$, is called a *negligible* event. ▲

Despite the simple statement, a zero-probability event *may* indeed occur as described in the following example.

Example 2.35. Consider a wood stick of a unit length, represented by $\Omega = [0, 1]$ and let a person blindly pinpoint a position $\omega \in \Omega$ on the stick. This resembles an instance of Example 2.17 with $n = 1$. We may consider the Borel σ-field $\mathcal{B}(\Omega)$ and a probability P over $(\Omega, \mathcal{B}(\Omega))$ in such a way that, for any interval $[a, b] \subset \Omega$, we have $P([a, b]) = b - a$. This means the probability that the pointed position ω belongs to a position $[a, b]$ is computed by the proportion of the length of $[a, b]$ to the whole unit length. One could simply observe that $P(\{\omega\}) = 0$ for any $\omega \in \Omega$, regardless the fact that one of them must be realized. ▲

The key takeaway from this paradoxical example is that zero-probability events *do* happen. In fact, this is a typical property when one considers continuous outcomes. This means the concept of negligibility is not the same as impossibility, and implies that *events with probability* 1 *are not certain events* eiter. This motivates the following terminology.

Definition 2.36. Let $(\Omega, \mathcal{F}, P)$ be a probability space. An event $A \in \mathcal{F}$ is called an *almost sure* (briefly *a.s.*) event if its complement $A^{\complement}$ is negligible. We also say that an event A occurs *almost surely* (also abbreviated to *a.s.*). ▲

Corollary 2.37. *Every a.s. event has a probability of* 1*.*

Proof. Follows from Proposition 2.20. ∎

To correctly understand zero-probability events, we look through the *frequentist* view:

> *A probability of an event conveys to an infinite number of independent trials as a limit of the ratio* $\dfrac{\#\textit{occurence}}{\#\textit{trials}}$ *as* $\#\textit{trials} \to \infty$.

With this intuition, a negligible event is simply an event in which this ratio tends to 0.

2.4. Independent events and conditional probability

Let us start with a definition.

Definition 2.38. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then the two events $A, B \in \mathcal{F}$ are called *independent* if

$$P(A \cap B) = P(A)P(B).$$ ▲

To illustrate the idea, we consider some examples.

Example 2.39. Recall Example 1.9. The events $\{X_1 = \mathsf{H}\}$ and $\{X_2 = \mathsf{H}\}$ are independent, as we have

$$P(\{X_1 = \mathsf{H}\} \cap \{X_2 = \mathsf{H}\}) = P(\{(\mathsf{H}, \mathsf{H})\}) = \frac{1}{4}$$

and

$$P(\{X_1 = \mathsf{H}\})P(\{X_2 = \mathsf{H}\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

We can similarly show that the events $\{X_1 = \mathsf{T}\}$ and $\{X_2 = \mathsf{T}\}$ are independent. On the other hand, the events $\{X_1 = \mathsf{H}\}$ and $\{X_1 = \mathsf{T}\}$ as well as $\{X_2 = \mathsf{H}\}$ and $\{X_2 = \mathsf{T}\}$ are not independent. ▲

Example 2.40. Recall Example 1.11. The events Odd and Even are not independent, as one may see that

$$P(\mathsf{Odd} \cap \mathsf{Even}) = P(\varnothing) = 0$$

while we have

$$P(\mathsf{Odd})P(\mathsf{Even}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Similarly, we can observe that neither two of the events Odd, Even, High and Low are independent. ▲

Example 2.41 (SOA Exam P Sample Question)**.** An actuary studying the insurance preferences of automobile owners makes the following conclusions:

(i) An automobile owner is twice as likely to purchase collision coverage as disability coverage.

(ii) The event that an automobile owner purchases collision coverage is independent of the event that he or she purchases disability coverage.

(iii) The probability that an automobile owner purchases both collision and disability coverages is 0.15.

Calculate the probability that an automobile owner purchases neither collision nor disability coverage.

Solution. Let A and B be events that an automobile owner purchases collision coverage and disability coverage, respectively. We know from the question that $P(A) = 2P(B)$ and that $P(A \cap B) = 0.15$. Since A and B are independent, it follows from Definition 2.38 that $0.15 = P(A \cap B) = P(A)P(B) = 2P(B)^2$. We obtain $P(B) = \sqrt{\frac{0.15}{2}}$ and consequently

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3P(B) - P(A \cap B) \approx 0.67.$$

Finally, it is concluded that $P((A \cup B)^{\complement}) = 1 - P(A \cup B) \approx 0.33$. ▲

We now turn to the concept of a conditional probability and some illustrative examples.

