

Textbooks for Tomorrow's Scientists

PROBABILITY
THEORY

Jiashan TANG

Elements of Stochastic Processes

**PROBABILITY
THEORY**

Elements of Stochastic Processes

Jiashan TANG

This book introduces the fundamental elements of stochastic processes, covering basic concepts, core knowledge, and essential methods. It is organized into seven chapters, which include: the foundations of probability, concepts of stochastic processes, Poisson processes, discrete-time Markov chains, continuous-time Markov chains, simple Markovian queueing models, and stationary processes. Each chapter features figures, examples, exercises, and either solutions or guidance to support learning. Written from a non-measure-theoretic perspective, the book is accessible and highly readable, making it especially suitable for self-study.

The intended audience includes senior undergraduate and junior graduate students, as well as teachers in fields such as engineering, economics, management science, operations research, and applied mathematics. It also serves as a valuable reference for researchers and professionals with an interest in stochastic processes.

Jiashan TANG is full professor of mathematics and director of the Stochastics Research Center at Nanjing University of Posts and Telecommunications (NUPT). He received his Ph.D. in mathematics from Beijing Normal University in 1997 and was a Postdoctoral Fellow in the School of Mathematics and Statistics at Carleton University (2002–2003). His research interests include probability theory, applied stochastic processes, queueing theory, signal and information processing, applied statistics, and data science.

ISBN : 978-2-7598-3918-6



ed sciences
www.edpsciences.org

SCIENCE PRESS

Textbooks for Tomorrow's Scientists

Jiashan TANG

Elements of Stochastic Processes

Printed in France

EDP Sciences – ISBN(print): 978-2-7598-3918-6 – ISBN(ebook): 978-2-7598-3919-3
DOI: 10.1051/978-2-7598-3918-6

All rights relative to translation, adaptation and reproduction by any means whatsoever are reserved, worldwide. In accordance with the terms of paragraphs 2 and 3 of Article 41 of the French Act dated March 11, 1957, “copies or reproductions reserved strictly for private use and not intended for collective use” and, on the other hand, analyses and short quotations for example or illustrative purposes, are allowed. Otherwise, “any representation or reproduction – whether in full or in part – without the consent of the author or of his successors or assigns, is unlawful” (Article 40, paragraph 1). Any representation or reproduction, by any means whatsoever, will therefore be deemed an infringement of copyright punishable under Articles 425 and following of the French Penal Code.

The printed edition is not for sale in Chinese mainland. Customers in Chinese mainland please order the print book from Science Press. ISBN of the China edition: Science Press 978-7-03-059620-8

© Science Press, EDP Sciences, 2025

For the Celebration of the 90th Birthday
of Professor Zikun Wang

Preface

Natural phenomena that occur in our real world can be categorized into two types: the deterministic type and the random type. As one of the subjects which are used to carry out the investigation for the random phenomenon, stochastic processes have attracted much attention since the beginning of the last century. It is not only important in scientific studies but also useful in the application areas, such as science, engineering, military science, economics, science of management, and medicine.

As more and more people have realized the importance of stochastic processes in our daily lives, stochastic processes have become a fundamental course in many majors in universities across the world. Tens of thousands of textbooks on stochastic processes have been published during the last hundred years. Chinese scholars have made great progress in the field of stochastic processes and have brought forth an important class of textbooks on stochastic processes with Chinese characteristics intertwined in them. Such kind of textbooks is particularly suitable for Chinese scholars and has played an important role in the cultivation of talents in the past.

As there have been developments in science and technology in China, more and more overseas young students have come to China to complete or expand their academic education. Textbooks of stochastic processes written in Chinese are not suitable for students who do not understand Chinese characters or have an insufficient understanding of mandarin. In addition, many western textbooks are not suitable for certain foreign students, either due to their personal field of study, background and/or environment. Evidence of this was seen when the author began to teach stochastic processes to these new students from the autumn of 2014 and found that the overseas students needed a new, more convenient textbook for learning. This made the author come to the idea of writing such a textbook.

The author has been teaching stochastic processes for undergraduate and graduate students for more than 20 years and has accumulated a wealth of teaching

experience. The principle of writing this book is to introduce the basic concepts, methods and practices clearly by using a sufficient amount of examples and/or figures, which increases the readability of the textbook. The book is written from a non-measure theory point of view and is suitable for readers majoring in engineering, economics, management science, operations research, applied mathematics, etc. In order to save space, the author assumes that readers have the basic knowledge of advanced mathematics, linear algebra, differential equations, elements of probability, etc. Based on this idea, the book consists of seven chapters.

Chapter 1 introduces the basis of probability, which includes random events, probabilities, random variables, numerical characteristics and limiting theorems. This is just a simplified version of the probability theory.

Chapter 2 gives the definition of stochastic processes and introduces some statistical features and properties as well. After this chapter, several important stochastic processes will be introduced in different chapters in the rest of the book.

Chapter 3 presents the introduction of Poisson processes, which includes definitions and measurements, random sequences of waiting time and interarrival time, conditional distributions and several extensions as well.

Chapter 4 is on the discrete-time Markov chains. The main content consists of definitions, finite-dimensional distributions, properties of a single state, classifications of state spaces, asymptotic behaviors of transition probabilities, stationary distributions and reversible Markov chains.

Chapter 5 is a continuous counterpart of Chapter 4. Besides the contents, which are similar to that in Chapter 4, the main attention is focused on the different parts, which include Q -matrixes, Kolmogorov differential equations, asymptotic behaviors and birth and death processes.

Chapter 6 deals with simple Markovian queueing models, which can be regarded as a part of stochastic processes or one class of its application. The content includes terminologies, Little's law, PASTA property, and some queueing models, such as $M/M/1$, $M/M/n$, $M^X/M/1$ and $M/G/1$ as well.

Chapter 7 is very important for engineering readers, especially for communication guys. The main content covers: definitions of strict/weak-sense stationary processes,

analytic properties of the weak-sense stationary processes, correlation functions and their spectra, ergodicities and stationary processes passing through a linear time-invariant system.

This book is suitable for senior undergraduate students, junior graduate students and related teachers in engineering, economics, management science, operations research and applied mathematics. It is also suitable for scientific workers who are interested in stochastic processes. Moreover, various parts of chapters can be selected to be used independently for different objectives. Due to the limitation of time and the author's knowledge level, there must exist many improper things in this book. Any comments and suggestions are warmly welcomed and appreciated!

When the author was a master and a PhD student studying in Beijing Normal University, his supervisors were Prof. Zikun Wang (Chinese academician of CAS) and Prof. Zhanbing Li. During his visit to Carleton University in Canada as a postdoc (Postal doctoral fellow), his supervisors were Prof. Donald A. Dawson (fellow of the Royal Society) and Prof. Yiqiang Q. Zhao. The author thanks his mentors for their guidance into the field of stochastic processes and its applications and extends warm thanks for their invaluable guidance in his life. The author thanks many predecessors and his colleges for their valuable concerns and supports. Thanks are also given to two master students, Ying Tang in Southeast University and Heidi Gabrielle Pluntke in Nanjing University of Posts and Telecommunications, for their pre-reading to improve the English level of the current version. During the procedure of publishing the book, the chief responsible editor Zhongxing Zhang in the Science Press has given a lot of help. Additionally, the book has been elected as a key textbook for universities in Jiangsu province and Nanjing University of Posts and Telecommunications as well, and has received financial support simultaneously. The author would like to take this opportunity to say thanks to them.

The author gives special thanks to his parents and his family members. The author was born and grew up in the countryside of Anhui province. His mother never went to school and his father had finished only the fourth grade of primary school. From the 1970s to the 1980s in the last century, people lived in a difficult environment in China. During those hard times, the author's parents, as peasants, still insisted on

providing chances for their children to go to school and further their education, which was an extremely difficult task, but a worthy one. To their delight, after many years of hard work, the author became to be known amongst his neighbours in the village as the first person to become a university student. The author wishes to express deep thanks and appreciation to his parents and family members for their motivation for continuing to improve his knowledge and providing encouragement in pushing his career forward. They are the most important people in the author's life and have had the most influence on his destiny.

