Hung T. Nguyen
Tonghui Wang

概率统计高级教程

Ⅰ 统计学的概率基础

A GRADUATE COURSE IN PROBABILITY AND STATISTICS

Volume Ⅰ
Essentials of Probability for Statistics

Tsinghua University Press

Hung T. Nguyen
Tonghui Wang

概率统计高级教程

Ⅰ 统计学的概率基础

A GRADUATE COURSE IN PROBABILITY AND STATISTICS

Volume Ⅰ
Essentials of Probability for Statistics

Tsinghua University Press

Bei Jing

图书在版编目(CIP)数据

概率统计高级教程. 1，统计学的概率基础＝A Graduate Course in Probability and Statistics. Volume I，Essentials of Probability for Statistics：英文/(美)源亨(Hung T. Nguyen)，(美)王通会(Tonghui Wang)编著. —北京：清华大学出版社，2008.4

ISBN 978-7-302-16726-6

Ⅰ. 概… Ⅱ. ①源… ②王… Ⅲ. ①概率论—教材—英文 ②数理统计—教材—英文 Ⅳ. O21

中国版本图书馆 CIP 数据核字(2008)第 001060 号

责任编辑：刘 颖
责任校对：王淑云
责任印制：孟凡玉
出版发行：清华大学出版社　　地　址：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　邮　编：100084
社 总 机：010-62770175　　邮　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京市清华园胶印厂
经　销：全国新华书店
开　本：165×230　印　张：20　字　数：431 千字
版　次：2008 年 4 月第 1 版　印　次：2008 年 4 月第 1 次印刷
印　数：1～2000
定　价：38.00 元

Preface

This is the first half of a text for a beginning graduate course in theoretical statistics. Since a strong background in statistics requires a strong background in probability theory, we divide the text into two volumes. This Volume I is devoted to probability while Volume II is devoted to statistics.

This is an introduction to probability and statistics from the ground up, designed for students who need a solid understanding of statistical theory in order to pursue higher education and research as well as using statistics in their careers.

Essentially, the material in this text is standard for an introductory course in statistics at all universities. As such, there exists a large number of similar texts. The reason for writing another text can be explained by the following distinctions with existing texts.

(i) This text is written for students. Of course, the instructors, when using this text, can provide additional topics or their favorite proof techniques, but we have students in mind in the hope that they will be able to read through the text without tears! This includes self-study. The main topic for students taking this course is statistics. As such, at the very beginning, it should be clearly explained why they need to study probability theory with strong emphasis in mathematics. We use the term "theoretical statistics" to classify this study, as opposed to "applied statistics". We avoid the term "mathematical statistics" for two reasons. First, although mathematics is the machinery needed to investigate statistical theory, there is no need to over-emphasize it. After all, we need mathematics in all fields of science. Second, we should not give the impression that statistics is reserved for mathematicians!

(ii) Introducing probability theory as the first step towards statistical analysis, we should make students appreciate the approach. The material presented in this first volume is also standard and sufficiently solid for students whose interests might not be in statistics, but in probability and related topics. For statistically oriented students, we bring in concepts and techniques from mathematics only when needed. We motivate every mathematical concept used. Our point of view is this: students should think about the material as interesting concepts and techniques for statistics, rather than a burden of heavy mathematics.

(iii) For students to read the text, line by line, with enthusiasm and interest, we write the text in the most elementary and simple manner. For example, a result can be proved by several different methods. We choose the simplest one, since, in our view, students need a simple proof that they can understand, at least in a first course on the topic. To assist them in their reading we do not hesitate to provide elementary arguments, as well as supplying review material as needed. Also, we trust that in reading the proofs of results, the students will learn ideas and proof techniques that are essential for further studies. For this to be efficient, proofs should be given with great care, keeping in mind that we are not in a hurry to give the shortest proof to get the result, but we are guiding our student readers in their learning of proof techniques.