Definition 2.42. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$ are two events with $P(B) > 0$. The *conditional probability of A given B* is the quantity

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$ ▲

Example 2.43. A survey group consists of 1000 people. Among them, 45% are male. Of all the people in this group, 120 males and 180 females used to have COVID-19. What is the probability that an infected member is a female ?
Solution. Let Ω be the sample space containing all the people in the survey group, and let $\mathcal{F} = 2^\Omega$ be its σ-field of events. We call M and F the sets of all male and female members, respectively, and C the set of members who used to have COVID-19. The question asked for $P(F|C)$, which could be calculated by

$$P(F|C) = \frac{P(F \cap C)}{P(C)} = \frac{0.18}{0.3} = \frac{3}{5}. \qquad \blacktriangle$$

For convenience, we consider an arbitrary probability space $(\Omega, \mathcal{F}, P)$ in the subsequent results unless otherwise specified. **Consider moving this**

Proposition 2.44. *Let A and B be two events with $P(B) > 0$. Then A and B are independent if and only if $P(A|B) = P(A)$.*

Proof. Directly follows from the definition. ∎

Theorem 2.45 (Bayes' rule of retrodiction)**.** *Let A and B be two events with $P(A), P(B) > 0$. Then*

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

Proof. Since $P(A \cap B) = P(A|B)P(B)$, we get

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}. \qquad \blacksquare$$

The above Bayes' rule (also known as the Bayes' theorem) plays a very important role in modern statistics, data science, and other data-driven subjects. It describes how an *arrival of new information* could impact the evaluation of a probability. Let us demonstrate such use in the following example.

Example 2.46. Suppose that a population is affected by a disease. Suppose that a probability of a person to be infected is known to be $P(\mathsf{Infected}) = 0.3$. The probability of a person to have a positive test result (including both true and false positives) from an self-test kit is $P(\mathsf{Positive}) = 0.32$, while the self-test kit is known to be 85% accurate on infected patients *i.e.* $P(\mathsf{Positive}|\mathsf{Infected}) = 0.9$.

Now, we repeat that, without knowing the self-test result, we only know that the probability of a person to be infected is $P(\mathsf{Infected}) = 0.3$. If this person has a positive self-test result (here comes a new piece of information), then the probability is updated to

$$P(\mathsf{Infected}|\mathsf{Positive}) = \frac{P(\mathsf{Positive}|\mathsf{Infected})P(\mathsf{Infected})}{P(\mathsf{Positive})} = \frac{0.9 \times 0.3}{0.32} = 0.84375.$$

This shows how an arrival of a new information could improve the confidence level of an event, using the Bayes' rule. ▲

2.5. Bayes' rule for partitions and scenario trees

In this section, we consider another important consequence of the Bayes' rule, which describes how a probability of an event can be decomposed into its partition. It is also the theorem behind the probability computation using a scenario tree.

Theorem 2.47 (Bayes' rule for partitions). *Let $\{B_i\}_{i\in\mathbb{N}}$ be a family of events that forms a partition of Ω, i.e. the collection $\{B_i\}_{i\in\mathbb{N}}$ is pairwise disjoint and $\Omega = \biguplus_{i=1}^{\infty} B_i$. Suppose that $P(B_i) > 0$ for all $i \in \mathbb{N}$. The the following formula holds*

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$$

for any event A.

Proof. Since $\{B_i\}_{i\in\mathbb{N}}$ is pairwise disjoint, so as the family $\{A \cap B_i\}_{i\in\mathbb{N}}$. It is also true that $A = \biguplus_{i=1}^{\infty}(A \cap B_i)$. Using the σ-additivity of P, we obtain

$$P(A) = P\left(\biguplus_{i=1}^{\infty}(A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i). \qquad \blacksquare$$

Let us illustrate how the above theorem is used in a scenario tree technique in the following examples. We start with a very simple problem of two consecutive coin flippings and card drawings.

Example 2.48. Consider the problem of flipping a fair coin twice, and we are looking for the event where a head (H) shows up in at least one of the two flips. The sample space is simply

$$\Omega := \{(\mathsf{H},\mathsf{H}),(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H}),(\mathsf{T},\mathsf{T})\}$$

and the σ-field of events is the whole power set, *i.e.* $\mathcal{F} := 2^{\Omega}$. The probability P of each event is computed based on the counting principle

$$P(A) := \frac{\#A}{\#\Omega}.$$

The event that we are after is represented by

$$A^* := \{(\mathsf{H},\mathsf{H}),(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H})\},$$

so the counting principle clearly yields $P(A^*) = \dfrac{3}{4}$.

Next, we show an alternative approach of computing $P(A^*)$ by using a scenario tree as shown below:

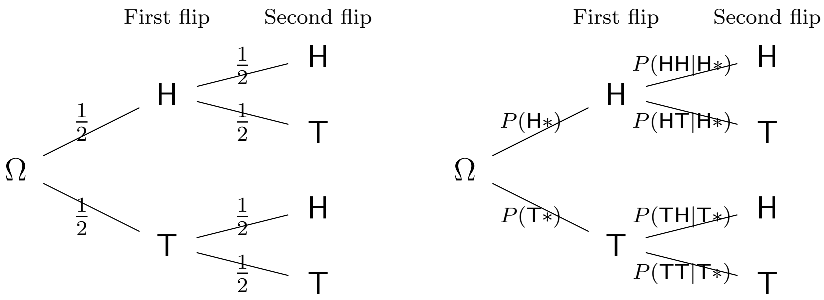

Figure 5.1: Scenario tree of flipping a coin twice.