Jiashan Tang
tangjs@ymail.com
25 July, 2018

Contents

Preface

Chapter 1	Basis of probability	1
1.1	Random events	1
1.1.1	Random experiments	2
1.1.2	Random events	3
1.1.3	Calculation among random events	6
1.1.4	σ -algebras	10
1.2	Probability	12
1.2.1	What is probability	12
1.2.2	Properties of probability	14
1.2.3	Two fundamental probability models	16
1.2.4	Conditional probability and three important formulas	19
1.2.5	Independence of events	23
1.3	Random variables	25
1.3.1	Definition	25
1.3.2	Cumulative distribution functions	27
1.3.3	Classification	30
1.3.4	Multiple-dimensional random variables	34
1.3.5	Distributions of functions of random variables	38
1.3.6	Conditional distributions and independences	44
1.3.7	Order statistics	47
1.4	Numerical characteristics	50
1.4.1	Definition of mathematical expectations	50
1.4.2	Properties	56
1.4.3	Other numerical characteristics of random variables	57
1.4.4	Conditional mathematical expectations	73
1.5	Limiting theorems	76
1.5.1	Convergence of random variables	76
1.5.2	Law of large numbers	80
1.5.3	Central limit theorems	84

Exercise 1	86
Answers or tips for Exercise 1	88
Chapter 2 Stochastic processes	95
2.1 Definition and classification	95
2.1.1 Definition	96
2.1.2 Classification	98
2.1.3 Examples	99
2.2 Statistical laws of stochastic processes	102
2.2.1 Finite dimensional distribution functions	102
2.2.2 Kolmogorov's theorem	103
2.3 Measurements of stochastic processes	104
2.3.1 Measurements for one stochastic process	104
2.3.2 Measurements for two stochastic processes	107
2.4 Further comments	108
Exercise 2	109
Answers or tips for Exercise 2	111
Chapter 3 Poisson processes	112
3.1 Definition and measurements	112
3.1.1 Definition	112
3.1.2 Measurements	123
3.2 Waiting times and interarrival times	126
3.2.1 Waiting times	127
3.2.2 Interarrival times	130
3.3 Conditional distributions	131
3.4 Extensions of Poisson processes	136
3.4.1 Non-homogeneous Poisson processes	137
3.4.2 Compound Poisson processes	142
3.4.3 Conditional Poisson processes	145
3.4.4 Renewal processes	147
Exercise 3	152
Answers or tips for Exercise 3	154
Chapter 4 Discrete-time Markov chains	157
4.1 Definition of discrete-time Markov chains	157
4.1.1 Definition	159
4.1.2 Chapman-Kolmogorov equations	163

4.2	Finite-dimensional distributions	165
4.3	Properties of a single state	168
4.3.1	Periods	168
4.3.2	Transience and recurrence	169
4.4	Decomposition of state space	177
4.4.1	Equivalence relation	177
4.4.2	Decomposition of state space	181
4.5	Asymptotic behaviors of transition probabilities $p_{ij}^{(n)}$	184
4.5.1	Case one: state j is transient or null-recurrent	185
4.5.2	Case two: state j is positive-recurrent	186
4.6	Stationary distributions	196
4.6.1	Definition of stationary distributions	196
4.6.2	How many stationary distributions a Markov chain may have?	199
4.6.3	Rates of convergence to stationary distributions	201
4.6.4	Stationary distributions of a censored Markov chain	201
4.6.5	Quasi-stationary distributions	203
4.7	Reversible Markov chains	206
	Exercise 4	209
	Answers or tips for Exercise 4	211
Chapter 5	Continuous-time Markov chains	213
5.1	Definition of continuous-time Markov chains	213
5.1.1	Definition	213
5.1.2	Chapman-Kolmogorov equations	218
5.2	Finite-dimensional distributions	218
5.3	Q -matrices	220
5.4	Kolmogorov differential equations	227
5.5	Asymptotic behaviors	236
5.5.1	Transience and recurrence	237
5.5.2	Limiting results	239
5.5.3	Stationary distributions	240
5.6	Birth and death processes	243
	Exercise 5	249
	Answers or tips for Exercise 5	251
Chapter 6	Simple Markovian queueing models	256
6.1	Terminology and notation	256

6.2	Little's law and PASTA property	259
6.2.1	Little's law	259
6.2.2	PASTA property	260
6.3	$M/M/1$ queueing model	261
6.3.1	Stationary distribution of queue length	261
6.3.2	Distributions of sojourn times and waiting times	266
6.3.3	Busy period distribution	270
6.3.4	Departure process	273
6.4	$M/M/n$ and state dependent $M/M/1$ queueing model	275
6.4.1	$M/M/n$ queueing model	275
6.4.2	State dependent $M/M/1$ queueing model	276
6.5	$M^X/M/1$ queueing model	279
6.6	$M/G/1$ queueing model	281
6.6.1	Embedded Markov chain	281
6.6.2	$M/E_r/1$ queueing model	286
	Exercise 6	290
	Answers or tips for Exercise 6	292
Chapter 7	Stationary processes	295
7.1	Definition	295
7.1.1	Strict-sense stationary processes	295
7.1.2	Wide-sense stationary processes	298
7.2	Analytic properties of wide-sense stationary processes	302
7.3	Correlation functions and their spectra	312
7.3.1	Properties of correlation functions	312
7.3.2	Spectral density functions	314
7.3.3	Properties of spectral density functions	323
7.3.4	Continuous-time white noise processes	325
7.4	Ergodicity	327
7.5	Passing through a linear time-invariant system	335
	Exercise 7	345
	Answers or tips for Exercise 7	346
	References	354
	Index	358

Chapter 1

Basis of probability

Main knowledge

1. Probability space.
2. Random variables.
3. Numerical measurements.
4. Limiting theorems.

1.1 Random events

In our natural world, there are many phenomena, which can be categorized into two cases. One is certain/deterministic phenomena, in which case, when we do something, if the condition is satisfied, then the unique result will happen. For example, water naturally flows downhill. The other is an uncertain/random case, in which, when we do something, although the condition is satisfied, the outcome still can not be forecasted.

Let's consider what the weather will be tomorrow. Although we know that it might be sunny, cloudy, rainy, etc. we do not know today what exactly the weather will be tomorrow. Similarly, the stock price tomorrow morning is also uncertain. In fact, there are many uncertain phenomena around us. In the modern scientific world, besides the classical mathematics, this phenomena can be analyzed by the disciplines, such as probability theory, statistics, stochastic processes, stochastic analysis and etc.

Among the subjects mentioned above, probability theory is a prerequisite course of stochastic processes. Therefore, before meeting the main content of this book beginning from Chapter 2, we need to revisit the knowledge of probability theory in the current chapter.

1.1.1 Random experiments

As a scientific approach, in order to investigate uncertain phenomena, we need to know the information or data of the phenomenon. There are two methods to collect those information: observation and experimentation. Apparently, observation is a passive behavior; how much information we can get only depends on the happening frequency of the phenomenon. Plenty of information can only be obtained through active experimentation.

Similar to the experiments implemented in physics and chemistry, we need to impose some conditions on the experiments in the probability theory.

- (1) The experiments (or trials) can be infinitely repeated under the same conditions;
- (2) All possible outcomes of the experiments are known at any time;
- (3) The outcome of a particular experiment can not be predicted before the implementation.

If an experiment satisfies the above three conditions, then it is called a **random experiment**.

Remark 1.1 *Condition (1) is a constrain condition which is imposed on all scientific experiments. For example, we may have had a dream yesterday evening. However, we can not have the same dream tonight. Therefore, we can not make dreams a scientific experiment. In fact, all scientific experiments can be repeated by any person, at any time, as long as the same conditions hold.*

Condition (2) implies that the experiment or phenomenon we are considering is a simple and post-modelling case in some sense. For example, if some outcomes of the experiment can not be observed or can not be observed clearly, then we can regard part of the results as a particular outcome of the experiment.