(iv) The material is presented in a logical order, connecting one topic nicely to another. We start from the ground up and guide the students, step by step, to get deeper into the analysis. We are not afraid to introduce what we could call "sophisticated mathematics" such as Caratheodory theorem, or σ-finite measures! This is so because, on the one hand, these are needed for a serious study of foundational statistics, and, on the other hand, we introduce them in a friendly way so that students just feel like learning a new but accessible concept in calculus! The benefit is twofold. Students go through the material smoothly without being forced to accept mysterious results, and will be exposed to these concepts at this level prior to further studies. After all, an introduction to probability for statistics or for other goals should start from simple things ssuch as "random samples", to more sophisticated things such as "sample means as Stieltjes integrals".

(v) In the above spirit, students will get to learn about Caratheodory's theorem, Fubini's theorem, Fatou's lemma, Lebesgue dominated convergence theorem, Lebesgue-Stieltjes theorem, and the Radon-Nikodym theorem. Only a few theorems such as Caratheodory's theorem, Fubini's theorem and the general decomposition of distribution functions are not proved, but complete references are provided. Whenever, a result is not proved, a reference is given.

(vi) Another point of our pedagogy is that students, even in a first course, should be exposed to some advances in statistical theory. An example is the concept of random sets. Biased by our own research interests, we choose to make this topic familiar to students in probability and statistics. Also, additional basic results for statistical theory should be included. As such, students will be exposed to copulas, Sklar's theorem, Choquet's theorem, capacity functionals of random sets, conditional events, large deviations, Glivenko-Cantelli theorem, Choquet integral, Kolmogorov consistency theorem, Portmamteau theorem, Paul Lévy theorem, etc. Note that the sections or parts with * are further materials for interested readers and may be skipped without interruputing the flow of the text.

In summary, the text is not a celebration of how great probability theory is, but simply a friendly guide for students to appreciate the contributions of mathematics to the field of statistics. The material presented in this Volume I is the minimum background, in our view, for the solid introduction to statistical theory in Volume II.

Giving the technological dominance in today's life style, it is again an opportunity to remind our students of the fundamental contributions of mathematics to all fields of science. The appreciation and understanding of such contributions are essential for any scientific career.

We thank our families for their love and support during the writing of this text. Our Department of Mathematical Sciences at New Mexico State University provided us with a constraint-free environment for carrying out this project. We thank Dr. Ying Liu of Tsinghua University Press for asking us to write this series of two-volume text for Tsinghua University Press. Finally, we thank all participants of our weekly Statistics Seminar at New Mexico State University, 2002-2005, for their discussions on statistics of random sets and especially for their insistence that we should include the topic of random sets in a first course in probability.

Hung T. Nguyen and **Tonghui Wang**
Las Cruces, New Mexico, USA
December, 2007.

Contents

Chapter 1

Models for Random Experiments

These beginning lectures aim at providing practical motivations for general mathematical models for random phenomena, namely probability spaces.

Essentially, statistics is a science for making inference from samples to populations. Data are observations obtained from samples. They are viewed as outcomes of random experiments.

An experiment is the making of an observation. An experiment is said to be *random* if its outcomes can not be predicted with certainty. For example, the experiment of tossing a fair coin three times is a random experiment, since the outcomes (heads or tails) cannot be predicted with certainty before each toss.

The first building block for a statistical science is the modeling of random experiments (or random phenomena). We will proceed to motivate *probability spaces* as general mathematical models for random phenomena.

1.1 Games of Chance

Let us first take a closer look at random experiments which are familiar to almost all of us, the games of chance.

Example 1.1 *Consider the experiment of rolling a pair of dice and observing the numbers shown. Suppose that you want to bet on "a sum of 7".* □

In a game of chance such as this, we are interested in the "chances" of "events" related to the game (since obviously, we are not flipping a coin to predict the weather of tomorrow). Thus, two related questions are: what is chance? what is an event? Now, in above Example 1.1, obviously, you want to know the "possibility" of winning, or more precisely, the "chance" of winning. In fact, while it is possible that an outcome of the experiment could result in a pair of numbers adding up to 7, you want to know more: how often does such a situation occur since such information could provide an idea about the chance of winning the bet? To obtain such information, we need to look at the structure of the experiment. While we cannot predict with certainty a specific outcome, we can list all possible outcomes of the experiment. They are pairs (x, y), x, $y = 1, 2, \cdots, 6$, where x denotes the number shown on the first die and y is the number shown on the second die. The collection of all possible outcomes of the experiment is called the *sample space* of that experiment, denoted by

$$\Omega = \{(x, y) : \quad x, \ y = 1, 2, \cdots, 6\}.$$

You are betting on the *event* that the outcome (x, y) is such that $x + y = 7$. Specifically, that event A is the collection of all outcomes (x, y) in Ω such that $x + y = 7$, i.e.