Explanation...

Example 2.49. Consider drawing two cards, one after another, from a full deck of standard playing cards. Find the probability that one of the two cards is a Queen of Hearts.

A standard deck of playing cards consists of 52 cards. The deck is divided into 4 suits, namely, Clubs (C), Diamonds (D), Hearts (H), and Spades (S). Each suit consists of 13 card faces, namely, $\mathsf{A}, \mathsf{2}, \mathsf{3}, \mathsf{4}, \mathsf{5}, \mathsf{6}, \mathsf{7}, \mathsf{8}, \mathsf{9}, \mathsf{10}, \mathsf{J}, \mathsf{Q}, \mathsf{K}$. We shall specify a card with a format of $\mathbb{FS}$, where $\mathbb{F}$ and $\mathbb{S}$ represents respectively the suit and face. For example, the Queen of Hearts is written QH.

The sample space of this problem is

$$\Omega := \left\{ (\mathbb{F}_1\mathbb{S}_1, \mathbb{F}_2\mathbb{S}_2) \;\middle|\; \begin{array}{c} \mathbb{F}_1, \mathbb{F}_2 \in \{\mathsf{A}, \mathsf{2}, \mathsf{3}, \mathsf{4}, \mathsf{5}, \mathsf{6}, \mathsf{7}, \mathsf{8}, \mathsf{9}, \mathsf{10}, \mathsf{J}, \mathsf{Q}, \mathsf{K}\} \\ \mathbb{S}_1, \mathbb{S}_2 \in \{\mathsf{C}, \mathsf{D}, \mathsf{H}, \mathsf{S}\} \\ \mathbb{F}_1\mathbb{S}_2 \neq \mathbb{F}_2\mathbb{S}_2 \end{array} \right\},$$

while the σ-field of events can simply be the power set, i.e. $\mathcal{F} := 2^{\Omega}$. The probability $P(A)$ of an event A is defined based on the counting principle

$$P(A) = \frac{\#A}{\#\Omega}.$$

Of course, the probability of having QH in one of the two draws corresponds to the event A^* defined by

$$A^* := A_1^* \uplus A_2^*, \quad A_1^* := \{(\mathsf{QH}, *) \,|\, * \neq \mathsf{QH}\}, \quad A_2^* := \{(*, \mathsf{QH}) \,|\, * \neq \mathsf{QH}\},$$

which gives $P(A^*) = \dfrac{\#A^*}{\#\Omega} = \dfrac{\#A_1^* + \#A_2^*}{\#\Omega} = \dfrac{51 + 51}{52 \times 51}$.

Now, we frame the problem in a more natural way (the structure of A^* is a hint in itself). In the first draw, we partition the possible outcomes into two disjoint cases — *getting* QH (the event A_1^*) and *not getting* QH (the event $(A_1^*)^{\complement}$). Next we focus on the second draw given that A_1^* or $(A_1^*)^{\complement}$ occurs. If A_1^* is known to occur, A^* is already accomplished so that

$$P(A^*|A_1^*) = 1.$$

On the other hand, if $(A_1^*)^\complement$ occurs, the occurence of A^* is equivalent to that the outcome of the second draw is QH. Hence we come up with

$$P(A^*|(A_1^*)^\complement) = P(\text{The second draw is } \mathsf{QH}.) = \frac{1}{51}.$$

Finally we conclude, using Theorem 2.47, that

$$P(A^*) = P(A^*|A_1^*)P(A_1^*) + P(A^*|(A_1^*)^\complement)P((A_1^*)^\complement) = \frac{1}{52} \times 1 + \frac{51}{52} \times \frac{1}{51} = \frac{51+51}{52 \times 51}. \quad \blacktriangle$$

The best way to actually come up with the latter calculation is to use a *scenario tree.*

Next, we demonstrate another way to utilize the Bayes' rule in actuarial situations.

Example 2.50 (SOA Exam P Sample Quesiton)**.** An auto insurance company isures drivers of all ages. An actuary compiled the following statistics on the company's insured drivers:

Age of Driver	Probability of Accident	Portion of Company's Insured Drivers
16–20	0.06	0.08
21–30	0.03	0.15
31–65	0.02	0.49
66–99	0.04	0.28

A randomly selected driver that the company insures has an accident. Calculate the probability that the driver was age 16–20.

Solution. Let Ω be the sample space containing all insured drivers and let A_1, A_2, A_3, A_4 be the event that a driver's age is 16–20, 21–30, 31–65, 66–99, respectively. It holds that $\Omega = A_1 \uplus A_2 \uplus A_3 \uplus A_4$.