*Condition (3) indicates that there are at least two outcomes of a random experiment. Otherwise, if there is only one outcome of an experiment, then the outcome of the experiment can be forecasted before the implementation. In this case, the experiment is called a **deterministic experiment**. A random experiment always has two or more results. Moreover, if a random experiment has exactly two possible outcomes, then it is called a **Bernoulli trial**.*

Example 1.1 Tossing a fair coin is a random experiment.

If we have a flat table and a fair coin, then we can do the experiment of tossing the coin. I can do it, and you can do it. Everybody can do it. We can do it today,

and of course, we can do it tomorrow. The experiment of tossing the coin can be repeated any times you want. It implies that condition (1) holds.

Before we flip the coin, we know all possible outcomes of the experiment. In this case, it maybe the head or the tail. Here we omit one case that the coin is cocked up. It implies that condition (2) holds.

For a particular flipping, the outcome can not be forecasted in advance, which means that condition (3) holds.

There are many examples of random experiment around us. For example, recording the number of incoming calls of a telephone exchange within one minute; considering the number of received calls of an ATM switch within one unit time; taking one from a batch of electronic components and testing its lifetime and etc.

In probability theory, the uncertain phenomena are investigated through the corresponding random experiments. All experiments are understood as random experiments in the rest of this chapter.

1.1.2 Random events

For random experiments introduced in the last subsection, we have the following definitions.

Definition 1.1 Sample point, sample space

For a random experiment, each simple outcome is called a **sample point**, which is denoted by ω . All sample points form a set, which is called a **sample space** and is denoted by Ω .

Example 1.2 For the random experiment of tossing a fair coin, we have two sample points: $\omega_1 = H$ and $\omega_2 = T$, where H means head and T means tail. Therefore, the sample space is $\Omega = \{\omega_1, \omega_2\} = \{H, T\}$.

Example 1.3 Consider a random experiment of tossing a fair dice which has six sides with faces numbered from 1 to 6, respectively. Apparently, we have six sample points, which are $\omega_1 = 1, \dots, \omega_6 = 6$. Moreover, the sample space is given by $\Omega = \{1, 2, \dots, 6\}$.

Example 1.4 We still consider the random experiment of tossing a fair dice as in Example 1.3. What we are interested in now is that the outcome for each toss is odd or even. In this case, “*odd*” is a simple outcome and “*even*” is another outcome. Therefore, the sample space is $\Omega = \{odd, even\}$.

Examples 1.3 and 1.4 indicate that even for the same random experiment, we may have different sample spaces, and the reason of which is depending on the facts that we are interested in. The cardinal numbers of the sample spaces in the above three examples all are finite. Let's have a look at examples whose cardinal numbers of the sample spaces are infinite.

Example 1.5 Let's consider the number of live fishes swimming in a section of the Yangtze River, which ranges from Wuhan City to Nanjing City in China. The worst thing is that there are no fish in this section, which means that 0 is a possible outcome. Similarly, 1 is also a possible outcome, and 2 is too. In fact, any non-negative integer number is a possible outcome. Therefore, the sample space is $\Omega = \{0, 1, 2, \dots\}$, which has infinitely many numbers. Realistically, we know that the number of fishes within the Yangtze River can not be infinite. However, any non-negative integer can be the number of fishes in that section, which means that the number of fishes can be, theoretically, arbitrarily large.

Example 1.6 Consider the water storage in a reservoir with the unit being m^3 . If there is no water, then the storage is 0, which means that 0 is a possible outcome. Suppose that the maximum storage of this reservoir is b and this number can not be reached. It's easy to see that the sample space can be written as $\Omega = \{x | 0 \leq x < b\}$.

In Example 1.3, we know that the sample space is $\Omega = \{1, 2, \dots, 6\}$. In a practical situation, one person might be interested in the tossing result which is equal to or less than 3, while another guy is interested in the result which is greater than 2. All these kinds of results are very important in probability theory and have a particular name, which is given in the following.

Definition 1.2 Classical random event

Suppose that Ω is a sample space for some random experiment. Any subset of the sample space, $A \subseteq \Omega$, is called a **random event** of the sample space. A random event is always denoted by a capital letter with or without a subscription. The family of all events is denoted by $\mathcal{F}_{all} = \{A | A \subseteq \Omega\}$.

Example 1.7 Consider the sample space $\Omega = \{H, T\}$ in Example 1.2. Since that the empty set is a subset of all sets, $\emptyset$ is a random event. In fact, we totally have four random events and they are: $A_1 = \emptyset$, $A_2 = \{H\}$, $A_3 = \{T\}$ and $A_4 = \{H, T\}$. Furthermore, $\mathcal{F}_{all} = \{A_1, A_2, A_3, A_4\}$. If the sample space is not complicated as in this example, then the family of events can also be written as $\mathcal{F}_{all} = \{\emptyset, \{H\}, \{T\}, \Omega\}$.

Consider tossing a fair dice as in Example 1.3. The sample space is $\Omega = \{1, 2, \dots, 6\}$. Apparently, both $A_1 = \{3\}$ and $A_2 = \{1, 3, 5\}$ are two random events. If in a particular performance, the outcome of the toss is 3, then we can say that the event, “ A_1 ”, occurred in this toss. Since $3 \in A_2$, we can say that the event, “ A_2 ”, occurred in this toss, too. Generally, if the outcome of a particular test is ω , then we can say that the event, which includes ω as its component, occurred in this test.

Example 1.8 If a sample space has three sample points $\Omega = \{\omega_1, \omega_2, \omega_3\}$, then all random events of this sample space are: $A_1 = \emptyset$, $A_2 = \{\omega_1\}$, $A_3 = \{\omega_2\}$, $A_4 = \{\omega_3\}$, $A_5 = \{\omega_2, \omega_3\}$, $A_6 = \{\omega_1, \omega_3\}$, $A_7 = \{\omega_1, \omega_2\}$, $A_8 = \{\omega_1, \omega_2, \omega_3\}$. Moreover, $\mathcal{F}_{all} = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$.

Generally, for sample spaces with finite cardinal numbers of sample points, we have the following proposition on the number of random events.

Proposition 1.1 If a sample space has n sample points, $\Omega = \{\omega_1, \dots, \omega_n\}$, then the total number of random events for this sample space is 2^n .

Proof Since that $\emptyset \subset \Omega$, $\emptyset$ is a random event, which corresponds to the subset that there is no sample points selected from the sample space. The number of this kind of subsets is $C_n^0 = 1$. This is the first case. For those subsets, each of which has only one sample point selected from the sample space, the total number is $C_n^1 = n$. This is the second case. The third case is that there are C_n^2 subsets, each of which has two sample points chosen from the sample space and so on. The last case is that the subset has all sample points chosen from the sample space, and the number is $C_n^n = 1$. Therefore, the total number of random events for the sample space having n sample points is

$$C_n^0 + C_n^1 + \dots + C_n^n = (1 + 1)^n = 2^n.$$

Example 1.9 Let's toss a dice two times (or toss a pair of dice). In this case, the simplest outcome reads a pair of numbers like $\omega = (i, j)$, which means that i is the number which appeared in the first toss and j is in the second. Since the value of i can be 1, 2, $\dots$, 6 and that of j is the same as that of i . Therefore, the sample space is $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 6)\}$. Consider some random events for this sample space.

- (1) Random event having no sample point: $A_1 = \emptyset$;
- (2) Random event with the property that at least one number is 6:

$$A_2 = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 5), (6, 4), (6, 3), (6, 2), (6, 1)\}.$$

(3) Random event that the summation of two numbers is 7:

$$A_3 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

1.1.3 Calculation among random events

For any sample space Ω , there are many random events. Sometimes, we know some random events, however, what we are interested in may be another event. Calculation among events can solve these kinds of problems.