$$A = \{(x, y) : \quad x + y = 7\}.$$

Thus, an event is a subset of the sample space Ω. If the outcome is $\omega = (2, 5)$, say, then $\omega \in A$, and A is realized or A occurs. Any subset of Ω defines an event. Thus, associated with the sample space Ω is the collection of events related to the experiment, and we denote it as $\mathcal{A}$. In this example, $\mathcal{A}$ is the collection of all possible subsets of Ω including the empty set $\varnothing$ (impossible event). This collection is called the *power set* of Ω, denoted by $\mathcal{P}(\Omega)$.

In games of chance such as the above, humans perform the experiment and we do know the structure of the game. In addition, aimed with an (naive/common sense) intuitive idea about "chance", we arrive, in a natural way, at describing the experiment by three basic components: the set of all possible outcomes of the experiment, denoted as Ω, and giving it a name, the *sample space* of the experiment; *events* related to this experiment are subsets of Ω so that the set of all subsets of Ω , denoted as $\mathcal{P}(\Omega)$ which we call it the power set of Ω, is the collection of events of interest; and we are able to write down a formula for computing "chances" for events to occur, namely a mapping P from $\mathcal{P}(\Omega)$ to the unit interval $[0, 1]$. In summary, our experiment is described as a triple $(\Omega, \mathcal{P}(\Omega), P)$. Let us take a closer look at this mathematical model. Since Ω is the set of all outcomes of a given experiment, it can be of various forms depending on experiments. For example, in games of chance, Ω could be a finite or infinitely countable subset of the set of non-negative integers $\mathbb{N} = \{0, 1, 2, \cdots\}$, and in an experiment like picking at random a number in the interval $[0, 1]$, Ω is the uncountable set $[0, 1]$. Next, $\mathcal{P}(\Omega)$, as the collection of events in games of chance with finite sample spaces as the example above, has some algebraic structure which allows us, among other things, to carry out computations of chances in some convenient way.

The basic structure of a collection of sets like $\mathcal{P}(\Omega)$ is as follows. First, let us specify basic notations in set theory. For A and B, subsets of Ω, the *complement* of A, denoted by A^c, is the set consisting of all elements in Ω that are not elements of A; the *union* of A and B, denoted as $A \cup B$, is the set consisting of all elements that are either in A or B or both, and the *intersection* of A and B, denoted as $A \cap B$, is the set consisting of all elements that are in both A and B. The empty set is denoted as $\varnothing$. Since the following properties of $\mathcal{P}(\Omega)$ can be shared by other smaller sub-collection $\mathcal{A}$ of subsets of Ω, let us put $\mathcal{A} = \mathcal{P}(\Omega)$ to achieve generality later. The following are obvious for $\mathcal{A} = \mathcal{P}(\Omega)$:

(i) $\Omega \in \mathcal{A}$;

(ii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$; and

(iii) if A and B are both in $\mathcal{A}$, then $A \cup B \in \mathcal{A}$.

Let us give a name to *any* collection $\mathcal{A}$ of subsets of Ω satisfying the above three properties: $\mathcal{A}$ is called a *field* (of subsets) or an *algebra* (of subsets).

Thus, we specify the structure of our experiment as a pair $(\Omega, \mathcal{A})$ with the meaning indicated in the above example. We are interested in finding the chances for the events of interest to happen. If this is possible, then we say that events are *measurable*, and hence we call $(\Omega, \mathcal{A})$ a *measurable space*. Next, to complete our mathematical description of our random experiment, we need to find a way to describe the random evolution of the experiment, i.e. a way to assign chances to events, so that, even if we cannot predict with certainty the occurrence of some specific event, we can at least specify its chance of occurrence.

The approach we are going to use for the example above is similar to all games of chance which form the core of the so-called *classical probability theory* that provides mathematical models for random experiments having the following characteristics:

(i) The sample space Ω is finite.

