Daguo XIONG

The Natural Axiom System of Probability Theory

Mathematical Model of the Random Universe

The Natural Axiom System of Probability Theory

Mathematical Model of the Random Universe

This page is intentionally left blank

The Natural Axiom System of Probability Theory

Mathematical Model of the Random Universe

Daguo XIONG

Beijing Institute of Technology, China

Translated From Chinese by

WU Jian

World Scientific

New Jersey • London • Singapore • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

THE NATURAL AXIOM SYSTEM OF PROBABILITY THEORY

Copyright © 2003 by World Scientific Publishing Co. Pte. Ltd.
All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-238-408-1

Printed in Singapore by World Scientific Printers (S) Pte Ltd

Preface

Since *Natural Axiom System of Probability Theory* (abbr. NAS) is published, the readers have given us a lot of support and encouragement, and put forward many questions to discuss, and invigorate us very much.

(1) The NAS is observed from probability theory

(a) *"Is it likely that the NAS considers the basic unit for probability theory to be studied to be random test?"*

Indeed really! Many preliminary students feel difficult on study of probability theory. What are the causes? We consider an important cause is that when probability theory is stated the purpose from the very beginning, the reader is demanded to find the probability of some event under ambiguous conditions, so that the preliminary students are at a loss what to do and believe that the method of probability theory is too mysterious to understand.

Improved method is as follows. The preliminary students are able to know that in order to understand an event, people must as much as possible know, and it is best to write down the random test which this event belongs to (**this fact is our principle I of the probability theory**, pp 5, 7, 79). At the same time they are able to know that there are many methods to assign probability values according people's wishes (**this fact is involved in our principle II of the probability theory**, pp 5, 8, 78). Of course, probability theory hopes readers to find the normal probability values tallying demand of question and received by people.

For emphasizing the random test, this book will use three sections to introduce random tests. In the following sections, it is emphasized that the probability is a measure on random test and the three groups of axioms of assignment with probability are given. For random variables, it is emphasized that their contents of probability are the random test generated by them and assignment with probability measure on it, the only difference is that at this time assignment with probability measure is represented as to find distribution function.

(b) *"It is likely no need to study the measure theory first for people can* *understand and grasp the NAS and probability theory."*

This just is the event which the author hopes to occur and to occur at those textbooks. In fact, suppose students have primary knowledge of set theory, i.e.

- the representation methods of a set;
- the operation: complement, union, intersection and difference of sets,

and to agree two basic concludes:

- there are uncountable sets, for example, the real number set; the intervals; the rectangles; the cubes and so on;
- the expansion theorem of probability measure,

then they can understand and grasp the contents of probability theory including the NAS and limitation theory, deeply and rigorously.

Now, the another cause for preliminary students to feel difficult on study of probability theory is probably that some teachers over demand students to solve difficult and eccentric problems using combinatorial method, so that study enthusiasm of the students is dampened. We consider such problems to be identical with the difficult and eccentric proof questions and the graphic questions in the elementary geometry and they are not necessary for the students who do not major in probability theory. They only need to grasp some representative probability spaces of classic type, and understand the thought and essential of probability theory through them.

The author carried out the ideas (a) and (b) above in probability theory teaching, and good effects are received.

(c) *"Can the NAS expand the research range of probability theory? Can it generate the new branches of probability theory?"*

This is most important concernment of probability scholars. It is also the problem to open and to have not answer until now. Here the author talks some points of view which are not mature.

The causation space is a mathematical model of random universe. It is a tremendous container and "living house" of all random tests and probability spaces. It describes the situations of various random phenomena to occur in natural world and human society, so that it is better to reflect complex reality than single probability space. For example, the recognition of human race consists of the recognitions of all individuals. Every person knows and is familiar with a part of events (this part of events may be denoted by $\mathcal{F}_{man}$) and the probability of every event A is estimated by $P_{man}(A)$ (some values are received objectively, some values are received subjectively). Suppose every person recognizes coincidentally front and back, then the recognition of this person to the random world can be denoted by a probability space $(\Omega_{man}, \mathcal{F}_{man}, P_{man})$. The recognition level of human race consists of several thousand million $(\Omega_{man}, \mathcal{F}_{man}, P_{man})$. Because every person has own past experience, knowledge, environment, psychology situations and so on, different person has different probability space. Even if for same person, his probability spaces at different periods are different, too. Those many various probability spaces just reflect numerous and confused recognitions of human race. From this many cooperations, contradictions, conflicts and various actions are generated.

To illustrate this we may use a little thing. The forecast of some stock is divided into five cases: big rise, little rise, unbiased, little fall, big fall. If a part of persons have $P_{man}(big\ raise) \approx 1$ and another part of persons have $P_{man}(big\ fall) \approx 1$, then front part of persons may buy this stock, latter part of persons may sale this stock. If all persons who hold this stock have $P_{man}(unbiased) \approx 1$, then the phenomena of this stock trade seldom occur.

We believe that there are so many probability spaces in the causation space, and these probability spaces are related by countless ties, so the positive answer should exist for the problem (c) above.

(2) The NAS is observed from mathematics

Table 0.1 in Preliminary displays that people may study concretely and vividly random phenomena in the causation space, just like people may study concretely and vividly geometrical figures in Euclidean space. Therefore **the causation space and the Euclidean space are two basic spaces in real world**. The introducing method and the structure of the NAS are very similar with the Hilbert's System of Axioms. **This kind of similarity reflects mathematical beauty, also displays that the NAS is an excellent answer of Hilbert's sixth problem.**

Mathematics circles (including the author) all consider that " the base of probability theory is the measure theory" and "the necessary condition of understanding and grasping probability theory deeply is to study measure theory well". But the study method of "abstractly first, concretely after" does likely not obey human's routine of recognizing. Therefor we insist that this situation should be reversed. In other words, people study concrete and vivid probability theory first, and study the abstract infinite dimension spaces and the measure theory latter. So probability theory can be the background material of abstract spaces, also can put advance new problems and provide new methods to abstract mathematics continuously.

(3) The probability theory is observed from science

Examining the getting process of scientific concludes, the importance of the thought and statistical method of probability theory is found. In fact, majority of laws of the experience science, such as the law of universal gravitation, the Boyle-Mariotte law of gas, the Mendel's heredity laws and so on, came from statistical treat of experiment data (as the Boyle-Mariotte law, the Mendel's heredity laws, etc.), or need statistical test by the experiment data, so the reliability of the laws is guaranteed.

Because there are random phenomena everywhere and people all concern their occurrences, if primary probability theory (including the primary knowledge of set theory, and the probability spaces of classic type, discrete type and geometric type) can be taught popularly, may we ask the probability theory can become the content of hight school in similar with algebra and geometry?

In fact, the thinking method which a general conclude is proved by several individual examples is bad. Now the society has so progressed that a general conclude is proved by statistical method basing experiment data. The thinking method has also changed.

It is evidence of occurring of various economic indexes and economic data, statistical data of investigation for popular will and medical treatment and other a lot of investigation data. Therefore the popularization of the thought and method of probability theory tallies with the demand of the society.

In a book of stating the new idea at first time there must be some defects such like incompleteness, leaks, ambiguous places and it is my hope with heartfelt emotion that the readers can point out shortcomings or give some complements.

Finally, I take this opportunity of expressing my sincere appreciation of that the conscientious translation of my classmate Prof. Wu Jian and the enthusiastic support and guidance of Prof. Gong Guanglu of Tsinghua University and Scientific Editor Dr. Lu Jitan of World Scientific Publishing Company.

Xiong Daguo
Dec. 28, 2002

Preface of Chinese Edition

The probability theory is a mathematical branch in which the random universe, i.e. various random phenomena in nature and human society, are studied. This theory is based on Kolmogorov Axiom System (abbr. KAS). In the KAS the probability space $(\Omega, \mathcal{F}, P)$ is the place which all research start from. In that place Ω is a certain set, its element is called an elementary event; $\mathcal{F}$ is a σ-field in which the element is called an event; P is a probability measure on $\mathcal{F}$, the value $P(A)$ is a size of the possibility of the event A to occur. Then the methods of set theory and measure theory are moved by force into probability theory and by those methods the theoretical system of probability is deduced.

There are at least five defects in this study mode.

(1) The whole image of random universe is not set up. Speaking in similitude, Euclidean space was set up as its whole image in the elementary geometry, so the study of geometrical figures and their measure is intuitive and active. We may ask: why does the probability theory do not so?

(2) By treating probability space absolutely the relationship between inside and outside of probability space is cut off, the intuition and appearance of random phenomena are broken down.

(3) It is not possible to illustrate why the methods of set theory can be used to study random phenomena.

(4) The interior link between condition and probability is not discussed, it is difficulty to set forth the essence of conditional probability and independence. In fact, the conditional probability in the KAS is in confusion.

(5) The core of probability theory was still in the vague. Speaking in similitude, however many kind of modern geometry were put forth, there always are some forms of Euclidean geometry in other kinds of geometries, and it is the source of generation of those geometries. People may ask, is there "Euclidean geometry" in probability theory? What is it?

This book puts forward advance the Natural Axiom System of Probability Theory (abbr. NAS). The NAS has overcome those defects, a kind of new research mode is developed. The NAS transform the random universe as the causation space. The random phenomena in the causation space are not only retained intuitive and appearance, but also obtain many kinds of representative forms. The methods of set theory in probability theory generated from many kinds of representations.

In the view of philosophy, the existence of the different representative forms origins from "the causes" and "the effects" in the law of causation. The methods of set theory and measure theory in probability theory were the mathematical abstract of thinking patterns about the law of causation.

There are various "graphes" in the causation space. The random tests and the probability spaces are main "graphes" in the NAS. Their cores are separately the discrete tests and Borel tests, discrete probability spaces and Kolmogorov's probability spaces. Now, not only single "graphic" can be studied, but also the relations among various "graphes" can be studied and mathematical calculation can be set up.

We reveal the essence of the probability, conditional probability and independence by the concept of random tests. And we discover that the conditional probability defined by Radon-Nikodym theorem is not conditional probability. In fact, its name is not suitable.

The random phenomena can almost be represented by random variables after mathematical treat, and all statistical knowledge and statistical laws of the random variables can be obtained by the statistical knowledge of Kolmogorov probability space. Therefore **the causation space, discrete tests, Borel tests, discrete probability spaces, Kolmogorov probability spaces and random variables** constitute the core of probability theory. By the core people can understand that the random universe is not the chaos which is difficult to ascertain, it is a reasonable world which varies according to statistical laws continuously.

Due to the limitation of the knowledge and the ability and stating first the new idea in this book, there must be some defects such like incompleteness, leaks, ambiguous places and it is my hope with heartfelt emotion that the readers can point out shortcomings or give some complements.

Having finished the book draft, the book is able to have devoted the readers in the great support given by Tsinghua University Press. The Prof. Qin Mingda of Beijing Since and Technology University peruse cautiously all draft of the book and he put forward advance many pertinent and valuable opinions, and raised the quality of it. The author obtained many helps of the Vice Chief Editor Zheng Guiquan of Education and Science Press and the book was promoted. Therefore I represent sincerely thanks here.

Xiong Daguo
Nov. 6, 1999.

Contents

Preliminary

The research object of probability theory is random phenomena in real world (they are called random universe in this book). A. N. Kolmogorov in the year 1933 published his special book[1] in which Kolmogorov's system of axioms was established (it is called Kolmogorov Axiom System, abbr. KAS). The KAS has become firm basis for the vigorous development of probability theory. It is a milestone in developmental history of probability theory.

The criticisms came mainly from the T. Bayes School of Thought. B. de Finetti put forward advance 7 topics for discussion[2]. Some authors established other systems of axioms being slightly different from the KAS[3-7]. Summary article[8] introduced the situation of establishing the axiom systems of subject probability.

We will take part in the rank of critics from another side, and point out 7 topics for discussion as follows.

(1) The system of axioms of probability theory should has completed a great task — **the establishment of the mathematical model of random universe.**

It should constitute an intuitive and strict logical system from the random universe that is look like chaos seemingly. That is, several essential attributes of random universe must first be drown out, a system of axioms should be abstracted from them, then logical deductive inference and mathematical calculation should be conducted in this system of axioms, finally the theoretical system of probability theory should be deduced. The purpose is to discover the more and deeper laws and attributes of random universe.

But the KAS did not put forward advance this task. It only studied some part of random universe, and abstracted the probability spaces from this part. The relations among those probability spaces were not clear.

(2) Why does the probability theory have not itself special research place?

The random phenomena in real world are so concrete and so widespread. But probability theory to study those phenomena has not own "living house"— the special research place[1]. This is not normal. Therefore some mathematicians consider probability theory as a part of measure theory. It is the reason that the elementary event space in the KAS is any set, and probability space is a kind of special measure spaces built at this set.

(3) The KAS breaks down the intuitive appearance of random universe:

Random Universe	**KAS**
(1)All sure events are equal, i.e. there is unique sure event.	(1) Mistake! There exist different sure events in different probability spaces.
(2)There exist the union, intersection, difference of two arbitrary events.	(2)No. It does so only in the same probability space.

It is not difficult to point out that the relations between inside and outside probability space are artificially cut off, and the research range of probability theory and mathematical statistics is reduced.

(4) Why is the event (for example, tomorrow will be clear) a set of elementary events in the KAS? Why are some sets consisted of elementary events the events, and another sets not the events?

This question corresponds to the first discussible topic of Finetti. Furthermore we can philosophically ask some questions as follows:

(5) Why have the methods of the set theory be able to become the basic research methods for random phenomena?

(6) Some of the elementary events in the KAS can be the events, and others can not be the events. If it is not, what is it?

(7) The interior relationship between condition and probability is not discussed in the KAS. The intuitive background of conditional probability and the independence is covered, so it is difficult to extend to general situation.

This question corresponds to the sixth discussible topic of Finetti.

The book has successfully answered those questions, and has extended the KAS as the Natural Axiom System of Probability Theory (abbr. NAS). The NAS describes the random universe as the causation space which is look like Euclidean space and can deduct mathematical inference. Table 0.1 indicates structure and meaning of the NAS by the analogism.

[1]The living house provided by the NAS is the causation space.

Table 0.1 Probability Theory Vs. Euclidean Geometry[9]

Subjects	**Euclidean geometry**	**Probability theory**
Intuitive objects of research	The shapes of the things in real world	The random phenomena in real world (the random universe)
Axiomatic method	The Hilbert's System of Axioms	The NAS
Elements	Points, Straight lines, Planes	Random events
Relations between elements	5 groups of axioms	6 groups of axioms
Products after axiomatize	The Euclidean Space	The Causation Space
Research objects	The geometrical figures The measure of figures	The random tests The probability spaces
Research tasks	The contents of research are to study geometrical figures and the measure of figures. The contents include: a) the properties of single figure and the single quantified figure. b) the relations among different figures or the various quantified figures.	The contents of research are to study random tests and to endow the random tests with probability measures. The contents include: a) the attributes of single random test and single probability space. b) the relations among different random tests and different probability spaces.

It follows from Table 0.1 that **the causation space is a mathematical model of random universe. The causation space and the Euclidean space are two basic spaces in real world.**

The NAS possesses two big characteristics. First characteristic is the requirement abiding by the modernization system of axioms, and is an axiom system of formalization.

The meaning of formalization is, firstly finding the basic object of probability theory — random events which are not regarded definable and is expressed by symbol $A, B, C, \cdots$; secondly, 6 groups of axioms were abstracted from the relations among the elements, and the axioms were symbolized, too; finally, the symbols were logically deducted and run mathematical calculation, new combinations of symbols — various concludes were obtained, furthermore the theoretical system of probability theory has formed.

Because these concrete meanings of symbols need be not taken into account during inference and calculation, thus the concludes are guaranteed for their reliability.

The second characteristic is that the NAS has the intuition and the appearance which is demanded by classic system of axioms. People should do his best to select the axioms for simple and intuitive. So the axioms will be accepted by people without the verification.

The probability theory has very concrete object for research — random universe. Therefore the intuitive appearance is the source to generate probability theory, and also is the basic of its existence. It not only inspires a thought into the scholar, and aids to think, but also provides new research tasks, new directions to research, and points out the developmental directions and stimulates probability theory with the vitality and vigor, so that many edge subjects have generated and very great branch of mathematics has formed.

The two explanations in the arrangement of the contents will be showed as follows:

1) We will firstly introduce the intuitive background of a group of axioms before we introduce it[2].

Although living in the random universe, we have obtained much intuitive, sensational knowledge, but the knowledge is scattered and is not organized, and most people consider it as all in a mess and that it is difficult to understand. Therefore it is our first task to show the intuitive background of every axiom. It is not the only one, we will introduce the intuitive background of random universe in one whole chapter[3]. The purpose is to handle random universe as reasonable and regularized world and to prepare the thought and the public opinion for the NAS to appear.

2) We promptly deduct some basic common conclusions after introduction of a group of axioms. Those functions are three aspects: ① It enables the contents as complete and systematic theory; ② The NAS can be understood by the readers who have acquired the basic knowledge of set theory. ③ It points out for the experts that the basic common conclusions occupy the place in the NAS.

This book is basic mathematical specialized book. To read this book, the reader only needs basic mathematical knowledge, but needs stronger ability of logic thinking.

---

[2]The establishment and training the intuitive appearance are a necessary task in the NAS, and are the combinative important part of this book.

[3]The contents of the chapter 1 has new point of view and belong to popular science.

Chapter 1

Real Background of Probability Theory

The people have accumulated a lot of simple observations and have discovered useful and veritable practical methods when they recognize the random universe — various random phenomena in nature and human society. The axiomatic method is that at first to draw some of the most essential attributes from those source materials, then critically and precisely to treat these attributes by the logical argumentation, finally to built a system of axioms of probability theory. Furthermore **we can recognize the random universe on the new high position.**

In this chapter two principles will be drawn from the most important source materials:

> **The principle I of the probability theory:** *A random event can be recognized only by the relations among it and the other events.*
>
> **The principle II of the probability theory:** *Under the rational planned (or given) condition the possibility of the occurrence of random event can be expressed by the unique real number p in [0,1].*

In the next chapter the first three groups of axioms will be built under the guidance of the principle I; the last three groups of axioms will be built under the guidance of the principle II.

1.1 Research object

The sun, the moon and the stars always are circulating. The atmosphere are always varying, it is raining, snowing, cloudy or clear, and always are alternating. A lot of water are always flowing along the rivers, are coming together into the sea finally. In the biologic world who can adapt oneself to circumstances could live and always propagate itself, become not extinct forever and ever. It is the creation of the kind of human — the spirit of the biological world that there are the war and peace; politics and economics; science, technology and culture; entertainment and sport in human society.

The nature and human society are always varying. The marks reflecting those variations and developments are appearance of the various phenomena, in another words, the occurrence of various events. For example:

(1) The sun rises from the east. (This phenomenon is denoted by A_1, and is called the event A_1.)

(2) The moon disappears from the sky. (the event A_2)

(3) There exist the attract forces between two bodies. (the event A_3)

(4) Under the standard atmospheric pressure the water is heat at 50^oC and is boiling up. (impossible event A_4)

(5) It will be raining tomorrow. (the event A_5)

(6) It will be cloudy tomorrow. (the event A_6)

(7) It will be fine tomorrow. (the event A_7)

(8) The air maximum temperature will be ξ^oC tomorrow. (the event A_8)

(9) The air minimum temperature will be η^oC tomorrow. (the event A_9)

(10) The fetus of some couple will be a boy. (the event A_{10})

(11) The fetus of some couple will be a girl. (the event A_{11})

(12) A plant cultivated from second generation of hybrid pea will flower with white petals. (the event A_{12})

(13) A plant cultivated from second generation of hybrid pea will flower with red petals. (the event A_{13})

The events A_{12} and A_{13} come from the famous inheritance experiment — hybrid pea experiment done by Mr. G. Mendel.

(14) In a pot there are n balls marked separately by the labels $1, 2, \cdots, n$. When any ball is taken up at random, it is i'th ball. (the event B_i, $i = 1, 2, \cdots, n$).

(15) There are 5 substandard products in 100 products, 3 products are taken up at random, they are all substandard products. (the event A_{15})

(16) The number of customers entering a shop on some day is just 5 thousand. (the event A_{16})

(17) The number of customers entering a shop on some day is below 5 thousand. (the event A_{17})

(18) When a coin is thrown on a desk, the head is up. (the event A_{18})

(19) When a coin is thrown on a desk, the tail is up. (the event A_{19})

(20) When a coin is thrown on a desk, the coin erects. (the event A_{20})

(21) When one plays at a dice, the point " i " appears. (the event $C_i, i = 1, 2, 3, 4, 5, 6$)

(22) When one plays at a dice, the even number appears. (the event A_{22}).

(23) When one plays at two dices, the point " i " appears in first dice and the point " j " appears in another dice. (the event $D_{ij},\ i, j = 1, 2, 3, 4, 5, 6$)

(24) When one plays at two dices, the double 6 appear. (the event A_{24})

(25) When one plays at two dices, the sum of two points to appear is larger than 8. (the event A_{25})

The events, such like those, are infinite, it is impossible to list.

All subjects in science study the events. The events are also studied in probability theory — a subject in science. The differences between probability theory and other subject in science are at four respects as follows:

1) The probability theory is only concerned whether the events occur.

2) The other contents of the event itself are not concerned in the probability theory.

3) The events are recognized by the relations among those events in probability theory.

For example, for the event A_7, " it will be fine " is not recognized by the concrete content of A_7 (for example, it is sunshine), and is recognized by that the event A_5 and event A_6 will not occur.

4) If the event is able to occur, then the value (or size) of possibility of its occurrence will be guessed.

Whether a phenomenon, an accident or an event to occur is controlled by various conditions (including many different random unforeseen elements). They are called random phenomenon, or random event because it is emphasized that their occurrences depend on " the chance ". In fact, even if this is the event like "the sun rises from the east", its occurrence depends on " the chance " also. Because if you stand on other stars, or you stand on the poles of the earth, the event A_1 needs not to appear.

Now a character can be abstracted from the random events. It is follows:

The principle I of the probability theory: *A random event can be recognized only by the relations among it and the other events.*

A random event is briefly called an event for the convenience of the language.

1.2 Research task

"The probability theory: a mathematical branch to study the quantitative laws of random phenomena." — the content is pointed at *Chinese Encyclopedia: Mathematics*, p 251. The quantitative law is a very generalized language. According to the discussion in section 1.1, the quantitative law should include the linkage among the random phenomena and the laws of the occurrences of random phenomena.

We would write a basic conviction down as a following principle.

The principle II of the probability theory: *Under the rational planned (or given) condition the possibility of the occurrence of random event can be expressed by the unique real number p in [0,1].*

Under this principle, the law of the occurrence of random phenomena is the useful language in everyday life: "under the given conditions some random phenomenon appears with the possibility $\alpha\%$." or "under the given conditions we infer with the confidence $\alpha\%$ some random phenomenon to appear".

The rational planned (or given) condition is denoted by C. Under the condition C, a concerned random event is denoted by A. For a change, the value p is denoted by symbol $P(A|C)$, it is called the (conditional) probability of the event A given the condition C. Therefore **probability theory has very pure task: to seek $P(A|C)$.**

In order to complete this task, we must transform the condition C, the event A and the probability $P(A|C)$ into the object which can be logically deducted and mathematically calculated. This just is what the NAS wants to do.

The three remarks will be given before ending of this section.

Note 1: What is the essence of randomness? Is the obeyed conviction rational? The probability theory refuses to answer the questions like those. Those belong to philosophy. Certainly, comparing the probability theory based the basic conviction and real random world, the degree of their coincidence can indirectly verify the rationality of the basic conviction and the requisition to correct.

Note 2: The statistical law is the useful language in the probability theory. The profound and collective law contained in a lot of random phenomena is recognized as the statistical law. It is clearer stated that *"there is few law in one test, but when the test is repeated again with 'a great number of times', some law is obeyed by the concludes of these tests, the law is called the statistical law, this kind of tests is called random tests, the phenomena represented in the tests are called random phenomena "*[10].

Note 3: In *the Modern Chinese Dictionary* the law is illustrated as *"the interior necessary linkage among the things "*. In the test playing a coin we have the conclude: "the possibility of the occurrence of the head is 50%. ". According to the illustration of the dictionary, this conclude is not a law; according to the illustration of note 2 this conclude is not a statistical law. But this conclude is certainly a law of occurrence of random phenomena. Therefore this book considers the law of the occurrence is a statistical law.

The knowledge contains more connotations than the law. The knowledge is used to be illustrated as the summation of the recognition and the experience of to research and to reform the world. In this book probability theory is defined as "**the probability theory is a mathematical branch to study the statistical knowledge of random phenomena**". i.e. We identify the statistical knowledge with the quantitative laws, it contains at least two terms: the linkages among random phenomena and statistical laws.

1.3 Probability

In order to understand the principle II, the implication of the probability must be discussed. This concept has widely been used in nature science and society science. It has already widely accepted by the people that the weather forecast with the mode such like that "the probability of tomorrow will be raining' will be 60% ". The concept of probability has likely no challenge again, but an examination of the facts proves quite the contrary. *"It is discovered that the understandings and illustrations of different people are different. This is necessary to influence the point of view of people for probability theory and mathematical statistics and their areas of application. The disputes have appeared not only today but also in its birthday from the point of view of the history of probability theory. During the different periods the disputes accompanied with probability theory has the different point of departures and the different focus, and influence the contents and the developing trend of the subject "*[11]. There is not an illustration which is accepted by all probability scholars until today.

The author's point of view is that every illustration has excavated the essence of the probability from some side. It is important that which condition some illustration is rational and acceptable under. In fact, **the coexistence of many illustrations not only enables us to understand "what is probability" correctly and comprehensively but also produces many applications and even obtains some discoveries which is surprised and is not anticipated.**

Let us to introduce four kinds of popular illustrations.

1.3.1 Classical definition of probability

P. S. Laplace at the year 1814 gave firstly out a definition of probability explicitly. The probability is defined by him as the proportion of the number of " the favorable outcomes" to the total number of " all possible outcomes". This is the classical definition of probability. In some situation the classical definition is clear and explicit.

Example 1 (The experiment of tossing a coin). *A coin is tossed on a smooth desk anyway.* (Experiment $\mathcal{E}_1$)

Solution: A outcome of the experiment is either the head to occur or the tail to occur. Therefore the all possible outcomes are

$$< head >, \qquad < tail > \tag{1.3.1}$$

The condition defining the experiment $\mathcal{E}_1$ is denoted by $\mathcal{C}_1$, the probability of head to occur is denoted by $P(< heads > | \mathcal{C}_1)$. It is due to the event of "head to occur" including only one favorite outcome that according to the classical definition of probability, we have

$$P(< head > | \mathcal{C}_1) = \frac{1}{2} \tag{1.3.2}$$

Similarly,

$$P(< tail > | \mathcal{C}_1) = \frac{1}{2} \tag{1.3.3}$$

Example 2 (The experiment of tossing two coins). *Two coins are tossed on a smooth desk.* (Experiment $\mathcal{E}_2$)

Solution: If the event of "the occurrence of α-side of the first coin and the occurrence of β-side of the second coin" is denoted by (α, β), then all possible outcomes are

$$< head, head >, < head, tail >, < tail, head >, < tail, tail > \tag{1.3.4}$$

The event of "the occurrence of at least one head" consists of $< head, head >$, $< head, tail >$, $< tail, head >$. Therefore we have

$$P(\text{ at least one head} \,|\, \mathcal{C}_2) = \frac{3}{4} \tag{1.3.5}$$

where $\mathcal{C}_2$ is the condition defining the experiment $\mathcal{E}_2$. Similarly, we have

$$P(\text{ two tails} \,|\, \mathcal{C}_2) = \frac{1}{4} \tag{1.3.6}$$

Now, the important ratio 3:1 in the genetics is obtained[1]. i.e.

$$\frac{P(\text{ at least one head} \,|\, \mathcal{C}_2)}{P(\text{ two tails} \,|\, \mathcal{C}_2)} = \frac{3}{1} \tag{1.3.7}$$

[1] To see the discussion of the end of this section.

Example 3 (The experiment of throwing a dice). *A well-regular dice is thrown on a smooth desk.* (Experiment $\mathcal{E}_3$).

Solution: At this time all possible outcomes are

$$<1>, <2>, <3>, <4>, <5>, <6> \tag{1.3.8}$$

here $< i >$ shows the occurrence of point i of the dice. The condition defining the experiment $\mathcal{E}_3$ is denoted by $\mathcal{C}_3$. By the classical definition of probability we have

$$P(<1> \mid \mathcal{C}_3) = P(<2> \mid \mathcal{C}_3) = \cdots = P(<6> \mid \mathcal{C}_3) = \frac{1}{6} \tag{1.3.9}$$

It is noted that only in the some special situations the classical definition of probability is rational and acceptable. For example, in Example 3 the condition $\mathcal{C}_3$ contains some conditions such as the dice is well-regular, the desk is smooth, the throw is random, and so on. Any changes of the condition $\mathcal{C}_3$ would enable Experiment $\mathcal{E}_3$ not to obey the classical definition of the probability, so the equality (1.3.9) is not obtained.

1.3.2 Geometric probability

When the probability theory was developed from the very beginning, people had noted that the number of favorite outcomes and the number of the possible outcomes can both be infinitude. Furthermore a few individual examples appeared, then another illustration for probability was promoted.

Those individual examples are able to generalized. Supposing the G is a region of a plane (or straight line, or space) and the g is a subregion of the G, now we at random throw a particle in the G (Experiment $\mathcal{E}$). What value is the probability of the particle falling into subregion g?

The answer of the question generates the geometric probability. The condition defining the experiment $\mathcal{E}$ is denoted by $\mathcal{C}$ and then the answer is

$$P(\text{ the particle falls into subregion } g \mid \mathcal{C}) = \frac{\mu(g)}{\mu(G)} \tag{1.3.10}$$

where the μ is regarded as the length, the area and the volume respectively when the G, g are regarded as regions of line, plane, space.

De Buffon discussed the geometric probability in detail (at the year 1777), and calculated the geometric probabilities of many problems.

Example 4 (The experiment of Buffon's throwing needle). *There is a plane drawing some parallel lines, all the distances between the contiguous lines equal to a. A needle with the length $l (l \leq a)$ is thrown at random on the plane (Experiment $\mathcal{E}_4$). Please find the probability of the intersection of the needle and any parallel line.*

Solution: The key of the solution still is to obtain "all possible outcomes" and "the favorite outcomes". In order to do so, the distance between the middle point of the needle and the nearest parallel line is denoted by x, the angle between the needle and the parallel line is denoted by ϕ (Figure 1.1(a)). Every pair (ϕ, x) satisfying $0 \le x \le \frac{a}{2}$, $0 \le \phi \le \pi$ defines an outcome of the needle falling on the plane. Therefore all possible outcomes constitute the region G (Figure 1.1(b)).

The needle intersects with the parallel line if and only if $x \le \frac{l}{2} \sin \frac{\phi}{2}$. All (ϕ, x) satisfying the inequality in G constitutes the region g (the shadow in Figure 1.1 (b)). i.e. The event "the needles have intersected with the parallel lines" constitute the region g.

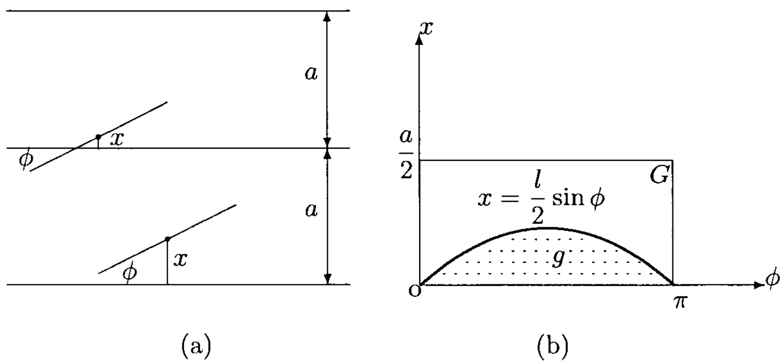

Figure 1. 1

The condition defining the experiment $\mathcal{E}_4$ is denoted by $\mathcal{C}_4$, and believing the geometric illustration of the probability to be rational, then we can obtain from (1.3.10)

$$P(\text{ the needle intersects with the parallel line} \mid \mathcal{C}_4) = \frac{\text{the area of } g}{\text{the area of G}}$$
$$= \frac{2}{\pi a} \int_0^{\pi} \frac{l}{2} \sin \phi \, d\phi = \frac{2l}{\pi a} \tag{1.3.11}$$

The geometric probability is rational and acceptable only in some special situation just like the case in classic definition of the probability.

1.3.3 Frequency illustration of probability

What a surprise! The people instinctively would accept following fact: the experiment in Example 1 is repeated. If the number of the occurrence of the head in the first n throw is denoted by $\mu_n(head)$, then when n becomes very large, we have

$$\frac{\mu_n(head)}{n} \approx \frac{1}{2} \tag{1.3.12}$$

We call $\frac{\mu_n(head)}{n}$ as the frequency of the occurrence of the heads in the throw sequence. Comparing (1.3.12) and (1.3.2), the people find that the frequency approaches the probability when n becomes very large. i.e.

$$P(\,head\,|\,\mathcal{C}_1) \approx \frac{\mu_n(head)}{n} \tag{1.3.13}$$

This equation can be verified by the experiment[2]. The records of some history experiments are showed in Table 1.1.

Table 1.1: The history materials of the experiment of tossing a coin.

The experimenters	**The number of throw** n	**The number of occurrence of head** $\mu_n(head)$	**The frequency** $\mu_n(head)/n$
de Buffon	4040	2048	0.5069
de Morgan	4092	2048	0.5005
Jevons	20480	10379	0.5068
Pearson	24000	12012	0.5005
Feller	10000	4979	0.4979
Romanovsky	80640	39699	0.4923

Now we generalize this thought. Given the experiment $\mathcal{E}$, suppose it is defined by condition $\mathcal{C}$. After the experiment $\mathcal{E}$ is performed, the event A can occur and can not occur also. The probability of the event A is denoted by $P(A|\mathcal{C})$. If the experiment $\mathcal{E}$ is repeated under the same condition $\mathcal{C}$, using $\mu_n(A)$ to denote the number of the occurrence of the event A during the n experiments in front, then

$$P(\,A\,|\,\mathcal{C}\,) \approx \frac{\mu_n(A)}{n} \tag{1.3.14}$$

and $\frac{\mu_n(A)}{n}$ is called the frequency of the event A in the experiment se-

[2]Describing the experiment sequence mathematically, the probability theory will prove the correction of the equation (1.3.13)(see the theorem 2.7.7 and the theorem 2.7.8), then the strange fact will have been become the scientific conclude.

quence. R. von Mises idealized this kind of the experiment sequence, and called it as an ideal infinite sequence. The axiom introduced by him is: the frequency in an ideal infinite sequence is convergent, its limited value is called the probability of the event A (Mises' the first axiom)[12]. By our marks it is expressed as follows:

$$P(A \mid C) = \lim_{n\to\infty} \frac{\mu_n(A)}{n} \tag{1.3.15}$$

A. N. Kolmogorov was the supporter of the frequency illustration. After he defined probability as a special kind of measures, he pointed out immediately that the frequency illustration is the basic to establish the linkage between abstract probability concept and the real world.

1.3.4 Subjective illustration of probability

This illustration believe that the probability exists in human subjective world, it reflects a degree of the human believing something, and it is a subjective inference for uncertainty of the thing. This subjective inference depends on personal past experience, knowledge, psychology factor, and so on. People call it the subject probability.

J. M. Keynes considers the probability as a proposition's rational belief based on the knowledge of other propositions. This belief is not able to be given some value and is only able to compare each other[13].

de Finetti considers that *" ... the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event"* [8, p 335].

The subjective illustration is adapted widely, convenient to use and habitable for people and special in the areas such as sociology and economics. For example, we introduce the various illustrations about the value 0.6 in "Tomorrow the probability to rain will be 60%".

1) The Meteorological Observatory believes it will be raining with possibility 60%, and it will not be raining with the possibility 40%. That is the subjective illustration of the probability right away.

2) Suppose it is possible that the meteorological condition is able to repeat continuously, and 0.6 truly reflects the value of the possibility to rain. After repeating to forecast n times, it will be probable that tomorrow the rain will appear $0.6n$ times, and tomorrow the rain will not appear $0.4n$ times. This is the frequency illustration of the probability.

3) Because it is impossible to forecast in infinite sequence for the supposition in 2), then another infinite sequence is designed. We arrange the dates which the forecast of the Meteorological Observatory to be "tomorrow the probability to rain will be 0.6" according the time, and an infinite sequence is obtained. If the forecast methods of the Meteorological Observatory has

not large change, we can consider it to be Mises' ideal sequence[3]. Therefore it will be probable that in front n dates there are $0.6n$ dates to appear rain tomorrow and there are $0.4n$ dates to appear no rain tomorrow.

If we accept the illustrations 1) and 3), then we can determine the accuracy of the meteorological report, in the front n dates the number of the dates of raining is the closer $0.6n$, then the accuracy of the meteorological report is the higher, the probability value 0.6 is the more believing.

The classic definition of the probability and the geometric probability are not effective here.

1.3.5 Prima facie of the application

A lot of convenience is given by the four illustrations of the probability. If two or more illustrations are rational, believing, then their estimates of the value of the probability must be coincident, otherwise at least one illustration of them is not rational. **On the one side the coincidence shows that the probability is a conception of science, on the other side the coincidence generates many applications.**

(1) There are the classic definition and frequency illustration of the probability in the experiment $\mathcal{E}_1$ (Example 1). Table 1.1 confirm the coincidence between the two illustrations.

In order to believe that the conception of the probability is scientific, the fancy readers can operate Example 2, Example 3 or other experiments similarly, and confirm the coincidence to be acceptable and rational.