1. Complement of an event

Let A be a random event of a sample space Ω , the **complement** of event A is denoted by A' (also by $\bar{A}$) and is defined by $A' = \Omega \setminus A = \{\omega : \omega \in \Omega, \omega \notin A\}$. From this definition, we know that the relationship between event A and its complement A' can be depicted in Fig. 1.1 as follows.

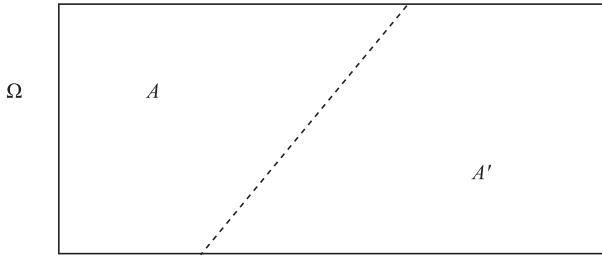


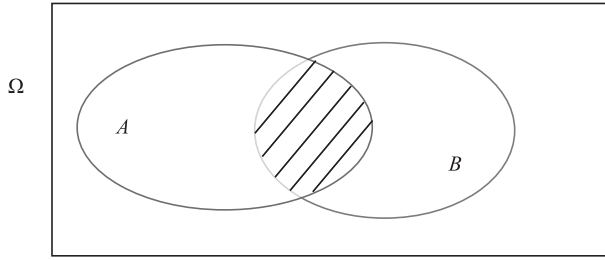
Fig. 1.1 Relationship between event A and its complement A'

One may note that a simple conclusion on the computation of complement of event A is that $(A')' = A$.

2. Intersection of events

Let A and B represent two events, the **intersection** of these two events consists of all outcomes that are contained within both events A and B , and is denoted by $A \cap B$ (also by AB). The intersection of two events can be depicted as the shadow area in Fig. 1.2.

Mathematically, we understand that $A \cap B = \{\omega | \omega \in A \text{ and } \omega \in B\}$, which means that the intersection of events A and B is the set consisting of the sample points, not only in A , but also in B .

Fig. 1.2 Intersection of events A and B

Example 1.10 Consider Example 1.9 again, and find the intersection of A_2 and A_3 .

Solution From the definition of the intersection of events and the values of events A_2 and A_3 , we know that

$$A_2 \cap A_3 = \{(1, 6), (6, 1)\}.$$

Let n be a positive integer, and $A_1, A_2, \dots, A_n$ be n events, and then the intersection of these n events is the event which consists of sample points belonging to all n events and is denoted by $A_1 \cap A_2 \cap \dots \cap A_n$, or $\bigcap_{\kappa=1}^n A_\kappa$. It means that $\bigcap_{\kappa=1}^n A_\kappa = \{\omega | \omega \in A_1 \text{ and } \dots \text{ and } \omega \in A_n\}$. The case of $n = 3$ is as the shadow area shown in Fig. 1.3.

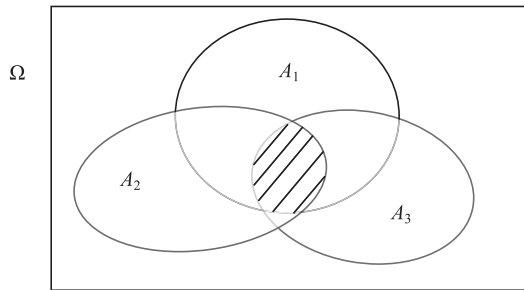


Fig. 1.3 Intersection of three events

It's easy to see that event A and its complement A' have no common sample points, and therefore $A \cap A' = \emptyset$. Generally, if events A and B have no common sample points, namely, $A \cap B = \emptyset$, then we call events A and B **mutually exclusive events**. See Fig. 1.4.

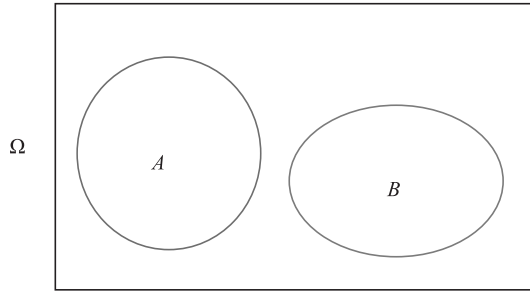
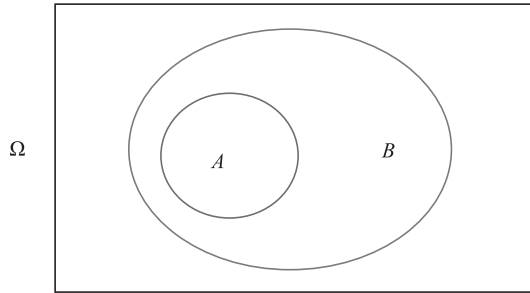


Fig. 1.4 Mutually exclusive events

As we know, a random event is a subset of a sample space. Sometimes, we may have that event A is **included** in event B , in which case, $A \cap B = A$. See Fig.1.5.

Fig. 1.5 Event A is included in event B

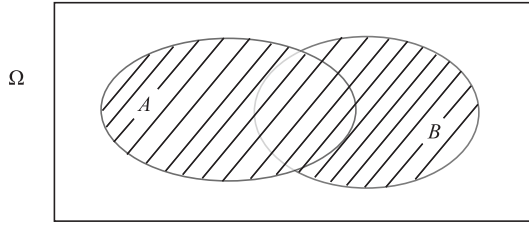
We should note that the relationship between two events has three cases, which are depicted in Fig. 1.4, Fig. 1.2 and Fig. 1.5, respectively.

3. Union of events

Let A and B represent two random events of a sample space, the **union** of events A and B is the set consisting of the outcomes that are contained within at least one of the events of A and B , and is denoted by the notation $A \cup B$. See Fig. 1.6 as in the following.

Similar to the computation of intersection, the union of n events $A_1, \dots, A_n$ is denoted by $\bigcup_{\kappa=1}^n A_{\kappa} = \{\omega | \omega \in A_1 \text{ or } \omega \in A_2 \text{ or } \dots \text{ or } \omega \in A_n\}$.

Based on the definition of complement, intersection and union, we have the following properties.

Fig. 1.6 Union of events A and B

Property 1.1 Let $A_1, \dots, A_n$ be n events of a sample space Ω . Then

- (1) $(A_1 \cap A_2)' = A_1' \cup A_2'$;
- (2) $(A_1 \cup A_2)' = A_1' \cap A_2'$;
- (3) $A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$;
- (4) $A_1 \cup (A_2 \cap A_3) = (A_1 \cup A_2) \cap (A_1 \cup A_3)$;
- (5) $\left(\bigcap_{\kappa=1}^n A_{\kappa} \right)' = \bigcup_{\kappa=1}^n A_{\kappa}'$;
- (6) $\left(\bigcup_{\kappa=1}^n A_{\kappa} \right)' = \bigcap_{\kappa=1}^n A_{\kappa}'$.

Proof We just give the proof of (3) and leave the proofs of the rest of the properties to readers.

In order to prove the equality of two events, the idea is to prove that the first event belongs to the second one and the second event belongs to the first one too. For this sake, take any sample point from the left hand side event of (3), namely, take $\omega \in A_1 \cap (A_2 \cup A_3)$, then $\omega \in A_1$ and $\omega \in A_2 \cup A_3$ simultaneously. It implies from $\omega \in A_2 \cup A_3$ that $\omega \in A_2$ or $\omega \in A_3$. Combine the property that $\omega \in A_1$, we have that $\omega \in A_1$ and $\omega \in A_2$, or $\omega \in A_1$ and $\omega \in A_3$, and therefore $\omega \in A_1 \cap A_2$ or $\omega \in A_1 \cap A_3$, which implies that $\omega \in (A_1 \cap A_2) \cup (A_1 \cap A_3)$. For the arbitrary of ω , we know that $A_1 \cap (A_2 \cup A_3) \subseteq (A_1 \cap A_2) \cup (A_1 \cap A_3)$.