Let B be the event that an insured driver has an accident. Then the table shows

$$\begin{aligned} P(A_1) &= 0.08, & P(B|A_1) &= 0.06, \\ P(A_2) &= 0.15, & P(B|A_2) &= 0.03, \\ P(A_3) &= 0.49, & P(B|A_3) &= 0.02, \\ P(A_4) &= 0.28, & P(B|A_4) &= 0.04. \end{aligned}$$

Using the Bayes' formula (Theorem 2.45), we have

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)}.$$

We have from Theorem 2.47 also

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4).$$

Combining the two, we obtain

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + P(B|A_4)P(A_4)} \\ &\approx 0.1584. \end{aligned}$$ ▲

Example 2.51 (SOA Exam P Sample Question)**.** An insurance company issues life insurance policies in three separate categories: Standard, Preferred, and Ultra-preferred. Of the company's policyholders, 50% are standard, 40% are preferred, and 10% are ultra-preferred. Each standard policyholder has probability 0.01 of dying in the nex year, each preferred policyholder has probability 0.005 of dying in the next year, and each ultra-preferred policyholder has probability 0.001 of dying in the next year.

A policyholder dies in the next year. Calculate the probability that the deceased policyholder was ultra-preferred.

Solution. We construct the following table to summarize the information given in the question:

Categories	Percentage of policyholders	Probability of dying
Standard	50%	0.01
Preferred	40%	0.005
Ultra-preferred	10%	0.001

Let A_1, A_2, and A_3 denote the events that a policyholder is Standard, Preferred, and Ultra-preferred, respectively. If Ω is the sample space containing all policyholders, then $\Omega = A_1 \uplus A_2 \uplus A_3$.

Let B be the event that a policyholder dies in the next year. Then we have

$$\begin{aligned} P(A_1) &= 0.5, & P(B|A_1) &= 0.01, \\ P(A_2) &= 0.4, & P(B|A_2) &= 0.005, \\ P(A_3) &= 0.1, & P(B|A_3) &= 0.001. \end{aligned}$$

Using Theorems 2.45 and 2.47, it follows that

$$\begin{aligned} P(A_3|B) &= \frac{P(B|A_3)P(A_3)}{P(B)} \\ &= \frac{P(B|A_3)P(A_3)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &\approx 0.0141. \end{aligned}$$ ▲

Practice problems 2

2.1. Prove Proposition 2.5.

2.2. Prove Proposition 2.6.

2.3. Suppose that $\Omega' = \{0, 1, 2\}$ be equipped with a σ-field $\mathcal{F}' = 2^{\Omega}$. Let $\Omega = \{(\mathsf{H}, \mathsf{H}), (\mathsf{H}, \mathsf{T}), (\mathsf{T}, \mathsf{H}), (\mathsf{T}, \mathsf{T})\}$ and define a map $f : \Omega \to \Omega'$ by

$$f((\mathsf{H}, \mathsf{H})) = 2, \quad f((\mathsf{H}, \mathsf{T})) = f((\mathsf{T}, \mathsf{H})) = 1, \quad f((\mathsf{T}, \mathsf{T})) = 0.$$

Calculate the σ-field $\mathcal{F}_f$.

2.4. Prove Proposition 2.13.

2.5. Prove Corollary 2.28.

2.6. Give a detailed proof of Proposition 2.30.

2.7. A doctor is studying the relationship between blood pressure and heartbeat abnormalities in her patients. She tests a random sample of her patients and notes their blood pressures (high, low, or normal) and their heartbeats (regular or irregular). She finds that:

(i) 14% have high blood pressure.
(ii) 22% have low blood pressure.
(iii) 15% have an irregular heartbeat.
(iv) Of those with an irregular heartbeat, one-third have high blood pressure.
(v) Of those with normal blood pressure, one-eighth have an irregular heartbeat.

Calculate the portion of the patients selected who have a regular heartbeat and low blood pressure.

2.8. A public health researcher examines the medical records of a group of 937 men who died in 1999 and discovers that 210 of the men died from causes related to heart disease. Moreover, 312 of the 937 men had at least one parent who suffered from heart disease, and, of these 312 men, 102 died from causes related to heart disease. Calculate the probability that a man randomly selected from this group died of causes related to heart disease, given that neither of his parents suffered from heart disease.

2.9. An auto insurance company has 10,000 policyholders. Each policyholder can be classfied as

(i) young or old;
(ii) male or female; and
(iii) married or single.

Of these policyholders, 3000 are young, 4600 are male, and 7000 are married. The policyholders can also be classified as 1320 young males, 3010 married males, and 1400 young married persons. Finally, 600 of the policyholders are Young Married Male persons.

Calculate the number of the company's policyholders who are young, female, and single.

2.10. Upon arrival at a hospital's emergency room, patients are categorized according to their condition as critical, serious, or stable. In the past year:

(i) 10% of the emergency room patients were critical;
(ii) 30% of the emergency room patients were serious;
(iii) the rest of the emergency room patients were stable;
(iv) 40% of the critical patients died;
(v) 10% of the serious patients died; and
(vi) 1% of the stable patients died.