(2) The author verify the coincidence by the following experiment. We pour the lead into the common angle among three planes "point 1", "point 2", "point 3" of a regular dice, and equip with a cylindrical tea pot which has the high to be 9.5cm and the diameter of the bottom plane to be 7.5cm. Put the dice into the pot, and shake the pot strongly (it is called the experiment of throwing the dice poured with lead or Experiment $\mathcal{E}_5$), then open the cover, and observe the number of points of the dice occurrence.

The condition defining the experiment $\mathcal{E}_5$ is denoted by $\mathcal{C}_5$. Obviously, the all possible outcomes are (1.3.8). The value of the probability of the event "point i to appear" expressed by $P(<i>|\mathcal{C}_5)\,(i = 1, 2, 3, 4, 5, 6)$. Differing from Example 3, those six values need not equal. Using the subjective illustration of the probability, we obtain following conclude:

① Due to the location pouring the lead, we certainly believe

$$P(<i>|\mathcal{C}_5) > P(<j>|\mathcal{C}_5) \qquad (i = 4, 5, 6;\ j = 1, 2, 3) \tag{1.3.16}$$

[3]In the fact, we can prove this sequence is a Mises' ideal sequence because it can be considered as a stationary stochastic sequence.

to be set up, and the difference between two sides is significant.

② Because the symmetry is not required on the angle when the lead is poured, we can not infer that the values of probability following

$$P(<1>|\mathcal{C}_5), \quad P(<2>|\mathcal{C}_5), \quad P(<3>|\mathcal{C}_5) \tag{1.3.17}$$

are equal.

③ Having regard to that the three plane "point 4", "point 5", "point 6" are separated in a long distance from the angle poured lead, therefore the homogeneity of pouring lead has almost the same effects on them. Therefore we believe that

$$P(<4>|\mathcal{C}_5) = P(<5>|\mathcal{C}_5) = P(<6>|\mathcal{C}_5) \tag{1.3.18}$$

should be right.

Towards the frequency illustration of the probability, we repeat the experiment $\mathcal{E}_5$ 2000 times under the same condition, and record the every figure to appear (there is the experiment records being neglected here). The number to appear point i in front n times of the experiment sequence is denied by $\nu_n(i)$, some frequencies of the infinite ideal sequence are showed on Table 1.2.

Table 1.2: The records of the experiment of throwing the dice poured with lead

No.	< 1 >		< 2 >		< 3 >	
	$\nu_n(1)$	$\nu_n(1)/n$	$\nu_n(2)$	$\nu_n(2)/n$	$\nu_n(3)$	$\nu_n(3)/n$
500	77	0.154	62	0.124	54	0.108
1000	145	0.145	116	0.116	117	0.117
1500	211	0.141	168	0.112	170	0.113
2000	286	0.143	219	0.110	229	0.114

No.	< 4 >		< 5 >		< 6 >	
	$\nu_n(4)$	$\nu_n(4)/n$	$\nu_n(5)$	$\nu_n(5)/n$	$\nu_n(6)$	$\nu_n(6)/n$
500	93	0.186	100	0.200	114	0.228
1000	209	0.209	204	0.204	209	0.209
1500	316	0.211	312	0.208	323	0.215
2000	431	0.215	410	0.205	425	0.214

Obviously, the frequency illustration supports the concludes ① and ③ of the subjective illustration, and replenishes the concludes ② by $P(<2> |\mathcal{C}_5) = P(<3>|\mathcal{C}_5)$.

(3) Repeat the experiment $\mathcal{E}_4$ in Example 4 under the same condition continuously. The number of the intersections between the needle and

the parallel lines in front n times throw is denoted by ν_n, the frequency of the event "the needle intersects with the parallel lines" is $\frac{\nu_n}{n}$. If we believe that the geometric probability and frequency illustrations both are rational and acceptable, then the $\frac{2l}{\pi a} \approx \frac{\nu_n}{n}$ is obtained by coincidence of two illustrations. So we have

$$\pi \approx \frac{2nl}{a\nu_n} \tag{1.3.19}$$

Therefore people obtain a method calculating π using the experiment data[4]. Conversely, the inference of the rationality and the belief about the two illustrations depends on whether the calculating value coincides with the true value. Table 1.3 is a history record of the experiment, it verifies again that the probability is a scientific conception.

Table 1.3: The history data of the experiment of Buffon's throwing the needle (a is converted to 1)

Experimenters	The year	Needle's length l	Times of the throw n	Number of intersection ν_n	Experiment value of π
Wolf	1850	0.8	5000	3532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
de Morgan	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

Note: Mr. Lazzerini obtained the excess exact value of π under the condition of throwing times to be not too large, statistical test showed his result not to be too reliable [14].

(4) Mendel's the experiment of the pea hybridization.

The flowers of the peas have two kinds of the color: the red flowers and white flowers. Mr. Mendel had done following jobs:

① Abstracting the purebreds of the red flower breeds and white flower breeds. The implication of the purebred is that after pollinating themselves these pea breeds, the later generations (without regard to how many generations) will always keep the same kind of color.

[4]The Monte-Carlo method had be developed after the generalization of this method.

② After "the hybrid of parental generations" of the red flower purebred and the white flower purebred[5], the generated seeds and the grown-up plants are called the first filial generations. It is discovered that the first filial generations all flower in red color.

③ After pollinating themselves of the first filial generations, the generated seeds and the grown-up plants are called the second filial generations. It is discovered that some flower in red color and some flower in white color in the second filial generations.

Mr. Mendel called the red flowers dominant character and called the white flower recessive character (due to ②), and called them a pair of relative characters. He recorded the numbers of the plants of the red flowers and the white flowers and tabulated them as Table 1.4.

Table 1.4: Mendel's experiment of the pea hybridization[6]

N.	Relative character		Num. of PSFG	The number of DC PSFG		The number of RC PSFG		Prop.
	D. C.	R. C.		Num.	%	Num.	%	
1	Red flower	White flower	929	705	75.89	224	24.11	3.15:1
2	Yellow cotyl.	Green cotyl.	8023	6022	75.06	2001	24.94	3.01:1
3	Plump seed	Wrinkle seed	7324	5474	74.74	1850	25.26	2.96:1
4	R.B.P. indiv.	R.B.P. divis.	1181	822	74.68	299	25.32	2.95:1
5	U.B.P. green	U.B.P. yellow	580	428	73.79	152	26.21	2.82:1
6	B.I.F. armpit	B.I.F. top	858	651	75.87	207	24.13	3.14:1
7	High plant	Lower plant	1064	787	73.96	277	26.04	2.84:1

Note: N. = No. Num. = Number Prop. = Proportion
D. C. = Dominant character
R. C. = Recessive character
PSFG = Plants in the second filial generation
DC PSFG = Dominant character plants in the second filial generation
RC PSFG = Recessive character plants in the second filial generation
Yellow cotyl. = Yellow cotyledon

[5]That is that the red flower is the farther and the white flower is mother, or the white flower is the farther and the red flower is mother copulate.

[6]Mr. Mendel had studied seven relative characters for 8 years.

Green cotyl. = Green cotyledon
R.B.P. indiv. = Ripe bean pod, indivisible node
R.B.P. divis. = Ripe bean pod, divisional node
U.B.P., green = Unripe bean pod, green
U.B.P., yellow = Unripe bean pod, yellow
B.I.F. armpit = Birth in flower, armpit
B.I.F. top = Birth in flower, top

"The proportion 3:1" appears in the last column at Table 1.4. It is called "*the magic proportion 3:1*" by the geneticist. i.e. The number of plants with the dominant character occupies $\frac{3}{4}$ and the number of plants with the recessive character occupies $\frac{1}{4}$. How does the magic proportion generate? It was written that *"thinking long and hard, Mr. Mendel suddenly saw the light at least. He put forth the hypothesis of separation which illustrates the cause of the proportion 3:1."* [27, p 16].

Let us to search *"the process to think long and hard and to see the light suddenly"*. Just as what the world of the geneticists considered commonly, people considered it was one of the reasons which he won tremendous successes that Mendel was more expert in the science of the mathematics and the physics than the other geneticists in the same time. We might guess that Mr. Mendel had compared the magic proportion and equality (1.3.7), and put forth Table 1.5.

Table 1.5: The table to compare the experiment of the pea hybridization and the experiment of tossing two coins

Relative character		The number of D. C.		The number of R. C.		Prop.
D. C.	R. C.	Num.	%	Num.	%	
One head at least appears (H,H) (H,T) (T,H)	Two tails appear (T,T)	3	75%	1	25%	3:1
Red flower	White flower	705	75.89%	224	24.11%	3.15:1
Yellow cotyledon	Green cotyledon	6022	75.06%	2001	24.94%	3.01:1

Note: D. C. = Dominant character
R. C. = Recessive character
H = Head
T = Tail
Num. = Number
Prop. = Proportion

Comparing the last column at Table 1.5, the question would generates naturally: Is it accidental or similitude between the mechanism of inheritance of the relative character and the mechanism of the experiment to throw two coins, so that proportions of the '3:1', ' 3.15:1 ', '3.01:1', etc. have appeared? The flash of light of Mr. Mendel's thinking located at where he believe firmly both are similar, and basing the experiment to throw two coins he described the mechanism of the inheritance of the relative character.

Therefore Table 1.5 inspired Mr. Mendel to approach the great discovery and to put forth **the hypothesis of the gene: There exist large quantities of "pair of genes" in the species, each of "pairs of genes" determined a relative character of the species. A pair of genes consists of two genes. The genes divide into two kinds: dominant and recessive. The dominant character obtains from the dominant genes, the recessive character obtains from the recessive genes, if the pair of genes consists of dominant gene and recessive gene, then the dominant character would appear.** In Table 1.5, the " H " (i.e. head) of the coin corresponds to the dominant gene, the " T " (i.e. tail) corresponds to recessive gene.

How do the species obtain the inheritance from the parents? **The first law of Mendel's genetics states: "The pairs of genes" determining the relative characters of parents divides into two genes according to the original state during copulation of the species, the child (correctly to say, the zygote) would take separately one gene from the male parent and female parent, and constitutes itself the pair of genes. The child's inheritances are determined by the "pairs of genes" that he or she obtains.**

The hypothesis of the gene and the first law of the genetics have satisfactorily illustrated Mr. Mendel's jobs ①–③. In the experiment the gene pair of the red purebred flower is CC, the gene pair of the white purebred flower is cc, here C, c represent separately the dominant and recessive gene. The inheriting process of the first filial generations and the second filial generations are showed in Figure 1.2.

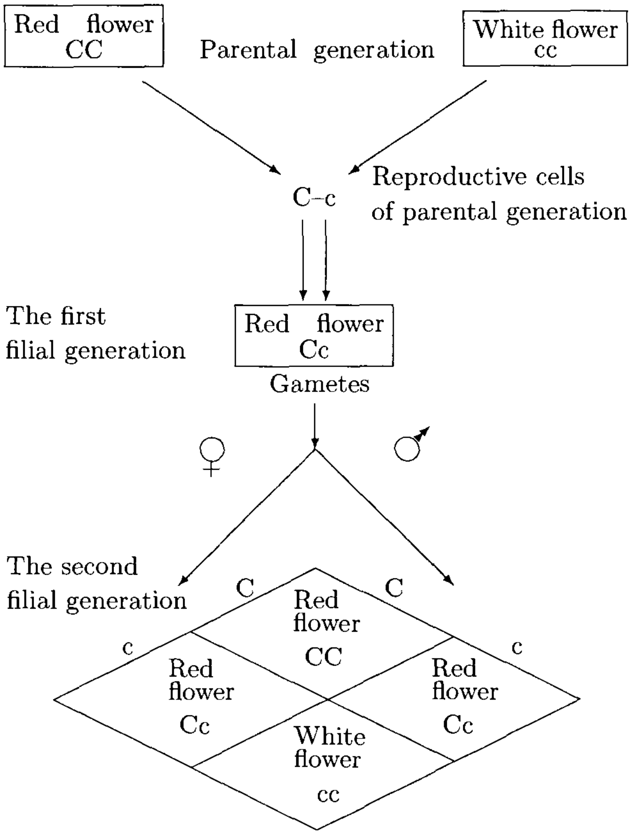

Fig. 1. 2 Mendel's hypothesis of gene and the first law of genetics

The other two among the three laws in genetics correspond firmly with the probability theory also. In order to not deviate the major subject, this book would not discuss them.

This page is intentionally left blank

Chapter 2

Natural Axiom System of Probability Theory

2.1 Element of probability theory and six groups of axioms

In the Natural Axiom System of Probability Theory the random event (called the event, for brief) is the only undefined basic concept[1]. It is only able to be described just as Section 1.1. The mark of the event to be described clearly, or "to be defined" is that people can determine the event to occur, or not to occur, and it must be one or the other.

In habit, the event is denoted by capital letters $A, B, C, \cdots$. The letter A not only denotes the event, but also implies that the event A occurs or the event A happens. The Natural Axiom System consists of six groups of axioms. They are

- Axiom $A_1 - A_3$: The group of axioms of the event space;
- Axiom $B_1 - B_3$: The group of axioms of the causal space;
- Axiom $C_1 - C_2$: The group of axioms of random test;
- Axiom $D_1 - D_3$: The group of axioms of probability measure;
- Axiom $E_1 - E_2$: The group of axioms of conditional probability measure;
- Axiom $F_1 - F_6$: The group of axioms of probability modelling.

[1]A concept is defined by some concepts which are defined by other concepts. Finally there always is a group of concepts which are not able to be defined and are called the meta-word.

The front three groups of axioms reflect the intuitive relations among the events. They abstract the random universe as the causation space which is figurative and calculable mathematically (see Table 0.1 in Preliminary). The fourth group and the fifth group of axioms discuss the interior relations between the condition and probability. They quantify the possibility of the events to occur, abstract the concepts of probability and conditional probability, unveil the essences of conditional probability and independence. This five groups of axioms constitute an axiom system of formalization of probability theory. On its basis the theoretical system of probability theory would be deduced by logical inference and mathematical calculation.

In the theoretical system of probability theory, the probability (i.e. probability measure) is some values with some properties (just as the distances in geometries and the quality in mechanics). When the theoretical system is established, we must suppose these values to exist and to be given, but need not know what are the concrete values and how to be obtain them. The general theory can be used to many applications, just as geometric theorems have become the basic for physics theories and engineering applications.

If the application requires to determine the concrete values of probability, there exist two kind of methods for use:

(1) relating the probability theory and the other theories;

(2) statistical treating the observational data[2].

The statistical methods often are smart. They are building on the probability profound theory. Therefore, only if the theory develops farther, more determining methods for the concrete values of probability would appear, the wider applications of probability theory could be seen.

In the simplified and idealistic situation, the principle of equally likely and the frequency illustration have stood a long-tested, they can determine a part of the probability values. We abstract them as the axioms in the sixth group and establish several classes of common, typical probability spaces which are very favorable for the application and deducing theoretical system.

---

[2]The joint usage of two methods deduce that the statistical methods have become an important scientific method to discover, to establish and to verify the other theories.

2.2 The first group of axioms: the group of axioms of event space

2.2.1 Event space

The probability theory does not touch the concrete contents of the events. Discussing an isolated event, the event would become an object which can not be understood. The event can only be recognized by the relations among it and other events (The Principle I of Probability Theory). Therefore what the probability theory wants first to do is that all events are put together and to find the most basic and the most essential relations among those events. This is the task what the first group of axioms wants to complete.

Definition 2.2.1. *The set which consists of all events is denoted by* $\mathcal{U}$. *i.e.*

$$\mathcal{U} = \{ A \mid A \text{ is an event } \} \tag{2.2.1}$$

If $\mathcal{U}$ *is endowed with the axioms* $A_1 - A_3$*, then* $\mathcal{U}$ *is called the event space.*

Axiom A_1 (Axiom of extensionality). *For any two events A and B, if the event A occurs then the event B must occur; and conversely, if B occurs then A must occur, then A and B are identical event.*

The events A and B in this axiom are denoted as $A = B$, reading as A equals to B.

Axiom A_2 (Axiom of opposite event). *For any event A, " the nonoccurrence of A " is an event also.*

The event "the nonoccurrence of A" is denoted by $\overline{A}$, is called the opposite event of A, or complement event, and is called the complement in brief.

In this book the letter I always represents the index set $\{1, 2, \cdots, n\}$ or $\{1, 2, 3, \cdots\}$. The former is a finite set, the latter is a countable set.

Axiom A_3 (Axiom of σ-union event). *For any finite or countable infinite events A_i, $i \in I$, "at least one event occurs, in those A_i, $i \in I$ " is an event also.*

The event in this axiom is denoted by $\bigcup_{i\in I} A_i$, is called the union event of A_i, $i \in I$, and is call the union in brief. Particularly, if $I = \{1, 2, 3, \cdots\}$, the $\bigcup_{i\in I} A_i$ is called σ-union, also written $\bigcup_{i=1}^{\infty} A_i$, or $A_1 \bigcup A_2 \bigcup \cdots$; if $I = \{1, 2, \cdots, n\}$, the $\bigcup_{i\in I} A_i$ is called finite union, also written as $\bigcup_{i=1}^{n} A_i$, or $A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n$.

Suppose A, B, $A_i \in \mathcal{U}\,(i \in I)$, by Axioms A_2 and A_3 we obtain that $\overline{A}_i$, $\bigcup_{i\in I} \overline{A}_i$, $\overline{\bigcup_{i\in I} \overline{A}_i}$ and $\overline{\overline{A} \bigcup B}$ all are events. In order to depict in detail the identical relation, and understand deeply the complements and unions, we would introduce a relation and two methods to generate new events.

Definition 2.2.2. *Suppose A and B are events and if A occurs then B must occur, then A is called a sub-event of B, or A is contained in B, or B contains A, written as*

$$A \subset B \quad or \quad B \supset A \tag{2.2.2}$$

Definition 2.2.3. *The event $\overline{\bigcup_{i\in I} \overline{A_i}}$ is called the intersection event of all A_i, $i \in I$, and is called the intersection in brief, written as $\bigcap_{i\in I} A_i$.*

In similitude with union, if $I = \{1, 2, , 3, \cdots\}$, the $\bigcap_{i\in I} A_i$ is called σ-intersection, also written $\bigcap_{i=1}^{\infty} A_i$, or $A_1 \bigcap A_2 \bigcap \cdots$, or simply $A_1A_2\cdots$; if $I = \{1, 2, \cdots, n\}$, the $\bigcap_{i\in I} A_i$ is called finite intersection, also written as $\bigcap_{i=1}^{n} A_i$, or $A_1 \bigcap A_2 \bigcap \cdots \bigcap A_n$, or simply $A_1A_2\cdots A_n$.

Definition 2.2.4. *The event $\overline{\overline{A} \bigcup B}$ is called the difference event between A and B, or is called the difference in brief, written as $A \setminus B$.*

Lemma 2.2.1. *Suppose $A, B \in \mathcal{U}$, then $A = B$ if and only if $A \subset B$ and $B \supset A$.*

Proof : The conclude is a direct inference of Axiom A_1 and Definition 2.2.2. ■

Lemma 2.2.2. *For any $A \in \mathcal{U}$, $\overline{\overline{A}} = A$ is set up.*

Proof : The attribute of event is that it occurs or it does not occur, and must be one or the other, so that the event "$\overline{A}$ does not occur" is the event " A occurs ". Using the symbol of opposite event and Axiom A_1, the equality $\overline{\overline{A}} = A$ is set up. ■

Lemma 2.2.3. *The intersection $\bigcap_{i\in I} A_i$ represents the event "all events A_i, $i \in I$ do occur together".*

Proof : " If and only if " is denoted by $\Longleftrightarrow$. Using Definition 2.2.3, we have

$$\bigcap_{i\in I} A_i \ occurs \quad \overset{Axiom\ A_2}{\Longleftrightarrow} \quad \bigcup_{i\in I} \overline{A_i} \text{ does not occur}$$

$$\overset{Axiom\ A_3}{\Longleftrightarrow} \quad \text{All events } \overline{A_i},\ i \in I \text{ do not occur}$$

$$\overset{Lemma 2.2.2}{\Longleftrightarrow} \quad \text{All events } A_i,\ i \in I \text{ occur together} \qquad ■$$

Lemma 2.2.4. *The difference $A \setminus B$ represents the event "A occurs and B does not occur".*

Proof : By Definition 2.2.4, we have

$$A \setminus B \quad occurs. \quad \overset{Axiom\ A_2}{\Longleftrightarrow} \quad \overline{A} \bigcup B \quad \text{does not occur.}$$

$$\overset{Axiom\ A_3}{\Longleftrightarrow} \quad \text{neither } \overline{A} \text{ nor } B \text{ occurs.}$$

$$\overset{Lemma 2.2.2}{\Longleftrightarrow} \quad A \text{ occurs and } B \text{ does not occur.} \qquad \blacksquare$$

Therefore Axioms $A_1 - A_3$, Definitions 2.2.2 – 2.2.4 definite six relations " $=$, $\subset$, $^-$, $\bigcup$, $\bigcap$, $\setminus$ " in the event space $\mathcal{U}$. It is convenient to represent the other relations among the events by six those relations. For example,

- The $\overline{A}\bigcap\overline{B}$ represents the event " both of A and B do not occur".
- The $A\bigcup B \subset C$ represents $A\bigcup B$ being the sub-event of C. On other words, if at least one of both A and B occurs then the event C must occur.
- The $(A\bigcap B)\setminus C = D$ represents " both A and B occur, but C does not occur " to be the event D.
- The $\bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i$ represents the event " there is a positive integer n such that all events A_i, $i \geq n$ occur together ". In fact, by Axiom A_3 we deduce that there is some positive integer n such that $\bigcap_{i=n}^{\infty} A_i$ occurs. Furthermore the conclude is obtained by using Lemma 2.2.3.
- By same reason, $\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i$ represents the event " at least there are infinite events in all A_i, $i \geq 1$ to occur ".

Please readers give more examples, because **knowing a lot of this kind of events would understand the event space deeply, and it is a key to grasp the probability theory.**

2.2.2 Sure event and impossible event

Definition 2.2.5. (1) *The X is called a sure event if X is an event and for any $A \in \mathcal{U}$, $A \subset X$ is true.*

(2) *The $\emptyset$ is called an impossible event if $\emptyset$ is an event and for any $A \in \mathcal{U}$, $\emptyset \subset A$ is true.*

Lemma 2.2.5. *Following four concludes are equivalent:*

(1) $A \subset B$; (2) $\overline{B} \subset \overline{A}$; (3) $A\bigcup B = B$; (4) $A\bigcap B = A$

Proof : "If A occurs then B must occur" identifies with "if B does not occur then A must do not occur". The latter is able to state as "if $\overline{B}$ occurs then $\overline{A}$ must occur", therefore (1) is identical with (2).

Suppose (3) is true. Axiom A_3 implies " if A occurs then $A \bigcup B$ must occur ". Now, if A occurs then $A \bigcup B = B$ must occur. Therefor the (1) is true.

Conversely, suppose (1) is true. If $A \bigcup B$ occurs, then at lest one of both A and B must occur. Therefore B must occur by (1), and $A \bigcup B \subset B$ is true. On the other hand, Axiom A_3 implies " if B occurs then $A \bigcup B$ must occur ", i.e. $A \bigcup B \supset B$. Combining two containing equations, we obtain the (3). It is proved that (1) is identical with (3).

Similarly, it can be proved that (1) is identical with (4). ■

Theorem 2.2.1. *There exist unique sure event and unique impossible event in the event space.*

Proof : Since $\mathcal{U}$ is not empty, taking any $B \in \mathcal{U}$, then there exists $\overline{B} \in \mathcal{U}$ by Axiom A_2. Let

$$X = B \bigcup \overline{B} \tag{2.2.3}$$

we have $X \in \mathcal{U}$ by Axiom A_3.

For any $A \in \mathcal{U}$, since there exists one and only one of both B and $\overline{B}$ to occur, so if A occurs then the event $B \bigcup \overline{B}$ must occur, therefore $A \subset X$. That has proved X to be a sure event. Let

$$\emptyset = \overline{X} \tag{2.2.4}$$

the $\emptyset \in \mathcal{U}$ is obtained by Axiom A_2. For any event A, we have $\overline{A} \subset X$. The $\overline{X} \subset \overline{\overline{A}} = A$ follows from Lemma 2.2.5 and Lemma 2,2,2, so that $\emptyset \subset A$ is true. We have proved that $\emptyset$ is an impossible event.

The uniqueness would be proved as follows. Let X_1 is another sure event, then $X_1 \subset X$ and $X \subset X_1$, and therefore $X_1 = X$ is obtained by Lemma 2.2.1. The uniqueness of the sure event has proved. The uniqueness of the impossible event is able to be proved similarly. ■

The intuitive background of Definition2.2.5: by (2.2.3) we obtain the event X which must occur in the event space; by (2.2.4) we obtain the event $\emptyset$ which must do not occur.

2.2.3 Rules of operations

There are the relations among the four methods (complement, union, intersection, difference) which generate new event in the event space. This leads up the four methods to become the four operations and the relations among them to become the rules of the operations.

Theorem 2.2.2 (The σ-commutative-associative law of the union and intersection). *If the index set I has subsets I_1 and I_2 (they can in-*

tersect), and $I_1 \bigcup I_2 = I$ [3] *, then*

$$\bigcup_{i\in I} A_i = [\bigcup_{j\in I_1} A_j] \bigcup [\bigcup_{k\in I_2} A_k] \qquad (2.2.5)$$

$$\bigcap_{i\in I} A_i = [\bigcap_{j\in I_1} A_j] \bigcap [\bigcap_{k\in I_2} A_k] \qquad (2.2.6)$$

Proof : Suppose the event $\bigcup_{i\in I} A_i$ occur. By Axiom A_3 we know that in all $A_i, i \in I$, at least one event occurs. We may suppose A_r occurs. The $r \in I$ implies that at least one of both $r \in I_1$ and $r \in I_2$ is set up, so at least one of both events $\bigcup_{j\in I_1} A_j$ and $\bigcup_{k\in I_2} A_k$ occurs. It follows from Definition 2.2.1 that

$$\bigcup_{i\in I} A_i \subset [\bigcup_{j\in I_1} A_j] \bigcup [\bigcup_{k\in I_2} A_k]$$

Conversely, suppose the event in the right of (2.2.5) occurs. By Axiom A_3 we know at least one of both events $\bigcup_{j\in I_1} A_j$ and $\bigcup_{k\in I_2} A_k$ occurs. We may suppose the event $\bigcup_{j\in I_1} A_j$ occurs. Secondarily using Axiom A_3, there is at least one $r(r \in I_1)$ such that A_r occurs. Noted that if $r \in I_1$, then $r \in I$, we obtain the $\bigcup_{i\in I} A_i$ occurs. Therefore

$$\bigcup_{i\in I} A_i \supset [\bigcup_{j\in I_1} A_j] \bigcup [\bigcup_{k\in I_2} A_k]$$

is set up. Combining two containing equations above we obtain (2.2.5). The (2.2.6) can be proved similarity. ■

Corollary. *The union and intersection operations have following properties :*

(1) The idempotent law :

$$A \bigcup A = A; \qquad A \bigcap A = A$$

(2) The commutative law :

$$A \bigcup B = B \bigcup A; \qquad A \bigcap B = B \bigcap A$$

(3) The associative law :

$$(A \bigcup B) \bigcup C = A \bigcup (B \bigcup C); \qquad (A \bigcap B) \bigcap C = A \bigcap (B \bigcap C)$$

[3]Note, this $\bigcup$ is the union operation in set theory. It is not accident that the "complement, union, intersection, difference" of the events and "complement, union, intersection, difference" of the sets have the identical names and symbols. See Section 2.3.

Theorem 2.2.3 (The σ-distributive law of the union and intersection). *If B, $A_i \in \mathcal{U}$ $(i \in I)$, then we have*

$$[\bigcup_{i\in I} A_i]\bigcap B = \bigcup_{i\in I}[A_i \bigcap B] \tag{2.2.7}$$

$$[\bigcap_{i\in I} A_i]\bigcup B = \bigcap_{i\in I}[A_i \bigcup B] \tag{2.2.8}$$

Proof : Suppose the event of the left of (2.2.7) occurs. By Lemma 2.2.3 we deduce that both of $\bigcup_{i\in I} A_i$ and B occur together. By Axiom A_3 we obtain that there is at least one $j \in I$ such that both of A_j and B occur together, that is the $A_j \bigcap B$ to occur. Therefore the event of the right of (2.2.7) occurs. So we have

$$[\bigcup_{i\in I} A_i]\bigcap B \subset \bigcup_{i\in I}[A_i \bigcap B]$$

Conversely, suppose the event of the right of (2.2.7) occurs, then there is at lest one $j \in I$ such that $A_j \bigcap B$ occurs (Axiom A_3), so that A_j and B occur together. Therefore both of $\bigcup_{i\in I} A_i$ and B occur together, we have

$$\bigcup_{i\in I}[A_i \bigcap B] \subset [\bigcup_{i\in I} A_i]\bigcap B$$

Combining both of the containing equation, we obtain (2.2.7). The (2.2.8) can be proved similarly. ■

Theorem 2.2.4. *The operations (complement, union, intersection, difference) have the properties: (1)*

$$A \setminus B = A \bigcap \overline{B} \tag{2.2.9}$$

(2) de Morgan's law:

$$\overline{\bigcup_{I\in I} A_i} = \bigcap_{I\in I} \overline{A}_i \tag{2.2.10}$$

$$\overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A}_i \tag{2.2.11}$$

Proof : (1) By Lemma 2.2.4 we know that $A\setminus B$ represents " A to occur and B not to occur ". The latter equals to "A and $\overline{B}$ occur together" by Axiom A_2. The (2.2.9) is obtained by using Lemma 2.2.3.

(2) By Definition 2.2.3, we have

$$\bigcap_{i\in I} A_i = \overline{\bigcup_{i\in I} \overline{A_i}} \tag{2.2.12}$$

Using $\overline{A_i}$ instead of A_i and Lemma 2.2.2, we obtain (2.2.10). Taking the opposite events for both side of (2.2.12) and using Lemma 2.2.2, we obtain (2.2.11). ■

2.2.4 Event fields and event σ-fields

Definition 2.2.6. *Suppose $\mathcal{F}$ is a set which is nonempty and consists of a part of events. $\mathcal{F}$ is called an event field if it satisfies the conditions (a) and (b):*

(a) It is closed under the formation of complements: if $A \in \mathcal{F}$, then $\overline{A} \in \mathcal{F}$;

(b) It is closed under the formation of finite unions: if A, $B \in \mathcal{F}$, then $A \bigcup B \in \mathcal{F}$.

Definition 2.2.7. *Suppose $\mathcal{F}$ is a set which is nonempty and consists of a part of events. $\mathcal{F}$ is called an event σ-field, if it satisfies the condition (a) above and the condition (c):*

(c) It is closed under the formation of σ-unions: if $A_i \in \mathcal{F}$, $i \in I$, then $\bigcup_{i\in I} A_i \in \mathcal{F}$.

Obviously, the event space $\mathcal{U}$ itself is an event field, and is an event σ-field, too.

Theorem 2.2.5. *(1) The event field contains the sure event X and the impossible event $\emptyset$, and it is closed under the formation of finite unions, finite intersections and differences.*

(2) The event σ-field is an event field, and it is closed under the formation of σ-unions, σ-intersections and differences.

Proof : (1) Since an event field is nonempty, so there exists an event $B \in \mathcal{F}$. It is from the conditions (a) and (b) that the sure event $X = B \bigcup \overline{B} \in \mathcal{F}$, therefore $\emptyset = \overline{X} \in \mathcal{F}$.

Suppose $A_i \in \mathcal{F}$ $(i = 1, 2, \cdots, n)$. We have $\bigcup_{i=1}^{3} A_i = [A_1 \bigcup A_2] \bigcup A_3 \in \mathcal{F}$. It is able to be proved by the mathematical induction that $\mathcal{F}$ is closed under the formation of finite unions.

Taking $I = \{1, 2, \cdots, n\}$ in the (2.2.12) and using the closeness under finite unions and the condition (a), we obtain that $\mathcal{F}$ is closed under the formation of finite intersections.

It follows from (2.2.9) and the closeness under finite intersections that $\mathcal{F}$ is closed under the formation of differences. That completes the proof of (1).

(2) Taking $I = \{1,2\}$, the condition (c) becomes the condition (b), so the event σ-field is an event field.

By (2.2.12), using the conditions (a) and (c), we obtain that $\mathcal{F}$ is closed under the formation of σ-intersections. ■

Definition 2.2.8. *Suppose $\mathcal{F}_1$ and $\mathcal{F}_2$ are two event σ-field (or the event field). If $\mathcal{F}_2 \subset \mathcal{F}_1$ is set up, then $\mathcal{F}_2$ is called an event σ-subfield of $\mathcal{F}_1$ (or an event subfield).*

Obviously, any event σ-field (or an event field) always is an event σ-subfield (or event subfield) of the event space $\mathcal{U}$; conversely, $\mathcal{F}_0 = \{\emptyset, X\}$ is an event σ-subfield (or event subfield) of any event σ-field (or event field). $\mathcal{F}_0$ may be called the minimum event σ-field (or the minimum event field).

2.2.5 Remark

(1) Using the sure event and the impossible event, the finite union and finite intersection are the particular cases of the countable union and countable intersection, respectively. The realization methods are

$$\bigcup_{i=1}^{n} A_i = A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n \bigcup \emptyset \bigcup \emptyset \bigcup \cdots \tag{2.2.13}$$

$$\bigcap_{i=1}^{n} A_i = A_1 A_2 \cdots A_n X X \cdots \tag{2.2.14}$$

(2) Suppose A and B are events. If $AB = \emptyset$, then A and B are called exclusive. It is very useful conclude that the union of events can be transformed to the union of exclusive events. In fact, suppose $\bigcup_{i \in I} A_i$ is any union event, let

$$B_1 = A_1,\ B_2 = A_2 \setminus B_1,\ \cdots,\ B_n = A_n \setminus (\bigcup_{i=1}^{n-1} B_i),\ \cdots \tag{2.2.15}$$

It is easy proved that all B_i $(i \in I)$ are exclusive (i.e. $B_i B_j = \emptyset$ if $i \neq j$) and

$$\bigcup_{i \in I} A_i = \bigcup_{i \in I} B_i \tag{2.2.16}$$

(3) The other operations can be introduced in the event space $\mathcal{U}$. For instance, the operational symbol $\triangle$ can be used to define

$$A \triangle B \overset{def}{=} (A \setminus B) \bigcup (B \setminus A) \tag{2.2.17}$$

which is called the symmetric difference of A and B.

(4) Suppose J is a certain index set, it may be finite, countable or uncountable. In this book, the index set I always belong to front two formers. The uncountable index sets have:

the set of real numbers $\mathbf{R}$;

the interval $[a, b]$;

the rectangle $\{ (x, y) \mid a \leq x \leq b;\ c \leq y \leq d \}$, and so on,

where a, b, c, d are real numbers and $a < b$, $c < d$.

Suppose $A_\alpha \in \mathcal{U}$ $(\alpha \in J)$. In probability theory people occasionally use the sentence " At least one of the events A_α, $\alpha \in J$ occurs. " and " All events A_α, $\alpha \in J$ occur. ", and written respectively as

$$\bigcup_{\alpha \in J} A_\alpha ; \quad \bigcap_{\alpha \in J} A_\alpha \tag{2.2.18}$$

It is noted that the two symbols represent the events, only if J is finite or countable. If J is uncountable set, they are not the operation in $\mathcal{U}$, are not guaranteed to be the events in $\mathcal{U}$. They are only two symbols.

(5) The countable infinite and finite locate in the same places in the KAS and the NAS, only uncountable infinite is " true infinite ". For emphasizing this, the term "σ" in the language usually does not omit.

2.3 The second group of axioms: the group of axioms of causal space

2.3.1 The questions and the methods

The event space $\mathcal{U}$ is a very enormous set of all random events. It is abstract, the reason is that we can not list its all elements, we even do not know what is the power of $\mathcal{U}$[4]. And it is concrete, too, the reason is that we can list many events. The question is: how do we clearly and visually recognize the event space $\mathcal{U}$?

In the science, people have encountered with similar question many times. For instance:

(1) In the world there are various materials. May we ask how do we recognize all materials?

Passing a lot of researches of many people, Demokritos declared that the origin of all things on earth is a thing called atom. The various in locations and orders arrangements of the atoms with various sizes and various shapes constitute the all things on the earth.

[4]The power of the set reflects how many its elements are. In special, the power of finite set equals to the number of its elements.

In order to recognize all materials, people have found atoms with various sizes and various shapes — more than hundred chemical elements. After that, people study how do the elements constitute all things on the earth — the various mixtures and the chemical compounds [5].

(2) The human being, the animals and the plants reproduce the latter generations. There are the same characters and the different characters among the individuals of the latter generations. The people may ask: how do we recognize the various inheritance characters?

G. Mendel declared that there exist large quantities of "pair of genes" in the species. The pair of genes consists of two genes. A pair of genes represents a character of inheritance. A child (correctly to say, the zygote) receives one gene from father and mother respectively and forms the pair of genes itself. The pairs of genes received by the latter generation determine its inheritance characters.

(3) There are various geometrical figures in the real world. May we ask how do we study the figures?

The geometry declare that people can assume the "minimum" geometrical figures to be a point. All assumed points have formed the Euclidean space. The curve and curved surface are some sets consisted of special points. People can besiege various figures by the curves and curved surfaces.

(4) There are various random events in real world. May we ask how do we study the random events?

Under the inspiration of the thought above the Natural Axiom System considers that people can assume the "minimum" random event to be a causal point. All assuming causal points form a causal space. A random event is a certain set consisted of special causal points. Using random events, the various "graphics"— the random tests and the probability spaces can be generated in the causal space.

The minimum unit — the elements and the genes are the materials in real world, they are observed by scientific methods. The points and the causal points are the products of abstract thinking of human being, they can be understood only by thinking of human being.