For another direction, take any sample point from the right hand side event of (3), namely, $\omega \in (A_1 \cap A_2) \cup (A_1 \cap A_3)$, and then $\omega \in A_1 \cap A_2$ or $\omega \in A_1 \cap A_3$. It implies that $\omega \in A_1$ and $\omega \in A_2$ or $\omega \in A_3$, which means that $\omega \in A_1$ and $\omega \in A_2 \cup A_3$. Therefore, we have that $\omega \in A_1 \cap (A_2 \cup A_3)$. Finally, we have that $(A_1 \cap A_2) \cup (A_1 \cap A_3) \subseteq A_1 \cap (A_2 \cup A_3)$.

Combining the above two facts together, we have the result of (3) in Property 1.1.

4. Calculation of infinitely many of random events

Sometimes, we happen to have infinitely many of random events, say $A_1, A_2, \dots$.

The intersection and the union of these infinitely many of events can be calculated similarly.

$$\bigcap_{\kappa=1}^{\infty} A_{\kappa} = \{\omega | \omega \in A_1 \text{ and } \omega \in A_2 \text{ and } \cdots\},$$

$$\bigcup_{\kappa=1}^{\infty} A_{\kappa} = \{\omega | \omega \in A_1 \text{ or } \omega \in A_2 \text{ or } \cdots\}.$$

Apparently, we have that

$$\bigcap_{\kappa=1}^{\infty} A_{\kappa} \subseteq \bigcup_{\kappa=1}^{\infty} A_{\kappa}.$$

We need to pay attention to two special cases. One case is that the sequence of random events satisfies $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$. In this case, we denote it by

$$\lim_{\kappa \rightarrow \infty} A_{\kappa} = \bigcap_{\kappa=1}^{\infty} A_{\kappa}.$$

Another case is that the sequence satisfies $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$. Similarly, it is denoted by

$$\lim_{\kappa \rightarrow \infty} A_{\kappa} = \bigcup_{\kappa=1}^{\infty} A_{\kappa}.$$

Based on these two facts, for the general sequence of random events: $\{A_1, A_2, A_3, \cdots\}$, we have the following result

$$\bigcup_{n=1}^{\infty} \bigcap_{\kappa=n}^{\infty} A_{\kappa} \subseteq \bigcap_{n=1}^{\infty} \bigcup_{\kappa=n}^{\infty} A_{\kappa}. \quad (1.1)$$

Intuitively, the event on the left hand side $\bigcup_{n=1}^{\infty} \bigcap_{\kappa=n}^{\infty} A_{\kappa}$ includes the sample points which do not belong to finitely many of events, while the event on the right hand side $\bigcap_{n=1}^{\infty} \bigcup_{\kappa=n}^{\infty} A_{\kappa}$ includes the sample points which belong to infinitely many of events.

1.1.4 σ -algebras

To investigate a random phenomenon, it is very important to know the possibility of an event in any experiment, the concept of which will be defined in the next section. This involves the definition of a random event. For a simple case, like in Definition 1.2, an event is defined as a subset of the sample space. These kinds of random events can be called classical random events. However, in a more complicated case, this classical definition may bring some difficulties. The worst case is that the possibility

of some events can not be well defined. To avoid this, we have a new family of sets called a σ -algebra.

Definition 1.3 σ -algebra

Let Ω be a sample space, $\mathcal{F}$ be a family of sets consisting of parts of all subsets of the sample space. If this family $\mathcal{F}$ has the following properties:

- (1) $\Omega \in \mathcal{F}$;
- (2) If $A \in \mathcal{F}$, then $A' \in \mathcal{F}$;
- (3) Take $A_\kappa \in \mathcal{F}$, $\kappa = 1, 2, \dots$, then $\bigcup_{\kappa=1}^{\infty} A_\kappa \in \mathcal{F}$.

Then $\mathcal{F}$ is called a **σ -algebra** of the sample space Ω .

For the concept of σ -algebras for a sample space Ω , we have the following statements.

(1) Define $\mathcal{F}_{\text{small}} = \{\emptyset, \Omega\}$, then $\mathcal{F}_{\text{small}}$ is a σ -algebra, which includes the smallest number of elements.

It is easy to verify that $\mathcal{F}_{\text{small}}$ is a σ -algebra. Note that, condition (1) indicates that $\Omega \in \mathcal{F}_{\text{small}}$, while condition (2) implies that $\emptyset = \Omega' \in \mathcal{F}_{\text{small}}$.

(2) Let A be a true subset of the sample space Ω , i.e. $A \subsetneq \Omega$, then a σ -algebra containing the subset A and having the smallest number of elements is $\mathcal{F}_A = \{\emptyset, \Omega, A, A'\}$.

We refer the readers to prove that $\mathcal{F}_A$ is a σ -algebra. Naturally, $A \in \mathcal{F}_A$. Moreover, condition (2) implies that $A' \in \mathcal{F}_A$.

(3) Using the same notation as above, $\mathcal{F}_{\text{all}}$ is a σ -algebra.

(4) If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two different σ -algebras, then $\mathcal{F}_1 \cap \mathcal{F}_2$ must be a σ -algebra, while $\mathcal{F}_1 \cup \mathcal{F}_2$ maybe not.

(5) Let $\mathcal{G}$ be a family of some subsets of the sample space, then the σ -algebra containing $\mathcal{G}$ with the smallest number of elements is always denoted by $\mathcal{F}_{\mathcal{G}} = \sigma(\mathcal{G}) =$

$\bigcap_{\tilde{\mathcal{F}}' \text{ is a } \sigma\text{-algebra, } \mathcal{G} \subseteq \tilde{\mathcal{F}}'} \tilde{\mathcal{F}}'$. It is called the σ -algebra generated by the family $\mathcal{G}$.

Particularly, if $\mathcal{G} = \{A\}$, then $\mathcal{F}_{\mathcal{G}} = \sigma(\mathcal{G}) = \sigma(\{A\}) = \{\emptyset, \Omega, A, A'\}$.

Statements (1)-(3) show that for one sample space Ω , we can have various different σ -algebras. Generally, for any σ -algebra $\mathcal{F}$, the notation $(\Omega, \mathcal{F})$ is often called a **measurable space**, which means that the possibility of any element in $\mathcal{F}$ can be well defined, or can be measured.

After giving the definition of a σ -algebra for a sample space Ω , we only call the elements of $\mathcal{F}$ **random events**. Let A be a subset of the sample space Ω , if $A \in \mathcal{F}$,

then we call A an event. However, if $A \notin \mathcal{F}$, then we can not call A an event even if A is a subset of the sample space. That is to say, the unique criteria of judging a subset is an event or not is that the subset belongs or doesn't belong to the σ -algebra.

Classical random events are a rough concept. It is only correct within a certain range. The most important thing of giving the definition of σ -algebra is that the barrier problem mentioned above can be avoided. In fact, for one sample space, we can have various σ -algebras. $\mathcal{F}_{all}$ is only one special case of those σ -algebras.

1.2 Probability

For a random phenomenon, it is very important to know the random events of the sample space corresponding to the random phenomenon. However, one more important thing is to know the possibilities of events in the future. For example, when we consider the weather for tomorrow, knowing the state, such as sunny, cloudy, rainy, of tomorrow is important. More importantly, it is to know the possibility of there being rain tomorrow. In the case of rain, an individual may bring an umbrella for himself. In probability theory, one may use the probability of the event to indicate this kind of indexes.

1.2.1 What is probability

Let's consider the random phenomenon of tossing a fair coin. If we just flip one time, then we may find that there is no clue about what we can get from this toss. However, if the coin is flipped more times, then some particular information may appears. In history, there were several scientists who had done the experiment of tossing fair coins. Some data are shown in Table 1.1.