(ii) It is reasonable to accept that all outcomes are similar as far as their chances of occurrence are concerned. Put it differently, outcomes are *equally likely*, i.e. all outcomes have the same chance to occur.

Our experiment of rolling a pair of dice makes the above characteristics apparent. In this context, the chance of our event A to happen can be quantified as the ratio of favorable outcomes over all possible outcomes, i.e. $\#(A)/\#(\Omega)$, where $\#(A)$ denotes the cardinality (number of elements) of the event A. Thus, the chance of A is defined to be the *probability* of A, denoted as $P(A) = \#(A)/\#(\Omega)$.

Now we arrive at a construction of an assignment of probabilities to events, i.e. P is a map, not from Ω, but from $\mathcal{A}$ to the unit interval $[0,1]$. Thus we can compute $P(A)$ by the above formula for any given $A \in \mathcal{A}$.

Putting P together with the measurable space $(\Omega, \mathcal{A})$, we obtain the triple $(\Omega, \mathcal{A}, P)$, called a *probability space*, in which P is called a *probability measure*.

A probability space is a mathematical model for a random experiment.

In the context of classical probability theory, the assignment of probabilities to events is based upon the above formula that involves *counting techniques* and hence falls into the area of *combinatorics*. Students are encouraged to practice their computations by doing as many as possible of the exercises at the end of the chapter.

Next, from $P(A) = \#(A)/\#(\Omega)$, we see easily that

(i) $P(\Omega) = 1$ and

(ii) if $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.

The property (ii) is called the *finite additivity* property of P, since it involves only a finite number of events. If fact, (ii) is equivalent to:

(ii*) If $A_1, A_2, \cdots, A_n$, $n \geqslant 1$, are pairwise disjoint, i.e. $A_i \cap A_j = \varnothing$, $i \neq j$, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i). \tag{1.1}$$

Indeed, (1.1) is true when $n = 2$ by (ii). We proceed by induction. Suppose that (1.1) is true for $n - 1$. Then

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left[\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right] = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) \quad \text{by (ii)} \\ &= \sum_{i=1}^{n-1} P(A_i) + P(A_n) = \sum_{i=1}^{n} P(A_i). \end{aligned}$$

In classical probability, we can take any map $P : \mathcal{A} \to [0, 1]$ satisfying properties (i) and (ii) above to be a candidate for quantifying the concept of chance. Such a map is called a *probability measure*.

Having a specific formula for computing $P(A)$, which we interpret as the probability of event A to occur, do we really capture the concept of chance? In fact, we have not even discussed the *meaning* of probability! we will not enter the discussions on this issue from the frequency viewpoint or from the subjective viewpoint, but instead, we follow an *axiomatic approach* to probability due to Kolmogorov (1933). However, evoking the frequency approach will provide a simple motivation for discussing $(\Omega, \mathcal{A}, P)$ for more complicated random experiments (see next section).

1.2 Experiments with Infinitely Many Outcomes

In using statistics to gain information about, say, lifetimes of some type of patients, we are in fact dealing with random experiments whose sample spaces

Ω are infinite. Any real number $t > 0$ is the possible lifetime of a patient, so that $\Omega = (0, \infty)$, which is uncountably infinite. Experiments of this type and more complicated ones are outside of the classical probability framework. The most obvious thing to observe is that we cannot use the formula $P(A) = \#(A)/\#(\Omega)$ to assign probabilities to events. Let us examine the following example.

Example 1.2 *Consider the random experiment consisting of picking a point at random from the interval* $[0, 1)$. *The sample space of this experiment is clearly* $\Omega = [0, 1)$. *In order to provide a (probabilistic) model for this experiment, we need to specify a way to assign probabilities to events of this experiment.*