2.3.2 Intuitive background of the causal point

One "minimum" random event is assumed as one visionary realization of total events. It is represented by following method. The occurrence and the nonoccurrence are denoted by 1 or 0, respectively. If the event A occurs then A is considered to correspond value 1, otherwise to correspond

[5] All the materials constitute a very enormous set, too. Similarly, we do not know what the power of the set is and how many chemical elements there are in the world.

value 0. Therefore a visionary realization of total events is a function which is defined on the event space $\mathcal{U}$ and taking 1 or 0, i.e.

$$u(A) = \begin{cases} 1, & A \text{ occurs.} \\ 0, & A \text{ does not occur.} \end{cases} \tag{2.3.1}$$

and $u(A)$ is called a causal point, written as u for short[6].

A causal point represents a state of the event space $\mathcal{U}$: All events B with $u(B) = 1$ occur; All C with $u(C) = 0$ do not occur. If the event space is located in the state that u represents, then the state u is called pseudo-occurrence[7].

The reason that we call u as a causal point is as follows. Using inversion of thinking, we can say that since the causal point u pseudo-occurs, so the event B with $u(B) = 1$ occurs and the event C with $u(C) = 0$ does not occur.

Note that a function defined on $\mathcal{U}$ taking values 1 or 0 need not be a causal point. For example, given an event D, the functions $u(A)$, $A \in \mathcal{U}$ with $u(D) = u(\overline{D}) = 1$ are not causal points. This reason is that there is not the state with D and $\overline{D}$ to occur together by Axiom A_2. In fact, by the first group of axioms we easily infer the causal points have following four properties:

- ① $u(X) = 1; \quad u(\emptyset) = 0.$
- ② $u(\overline{A}) = 1 - u(A), \quad$ for any $A \in \mathcal{U}$.
- ③ $u(\bigcup_{i\in I} A_i) = \sup_{i\in I}\{\ u(A_i)\ \}, \quad$ for any $A_i \in \mathcal{U}$ $(i \in I)$.
- ④ $u(\bigcap_{i\in I} A_i) = \inf_{i\in I}\{\ u(A_i)\ \}, \quad$ for any $A_i \in \mathcal{U}$ $(i \in I)$.

[6] The formal definition of the causal point will be given in Subsection 2.3.3. By the opportunity we declare that any concepts and any symbols of the NAS are not able to define in the chapter, the section and the subsection with intuitive background. That chapter is Chapter 1; that section is Section 3.1; those subsections are Subsection 2.3.2, 2.4.1, 2.7.1 and 2.9.1. Chapter 1 is the intuitive background of the first, the third, the fourth and the sixth groups of axioms.

In those chapter, section and subsections, we have abstracted the essence from the random phenomena, and formed the prototype of the concepts, and guided the NAS to define those concepts and to introduce the axioms. Therefore they are not absent for building the NAS.

Because those contents are the linkages between abstract axiom system and the real world. So they are necessary knowledge to understand the NAS and to do application of probability theory.

[7] A causal point u is not an event. In order to determine whether the event space is located in the state u, we need know what events have occur and what events have not occur. Obviously, this is impossible. That is to say, we can not determine "phenomenon u" occurs or not, so u is not an event.

and the properties ①, ④ can be deducted from ②, ③[8].

Now fixing an event B, let us define a set which consists of some causal points:

$$\{ u \mid u \text{ is causal point and } u(B) = 1\}$$

Intuitive, the set represents a group of causes. We call this group of causes appearance if some causal point in the set pseudo-occurs. Therefore if this group of causes appears then the event B must occur; if this group of causes does not appear then the event B must do not occur; and under obeying the properties ②and ③ the other events can freely select occurrence or nonoccurrence. So **we can identify this group of causes with it's effect — the event B. That is**

$$B = \{ u \mid u \text{ is a causal point and } u(B) = 1 \}$$

and **the event B** occurs if and only if a causal point u in **the set B** pseudo-occurs.

Sofar, we have got an important conclude: **an event is a set consisted of some causal points.**

In particular, if $B = \emptyset$, then the right side of above equality is empty, therefore the impossible event is identified with the empty set. If $B = X$, then above equality becomes

$$X = \{ u \mid u \text{ is a causal point } \}$$

It is a set consisted of all causal points. So the sure event is identified with the causal space.

Brief summary: the sequence of the concepts introduced by us is: the event $\Longrightarrow$ the event space $\Longrightarrow$ the causal point $\Longrightarrow$ an event is a set which consists of causal points $\Longrightarrow$ the causal space is the sure event.

It is very importance that **an event becomes a set in the causal space**. We will prove that the six relations — equality, contain, complement, union, intersection, difference of events in the event space identify with equality, contain, complement, union, intersection, difference of sets in the causal space, respectively, and the pair $(X, \mathcal{U})$ becomes a measurable space. Therefore **the probability theory has found the most effective research tool — the methods of set theory.**

J. Venn's graphics give the intuitive appearance of the six relations in set theory. Therefore they give the intuitive appearance of the six relations

[8] After deleting the probability meaning of the symbols and the operations, $\mathcal{U}$ is a Boolean σ-algebra in the algebra, the set $\{0, 1\}$ given adaptable operations is an algebra. The function $u(A)$, $A \in \mathcal{U}$ satisfying the properties ②, ③ is called a homomorphism[15][16].

in probability theory (Figure 2.1).

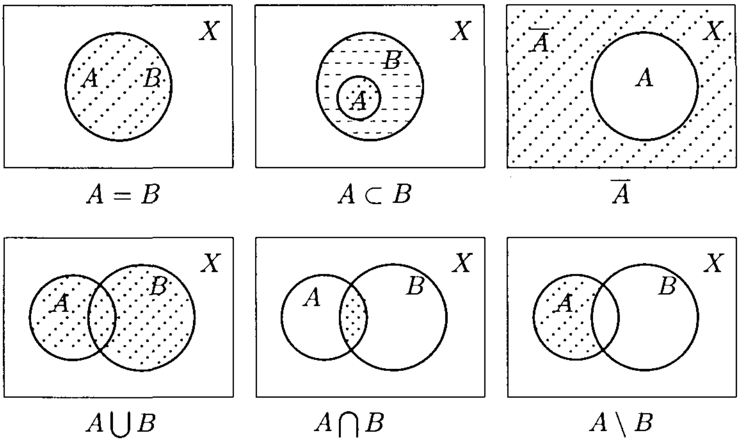

Fig. 2.1 The six relations of events in the causal space X

2.3.3 The group of axioms of causal space

Definition 2.3.1. Suppose $u(A)$, $A \in \mathcal{U}$ (written as u) is a function defined on the event space $\mathcal{U}$ and taking values 1 or 0. The u is called a causal point if it has the properties ①–④ in Subsection 2.3.2.

Axiom B_1 (Axiom of Causal point). *There exist causal points. A causal point u represents a state of the event space $\mathcal{U}$: all events B with $u(B) = 1$ occur; all events C with $u(C) = 0$ do not occur.*

The set which consists of all causal points is denoted by X. i.e.

$$X = \{\, u \mid u \text{ is a causal point } \} \tag{2.3.2}$$

and X is called the causal space. The causal point u is called pseudo-occurrence if the event space is located in the state u.

Axiom B_2 (Axiom of pseudo-occurrence). *There exists one and only one causal point to pseudo-occur.*

Axiom B_3 (Axiom of event). *The event B is a set in the causal space X. It is*

$$B = \{\, u \mid u \textit{ is a causal point and } u(B) = 1 \,\} \tag{2.3.3}$$

In addition the occurrence of the event B means that some causal point in the set B pseudo-occurs.

Definition 2.3.2. *The pair $(X, \mathcal{U})$ is called the causation space, if $\mathcal{U}$ is the event space and X is the causal space.*

Theorem 2.3.1. *In the causation space $(X, \mathcal{U})$, the events and their operations in $\mathcal{U}$ and the sets and their operations in X take on the identity that Table 2.1 represents.*

Proof : The second group of axioms sets up the correspondence among the basic elements in Table 2.1.

The correspondence among the basic relation will be proved as follows.

(1) Suppose the $B \subset C$ is the containing equation of the events, then for any causal point $u(A)$, $A \in \mathcal{U}$, using $C = B \bigcup C$ and the property ③ of the causal points, we have

$$u(C) = u(B \bigcup C) = \sup\{\, u(B),\, u(C)\,\}.$$

Now, suppose the causal point $u^* \in B$. By Axiom B_3, we know $u^*(B) = 1$. It is from above equality that $u^*(C) = 1$, then $u^* \in C$ by using Axiom B_3 again. So the $B \subset C$ is a containing equation in set theory.

On the contrary, suppose the $B \subset C$ is a containing equation in set theory, and A and B are events. Now, if B occurs, then some causal point u^* in the B pseudo-occurs by Axiom B_3. It is from the equation $B \subset C$ in set theory that $u^* \in C$, then the event C occurs by Axiom B_3 again, therefor the $B \subset C$ is also a containing equation in probability theory. So the identity of containing relation is proved.

(2) Because there is the similarity of Lemma 2.2.1 in set theory, it is from above (1) and Lemma 2.2.1 that the identity of equality relation is true.

(3) Suppose B and $\overline{B}$ are a pair of opposite events. We know B to have expression (2.3.3) and $\overline{B}$ to have expression

$$\overline{B} = \{\, u \,|\, u \text{ is a causal point and } u(\overline{B}) = 1\,\} \tag{2.3.4}$$

by Axiom B_3. Noting that causal points are the functions taking values 1 and 0, and using the property ② of the causal point we get for any $u^* \in X$

$$u^* \in B \Longleftrightarrow u^*(B) = 1 \Longleftrightarrow u^*(\overline{B}) = 0 \Longleftrightarrow u^* \notin \overline{B}.$$

So we have proved $\overline{B} = X \setminus B$ in set theory, i.e. the $\overline{B}$ is the complement of B in set theory.

On the contrary, suppose the B is a set in the causal space X, and also is an event. If the complement of B in set theory is denoted by B^*, and the opposite event of B is noted by $\overline{B}$, from (2.3.3), and the property which the causal point is a function with values 0 ,1, and the definition of complement in set theory we have

$$B^* = \{\, u \,|\, u \text{ is the causal point with } u(B) = 0\,\} \tag{2.3.5}$$

Table 2.1 Probability Theory Vs. Set Theory

The Causation Space $(X, \mathcal{U})$			
	subject	**Set Theory**	**Probability Theory**
Sym.		Causal Space X	Event Space $\mathcal{U}$
Basic Elements	u	Causal point	Absent
	G	Any set	Absent
	A	Special set (A group of causes)	Event
	$\emptyset$	Empty set	Impossible event
	X	Causal space	Sure event
	$\mathcal{U}$	A special σ-field	Event space
	$(X, \mathcal{U})$	A measurable space	Causation space
As follows: all A, B, A_i, A_j are events; I is finite or countable index set; J is uncountable index set.			
Basic Relations	$u \in A$	Causal point u is an element in A.	Pseudo-occurrence of u causes event A to occur.
	$A = B$	A and B consist of the same causal points.	If A occurs then B must occur and If B occurs then A must occur.
	$A \subset B$	Every causal point of A belongs to B.	If A occurs then B must occur.
	$\overline{A}$	The set consists of all causal points which do not belong to A.	The complement event: "A does not occur".
	$\bigcup_{i\in I} A_i$	The set consists of all causal points of all $A_i\,(i \in I)$.	The union event: "at least one event of all $A_i\,(i \in I)$ occurs".
	$\bigcup_{j\in J} A_j$	The set consists of all causal points of all $A_j\,(j \in J)$.	No meaning.
	$\bigcap_{i\in I} A_i$	The set consists of all causal points which belong to every $A_i\,(i \in I)$.	The intersection event: "all events $A_i\,(i \in I)$ occur together".
	$\bigcap_{j\in J} A_j$	The set consists of all causal points which belong to every $A_j\,(j \in J)$.	No meaning.
	$A \setminus B$	The set consists of all causal points which belong to A, but not to B.	The difference event: "A occurs, but B does not occur".
	$A \bigcap B = \emptyset$	The sets A and B are disjoint.	The events A and B are exclusive.

It is from the property ② of causal points that

$$B^* = \{\, u \,|\, u \text{ is the causal point with } u(\overline{B}) = 1 \,\}$$

So by Axiom B_3 we have $B^* = \overline{B}$, i.e. B^* is the opposite event of B. The identity of complement relation is proved.

(4) Suppose $B_i\,(i \in I)$ is an event, so $\bigcup_{i \in I} B_i$ is an event also. By Axiom B_3 we have

$$B_i = \{\, u \,|\, u \text{ is the causal point with } u(B_i) = 1 \,\},\ (i \in I) \quad (2.3.6)$$

$$\bigcup_{i \in I} B_i = \{\, u \,|\, u \text{ is the causal point with } u(\bigcup_{i \in I} B_i) = 1 \,\} \quad (2.3.7)$$

Form the property ③ of causal points we deduct that $u(\bigcup_{i \in I} B_i) = 1$ if and only if at least there is some B_i such that $u(B_i) = 1$. Therefore the $\bigcup_{i \in I} B_i$ is a set consisted of all causal points of B_i, $i \in I$, i.e. the $\bigcup_{i \in I} B_i$ is the union of all B_i, $i \in I$ in set theory.

On the contrary, suppose for any $i\,(i \in I)$, B_i is a set in the X, and they are all events. It is from the latter supposition that the (2.3.6) is true.

Now, the union of all B_i, $i \in I$ in set theory is denoted by B^*; the union in probability theory is denoted by $\bigcup_{i \in I} B_i$. It is from the definition of the union in set theory that

$$B^* = \{\, u \,|\, u \text{ is the causal point, and there is some } B_i\,(i \in I) \text{ such that } u(B_i) = 1 \,\}$$

so using the conclude about the causal points' property after (2.3.7) we deduct

$$B^* = \{\, u \,|\, u \text{ is a causal point such that } u(\bigcup_{i \in I} B_i) = 1 \,\}$$

Therefore, by Axiom B_3 we know that B^* is an event, and $B^* = \bigcup_{i \in I} B_i$. That has proved the identity of union relation.

(5) Since de Morgan's law is right in both of probability theory and set theory, the identity of the intersection relation is proved by using (2.2.12) and proved (3), (4).

(6) Because the $A \setminus B = A \bigcap \overline{B}$ is right in both of probability theory and set theory, the identity of the difference relation is proved by using $A \setminus B = A \bigcap \overline{B}$ and proved (3), (5). ■

2.3.4 Remark

(1) The philosophical illustration of the causation space

The event B has the representative (2.3.3), i.e.

$$B = \{\, u \,|\, u \text{ is a causal point and } u(B) = 1 \,\} \tag{2.3.3}$$

The right is a special set in the causal space X, it represents a group of causes; the left is an element in the event space $\mathcal{U}$, it represents the effect of the group of causes. Therefore both sides constitute a relation of causation, and it is represented intuitively as

$$\text{the effect} = \text{the group of causes} \tag{2.3.8}$$

and the effect appears if and only if some causal point in the group of causes appears (pseudo-occurs). In particular, suppose $B = X$, the above equality becomes

$$\text{the sure event} = \text{all of causes} \tag{2.3.9}$$

Consequently, (2.3.3) is the mathematical expression of the usual language " The cause leads to the effect, the effect searches the causes".

In reality people's the methods of thinking about the law of causation contain two terms of contents:

① The effect is deduced from the causes; or the causes are searched by the effect.

② The correspond changes of the effect are deduced from the changes of the causes.

The mathematical abstract of the front term is the equality (2.3.3). The mathematical abstract of the latter term are the mathematical operations of the (2.3.3), i.e. the changes of the causes are abstracted as the set operation in the right side; the changes of the effects are abstracted as the event operation in the left side. Therefore the NAS abstracts the law of causation as the quantitative philosophical law with mathematica operations.

(2) Table 2.1 indicates that the uncountable union $\bigcup_{j \in J} A_j$ of events and uncountable intersection $\bigcap_{j \in J} A_j$ are operations without meaning in probability theory. But they are the operations with meaning in set theory and generate a new set in the causal space. It must be careful to determine whether the new set is an event. Generally it is not an event, but in special cases (for instance, for any $j \in J$, $A_j = X$), it can be an event.

(3) Axiom B_1 guarantees the existence of causal points. Axiom B_3 guarantees that if $B \neq \emptyset$ then the right side in the (2.3.3) is not empty. These two pledges are not only beliefs but also verifiable mathematical concludes.

In fact, they are the concludes of M. H. Stone's famous representative theorem in the Boole algebra[15][16].

The reason of regarding those two pledge as the axioms is their intuition. We do not want to involve too much the algebraic knowledge, furthermore we do not want to cover the intuition and appearance of probability theory with the algebraic abstract theorems.

(4) The concept of partition will be using at next section. It is introduced here briefly.

Definition 2.3.3. *Let E is a set, J is an index set, $\Pi = \{\pi_\alpha | \alpha \in J\}$ is a class of sets in the E. If the class Π satisfies following three conditions:*

(a) $\pi_\alpha \neq \emptyset$, for any $\alpha \in J$;

(b) $\pi_\alpha \bigcap \pi_\beta = \emptyset$, for any α, $\beta \in J$ and $\alpha \neq \beta$;

(c) $\bigcup_{\alpha \in J} \pi_\alpha = E$,

then Π is called a partition of the E, and π_α is called a partition block. If Π satisfies conditions (b) and (c), then Π is called a broad partition.

Definition 2.3.4. *Let $\Pi_1 = \{\pi^{(1)}_\alpha | \alpha \in J_1\}$, $\Pi_2 = \{\pi^{(2)}_\beta | \beta \in J_2\}$ are two partitions of E. If for any $\pi^{(2)}_\beta$, there is a $\pi^{(1)}_\alpha$ in Π_1 such that*

$$\pi^{(2)}_\beta \subset \pi^{(1)}_\alpha \tag{2.3.10}$$

then Π_2 is finer than Π_1, and Π_1 is coarser than Π_2.

Example 1. Let $A_i \, (i \in I)$ is an event which is not empty, all $A_i, i \in I$ are disjoint (i.e. $A_i A_j = \emptyset$ for $i, j \in I, i \neq j$), and $\bigcup_{i \in I} A_i = X$. Obviously $\{A_i | i \in I\}$ is a partition of the causal space X.

Example 2. Let $A_\alpha (\alpha \in J, J$ is uncountable) is an event, all $A_\alpha, \alpha \in J$ are disjoint (i.e. $A_\alpha A_\beta = \emptyset$ for $\alpha, \beta \in J$, $\alpha \neq \beta$). If all A_α, $\alpha \in J$ satisfy $\bigcup_{\alpha \in J} A_\alpha = X$ in set theory, then $\{A_\alpha | \alpha \in J\}$ is a broad partition of the causal space X. In addition, if $A_\alpha (\alpha \in J)$ is not empty, then $\{A_\alpha | \alpha \in J\}$ is a partition of the causal space X.

Example 3. Let $\mathbf{R} = (-\infty, +\infty)$. Obviously,

$$\Pi_1 = \{\, \mathbf{R} \,\} \tag{2.3.11}$$

$$\Pi_2 = \{\, (n, n+2] \,|\, n \text{ is an odd } \} \tag{2.3.12}$$

$$\Pi_3 = \{\, (n, n+2] \,|\, n \text{ is an even } \} \tag{2.3.13}$$

$$\Pi_4 = \{\, (n, n+1] \,|\, n \text{ is an integer } \} \tag{2.3.14}$$

$$\Pi_5 = \{\, \{x\} \,|\, x \in \mathbf{R} \,\} \tag{2.3.15}$$

all are partition of $\mathbf{R}$. And

(a) Π_1 is the coarsest partition. i.e. There is not a coarser partition than Π_1. In other words, any partition of $\mathbf{R}$ is finer than Π_1.

(b) Π_5 is the finest partition. i.e. There is not a finer partition than Π_5. In other words, any partition of **R** is coarser than Π_5.

(c) Π_4 is finer than Π_2, and is finer than Π_3 also.

(d) Π_2 and Π_3 can not compare with fining or coarseness.

Lemma 2.3.1. *Supposing* $\Pi_1 = \{\pi_\alpha^{(1)} \mid \alpha \in J_1\}$, $\Pi_2 = \{\pi_\beta^{(2)} \mid \beta \in J_2\}$ *are two partitions of* E, *and* Π_2 *is finer than* Π_1, *then for any* $\pi_\alpha^{(1)} \in \Pi_1$, *there exists a subset* J_α *of* J_2 *such that*

$$\pi_\alpha^{(1)} = \bigcup_{\beta \in J_\alpha} \pi_\beta^{(2)} \tag{2.3.16}$$

Lemma 2.3.2. *Supposing* $\Pi_1 = \{\pi_\alpha^{(1)} \mid \alpha \in J_1\}$, $\Pi_2 = \{\pi_\beta^{(2)} \mid \beta \in J_2\}$ *are two partitions of* E, *then*

$$\Pi_1 * \Pi_2 \overset{def}{=} \{\, \pi_\alpha^{(1)} \bigcap \pi_\beta^{(2)} \mid \alpha \in J_1, \beta \in J_2 \,\} \tag{2.3.17}$$

is broad partition of E.

Definition 2.3.5. *The* $\Pi_1 * \Pi_2$ *removed the empty set is called a crossed partition of* Π_1 *and* Π_2*(or a cross for short), written as* $\Pi_1 \odot \Pi_2$. *If* $\Pi_1 * \Pi_2$ *contain no empty set, then* $\Pi_1 \odot \Pi_2 (= \Pi_1 * \Pi_2)$ *is call a true crossed partition.*

Example 4. For the partitions in Example 3, we give four conclusions:

(α) Supposing Π is any partition of **R**, then

$$\Pi_1 \odot \Pi = \Pi_1 * \Pi = \Pi \tag{2.3.18}$$

is a true cross.

(β) Supposing Π is any partition of **R**, but $\Pi \neq \Pi_1$, then $\Pi_5 * \Pi$ is a broad partition, but not a partition. Therefore

$$\Pi_5 \odot \Pi = \Pi_5 \tag{2.3.19}$$

is a cross of Π_5 and Π, but not a true cross.

(γ) The $\Pi_2 * \Pi_3 = \{(n, n+2] \bigcap (m, m+2] \mid n \text{ is odd }, m \text{ is even } \}$ is a broad partition of **R**, but not a partition. So

$$\Pi_2 \odot \Pi_3 = \Pi_4 \tag{2.3.20}$$

is a crossed partition of Π_2 and Π_3, but not a true crossed partition.

(δ) The $\Pi_3 * \Pi_4 = \{(n, n+2] \bigcap (m, m+1] \mid n \text{ is even}, m \text{ is integer}\}$ is a broad partition, but not a partition. So

$$\Pi_3 \odot \Pi_4 = \Pi_4 \tag{2.3.21}$$

is a crossed partition of Π_3 and Π_4, but not a true crossed partition.

Example 5. Given the Euclidean plane $\mathbf{R}^2 = \{(x,y)|x,y \in \mathbf{R}\}$, for any fix real number x and y, let

$$(x,\mathbf{R}) = \{(x,z)\,|\,z \in \mathbf{R}\} \tag{2.3.22}$$

$$(\mathbf{R},y) = \{(z,y)\,|\,z \in \mathbf{R}\} \tag{2.3.23}$$

Obviously

$$\Pi_1 = \{(x,\mathbf{R})\,|\,x \in \mathbf{R}\} \tag{2.3.24}$$

$$\Pi_2 = \{(\mathbf{R},y)\,|\,y \in ,\mathbf{R}\} \tag{2.3.25}$$

are two partition of $\mathbf{R}^2$, and

$$\Pi_1 \odot \Pi_2 = \Pi_1 * \Pi_2 = \mathbf{R}_2 \tag{2.3.26}$$

is a true cross of Π_1 and Π_2.

Example 6. Supposing

$$\Omega_1 = \{\omega_\alpha^{(1)}\,|\,\alpha \in J_1\} \tag{2.3.27}$$

$$\Omega_2 = \{\omega_\beta^{(2)}\,|\,\beta \in J_2\} \tag{2.3.28}$$

are two partitions of the causal space X, where J_1, J_2 are two index sets which may be finite, countable or uncountable, then

$$\Omega_1 \odot \Omega_2 = \{(\omega_\alpha^{(1)} \cap \omega_\beta^{(2)})\,|\,\alpha \in J_1, \beta \in J_2, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \emptyset\} \tag{2.3.29}$$

is a cross of Ω_1 and Ω_2.

Further supposing for any $\alpha \in J_1$ and $\beta \in J_2$ such that $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \emptyset$, then

$$\Omega_1 \odot \Omega_2 = \{(\omega_\alpha^{(1)} \cap \omega_\beta^{(2)})\,|\,\alpha \in J_1, \beta \in J_2\} \tag{2.3.30}$$

is a true cross of Ω_1 and Ω_2.

2.4 The third group of axioms: the group of axioms of random test

2.4.1 Intuitive background

The front two groups of axioms abstract the random universe as the causation space $(X,\mathcal{U})$. It is pity that neither we list all elements of X, nor list all elements of $\mathcal{U}$ [9]. Generally, people can only study an isolating and idealistic part, and the relations among different parts.

The experience tells us that the so-called "part" is a few of events concerned by us. Their number can be finite, countable or uncountable. **The part is denoted by $\mathcal{F}$**, and supposed

$$\mathcal{F} = \{A, A_i\,(i \in I), B, C, \cdots\}$$

Intuitive, the events generated with complement, union, intersection and difference also are the events concerned by us. i.e.

$$\overline{A},\ \bigcup_{i\in I} A_i,\ \bigcap_{i\in I} A_i,\ B \setminus C,\ \cdots$$

belong to $\mathcal{F}$. Therefore $\mathcal{F}$ is closed under the formation of operations "complement, union, intersection, difference"[10]. **In brief, $\mathcal{F}$ is an event σ-field.**

In order to study this part, at first we reduce the research objects from $(X,\mathcal{U})$ to $(X,\mathcal{F})$, where all elements of $\mathcal{F}$ can be listed or can be described generally.

The secondly, We discuss similar to Subsection 2.3.2 instead $\mathcal{U}$ by $\mathcal{F}$. For any $A \in \mathcal{F}$, let

$$\omega(A) = \begin{cases} 1, & A \text{ occurs.} \\ 0, & A \text{ does not occur.} \end{cases} \tag{2.4.1}$$

$$\Omega = \{\,\omega \mid \omega(A), A \in \mathcal{F} \text{ is a function with the properties ①–④ in Subsection 2.3.2 (of cause, } \mathcal{F} \text{ is instead of } \mathcal{U}\text{)}\} \tag{2.4.2}$$

Intuitively, ω represents that $\mathcal{F}$ is located in such a state: the event B in $\mathcal{F}$ occurs if $\omega(B) = 1$, and the event C does not occur if $\omega(C) = 0$.

By the same reason we get that every event in $\mathcal{F}$ is a subset of Ω. i.e. If $B \in \mathcal{F}$, we have

$$B = \{\,\omega \mid \omega \in \Omega \text{ and } \omega(B) = 1\,\} \tag{2.4.3}$$

The third, we discuss the relations between Ω and the causal space X. Comparing (2.3.1), (2.3.2) and (2.4.1), (2.4.2) we get in a sense of set

[9] The set consisted of all chemical elements is denoted by E, the set of all materials is denoted by $\mathcal{M}$. Therefore $(E, \mathcal{M})$ corresponds $(X,\mathcal{U})$, when the people study the materials, neither they list all elements of E nor list all elements of $\mathcal{M}$.

[10] The third group of axioms is an abstract of this intuitive fact.

operations

$$\omega = \{u \mid \text{u is a causal point with } u(A) = \omega(A),\ A \in \mathcal{F}\} \quad (2.4.4)$$

$$X = \bigcup_{\omega \in \Omega} \omega \quad (2.4.5)$$

Supposing ω_1 and ω_2 are different elements in Ω. Form (2.4.4) we get that ω_1 and ω_2 do not contain same causal point. Therefore, Ω is a partition of the causal space X.

The forth, the event B can now represents in three ways. That is the capital B in $\mathcal{U}$, the right of (2.3.3) in X, the right of (2.4.3) in Ω. That is

$$B = \{\omega \mid \omega \in B\} = \{u \mid u \in B\} \quad (2.4.6)$$

where the front equation is brief of (2.4.3); $B = \{u \mid u \in B\}$ is the brief of (2.3.3); the latter equation is brief of set's operation

$$\{\omega \mid \omega \in B\} \stackrel{def}{=} \bigcup_{\omega \in B} \omega = \{u \mid u \in B\}$$

By the discussion above, **we may reduce the research object from $(X, \mathcal{U})$ to $(\Omega, \mathcal{F})$. The important of this change is that all elements of both Ω and $\mathcal{F}$ can be listed, or can be described generally.**

2.4.2 The group of axioms of random test

Definition 2.4.1. *If $\Omega = \{\omega_\alpha \mid \alpha \in J\}$ is a partition of the causal space X, then Ω is called a sample space, and ω_α is called a sample point. The sample point ω_α is called pseudo-occurs if some causal point in ω_α pseudo-occurs.*

Definition 2.4.2. *Supposing B is an event, Ω is the sample space in Definition 2.4.1. If B is a subset of Ω, i.e. there exists a subset J_1 of J such that*

$$B = \bigcup_{\alpha \in J_1} \omega_\alpha \stackrel{def}{=} \{\omega_\alpha \mid \alpha \in J_1\} \quad (2.4.7)$$

then B is called an event on Ω.

Definition 2.4.3. *The pair $(\Omega, \mathcal{F})$ is called a random test if Ω is a sample space and $\mathcal{F}$ is a nonempty set which consists of some events on Ω and obeys the Axiom C_1 and Axiom C_2.*

Axiom C_1 (Axiom of complement closing). *For any $A \in \mathcal{F}$, the complement $\overline{A} \in \mathcal{F}$ is true.*

Axiom C_2 (Axiom of σ-union closing). *For any finite or countable $A_i \in \mathcal{F}, i \in I$, the union $\bigcup_{i \in I} A_i \in \mathcal{F}$ is true.*

By Definition 2.2.7 we know $\mathcal{F}$ is a special event σ-field which every element is an event on Ω. Such event σ-field is called that on Ω. Therefore, Definition 2.4.3 can be stated briefly.

Definition 2.4.3′ *The pair* $(\Omega, \mathcal{F})$ *is called a random test if* Ω *is a sample space and* $\mathcal{F}$ *is an event* σ*-field on* Ω.

For convenience of language, the random test is briefly the test[11]. We first write down several clear facts as

Lemma 2.4.1. *(1) For any event* σ*-field* $\mathcal{F}$*, the* $(X, \mathcal{F})$ *is a test.*

(2) Given an event σ*-field* $\mathcal{F}$*, if* Ω *is defined by (2.4.2), then the* $(\Omega, \mathcal{F})$ *is a test.*

(3) Supposing the $(\Omega, \mathcal{F})$ *is a test, if a partition* Ω^* *is finer than* Ω*, then the* $(\Omega^*, \mathcal{F})$ *is a test.*

This lemma has declared an important fact: **The sample space is only a tool to study the event σ-field. It can be selected freely in some meaning.** The $(\Omega, \mathcal{F})$ in (2) is called a living test.

Theorem 2.4.1. *Supposing the* $(\Omega, \mathcal{F})$ *is a test, then*

(1) There is one and only one sample point to pseudo-occur in Ω.

(2) The event B *in* $\mathcal{F}$ *occurs if and only if its some sample point pseudo-occurs.*

Proof: (1) Supposing $\Omega = \{\omega_\alpha \mid \alpha \in J\}$ is a sample space. By Axiom B_2 we know that there exists one and only one causal point to pseudo-occur in the causal space X. Since Ω is a partition of X, so this pseudo-occurrent causal point belongs and only belongs to some sample point ω_α. By Definition 2.4.1 we have got that there is the sample point ω_α and only one to pseudo-occur in the Ω.

(2) Supposing B has the form (2.3.3) and (2.4.7). If the B occurs, then the some causal point u^* in B pseudo-occurs. By (2.4.7) we deduce that the u^* belongs to some ω_α in B, so the sample point ω_α pseudo-occurs. On the contrary, If a sample point ω_α in B pseudo-occurs, then there is some $u^* \in \omega_\alpha$ to pseudo-occurs (Definition 2.4.1). It is from $\omega_\alpha \in B$ that $u^* \in B$ and B occurs. The proof is completed. ■

We get the following theorem By the discussion which is similar to Theorem 2.3.1.

Theorem 2.4.2. *In the random test* $(\Omega, \mathcal{F})$*, the events and their operations in* $\mathcal{F}$ *and the sets and their operations in* Ω *take on the identity that Table 2.2 represents.*

---

[11] There are not the differences between the experience and the test in probability theory, their definitions were not given exactly. Now, the test has an exact definition in the NAS.

Table 2.2 Probability Theory Vs. Set Theory
(In Random Test)

The Random Test $(\Omega, \mathcal{F})$			
	subject	**Set Theory**	**Probability Theory**
Sym.		Sample space Ω	Event σ-field $\mathcal{F}$ on Ω
Basic Elements	ω	Sample point	May be an event or may not
	F	Any set in Ω	May be an event or may not
	A	Special set in Ω (A group of causes)	Event in $\mathcal{F}$
	$\emptyset$	Empty set	Impossible event
	Ω	Sample space	Sure event
	$\mathcal{F}$	A special σ-field	Event σ-field on Ω
	$(\Omega, \mathcal{F})$	A measurable space	Random test
As follows: all A, B, A_i, A_j are events; I is finite or countable index set; J is uncountable index set.			
Basic Relations	$\omega \in A$	Sample point ω is an element in A.	Pseudo-occurrence of ω causes event A to occur.
	$A = B$	A and B consist of the same sample points.	If A occurs then B must occur and If B occurs then A must occur.
	$A \subset B$	Every sample point of A belongs to B.	If A occurs then B must occur.
	$\overline{A}$	The set consists of all sample points which do not belong to A.	The complement event: "A does not occur".
	$\bigcup_{i\in I} A_i$	The set consists of all sample points of all $A_i\,(i \in I)$.	The union event: "at least one event of all $A_i\,(i \in I)$ occurs".
	$\bigcup_{j\in J} A_j$	The set consists of all sample points of all $A_j\,(j \in J)$.	No meaning.
	$\bigcap_{i\in I} A_i$	The set consists of all sample points which belong to every $A_i\,(i \in I)$.	The intersection event: "all events $A_i\,(i \in I)$ occur together".
	$\bigcap_{j\in J} A_j$	The set consists of all sample points which belong to every $A_j\,(j \in J)$.	No meaning.
	$A \setminus B$	The set consists of all sample points which belong to A, but not to B.	The difference event: "A occurs, but B does not occur".

2.4.3 Random sub-tests

Definition 2.4.4. *Suppose* $(\Omega_i, \mathcal{F}_i)(i = 1, 2)$ *are two tests. If* $\mathcal{F}_2 \subset \mathcal{F}_1$, *then* $(\Omega_2, \mathcal{F}_2)$ *is called a random sub-test of* $(\Omega_1, \mathcal{F}_1)$, *or sub-test in brief. In particular, if* $\Omega_1 = \Omega_2$, *then it is called a sub-test with same causes.*

Obviously, the causation space $(X, \mathcal{U})$ itself is a test. And any test is its sub-tests. In addition,

(1) $\Omega_0 = \{X\}$ is the coarsest partition of the causal space, $\mathcal{F}_0 = \{\emptyset, X\}$ is the minimum event σ-field. Obviously, $(\Omega_0, \mathcal{F}_0)$ is a test, and is a sub-test of any test. The $(\Omega_0, \mathcal{F}_0)$ is called a degeneration test.

(2) Suppose A is any event. Obviously $\Omega = \{A, \overline{A}\}$ is a partition of causal space. The $\mathcal{F} = \{\emptyset, A, \overline{A}, \Omega\}$ is an event σ-field on Ω. Therefore $(\Omega, \mathcal{F}\}$ is a test. We call it the test generated by A, or Bernoulli's test.

In Table 0.1 we gave the similarity of the causation space to the Euclidean space, and the similarity of the tests to the geometrical figures. Therefore it can not be absent for cultivating the appearance of the causation space and studying probability theory to familiarize a lot of "the graphics"— the random tests. From now on until the end of Section 2.6 we will have done this job.

2.4.4 Modelling of random test(1): Listing method

The subject of probability theory contains two kinds of works. The first is **building theoretical system by the axiom system**. The used methods are logical deduction and mathematical calculation. For this kind of work, if the premise is right, then the conclude deducted is certainly right.

The another kind of word is that **the random phenomena concerned by people in real world are abstracted as the research object of the theoretical system.** We call this job as **probability theory modelling**, or modelling in brief; After the modelling, the first kind of work is be deducted, and various concludes are got; Finally the statistical laws implied in the random phenomena are represented by the obtaining concludes.

Modelling is a kind of relations between the reflecting and the reflected. People can not talk whether modelling is right or not, can only talk whether modelling is rational or not, and is success or not. Three criterions to decide the rationality of modelling are as follows:

Criterion 1: The model retains some essence of the real problem and is adaptable for user, so people have received this model.

Criterion 2: Many various models may be abstracted from one real problem. The reason is that various models emphasizes separately some aspects of the real problem.

Criterion 3: The mode reflects a kind of idealistic aspect. The degree of the coincidence between it and the real problem needs to be verified by practice.

There are modelling works always in use. In use of geometry the heavenly bodies such as the sun, the earth, etc. are regarded as the particles sometimes, and are regarded as the balls or the elliptic sometimes. A section of surrounding wall built on the ground is regarded as a curve on the plane sometimes, and is regarded as a belt shape sometimes. This kind of job belongs to modelling of geometry.

Because geometric modelling is over-agile and always varying, furthermore people have the "inborn" modelling genius, so the geometry refuses to absorb the contents of modelling as a part of itself. The probability theory is different, people familiarize with neither the research object in theoretical system nor modelling methods in use. So we must familiarize with the research object and modelling methods by a lot of typical examples.