Table 1.1 Historical experiment of tossing coin

Experimenters	Numbers of tossing	Numbers of head upside	Frequencies of head upside
De Morgan	4092	2048	0.5005
Buffon	4040	2048	0.5069
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005
Romanovsky	80640	39699	0.4923

From this data, we may find that if the number of toss is big enough, then the frequencies of head (H) upside of these tosses are approximately 0.5. Apparently, if

the frequency of head upside is 0.5, then that of tail (T) upside is also 0.5. These two 0.5 and 0.5 are the possibilities of the head and the tail in one toss, respectively. We call them probabilities and denote them by: $P(H) = 0.5$ and $P(T) = 0.5$. We should note that the formula $P(H) = 0.5$ indicates that there is a possibility of reoccurring heads in one toss of *this particular coin*.

If someone chooses another new coin to do the same experiment, then the result might be different. Say, $P(H) = 0.7$ and $P(T) = 0.3$. These 0.7 and 0.3 are also called probabilities of tossing with this new coin.

Apparently, the probabilities of occurring head, or tail, in one toss corresponding to different coins may be different. Generally, from mathematical point of view, the concept of probabilities is to assign a nonnegative value to each event (outcome) with some properties. The concrete definition is as follows.

Definition 1.4 Probability

Suppose that $(\Omega, \mathcal{F})$ is a measurable space. For each $A \in \mathcal{F}$, there corresponds a nonnegative number $P(A)$, moreover, if these values $\{P(A), A \in \mathcal{F}\}$ satisfy that

(1) For any $A \in \mathcal{F}$, $0 \leq P(A) \leq 1$;

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

(3) For any sequence $A_1, A_2 \cdots \in \mathcal{F}$, if $A_i \cap A_j = \emptyset$ for $i \neq j$, then $P(\bigcup_{\kappa=1}^{\infty} A_{\kappa}) = \sum_{\kappa=1}^{\infty} P(A_{\kappa})$, then $P = P(\cdot)$ is called a **probability**. $P(A)$ is called the probability of event A .

Remark 1.2 *In the definition of probability, there is a family of events $\mathcal{F}$. In the classical case in this book (see Definition 1.2), $\mathcal{F} = \mathcal{F}_{all}$ includes all subsets (events) of the sample space. In a general case, the family of sets, $\mathcal{F}$, is defined as a σ -algebra. Above all, Ω is a sample space, $(\Omega, \mathcal{F})$ is a measurable space and $(\Omega, \mathcal{F}, P)$ is called a **probability space**.*

Remark 1.3 *The probability $P(\cdot)$ can be regarded as a mapping from $\mathcal{F}$ to $[0, 1]$,*

$$P(\cdot) : \mathcal{F} \mapsto [0, 1].$$

One may note that a probability $P(\cdot)$ is a set function, while those functions we have learned in the classical mathematics, such as advanced mathematics, calculus, algebra, are point functions. Therefore, results on point functions can not be applied to the probability. In order to solve this kind of problem, we need to construct a bridge

between set functions and point functions. Readers will find this part of contents in the following Section 1.3 entitled random variables.

1.2.2 Properties of probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space, statements of properties of probability are as the following

$$(1) P(\emptyset) = 0.$$

Consider a sequence of random events $A_1, A_2, \dots$, and $A_\kappa = \emptyset$, $\kappa = 1, 2, \dots$. Apparently, $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\bigcup_{\kappa=1}^{\infty} A_\kappa = \emptyset$. It implies from condition (3) of the definition of probability that

$$\begin{aligned} P(\emptyset) &= P\left(\bigcup_{\kappa=1}^{\infty} A_\kappa\right) \\ &= \sum_{\kappa=1}^{\infty} P(A_\kappa) \\ &= P(\emptyset) + P(\emptyset) + \dots \end{aligned}$$

Due to the fact that $P(\emptyset) \geq 0$, it implies from the above formula that $P(\emptyset) = 0$.

(2) For any positive integer n , and random events $A_1, \dots, A_n \in \mathcal{F}$. If $A_i \cap A_j = \emptyset$ for any $i \neq j$, then $P(\bigcup_{\kappa=1}^n A_\kappa) = \sum_{\kappa=1}^n P(A_\kappa)$.

For the random events $A_1, \dots, A_n$, consider events $A_{n+1} = \emptyset$, $A_{n+2} = \emptyset$, $\dots$. Namely, $A_{n+k} = \emptyset$ for $k \geq 1$. In this case, we still have the property that $A_i \cap A_j = \emptyset$ for any $i \neq j$. Note that $P(\emptyset) = 0$, then we have that

$$\begin{aligned} P\left(\bigcup_{\kappa=1}^n A_\kappa\right) &= P\left(\bigcup_{\kappa=1}^{\infty} A_\kappa\right) \\ &= \sum_{\kappa=1}^{\infty} P(A_\kappa) \\ &= \sum_{\kappa=1}^n P(A_\kappa) + \sum_{\kappa=n+1}^{\infty} P(A_\kappa) \\ &= \sum_{\kappa=1}^n P(A_\kappa). \end{aligned}$$

(3) If $A, B \in \mathcal{F}$ and $A \subseteq B$, then $P(A) \leq P(B)$.

Since that $A \subseteq B$, we have that $B = A \cup (B \cap A')$, moreover $A \cap (B \cap A') = \emptyset$. It means that $P(B) = P(A \cup (B \cap A')) = P(A) + P(B \cap A') \geq P(A)$, in which we use the fact that $P(B \cap A') \geq 0$.

(4) For any events $A, B \in \mathcal{F}$, we have that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Note that $A \cup B = A \cup (B \cap A')$ and $A \cap (B \cap A') = \emptyset$. $B = (A \cap B) \cup (B \cap A')$ and $(A \cap B) \cap (B \cap A') = \emptyset$. Therefore,

$$\begin{aligned} P(B) &= P(A \cap B) + P(B \cap A'), \\ P(A \cup B) &= P(A \cup (B \cap A')) \\ &= P(A) + P(B \cap A') \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

(5) Let $A_\kappa \in \mathcal{F}$, $\kappa = 1, 2, \dots$ be a sequence of events, and $A_1 \supseteq A_2 \supseteq \dots$, then we have that

$$P\left(\lim_{\kappa \rightarrow \infty} A_\kappa\right) = P\left(\bigcap_{\kappa=1}^{\infty} A_\kappa\right) = \lim_{\kappa \rightarrow \infty} P(A_\kappa),$$

which is called the **upper-continuous property of probability**.

(6) Let $A_\kappa \in \mathcal{F}$, $\kappa = 1, 2, \dots$ be a sequence of events, and $A_1 \subseteq A_2 \subseteq \dots$, then we have that

$$P\left(\lim_{\kappa \rightarrow \infty} A_\kappa\right) = P\left(\bigcup_{\kappa=1}^{\infty} A_\kappa\right) = \lim_{\kappa \rightarrow \infty} P(A_\kappa),$$

which is called the **lower-continuous property of probability**.

Remark 1.4 Under the frame of probability space, we should note that only events (elements) of the σ -algebra, say $A \in \mathcal{F}$, there is $P(A)$ corresponding to it. Otherwise, even if $A \subset \Omega$, as long as $A \notin \mathcal{F}$, we must realize that $P(A)$ does not make sense. This will bring a peculiar problem. Suppose that $B \in \mathcal{F}$ and $P(B) = 0$, and that $A \subset B$ and $A \notin \mathcal{F}$. On one hand, according to the property of probability that $0 \leq P(A) \leq P(B) = 0$, which results in $P(A) = 0$. On the other hand, $A \notin \mathcal{F}$, $P(A)$ does not make sense.

In order to solve this problem, one may extend the set of $\mathcal{F}$, the domain of probability $P(\cdot)$. In fact, for a general probability space $(\Omega, \mathcal{F}, P)$, define

$$\begin{aligned} \tilde{\mathcal{F}} &= \sigma(\mathcal{F} \cup \{A | A \subset B, B \in \mathcal{F}, P(B) = 0\}) \\ &= \{A | A \subset \Omega, B_1 \subset A \subset B_2, B_1 \in \mathcal{F}, B_2 \in \mathcal{F}, P(B_2 \setminus B_1) = 0\}, \\ \tilde{P}(A) &= P(B_1), \text{ if } B_1 \subset A \subset B_2 \text{ and } P(B_2 \setminus B_1) = 0. \end{aligned}$$

Then, $(\Omega, \tilde{\mathcal{F}}, \tilde{P})$ is usually called the **completion** of $(\Omega, \mathcal{F}, P)$ with respect to P . If $\tilde{\mathcal{F}} = \mathcal{F}$, then $(\Omega, \mathcal{F}, P)$ is a complete probability space. For a complete probability space, peculiar problems mentioned above do not exist anymore.