Given that a patient survived, calculate the probability that the patient was categorized as serious upon arrival.

Chapter 3. Discrete random variables

A random variable is a representation of an uncertain event, often in a numerical manner. Eventhough a random variable is not really a variable, it is usually treated as if it is a variable, but its value depends on a random event. We actually have used the idea of a random variable before, without realizing or calling it a random variable.

3.1. Definitions and examples

We have already exploited unofficially the concept of a random variable earlier in various examples. Let us now give state its formal definition of a discrete variable.

Definition 3.1. Let $(\Omega, \mathcal{F})$ be a measurable space. A function $X : \Omega \to E$, with a countable set E, is called a *discrete random variable* (or briefly a *discrete r.v.*) on $(\Omega, \mathcal{F})$ if the preimage

$$\{X = x\} := X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}$$

is an event in $\mathcal{F}$, for all $x \in E$. ▲

Remark. The following remarks are in order.

- The countable set E could be any abstract set. In many applications (including *dice rolling* or *counting heads* in coin tosses), it is natural to set $E = \mathbb{Z}$ (or more commonly, $E = \mathbb{N}$) which has an advantage of being computable. In some other applications (like *head-or-tail* or *card faces*), there is no natural way to represent an event numerically. One would see that having $E = \mathbb{Z}$ (or $E = \mathbb{N}$) has a huge benefit that allows the study of *expectation*.
- The discrete (or later, continuous) nature of a random variable is usually understood directly from the context (the countability of E) and the term *random variable* (or *r.v.*) is usually used with its prefix omitted when there is no possible confusion. ▲

In this chapter, every r.v.s are referred to the discrete ones. The following examples shall illustrate the use of a random variable. The first example of rolling a dice is perhaps the most natural one, since we have already interpret outcomes that are represented by numbers.

Example 3.2. Consider rolling a dice. The sample space is then $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$ and the σ-field of all events is the power set $\mathcal{F} = 2^{\Omega}$. We define $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$, representing numerically an outcome, by

$$X(\epsdice{1}) = 1, \quad X(\epsdice{2}) = 2, \quad X(\epsdice{3}) = 3, \quad X(\epsdice{4}) = 4, \quad X(\epsdice{5}) = 5, \quad X(\epsdice{6}) = 6. \tag{3.1}$$

Then X is a r.v. ▲

In the following example, we still consider rolling a dice but we are only able to distinguish odd and even faces. One would see that the same r.v. function is no longer available as a r.v.

Example 3.3. Consider $\Omega = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$ and $\mathcal{F} = \{\varnothing, \Omega, \{\epsdice{1}, \epsdice{3}, \epsdice{5}\}, \{\epsdice{2}, \epsdice{4}, \epsdice{6}\}\}$. Then the function $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ defined as (3.1) is not a r.v. For instance, one could notice that $X^{-1}(1) = \{\epsdice{1}\} \notin \mathcal{F}$. ▲

The next is an example of a r.v. that does not take numerical values.

Example 3.4. Consider tossing a coin twice, so that $\Omega = \{(\mathsf{H}, \mathsf{H}), (\mathsf{H}, \mathsf{T}), (\mathsf{T}, \mathsf{H}), (\mathsf{T}, \mathsf{T})\}$ and $\mathcal{F} = 2^\Omega$. Set $E = \{\mathsf{H}, \mathsf{T}\}$ and define $X_1, X_2 : \Omega \to E$ by

$$\begin{aligned}
X_1((\mathsf{H}, \mathsf{H})) = X_1((\mathsf{H}, \mathsf{T})) = \mathsf{H}, \\
X_1((\mathsf{T}, \mathsf{H})) = X_1((\mathsf{T}, \mathsf{T})) = \mathsf{T}, \\
X_2((\mathsf{H}, \mathsf{H})) = X_2((\mathsf{T}, \mathsf{H})) = \mathsf{H}, \\
X_2((\mathsf{H}, \mathsf{T})) = X_2((\mathsf{T}, \mathsf{T})) = \mathsf{T}.
\end{aligned}$$

We could see that X_1 and X_2 are r.v.s representing the outcomes of the first and second toss, respectively. These r.v.s are examples of those in which numerical representation is not natural. ▲

Under some circumstances, it also makes sense to assign numerical values to these non-numerical outcomes. We shall see this in the following example.