Example 1 (The experiment of tossing a coin). *Throwing a coin on a desk (Experiment $\mathcal{E}_1$), try to find the random test generated by the $\mathcal{E}_1$.*

Solution: After any performance of Experiment $\mathcal{E}_1$, if the head (or tail) does not occur then the tail (or head) must occur. Supposing H and T represent separately "head" and "tail', therefore two events H and T are a pair of opposite events, and

$$\Omega_1 = \{ H,\ T \} \tag{2.4.8}$$

is the sure event. So Ω_1 is a partition of the causal space, and is a sample space[12]. Let

$$\mathcal{F}_1 = \{ \emptyset,\ < H >,\ < T >,\ \Omega_1 \} \tag{2.4.9}$$

Obviously it is an event σ-field on Ω_1. Therefore $(\Omega_1, \mathcal{F}_1)$ is the random test generated by Experiment $\mathcal{E}_1$. ■

The answer $(\Omega_1, \mathcal{F}_1)$ has a clear intuitive background. When people ask: "What result will appear if the people throw a coin"? Some people many answer: " I do not know "; Another people may answer: " If the head does not appear then the tail must appear "; There are some people who answer: " **The result to appear is and only is an element event in Ω_1, all events concerned by us constitute an event σ-field $\mathcal{F}_1$ on Ω_1** ".

The three answers in life all are rational and acceptable. But only the third answer is scientific. The reason is that the random test $(\Omega_1, \mathcal{F}_1)$

[12] The sample points in the application problems almost are the events. It is possible reason that Kolmogorov called the sample points introduced by Mises as elementary events[1]. Therefore people call the sample space as elementary event space also. Even so, the elementary events need not belong to $\mathcal{F}$ in some test $(\Omega, \mathcal{F})$, see Example 6 and Example 7.

comprehensively, exactly reflects the results of Experiment $\mathcal{E}_1$; **The most important reason is that the results of any experiment always may be described comprehensively, exactly by a random tests.**

Example 2 (The experiment of tossing a coin). *Throwing a coin on a desk (Experiment $\mathcal{E}_2$), try to find the random test generated by the $\mathcal{E}_2$.*

Solution: After any performance of Experiment $\mathcal{E}_2$, the head (abbr. H) can occur, the tail (abbr. T) can occur also, and " the stand state (abbr. S) " can also occur. Furthermore it is believed that there is one and only one to occur in three cases. So .

$$\Omega_2 = \{ H,\ T,\ S \} \tag{2.4.10}$$

is the sure event, also is a partition of the causal space, and is a sample space. Let

$$\mathcal{F}_2 = \text{the set consisted of all subsets of } \Omega_2 \tag{2.4.11}$$

It is easy to verify $\mathcal{F}_2$ is an event σ-field on Ω_2, and $(\Omega_2, \mathcal{F}_2)$ is the random test generated by Experiment $\mathcal{E}_2$. ∎

The word " experiment " sometimes is clear and sometimes is vague in life. So-called " clear " means that the conditions obeyed by the experiment are " self-evident " and considered commonly by people. So-called " vague" means that people can not list all conditions without any loss. So we can transfer one experiment into two experiments with the lost conditions. Now, the test throwing a coin on the desk is an example of this kind of cases: When we obey the conditions of " self-evident ", it is Experiment $\mathcal{E}_1$; When there are crevices on the desk, it is Experiment $\mathcal{E}_2$.

One task of science is to delete this kind of ambiguity. The random test $(\Omega_i, \mathcal{F}_i)$ defines a part of condition of experiment $\mathcal{E}_i$. So when doing theoretical research, we may depict a part of conditions of Experiment $\mathcal{E}_i$ by using the test $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2)$; When doing probability theory modelling, we use the conditions of " self-evident " as the conditions of performance of experiment. i.e. **In application of probability theory we regard the "self-evident" conditions as the conditions of performance of experiment $\mathcal{E}$ so that we obtain the test $(\Omega, \mathcal{F})$ generated by $\mathcal{E}$; When doing theoretical research, we may depict a part of conditions of performance of experiment $\mathcal{E}$ by the test $(\Omega, \mathcal{F})$[13].**

Example 3. *Knowing nothing, forecast the sex of some fetus (Experiment $\mathcal{E}_3$), try to find the test generated by $\mathcal{E}_3$.*

Solution: Since the sex of the fetus is either male or female, so

$$\Omega_3 = \{ \text{ male, female } \} \tag{2.4.12}$$

[13] i.e. The experiment $\mathcal{E}$ generates the test $(\Omega, \mathcal{F})$. The probability space $(\Omega, \mathcal{F}, P)$ in Section 2.7 will describe comprehensively, exactly all conditions of experiment $\mathcal{E}$.

is a sample space generated by $\mathcal{E}_3$. Let

$$\mathcal{F}_3 = \{\emptyset, < male >, < female >, \Omega_3\} \tag{2.4.13}$$

Obviously it is an event σ-field on Ω_3, and $(\Omega_3, \mathcal{F}_3)$ is the random test generated by Experiment $\mathcal{E}_3$. ■

Example 4. *Forecast the flower's color of a plant of the second filial generation of pea hybrid (Experiment $\mathcal{E}_4$), try to find the random test generated by $\mathcal{E}_4$.*

Solution: From Subsection 1.4.5(4) we know that the flower's color of the plants are either red or white. So

$$\Omega_4 = \{ \text{ red, white } \} \tag{2.4.14}$$

is a sample space generated by $\mathcal{E}_4$. Let

$$\mathcal{F}_4 = \{\emptyset, < red >, < white >, \Omega_4\} \tag{2.4.15}$$

Obviously it is an event σ-field on Ω_4, and $(\Omega_4, \mathcal{F}_4)$ is the random test generated by Experiment $\mathcal{E}_4$. ■

Example 5 (The experiment of throwing a dice). *Throwing a dice on a desk (Experiment $\mathcal{E}_5$), try to find the random test generated by the $\mathcal{E}_5$.*

Solution: Since the point number of the dice to occur is only one in 1, 2, 3, 4, 5, 6. So

$$\Omega_5 = \{ 1, 2, 3, 4, 5, 6 \} \tag{2.4.16}$$

is the sure event, also is a partition of the causal space X. i.e. It is a sample space generated by $\mathcal{E}_5$. Let

$$\mathcal{F}_5 = \text{ the set consisted of all subsets of } \Omega_5 \tag{2.4.17}$$

Obvious it is an event σ-field on Ω_5. Therefore $(\Omega_5, \mathcal{F}_5)$ is the random test generated by Experiment $\mathcal{E}_5$. ■

Example 6. *What we concern is only odd or even of the point number to occur in Example 5 (Experiment $\mathcal{E}_6$), try to find the random test generated by the $\mathcal{E}_6$.*

Solution 1: The symbols at the example above is used here. The subsets $\{1,3,5\}$ and $\{2,4,6\}$ are separately the events "the odd occurs" and "the even occurs". Let

$$\mathcal{F}_6 = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega_5\} \tag{2.4.18}$$

Obviously it is an event σ-field on Ω_5, and it contains all events concerned by Experiment $\mathcal{E}_6$. So $(\Omega_5, \mathcal{F}_6)$ is the random test generated by Experiment $\mathcal{E}_6$.

Solution 2: By ω_o and ω_e separately denoted the events "the odd occurs" and "the even occurs". Since result for the dice to occur is either odd or even, so

$$\Omega_6 = \{\, \omega_o,\ \omega_e \,\} \tag{2.4.19}$$

is the sure event, also is a partition of the causal space. i.e. The Ω_6 is a sample space generated by $\mathcal{E}_6$. Let

$$\mathcal{F}_6 = \{\, \emptyset,\ <\omega_o>,\ <\omega_e>,\ \Omega_6 \,\} \tag{2.4.20}$$

It is an event σ-field on Ω_6. So $(\Omega_6, \mathcal{F}_6)$ is the random test generated by Experiment $\mathcal{E}_6$. ■

Example 7. *In Example 5, the points 5, 6 are regarded as the large point, the points 1, 2, 3, 4 are regarded as the small point. When we concern only either "the large point to occur" or "the small point to occur" (Experiment $\mathcal{E}_7$), try to find the random test generated by $\mathcal{E}_7$.*

Solution: The events "the large point to occur" and "the small point to occur" separately are denoted by ω_l and ω_s. After any performance of Experiment $\mathcal{E}_7$, there is one and only one event of both "the large point" and "the small point" to occur. So

$$\Omega_7 = \{\, \omega_l,\ \omega_s \,\} \tag{2.4.21}$$

is the sure event, also is a partition of the causal space. i.e. The Ω_7 is a sample space generated by $\mathcal{E}_7$. Let

$$\mathcal{F}_7 = \{\, \emptyset,\ <\omega_l>,\ <\omega_s>,\ \Omega_7 \,\} \tag{2.4.22}$$

Obviously it is an event σ-field on Ω_7. So $(\Omega_7, \mathcal{F}_7)$ is the random test generated by Experiment $\mathcal{E}_7$. ■

It is easy to see that the partition Ω_5 of the causal space is finer than Ω_7. So the $(\Omega_5, \mathcal{F}_7)$ is a random test by Lemma 2.4.1. Obviously, it is also the random test generated by Experiment $\mathcal{E}_7$.

Intuitively, Experiment $\mathcal{E}_6$ and $\mathcal{E}_7$ are a part of Experiment $\mathcal{E}_5$, they may be called the sub-experiments of $\mathcal{E}_5$. In the NAS this fact is represented as that $(\Omega_6, \mathcal{F}_6)$ and $(\Omega_7, \mathcal{F}_7)$ is the sub-test of $(\Omega_5, \mathcal{F}_5)$; and $(\Omega_5, \mathcal{F}_6)$ and $(\Omega_5, \mathcal{F}_7)$ is the sub-test with same causes of $(\Omega_5, \mathcal{F}_5)$. In the test $(\Omega_5, \mathcal{F}_5)$ we have

$$<\omega_o> \;=\; \{\,1,\ 3,\ 5\,\} \qquad <\omega_e>=\{\,2,\ 4,\ 6\} \tag{2.4.23}$$

$$<\omega_l> \;=\; \{\,5,\ 6\,\} \qquad <\omega_s>=\{\,1,\ 2,\ 3,\ 4\,\} \tag{2.4.24}$$

The random tests $(\Omega_5, \mathcal{F}_6)$ and $(\Omega_5, \mathcal{F}_7)$ have verified the fact in the footnote 12: the element events need not belong to $\mathcal{F}$ in some test $(\Omega, \mathcal{F})$.

Example 8. *There are n balls in a bag which labelling separately $1, 2, \cdots, n$. Now we fetch a ball from the bag anyway (Experiment $\mathcal{E}_8$), try to find the random test generated by $\mathcal{E}_8$.*

Solution: The event " fetching out the ball i " is denoted by ω_i. Since all nonempty events ω_i $(1 \le i \le n)$ are exclusive each other and their union is the sure event, so

$$\Omega_8 = \{\omega_1, \omega_2, \cdots, \omega_n\} \tag{2.4.25}$$

is a partition of the causal space. i.e. The Ω_8 is a sample space generated by $\mathcal{E}_8$. Let

$$\mathcal{F}_8 = \text{ the set consisted of all subsets of } \Omega_8 \tag{2.4.26}$$

Obviously it is an event σ-field on Ω_8, and $(\Omega_8, \mathcal{F}_8)$ is the random test generated by Experiment $\mathcal{E}_8$. ■

Example 9. *Throwing a coin continuously, it does not stop until the head occurs first (Experiment $\mathcal{E}_9$), try to find the random test generated by the $\mathcal{E}_9$.*

Solution: The event "front $n-1$ times the tail occurs and the nth time the head occurs" is denoted by $\omega_n \equiv (\underbrace{T, T, \cdots, T}_{n-1\, times}, H)$, and the event "throw always without stop" is denoted by $\omega_\infty \equiv (T, T, \cdots)$. Since all nonempty events ω_n $(1 \le n \le \infty)$ are exclusive each other and their union is the sure event, so

$$\Omega_9 = \{\omega_1, \omega_2, \cdots, \omega_\infty\} \tag{2.4.27}$$

is a partition of the causal space. i.e. The Ω_9 is a sample space generated by $\mathcal{E}_9$. Let

$$\mathcal{F}_9 = \text{the set consisted of all subsets of } \Omega_9 \tag{2.4.28}$$

Obviously it is an event σ-field on Ω_9, and $(\Omega_9, \mathcal{F}_9)$ is the random test generated by Experiment $\mathcal{E}_9$. ■

Example 10. *A particle is thrown on the line* $\mathbf{R} = (-\infty, +\infty)$ *(Experiment $\mathcal{E}_{10}$), try to find the random test generated by $\mathcal{E}_{10}$.*

Solution: The event "the particle falls on the point with the coordinate x" is denoted by x. Obviously, all events x $(x \in \mathbf{R})$ are exclusive each other (i.e. for any real numbers $x \ne y$, the event's operation has $x \cap y = \emptyset$), and there is one and only one event to occur in all events. So the set consisted of these events is the sure event.[14] Therefore

$$\mathbf{R} = \{x \mid -\infty < x < +\infty\} \tag{2.4.29}$$

[14] Here the sentence "the union of all events is the sure event" can not be said. The reason is that the number of those events is uncountable and the uncountable union of events is meaningless.

is a partition of the causal space. i.e. The $\mathbf{R}$ is a sample space generated by $\mathcal{E}_{10}$.

Let A is any subset of $\mathbf{R}$. The A also denotes the event " the particle falls in A". Supposing

$$\mathcal{F}_{10} = \{ A \mid A \subset \mathbf{R}\} \tag{2.4.30}$$

obviously it is an event σ-field on $\mathbf{R}$, and $(\mathbf{R}, \mathcal{F}_{10})$ is the random test generated by Experiment $\mathcal{E}_{10}$. ■

Example 11. *A particle is thrown in the n-dimensions Euclidean space* $\mathbf{R}^n$ *(Experiment* $\mathcal{E}_{11}$*), try to find the random test generated by* $\mathcal{E}_{11}$.

Solution: Let $(x_1, x_2, \cdots, x_n)$ is the coordinate of points. It denotes also the event "the particle falls on the point with the coordinate $(x_1, x_2, \cdots, x_n)$". The subset A of $\mathbf{R}^n$ denotes the event " the particle falls in A". Supposing

$$\mathbf{R}^n = \{ (x_1, x_2, \cdots, x_n) \mid -\infty < x_1, x_2, \cdots, x_n < +\infty\} \tag{2.4.31}$$

$$\mathcal{F}_{11} = \{ A \mid A \subset \mathbf{R}^n\} \tag{2.4.32}$$

then $(\mathbf{R}^n, \mathcal{F}_{11})$ is the random test generated by Experiment $\mathcal{E}_{11}$ by the discussion similar to Example 10. ■

2.4.5 Modelling of random test(2): Expansionary method

Suppose Ω is a sample space and $\mathcal{A}$ is a nonempty set which consists of some events on Ω.

Definition 2.4.5. $\mathcal{F}_0$ *is called as the event* σ*-field (or event field) generated by* $\mathcal{A}$*, if* $\mathcal{F}_0$ *is an event* σ*-field (or event field) on* Ω *implying* $\mathcal{A}$*, and for any event* σ*-field (or event field)* $\mathcal{F}$ *to set up* $\mathcal{F} \supset \mathcal{F}_0$.

The event σ-field $\mathcal{F}_0$ usually is denoted by $\sigma(\mathcal{A})$, the event field $\mathcal{F}_0$ usually is denoted by $r(\mathcal{A})$. The $\sigma(\mathcal{A})$ and $r(\mathcal{A})$ also are called smallest event σ-field and smallest event field containing $\mathcal{A}$, respectively.

First we will prove the existence of $\sigma(\mathcal{A})$ and $r(\mathcal{A})$. In the NAS we can not use the method which is popular now in probability theory. The popular method transform the proof of existence into that in set theory. The latter has used the conclude "all subset of Ω is an σ-field(or field)". Since this σ-field (or field) may not an event σ-field (or event field), we can not guarantee that all elements in $\sigma(\mathcal{A})$ and $r(\mathcal{A})$ are events (see Table 2.2).

We give another proof of existence [17]. This is a constructive proof, so the construction of $\sigma(\mathcal{A})$ and $r(\mathcal{A})$ are displayed clearly.

Theorem 2.4.3. *Supposing* $\mathcal{A}$ *is a nonempty set which consists of some events on* Ω*, then there exist unique* $\sigma(\mathcal{A})$ *and unique* $r(\mathcal{A})$.

Proof : (1) the existence and unique of $r(\mathcal{A})$
Let $\mathcal{D}$ is a set which consists of some events on Ω. Suppose

$$\mathcal{D}^* = \{\bigcup_{i=1}^{m} A_i,\ \bigcap_{i=1}^{m} A_i \,|\, m \geq 1,\ A_i \ \mathit{or}\ \overline{A}_i \in \mathcal{D},\ 1 \leq i \leq m\} \tag{2.4.33}$$

That is that $\mathcal{D}^*$ is the set which is generated by below means: first the complements of events in $\mathcal{D}$ are added to $\mathcal{D}$ (the new set is denoted by $\widetilde{\mathcal{D}}$), afterwards the finite unions and finite intersections of events in $\widetilde{\mathcal{D}}$ are added to $\widetilde{\mathcal{D}}$ and finally $\mathcal{D}^*$ is obtained. The method obtaining $\mathcal{D}^*$ from $\mathcal{D}$ may be called the operation "*". Now we perform a sequence of operations "*" on $\mathcal{A}$. Suppose

$$\mathcal{R}_1 = \mathcal{A}^*, \qquad \mathcal{R}_n = \mathcal{R}_{n-1}^* \quad (n = 2, 3, \cdots), \tag{2.4.34}$$

$$\mathcal{R}_0 = \bigcup_{n=1}^{\infty} \mathcal{R}_n \tag{2.4.35}$$

Obviously, the elements of $\mathcal{R}_0$ all are the events on Ω. We show that it is an event field as follows. In fact, supposing $A \in \mathcal{R}_0$, by (2.4.35) we deduct that there exists an integral number n such that $A \in \mathcal{R}_n$. It is from the definition of the operation "*" that $\overline{A} \in \mathcal{R}_{n+1}$, so $\overline{A} \in \mathcal{R}_0$ is proved. Furthermore supposing $A_1, A_2, \cdots, A_n \in \mathcal{R}_0$, then there exist the integral numbers $s_1, s_2, \cdots, s_n$ such that $A_i \in \mathcal{R}_{s_i}$ $(i = 1, 2, \cdots n)$. Let $m = \max\{s_1, s_2, \cdots, s_n\}$, It is from the definition of the operation "*" that $\bigcup_{i=1}^{n} A_i \in \mathcal{R}_{m+1}$ and $\bigcup_{i=1}^{n} A_i \in \mathcal{R}_0$. So we have proved that $\mathcal{R}_0$ is an event field.

Suppose $\mathcal{R}$ is any event field containing $\mathcal{A}$. It is from the definition of the operation "*" that $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \cdots$ all are contained in $\mathcal{R}$. So $\mathcal{R}_0 \subset \mathcal{R}$, i.e. $\mathcal{R}_0$ is unique smallest event field containing $\mathcal{A}$ and the proof of (1) is completed.

(2) the existence and unique of $\sigma(\mathcal{A})$
Suppose

$$\mathcal{D}^{**} = \{\bigcup_{i \in I} A_i,\ \bigcap_{i \in I} A_i \,|\, A_i \ \mathit{or}\ \overline{A}_i \in \mathcal{D},\ \ i \in I\ \} \tag{2.4.36}$$

That is that $\mathcal{D}^{**}$ is the set which is generated by below means: first the complements of events in $\mathcal{D}$ are added to $\mathcal{D}$ (the new set is denoted by $\widetilde{\mathcal{D}}$), afterwards the σ-unions and σ-intersections of events in $\widetilde{\mathcal{D}}$ are added to $\widetilde{\mathcal{D}}$ and finally $\mathcal{D}^{**}$ is obtained. The method obtaining $\mathcal{D}^{**}$ from $\mathcal{D}$ may be called the operation "**". Now we perform a sequence of operations "**" on $\mathcal{A}$. Let

$$\mathcal{F}_1 = \mathcal{A}^{**}, \qquad \mathcal{F}_n = \mathcal{F}_{n-1}^{**} \quad (n = 2, 3, \cdots), \tag{2.4.37}$$

$$\mathcal{F}_w = \bigcup_{n=1}^{\infty} \mathcal{F}_n, \quad \mathcal{F}_{w+1} = \mathcal{F}_w^{**}, \quad \mathcal{F}_{w+2} = \mathcal{F}_{w+1}^{**}, \cdots \tag{2.4.38}$$

where w is the first infinite ordinal number[15].

Generally, supposing α is any countable ordinal number, and $\mathcal{F}_\beta$ $(\beta < \alpha)$ have all defined clearly. If α is an isolated ordinal number, then there is an ordinal number α_1 such that $\alpha_1 + 1 = \alpha$. At that time we define

$$\mathcal{F}_\alpha = \mathcal{F}_{\alpha_1}^{**} \tag{2.4.39}$$

If α is a limit ordinal number, then at that time we define

$$\mathcal{F}_\alpha = \bigcup_{\beta<\alpha} \mathcal{F}_\beta \tag{2.4.40}$$

By mathematical superinduction for all countable ordinal number α we have defined $\mathcal{F}_\alpha$. Therefore we have got the super-sequence of the event sets

$$\mathcal{F}_1,\ \mathcal{F}_2,\ \cdots,\ \mathcal{F}_n,\ \cdots,\ \mathcal{F}_w,\ \mathcal{F}_{w+1},\ \cdots,\ \mathcal{F}_\alpha,\ \mathcal{F}_{\alpha+1},\ \cdots \tag{2.4.41}$$

As we know, for any α, $\{\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_\alpha\}$ is a countable set, but the total number of (2.4.41) is uncountable. Let

$$\mathcal{F}_0 = \bigcup_{\alpha} \mathcal{F}_\alpha \tag{2.4.42}$$

Obviously, the elements of $\mathcal{F}_0$ all are the events on Ω. We show it is an event σ-field as follows. In fact, supposing $A \in \mathcal{F}_0$, by (2.4.42) we deduct that there exists an ordinal number α such that $A \in \mathcal{R}_\alpha$. It is from the definition of the operation "**" that $\overline{A} \in \mathcal{R}_{\alpha+1}$, so $\overline{A} \in \mathcal{F}_0$ is proved. Furthermore supposing $A_i \in \mathcal{F}_0\,(i \in I)$, then there exist the countable ordinal α_i such that $A_i \in \mathcal{F}_{\alpha_i}\,(i \in I)$. By β we denote the first ordinal number after the finite or countable ordinal numbers $\alpha_i, i \in I$. We know β is also a countable ordinal number. It is from the definition of the operation "**" that $\bigcup_{i\in I} A_i \in \mathcal{F}_{\beta+1}$, and $\bigcup_{i\in I} A_i \in \mathcal{F}_0$. So we have proved that $\mathcal{F}_0$ is an event σ-field.

[15] See [17] about the knowledge of the ordinal number. In set theory the first infinite ordinal number is denoted by ω. Because the sample point is denoted by ω in probability theory, we denote the first infinite ordinal number by w.

Suppose $\mathcal{F}$ is any event σ-field containing $\mathcal{A}$. It is from the mathematical superinduction and the definition of the operation "**" that $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_\alpha$, $\cdots$ all are contained in $\mathcal{F}$. So $\mathcal{F}_0 \subset \mathcal{F}$. i.e. $\mathcal{F}_0$ is unique smallest event σ-field containing $\mathcal{A}$ and the proof of (2) is completed. ■

Corollary: *Suppose both of $\mathcal{A}_1$ and $\mathcal{A}_2$ are nonempty sets which consists of some events on Ω. If the events of $\mathcal{A}_1$ all can be represented by the complements, unions, intersections and differences of the events of $\mathcal{A}_2$, then*

$$\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_2) \tag{2.4.43}$$

It is easy to point out that when $\sigma(\mathcal{A})$ is constructed by the operation "**", may appear following situation: there exists some countable ordinal number α such that $\mathcal{F}_\alpha = \mathcal{F}_{\alpha+1} = \cdots$. For example, when $\mathcal{A}$ itself is an event σ-field, we have $\mathcal{F}_1 = \mathcal{F}_2 = \cdots$. But there exist such Ω and $\mathcal{A}$ that for any countable ordinal number α, $\mathcal{F}_0 \neq \bigcup_{\beta \leq \alpha} \mathcal{F}_\beta$ is set up[17].

Definition 2.4.6. *If the events in $\sigma(\mathcal{A})$ are total events concerned by us, then $\mathcal{A}$ is called a bud set of events, or a bud set in brief. The event in $\mathcal{A}$ is called the bud event.*

Example 12. *Throw a particle on the line $\mathbf{R} = (-\infty, +\infty)$, the bud set is*

$$\mathcal{A}_{12} = \{ (a, b] \mid -\infty < a < b < +\infty \} \tag{2.4.44}$$

where $(a, b]$ represents the event "the particle falls into the interval $(a, b]$" (Experiment $\mathcal{E}_{12}$), try to find the random test generated by $\mathcal{E}_{12}$.

Solution: What $\mathbf{R}$ is the sample space of $\mathcal{E}_{12}$ is known by the discussion of Example 10. Obviously, the bud events all are the events on $\mathbf{R}$. By the meaning of the problem, total events concerned by us constitute $\sigma(\mathcal{A}_{12})$. So $(\mathbf{R}, \sigma(\mathcal{A}_{12}))$ is the random test generated by $\mathcal{E}_{12}$. ■

Example 13. *Throw a particle on the line $\mathbf{R} = (-\infty, +\infty)$, the bud set is*

$$\mathcal{A}_{13} = \{ (a, a+1] \mid a \text{ is an integer} \} \tag{2.4.45}$$

where $(a, a+1]$ represents the event "the particle falls into the interval $(a, a+1]$" (Experiment $\mathcal{E}_{13}$), try to find the random test generated by $\mathcal{E}_{13}$.

Solution: If we select $\mathbf{R}$ as the sample space, Obviously, the bud events all are the events on $\mathbf{R}$. Since all events concerned by us constitute $\sigma(\mathcal{A}_{13})$, by the meaning of the problem, $(\mathbf{R}, \sigma(\mathcal{A}_{13}))$ is the random test generated by $\mathcal{E}_{13}$. ■

Example 14. *Throw a particle on the Euclidean plane $\mathbf{R}^2 = \{(x, y) \mid x, y \in \mathbf{R}\}$, the bud set is*

$$\mathcal{A}_{14} = \{ (a_1, b_1] \times (a_2, b_2] \mid -\infty < a_i < b_i < +\infty, i = 1, 2\} \tag{2.4.46}$$

where $(a_1, b_1] \times (a_2, b_2]$ represents the event "the particle falls into the rectangle $(a_1, b_1] \times (a_2, b_2]$" (Experiment $\mathcal{E}_{14}$), try to find the random test generated by $\mathcal{E}_{14}$.

Solution: By the same reason as Example 10 we get that $\mathbf{R}^2$ is the sample space of $\mathcal{E}_{14}$. Obviously, By the meaning of the problem, all events concerned by us constitute $\sigma(\mathcal{A}_{14})$. So $(\mathbf{R}^2, \sigma(\mathcal{A}_{14}))$ is the random test generated by $\mathcal{E}_{14}$. ■

Example 15. *Throw a particle in the Euclidean space $\mathbf{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbf{R}\}$, the bud set is*

$$\mathcal{A}_{14} = \{ \prod_{i=1}^{3} (a_i, b_i] \mid -\infty < a_i < b_i < +\infty, i = 1, 2, 3\} \qquad (2.4.47)$$

where $\prod_{i=1}^{3}(a_i, b_i]$ represents the event "the particle falls into the cuboid $\prod_{i=1}^{3}(a_i, b_i]$" (Experiment $\mathcal{E}_{15}$), try to find the random test generated by $\mathcal{E}_{15}$.

Solution: By the same reason as Example 10 we get that $\mathbf{R}^3$ is the sample space of $\mathcal{E}_{15}$. Obviously, By the meaning of the problem, all events concerned by us constitute $\sigma(\mathcal{A}_{15})$. So $(\mathbf{R}^3, \sigma(\mathcal{A}_{15}))$ is the random test generated by $\mathcal{E}_{15}$. ■

2.4.6 Remark

(1) The knowledge of the structures of $r(\mathcal{A})$ and $\sigma(\mathcal{A})$ are very helpful to understand the probability theory. Furthermore their structures declare the countable is identified with the finite and the uncountable is really "the true infinite" in probability theory.

(2) Example 12 and Example 10 are two different random tests. We will display their essential difference, referencing Theorem 2.5.8 in the next section and Subsection 2.8.5.

2.5 Several kinds of typical random tests

We can introduce a lot of random tests by use of the listing and the expansionary method. Many examples can be found in various introduction books of probability theory. In this section we classify common random tests as several kinds by the isomorphism in the algebra.

2.5.1 σ-isomorphism; Standard tests

Definition 2.5.1. *Suppose $\mathcal{F}_1$ and $\mathcal{F}_2$ are two event σ-fields. If there exists a correspondence φ of an one to one between $\mathcal{F}_1$ and $\mathcal{F}_2$,*

$$\varphi: \quad \mathcal{F}_1 \ni A \longleftrightarrow \varphi(A) \in \mathcal{F}_2 \tag{2.5.1}$$

such that for any A, $A_i \in \mathcal{F}_1$ $(i \in I)$ we have

$$\varphi(\overline{A}) = \overline{\varphi(A)} \tag{2.5.2}$$

$$\varphi(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} \varphi(A_i) \tag{2.5.3}$$

then $\mathcal{F}_1$ and $\mathcal{F}_2$ are called as σ-isomorphic, and φ is called as a σ-isomorphic mapping.

Definition 2.5.2. *The random tests $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ is called as σ-isomorphic, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-isomorphic event σ-field.*

Lemma 2.5.1. *Suppose φ is a σ-isomorphic mapping between $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, then for any $A, B, A_i \in \mathcal{F}_1$ $(i \in I)$ we have*

$$(1) \qquad \varphi(\Omega_1) = \Omega_2; \quad \varphi(\emptyset) = \emptyset \tag{2.5.4}$$

$$(2) \qquad \varphi(\bigcap_{i\in I} A_i) = \bigcap_{i\in I} \varphi(A_i) \tag{2.5.5}$$

$$(3) \qquad \varphi(A \backslash B) = \varphi(A) \backslash \varphi(B) \tag{2.5.6}$$

The proof is easy and left to the readers. Obviously, the tests of Examples 1, 3, 4, 6, 7 in Section 2.4 all are mutually σ-isomorphic.

Definition 2.5.3. *The random test $(\Omega, \mathcal{F})$ is called standard, if for any $\omega \in \Omega$, we have $\{\omega\} \in \mathcal{F}$.*

Usually, the random tests generated by the experiments often are standard. For example, $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2, \cdots, 9)$, $(\mathbf{R}, \mathcal{F}_{10})$, $(\mathbf{R}^n, \mathcal{F}_{11})$, $(\mathbf{R}, \sigma(\mathcal{A}_{12}))$, $(\mathbf{R}^2, \sigma(\mathcal{A}_{14}))$ and $(\mathbf{R}^3, \sigma(\mathcal{A}_{15}))$ in Section 2.4 all are standard. But sometimes nonstandard tests may be got. For example, $(\Omega_5, \mathcal{F}_6), (\Omega_5, \mathcal{F}_7)$ and $(\mathbf{R}, \mathcal{F}_{13})$ in Section 2.4 all are nonstandard.

We should like to ask: for any random test $(\Omega, \mathcal{F})$, is there the sample space Ω^* such that $(\Omega^*, \mathcal{F})$ is standard?

The answer is negative. In fact, there is not the conclude like this for the causation space $(X, \mathcal{U})$ (see the footnote 7 and remak (3) in Subsection 2.3.4). In this section we prove that common typical tests all are standard or can be transformed into standard tests.

2.5.2 n-type tests

Definition 2.5.4. *The $(\Omega, \mathcal{F})$ is called n-type test if Ω contains the n elements.*

Theorem 2.5.1. *A n-type test is standard if and only if $\mathcal{F}$ is the set which consists of all subset of Ω.*

Proof : The sufficiency is obvious. Suppose $(\Omega, \mathcal{F})$ is a n-type standard test. Without loss of generality, we may suppose

$$\Omega = \{\, \omega_1,\, \omega_2,\, \cdots,\, \omega_n \,\} \tag{2.5.7}$$

$\{\omega_i\} \in \mathcal{F}$ is obtained by the supposition of standard. Now, supposing A is any subset of Ω, since $A = \bigcup_{\omega \in A} \{\omega\}$ is a finite union, so $A \in \mathcal{F}$. The necessary has been proved. ■

Theorem 2.5.2. *The n-type standard tests are σ-isomorphic each other.*

Proof : Suppose $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ two standard test. We may suppose Ω has the formation (2.5.7), Ω^* has the formation

$$\Omega^* = \{\, \omega_1^*,\, \omega_2^*,\, \cdots,\, \omega_n^* \,\} \tag{2.5.8}$$

Taking any correspondence of an one to one between Ω and Ω^*, for example, taking

$$\varphi: \quad \omega_i \longleftrightarrow \omega_i^* \tag{2.5.9}$$

Let

$$\varphi: \quad \mathcal{F} \ni A \longleftrightarrow \varphi(A) = \{\, \omega_i^* \,|\, \omega_i \in A\} \in \mathcal{F}^* \tag{2.5.10}$$

It is easy to check that φ is a σ-isomorphic mapping from $\mathcal{F}$ to $\mathcal{F}^*$. So $(\Omega, \mathcal{F})$ and $(\Omega^*, \mathcal{F}^*)$ are σ-isomorphic. ■

The n-type standard tests have clear construction:

1) The sample space Ω contains n elements. The general formation is (2.5.7).

2) The event σ-field $\mathcal{F}$ contains 2^n elements. They can be classified as $n+1$ classes. The 0th class only contains one event — impossible event. The ith$(1 \le i \le n)$ class contains the C_n^i events, and each event contains i sample points, they are:

$$\{\omega_1, \omega_2, \cdots, \omega_i\},\ \{\omega_1, \cdots, \omega_{i-1}, \omega_{i+1}\}\ \cdots\ \{\omega_{n-i+1}, \omega_{n-i+2}, \cdots, \omega_n\} \tag{2.5.11}$$

In order to display the construction of n-type nonstandard tests, we need introduce

Definition 2.5.5. *Supposing $(\Omega, \mathcal{F})$ is any test and B is a nonempty event in $\mathcal{F}$. The B is called an atomic event of the $(\Omega, \mathcal{F})$, or an atom* *in brief, if for any $A \in \mathcal{F}$ and $A \subset B$, we have either $A = B$ or $A = \emptyset$. Otherwise B is called non atom.*

Obviously, difference atomic events B_1 and B_2 are disjoint. In fact, if $B_1 \bigcap B_2 \neq \emptyset$, it is from the atom's definition that $B_1 \bigcap B_2 = B_1 = B_2$, that is impossible.

Theorem 2.5.3. *Supposing $(\Omega, \mathcal{F})$ is a n-type test, then there exists a sample space Ω^* such that $(\Omega^*, \mathcal{F})$ is m-type standard, where $m \leq n$.*

Proof : Let us prove that if A is a nonempty event in $\mathcal{F}$, then there is some atomic event B such that $B \subset A$. In fact, if A itself is atomic, the conclude is set up. Otherwise there are nonempty true subsets A_1 and $A \backslash A_1$. If one of both is atomic, then taking it as B and the conclude is set up. Otherwise A_1 is operation similarly. Operating continuatively, when the operations stop, the needed B has been obtained. Since A is a finite set, the operations can end within finite steps.

By $B_1, B_2, \cdots, B_m$ represent all atomic events of $(\Omega, \mathcal{F})$. Let

$$\Omega^* = \{ B_1, \ B_2, \ \cdots, \ B_m \} \tag{2.5.12}$$

Obviously, all $B_i \, (1 \leq i \leq m)$ are nonempty and disjoint each other, so $m \leq n$. We show

$$\Omega = \bigcup_{i=1}^{m} B_i \tag{2.5.13}$$

In fact, if it is not true, then $\Omega \backslash \bigcup_{i=1}^{m} B_i$ is a nonempty event of $\mathcal{F}$. Form conclude first proved we know $\Omega \backslash \bigcup_{i=1}^{m} B_i$ contains an atomic event B. Obviously B is different with all $B_i (1 \leq i \leq m)$. That is impossible, so (2.5.13) is set up.

It is from (2.5.12) and (2.5.13) that Ω^* is a partition of the causal space and is coarser than Ω. For any $A \in \mathcal{F}$, we have

$$A = A \bigcap [\bigcup_{i=1}^{m} B_i] = \bigcup_{i=1}^{m} AB_i \tag{2.5.14}$$

Since B_i is atomic, we deduct either $AB_i = B_i$ or $AB_i = \emptyset$. So we have proved the A is an event on Ω^* and $\mathcal{F}$ is the event σ-field on Ω^*. Therefore $(\Omega^*, \mathcal{F})$ is a m-type standard test. ■

2.5.3 Countable type tests

Definition 2.5.6. *The $(\Omega, \mathcal{F})$ is called countable type test if Ω contain countable elements.*

From the discussion similarly to the n-type tests we have

Theorem 2.5.4. *The countable type test $(\Omega, \mathcal{F})$ is standard if and only if $\mathcal{F}$ is the set which is consists of all subsets of Ω.*

Theorem 2.5.5. *The countable type standard test all are σ-isomorphic.*

Similarly, the countable type standard tests have the clear constructions:

1) The sample space Ω contains the countable elements. We may suppose

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n, \cdots\} \tag{2.5.15}$$

2) The event σ-field $\mathcal{F}$ contains uncountable elements. They can be classified as countable classes:

① The ith$(0 \leq i < \infty)$ class consists of the events containing i sample points. The 0th class only contains one event — impossible one. The ith$(1 \leq i < \infty)$ contains the countable events, they are:

$$\{\omega_1, \cdots, \omega_{i-1}, \omega_i\}, \{\omega_1, \cdots, \omega_{i-1}, \omega_{i+1}\}, \cdots\cdots \tag{2.5.16}$$

② The wth class consists of the events contained countable sample points. It contains uncountable events.