1.2.3 Two fundamental probability models

A probability space $(\Omega, \mathcal{F}, P)$ is often regarded as a stage on which we will deal with all random phenomena. However, a probability space is always an abstract concept. Fortunately, for some particular cases, it has some simplified forms. We consider two simplified probability models in this subsection.

1. Classical probability model

Suppose that the cardinal number of the sample space for one random phenomenon is finite and that all sample points are equally likely to occur. The probability model set up for this random phenomenon is called a **classical probability model**. In this case, the family of all subsets of the sample space is often taken as the σ -algebra, $\mathcal{F}_{\text{all}}$.

Let n be a positive integer and $\Omega = \{\omega_1, \dots, \omega_n\}$. From the description mentioned above, we know that $P(\{\omega_1\}) = \dots = P(\{\omega_n\})$. Combining with $P(\Omega) = 1$, we have that

$$P(\{\omega_1\}) = \dots = P(\{\omega_n\}) = \frac{1}{n}.$$

If we denote the cardinal number of a set D by $|D|$, then the probability of event A can be computed as

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

Example 1.11 Toss two fair dice simultaneously and find the probability of the event that the summation of two tossed numbers is 7.

Solution It is easy to see that the sample space is $\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}$ and $|\Omega| = 36$. Moreover, since two dice are fair, it is the case of a classical probability model. The event we are considering is

$$A = \{(1, 6), (2, 5), \dots, (6, 1)\}.$$

Therefore $|A| = 6$, and we have that

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

2. Geometrical probability model

The finiteness of the number of sample points in the classical probability model limits its applications. In order to solve this problem, one method is to relax the finiteness of the number of sample points in the last model and to keep the identical possibility of each elementary event.

Let's consider a random phenomenon, each outcome of which has a one to one relationship with some point within a bounded region in an Euclidean space of some dimension d , say $d = 2$. Therefore, once we proceed an experiment, it is similar to dropping a point in the region randomly. Moreover, if the possibility of dropping a point in some subregion is only depending on the area of the subregion and has no relationship with the shape and the location of the subregion, then the model is called a **geometrical probability model**.

See Fig. 1.7, and suppose that B and C are two disjoint subsets of the bounded domain D . Moreover, we may suppose that $P(B) = f(\text{Area}(B))$, where $f(\cdot)$ is a function valued in $[0, 1]$ and $\text{Area}(B)$ is the area of subset B . According to the properties of probability and the above hypotheses, we have that

$$\begin{aligned} f(\text{Area}(B) + \text{Area}(C)) &= f(\text{Area}(B \cup C)) \\ &= P(B \cup C) \\ &= P(B) + P(C) \\ &= f(\text{Area}(B)) + f(\text{Area}(C)). \end{aligned}$$

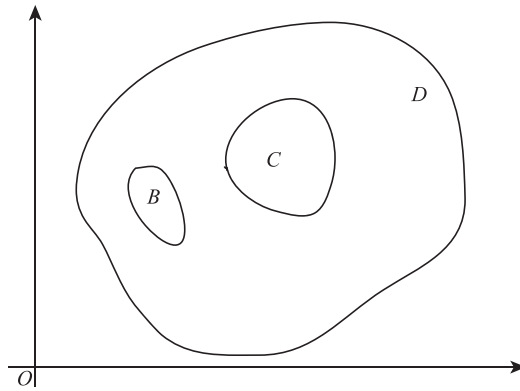


Fig. 1.7 Geometrical probability model

If we suppose that $x = \text{Area}(B)$ and $y = \text{Area}(C)$, the arbitrary of subsets B and C indicates that $f(x + y) = f(x) + f(y)$ for any $x, y \geq 0$. Furthermore, the monotone property of probability implies the same property holds for function $f(\cdot)$. It is easy to see that the structure of function $f(\cdot)$ is a linear function, $f(x) = \lambda x$, where λ is a positive constant. Therefore, for any subset $B \subset D$, we have

$$P(B) = f(\text{Area}(B)) = \lambda \text{Area}(B).$$

Note that $1 = P(D) = f(\text{Area}(D)) = \lambda \text{Area}(D)$ indicates that $\lambda = \frac{1}{\text{Area}(D)}$. Therefore, the probability of any subset B of domain D can be computed as

$$P(B) = \frac{\text{Area}(B)}{\text{Area}(D)}.$$

Example 1.12 Two friends make an appointment that they will meet at the lobby in the library during the time period from 15:00 to 16:00 tomorrow afternoon and agree that the first arriving one can leave after 20 minutes of waiting for the other. Suppose that these two friends came to the lobby randomly during the appointed time period. Find the probability that these two friends can meet each other during this designated time.

Solution We use one minute as the time unit. After regarding 15:00 as 0 and 16:00 as 60, we can construct a coordinate system shown in Fig. 1.8. Suppose that the time epoch at which one friend arrives is x and that the other friend arrives is y . Then, the sample space can be written as

$$D = \{(x, y) | 0 \leq x \leq 60, 0 \leq y \leq 60\}.$$

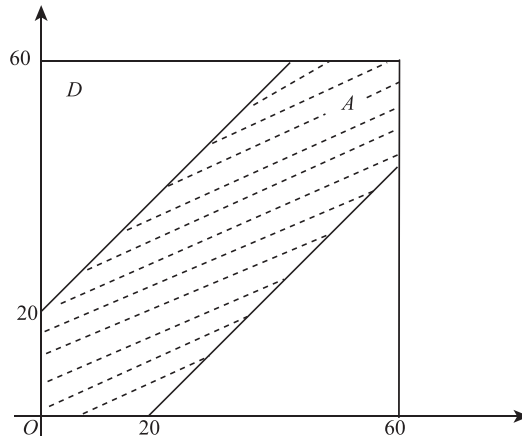


Fig. 1.8 An example of geometrical probability model

We use A to denote the event that the two friends can meet each other. Then event A can be written as

$$A = \{(x, y) | 0 \leq x \leq 60, 0 \leq y \leq 60, |x - y| \leq 20\}.$$

According to the computation formula, we know that

$$\begin{aligned}
 P(A) &= \frac{\text{Area}(A)}{\text{Area}(D)} \\
 &= \frac{60 * 60 - (60 - 20) * (60 - 20)}{60 * 60} \\
 &= \frac{36 - 16}{36} \\
 &= \frac{5}{9}.
 \end{aligned}$$

Remark 1.5 *Symbol “*” used above stands for a multiplication sign. It is also used for an operator of convolution between two functions in the sequel. Readers may find the meaning of the symbol according to contents.*

1.2.4 Conditional probability and three important formulas

In this subsection, we will discuss conditional probabilities of events and three classical formulas, which play an important role in probability computation.

1. Conditional probability

The physical meaning of the probability of an event is to measure the possibility of the event in one experiment. This kind of probability is always called an **unconditional probability**. For example, $P(A)$ is referred as the unconditional probability of event A . Sometimes, before we discuss the probability of some event, say A , we have already known that some other event had already occurred, say B . This kind of probability is always called a conditional probability and is denoted by $P(A|B)$.