Example 3.5. Consider tossing a coin twice and let Ω be the sample space with $\mathcal{F} = 2^\Omega$ being its σ-field of events. We shall now consider a r.v. X that represents the number of H occuring in the two tosses. This X is defined by

$$X((\mathsf{T}, \mathsf{T})) = 0, \quad X((\mathsf{H}, \mathsf{T})) = X((\mathsf{T}, \mathsf{H})) = 1, \quad X((\mathsf{H}, \mathsf{H})) = 2.$$

With this notion, we could consider probabilities like

$$P(X = 0) = P(\{X = 0\}) = P(\{\omega \in \Omega \mid X(\omega) = 0\}) = P(\{(\mathsf{T}, \mathsf{T})\}) = \frac{1}{4}.$$ ▲

Example 3.6. Consider rolling a dice twice, and let $\Omega = \{(i, j) \mid i, j \in \{1, 2, \ldots, 6\}\}$ be equipped with the σ-field $\mathcal{F} = 2^\Omega$. We shall construct a r.v. representing the difference of the outcomes of the two dices. Hence, we define $X : \Omega \to \{1, 2, 3, 4, 5\}$ by

$$X((i, j)) = |i - j|.$$ ▲

Notice that the gross σ-fields, $\mathcal{F} = 2^{\Omega}$, were considered in all the examples above. This guarantees that all the sets $\{X = x\}$ belongs to $\mathcal{F}$. If this is not the case, then an explicit proof justifying the condition is necessary. We emphasize this fact in the following example.

Example 3.7. Let us consider again the situation of tossing two identical coins simultaneously. Recall that the sample space is $\Omega = \{(\mathsf{H},\mathsf{H}),(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H}),(\mathsf{T},\mathsf{T})\}$ and it is equipped with the σ-field $\mathcal{F} = \{\varnothing,\Omega,\{(\mathsf{H},\mathsf{H})\},\{(\mathsf{T},\mathsf{T})\},\{(\mathsf{H},\mathsf{T}),(\mathsf{T},\mathsf{H})\}\}$. Consider the following function $X_1, X_2 : \Omega \to \mathbb{Z}$ defined by

$$X_1((\mathsf{H},\mathsf{H})) = X_1((\mathsf{T},\mathsf{T})) = 0, \quad X_1((\mathsf{H},\mathsf{T})) = X_1((\mathsf{T},\mathsf{H})) = 1,$$

and

$$X_1((\mathsf{H},\mathsf{H})) = X_1((\mathsf{T},\mathsf{T})) = 0, \quad X_1((\mathsf{H},\mathsf{T})) = -1, \quad X_1((\mathsf{T},\mathsf{H})) = 1.$$

Then X_1 is a random variable, while X_2 is not. ▲

The following result hints some construction of a new r.v. from the given ones.

Theorem 3.8. *Let $(\Omega,\mathcal{F})$ be a measurable space, E and F be two countable sets. If $X : \Omega \to E$ is a r.v. and $f : E \to F$, then $Y := f(X) = f \circ X$ is also a r.v.*

Proof. Let $y \in F$. Observe that

$$\begin{aligned} Y^{-1}(\{y\}) &= \{\omega \in \Omega \mid f(X(\omega)) = y\} \\ &= \bigcup_{x \in f^{-1}(\{y\})} X^{-1}(\{x\}). \end{aligned} \tag{3.2}$$

Since $f^{-1}(\{y\}) \subset E$, it is countable and (3.2) shows that $Y^{-1}(\{y\})$ is a countable union of events in $\mathcal{F}$. This shows that Y is a r.v. ∎

Theorem 3.9. *Let $(\Omega,\mathcal{F})$ be a measurable space, E_1 and E_2 be two countable sets, and $X_1 : \Omega \to E_1$ and $X_2 : \Omega \to E_2$ are two r.v.s. Then $X := (X_1, X_2) : \Omega \to E_1 \times E_2$ is also a r.v.*

Proof. Let $x = (x_1, x_2) \in E_1 \times E_2$, then

$$\begin{aligned} X^{-1}(\{x\}) &= \{\omega \in \Omega \mid X_1(\omega) = x_1,\ X_2(\omega) = x_2\} \\ &= \{\omega \in \Omega \mid X_1(\omega) = x\} \cap \{\omega \in \Omega \mid X_2(\omega) = x_2\} \\ &= X_1^{-1}(\{x_1\}) \cap X_2^{-1}(\{x_2\}) \ \in \mathcal{F}. \end{aligned}$$ ∎

Now that we translate events into r.v.s, one could also speak of probability. Consider a probability space $(\Omega,\mathcal{F},P)$ and take any r.v. $X : \Omega \to E$, for some countable set E. We could consider the gross σ-field 2^E and on this measurable space $(E, 2^E)$ we define the *push-forward* probability measure, $P_X : 2^E \to [0,1]$, by

$$P_X(B) := P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\}).$$

When $B = \{x\}$ is a singleton, we usually write $P(X = x)$ instead of $P_X(\{x\})$. This defines a *distribution function*.

Definition 3.10. Let $X : \Omega \to E$ be a discrete r.v., then its *distribution function* of X is the function $\pi : E \to [0, 1]$ defined by

$$\pi(x) := P(X = x). \qquad \blacktriangle$$

3.2. Independent and Identically Distributed Sequences

The independence notion for events translates naturally into the setting of r.v.s. Unless otherwise specified, suppose throughout the section that the r.v.s are defined on a given probability space.