The discussion about countable type nonstandard tests shows as follows. We first prove

Lemma 2.5.2. *If $(\Omega, \mathcal{F})$ is a countable type test, then for any $\omega \in \Omega$ there exists unique atomic event B such that $\omega \in B$.*

Proof : First we prove the uniqueness. Let B_1 and B_2 are two atomic contained ω. Since $B_1 \bigcap B_2 \supset \{\omega\} \neq \emptyset$, it is from $B_1 \bigcap B_2 \subset B_i (i = 1, 2)$ and the atom's definition that $B_1 \bigcap B_2 = B_1 = B_2$.

The existence will be prove as follows. If Ω is atomic, then taking $B = \Omega$ and the lemma's conclude is got. Otherwise, there are nonempty sub-events A_1 and $\Omega \backslash A_1$ in $\mathcal{F}$. One of both $\omega \in A_1$ and $\omega \in \Omega \backslash A_1$ is true. Without loss of generality, suppose $\omega \in A_1$.

If A_1 is atomic, then taking $B = A_1$, the conclude is got. Otherwise, using A_1 instead of Ω, the operations is run, the A_2 is got. Going on such like this, we have got $A_2, A_3, A_4, \cdots$ and $\omega \in A_i$ $(i \geq 2)$. If the operations stop at natural number k, taking $B = A_k$, the conclude has been got. Otherwise, by mathematical induction we have got a sequence of the events

$$\Omega, A_1, A_2, \cdots, A_n, \cdots \tag{2.5.17}$$

Now, let $A_w = \bigcap_{n=1}^{\infty} A_n$. Obviously $\omega \in A_w$. If A_w is atomic, then taking $B = A_w$ and the conclude has been got. Otherwise, by mathematical superinduction the operations continue: if α is an isolate countable ordinal number, then A_α is obtained like the A_1; If α is a limit countable ordinal number, then A_α is obtained like the A_w. So we have got a transfinite sequence

$$\Omega, A_1, A_2, \cdots, A_w, A_{w+1}, \cdots, A_\alpha, A_{\alpha+1}, \cdots \tag{2.5.18}$$

The $\omega \in A_\alpha$ and $A_\alpha \backslash A_{\alpha+1}$ are guaranteed nonempty by the operations. Note that Ω is a countable set and the number of elements in (2.5.18) is uncountable, so there exists a ordinal number β such that the operation stop at the β. That is, A_β is atomic, $\omega \in A_\beta$ and $A_{\beta+1} = \emptyset$. Finally taking $B = A_\beta$ and we have got the conclude. ■

Theorem 2.5.6. *Supposing $(\Omega, \mathcal{F})$ is a countable type test, then there exists a sample space Ω^* such that $(\Omega^*, \mathcal{F})$ is n-type or countable type standard test.*

Proof : Lemma 2.5.2 guarantees that there is at least one atomic event in $\mathcal{F}$. All atomic events are represented by $B_1, B_2, \cdots, B_n, \cdots$. Note that all $B_i\,(i \in I)$ are nonempty and disjoint each other, so the number of $B_i\,(i \in I)$ is finite (may suppose as n) or countable. Let

$$\Omega^* = \{\, B_1,\ B_2,\ \cdots,\ B_n,\ \cdots\} \tag{2.5.19}$$

Similar to the discussions of Theorem 2.5.3, we have got that Ω^* is a sample space and $(\Omega^*, \mathcal{F})$ is a standard test. ■

It is easy to point out that n-type and countable type have similar properties and the research methods. Therefore people call those two type of tests as discrete type of tests.

2.5.4 n-dimensional É. Borel tests

Definition 2.5.7. (1) *$(\mathbf{R}^n, \mathcal{B}^n)$ is called n-dimensional Borel test if $\mathbf{R}^n$ is n-dimensional Euclidean space and $\mathcal{B}^n = \sigma(\mathcal{A}_n)$, where the bud set is*

$$\mathcal{A}_n = \{\, \prod_{i=1}^{n} (a_i, b_i] \,|\, -\infty < a_i < b_i < +\infty,\ i = 1, 2, \cdots, n\} \tag{2.5.20}$$

(2) *$(D, \mathcal{B}^n(D))$ is called the Borel test on D, if $D \in \mathcal{B}^n$ and*

$$\mathcal{B}^n(D) = \{\, B \,|\, B \in \mathcal{B}^n,\ B \subset D\} \tag{2.5.21}$$

Examples 12, 14, 15 in Section 2.4 are 1-, 2-, 3-dimensional Borel tests, respectively. Any one of four kinds of $(\cdots)$, $(\cdots]$, $[\cdots)$, $[\cdots]$ is denoted by $< \cdots >$. Let

$$\mathcal{A}_n^* = \{\, \prod_{i=1}^{n} < a_i, b_i > \,|\, -\infty < a_i < b_i < +\infty,\ i = 1, 2, \cdots, n\} \tag{2.5.22}$$

Theorem 2.5.7. *$(\mathbf{R}^n, \mathcal{B}^n)$ is a standard test, and $\mathcal{B}^n = \sigma(\mathcal{A}_n^*)$.*

Proof : For any $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$, we have

$$\{\, (x_1, x_2, \cdots, x_n)\,\} = \bigcap_{m=1}^{\infty} (\prod_{i=1}^{n} (x_i - \frac{1}{m},\ x_i]\,) \tag{2.5.23}$$

By (2.5.20), we have got $\{(x_1, x_2, \cdots, x_n)\} \in \mathcal{B}^n$ and $(\mathbf{R}^n, \mathcal{B}^n)$ is standard.

It is easy to prove that

$$\prod_{i=1}^{n}(a_i, b_i) = \bigcup_{m=1}^{\infty}(\prod_{i=1}^{n}(a_i, b_i - \frac{1}{m}]) \tag{2.5.24}$$

$$\prod_{i=1}^{n}[a_i, b_i) = \bigcap_{m=1}^{\infty}(\prod_{i=1}^{n}(a_i - \frac{1}{m}, b_i)) \tag{2.5.25}$$

$$\prod_{i=1}^{n}[a_i, b_i] = \bigcap_{m=1}^{\infty}(\prod_{i=1}^{n}(a_i - \frac{1}{m}, b_i]) \tag{2.5.26}$$

Various cylinders of mixed type (for example, $(a_1, b_1) \times (a_2, b_2] \times [a_3, b_3] \times \cdots \times [a_n, b_n)$, and so on) have also similar equalities. That is to say that the events of $\mathcal{A}_n^*$ all can be represented by the complements, unions, intersections and differences of the events in $\mathcal{A}_n$. By use of Corollary of Theorem 2.4.3 we have $\sigma(\mathcal{A}_n^*) \subset \sigma(\mathcal{A}_n) = \mathcal{B}_n$. On the other hand, since $\mathcal{A}_n \subset \mathcal{A}_n^*$, we have $\sigma(\mathcal{A}_n) \subset \sigma(\mathcal{A}_n^*)$. Linking the two containing equations, we have $\mathcal{B}_n = \sigma(\mathcal{A}_n^*)$. ■

This theorem declares $\mathcal{B}^n$ to imply various cylinders (or cylinder events), therefor to imply the areas generated by complements, unions, intersections and differences of these cylinders. In fact, all common areas belong to $\mathcal{B}^n$. It is difficult to find an area which does not belong to $\mathcal{B}^n$. But there really exist this kind of areas.

Theorem 2.5.8. *there exists a subset M of* $\mathbf{R}^n$ such that $M \notin \mathcal{B}^n$.

This conclude is very important for us to understand Borel test. Its proof uses the special knowledge about measure theory[16][17], and is omitted here.

2.5.5 Countable dimensional and any dimensional Borel tests

Suppose $\mathbf{R}$ is a number line, $\mathcal{B}$ is the Borel field on $\mathbf{R}$.

$$\mathbf{R}^\infty = \{(x_1, x_2, \cdots, x_n, \cdots) \mid x_i \in \mathbf{R},\, i = 1, 2, 3, \cdots\} \tag{2.5.27}$$

is called as a countable dimensional Euclidean space. Obviously, for any $A_i \in \mathcal{B}\,(i = 1, 2, 3, \cdots)$,

$$\prod_{i=1}^{\infty} A_i = \{(x_1, x_2, \cdots, x_n, \cdots) \mid x_i \in A_i,\, i = 1, 2, 3, \cdots\} \tag{2.5.28}$$

is a subset on $\mathbf{R}^\infty$. If all A_i, except finite A_i (for example to say, $A_{\alpha_1}, A_{\alpha_2}, \cdots, A_{\alpha_n}$), have $A_i = \mathbf{R}$, then the $\prod_{i=1}^{\infty} A_i$ is called as a n-dimensional cylinder, and written in brief as

$$\prod_{i=1}^{n} A_{\alpha_i} \times \mathbf{R} \tag{2.5.29}$$

The set consisted of all cylinders is denoted by $\mathcal{A}_\infty$, i.e.

$$\mathcal{A}_\infty = \{ \prod_{i=1}^{n} A_{\alpha_i} \times \mathbf{R} \,|\, n \geq 1,\ \alpha_i \text{ is a natural number },$$
$$A_{\alpha_i} \in \mathcal{B},\ i = 1, 2, \cdots, n \} \tag{2.5.30}$$

Definition 2.5.8. *The $(\mathbf{R}^\infty, \mathcal{B}^\infty)$ is called a countable dimensional Borel test if $\mathbf{R}^\infty$ is a countable dimensional Euclidean space and $\mathcal{B}^\infty = \sigma(\mathcal{A}_\infty)$.*

Theorem 2.5.9. *$(\mathbf{R}^\infty, \mathcal{B}^\infty)$ is a standard test.*

Proof : For any $(x_1, x_2, \cdots, x_n, \cdots) \in \mathbf{R}^n$, we have

$$\{ (x_1, x_2, \cdots, x_n, \cdots) \} = \bigcap_{n=1}^{\infty} \bigcap_{m=1}^{\infty} (\prod_{i=1}^{n} (x_i - \frac{1}{m}, x_i] \times \mathbf{R}) \tag{2.5.31}$$

and $\{(x_1, x_2, \cdots, x_n, \cdots)\} \in \mathcal{B}^\infty$. The proof is complete. ■

It is unforeseen simple to extend countable dimensional Borel tests to any dimensional ones. We denote any index set by T. Let

$$\mathbf{R}^T = \{ x(t), t \in T | \text{ for any fixed } t, \text{ we have } x(t) \in \mathbf{R}\} \tag{2.5.32}$$

Obviously, $\mathbf{R}^T$ consists of all real value functions with domain T and is called T-dimensional Euclidean space. For any $A_t \in \mathcal{B}\,(t \in T)$, let

$$\prod_{t \in T} A_t = \{ x(t), t \in T | \text{ for any fixed } t \in T, \text{ we have } x(t) \in A_t\} \tag{2.5.33}$$

It is a subset on $\mathbf{R}^T$. If all A_t, except finite A_t (for example to say, A_{t_1}, A_{t_2}, $\cdots$, A_{t_n}), have $A_t = \mathbf{R}$, then the $\prod_{t \in T} A_t$ is called as a n-dimensional cylinder, and written in brief as

$$\prod_{i=1}^{n} A_{t_i} \times \mathbf{R} \tag{2.5.34}$$

The set consisted of all cylinders is denoted by $\mathcal{A}_T$, i.e.

$$\mathcal{A}_T = \{ \prod_{i=1}^{n} A_{t_i} \times \mathbf{R} \,|\, n \geq 1, t_i \in T, A_{t_i} \in \mathcal{B},\ i = 1, 2, 3, \cdots, n\} \tag{2.5.35}$$

Definition 2.5.9. *The* $(\mathbf{R}^T, \mathcal{B}^T)$ *is called a* T*-dimensional Borel test if* $\mathbf{R}^T$ *is a* T*-dimensional Euclidean space and* $\mathcal{B}^T = \sigma(\mathcal{A}_T)$.

Now, it is different to the countable dimensional Borel test that the $(\mathbf{R}^T, \mathcal{B}^T)$ is nonstandard when T is an uncountable index set.

2.5.6 Remark

The standard character of the random tests has important meaning. In the Kolmogorov's axiom system rose similar question: giving a probability space $(\Omega, \mathcal{F}, P)$, is there a σ-isomorphic probability space $(\Omega^*, \mathcal{F}^*, P^*)$ such that for any $\omega^* \in \Omega^*$, we have $\{\omega^*\} \in \mathcal{F}^*$?

This question has got a positive answer for discrete probability space[18]. For general probability spaces, if the essential isomorph instead of the σ-isomorph, then the positive answer can be obtain also[19]. The another way to standardize probability spaces will be given in Subsection 2.7.10.

2.6 Joint random tests

In the causation space there is a great number of random tests. Those tests are not isolation. There are loose or close relation among them. The joint random tests are important tool to study this kind of the relations.

2.6.1 Joint event σ-fields

Lemma 2.6.1. *Suppose* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are two event* σ*-fields. Let*

$$\mathcal{F}_1 * \mathcal{F}_2 = \{ A \bigcap B \mid A \in \mathcal{F}_1,\ B \in \mathcal{F}_2 \} \tag{2.6.1}$$

Then we have

(1) $\mathcal{F}_i \subset \mathcal{F}_1 * \mathcal{F}_2$ $(i = 1, 2)$;

(2) $\mathcal{F}_1 * \mathcal{F}_2$ *is a* π*-class. i.e. If* $C, D \in \mathcal{F}_1 * \mathcal{F}_2$, *then* $C \bigcap D \in \mathcal{F}_1 * \mathcal{F}_2$;

(3) $\sigma(\mathcal{F}_1 * \mathcal{F}_2) = \sigma(\mathcal{F}_1 \bigcup \mathcal{F}_2)$.

This proof is not difficult and is left to readers.

Hereafter $\sigma(\mathcal{F}_1 * \mathcal{F}_2)$ is written as $\mathcal{F}_1 \otimes \mathcal{F}_2$ in brief. The lemma declares that $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the smallest event σ-field containing $\mathcal{F}_1$ and $\mathcal{F}_2$.

Definition 2.6.1. *The* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *is called the joint event* σ*-field of* $\mathcal{F}_1$ *and* $\mathcal{F}_2$. *The* $\mathcal{F}_1 * \mathcal{F}_2$ *is called the cylinder event set of* $\mathcal{F}_1$ *and* $\mathcal{F}_2$, *its element* $A \bigcap B$ *is called a cylinder event, and* A *and* B *are called sides of cylinder event.*

Joint event σ-field has established a kind of intuitive figure: in the causation space there exist relations among all events. "The relations" is represented by complements, unions, intersections and differences in the first group of axioms. So the relations between $\mathcal{F}_1$ and $\mathcal{F}_2$ is represented by complements, unions, intersections and differences of events in $\mathcal{F}_1$ and $\mathcal{F}_2$. Because $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the smallest σ-field containing those complements, unions, intersections and differences, $\mathcal{F}_1 \otimes \mathcal{F}_2$ becomes a convenient place to study the relations between $\mathcal{F}_1$ and $\mathcal{F}_2$.

2.6.2 Joint random tests

This paragraph demonstrates that the most convenient place to study the relations between $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ is their joint random test.

Theorem 2.6.1. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two random tests. Then $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is a random test, and $(\Omega_i, \mathcal{F}_i)(i = 1, 2)$ are its sub-tests, where $\Omega_1 \odot \Omega_2$ is the cross partition of Ω_1 and Ω_2, the $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the joint event σ-field of $\mathcal{F}_1$ and $\mathcal{F}_2$.*

Proof : It is only need to prove $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is a random test by Lemma 2.6.1. Without loss of generality, let

$$\Omega_1 = \{\,\omega_\alpha^{(1)} \,|\, \alpha \in J_1\}\,; \qquad \Omega_2 = \{\,\omega_\beta^{(2)} \,|\, \beta \in J_2\} \tag{2.6.2}$$

where J_1 and J_2 are two index sets. From Definition 2.3.5, we have

$$\Omega_1 \odot \Omega_2 = \{\,\omega_\alpha^{(1)} \bigcap \omega_\beta^{(2)} \,|\, \alpha \in J_1,\, \beta \in J_2,\, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \emptyset\,\} \tag{2.6.3}$$

Taking any $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, because A and B are the events on Ω_1 and on Ω_2 respectively, without loss of generality we may suppose

$$A = \{\,\omega_\alpha^{(1)} \,|\, \alpha \in J_{1A}\}; \qquad B = \{\,\omega_\beta^{(2)} \,|\, \beta \in J_{2B}\} \tag{2.6.4}$$

where J_{1A} and J_{2B} are the subsets of J_1 and J_2, respectively. The intersection of A and B in the causal space is

$$A \bigcap B = \{\,\omega_\alpha^{(1)} \bigcap \omega_\beta^{(2)} \,|\, \alpha \in J_{1A},\, \beta \in J_{2B},\, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \neq \emptyset\,\} \tag{2.6.5}$$

From this equation and (2.6.3) we get $A \bigcap B$ is an event on $\Omega_1 \odot \Omega_2$. So $\sigma(\mathcal{F}_1 * \mathcal{F}_2)$ is an event σ-field on the $\Omega_1 \odot \Omega_2$ and $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is a random test. ■

Definition 2.6.2. *The $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is called a joint random test of $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, or a joint test in brief.*

Obviously, if $(\Omega_2, \mathcal{F}_2)$ is a sub-test of $(\Omega_1, \mathcal{F}_1)$, then their joint test is $(\Omega_1 \odot \Omega_2, \mathcal{F}_1)$; If the sub-test is the same causes, then the joint test is $(\Omega_1, \mathcal{F}_1)$.

Example 1. *To find the joint test of $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_3, \mathcal{F}_3)$ in Section 2.4.*

Solution: The joint test to find is the $(\Omega_1 \odot \Omega_3, \mathcal{F}_1 \otimes \mathcal{F}_3)$, where

$$\Omega_1 \odot \Omega_3 = \{H \bigcap male,\ H \bigcap female,\ T \bigcap male,\ T \bigcap female\} \qquad (2.6.6)$$

$$\mathcal{F}_1 \otimes \mathcal{F}_3 = \text{the set which consists of all subsets of } \Omega_1 \odot \Omega_3 \qquad (2.6.7)$$

■

In the reality the test of throw a coin and the test of bearing baby are two no relative tests. But their joint test still has the realistic meaning. For example, in common life we sometimes meet the situation as follows. Some father who guesses the sex of the baby to be born says: " I throw a coin. If the head occurs, then the baby is a boy. Otherwise, the baby is a girl ". Would we like to ask how do describe the father's action? The NAS maintains that the father's action is to do an experiment which generates the joint test in the example above. But what the father really concerns is the sub-test $(\Omega_1 \odot \Omega_3,\ \mathcal{F}^*)$, where $\mathcal{F}^* = \{\emptyset,\ A,\ \overline{A},\ \Omega_1 \odot \Omega_3\}$. The events

$$A = \{H \cap male,\ T \cap female\}; \quad \overline{A} = \{H \cap female,\ T \cap male\} \qquad (2.6.8)$$

denotes the events " the father guesses right " and " the father guesses wrong ", respectively.

Example 2. *Verify that $(\Omega_5, \mathcal{F}_5)$ is the joint test of $(\Omega_5, \mathcal{F}_5)$ and $(\Omega_6, \mathcal{F}_6)$ in Section 2.4.*

Solution: From (2.4.23) we deduct that the elements of $\Omega_5 * \Omega_6$ are

$$\begin{cases} i \cap \omega_o = i, \quad i \cap \omega_e = \emptyset, & i = 1, 3, 5 \\ i \cap \omega_o = \emptyset, \quad i \cap \omega_e = i, & i = 2, 4, 6 \end{cases} \qquad (2.6.9)$$

We get $\Omega_5 \odot \Omega_6 = \Omega_5$ after deleting the empty. Obviously $\mathcal{F}_5 \otimes \mathcal{F}_6 = \mathcal{F}_5$. So the test to find is the $(\Omega_5, \mathcal{F}_5)$. ■

Example 3. *Should like to find the joint test of $(\Omega_6, \mathcal{F}_6)$ and $(\Omega_7, \mathcal{F}_7)$ in Section 2.4.*

Solution: The finding test is $(\Omega_6 \odot \Omega_7, \mathcal{F}_6 \otimes \mathcal{F}_7)$, where

$$\Omega_6 \odot \Omega_7 = \{\omega_o \cap \omega_l,\ \omega_o \cap \omega_s,\ \omega_e \cap \omega_l,\ \omega_e \cap \omega_s\} \qquad (2.6.10)$$

$$\mathcal{F}_6 \otimes \mathcal{F}_7 = \text{the set which consists of all subsets of } \Omega_6 \odot \Omega_7 \qquad (2.6.11)$$

■

Note: If we maintain $(\Omega_5, \mathcal{F}_6)$ and $(\Omega_5, \mathcal{F}_7)$ are the random tests generated by Examples 6 and 7 in Section 2.4, respectively, then their joint test is $(\Omega_5, \mathcal{F}_6 \otimes \mathcal{F}_7)$. In this test we have

$$\begin{cases} \omega_o \cap \omega_l = \{5\}; & \omega_o \cap \omega_s = \{1,3\}; \\ \omega_e \cap \omega_l = \{6\}; & \omega_e \cap \omega_s = \{2,4\} \end{cases} \tag{2.6.12}$$

Example 4. *Repeating twice the experiment $\mathcal{E}_1$ of Example 1 in Section 2.4, the new experiment $\mathcal{E}^*$ is got. Should like to find the random test generated by $\mathcal{E}^*$.*

Solution: In the experiment $\mathcal{E}^*$, the sentence like " the head occurs" is ambiguous, and can not be used to describe an event. In fact, there are at lest four kinds of understanding: ① the head occurs in the first throw; ② the head occurs in the secondly throw; ③ at least one head occurs in twice throws; ④ the head occurs twice in twice throws

In order to delete the ambiguity, note that throw first the coin and throw secondly the coin are two different experiments, we regard i time throw in the experiment $\mathcal{E}^*$ as the experiment $\mathcal{E}_{1i}$ which generates the test $(\Omega_{1i}, \mathcal{F}_{1i})$ $(i = 1, 2)$, where

$$\Omega_{1i} = \{ H_i, T_i \} \tag{2.6.13}$$

$$\mathcal{F}_{1i} = \{ \emptyset, < H_i >, < T_i >, \Omega_{1i} \} \tag{2.6.14}$$

By the new symbols we can exactly represent the random test generated by $\mathcal{E}^*$, it is the joint test $(\Omega_{11} \odot \Omega_{12}, \mathcal{F}_{11} \otimes \mathcal{F}_{12})$, where

$$\Omega_{11} \odot \Omega_{12} = \{ (H_1, H_2), (H_1, T_2), (T_1, H_2), (T_1, T_2) \} \tag{2.6.15}$$

$$\mathcal{F}_{11} \otimes \mathcal{F}_{12} = \text{the set which consists of all subsets of } \Omega_{11} \odot \Omega_{12} \tag{2.6.16}$$

■

2.6.3 Product random tests

The product tests are important special situation of the joint tests. It is used widely in probability theory.

Definition 2.6.3. *If $\Omega_1 \odot \Omega_2$ is a true cross partition, then the joint test $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is called as product test.*

Lemma 2.6.2. *Suppose $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is the product test of $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$. We have*

(1) If Ω_1 (or Ω_2) is not the coarsest partition, then for any $\omega_\alpha^{(1)} \in \Omega_1$, $\omega_\beta^{(2)} \in \Omega_2$, the set $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ is a true subset of $\omega_\beta^{(2)}$ (or $\omega_\alpha^{(1)}$);

(2) The cylinder event $A \bigcap B = \emptyset$ if and only if at least one of $A = \emptyset$ and $B = \emptyset$ is set up;

(3) Nonempty cylinder event $A\bigcap B \subset C\bigcap D$ if and only if $A \subset C$ and $B \subset D$;

(4) Nonempty cylinder event $A\bigcap B = C\bigcap D$ if and only if $A = C$ and $B = D$;

(5) $\mathcal{F}_1\bigcap\mathcal{F}_2 = \{\emptyset, X\}$;

(6) If $A \in \mathcal{F}_1, B \in \mathcal{F}_2$ and A and B are not $\emptyset$ or X, then the cylinder event $A\bigcap B$ belongs neither $\mathcal{F}_1$ nor $\mathcal{F}_2$, i.e.

$$A\bigcap B \notin \mathcal{F}_1; \qquad A\bigcap B \notin \mathcal{F}_2 \tag{2.6.17}$$

Proof : Without loss of generality, we may suppose Ω_1 and Ω_2 have expressions (2.6.2). The conditions of the lemma imply

$$\Omega_1 \odot \Omega_2 = \{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \mid \alpha \in J_1, \beta \in J_2\} \tag{2.6.18}$$

(1) If Ω_1 is not the coarsest partition, then $J_1\backslash\{\alpha\}$ is nonempty. By (2.6.18) we deduct that $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ and $\{\omega_\gamma^{(1)} \cap \omega_\beta^{(2)} | \gamma \in J_1\backslash\{\alpha\}\}$ are disjoint nonempty subsets of the causal space. Obviously, they are subsets of $\omega_\beta^{(2)}$, so we have proved that $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ is a true subset of $\omega_\beta^{(2)}$.

(2) The sufficiency is obvious. If $A \neq \emptyset$ and $B \neq \emptyset$, then there is a sample point $\omega_\alpha^{(1)} \in A$ in Ω_1 and a sample point $\omega_\beta^{(2)} \in B$ in Ω_2. So we get $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \in A\bigcap B$. Since $\Omega_1 \odot \Omega_1$ is a true cross partition, we get $A\bigcap B$ being nonempty, and the necessity is proved.

(3) The sufficiency is obvious. Let the $A\bigcap B \subset C\bigcap D$ is true. If supposing at least one of both $A \subset C$ and $B \subset D$ does not set up, we may suppose $A \subset C$ does not set up. Since $A\bigcap B$ is nonempty, we get $A \neq \emptyset$. So there is some sample point $\omega_\alpha^{(1)}$ such that $\omega_\alpha^{(1)} \in A$ and $\omega_\alpha^{(1)} \notin C$.

By same reason, $B \neq \emptyset$. So there is some sample point $\omega_\beta^{(2)}$ such that $\omega_\beta^{(2)} \in B$. It follows that

$$\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \in A\bigcap B; \qquad \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \notin C\bigcap D$$

Note that $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ is nonempty, we have $A\bigcap B \subset C\bigcap D$ does not set up. There is contradictory with the supposition. Thus the $A \subset C$ do set up and the necessity is proved.

(4) The conclude is deducted by (3).

(5) It is easy to verify $\mathcal{F}_1\bigcap\mathcal{F}_2$ is an event σ- field. By Theory 2.2.5 we have $\{\emptyset, X\} \subset \mathcal{F}_1\bigcap\mathcal{F}_2$. The proof of $\mathcal{F}_1\bigcap\mathcal{F}_2 = \{\emptyset, X\}$ is below.

Supposing $\mathcal{F}_1\bigcap\mathcal{F}_2 \neq \{\emptyset, X\}$, then there is an event A such that

$$A \in \mathcal{F}_1\bigcap\mathcal{F}_2, \quad A \neq \emptyset, \quad A \neq X$$

From this we deduct

$$\overline{A} \in \mathcal{F}_1 \bigcap \mathcal{F}_2, \quad \overline{A} \neq \emptyset, \quad \overline{A} \neq X$$

Now, by $A \in \mathcal{F}_1$ we know that there is a $\omega_\alpha^{(1)}$ in Ω_1 such that $\omega_\alpha^{(1)} \in A$; by $\overline{A} \in \mathcal{F}_2$ we know that there is a $\omega_\beta^{(2)}$ in Ω_2 such that $\omega_\beta^{(2)} \in \overline{A}$. So we have

$$\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \in A \bigcap \overline{A} = \emptyset$$

This equation is contradictory with that $\Omega_1 \odot \Omega_2$ is a true cross partition. Therefore the conclude is prove.

(6) Without loss of generality, suppose A and B have expressions (2.6.4). Because A and B are not $\emptyset$ or X, so J_{1A} and J_{2B} are two nonempty true subsets of J_1 and J_2, respectively.

Suppose $\omega_\alpha^{(1)} \in A$ is a sample point in Ω_1. By the proved (1) we get that two sets of the causal points

$$\{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \,|\, \beta \in J_{2B}\} \quad and \quad \{\omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \,|\, \beta \in J_2 \backslash J_{2B}\}$$

are nonempty true subsects of $\omega_\alpha^{(1)}$. Obviously, the former belong to $A \bigcap B$, the latter does not belong to $A \bigcap B$. Therefore $A \bigcap B$ is not an event on Ω_1 and $A \bigcap B \notin \mathcal{F}_1$ has been proved. By the same reason, we have $A \bigcap B \notin \mathcal{F}_2$. ■

Let us recall the knowledge of product measurable space in the measure theory[16]. Suppose $(Y, \mathcal{S})$ and $(Z, \mathcal{T})$ are two measurable spaces. We call $(y, z)\,(y \in Y, z \in Z)$ as an order pair. The set which consists of all order pairs is written as $Y \times Z$, i.e.

$$Y \times Z = \{(y, z) \,|\, y \in Y,\ z \in Z\} \tag{2.6.19}$$

$Y \times Z$ is called Descartes product space of Y and Z.

Let $A \in \mathcal{S}$, $B \in \mathcal{T}$,

$$A \times B = \{(y, z) \,|\, y \in A,\ z \in B\} \tag{2.6.20}$$

is called a rectangle, and A and B are called its sides. We denote the set consisted of all the rectangle by $\mathcal{C}$, i.e.

$$\mathcal{C} = \{A \times B \,|\, A \in \mathcal{S},\ B \in \mathcal{T}\} \tag{2.6.21}$$

The $\mathcal{C}$ is called as the rectangle set generated by $\mathcal{S}$ and $\mathcal{T}$. the smallest σ-field containing $\mathcal{C}$ is denoted by $\mathcal{S} \times \mathcal{T}$. i.e.

$$\mathcal{S} \times \mathcal{T} = \sigma(\mathcal{C}) \tag{2.6.22}$$

It is called the product σ-field of $\mathcal{S}$ and $\mathcal{T}$.

The pair $(Y \times Z, \mathcal{S} \times \mathcal{T})$ is called product measurable space of $(Y, \mathcal{S})$ and $(Z, \mathcal{T})$.

Theorem 2.6.2. *Suppose $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two random tests and $\Omega_1 \odot \Omega_2$ is a true cross partition. If the sample point $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ in $\Omega_1 \odot \Omega_2$ is denoted by $(\omega_\alpha, \omega_\beta)$, then the product test $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ becomes $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$.*

Proof: Let the φ denotes the operation which rewrite the sample point $\omega_\alpha^{(1)} \cap \omega_\beta^{(2)}$ in $\Omega_1 \odot \Omega_2$ as $(\omega_\alpha, \omega_\beta)$. i.e.

$$\varphi(\omega_\alpha^{(1)} \bigcap \omega_\beta^{(2)}) = (\,\omega_\alpha,\ \omega_\beta\,) \tag{2.6.23}$$

Comparing (2.6.18) and (2.6.19), we get that φ is an one to one mapping from $\Omega_1 \odot \Omega_2$ to $\Omega_1 \times \Omega_2$. For any $D \in \mathcal{F}_1 \otimes \mathcal{F}_2$, its range under the mapping φ is

$$\varphi(D) = \{\,(\omega_\alpha, \omega_\beta) \,|\, \omega_\alpha^{(1)} \cap \omega_\beta^{(2)} \in D\} \tag{2.6.24}$$

If we can prove that φ is a σ-isomorph between $\mathcal{F}_1 \otimes \mathcal{F}_2$ and $\mathcal{F}_1 \times \mathcal{F}_2$, then the theorem's conclude is proved.

Comparing (2.6.1) and (2.6.21), we get that φ is an one to one correspondence between $\mathcal{F}_1 * \mathcal{F}_2$ and $\mathcal{C}$. And satisfy

(1) For any $A_i \bigcap B_i \in \mathcal{F}_1 * \mathcal{F}_2\,(i \in I)$, we have

$$\varphi(\bigcup_{i\in I}(A_i \bigcap B_i)) = \bigcup_{i\in I}(A_i \times B_i) \tag{2.6.25}$$

(2) For any $A \bigcap B \in \mathcal{F}_1 * \mathcal{F}_2$, we have

$$\varphi(\overline{A \bigcap B}) = \overline{A \times B} \tag{2.6.26}$$

In fact, (2.6.25) is obviously. Using (2.6.25), we have

$$\begin{aligned}\varphi(\overline{A \bigcap B}) &= \varphi\left[(A \bigcap \overline{B}) \bigcup (\overline{A} \bigcap B) \bigcup (\overline{A} \bigcap \overline{B})\right] \\ &= (A \times \overline{B}) \bigcup (\overline{A} \times B) \bigcup (\overline{A} \times \overline{B}) = \overline{A \times B}\end{aligned}$$

So (2.6.26) is proved.

Now, it is from (2.6.25) and (2.6.26) that the ranges of complements, unions, intersections and differences of the events in $\mathcal{F}_1 * \mathcal{F}_2$ under the mapping φ are complements, unions, intersections and differences of the corresponding rectangles, respectively. So the φ defined by (2.6.24) is an one to one correspondence from $(\mathcal{F}_1 * \mathcal{F}_2)^{**}$ to $\mathcal{C}^{**}$ and satisfy (2.6.25) and

(2.6.26), where "**" is the operation in Theorem 2.4.3. Note that $\mathcal{F}_1 \otimes \mathcal{F}_2$ and $\mathcal{F}_1 \times \mathcal{F}_2$ are constructed by the operation "**", using the mathematical superinduction we get φ is a σ-isomorph between $\mathcal{F}_1 \otimes \mathcal{F}_2$ and $\mathcal{F}_1 \times \mathcal{F}_2$. ■

Form now on the product test of $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ is denoted by $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$. New symbol will bring the convenience in writing. For example, when the solution of Example 4 is given, we may simply say, the random test generated by $\mathcal{E}^*$ is $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$. The trouble of adding the below index "i" is omitted. Note that the below index must be added in the original solution. Otherwise, the answer is $(\Omega_1 \odot \Omega_1, \mathcal{F}_1 \otimes \mathcal{F}_1)$. According to the rules of the operations "$\odot$" and "$\otimes$", it equals to $(\Omega_1, \mathcal{F}_1)$. So it is not the correct answer of the question.

2.6.4 Duplicate random tests

Definition 2.6.4. *We call $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$ as duplicate test and it is written $(\Omega^2, \mathcal{F}^2)$ in brief.*

Theorem 2.6.3. *Suppose that the experiment $\mathcal{E}$ generates standard test $(\Omega, \mathcal{F})$. If the experiment $\mathcal{E}$ is repeated twice, then we obtain the experiment $\mathcal{E}^*$ and the standard test $(\Omega^2, \mathcal{F}^2)$ generated by $\mathcal{E}^*$.*

Proof: Without loss of generality, we can suppose

$$\Omega = \{\, \omega_\alpha \mid \alpha \in J \,\} \tag{2.6.27}$$

where J is an index set. Now, we add the index "i" to all products generated by the experiment $\mathcal{E}$. The things with index "i" ($i = 1$ or 2) represent as products generated by i-th time performance of experiment $\mathcal{E}$. Particularly, $(\Omega_i, \mathcal{F}_i)$ is the random test generated by i-time performance of experiment $\mathcal{E}$. Therefore the experiment $\mathcal{E}^*$ generated a test $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

By the theorem 2.6.2, in order to complete the proof of the theorem, it is needed only that $\Omega_1 \odot \Omega_2$ is a true cross partition. Obviously

$$\Omega_1 \odot \Omega_2 = \{\, \omega_{\alpha 1} \bigcap \omega_{\beta 2} \mid \alpha, \beta \in J,\ \omega_{\alpha 1} \cap \omega_{\beta 2} \neq \emptyset \,\} \tag{2.6.28}$$

Form the supposition of standard we get $\{\omega_{\alpha 1}\} \in \mathcal{F}_1$ and $\{\omega_{\beta 2}\} \in \mathcal{F}_2$. So $\omega_{\alpha 1} \bigcap \omega_{\beta 2} \in \mathcal{F}_1 \otimes \mathcal{F}_2$ represents the event "the element event $\omega_{\alpha 1}$ occurs in the first performance of experiment $\mathcal{E}$ and the element event $\omega_{\beta 2}$ occurs in the second performance of experiment $\mathcal{E}$". This is the event possible to occur. So $\omega_{\alpha 1} \bigcap \omega_{\beta 2}$ is not empty event and $\Omega_1 \odot \Omega_2$ is a true cross partition. The proof is complete. ■

Example 5. *Two dimension Borel test $(\mathbf{R}^2, \mathcal{B}^2)$ is the duplicate test of the one dimension Borel test $(\mathbf{R}, \mathcal{B})$.*

2.6.5 n-dimensional joint tests

It is easy that two dimensional joint tests are extended to n-dimensional joint tests. Suppose $(\Omega_i, \mathcal{F}_i)(i = 1, 2, \cdots, n)$ are random tests. Without loss of generality, suppose

$$\Omega_i = \{\, \omega_{i_\alpha}^{(i)} \,|\, i_\alpha \in J_i \,\} \quad (\, i = 1, 2, \cdots, n\,) \tag{2.6.29}$$

where J_i is the index set. We easy prove that

$$\Omega_1 * \Omega_2 * \cdots * \Omega_n = \{\, \omega_{1_\alpha}^{(1)} \bigcap \omega_{2_\alpha}^{(2)} \bigcap \cdots \bigcap \omega_{n_\alpha}^{(n)} \,|\, i_\alpha \in J_i,\ i = 1, 2, \cdot, n\,\} \tag{2.6.30}$$

is a broad partition. After deleting the empty set, it is written as $\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n$, and is called the cross partition of Ω_i $(i = 1, 2, \cdots, n)$. Particularly, when $\Omega_1 * \Omega_2 * \cdots * \Omega_n$ does not contain empty set, $\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n$ is called a true cross partition.