Definition 1.5 Conditional probability

Suppose that $B \in \mathcal{F}$ is an event and $P(B) > 0$. For any event $A \in \mathcal{F}$, the **conditional probability** of A given event B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1.2)$$

Remark 1.6 *In the procedure of discussing conditional probabilities, note that we have already known that event B occurred, and we can regard event B as a new sample space $\Omega' = B$ (Fig. 1.9).*

Using the idea similar to the geometrical probability model, to compute the conditional probability of A given B is to compute the probability of $A \cap B$ under the new

sample space $\Omega' = B$. Therefore,

$$\begin{aligned} P(A|B) &= \frac{\text{Area}(A \cap B)}{\text{Area}(B)} \\ &= \frac{\text{Area}(A \cap B)/\text{Area}(\Omega)}{\text{Area}(B)/\text{Area}(\Omega)} \\ &= \frac{P(A \cap B)}{P(B)}. \end{aligned}$$

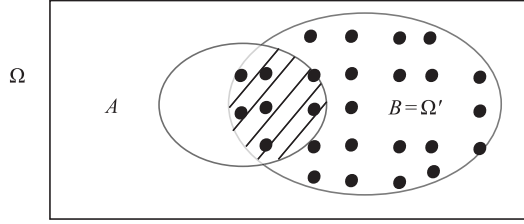


Fig. 1.9 Figure for conditional probability

2. Multiplication formula for probabilities

From Definition 1.5, we know that for two events A and B , if $P(B) > 0$, then by multiplying both sides of equation (1.2) by $P(B)$, we have that

$$P(AB) = P(B)P(A|B).$$

This equality also holds for the case that $P(B) = 0$. In fact, if $P(B) = 0$, then the right hand side is 0, while the complete probability space indicates that the left hand side is also 0. Symmetric property indicates that $P(AB) = P(A)P(B|A)$. It is called a **multiplication formula** for probabilities of two events. For the case of multiple events, if $n \geq 2$ and $A_1, A_2, \dots, A_n \in \mathcal{F}$, we have that

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}). \quad (1.3)$$

Example 1.13 (Example 1 on page 24 in Ref. [1]) Consider a family with two children. We ask for the probability that both children are boys, assuming that

- (a) The older child is a boy;
- (b) At least one of the children is a boy.

Solution The sample space is

$$\Omega = \{BB, BG, GB, GG\},$$

where BG means that the older child is a boy and the younger is a girl, etc.

Suppose that all sample points are equally probable:

$$P(BB) = P(BG) = P(GB) = P(GG) = \frac{1}{4}.$$

Let C be the event that the older child is a boy, and D that the younger child is a boy. Then $C \cup D$ is the event that at least one child is a boy, and CD is the event that both children are boys. In question (a) we want to find the conditional probability $P(CD|C)$, and in (b), the conditional probability $P(CD|C \cup D)$.

It is easy to see that

$$\begin{aligned} P(CD|C) &= \frac{P(CD)}{P(C)} = \frac{1/4}{1/2} = \frac{1}{2}, \\ P(CD|C \cup D) &= \frac{P(CD)}{P(C \cup D)} = \frac{1/4}{3/4} = \frac{1}{3}. \end{aligned}$$

3. Formula for total probability

Consider a *complete set of disjoint events* $\mathcal{D} = \{A_1, \dots, A_n\} \subset \mathcal{F}$, which means that: (i) $A_i \cap A_j = \emptyset$ for $1 \leq i \neq j \leq n$; (ii) $\bigcup_{\kappa=1}^n A_\kappa = \Omega$. For any event $B \in \mathcal{F}$, we have that

$$\begin{aligned} P(B) &= P(BA_1 + \dots + BA_n) \\ &= P(BA_1) + \dots + P(BA_n) \\ &= P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n) \\ &= \sum_{\kappa=1}^n P(A_\kappa)P(B|A_\kappa). \end{aligned} \tag{1.4}$$

It is called the **formula for total probability**.

Example 1.14 (Example 2 on page 26 of Ref. [1]) An box contains M balls, m of which are “lucky”. We ask for the probability that the second ball drawn is lucky (assuming that the result of the first draw is unknown, that a sample of size 2 is drawn without replacement, and that all outcomes are equally probable).

Solution Let A be the event that the first ball is lucky, B the event that the second ball is lucky. Then the conditional probability indicates that

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{m(m-1)}{M(M-1)}}{\frac{m}{M}} = \frac{m-1}{M-1},$$

$$P(B|\bar{A}) = \frac{P(\bar{A}B)}{P(\bar{A})} = \frac{\frac{m(M-m)}{M(M-1)}}{\frac{M-m}{M}} = \frac{m}{M-1}.$$

It implies from the formula for the total probability that

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) \\ &= \frac{m}{M} \cdot \frac{m-1}{M-1} + \frac{M-m}{M} \cdot \frac{m}{M-1} = \frac{m}{M}. \end{aligned}$$

It's easy to see that $P(B) = P(A) = \frac{m}{M}$, which means that if the result of the first ball is unknown, it does not affect the probability that the second ball is lucky. In other words, whether a drawn ball is lucky or not does not depend on the drawing order. This is the reason why drawing lots is often applied to our daily lives.

4. Bayes's formula

Let $\mathcal{D} = \{A_1, \dots, A_n\}$ be a complete set of disjoint events as in the former paragraph and $B \in \mathcal{F}$. Then the **Bayes's formula** is as the following, for $1 \leq i \leq n$,

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i B)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)}. \end{aligned} \quad (1.5)$$

In the application area, $P(A_i)$ is often regarded as a **priori probability** of A_i , while $P(A_i|B)$ is called a **posteriori probability** of A_i after the occurrence of event B , $1 \leq i \leq n$.

Example 1.15 (Example 3 on page 27 of Ref. [1]) Let a box contain two coins: A_1 , a fair coin with probability $\frac{1}{2}$ of falling H ; and A_2 , a biased coin with probability $\frac{1}{3}$ of falling H . A coin is drawn at random and tossed. Suppose that it falls head. We ask for the probability that the fair coin was selected.

Solution We use A_1 to indicate that a fair coin is selected and A_2 to indicate that the biased coin is selected. H and T indicate that head and Tail fall respectively. This is represented as

$$P(A_1) = P(A_2) = \frac{1}{2},$$

and

$$P(H|A_1) = \frac{1}{2}, \quad P(H|A_2) = \frac{1}{3}.$$

Multiplication formula indicates that

$$\begin{aligned} P(A_1H) &= P(A_1)P(H|A_1) = \frac{1}{4}, & P(A_1T) &= \frac{1}{4}, \\ P(A_2H) &= \frac{1}{6}, & P(A_2T) &= \frac{1}{3}. \end{aligned}$$

Then, Bayes's formula implies that

$$P(A_1|H) = \frac{P(A_1)P(H|A_1)}{P(A_1)P(H|A_1) + P(A_2)P(H|A_2)} = \frac{3}{5},$$

and therefore

$$P(A_2|H) = 1 - P(A_1|H) = \frac{2}{5}.$$

Since, $P(A_1|H) = \frac{3}{5} > \frac{2}{5} = P(A_2|H)$, we always believe that the fair coin is selected.

1.2.5 Independence of events

From the previous subsection, we know that for any two events, $A, B \in \mathcal{F}$, unconditional probability $P(A)$ is a priori probability and conditional probability $P(A|B)$, provided $P(B) > 0$, is a posteriori probability. Generally, $P(A)$ is not necessarily equal to $P(A|B)$. However, if $P(A) = P(A|B)$, then we say that event A is independent of event B . It is easy to see that event A being independent of event B means that whether event B occurred or not has no effect on the probability of event A .

Applying the multiplication formula we know that event A being independent of event B means that

$$P(AB) = P(A)P(B).$$

Suppose that $P(A) > 0$, then event B is independent of A if $P(B) = P(B|A)$. Similar arguments indicate that $P(AB) = P(A)P(B)$. This formula also holds for the case that $P(B) = 0$ or $P(A) = 0$. Above all, we have the following definition.

Definition 1.6 Independence or statistically independence of two events

Let $A, B \in \mathcal{F}$, events A and B are called **independent** or **statistically independent** (with respect to the probability P) if

$$P(AB) = P(A)P(B). \quad (1.6)$$

The concept of independence of two events can be extended to multiple events.