Definition 3.11. Let $(\Omega, \mathcal{F})$ be a measurable space, E_1 and E_2 are two countable sets. Two r.v.s $X_1 : \Omega \to E_1$ and $X_2 : \Omega \to E_2$ are called *independent* if

$$P(X_1 = x_1,\ X_2 = x_2) := P(\{X_1 = x_1\} \cap \{X_2 = x_2\}) = P(X_1 = x_1)P(X_2 = x_2),$$

for all $x_1 \in E_1$ and all $x_2 \in E_2$.

This concept of independence generalizes naturally to any finite case.

Definition 3.12. Let $(\Omega, \mathcal{F})$ be a measurable space, $E_1, \dots, E_n$ are two countable sets. For each $i = 1, \dots, n$, let $X_i : \Omega \to E_i$ be r.v.s. We say that $X_1, \dots, X_n$ are *independent* if

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n),$$

for all $x_i \in X_i$ and $\ i = 1, \dots, n$.

Example 3.13. Consider the switching network shown in Figure 2.1, where there are two switches, s_1 and s_2. Each switch has two states, namely, Working and Failed. The switches s_1 and s_2 have failure rates of 0.02 and 0.05, respectively.

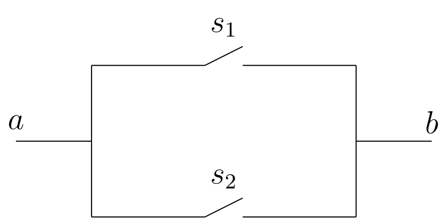

Figure 2.1: Switching network with two switches.

In order to have the electricity flows between points a and b, at least one of the two switches must be in the Working state. What is the probability that no electricity flows between a and b ?

Solution. For $i = 1, 2$, let X_i be the r.v. describing the position of the switch s_i, *i.e.*, $X_i = \mathsf{Working}$ ($X_i = \mathsf{Failed}$) if the switch s_i is in the Working (Failed) position. Then $P(X_1 = \mathsf{Failed}) = 0.02$ and $P(X_2 = \mathsf{Failed}) = 0.05$. The failure of each switche is independent of one another, so we have

$$P(X_1 = \mathsf{Failed},\ X_2 = \mathsf{Failed}) = P(X_1 = \mathsf{Failed})P(X_2 = \mathsf{Failed}) = 0.02 \times 0.05.$$

The electricity flows between a and b if both switches do not fail. This is the complement of $\{X_1 = \mathsf{Failed},\ X_2 = \mathsf{Failed}\}$, therefore has a probability of

$$1 - P(X_1 = \mathsf{Failed},\ X_2 = \mathsf{Failed}) = 1 - 0.001 = 0.999. \qquad \blacktriangle$$

Example 3.14. Consider the switching network shown in Figure 2.2, where there are four switches, s_1, s_2, s_3, s_4. Each switch can be in one of the two positions, On or Off, with a probability of $\frac{1}{2}$ of being in the On position.

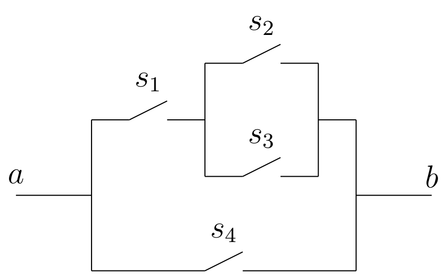

Figure 2.2: Switching network with four switches.

What is the probability that electricity flows between a and b ?
Solution. COMPLETE

The concept of independence could be even extended to sequences of r.v.s.

Definition 3.15. A sequence (X_n) of discrete r.v.s is said to be *independent* if any finite collection $X_{i_1}, \cdots, X_{i_r}$ extracted from this sequence are independent.

Definition 3.16. A sequence (X_n) of discreate r.v.s is said to be *independent and identically distributed* (briefly *i.i.d.*) if

- all X_n's take their values in the same set E,
- the sequence (X_n) is independent as of Definition 3.15,
- the probability distribution functions of all X_n's are the same.

Natural examples of i.i.d. sequences are those related to infinite random experiments.

Example 3.17. Consider tossing a single coin repeatedly and indefinitely. If X_n is the r.v. representing the outcome of the n^{th} toss, prove that the sequence (X_n) is i.i.d.
Solution. COMPLETE

The following is a more advanced example raised in indefinite coin tossing. It also serves as a counterexample showing that the sum of r.v.s in an i.i.d. sequence is no longer i.i.d.

Example 3.18. Consider again an indefinite coin tossing where (X_n) is an i.i.d. sequence of r.v. and for each $n \in \mathbb{N}$, X_n represents the outcome of the n^{th} toss (0 representing T and 1 representing H). Now, for each $n \in \mathbb{N}$, we consider an r.v. $S_n : \Omega \to \mathbb{R}$ defined by

$$S_n = X_1 + \cdots + X_n.$$

This r.v. is the number of heads appearing during the first n tosses.