Taking any $A_i \in \mathcal{F}_i$ $(i = 1, 2, \cdots, n)$, obviously

$$A_1 \bigcap A_2 \bigcap \cdots \bigcap A_n = \{\, \omega_{1_\alpha}^{(1)} \bigcap \omega_{2_\alpha}^{(2)} \bigcap \cdots \bigcap \omega_{n_\alpha}^{(n)} \,|\, \omega_{i_\alpha}^{(i)} \in A_i,\ i = 1, 2, \cdot, n\,\} \tag{2.6.31}$$

is an event on $\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n$. We called it as a n-dimensional cylinder event, and $A_i (1 \le i \le n)$ is called as its side. The set which consists of all cylinder events is denoted by $\mathcal{F}_1 * \mathcal{F}_2 * \cdots * \mathcal{F}_n$, i.e.

$$\mathcal{F}_1 * \mathcal{F}_2 * \cdots * \mathcal{F}_n = \{A_1 \bigcap A_2 \bigcap \cdots \bigcap A_n | A_i \in \mathcal{F}_i, i = 1, 2, \cdots, n\} \tag{2.6.32}$$

It is called the cylinder event set generated by $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$. The smallest σ-field containing all cylinder events is denoted by $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n$. i.e.

$$\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\mathcal{F}_1 * \mathcal{F}_2 * \cdots * \mathcal{F}_n) \tag{2.6.33}$$

Obviously, $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n$ is an event σ-field on the sample space $\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n$, and also is the smallest event σ-field containing all $\mathcal{F}_i$ $(i = 1, 2, \cdots, n)$. We call it as the joint event σ-field of $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$.

It is not difficult to prove that the discussions of two dimensional situation is adaptive for n-dimensional.

Definition 2.6.5. $(\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n,\ \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n)$ *is called n-dimensional joint random test of all tests* $(\Omega_i, \mathcal{F}_i)(i = 1, 2, \cdots, n)$, *and called as the joint test in brief.*

Two important special situations are:

(1) If $\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n$ is a true cross partition, then the joint test is called as the n-dimensional product test. At present Theorem 2.6.2 may be extended to n-dimensional situation. Therefore we represent the product

tests by n-dimensional product measurable spaces $(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n,\ \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n)$ in the measure theory.

(2) $(\underbrace{\Omega \times \Omega \times \cdots \times \Omega}_{n\ times},\ \underbrace{\mathcal{F} \times \mathcal{F} \times \cdots \times \mathcal{F}}_{n\ times})$ is called as n-duplicate test of $(\Omega, \mathcal{F})$, and written as $(\Omega^n, \mathcal{F}^n)$ in brief. At present Theorem 2.6.3 can be extended to n-duplicate situation.

Example 6. *n-dimension Borel test $(\mathbf{R}^n, \mathcal{B}^n)$ is the n-duplicate test of one dimension Borel test.*

2.6.6 Countable dimensional and any dimensional joint tests

We represent an infinite index set by T. The T may be a countable set or an uncountable set. Suppose $(\Omega_t, \mathcal{F}_t)(t \in T)$ are a class of random tests. Without loss of generality, suppose

$$\Omega_t = \{\, \omega_{t(\alpha)}^{(t)} \,|\, t(\alpha) \in J_t\} \quad (t \in T) \tag{2.6.34}$$

where J_t is the index set. We easy prove that

$$*_{t\in T}\Omega_t = \{\, \bigcap_{t\in T} \omega_{t(\alpha)}^{(t)} \,|\, t(\alpha) \in J_t,\ t \in T\,\} \tag{2.6.35}$$

is a broad partition of the causal space. After deleting the empty set, it is written as $\odot_{t\in T}\Omega_t$, and is called the cross partition of $\Omega_t\, (t \in T)$. Particularly, when $*_{t\in T}\Omega_t$ does not contain empty set, $\odot_{t\in T}\Omega_t$ is called a true cross partition.

Taking any $A_t \in \mathcal{F}_t\, (t \in T)$, if except finite (for example to say, $A_{t_1}, A_{t_2}, \cdots, A_{t_n}$) we have $A_t = \Omega_t\, (t \neq t_1, t_2, \cdots, t_n)$, then the event

$$\bigcap_{t\in T} A_t \overset{def}{=} \bigcap_{i=1}^{n} A_{t_i} = \{\, \bigcap_{t\in T} \omega_{t(\alpha)}^{(t)} \,|\, \omega_{t(\alpha)}^{(t)} \in A_t,\ t \in T\,\} \tag{2.6.36}$$

is called as a n-dimensional cylinder event, and $A_{t_i} (i = 1, 2, \cdots, n)$ is called as its sides. The n-dimension cylinder event is written in brief as

$$\bigcap_{i=1}^{n} A_{t_i} \bigcap \Omega_t \tag{2.6.37}$$

Obviously it is an event on the sample space $\odot_{t\in T}\Omega_t$. The set which consists of all cylinder events is denoted by $*_{t\in T}\mathcal{F}_t$, i.e.

$$*_{t\in T}\mathcal{F}_t = \{\bigcap_{i=1}^{n} A_{t_i} \bigcap \Omega_t \,|\, n \geq 1, A_{t_i} \in \mathcal{F}_{t_i}, i = 1, 2, \cdots, n\} \tag{2.6.38}$$

It is called the cylinder event set generated by all $\mathcal{F}_t\,(t \in T)$. The smallest σ-field containing all cylinder events is denoted by $\otimes_{t\in T}\mathcal{F}_t$, i.e.

$$\otimes_{t\in T}\mathcal{F}_t = \sigma(*_{t\in T}\mathcal{F}_t) \tag{2.6.39}$$

Obviously, it is an event σ-field on the sample space $\odot_{t\in T}\Omega_t$, and also is the smallest event σ-field containing all $\mathcal{F}_t\,(t \in T)$. We call it as the joint event σ-field of all $\mathcal{F}_t\,(t \in T)$.

Similarly, the discussions of two dimensional situation can be extended to the T-dimensional situation.

Definition 2.6.6. *$(\odot_{t\in T}\Omega_t, \otimes_{t\in T}\mathcal{F}_t)$ is called T-dimensional joint random test of all tests $(\Omega_t, \mathcal{F}_t)(t \in T)$, and called as the joint test in brief.*

Two important special situations are:

(1) If $\odot_{t\in T}\Omega_t$ is a true cross partition, then the joint test is called as the T-dimensional product test. At that time Theorem 2.6.2 may be extended to T-dimensional situation. Therefore we represent the product test by T-dimensional product measurable spaces $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t)$ in the measure theory.

Therefor, the T-dimensional product test provides a kind of reality background materials to the abstract infinite dimensional measurable space.

(2) $(\prod_{t\in T}\Omega, \prod_{t\in T}\mathcal{F})$ is called as T-duplicate test of $(\Omega, \mathcal{F})$, and written as $(\Omega^T, \mathcal{F}^T)$ in brief. At present Theorem 2.6.3 can only be extended to countable duplicate situation.

Example 7. *T-dimension Borel test $(\mathbf{R}^T, \mathcal{B}^T)$ is the T-duplicate test of one dimension Borel test.*

Example 8. Let $(\Omega, \mathcal{F})$ is Bernoulli test (see Subsection 2.4.3). i.e.

$$\Omega = \{A, \overline{A}\} \tag{2.6.40}$$

$$\mathcal{F} = \{\emptyset, A, \overline{A}, \Omega\} \tag{2.6.41}$$

Then countable duplicate Bernoulli test is $(\Omega^\infty, \mathcal{F}^\infty)$, where

$$\Omega^\infty = \{(\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots) \,|\, \alpha_i = A \text{ or } \overline{A},\ i = 1, 2, 3, \cdots\} \tag{2.6.42}$$

$$\mathcal{F}^\infty = \sigma(\mathcal{A}) \tag{2.6.43}$$

$$\mathcal{A} = \{\prod_{i=1}^{n} B_i \times \Omega \,|\, n \geq 1,\ B_i \in \mathcal{F},\ i = 1, 2, \cdots, n\} \tag{2.6.44}$$

2.6.7 Remark

(1) In the operations of the joint tests, the operations about the sample space are the operations in set theory, so uncountable unions and uncountable intersections are permitted to be used.

(2) We represent the joint test of all tests $(\Omega_t, \mathcal{F}_t)(t \in T)$ by the symbols $(\odot_{t\in T}\Omega_t, \otimes_{t\in T}\mathcal{F}_t)$. New symbols have concise meaning: $\odot_{t\in T}\Omega_t$ is the cross partition of all $\Omega_t\,(t \in T)$; $\otimes_{t\in T}\mathcal{F}_t$ is the smallest event σ-field containing all $\mathcal{F}_t\,(t \in T)$. Particularly, if $T = \{1, 2, \cdots, n\}$ then it is $(\Omega_1 \odot \Omega_2 \odot \cdots \odot \Omega_n,\ \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n)$; if $\odot_{t\in T}\Omega_t$ is a true cross partition, then it is the product measurable space $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t)$.

(3) Fining the sample space Ω_1 to $\Omega_1 \odot \Omega_2$, it is from (2.4.3) and (2.6.18) that the partition block $\omega_\alpha^{(1)}$ is split as

$$\omega_\alpha^{(1)} = \bigcup_{\beta\in J_2} (\omega_\alpha^{(1)} \bigcap \omega_\beta^{(2)}) \tag{2.6.45}$$

i.e. Every partition block of Ω_1 all are similarly split as the partition block of $\Omega_1 \odot \Omega_2$.

Generally, we can split some partition blocks of Ω_1 into finer partition blocks by a way; the another partition blocks of Ω_1 into finer partition blocks by another way; and so on. By this manner we can obtain the tests which are more complex tests than the joint ones. The author has used this complex method when all right continuous Q-processes have been constructed[20].

2.7 The forth group of axioms: the group of axioms of probability measure

2.7.1 Intuitive background

Without the doubt, the probability is the most essential and the most crucial concept in probability theory. In order to put it into the NAS, in Section 1.2 according to common knowledge we abstract that:

The principle II of the probability theory: *Under the rational planned (or given) conditions C, the possibility of the occurrence of random event A can be expressed by the unique real number $P(A|C)$* in [0,1].

Then in Section 1.3 the probability was discuss elementarily. We introduced four illustrations about the probability, and point out the theoretical meaning and the useful value of the coexist of many illustrations. In fact, the classic probability theory and analytic probability theory in the probability history are founded on the four illustrations.

Now, we would indicate that four illustrations imply a big defect. They only consider the probability as separate values, do not emphasize that the probability is a family of values under the giving conditions. Therefore they neither operate "the family of values" as a whole logically and mathematically nor discuss the interior relations between "the giving conditions" and "the family of value".

In order to express the defect more clearly, we may agree in advance the probability to be a function (see the expression (2.7.1)). The defect of the four illustrations is that they only give the methods defining some function values (we abstract the axioms of the sixth group from these methods), have ignored the existence of these function values as a whole. Therefore it is difficult to operate the probability as a function logically and mathematically, and it is more difficult to discuss the interior relations between "the giving conditions" and the function.

The method to overcome this defect is to complete the task in Section 1.2. i.e. The condition $\mathcal{C}$, the event A and the probability $P(A|\mathcal{C})$ are transformed into the object which can be logically deducted and mathematically calculated.

The front three groups of axioms of the NAS draw a picture: the random universe in reality is originally a causation space. In this space there are various "graphics"— causal points, sample points, events, event fields, event σ-fields and random tests, and so on. It is peculiar that in order to understand "the small graphics", we must study "the large graphics" containing "the small graphics". For example, the principle I of the probability theory now is substantiated as

The principle I of the probability theory (in a narrow sense): *A particular event exists in some random test, only if this test is recognized then the particular event is understand.*

Where the content to be recognized includes the relations among the particular event and other events in the test. From this we have got that **the study of "graphics" starts from the random tests, then the concrete events and "sub-graphics" in the tests are studied**.

In reality the given conditions $\mathcal{C}$ are illustrated as the conditions of the performance of an experiment $\mathcal{E}$. According the discussion at Section 2.4, the experiment $\mathcal{E}$ generates a random test $(\Omega, \mathcal{F})$. Therefore the principle II implies that if the event $A \in \mathcal{F}$, then the size of possibility for A to occur is some real number $P(A|\mathcal{C})$. That is to say that the conditions $\mathcal{C}$ generate a set function

$$P(A|\mathcal{C}), \quad A \in \mathcal{F} \tag{2.7.1}$$

The reason to call it set function is that its domain is a set of sets.

Similar to the study of "graphics", **the study of probability starts from "set function", not from the probability values**. That is to say that at first we recognize the whole which consist of both of the domain and the taking values; Then the domain $\mathcal{F}$ must be known clearly (it is easy neglected by people); Finally the taking values is found as far as possible.

Another primary fact also verifies this point of view. People often use the proposition "the possibility of occurrence of the event A is value p". In order to understand this proposition, no matter whether he is conscious, there exists another proposition "the possibility of occurrence of the opposite event $\overline{A}$ is $1-p$" at once in his subconsciousness. Two propositions form a function which domain is $\mathcal{F} = \{\emptyset, A, \overline{A}, \Omega\}$ and corresponding law is the (2.7.7). In other words, **in order to understand the proposition "the possibility of occurrence of the event A is value p", he uses the Bernoulli test and a probability function (2.7.7) on the test.**

At last we recognize the true face of the probability — it is a set function defined on the event σ-field $\mathcal{F}$ taking values in [0,1] ! Since there are close relation among the events in $\mathcal{F}$, the family of probability values $\{P(A|\mathcal{C})|A \in \mathcal{F}\}$ reflecting the possibility of occurrence of those events must depend each other. **A. N. Kolmogorov abstracts three axioms from such dependence, i.e. the group of axioms of probability measure in this section.**

2.7.2 The group of axioms of probability measure

Definition 2.7.1. *Suppose $(\Omega, \mathcal{F})$ is a random test, $P(A), A \in \mathcal{F}$ is a real value function on $\mathcal{F}$. We call $P(A), A \in \mathcal{F}$ as a probability measure if it satisfies three axioms as follows:*

Axiom D_1 (Axiom of nonnegativity). $P(A) \geq 0$, *for every* $A \in \mathcal{F}$.

Axiom D_2 (Axiom of normalized). $P(\Omega) = 1$.

Axiom D_3 (Axiom of σ-additive). *If $A_i\ (i \in I)$ are finite or countable disjoint events, then*

$$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i) \tag{2.7.2}$$

Usually, the probability measure is called the probability in brief, and is written as $P(\mathcal{F})$ or P in brief [16].

Definition 2.7.2. *The triad $(\Omega, \mathcal{F}, P)$ is called a probability space if $(\Omega, \mathcal{F})$ is a random test and P is a probability measure on $\mathcal{F}$.*

Definition 2.7.3. *Suppose $(\Omega_i, \mathcal{F}_i, P_i)(i = 1, 2)$ is two probability spaces. The $(\Omega_2, \mathcal{F}_2, P_2)$ is called a probability subspace of $(\Omega_1, \mathcal{F}_1, P_1)$ if they satisfy two conditions as follows:*

1) $\mathcal{F}_2 \subset \mathcal{F}_1$;

2) $P_2(A) = P_1(A)$, *for any* $A \in \mathcal{F}_2$.

[16]From now on any function $f(x), x \in D$ can be written as $f(D)$ in brief.

Particularly, if $\Omega_1 = \Omega_2$, then it is called a probability subspace with same causes.

Before the probability spaces are studied, let us to give the intuitive meaning of Definition 2.7.2. If a probability space $(\Omega, \mathcal{F}, P)$ is given, then we have the conclude which *the possibility of occurrent of the event $A\,(A \in \mathcal{F})$ is probability value $P(A)$ under the condition of given $(\Omega, \mathcal{F}, P)$.* Contrasting it and the principle II, we discover the probability space $(\Omega, \mathcal{F}, P)$ depicts exactly and overall the given conditions $\mathcal{C}$ in the principle II. i.e.

$$\text{given conditions } \mathcal{C} \Longleftrightarrow (\Omega, \mathcal{F}, P) \tag{2.7.3}$$

Therefore, the given conditions $\mathcal{C}$ is abstracted as $(\Omega, \mathcal{F}, P)$; the events concerned with $\mathcal{C}$ is abstracted as the event σ-field $\mathcal{F}$; the possibility of occurrence of events is abstracted as probability $P(A), A \in \mathcal{F}$. These products can all be operated logically and mathematically.

Note that the (2.7.3) declares that Definition 2.7.2 defines exactly the conditions $\mathcal{C}$, and the probability symbol (2.7.1) becomes $P(A), A \in \mathcal{F}$.

In application of probability theory (particularly, the job of modelling) the given $\mathcal{C}$ is still the "self-evident" conditions of performance of the experiment. In Section 2.4 the condition $\mathcal{C}$ generates a random test, and now it generates a probability space $(\Omega, \mathcal{F}, P)$. i.e.

$$\text{Experiment } \mathcal{E} \text{ (under given conditions } \mathcal{C}) \Longrightarrow (\Omega, \mathcal{F}, P) \tag{2.7.4}$$

Obvious, the (2.7.4) constructs a pair relation of cause and effect in real-life. The condition $\mathcal{C}$ is cause; the probability space $(\Omega, \mathcal{F}, P)$ is effect.

Form now on, **in application of probability theory we regard the " self-evident " conditions $\mathcal{C}$ as the conditions of performance of experiment $\mathcal{E}$ so that we obtain the probability space $(\Omega, \mathcal{F}, P)$ generated by $\mathcal{E}$; When doing theoretical research, we may depict exactly conditions $\mathcal{C}$ of performance of experiment $\mathcal{E}$ by the probability space $(\Omega, \mathcal{F}, P)$.**

2.7.3 Existence of probability measure

Is there the probability measure? Is it unique? This is the problem which the fourth group of axioms is faced. Kolmogorov answered: *"Our system of axioms A – E is compatible (i.e. without contradiction). It can be verified by the example as follows: E consists of one element ξ, $\mathcal{F}$ consists of E and the empty $\emptyset$, where supposing $P(E) = 1, P(\emptyset) = 0$"*[1][17]. Afterwards he cited an example where $P(\mathcal{F})$ is not only one, this declared the system of axioms is not complete.

[17]This example is the degenerative probability space defined by (2.7.5).

The seventh topic for discussion put forth by de Finetti is that: *"Is the above-mentioned proof of compatibility satisfactory?"*. Afterwards he declared the unsatisfactory reason. We made following lemma as the answer to this topic.

Lemma 2.7.1. *For any random test $(\Omega, \mathcal{F})$, we can always assign a probability measure $P(\mathcal{F})$ to this test. The $P(\mathcal{F})$ is unique if and only if the $(\Omega, \mathcal{F})$ is the degenerative test.*

Proof: Suppose $(\Omega, \mathcal{F})$ is the degenerative test. At that time $\mathcal{F} = \{\emptyset, \Omega\}$. Let

$$P(A) = \begin{cases} 0, & A = \emptyset \\ 1, & A = \Omega \end{cases} \tag{2.7.5}$$

Obviously, the $P(\mathcal{F})$ is a probability measure and it is unique.

Now suppose $(\Omega, \mathcal{F})$ is not degenerative. At that time, there is $B \in \mathcal{F}$ such that $B \neq \emptyset$ and $B \neq \Omega$. Therefore there are the sample points ω_1 and ω_2 such that $\omega_1 \in B$, $\omega_2 \in \overline{B}$. For any $A \in \mathcal{F}$, let

$$P(A) = \begin{cases} 0, & \omega_1 \notin A \text{ and } \omega_2 \notin A \\ p, & \omega_1 \in A \text{ and } \omega_2 \notin A \\ q, & \omega_1 \notin A \text{ and } \omega_2 \in A \\ 1, & \omega_1 \in A \text{ and } \omega_2 \in A \end{cases} \tag{2.7.6}$$

where $p \geq 0$, $q \geq 0$, $p + q = 1$. It is easy to prove $P(\mathcal{F})$ to be a probability measure. It is from the p, q to be not unique that the $P(\mathcal{F})$ is not unique. ■

Example 1. Suppose $(\Omega, \mathcal{F})$ is Bernoulli test defined by (2.6.40) and (2.6.41). Let

$$P(B) = \begin{cases} 0, & B = \emptyset \\ p, & B = A \\ q, & B = \overline{A} \\ 1, & B = \Omega \end{cases} \tag{2.7.7}$$

where $p \geq 0$, $q \geq 0$, $p + q = 1$. Obviously, the $P(\mathcal{F})$ is a probability measure. ■

Form now on, the $(\Omega, \mathcal{F}, P)$ defined by (2.7.5) is called the degenerative probability space; The $(\Omega, \mathcal{F}, P)$ in Example 1 is called Bernoulli probability space.

2.7.4 Probability spaces of classical type

The following lemma declares the classical definition of probability satisfying the fourth group of axioms.

Lemma 2.7.2. *Suppose $(\Omega, \mathcal{F})$ is a test of n-type. For any $A \in \mathcal{F}$, let*

$$P(A) = \frac{\text{the number of sample points in } A}{n} \tag{2.7.8}$$

Then the $P(\mathcal{F})$ is a probability measure on the $(\Omega, \mathcal{F})$.

Proof : Obviously the $P(\mathcal{F})$ satisfies Axioms D_1 and D_2. We would like to prove it to satisfy Axiom D_3. Suppose $A_1, A_2, \cdots, A_m$ are disjoint events. Because all A_i $(1 \leq i \leq m)$ have no common sample point, therefore

$$\begin{aligned} P(\bigcup_{i=1}^{m} A_i) &= \frac{\text{the number of sample points in } \bigcup_{i=1}^{m} A_i}{n} \\ &= \frac{\text{the number of sample points in } A_1}{n} + \cdots \\ &\quad + \frac{\text{the number of sample points in } A_m}{n} \\ &= \sum_{i=1}^{m} P(A_i) \end{aligned} \tag{2.7.9}$$

Now suppose A_i $(i = 1, 2, \cdots)$ are countable disjoint events. There are at most 2^n elements in $\mathcal{F}$, therefore there are only finite nonempty events in A_i $(i = 1, 2, \cdots)$. We may suppose that they are $A_1, A_2, \ldots, A_m$, and $A_i = \emptyset$ $(i \geq m + 1)$. Obviously (2.7.8) implies $P(\emptyset) = 0$, so

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{m} A_i) = \sum_{i=1}^{m} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$

It is proved that the $P(\mathcal{F})$ satisfies the axiom of σ-additive. ∎

Corollary. *Suppose $(\Omega, \mathcal{F})$ is standard test of n-type, and $P(\mathcal{F})$ is a probability measure. Then (2.7.8) is set up if and only if for any $\omega \in \Omega$ we have*

$$P(<\omega>) = \frac{1}{n} \tag{2.7.10}$$

where $<\omega>$ is the event consisted of single sample point.

Definition 2.7.4. *The $(\Omega, \mathcal{F}, P)$ in Lemma 2.7.2 is called the probability spaces of classical type, and P is called the probability measure of classical type, or classical probability (measure).*

2.7.5 Probability spaces of geometric type

The following lemma declares the geometric probability satisfying the fourth group of axioms. The Lebesgue measure on n- dimension Borel test $(\mathbf{R}^n, \mathcal{B}^n)$

is denoted by $\mu(\mathcal{B}^n)$. As everybody knows, $\mu(\mathcal{B})$ is the length; $\mu(\mathcal{B}^2)$ is the area; $\mu(\mathcal{B}^3)$ is the volume.

Lemma 2.7.3. *Suppose $(D, \mathcal{B}^n(D))$ is a Borel test on D(see Definition 2.5.7), and $0 < \mu(D) < +\infty$. For any $A \in \mathcal{B}^n(D)$, let*

$$P(A) = \frac{\mu(A)}{\mu(D)} \tag{2.7.11}$$

then P is a probability measure on $(D, \mathcal{B}^n(D))$.

Proof : Obviously P satisfies Axioms D_1 and D_2. It is from the σ-additivity of the Lebesgue measure that P satisfies Axiom D_3. ■

Definition 2.7.5. *$(D, \mathcal{B}^n(D), P)$ in Lemma 2.7.3 is called the probability space of geometric type, and P is called the probability measure of geometric type, or the geometric probability (measure).*

The probability spaces of classic type and geometric type have importance roles during the early development period of probability theory. We would not give some concrete example here, since a lot of examples can be found in the primer probability books. Why do the these examples generate the probability spaces of classical type and geometric type at the time of modelling? We will discuss this problem in the sixth group of axioms (see Theorems 2.11.3 and 2.11.5 of Section 2.11).

2.7.6 Whether do the frequency illustration and subjection illustration satisfy the fourth group of axioms?

For stating easily and smooth, we introduce two axioms first. Suppose $(\Omega, \mathcal{F})$ is a random test and $P(A), A \in \mathcal{F}$ is a real value function.

Axiom D_4 (Axiom of finite additive). *If A and B are two disjoint events in $\mathcal{F}$, then*

$$P(A \bigcup B) = P(A) + P(B) \tag{2.7.12}$$

Axiom D_5 (Axiom of continuity). *If $A_n \in \mathcal{F}\,(n \geq 1)$ and $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$, $\bigcap_{n=1}^{\infty} A_n = \emptyset$, then*

$$\lim_{n\to\infty} P(A_n) = 0 \tag{2.7.13}$$

Afterwards we introduce two concludes to be proved in this section.

Conclude 1 (Theorem 2.7.1(6)). *If $A_i\,(i = 1, 2, \cdots, n)$ are disjoint events in $\mathcal{F}$, then the (2.7.12) is equivalent to*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{2.7.14}$$

Conclude 2 (Theorem 2.7.3). *Suppose* $(\Omega, \mathcal{F})$ *is a random test. Also suppose* $P(A), A \in \mathcal{F}$ *is a real value set function and satisfies Axioms* D_1, D_2 *and* D_4. *Then a necessary and sufficient for* $P(\mathcal{F})$ *to satisfy Axiom* D_3 *is it satisfies Axiom* D_5.

From (1.3.14) and (1.3.15) we make out that the frequency illustration of the probability satisfies Axioms D_1, D_2 and D_4 (thereby also satisfies the conclude 1) and does not reject Axiom D_3 (or Axiom D_5).

Similarly, obviously the subject illustration of the probability satisfies Axioms D_1 and D_2. A belief of the subject illustration require the equivalence of before and after judgement, i.e. compatibility and consistence. The requirement of compatibility leads to obey Axiom D_4 (also to obey the conclude 1), not to reject Axiom D_3 (or Axiom D_5)[2].

The NAS considers that the finite and the countable locate at same place. Since the frequency illustration and subjection illustration all support Axiom D_4, therefore we also consider that those two illustration satisfy Axiom D_3 (or Axiom D_5).

Now the real situation is that de Finetti and a group of scholars do not support Axiom D_3. They maintain that Axiom D_4 should be used instead of Axiom D_3[2][8]. In order to reflect their point of view, we give

Definition 2.7.6. *Suppose* $(\Omega, \mathcal{F})$ *is a random test, and* $P(A)$, $A \in \mathcal{F}$ *is a real value set function. The* $P(A)$, $A \in \mathcal{F}$ *is called finite additive probability, if it satisfies Axioms* D_1, D_2 *and* D_4.

Distinguishing from this, the P defined in Definition 2.7.1 may be called σ-additive probability. Obviously, a σ-additive probability is a finite additive probability, the reversal not to be right.

2.7.7 Basic properties of probability

Theorem 2.7.1. *Let* $(\Omega, \mathcal{F})$ *is a random test. If* $P(A), A \in \mathcal{F}$ *is a finite additive probability, then* $P(\mathcal{F})$ *has the properties as follows:*

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

(2) $0 \leq P(A) \leq 1 \ (A \in \mathcal{F})$;

(3) Complementarity: $P(A) + P(\overline{A}) = 1 \ (A \in \mathcal{F})$;

(4) Monotonicity: If $A \subset B$, *then* $P(A) \leq P(B)$;

(5) Implying subtraction: If $A \subset B$, *then* $P(B \setminus A) = P(B) - P(A)$;

(6) Finite additivity: If $A_1, A_2, \cdots, A_n \ (n \geq 2)$ *are disjoint events, then the (2.7.4) is set up.*

(7) Subadditivity: Suppose $A_1, A_2, \cdots, A_n \ (n \geq 2)$ *are any events, then*

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) \tag{2.7.15}$$

(8) General addition formula: Suppose $A_1, A_2, \cdots, A_n\,(n \geq 2)$ are any events in $\mathcal{F}$, we have

$$P(\bigcup_{i=1}^{n} A_i) = s_1 - s_2 + s_3 - \cdots + (-1)^{n-1} s_n \qquad (2.7.16)$$

where

$$\begin{aligned} s_1 &= \sum_{i=1}^{n} P(A_i), \quad s_2 = \sum_{i<j\leq n} P(A_iA_j), \\ s_3 &= \sum_{i<j<k\leq n} P(A_iA_jA_k), \ \cdots, \ s_n = P(A_1A_2\cdots A_n) \end{aligned}$$

Proof: (1) Since $\emptyset = \emptyset \bigcup \emptyset$, from Axiom D_4 we get $P(\emptyset) = P(\emptyset) + P(\emptyset)$, so $P(\emptyset) = 0$.

(3) Since $A \bigcup \overline{A} = \Omega$, the conclude is get by Axioms D_4 and D_2.

(2) Using (3) and Axiom D_1, the conclude is obtained.

(5) Supposing $A \subset B$, it is easy to verify that

$$A \bigcup (B \setminus A) = B; \qquad A \bigcap (B \setminus A) = \emptyset \qquad (2.7.17)$$

By Axiom D_4 we get $P(A) + P(B \setminus A) = P(B)$. That completes the proof of the implying subtraction.

(4) The conclude is get by (5) and Axiom D_1.

(6) Regarding that A_n is disjoint with $\bigcup_{i=1}^{n-1} A_i$, by Axiom D_4 we get

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{n-1} A_i) + P(A_n)$$

This operation is going on for the first term of the right, and the (2.7.14) is got after $n-1$ times those operations.

(7) We will prove the situation $n = 2$. Obviously we have

$$A_1 \bigcup A_2 = A_1 \bigcup (A_2 \setminus A_1A_2); \quad A_1 \bigcap (A_2 \setminus A_1A_2) = \emptyset \qquad (2.7.18)$$

By Axiom D_4 and the monotonicity of probability we get

$$P(A_1 \bigcup A_2) = P(A_1) + P(A_2 \setminus A_1A_2) \leq P(A_1) + P(A_2) \qquad (2.7.19)$$

So when $n = 2$ the (2.7.15) is set up. Now suppose when $n \leq r$ the (2.7.15) is true, we have

$$P(\bigcup_{i=1}^{r+1} A_i) \leq P(\bigcup_{i=1}^{r} A_i) + P(A_{r+1}) \leq \sum_{i=1}^{r+1} P(A_i)$$

By the mathematical deduction the (2.7.15) is set up for any positive integer.

(8) Using the implying subtraction for the equality in (2.7.19), we have

$$P(A_1 \bigcup A_2) = P(A_1) + P(A_2) - P(A_1A_2) \tag{2.7.20}$$

So when $n = 2$ the (2.7.16) is set up. Now supposing when $n \leq r$ the (2.7.16) is true, particularly, we have

$$\begin{aligned} P(\bigcup_{i=1}^{r} A_i) &= \sum_{i=1}^{r} P(A_i) - \sum_{i<j\leq n} P(A_iA_j) \\ &\quad + \cdots + (-1)^r P(A_1A_2\cdots A_r) \\ P(\bigcup_{i=1}^{r}(A_iA_{r+1})) &= \sum_{i=1}^{r} P(A_iA_{r+1}) - \sum_{i<j\leq n} P(A_iA_jA_{r+1}) \\ &\quad + \cdots + (-1)^r P(A_1A_2\cdots A_rA_{r+1}) \end{aligned}$$

On another side, from (2.7.20) we get

$$P(\bigcup_{i=1}^{r+1} A_i) = P(\bigcup_{i=1}^{r} A_i) + P(A_{r+1}) - P[\bigcup_{i=1}^{r}(A_iA_{r+1})]$$

Substituting the front two equalities into the equality above, and passing simple rearrangement we get (2.7.16) when $n = r+1$. By the mathematical deduction the (2.7.16) is set up for any positive integer. ■

Theorem 2.7.2 (Continuity theorem). *Let $(\Omega, \mathcal{F})$ is a random test. If $P(\mathcal{F})$ is a σ-additive probability, then $P(\mathcal{F})$ has the properties as follows:*

(1) Continuity from below: Suppose $A_i \in \mathcal{F}\,(i \in I)$. If $A_1 \subset A_2 \subset \cdots \subset A_i \subset \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$, then

$$P(A) = \lim_{i\to\infty} P(A_i) \tag{2.7.21}$$

(2) Continuity from above: Suppose $A_i \in \mathcal{F}\,(i \in I)$. If $A_1 \supset A_2 \supset \cdots \supset A_i \supset \cdots$ and $A = \bigcap_{i=1}^{\infty} A_i$, then

$$P(A) = \lim_{i\to\infty} P(A_i) \tag{2.7.22}$$

Proof : (1) Suppose $A_0 = \emptyset, B_i = A_i \setminus A_{i-1}\,(i = 1, 2, 3, \cdots)$. It is easy to verify all $B_i\,(i \in I)$ are disjoin each other, and

$$A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \tag{2.7.23}$$

By Axiom D_3 and the implying subtraction of probability we get

$$P(A) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{\infty}[P(A_i) - P(A_{i-1}] = \lim_{i\to\infty} P(A_i)$$

(2) Suppose $B_i = \overline{A}_i\,(i = 1,2,3,\cdots)$. By Lemma 2.2.3 we get that $B_1, B_2, B_3, \cdots$ is a sequence satisfying conditions in the (1), and $\overline{A} = \bigcup_{i=1}^{\infty} B_i$ (De Morgan law). By (2.7.21) we get

$$P(\overline{A}) = \lim_{i\to\infty} P(B_i) = \lim_{i\to\infty} P(\overline{A_i}) \tag{2.7.24}$$

Regarding $P(\overline{A}) = 1 - P(A)$, $P(\overline{A_i}) = 1 - P(A_i)$ and substituting into (2.7.24), then (2.7.22) is got. ■

Theorem 2.7.3. *Let $(\Omega, \mathcal{F})$ is a random test. A necessary and sufficient condition that $P(\mathcal{F})$ be a σ-additive probability is that $P(\mathcal{F})$ be a finite additive probability and satisfy Axiom D_5.*

Proof : The necessary is deducted from the continuity from above of σ-additive probability. We would prove the sufficiency. Suppose $A_i\ (i \in I)$ are the events in Axiom D_3. Let

$$A = \bigcup_{i=1}^{\infty} A_i; \quad B_n = \bigcup_{i=1}^{n} A_i; \quad C_n = \bigcup_{i=n+1}^{\infty} A_i \tag{2.7.25}$$

Obviously, B_n and C_n are disjoint each other and $A = B_n \bigcup C_n$. It is from Axiom D_4 and (2.7.14) that

$$P(A) = P(B_n) + P(C_n) = \sum_{i=1}^{n} P(A_i) + P(C_n) \tag{2.7.26}$$

Setting $n \to \infty$, we get $P(A) = \sum_{i=1}^{\infty} P(A_i) + \lim_{n\to\infty} P(C_n)$. In order to finish the proof, $\lim_{n\to\infty} P(C_n) = 0$ is only needed.

It is from $C_n = A \setminus B_n = A\overline{B}_n$ that $\overline{C}_n = \overline{A} \bigcup B_n$, so

$$\overline{\bigcap_{n=1}^{\infty} C_n} = \bigcup_{n=1}^{\infty} \overline{C}_n = \overline{A} \bigcup (\bigcup_{n=1}^{\infty} B_n) = \overline{A} \bigcup A = \Omega$$

That is $\bigcap_{n=1}^{\infty} C_n = \emptyset$. Obviously $C_1 \supset C_2 \supset \cdots \supset C_n \cdots$. Therefore $C_1, C_2, \cdots, C_n \cdots$ is an event sequence satisfying the conditions in Axiom D_5. Using this axiom, we have proved $\lim_{n\to\infty} P(C_n) = 0$. ■

2.7.8 Extension of probability measure

It is little that the probability measure is defined entirely just like (2.7.8) and (2.7.11). Usually the situation is that supposing $\mathcal{F} = \sigma(\mathcal{A})$, we can assigned the values of probability

$$P(A) \qquad (A \in \mathcal{A}) \tag{2.7.27}$$

on all sets of $\mathcal{A}$. We would like to ask, what conditions about $\mathcal{A}$ and $P(\mathcal{A})$ would guarantee to exist unique probability measure $P(A), A \in \mathcal{F}$ such that the values of $P(\mathcal{F})$ on $\mathcal{A}$ are (2.7.27)? In brief, what conditions can we extend $P(\mathcal{A})$ to $P(\mathcal{F})$ under?

In fact, the extension problem is one basic problem in the measure theory. In order to state this problem, we need introduce

Definition 2.7.7. *Let $\mathcal{R}$ is an event field in the causation space, $P(A), A \in \mathcal{R}$ is a real value function defined on $\mathcal{R}$. The $P(A), A \in \mathcal{R}$ is called a probability measure on $\mathcal{R}$ if $P(\mathcal{R})$ satisfies Axioms D_1 and D_2, and the condition: supposing $A_i\,(i \in I)$ are finite or countable disjoint events in $\mathcal{R}$, and $\bigcup_{i\in I} A_i \in \mathcal{R}$, then*

$$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i) \tag{2.7.28}$$

Theorem 2.7.4 (The extension theorem of probability measure). *Let $(\Omega, \mathcal{F})$ is a random test, $\mathcal{R}$ is an event field and $\mathcal{F} = \sigma(\mathcal{R})$. If $P(\mathcal{R})$ is a probability measure on $\mathcal{R}$, then there exists unique probability measure $P^*(\mathcal{F})$ satisfying*

$$P^*(A) = P(A), \qquad A \in \mathcal{R} \tag{2.7.29}$$

This theorem is the special case of the extension theorem of measure in the measure theory, its proof can be seen in the reference[16]. Usually $P^*(\mathcal{F})$ is still written as $P(\mathcal{F})$, and $P(\mathcal{F})$ is called the extension of $P(\mathcal{R})$.