This experiment has a flavor of a characteristic of experiment in classical probability theory, namely that "outcomes are equally likely". Thus instead of cardinality of a subset $A \subseteq \Omega$ (viewed as the event that outcome lies in A), we can use the "length" of A. While it is clear that the length of Ω, i.e. the length of the interval $[0, 1)$, is one what it is not clear is the length of an arbitrary subset A of Ω! Intuition would suggest to define $P(A)$ as the length of A. But it is here that a technical difficulty arises. The length of an interval $A = [a, b]$ is $b - a$, and the length of a subset A which is a countable union of disjoint intervals, say, $A = \bigcup_{n \geqslant 1}[a_n, b_n]$, is $\sum_{n \geqslant 1}(b_n - a_n)$. Then, roughly speaking, the question is: what is the largest class $\mathcal{B}$ of subsets of Ω for which their lengths can be defined? These subsets will be called *measurable sets.* They correspond to events where probabilities can be assigned. Of course, any subset A of Ω is an event in the real sense of the term, namely "the outcome lies in A", but we are obviously interested only in events for which we can assign probabilities.

The above technical problem is solved in analysis, in fact in Lebesgue measure theory. We will return to this in more details in Chapter 2. For now, it suffices to say this. The class $\mathcal{B}$ is strictly smaller than $\mathcal{P}(\Omega)$. Thus our collection of events (of interest) should be $\mathcal{A} = \mathcal{B}$. Moreover, similar to $\mathcal{P}(\Omega)$, this $\mathcal{A}$ satisfies the following properties:

(i) $\Omega \in \mathcal{A}$;

(ii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$; and

(iii) if $A_n \in \mathcal{A}, n \geqslant 1$, then $\bigcup_{n \geqslant 1} A_n \in \mathcal{A}$.

The property (iii) is stronger than the corresponding property of a field (or an algebra).

A collection $\mathcal{A}$ of subsets of Ω satisfying (i), (ii), and (iii) above is called a *σ-algebra* (or *σ-field*) of subsets of Ω. The prefix σ refers to the infinitely countable property in (iii). Thus we obtain a *measurable space* $(\Omega, \mathcal{A})$, where Ω is a set and $\mathcal{A}$ is a σ-field of subsets of Ω.

Next, we need to specify $P : \mathcal{A} \to [0, 1]$ to complete our description of our experiment of Example 1.2. As stated earlier, for $A \in \mathcal{A}$, $P(A)$ will be taken to be the length of A. Due to properties of the concept of length, P satisfies the following properties:

(a) $P(\Omega) = 1$ and

(b) If $\{A_n,\, n \geqslant 1\}$ is a sequence of finite or infinitely countable pairwise disjoint elements of $\mathcal{A}$, then

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

The property (b) is called the *σ-additivity* of P.

A *probability measure* is a map $P : \mathcal{A} \to [0,1]$, satisfying (a) and (b) above. The triple $(\Omega, \mathcal{A}, P)$ is called a *probability space* which provides a model for a random experiment. Note that this description covers the case of classical probability, and in fact, is general enough for all types of random experiments.

In summary, the *probabilistic model* of an arbitrary random experiment is *a probability space* $(\Omega, \mathcal{A}, P)$, where Ω is a set, $\mathcal{A}$ is a *σ-field* of subsets of Ω, and P is a *probability measure* defined on $\mathcal{A}$.

Remark. The sample space $[0,1)$ is the same as the sample space $\{0,\,1\}^{\mathbb{N}}$ of the experiment of tossing a fair coin indefinitely, i.e. Ω is uncountable. In such an experiment, we are interested in a subset of Ω of the form

$$A = \bigcup_{n=1}^{\infty} \bigcap_{k\geqslant n} \left\{\omega : \left|x_k(\omega) - \frac{1}{2}\right| < \varepsilon\right\},$$

where $\varepsilon > 0$, and $x_k(\omega)$ denotes the proportion of number of 1's (heads) in the first k tosses in ω. The reason is that A is the subset of Ω on which $x_k(\omega) \to 1/2$ as $k \to \infty$. Thus we would like to compute $P(A)$. For that, we need A to be in $\mathcal{A}$ as well as extending the finite additivity property of P to σ-additivity. □

Specifically, a stronger property for $\mathcal{A}$ is:

$$\bigcup_{n\geqslant 1} A_n \in \mathcal{A} \qquad \text{for all} \quad A_n \in \mathcal{A}, \quad n \geqslant 1.$$

And the (finite) additivity property of P should be strengthened to the countable case, namely, for all $A_n \in \mathcal{A}$, $n \geqslant 1$, pairwise disjoint, i.e. $A_n \cap A_m = \varnothing$, $n \neq m$,

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

Thus, we arrive at the probability space $(\Omega, \mathcal{A}, P)$, with the new properties, as a mathematical model for general random experiments. With this model, it can be shown that, for an uncountable sample space like $\Omega = [0,1)$, it is impossible to define probability measures on its power set $\mathcal{P}(\Omega)$, so that the domain of P (i.e. events) is strictly contained in $\mathcal{P}(\Omega)$. See Billingsley (pages 45-46, 1995) for details.