(a) Show that (S_n) is not i.i.d.

(b) Compute the probability of the event $S_n = k$ for $k = \{1, \dots, n\}$.

Solution. COMPLETE

3.3. Expectation and variance

Expectation of a r.v. could be thought of as an *average* or a *mean* of all possible outcomes weighted by the probability of the given value. To be able to make an average, it is required either that the r.v. takes numeric values or that there is a function that maps the r.v. images into numeric values.

Again, always consider a probability space $(\Omega, \mathcal{F}, P)$.

Definition 3.19. Let X be a discrete r.v. with values in $E \subset \mathbb{R}$ with the condition that

$$\text{either} \quad E \subset \mathbb{R}_+ \quad \text{or} \quad \sum_{x \in E} |x| \, P(X = x) < \infty. \tag{3.3}$$

Then the *expectation* (or *expected value*) of X, denoted with $\mathbb{E}X$ (or $\mathbb{E}[X]$), is defined by

$$\mathbb{E}X := \sum_{x \in E} x P(X = x). \qquad \blacktriangle$$

Remark. If X is a general discrete r.v. with values in E (not necessarily a subset in $\mathbb{R}$), then we have to map its image into $\mathbb{R}$ with some function $g : E \to \mathbb{R}$ with the condition that

$$\text{either} \quad g(x) \geq 0 \text{ for all } x \in E \quad \text{or} \quad \sum_{x \in E} |g(x)| \, P(X = x) < \infty. \tag{3.4}$$

This means $g(X)$ is a discrete r.v. satisfyin (3.3), and hence we may deduce its expectation

$$\mathbb{E}[g(X)] = \sum_{x \in E} g(x) P(X = x). \qquad \blacktriangle$$

Note that the conditions (3.3) and (3.4) are adopted to guarantee that the sum in the definition of an expectation is well-defined. From now on, a r.v. enjoying such a property is said to be *integrable*. It should be noted that an integrable r.v. could still have an infinite expectation.

Let us first consider the numeric case of dice rolling.

Example 3.20. Consider coin tossing from Example 3.2 and the constructions therein. Then we have

$$\mathbb{E}X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

One should note that the value of $\mathbb{E}X$ might not be included in the image of X. ▲

Example 3.21. Consider the story of Example 3.18 and take the r.v. S_n. Calculate the expected value $\mathbb{E}[S_n]$.
Solution. COMPLETE

An *indicator function* of an event is considered a useful trick concerning the theory of expection.

Example 3.22. The the indicator function of an event A is the function $\mathbf{1}_A : \Omega \to \{0, 1\}$ defined by

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Prove that $\mathbf{1}_A$ is a r.v. and that $\mathbb{E}[\mathbf{1}_A] = P(A)$.
Solution. COMPLETE

In the next example, we show that an r.v. taking finite values could have an infinite expectation.

Example 3.23. Let X be a r.v. taking values in $E = \mathbb{N}$ with the distribution

$$P(X = n) = \frac{1}{cn^2}$$

for each $n \in \mathbb{N}$, where $c = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. Then P is justified as a probability measure, since

$$P(E) = P(X = 1) + P(X = 2) + \cdots = \sum_{n=1}^{\infty} \frac{1}{cn^2} = \frac{\sum_{n=1}^{\infty} 1/n^2}{\sum_{n=1}^{\infty} 1/n^2} = 1.$$

On the other hand, we have

$$\mathbb{E}X = \sum_{n=1}^{\infty} nP(X = n) = \frac{1}{c}\left(\sum_{n=1}^{\infty} \frac{1}{n}\right) = \infty. \qquad ▲$$

Next, we study the properties of an expectation.

Theorem 3.24. *The following properties hold.*

(a) *The expectation is linear, i.e.*

$$\mathbb{E}[\lambda_1 X_1 + \cdots + \lambda_n X_n] = \lambda_1 \mathbb{E}X_1 + \ldots \lambda_n \mathbb{E}X_n.$$

for any $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, *where* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *are discrete r.v.s.*

(b) *The expectation is monotone, i.e. for any discrete r.v.s* $X_1, X_2 : \Omega \to \mathbb{R}$ *such that* $X_1 \leq X_2$, *then*

$$\mathbb{E}X_1 \leq \mathbb{E}X_2.$$

(c) *The expectation satisfies the triangle inequality, i.e.*

$$|\mathbb{E}X| \leq \mathbb{E}\,|X|$$

for a discrete r.v. $X : \Omega \to \mathbb{R}$.

Proof. The proof follows from the elementary properties of series. ∎

Bibliography

[1] René L. Schilling. *Measures, integrals and martingales.* Cambridge: Cambridge University Press, 2nd edition edition, 2017. ISBN 978-1-316-62024-3; 978-0-511-81088-6. doi: 10.1017/CBO9780511810886.