Theorem 2.7.5. *Let $(\odot_{t\in T}\Omega_t, \otimes_{t\in T}\mathcal{F}_t, P)$ is a probability space, where $(\odot_{t\in T}\Omega_t, \otimes_{t\in T}\mathcal{F}_t)$ is a joint random test of $(\Omega_t, \mathcal{F}_t)$ $(t \in T)$. Then the probability measure $P(\otimes_{t\in T}\mathcal{F}_t)$ is uniquely defined only by its part of the values*

$$P(A), \quad A \in *_{t\in T}\mathcal{F}_t \tag{2.7.30}$$

where $*_{t\in T}\mathcal{F}_t$ is the cylinder event set defined by (2.6.38).

Proof : Suppose

$$\mathcal{R} = \{\bigcup_{i=1}^{m} B_i \,|\, m \geq 1,\ B_1, B_2, \cdots, B_m \text{ are disjoint cylinder events}\} \tag{2.7.31}$$

We would prove $\mathcal{R}$ is an event field. At first $\mathcal{R}$ is closed for the complement. In fact, supposing $B = A_{t_1} \bigcap A_{t_2} \bigcap \Omega_t$ is a 2- dimension cylinder event, we have

$$\overline{B} = (A_{t_1} \bigcap \overline{A}_{t_2} \bigcap \Omega_t) \bigcup (\overline{A}_{t_1} \bigcap A_{t_1} \bigcap \Omega_t) \bigcup (\overline{A}_{t_1} \bigcap \overline{A}_{t_2} \bigcap \Omega_t) \quad (2.7.32)$$

Obviously, the right is three disjoint cylinder events each other, so $\overline{B} \in \mathcal{R}$. Similarly, if B is a n- dimension cylinder event, then $\overline{B} \in \mathcal{R}$.

Furthermore suppose $B = \bigcup_{i=1}^{m} B_i$ is any element in $\mathcal{R}$. That is that $B_1, B_2, \cdots B_m$ are disjoint each other cylinder events. By the De Morgan low we get

$$\overline{B} = \bigcap_{i=1}^{m} \overline{B}_i \quad (2.7.33)$$

Form the proved conclude we get $\overline{B}_i$ is finite union of disjoint cylinder events. Using the distributive law of union and intersection (Theorem 2.2.3), we get $\overline{B}$ is finite union of some disjoint cylinder events, so $\overline{B} \in \mathcal{R}$. We have got $\mathcal{R}$ is closed for complement. Similarly, we can prove $\mathcal{R}$ to be closed for difference.

Obviously $\mathcal{R}$ is closed for finite union of disjoint events. By the remark (2) in Subsection 2.2.5, we get $\mathcal{R}$ to be closed for finite union. What $\mathcal{R}$ is an event field is proved.

Finally, it is from (2.7.14) that the family of values (2.7.30) defines uniquely the $P(\mathcal{R})$, so that it defines uniquely the probability measure $P(\otimes_{t\in T}\mathcal{F}_t)$ by Theorem 2.7.4. ■

Note that the elements of different forms in $*_{t\in T}\mathcal{F}_t$ can express the same event. Therefor the family of values in (2.7.30) are not given in advance, but are assigned by us, then in order to guarantee the one-value of $P(*_{t\in T}\mathcal{F}_t)$, we must grasp the structure of $*_{t\in T}\mathcal{F}_t$. The importance of Lemma 2.6.2 located that it declares the product random test $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t)$ to have simple, clear structure. At that time only empty event has expression in various forms. Therefore it is clear that $P(\prod_{t\in T}\mathcal{F}_t)$ defined by (2.7.34) following is an one-valued function.

2.7.9 Independent product probability space

Theorem 2.7.6. *Let $(\Omega_t, \mathcal{F}_t, P_t)(t \in T)$ are a group of probability spaces. Suppose also the joint test of $(\Omega_t, \mathcal{F}_t)(t \in T)$ is the product test $(\prod_{t\in T}\Omega_t,$ $\prod_{t\in T}\mathcal{F}_t)$. If we assign the family of probability values*

$$P(\prod_{i=1}^{n} A_{t_i} \times \Omega_t) = \prod_{i=1}^{n} P_{t_i}(A_{t_i}) \quad (2.7.34)$$

on the cylinder event set $*_{t\in T}\mathcal{F}_t$ *(see (2.6.38)), then* $P(*_{t\in T}\mathcal{F}_t)$ *can be extended to probability measure* $P(\prod_{t\in T}\mathcal{F}_t)$.

The proof can be found in the reference[16].

Definition 2.7.8. *The* $P(\prod_{t\in T}\mathcal{F}_t)$ *in theorem above is called as the independent product probability measure of the* $P_t(\mathcal{F}_t)(t\in T)$*, and is written as* $\prod_{t\in T}P_t$ *in brief, and* $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t)$ *is called as the independence product probability space of* $(\Omega_t, \mathcal{F}_t, P_t)(t\in T)$.

Particular, if $(\Omega_t, \mathcal{F}_t, P_t)$ do not depends on t, it is written as $(\Omega^T, \mathcal{F}^T, P^T)$, and is called an independent duplicate probability space.

It is easy to point out that $(\Omega_t, \mathcal{F}_t, P_t)$ is a probability subspace of $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t)$, and $(\prod_{t\in T}\Omega_t, \mathcal{F}_t, P_t)$ is a probability subspace with same causes.

In order to echo with Subsection 2.7.1, let us indicate the conclusion here to be different with the conclusions about probability in Subsection 2.7.7. The latter (Theorems 2.7.1 – 2.7.3) is the conclusions among the taking values of a probability measure; The front (Theorem 2.7.6) is the performance of mathematical calculation among a lot of probability measures.

Example 2. Let $(\Omega, \mathcal{F}, P)$ is the Bernoulli probability spaces in Example 1. We call $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$ as a countable duplicate Bernoulli probability space, where $(\Omega^\infty, \mathcal{F}^\infty)$ is defined by (2.6.42)– (2.6.44), P^∞ is the independent product probability measure.

For $\omega \stackrel{\text{def}}{=} (\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots)$ in (2.6.42), let

$$\nu_{nA}(\omega) = \text{ the number of } \alpha_1, \alpha_2, \cdots, \alpha_n \textit{ to be } A \tag{2.7.35}$$

Obviously for fixed n, the set

$$\{\,|\frac{\nu_{nA}(\omega)}{n} - p\,| < \varepsilon\} \stackrel{\text{def}}{=} \{\omega\,|\,|\frac{\nu_{nA}(\omega)}{n} - p\,| < \varepsilon\} \tag{2.7.36}$$

is the finite union of the cylinder events in (2.6.44). So

$$\{\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n} = p\} \stackrel{\text{def}}{=} \{\,\omega\,|\lim_{n\to\infty}\frac{\nu_{nA}(\omega)}{n} = p\} \tag{2.7.37}$$

is an event in $\mathcal{F}^\infty$. We have

Theorem 2.7.7 (Bernoulli theorem of large numbers). *In a countable duplicate Bernoulli probability space* $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$*, for any* ε $(\varepsilon > 0)$ *we have*

$$\lim_{n\to\infty} P^\infty(\,|\frac{\nu_{nA}(\omega)}{n} - p\,| < \varepsilon\,) = 1 \tag{2.7.38}$$

Theorem 2.7.8 (Borel theorem of large numbers). *In a countable duplicate Bernoulli probability space* $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$*, we have*

$$P^\infty(\lim_{n\to\infty} \frac{\nu_{nA}(\omega)}{n} = p\} = 1 \tag{2.7.39}$$

The both proofs is omitted for saving. The adaptive here primary proof can be found in the reference[21]. These two theorems put forth the scientific foundation for the frequency illustration of the probability (Subsection 1.3.3).

2.7.10 Remark

(1) The theorems of large numbers stated by two probability measures, such as Theorems 2.7.7 and 2.7.8, show clearly us the implications of laws of large numbers.

(2) Let $(\Omega, \mathcal{F}, P)$ is a probability space. Since the subset of the set with zero probability need not to be event; even if to be an event, it need not to belong to $\mathcal{F}$. This fact has brought unnecessary trouble for the research work. The method to delete trouble is that we believe every subset of the set with zero probability all is a event, belong to $\mathcal{F}$ and its probability equals to zero. The way to realize these things is that let

$$\widetilde{\mathcal{F}} = \{A \triangle N | A \in \mathcal{F},\ N \text{ is a subset of zero probability set}\} \tag{2.7.40}$$

$$\widetilde{P}(A \triangle N) = P(A) \tag{2.7.41}$$

where $\triangle$ is the symmetric difference introduced in Subsection 2.2.5. Habitually $\widetilde{P}(\widetilde{\mathcal{F}})$ is written as $P(\widetilde{\mathcal{F}})$. It is easy to verify $(\Omega, \widetilde{\mathcal{F}}, P)$ is a probability space, and $(\Omega, \mathcal{F}, P)$ is its probability subspace with same causes.

The $P(\widetilde{\mathcal{F}})$ is called the completion of probability measure $P(\mathcal{F})$, $(\Omega, \widetilde{\mathcal{F}}, P)$ is called the complemental probability space.

(3) Supposing $(\Omega, \mathcal{F}, P)$ is a probability space, then there are at most $n-1$ atomic events with the probability going beyond $1/n$. Therefore there are at most countable disjoint atom events with the positive probability in $\mathcal{F}$.

All atomic events with the positive probability is denoted by B_i $(i \in I)$, let

$$\widetilde{\Omega} = \{B_i,\ \omega \,|\, i \in I,\ \omega \in \Omega \setminus \bigcup_{i\in I} B_i\} \tag{2.7.42}$$

It is easy to prove that $\widetilde{\Omega}$ is coarser than Ω, $(\widetilde{\Omega}, \mathcal{F}, P)$ is a probability space, and its complement probability space is $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, P)$. Obviously, $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ is a standard random test. Therefore the method of completion has completed the standardization task mentioned in Subsection 2.5.6, and it has become a popular method.

(4) In the debate about the finite additive probability and the σ-additive probability, the author agrees with the point of view of Fishburn: *"My own attitude toward issue is pragmatic. Much like the Axiom of Choice in set theory, if I can do without countable additivity to get where I want to go, so much the better. But I will not hesitate to invoke it (i.e. Axiom D_3 or D_5 in this book) when its denial would create mathematical complexities of little interest to the topic at hand"*[8].

(5) The NAS takes in anything and everything in the debate above. It believes the finite additive and σ-additive probability spaces all are "the graphics" in the causation space, and encourages to introduce more "graphics". It is important to discover the statistical ideas and statistical laws and the relations among the "graphics".

2.8 Point functions on random test

In a probability space $(\Omega, \mathcal{F}, P)$, the $\mathcal{F}$ is a set consisted of all events concerned by us, the $P(\mathcal{F})$ is sizes (or a measure) of possibility of these events to occur. Therefore $(\mathcal{F}, P)$ is the main body of the probability space. The third group of axioms states that Ω is a important tool to recognize $\mathcal{F}$. We expound what Ω is also a important tool to recognize $P(\mathcal{F})$ in this section.

There are two kind of functions on a random test $(\Omega, \mathcal{F})$. The first kind is the set functions defined on $\mathcal{F}$. The probability measures are special set functions. The another kind is the point functions defined on Ω. Usually the point function is simpler than the set function.

The task of this section is to introduce some popular kind of point functions. By them we can determine all probability measures of the typical random tests. Therefore we get some kind of typical probability spaces. They have played the essential role in the theoretical system and application of probability theory.

2.8.1 Probability spaces of discrete type

Definition 2.8.1. *A probability space $(\Omega, \mathcal{F}, P)$ is called discrete type (or n-type, or countable type), if the random test $(\Omega, \mathcal{F})$ is discrete type (or n-type, or countable type).*

Definition 2.8.2. *Let $(\Omega, \mathcal{F})$ is discrete type test, and $f(\omega), \omega \in \Omega$ is a real value function. The $f(\omega), \omega \in \Omega$ is called as a distribution law on $(\Omega, \mathcal{F})$, or probability mass function, if it is not negative and $\sum_{\omega \in \Omega} f(\omega) = 1$.*

The sample points of discrete type test can be listed one by one. Without loss of generality, we may suppose

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_i, \cdots\} = \{\omega_i \mid i \in I\} \tag{2.8.1}$$

It may be a finite or countable set. At that time the distribution law $f(\omega), \omega \in \Omega$ can express by the tabulation form of function as

$$f : \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_i & \cdots \\ f(\omega_1) & f(\omega_2) & \cdots & f(\omega_i) & \cdots \end{pmatrix} \tag{2.8.2}$$

Therefore $f(\omega), \omega \in \Omega$ is also called a distribution column.

Theorem 2.8.1. *Let $(\Omega, \mathcal{F})$ is a discrete type test. Then the distribution law $f(\omega)$ determines a probability measure $P(\mathcal{F})$ by the formula*

$$P(A) = \sum_{\omega \in A} f(\omega), \quad A \in \mathcal{F} \tag{2.8.3}$$

On the contrary, supposing $P(\mathcal{F})$ is a probability measure, then there is a distribution law $f(\omega)$ such that (2.8.3) is set up. This $f(\omega)$ is unique if and only if there is not the atomic event B such that $P(B) > 0$ and B contains at least two sample points.

Proof: Suppose $f(\omega)$ is a distribution law, and $P(\mathcal{F})$ is defined by (2.8.3). Obviously $P(\mathcal{F})$ satisfies Axioms D_1 and D_2. Let $A_1, A_2, \cdots, A_k, \cdots$ are finite or countable disjoint events in $\mathcal{F}$, we may suppose

$$A_k = \{\omega_i \mid i \in I_k\} \quad (k = 1, 2, 3, \cdots)$$

where I_k is a subset of I (see (2.8.1)). Obviously all $I_k (k \geq 1)$ disjoint each other. Supposing $I_0 = \bigcup_{k \geq 1} I_k$, by the commutative law and combine law of the convergent positive series (since $\sum_{i \in I_0} f(\omega_i) \leq 1$), from (2.8.3) we get

$$P(\bigcup_{k \geq 1} A_k) = \sum_{i \in I_0} f(\omega_i) = \sum_{k \geq 1} \left[\sum_{i \in I_k} f(\omega_i)\right] = \sum_{k \geq 1} P(A_k)$$

So $P(\mathcal{F})$ satisfies Axiom D_3, the front half of theorem is proved.

The remainder part will be proved as follows. By Theorems 2.5.3 and 2.5.6 we know that there are finite or countable disjoint atomic events B_k $(k = 1, 2, 3, \cdots)$ such that

$$\Omega = \bigcup_{k \geq 1} B_k \tag{2.8.4}$$

Obviously there exists a real value function $f(\omega)$ which is not negative and such that

$$P(B_k) = \sum_{\omega \in B_k} f(\omega) \quad (k = 1, 2, 3, \cdots) \tag{2.8.5}$$

Due to

$$\sum_{\omega \in \Omega} f(\omega) = \sum_{k \geq 1} \left[\sum_{\omega \in B_k} f(\omega) \right] = \sum_{k \geq 1} P(B_k) = 1$$

So $f(\omega)$ is a distribution law.

For any $A \in \mathcal{F}$, from (2.8.4) we get

$$A = \bigcup_{k \geq 1} (AB_k) \tag{2.8.6}$$

Since all events are disjoint in the right, from σ-additivity of probability we get

$$P(A) = \sum_{k \geq 1} P(AB_k)$$

Noting that $AB_k (k \geq 1)$ is B_k or the empty event, we have

$$P(A) = \sum_{k \geq 1} \left[\sum_{\omega \in AB_k} f(\omega) \right] = \sum_{\omega \in A} f(\omega)$$

So (2.8.3) is proved.

The proof of unique is follows. If there is an atomic event B such that $P(B) > 0$ and B contains at least two sample points, obviously there are infinite $f(\omega)$ satisfying (2.8.5), what $f(\omega)$ is not unique is proved. If there is not such B, the function $f(\omega)$ exists and only one, it is

$$f : \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_i \cdots & \\ P(<\omega_1>) & P(<\omega_2>) & \cdots & P(<\omega_i>) & \cdots \end{pmatrix} \tag{2.8.7}$$

where we appoint that if $<\omega_i> \notin \mathcal{F}$ then $P(<\omega_i>) = 0$. ■

Corollary. *If $(\Omega, \mathcal{F})$ is a standard test, then the $P(\mathcal{F})$ and the distribution law (2.8.7) determine uniquely each other.*

Definition 2.8.3 *The point function $f(\omega)$ defined by (2.8.5) is called as the distribution law (or mass function) of probability measure $P(\mathcal{F})$.*

2.8.2 Kolmogorov's probability spaces

Definition 2.8.4. *A probability space* $(\mathbf{R}, \mathcal{B}, P)$ *is called Kolmogorov's*[18] *if* $(\mathbf{R}, \mathcal{B})$ *is one dimensional Borel test.*

Definition 2.8.5. *A real value function* $F(x)$ *defined on* $\mathbf{R}$ *is called a distribution function if* $F(x)$ *has three properties as follows:*

(1) Monotonic nondecreasing: $F(x_1) \leq F(x_2)$ *if* $x_1 \leq x_2$.

(2) Continuous from the right: $\lim_{x \to a^+} F(x) = F(a)$ *for any* $a \in \mathbf{R}$.

(3) Normality:

$$F(-\infty) \stackrel{\text{def}}{=} \lim_{x \to -\infty} F(x) = 0\,; \quad F(+\infty) \stackrel{\text{def}}{=} \lim_{x \to +\infty} F(x) = 1 \tag{2.8.8}$$

Theorem 2.8.2. *Let* $(\mathbf{R}, \mathcal{B})$ *is one dimensional Borel test. Then the distribution function* $F(x)$ *determines uniquely a probability measure* $P(\mathcal{B})$ *by the formula*

$$P((a, b]) = F(b) - F(a) \qquad (-\infty < a \leq b < +\infty) \tag{2.8.9}$$

Conversely, supposing $P(\mathcal{B})$ *is a probability measure, then there is unique distribution function* $F(x)$ *such that (2.8.9) is set up.*

Proof : We first prove the front half of the theorem. Let

$$\begin{aligned} \mathcal{A} &= \{(a, b] \mid -\infty \leq a \leq b \leq +\infty\} \qquad (2.8.10) \\ \mathcal{D} &= \{\bigcup_{i=1}^{m} (a_i, b_i] \mid m \geq 1,\ (a_1, b_1], (a_2, b_2], \\ &\qquad \cdots, (a_m, b_m] \textit{ are disjoint each other}\} \qquad (2.8.11) \end{aligned}$$

then $\mathcal{B} = \sigma(\mathcal{A}) = \sigma(\mathcal{D})$. It is easy to verify $\mathcal{D}$ to be an event field.

Supposing $F(x)$ is a distribution function, then we may define $P(\mathcal{A})$ by (2.8.9). For any event $\bigcup_{i=1}^{m} (a_i, b_i]$ in $\mathcal{D}$, let

$$P(\bigcup_{i=1}^{m} (a_i, b_i]\,) = \sum_{i=1}^{m} [F(b_i) - F(a_i)] \tag{2.8.12}$$

We may verify $P(\mathcal{D})$ to be an one -valued set function, and it is a probability measure on $\mathcal{D}$[16]. Now the conclude of the front half has got by using Theorem 2.7.4.

[18] The finite, countable and any dimensional Kolmogorov's probability spaces have been used widely and have played essential role in probability theory. Naming them as K's is the author makes up.

Therefore we prove the converse part . Suppose $P(\mathcal{B})$ is the given probability measure. For any real number x, let

$$F(x) = P((-\infty, x]\,) \tag{2.8.13}$$

Since $(a, b] = (-\infty, b] \setminus (-\infty, a]$, by using the implying subtraction of probability we get the (2.8.9).

From (2.8.9) and the property of nonnegative of probability we get that $F(x)$ is nonnegative, monotonic nondecreasing function. Noting that $(-\infty, x] = \bigcap_{m=1}^{\infty}(-\infty, x + \varepsilon_m]$, where ε_m is nonnegative, $\varepsilon_m \downarrow 0$. From the continuity from above of probability (Theorem 2.7.2) we have

$$\begin{aligned} F(x) &= P(\bigcap_{m=1}^{\infty}(-\infty, x + \varepsilon_m]\,) \\ &= \lim_{m\to\infty} P((-\infty, x + \varepsilon_m]\,) = \lim_{m\to\infty} F(x + \varepsilon_m) \end{aligned}$$

So we obtain $F(x)$ to be continuous from the right. Similarly we can prove $F(x)$ to satisfy the condition of the normality. Therefore $F(x)$ is a distribution function. Since (2.8.9) contains (2.8.13), so $F(x)$ is unique. The proof is completed. ■

From the knowledge of the Lebesgue-Stieltjes integral we have

Corollary. *The probability measure $P(\mathcal{B})$ and the distribute function $F(x)$ on a Borel test $(\mathbf{R}, \mathcal{B})$ are uniquely determined each other by (2.8.13) and*

$$P(A) = \int_A dF(x) \qquad (A \in \mathcal{B}) \tag{2.8.14}$$

where the integral is Lebesgue-Stieltjes integral.

Definition 2.8.6. *The point function $F(x)$ defined by (2.8.13) is called as the distribute function of probability measure $P(\mathcal{B})$.*

2.8.3 Kolmogorov's probability spaces: n-dimensional case

Definition 2.8.7. *A probability space $(\mathbf{R}^n, \mathcal{B}^n, P)$ is called as the n-dimensional Kolmogorov's if $(\mathbf{R}^n, \mathcal{B}^n)$ is a n-dimensional Borel test.*

Definition 2.8.8. *A real value function $F(x_1, x_2, \cdots, x_n)$ defined on n-dimensional Euclidean space $\mathbf{R}^n$ is called a distribution function of n-variables if $F(x_1, x_2, \cdots, x_n)$ has four properties as follows:*

(1) $F(x_1, x_2, \cdots, x_n)$ is the monotonic nondecreasing in all its arguments.

(2) $F(x_1, x_2, \cdots, x_n)$ is continuous from the right in all its arguments.
(3) For any $i\,(1 \le i \le n)$, if $x_i \to -\infty$, then

$$F(x_1, x_2, \cdots, x_n) \longrightarrow 0 \tag{2.8.15}$$

If $x_1 \to +\infty,\ x_2 \to +\infty,\ \cdots,\ x_n \to +\infty$, then

$$F(x_1, x_2, \cdots, x_n) \longrightarrow 1 \tag{2.8.16}$$

(4) For any $-\infty < a_i < b_i < +\infty\ (i = 1, 2, \cdots, n)$, we have

$$F_0 - \sum_{i=1}^{n} F_{ij} + \sum_{1 \le i < j \le n} F_{ij} - \sum_{1 \le i < j < k \le n} F_{ijk} + \cdots + (-1)^n F_{12\cdots n} \ge 0 \tag{2.8.17}$$

where $F_0 = F(b_1, b_2, \cdots, b_n)$, and for any $1 \le i < j < \cdots < m \le n$, we have

$$F_{ij\cdots m} = F(x_1, x_2, \cdots, x_n)\Bigg|\begin{array}{l} x_r = a_r,\ r = i, j, \cdots, m \\ x_s = b_s,\ s \ne i, j, \cdots, m \end{array} \tag{2.8.18}$$

Theorem 2.8.3. *Let $(\mathbf{R}^n, \mathcal{B}^n)$ is a n-dimensional Borel test. Then the distribution function of n-variables $F(x_1, x_2, \cdots, x_n)$ determines uniquely a probability measure $P(\mathcal{B}^n)$ by the formula*

$$P(\prod_{i=1}^{n}(a_i, b_i]) = F_0 - \sum_{i=1}^{n} F_{ij} + \sum_{1 \le i < j \le n} F_{ij} - \sum_{1 \le i < j < k \le n} F_{ijk} + \cdots + (-1)^n F_{12\cdots n} \tag{2.8.19}$$

Conversely, supposing $P(\mathcal{B}^n)$ is a probability measure, then there is unique distribution function $F(x_1, x_2, \cdots, x_n)$ such that (2.8.19) is set up.

Corollary. *The probability measure $P(\mathcal{B}^n)$ and the distribute function $F(x_1, x_2, \cdots, x_n)$ on n-dimensional Borel test $(\mathbf{R}^n, \mathcal{B}^n)$ are uniquely determined each other by the formulas*

$$F(x_1, x_2, \cdots, x_n) = P(\prod_{i=1}^{n}(-\infty, x_i]\,) \tag{2.8.20}$$

$$P(A) = \int_A d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n) \quad (A \in \mathcal{B}^n) \tag{2.8.21}$$

where the integral is n-dimensional Lebesgue-Stieltjes integral.

Theorems 2.8.2 and 2.8.3 are classic theorem in the Lebesgue-Stieltjes integral. The proof of n-dimensional situation is similar with the one dimensional situation (see [22]), and is omitted.

Definition 2.8.9. *The point function $F(x_1, x_2, \cdots, x_n)$ defined by (2.8.20) is called as the distribute function of n-variables of probability measure $P(\mathcal{B}^n)$.*

2.8.4 Kolmogorov's probability spaces: infinite dimensional case

Suppose T is an infinite index set.

Definition 2.8.10. *A probability space $(\mathbf{R}^T, \mathcal{B}^T, P)$ is called as the T-dimensional Kolmogorov's if $(\mathbf{R}^T, \mathcal{B}^T)$ is a T-dimensional Borel test.*

Definition 2.8.11. *For any fixed $t_1, t_2, \cdots, t_n$ in T, we introduce a distribute function of n-variables $F(t_1, x_1; t_2, x_2; \cdots; t_n, x_n)$. We call*

$$F \equiv \{F(t_1, x_1; t_2, x_2; \cdots; t_n, x_n) \,|\, n \geq 1,\ t_1, t_2, \cdots, t_n \in T\} \qquad (2.8.22)$$

as a family of finite dimensional distribution functions on $(\mathbf{R}^T, \mathcal{B}^T)$.

Definition 2.8.12. *Suppose F is a family of finite dimensional distribution functions. It is called consistent, if F satisfies two condition as follows:*

(1) If $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ is any arrangement of $(1, 2, \cdots, n)$, then

$$\begin{aligned} &F(t_{\alpha_1}, x_{\alpha_1}; t_{\alpha_2}, x_{\alpha_2};\ \cdots;\ t_{\alpha_n}, x_{\alpha_n}) \\ &\quad = F(t_1, x_1; t_2, x_2;\ \cdots;\ t_n, x_n) \end{aligned} \qquad (2.8.23)$$

(2) If $m < n$, then

$$\begin{aligned} &F(t_1, x_1; t_2, x_2;\ \cdots;\ t_m, x_m) \\ = &\lim_{x_{m+1}, \cdots, x_n \to +\infty} F(t_1, x_1; t_2, x_2;\ \cdots;\ t_n, x_n) \end{aligned} \qquad (2.8.24)$$

Theorem 2.8.4. *Let $(\mathbf{R}^T, \mathcal{B}^T)$ is a T-dimensional Borel test. Then the consistent family F of finite dimensional distribution functions on $(\mathbf{R}^T, \mathcal{B}^T)$ uniquely determines a probability measure $P(\mathcal{B}^T)$ by the formula*

$$\begin{aligned} P(\prod_{i=1}^{n}(a_{t_i}, b_{t_i}] \times \mathbf{R}) &= F_0 - \sum_{i=1}^{n} F_{ij} + \sum_{1 \leq i < j \leq n} F_{ij} \\ &- \sum_{1 \leq i < j < k \leq n} F_{ijk} + \cdots + (-1)^n F_{12\cdots n} \end{aligned} \qquad (2.8.25)$$

where the event on the left is the cylinder event defined by (2.6.37), the right is the value defined by (2.8.17), but ought to use $F(t_1, x_1; t_2, x_2; \cdots; t_n, x_n)$ instead of $F(x_1, x_2; \cdots, x_n)$, and a_{t_i}, b_{t_i} instead of a_i, b_i $(1 \leq i \leq n)$, respectively.

Conversely, suppose $P(\mathcal{B}^n)$ is a probability measure. For any real number $x_1, x_2, \cdots, x_n$ and any $t_1, t_2, \cdots, t_n$ in T, let

$$F(t_1, x_1; t_2, x_2; \cdots; t_n, x_n) = P(\prod_{i=1}^{n}(-\infty, x_i] \times \mathbf{R}) \tag{2.8.26}$$

$$F = \{F(t_1, x_1; t_2, x_2; \cdots; t_n, x_n) \mid n \geq 1, t_1, t_2, \cdots, t_n \in T)\} \tag{2.8.27}$$

then F is a consistent family of finite dimensional distribution functions and such that (2.8.25) is set up.

This is famous existence theorem of Kolmogorov[1]. It is foundation of the stochastic process theory.

Definition 2.8.13. *The F defined by (2.8.27) is called as the family of finite dimensional distribution functions of probability measure $P(\mathcal{B}^T)$.*

2.8.5 Remark

A conclude in the measure theory, as is known to all, is that if a distribution function of n-variables is not discrete type, then the probability measure $P(\mathcal{B}^n)$ defined by it can not extend to $(\mathbf{R}^n, \mathcal{F}_{11})$ (Example 11 in Section 2.4). That is the reason why the random test $(\mathbf{R}^n, \mathcal{F}_{11})$ has few of use.

2.9 The fifth group of axioms: the group of axioms of conditional probability measure

We can assign a lot of probability measures to a random test $(\Omega, \mathcal{F})$. It is an important task of probability theory to study the relations among these probability measures. The fifth group of axioms inquires into the interior relations between the conditions $\mathcal{C}$ and probability $P(\mathcal{F})$, and abstracts the concepts of conditional probability and independence.

The random tests $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are the sub-tests of the joint test $(\Omega_1 \odot \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Therefore the conditional probability and independence are an important tool to study the interior relations among probability spaces, too.

2.9.1 Intuitive background

In this section we always suppose that the $(\Omega, \mathcal{F}, P)$ is a probability space generated by (2.7.4), i.e.

$$\text{Experiment } \mathcal{E} \text{ (under given conditions } \mathcal{C}) \Longrightarrow (\Omega, \mathcal{F}, P) \tag{2.9.1}$$

Let us examine how does the probability space vary when the condition varies.

(1) $(\Omega, \sigma(\mathcal{F}_B), P^*)$

Suppose $B \in \mathcal{F}$, $P(B) > 0$. If "the event B has occurred" is added to the condition $\mathcal{C}$, then we obtain a new condition which is denoted by $\mathcal{C}^*$.

According to the general knowledge, **under the condition $\mathcal{C}^*$ we select the bud events as all sub-events of B**, i.e.

$$\mathcal{F}_B = \{ A \mid A \in \mathcal{F},\ A \subset B \} \tag{2.9.2}$$

Therefore the condition $\mathcal{C}^*$ generates a random test $(\Omega, \sigma(\mathcal{F}_B))$. It is easy to point out $\overline{B}$ to be an atomic event of $\sigma(\mathcal{F}_B)$. Instead of (2.9.1) (i.e.(2.7.4)) we have

$$\text{Experiment } \mathcal{E} \text{ (under given conditions } \mathcal{C}^*) \Longrightarrow (\Omega, \sigma(\mathcal{F}_B), P^*) \tag{2.9.3}$$

where P^* is the probability measure to be inquire into. Obviously the condition $\mathcal{C}^*$ implies

$$P^*(B) = 1\,; \qquad P^*(\overline{B}) = 0$$

At first, let us discuss the probability values of the events in $\mathcal{F}_B$. The experience tell us that if the condition $\mathcal{C}$ is strengthened as $\mathcal{C}^*$, then the possibilities of occurrence of the events in $\mathcal{F}_B$ all increase, i.e. $P^*(A) \geq P(A) (A \in \mathcal{F}_B)$. How much is the ranger to increase? **A rational supposition is that they increase according the ratio**. i.e.

$$\frac{P^*(A)}{P(A)} = \frac{P^*(A_1)}{P(A_1)} \quad (A,\ A_1 \in \mathcal{F}_B, P(A) > 0, P(A_1) > 0) \tag{2.9.4}$$

Particularly, taking $A_1 = B$, it becomes

$$P^*(A) = \frac{P(A)}{P(B)} \qquad (A \in \mathcal{F}_B) \tag{2.9.5}$$

Afterwards, we discuss the probability values of the events in $\sigma(\mathcal{F}_B)$. For any $A \in \sigma(\mathcal{F}_B)$, let $D = A \bigcap B$, then $D \in \sigma(\mathcal{F}_B)$. Due to the atomicity of $\overline{B}$, we get that one and only one of both $A = D$ and $A = D \bigcup \overline{B}$ is set up. Regardless which case, it is rational that let $P^*(A) = P^*(D) = \dfrac{P(D)}{P(B)}$. Therefore (2.9.5) is extended as

$$P^*(A) = P^*(D) = \frac{P(AB)}{P(B)} \qquad (A \in \sigma(\mathcal{F}_B)) \tag{2.9.6}$$

And that realizes the relation of causation reflected by (2.9.3).

(2) The "completion" of $(\Omega, \sigma(\mathcal{F}_B), P^*)$

In the probability space $(\Omega, \sigma(\mathcal{F}_B), P^*)$ we have $P(\overline{B}) = 0$. Let

$$\mathcal{F}_{\overline{B}} = \{ A \mid A \in \mathcal{F}, A \subset \overline{B} \} \tag{2.9.7}$$

$$P^*(A) = 0 \qquad (A \in \mathcal{F}_{\overline{B}}) \tag{2.9.8}$$

Obviously $\mathcal{F} = \sigma(\mathcal{F}_B \bigcup \mathcal{F}_{\overline{B}})$, and for any $A \in \mathcal{F}$, we have

$$A = (AB) \bigcup (A\overline{B})$$

So it is rational that

$$P^*(A) = P^*(AB) + P^*(A\overline{B}) = \frac{P(AB)}{P(B)} \qquad (A \in \mathcal{F}) \tag{2.9.9}$$

Therefore we also extend $P^*(\sigma(\mathcal{F}_B))$ to $P^*(\mathcal{F})$. It is easy to verify $(\Omega, \sigma(\mathcal{F}_B), P^*)$ is a probability subspace of $(\Omega, \mathcal{F}, P^*)$. Instead of (2.9.3), we have

$$\text{Experiment } \mathcal{E} \text{ (under given conditions } \mathcal{C}^*) \Longrightarrow (\Omega, \mathcal{F}, P^*) \tag{2.9.10}$$

(3) We rewrite $P^*(A), A \in \mathcal{F}$ and $P^*(\mathcal{F})$ as $P(A|B), A \in \mathcal{F}$ and $P(\mathcal{F}|B)$, respectively. After comparing (2.9.1) and (2.9.10) we naturally call $P(\mathcal{F}|B)$ as a conditional probability (or a conditional probability measure) of $\mathcal{F}$ given event B. The $P(A|B)$ is called as the conditional probability value of A given B. Under new symbols the (2.9.9) has become an equality which is known to all:

$$P(A|B) = \frac{P(AB)}{P(B)} \qquad (A \in \mathcal{F}) \tag{2.9.11}$$

(4) The four illustrations of probability all satisfy (2.9.11).

① In the classical definition of probability, from (2.7.8) we get

$$P(A|B) = \frac{\text{the number of sample points in } AB}{\text{the number of sample points in } B} \tag{2.9.12}$$

It declares that if the condition $\mathcal{C}$ is strengthened as $\mathcal{C}^*$, then B plays the role of the sure event Ω, and the classical method of assignment probability is keeping still.

② In the geometric probability, from (2.7.11) we get

$$P(A|B) = \frac{\mu(AB)}{\mu(B)} \tag{2.9.13}$$

It declares that if the condition C is strengthened as C^*, then B plays the role of the sure event D, and the method of assignment geometric probability is keeping still.

③ In the frequency illustration of probability, from (1.3.15) we get

$$P(A|B) = \lim_{\nu_n(B)\to\infty} \frac{\nu_n(AB)}{\nu_n(B)} \tag{2.9.14}$$

Since $P(B) = \lim_{n\to\infty} \dfrac{\nu_n(B)}{n} > 0$, so it is rational where to use $\nu_n(B) \to \infty$. The (2.9.14) declares that after delating the terms that the B does not occur in the idealistic infinite sequence, the left terms still forms an idealistic infinite sequence. At that time, the B plays the role of the sure event Ω, and the frequency method of assignment probability is keeping still.

④ In subjective illustration of probability, the conditional probability is defined directly just like the probability. The (2.9.11) is the result of the consistency requirement.

2.9.2 The group of axioms of conditional probability measure

Lemma 2.9.1. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. Let*

$$\mathcal{F}^* = \{\, B \mid B \in \mathcal{F}, P(B) > 0 \,\}^{19} \tag{2.9.15}$$

then set function of two variables

$$P(A|B) \overset{def}{=} \frac{P(AB)}{P(B)}, \qquad A \in \mathcal{F},\ B \in \mathcal{F}^* \tag{2.9.16}$$

has three properties as follows:

(1) If $A \in \mathcal{F}, B \in \mathcal{F}^*$, then

$$0 \le P(A|B) \le 1 \tag{2.9.17}$$

(2) If $B \in \mathcal{F}^*$, then

$$P(\Omega|B) = 1 \tag{2.9.18}$$

(3) For any $B \in \mathcal{F}^*$ and finite or countable disjoint events A_i $(i \in I)$, we have

$$P(\bigcup_{i\in I} A_i|\ B) = \sum_{i\in I} P(A_i|B) \tag{2.9.19}$$

[19] Afterwards the right top index "*" of the event σ-field always plays the role like this.