1.3 Structure and Properties of Probability Spaces

Motivated by extreme cases for random experiments, as above, we arrive at *abstract models* (general models) for arbitrary random experiments, namely probability spaces $(\Omega, \mathcal{A}, P)$! In Chapter 2, we will give examples of probability spaces in many applications. But here, we need to take a closer look at probability spaces, in the abstract setting, for computation purposes in applications.

Let Ω be a set which is intended to be the sample space of some random experiment. The collection of events is some σ-field $\mathcal{A}$. How $\mathcal{A}$ is chosen depends very much on the structure of Ω. There are two trivial σ-fields of subsets of Ω, namely $\mathcal{P}(\Omega)$ and $\{\varnothing, \Omega\}$, for which

$$\{\varnothing, \Omega\} \subseteq \mathcal{A} \subseteq \mathcal{P}(\Omega).$$

That is to say $\{\varnothing, \Omega\}$ is a *sub-σ-field* of $\mathcal{A}$, which is, in turn, a sub-σ-field of $\mathcal{P}(\Omega)$. Later, in applications, we will deal with many non-trivial sub-σ-fields of $\mathcal{A}$, each such sub-σ-field represents some information obtained, say, up to a certain time, and will play an essential role in prediction problems.

Note that σ-fields are domains of probability measures and their constructions are in general delicate. For each sample space Ω, we need to specify a σ-field $\mathcal{A}$ of its subsets. Then any $P : \mathcal{A} \to [0, 1]$ will be viewed as a model for the random experiment under study. As we will see, in general it is quite plausible to be able to assign probabilities to simple subsets of Ω , e.g. intervals when $\Omega = \mathbb{R}$. Then the specification of an appropriate $\mathcal{A}$ can be done as follows. Let $\mathcal{C}$ be a collection of (simple) subsets of Ω. Since $\mathcal{C} \subseteq \mathcal{P}(\Omega)$, the collection of σ-fields which contain $\mathcal{C}$ is not empty. Moreover, it is easy to verify that any *arbitrary intersection* of σ-fields is a σ-field. Thus, we have the following definition.

Definition 1.1 *The* **σ-field generated** *by a class $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ is the smallest σ-field containing $\mathcal{C}$.*

Remark. The partial order relation between σ-fields is, of course, inclusion. The σ-field generated by $\mathcal{C}$, usually denoted as $\sigma(\mathcal{C})$, is precisely the intersection of all σ-fields containing $\mathcal{C}$. In the next chapter, we will consider the case where $\Omega = \mathbb{R}$, the set of all real numbers. □

The following analysis is useful for constructing $\sigma(\mathcal{C})$. First, observe that here we are dealing with sets rather then points (say, of Euclidean spaces).

The space $\mathcal{P}(\Omega)$ is partially ordered by the inclusion $\subseteq$ (contained in). By analogy with limit concepts of real numbers, we consider limits of sequences of sets as follows.

Let $\{A_n, n \geqslant 1\}$ be a sequence of subsets of Ω.

(i) If $\{A_n, n \geqslant 1\}$ is said to be *non-decreasing*, i.e. $A_n \subseteq A_{n+1}$ for $n \geqslant 1$, in symbol $A_n \nearrow$, then define

$$\lim_{n\to\infty} A_n = \bigcup_{n\geqslant 1} A_n.$$

Remark. This concept of limit for sets can be also expressed in terms of their *indicator functions*:

For $A \subseteq \Omega$, we put it in a bijective correspondence with the function $I_A : \Omega \to \{0, 1\}$, called the indicator function of the set A, defined as

$$I_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A. \end{cases}$$

Then, $\lim_{n\to\infty} A_n$ corresponds to

$$\lim_{n\to\infty} I_{A_n}(\omega) = I_{\bigcup_{n\geqslant 1} A_n}(\omega) \qquad \text{for all} \quad \omega \in \Omega.$$

Note that in mathematics indicator functions of sets are usually called *characteristics functions.* In probability theory, we use another terminology since characteristic functions are reserved for Fourier transforms of probability distributions (see Chapter 4). □

(ii) If $\{A_n, n \geqslant 1\}$ is said to be *non-increasing*, i.e. $A_{n+1} \subseteq A_n$ for $n \geqslant 1$, in symbol $A_n \searrow$, then define

$$\lim_{n\to\infty} A_n = \bigcap_{n\geqslant 1} A_n.$$

(iii) For arbitrary $\{A_n, n \geqslant 1\}$, we let

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} A_k \quad \text{and} \quad \limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k.$$

If

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n,$$

then this common set is taken to be $\lim_{n\to\infty} A_n$. Alternatively,

$$\lim_{n\to\infty} I_{A_n}(\omega) = I_{\lim_{n\to\infty} A_n}(\omega), \quad \text{for all} \quad \omega \in \Omega.$$

Obviously, any σ-field $\mathcal{A}$ is a *monotone class* of sets, i.e. if $A_n \in \mathcal{A}$, $n \geqslant 1$ and either $\{A_n,\, n \geqslant 1\}$ is non-decreasing or non-increasing, then $\lim_{n\to\infty} A_n \in \mathcal{A}$. Indeed, if $\{A_n,\, n \geqslant 1\}$ is non-decreasing, then

$$\lim_{n\to\infty} A_n = \bigcup_{n\geqslant 1} A_n \in \mathcal{A}$$

by definition of $\mathcal{A}$. If $\{A_n,\, n \geqslant 1\}$ is non-increasing, then

$$\lim_{n\to\infty} A_n = \bigcap_{n\geqslant 1} A_n = \left(\bigcup_{n\geqslant 1} A_n^c\right)^c,$$

by *DeMorgan's law* (see Exercise 1.8), and hence is in $\mathcal{A}$.

A field may not be a σ-field, but a field $\mathcal{A}$ which is also a monotone class is a σ-field. Indeed, let $A_n \in \mathcal{A}$, $n \geqslant 1$. Since $\mathcal{A}$ is a field, we have $B_n = \bigcup_{i=1}^n A_i \in \mathcal{A}$, for all $n \geqslant 1$. But $\{B_n, n \geqslant 1\}$ is non-decreasing, so that

$$\lim_{n\to\infty} B_n = \bigcup_{n\geqslant 1} B_n \quad \in \mathcal{A},$$

since $\mathcal{A}$ is also a monotone class, by hypothesis. But

$$\bigcup_{n\geqslant 1} B_n = \bigcup_{n\geqslant 1} A_n.$$

The following results is useful for constructing generated σ-fields.

Theorem 1.1 **(Monotone Class Theorem)** *Let $\mathcal{C}$ be a field of subsets of Ω. Then $\sigma(\mathcal{C})$ is precisely the smallest monotone class containing $\mathcal{C}$.*

Proof. Let $\mathcal{M}$ denote the smallest monotone class containing $\mathcal{C}$. Since $\sigma(\mathcal{C})$ is a monotone class containing $\mathcal{C}$, we have that $\mathcal{M} \subseteq \sigma(\mathcal{C})$.

To obtain the reverse inclusion, it suffices to show that $\mathcal{M}$ is a field, in view of the previous analysis. First, for $A \subseteq \Omega$, let

$$\mathcal{B}(A) = \{B \in \mathcal{P}(\Omega) :\, A \cup B,\, A \cap B^c,\, B \cap A^c \in \mathcal{M}\}.$$

Then $\mathcal{B}(A)$ is a monotone class. Indeed, let $B_n \in \mathcal{B}(A)$, $n \geqslant 1$, be non-increasing, then

$$A \cup \left(\bigcap_{n\geqslant 1} B_n\right) = \bigcap_{n\geqslant 1} (A \cup B_n) \quad \in \mathcal{M},$$