Proof : (1) and (2) is obvious. (3) is obtained by (2.9.16) and Axiom D_3. ■

Corollary. *For any fix $B \in \mathcal{F}^*$, the set function of one variable $P(\mathcal{F}|B)$ is a probability measure on $(\Omega, \mathcal{F})$. The set function of two variables $P(\mathcal{F}|\mathcal{F}^*)$ is a family of probability measures $\{P(\mathcal{F}|B) |\ B \in \mathcal{F}^*\}$ on $(\Omega, \mathcal{F})$.*[20]

As is known to all the two essential factors of defining function are the domain and the corresponding law. Even if the corresponding law are identical, the different domains represent the different functions. In order to emphasize the important of the domain, we use the footnote 16 and the symbols $P(\mathcal{F}|B)$ and $P(\mathcal{F}|\mathcal{F}^*)$ in the corollary above. For convenience of writing, afterwards the $P(\mathcal{F}|\mathcal{F}^*)$ is written $P(\mathcal{F}|\mathcal{F})$. i.e. Please readers judge yourself whether $\mathcal{F}$ has right top index "*".

Suppose $y = f(x), x \in \mathcal{D}$ is a function. The operation which makes up a new function $y = f(x), x \in \mathcal{D} \bigcap \mathcal{D}_1$ is called a restriction of $f(\mathcal{D})$ on $\mathcal{D}_1$ and is written as $f(\mathcal{D}) \restriction \mathcal{D}_1$. i.e.

$$f(\mathcal{D} \bigcap \mathcal{D}_1) = f(\mathcal{D}) \restriction \mathcal{D}_1 \tag{2.9.20}$$

Particularly, if $\mathcal{D}_1 \subset \mathcal{D}$, then it becomes $f(\mathcal{D}_1) = f(\mathcal{D}) \restriction \mathcal{D}_1$.

Now suppose $\mathcal{F}_1$ and $\mathcal{F}_2$ are two any event σ-subfields of $\mathcal{F}$. Using the restriction operation, we may obtain three classes of set functions by $P(\mathcal{F}|\mathcal{F})$ as follows:

$$P(\mathcal{F}_1|\mathcal{F}_2), \quad i.e. \quad P(A|B),\ A \in \mathcal{F}_1, B \in \mathcal{F}_2^* \tag{2.9.21}$$

$$P(A|\mathcal{F}_2), \quad i.e. \quad P(A|B),\ B \in \mathcal{F}_2^* \tag{2.9.22}$$

$$P(\mathcal{F}_1|B), \quad i.e. \quad P(A|B),\ A \in \mathcal{F}_1 \tag{2.9.23}$$

The fifth group of axioms reveals the interior relations between the condition and the probability by these three classes of set functions. In this group of axioms the $(\Omega, \mathcal{F}, P)$ is a probability space generated by (2.9.1), $\mathcal{F}_1$ and $\mathcal{F}_2$ are two any event σ-subfields of $\mathcal{F}$, B is an event and $P(B) > 0$.

Axiom E_1 (Axiom of event's condition). *Let $B \in \mathcal{F}^*$. Adding "the event B has occurred" to the condition C, we get new condition C^*. Then the condition C^* assignments a probability measure $P(\mathcal{F}_1|B)$ to the sub-test $(\Omega, \mathcal{F}_1)$.*

Axiom E_2 (Axiom of sub-test's condition). *Let $\mathcal{F}_2 \subset \mathcal{F}$. Adding "the sub-test $(\Omega, \mathcal{F}_2)$ has performed" to the condition C, we get new condition C^{**}. Then the condition C^{**} assignments a family of probability measures $P(\mathcal{F}_1|\mathcal{F}_2)$ to the sub-test $(\Omega, \mathcal{F}_1)$.*

[20]The $P(\mathcal{F}|B)$ is the abbreviation of $P(A|B), A \in \mathcal{F}$. The $P(\mathcal{F}|\mathcal{F}^*)$ is abbreviation of $P(A|B), A \in \mathcal{F}, B \in \mathcal{F}^*$. See footnote 16.

Definition 2.9.1. *The probability space* $(\Omega, \mathcal{F}, P)$ *is given. Supposing* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *are the event* σ*-subfields of* $\mathcal{F}$, $A \in \mathcal{F}$, $B \in \mathcal{F}^*$, *then*

(1) The $P(\mathcal{F}_1|\mathcal{F}_2)$ *is called the family of conditional probability measures on the sub-test* $(\Omega, \mathcal{F}_1)$ *given the sub-test* $(\Omega, \mathcal{F}_2)$. *It is called the conditional probability of* $\mathcal{F}_1$ *given* $\mathcal{F}_2$ *in brief.*

(2) The $P(A|\mathcal{F}_2)$ *is called the family of conditional probability values of the event* A *given the sub-test* $(\Omega, \mathcal{F}_2)$. *It is called the conditional probability of* A *given* $\mathcal{F}_2$ *in brief.*

(3) The $P(\mathcal{F}_1|B)$ *is called the conditional probability measure of the sub-test* $(\Omega, \mathcal{F}_1)$ *given the event* B. *It is called the conditional probability of* $\mathcal{F}_1$ *given* B *in brief.*

(4) The $P(A|B)$ *is called the conditional probability value of the event* A *given the event* B. *It is called the conditional probability of* A *given* B *in brief.*

Until to now the front two classes of conditional probability have not been introduced in probability theory. Although the third class occasionally has been used, but has not been formally defined. Since conditional probability value $P(A|B)$ is solely and generally used, the conditional probability has become a concept with the ambiguous intuition.

The front two classes of conditional probability must be involved in deep development of probability theory. Therefore the modern probability theory has the aid of the concept of the random variable, has introduced two classes of the point function by using an abstract theorem in the measure theory. They instead of the front two classes of conditional probability here. Because these two classes of the point functions determine uniquely the front two classes of conditional probability under very general condition, the substitution has taken a great success. But people call these two classes of the point functions as conditional probability, so the conditional probability becomes a concept which is difficult to understand. The "conditional probabilities" of the point functions are a basic concept in the modern probability theory, and is one of most important calculating tools. The fifth group of axioms provides intuitive and appearance illustration for them (see Section 2.10).

2.9.3 Independence of two sub-tests

Due to the intuitive background and the important role people admire that "the independence is a special concept distinguishing probability theory from the measure theory"[21]. But the original definition of the independence is crude and artificial, and does not match with its intuitive and appearance.

[21]See *Chinese Encyclopedia: Mathematics*, P 243.

It is the hope changing unreasonable phenomena to introduce the fifth group of axioms.

The conditional probability $P(\mathcal{F}_1|\mathcal{F}_2)$ and $P(\mathcal{F}_2|\mathcal{F}_1)$ reflect the relations between the probability subspace $(\Omega, \mathcal{F}_1, P(\mathcal{F}_1))$ and $(\Omega, \mathcal{F}_2, P(\mathcal{F}_2))$. The simplest linkage is "having no relation". i.e. The occurrence of events in $\mathcal{F}_1$ has nothing to do with whether the sub-test $(\Omega, \mathcal{F}_2)$ is performed; similarly the occurrence of events in $\mathcal{F}_2$ has nothing to do with whether the sub-test $(\Omega, \mathcal{F}_1)$ is performed. By Axiom E_2, the "having no relation" can be represented by mathematical expression

$$P(\mathcal{F}_1|\mathcal{F}_2) = P(\mathcal{F}_1); \qquad P(\mathcal{F}_2|\mathcal{F}_1) = P(\mathcal{F}_2) \tag{2.9.24}$$

So the conception about independence generates.

Definition 2.9.2 (Independence of two sub-tests). *The sub-test $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$ are called as independence each other about probability measure $P(\mathcal{F})$, and called independence in brief, if we have $P(\mathcal{F}_1|\mathcal{F}_2) = P(\mathcal{F}_1)$, that is*

$$P(A|B) = P(B) \quad (A \in \mathcal{F}_1, B \in \mathcal{F}_2^*) \tag{2.9.25}$$

Definition 2.9.3 (Independence of two events). *The events A and B are called independence each other about probability measure $P(\mathcal{F})$, and called independence in brief, if there exist independent sub-tests $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$ such that $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$.*

By the two definitions we directly deduct

Lemma 2.9.2. *(1) The degeneration test $(\Omega_0, \mathcal{F}_0)$ and any sub-test are independent.*

(2) $\emptyset$ or Ω and any event are independent.

(3) If A and B are independent, then each of the three pair of events $\overline{A}$ and B, A and $\overline{B}$, $\overline{A}$ and $\overline{B}$ is independent too.

Theorem 2.9.1. *Let $(\Omega, \mathcal{F}, P)$ is a probability space. Supposing $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$ are two event σ-subfields of $\mathcal{F}$, then $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$ are independent if and only if*

$$P(AB) = P(A)P(B) \quad (A \in \mathcal{F}_1, \ B \in \mathcal{F}_2) \tag{2.9.26}$$

Proof : We prove first the necessity of the conditions. If $P(B) = 0$, then (2.9.26) holds obviously. We suppose $P(B) > 0$. From (2.9.16) we get $P(AB) = P(A|B)P(B)$. The (2.9.26) is obtained from (2.9.25).

We turn to the proof of sufficiency. If $B \in \mathcal{F}^*$, then (2.9.26) implies $P(A) = \dfrac{P(AB)}{P(B)}$. The (2.9.25) is obtain from (2.9.16). ■

Corollary 1. $P(\mathcal{F}_1|\mathcal{F}_2) = P(\mathcal{F}_1)$ *if and only if* $P(\mathcal{F}_2|\mathcal{F}_1) = P(\mathcal{F}_2)$.

Corollary 2 (0-1 law). *Supposing $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent, then*

$$\mathcal{F}_1 \bigcap \mathcal{F}_2 \subset \{ A \,|\, P(A) = 0 \ \ or \ 1 \} \tag{2.9.27}$$

Corollary 3 (0-1 law). *If $\mathcal{F}_1$ with itself are independent, then*

$$\mathcal{F}_1 \subset \{ A \mid P(A) = 0 \ \ or\ 1\} \tag{2.9.28}$$

Corollary 4. *Supposing $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$ to be an independent product probability space, then the sub-sets $(\Omega_1 \times \Omega_2, \mathcal{F}_1)$ and $(\Omega_1 \times \Omega_2, \mathcal{F}_2)$ are independent.*

Corollary 5. *Let $(\Omega, \mathcal{F}, P)$ is a probability space. Supposing $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent about probability $P(\mathcal{F})$, then $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent in the subspace $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$.*

Therefore, provided $(\Omega, \mathcal{F}, P)$ contains two independent events, then it implies two independent sub-sets (we may suppose they are $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$), so it implies a probability subspace $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$.

Theorem 2.9.2. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space, $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent event σ-subfields. Then the probability subspace $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$ has properties follows: If $A \in \mathcal{F}_1, B \in \mathcal{F}_2$, then*

(1) (2.9.26) is set up, and $P(\mathcal{F}_1 \otimes \mathcal{F}_2)$ is determined only by $P(\mathcal{F}_1)$ and $P(\mathcal{F}_2)$;

(2) $P(A \bigcap B) = 0$ if and only if $P(A) = 0$ or $P(B) = 0$;

(3) $\mathcal{F}_1 \bigcap \mathcal{F}_2 \subset \{ C \mid P(C) = 0 \ \ or \ \ 1\}$.

(4) If $P(A)$ and $P(B)$ all do not equal to 0 or 1, then

$$A \bigcap B \notin \mathcal{F}_1 \ ; \qquad A \bigcap B \notin \mathcal{F}_2$$

Proof : (1) (2.9.26) is got by Theorem 2.9.1. Remarking that the bud event set $\mathcal{F}_1 * \mathcal{F}_2$ of $\mathcal{F}_1 \otimes \mathcal{F}_2$ is definld by (2.6.1), so $P(\mathcal{F}_1 * \mathcal{F}_2)$ is determined uniquely by $P(\mathcal{F}_1)$ and $P(\mathcal{F}_2)$. The conclude (1) is an immediate consequence of Theorem 2.7.5 .

(2) The conclude is caused directly by (2.9.26).

(3) It is the 0-1 law.

(4) If the conclude does not set up. Suppose $A \bigcap B \in \mathcal{F}_1$. By using (2.9.26) we get

$$P(A \bigcap B) = P([A \bigcap B] \bigcap B) = P(A \bigcap B) P(B)$$

It is contradiction with $P(B) \neq 0$ or 1, so $A \bigcap B \notin \mathcal{F}_1$. Similarly $A \bigcap B \notin \mathcal{F}_2$ ■

Comparing this theorem with Lemma 2.6.2 and Theorem 2.7.6 (in case $n = 2$), the $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$ is similar with the independent product probability space, where instead of the exact equality and containing expression by the probability's equality and inequality. Therefore we can consider $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2, P)$ to be extension of the independent product probability space.

Theorem 2.9.3. *Supposing $(\Omega, \mathcal{F}, P)$ is a probability space, $A, B \in \mathcal{F}$, then A and B are independent if and only if*

$$P(AB) = P(A)P(B) \tag{2.9.29}$$

Proof : The necessity is get by Theorem 2.9.1. We would prove the sufficiency as follows. Suppose A and B satisfies (2.9.29). Using the implying subtraction of the probability, we get

$$P(\overline{A}B) = P(B) - P(AB) = [1 - P(A)]P(B) = P(\overline{A})P(B) \tag{2.9.30}$$

From symmetry of A and B we get $P(A\overline{B}) = P(A)P(\overline{B})$. Instead of B in (2.9.30) by $\overline{B}$, we get $P(\overline{A}\,\overline{B}) = P(\overline{A})P(\overline{B})$. Now let

$$\mathcal{F}_1 = \{\emptyset, A, \overline{A}, \Omega\}; \quad \mathcal{F}_2 = \{\emptyset, B, \overline{B}, \Omega\} \tag{2.9.31}$$

From Theorem 2.9.1 we deduct $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent. So A and B are independent. ■

2.9.4 Independence of n sub-tests

Any arrangement of $(1, 2, \cdots, n)$ is denoted by the $(t_1, t_2, \cdots, t_n)$.

Definition 2.9.4 (Independence of n sub-tests). *The probability space $(\Omega, \mathcal{F}, P)$ is given, and suppose $(\Omega_i, \mathcal{F}_i)(i = 1, 2, \cdots, n)$ are n sub-tests. If for any $k\,(1 \leq k < n)$, the sub-test $(\Omega, \mathcal{F}_{t_1} \otimes \mathcal{F}_{t_2} \otimes \cdots \otimes \mathcal{F}_{t_k})$ and $(\Omega, \mathcal{F}_{t_{k+1}} \otimes \mathcal{F}_{t_{k+2}} \otimes \cdots \otimes \mathcal{F}_{t_n})$ are independent, then n sub-tests $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), \cdots, (\Omega_n, \mathcal{F}_n)$ are called mutually independent about probability measure $P(\mathcal{F})$, or n event σ-subfields $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$ are mutually independence in brief.*

Definition 2.9.5 (Independence of n events). *Let $A_i\,(i = 1, 2, \cdots, n)$ is n events in $(\Omega, \mathcal{F}, P)$. n events $A_i\,(i = 1, 2, \cdots, n)$ are called mutually independent about probability measure $P(\mathcal{F})$, or n event $A_i\,(i = 1, 2, \cdots, n)$ are mutually independence in brief, if there are n mutually independent event σ-subfields $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$ such that $A_i \in \mathcal{F}_i\,(i = 1, 2, \cdots, n)$.*

Lemma 2.9.3. *Suppose events $A_i\,(i = 1, 2, \cdots, n)$ are mutually independent, then for any k the events $A_{t_1}, A_{t_2}, \cdots, A_{t_k}, \overline{A}_{t_{k+1}}, \cdots, \overline{A}_{t_n}$ are mutually independent too.*

Proof : It is direct inference of Definition 2.9.5. ■

Lemma 2.9.4. *Let $(\Omega, \mathcal{F}, P)$ is a probability space, and $B \in \mathcal{F}$. The set of events which are independent with B is denoted by $\mathcal{D}_B$, i.e.*

$$\mathcal{D}_B = \{A \,|\, A \in \mathcal{F},\ P(AB) = P(A)P(B)\} \tag{2.9.32}$$

Then $\mathcal{D}_B$ is a λ-class, i.e. $\mathcal{D}_B$ has four properties as follows :

(1) $\emptyset$, $\Omega \in \mathcal{D}_B$;

(2) If $A_1, A_2 \in \mathcal{D}_B$ *and* $A_1 \bigcap A_2 = \emptyset$, *then* $A_1 \bigcup A_2 \in \mathcal{D}_B$;

(3) If $A_1, A_2 \in \mathcal{D}_B$ *and* $A_1 \supset A_2$, *then* $A_1 \setminus A_2 \in \mathcal{D}_B$;

(4) If $A_i \in \mathcal{D}_B$ $(i = 1, 2, \cdots, n, \cdots)$ *and* $A_1 \subset A_2 \subset \cdots \subset A_i \subset \cdots$, *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B$.

Proof : (1) Obviously.

(2) It is from $(A_1B)\bigcap(A_2B) = \emptyset$ that

$$\begin{aligned} P\left([A_1 \bigcup A_2] \bigcap B\right) &= P(A_1B) + P(A_2B) \\ &= [P(A_1) + P(A_2)]P(B) \\ &= P(A_1 \bigcup A_2)P(B) \end{aligned}$$

It follows that $A_1 \bigcup A_2 \in \mathcal{D}_B$.

(3) Noting $A_1B \supset A_2B$, by the implying subtraction of probability we get

$$\begin{aligned} P\left([A_1 \setminus A_2] \bigcap B\right) &= P(A_1B \setminus A_2B) \\ &= P(A_1B) - P(A_2B) \\ &= [P(A_1) - P(A_2)]P(B) \\ &= P(A_1 \setminus A_2)P(B) \end{aligned}$$

It follows that $A_1 \setminus A_2 \in \mathcal{D}_B$.

(4) Since all $A_i \setminus A_{i-1}$ $(i = 1, 2, 3, \cdots;\ A_0 = \emptyset)$ are disjoint, and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty}[A_i \setminus A_{i-1}]$, by proved (3) and σ-additivity of probability we get

$$\begin{aligned} P\left([\bigcup_{i=1}^{\infty} A_i] \bigcap B\right) &= P\left(\bigcup_{i=1}^{\infty}[A_i \setminus A_{i-1}] \bigcap B\right) \\ &= \sum_{i=1}^{\infty} P\left([A_i \setminus A_{i-1}] \bigcap B\right) \\ &= \sum_{i=1}^{\infty} P(A_i \setminus A_{i-1})P(B) \\ &= P(\bigcup_{i=1}^{\infty} A_i)P(B) \end{aligned}$$

It follows that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B$. ■

Theorem 2.9.4. *Let* $(\Omega, \mathcal{F}, P)$ *is a probability spaces,* $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2, \cdots, n)$ *are* n *sub-tests. Then a necessary and sufficient condition of*

$(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), \cdots, (\Omega_n, \mathcal{F}_n)$ *to be mutually independent is for any* $A_i \in \mathcal{F}_i \, (i = 1, 2, \cdots, n)$ *we have*

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n) \tag{2.9.33}$$

Proof : We prove first the necessity of the conditions. Let $A_i \in \mathcal{F}_i \, (i = 1, 2, \cdots, n)$, then $A_2 A_3 \cdots A_n \in \mathcal{F}_2 \otimes \mathcal{F}_3 \otimes \cdots \otimes \mathcal{F}_n$. Since $\mathcal{F}_1$ and $\mathcal{F}_2 \otimes \mathcal{F}_3 \otimes \cdots \otimes \mathcal{F}_n$ are independent, it is form Theorem 2.9.1 that

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 A_3 \cdots A_n)$$

Similarly we can prove $P(A_2 A_3 \cdots A_n) = P(\Omega A_2)P(A_3 A_4 \cdots A_n)$. Duplicating this operation $n - 1$ times, we have got (2.9.33).

We turn to the proof of sufficiency. Suppose (2.9.33) is set up. We would prove that for any $k, \mathcal{F}_{t_1} \otimes \mathcal{F}_{t_2} \otimes \cdots \otimes \mathcal{F}_{t_k}$ and $\mathcal{F}_{t_{k+1}} \otimes \mathcal{F}_{t_{k+2}} \otimes \cdots \otimes \mathcal{F}_{t_n}$ are independent. From Theorem 2.9.1 we known that in order to complete theorem's proof only to prove that for any

$$\left.\begin{aligned} &E \in \mathcal{F}_{t_1} \otimes \mathcal{F}_{t_2} \otimes \cdots \otimes \mathcal{F}_{t_k} \\ &F \in \mathcal{F}_{t_{k+1}} \otimes \mathcal{F}_{t_{k+2}} \otimes \cdots \otimes \mathcal{F}_{t_n} \end{aligned}\right\} \tag{2.9.34}$$

the equality $P(EF) = P(E)P(F)$ is set up.

Selecting any cylinder event $C = A_{t_1} A_{t_2} \cdots A_{t_k} \in \mathcal{F}_{t_1} * \mathcal{F}_{t_2} * \cdots * \mathcal{F}_{t_k}$, it follows from Lemma 2.9.4 that the set

$$\mathcal{D}_C = \{A \mid A \in \mathcal{F}, \; P(AC) = P(A)P(C)\}$$

is a λ-class. On the other side, (2.9.33) implies $\mathcal{D}_C \supset \mathcal{F}_{t_{k+1}} * \mathcal{F}_{t_{k+2}} * \cdots * \mathcal{F}_{t_n}$. Lemma 2.6.1 guarantees $\mathcal{F}_{t_{k+1}} * \mathcal{F}_{t_{k+2}} * \cdots * \mathcal{F}_{t_n}$ is a π-class. By the methods of λ-class we get

$$\mathcal{D}_C \supset \mathcal{F}_{t_{k+1}} \otimes \mathcal{F}_{t_{k+2}} \otimes \cdots \otimes \mathcal{F}_{t_n} \tag{2.9.35}$$

Now for the F in (2.9.34) we introduce a set of events

$$\mathcal{D}_F = \{A \mid A \in \mathcal{F}, \; P(AF) = P(A)P(F)\}$$

From Lemma 2.9.4 we know that $\mathcal{D}_F$ is a λ-class. From (2.9.35) we have $C \in \mathcal{D}_F$. Noting C may be selected anyway, so $\mathcal{D}_F \supset \mathcal{F}_{t_1} * \mathcal{F}_{t_2} * \cdots * \mathcal{F}_{t_k}$. Since its right side is a π-class, using the method of λ-class again we get

$$\mathcal{D}_F \supset \mathcal{F}_{t_1} \otimes \mathcal{F}_{t_2} \otimes \cdots \otimes \mathcal{F}_{t_k}$$

This expression implies that for any E and$F$ in (2.9.34) the equality $P(EF) = P(E)P(F)$ hold. ■

Corollary 1. *If n event σ-subfields $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_n$ are independent each other, then any $m\,(m < n)$ event σ-subfields in it are independent each other, too.*

Corollary 2. *If $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{F}_i, \prod_{i=1}^n P_i)$ is an independent product probability space. Then n sub-tests $(\prod_{i=1}^n \Omega_i,\ \mathcal{F}_i)\,(i = 1, 2, \cdots, n)$ are independent each other.*

Corollary 3. *If $(\Omega, \mathcal{F}, P)$ is a probability spaces, $\mathcal{F}_i\,(i = 1, 2, \cdots, n)$ are n independent event σ-subfields. Then in probability subspace $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n,\ P)$, the $\mathcal{F}_i\,(i = 1, 2, \cdots, n)$ are independent event σ-subfields about probability $P(\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n)$, and the $P(\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n)$ is determined only by n probability measures $P(\mathcal{F}_i)\,(i = 1, 2, \cdots, n)$.*

Similarly with the case of two independent sub-tests, Corollaries 2 and 3 reflect how the n independent sub-tests exist in $(\Omega, \mathcal{F}, P)$. Roughly speaking, if $(\Omega, \mathcal{F}, P)$ contains n independent events, then it contains n independent sub-tests (we may suppose they are $(\Omega, \mathcal{F}_1), \cdots, (\Omega, \mathcal{F}_n)$), therefore it has a probability subspace $(\Omega, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n, P)$ which is similar with n-dimension independent product probability space.

Theorem 2.9.5. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces, $A_i \in \mathcal{F}\,(i = 1, 2, \cdots, n)$. Then a necessary and sufficient condition that n events $A_1, A_2, \cdots, A_n$ to be independent is that for any $k\,(2 \le k \le n)$, we have*

$$P(A_{t_1} A_{t_2} \cdots A_{t_k}) = P(A_{t_1}) P(A_{t_2}) \cdots P(A_{t_k}) \tag{2.9.36}$$

Note: the (2.9.36) contains $2^n - n - 1$ equalities all together. They are

$$\left.\begin{array}{ll} P(A_{t_1} A_{t_2}) = P(A_{t_1}) P(A_{t_2}) & (\textit{total } C_n^2) \\ P(A_{t_1} A_{t_2} A_{t_3}) = P(A_{t_1}) P(A_{t_2}) P(A_{t_3}) & (\textit{total } C_n^3) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ P(A_{t_1} A_{t_2} \cdots A_{t_n}) = P(A_{t_1}) P(A_{t_2}) \cdots P(A_{t_n}) & (\textit{total } C_n^n) \end{array}\right\} \tag{2.9.37}$$

If the first group of equalities is set up, then $A_1, A_2, \cdots, A_n$ is called pairwise independent. As is known by all, $A_1, A_2, \cdots, A_n$ may not be independent when they are pairwise independent. The causes is that the first group of equalities can not deduct the other groups in general case.

Proof : We prove first the necessity of the conditions. Suppose n events $A_1, A_2, \cdots, A_n$ to be mutually independent. Then there are n independent sub-tests $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), \cdots, (\Omega_n, \mathcal{F}_n)$ such that $A_i \in \mathcal{F}_i\,(i = 1, 2, \cdots, n)$. Therefore using Theorem 2.9.4 for $A_{t_1}, A_{t_2}, \cdots, A_{t_k}$ and $A_{t_j} = \Omega\,(k + 1 \le j \le n)$ we have (2.9.36).

We turn to the proof of sufficiency. Suppose (2.9.36) is held. Firstly we deduct 2^n equalities by (2.9.36).

$$\left.\begin{array}{ll} P(A_{t_1}A_{t_2}\cdots A_{t_n}) = P(A_{t_1})P(A_{t_2})\cdots P(A_{t_n}) & (total\ C_n^0) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ P(A_{t_1}\cdots A_{t_{n-k}}\overline{A}_{t_{n-k+1}}\cdots\overline{A}_{t_n}) & \\ \quad = P(A_{t_1})\cdots P(A_{t_{n-k}})P(\overline{A}_{t_{n-k+1}})\cdots P(\overline{A}_{t_n}) & (total\ C_n^k) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ P(\overline{A}_{t_1}\overline{A}_{t_2}\cdots\overline{A}_{t_n}) = P(\overline{A}_{t_1})P(\overline{A}_{t_2})\cdots P(\overline{A}_{t_n}) & (total\ C_n^n) \end{array}\right\} \quad (2.9.38)$$

In fact, the first group of equalities in (2.9.38) is the last group in (2.9.37). Writing down $B = A_{t_1}, A_{t_2}, \cdots, A_{t_{n-1}}$ and using (2.9.36), we have

$$\begin{aligned} P(BA_{t_n}) &= P(A_{t_1}A_{t_2}\cdots A_{t_{n-1}}A_{t_n}) \\ &= P(A_{t_1})P(A_{t_2})\cdots P(A_{t_{n-1}})P(A_{t_n}) \\ &= P(A_{t_1}A_{t_2}\cdots A_{t_{n-1}})P(A_{t_n}) \\ &= P(B)P(A_{t_n}) \end{aligned}$$

It follows from (2.9.30) and (2.9.36) that

$$\begin{aligned} P(B\overline{A}_{t_n}) &= P(B)P(\overline{A}_{t_n}) \\ &= P(A_{t_1})P(A_{t_2})\cdots P(A_{t_{n-1}})P(\overline{A}_{t_n}) \end{aligned}$$

the second group in (2.9.38) is proved. The third group can be proved by the similar manner. Simply copy down we get (2.9.38).

Now let

$$\mathcal{F}_i = \{\, \emptyset,\ A_i,\ \overline{A}_i,\ \Omega \,\} \quad (i = 1, 2, \cdots, n) \tag{2.9.39}$$

The (2.9.38) guarantees $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2), \cdots, (\Omega_n, \mathcal{F}_n)$ are mutually independent. So $A_1, A_2, \cdots, A_n$ are mutually independent. ■

Corollary. *If n events $A_1, A_2, \cdots, A_n$ are mutually independent, then any $m\,(m < n)$ events in it are mutually independent, too.*

2.9.5 Independence of a family of sub-tests

An infinite index set is denoted by T.

Definition 2.9.6 (Independence of a family of sub-tests). *Let $(\Omega, \mathcal{F}, P)$ is a probability space. Suppose $\{(\Omega, \mathcal{F}_t) | t \in T\}$ is a family of sub-tests. If every finite sub-tests $(\Omega, \mathcal{F}_{t_1}), (\Omega, \mathcal{F}_{t_2}), \cdots, (\Omega, \mathcal{F}_{t_n})(t_1, t_2, \cdots, t_n \in T)$ are mutually independent, then the family of sub-tests $\{(\Omega, \mathcal{F}_t)\ | t \in T\}$*

is called independent about probability measure $P(\mathcal{F})$, or the family $\{\mathcal{F}_t\,|t \in T\}$ of event σ-subfields is called independent in brief.

Definition 2.9.7 (Independence of a family of events). *Let $(\Omega, \mathcal{F}, P)$ is a probability space and $A_t \in \mathcal{F}\,(t \in T)$. The event family $\{A_t|\,t \in T\}$ is called independent about probability measure $P(\mathcal{F})$, or the family $\{A_t|\,t \in T\}$ is called independent in brief, if there is an independent family $\{\mathcal{F}_t|\,t \in T\}$ of sub-tests such that $A_t \in \mathcal{F}_t$ $(t \in T)$.*

Following two theorems are direct inference of this two definitions.

Theorem 2.9.6. *Let $(\Omega, \mathcal{F}, P)$ is a probability space, $\mathcal{F}_t\,(t \in T)$ are a family of sub-tests. Then following three concludes are equivalent:*

(1) The family $\{\mathcal{F}_t|t \in T\}$ of sub-tests are independent;

(2) Any subfamily $\{\mathcal{F}_t|t \in T_1\}\,(T_1 \subset T)$ are independent;

(3) For any $t_1, t_2, \cdots, t_n \in T$ and $A_{t_1} \in \mathcal{F}_1, A_{t_2} \in \mathcal{F}_2, \cdots, A_{t_n} \in \mathcal{F}_n$, we have

$$P(A_{t_1}A_{t_2}\cdots A_{t_n}) = P(A_{t_1})P(A_{t_2})\cdots P(A_{t_n}) \tag{2.9.40}$$

Corollary 1. *Let $(\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t)$ is an independent product probability space. Then the family $\{(\prod_{t\in T}\Omega_t, \mathcal{F}_t)|t \in T\}$ of sub-tests are independent.*

Corollary 2. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces, $\{\mathcal{F}_t|t \in T\}$ are an independent family of event σ-subfields. Then in probability subspace $(\Omega, \otimes_{t\in T}\mathcal{F}_t, P)$, the family $\{\mathcal{F}_t|t \in T\}$ are independent about probability $P(\otimes_{t\in T}\mathcal{F}_t)$, and the $P(\otimes_{t\in T}\mathcal{F}_t)$ is determined only by a family of probability measures $P(\mathcal{F}_t)(t \in T)$.*

Similarly with the case of n independent sub-tests, Corollaries 1 and 2 reflect how the independent family of sub-tests exists in $(\Omega, \mathcal{F}, P)$. Roughly speaking, if $(\Omega, \mathcal{F}, P)$ contains an independent family $\{\,A_t\,|\,t \in T\}$ of events, then it contains an independent family $\{\,(\Omega, \mathcal{F}_t)\,|\,t \in T\}$ of sub-tests, therefore it has a probability subspace $(\Omega, \otimes_{t\in T}\mathcal{F}_t, P)$ which is similar with T-dimension independent product probability space.

Theorem 2.9.7. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces, $A_t \in \mathcal{F}\,(t \in T)$. Then following four concludes equivalent:*

(1) The family of events $\{A_t|t \in T\}$ is independent.

(2) Any subfamily of events $\{A_t|t \in T_1\}$ $(T_1 \subset T)$ is independent.

(3) Any finite event in the family $\{A_t|t \in T\}$ are independent.

(4) For any $n \geq 2$ and $t_1, t_2, \cdots, t_n \in T$, we have

$$P(A_{t_1}A_{t_2}\cdots A_{t_n}) = P(A_{t_1})P(A_{t_2})\cdots P(A_{t_n}) \tag{2.9.41}$$

2.9.6 Basic theorems of conditional probability values

Theorem 2.9.8 (The multiplication theorem). *Let $(\Omega, \mathcal{F}, P)$ is a prob-*

ability spaces, $A_i \in \mathcal{F}\,(i = 1, 2, \cdots, n)$. If $P(A_1A_2\cdots A_n) > 0$, then

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \times \cdots \times P(A_n|A_1A_2\cdots A_{n-1}) \qquad (2.9.42)$$

Proof: Since $P(A_1) \geq P(A_1A_2) \geq \cdots \geq P(A_1A_2\cdots A_n) > 0$, the conditional probabilities of the right of (2.9.42) all have meaning. If we substitute conditional probability values for the right of (2.9.42), we obtain

$$P(A_1A_2\cdots A_n) = P(A_1)\frac{P(A_1A_2)}{P(A_1)}\frac{P(A_1A_2A_3)}{P(A_1A_2)}\cdots\frac{P(A_1A_2\cdots A_n)}{P(A_1A_2\cdots A_{n-1})}$$

It is an identity, so (2.9.42) is set up. ■

Theorem 2.9.9 (The total probability theorem). *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces, and $\mathcal{H} = \{H_i \,|\, i \in I\}$ is a partition of Ω. Suppose $H_i \in \mathcal{F}$, $P(H_i) > 0\ (i \in I)$. Then for any $A \in \mathcal{F}$* we have

$$P(A) = \sum_{i \in I} P(A|H_i)P(H_i) \qquad (2.9.43)$$

Proof: Since all $AH_i\ (i \in I)$ are finite or countable disjoint events and $\bigcup_{i\in I} AH_i = A$, so

$$P(A) = \sum_{i\in I} P(AH_i)$$

Now for $P(AH_i)$ using the multiplication theorem we get (2.9.43). ■

Theorem 2.9.10 (The Bayes' theorem). *Under the conditions of Theorem 2.9.9, If $A \in \mathcal{F}^*$ (i.e. $A \in \mathcal{F}, P(A) > 0$), then*

$$P(H_k|A) = \frac{P(A|H_k)P(H_k)}{\sum_{i\in I} P(A|H_i)P(H_i)} \quad (k \in I) \qquad (2.9.44)$$

Proof: We have $P(H_k|A) = \dfrac{P(AH_k)}{P(A)}\ (k \in I)$. Now for the numerator using the multiplication theorem, for the denominator using the total probability theorem, we get (2.9.44). ■

Customary, (2.9.42), (2.9.43) and (2.9.44) are called the multiplication formula, the total probability formula and Bayes' formula, respectively.

2.9.7 Remark

(1) The linkage between the independence and the real word.

In real world there are a great number of experiments $\mathcal{E}_t$ ($t \in T$) with no relation each other. Every $\mathcal{E}_t$ generates a relation of cause and effect:

$$\text{Experiment } \mathcal{E}_t \text{ (under given conditions } \mathcal{C}_t) \Longrightarrow (\Omega_t, \mathcal{F}_t, P_t) \qquad (2.9.45)$$

Gathering together those experiments, a new experiment $\widetilde{\mathcal{E}}$ is formed. Supposing the $\odot_{t\in T}\Omega_t$ is a true crossed partition (in application the purpose always can be got by selecting Ω_t), then $\widetilde{\mathcal{E}}$ generates a new relation of cause and effect[22]:

$$\widetilde{\mathcal{E}} \text{ (under independent condition group } \mathcal{C}_t, t \in T)$$
$$\Longrightarrow (\prod_{t\in T}\Omega_t, \prod_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t) \qquad (2.9.46)$$

If the $\odot_{t\in T}\Omega_t$ only is a crossed partition, at that time we only can get

$$\widetilde{\mathcal{E}} \text{ (under independent condition group } \mathcal{C}_t, t \in T)$$
$$\Longrightarrow (\bigodot_{t\in T}\Omega_t, \bigotimes_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t) \qquad (2.9.47)$$

Similarly with the argument after Theorem 2.9.6, intuitively we can believe $(\odot_{t\in T}\Omega_t, \otimes_{t\in T}\mathcal{F}_t, \prod_{t\in T}P_t)$ to be similar with an independent product probability space.

Usually people study a big experiment $\mathcal{E}$ in reality. The $\mathcal{E}$ generates a relation of cause and effect:

$$\mathcal{E} \text{ (under condition } \mathcal{C}) \Longrightarrow (\Omega, \mathcal{F}, P) \qquad (2.9.48)$$

The season for $\mathcal{E}$ to be called a big experiment is that it contains many child experiments $\mathcal{E}_t$ ($t \in T^*$). Some of these child experiments may be mutually independent, some others may be depend each other. Suppose mutually independent child experiments are $\mathcal{E}_t$ ($t \in T$, $T \subset T^*$) and $\mathcal{E}_t$ generates probability subspace $(\Omega, \mathcal{F}_t, P)$, then $(\Omega, \mathcal{F}, P)$ contains a subspace $(\Omega, \otimes_{t\in T}\mathcal{F}_t, P)$ to be similar with an independent product probability space. (see Corollaries of Theorems 2.9.1, 2.9.4 and 2.9.6).

The essential of independence is that as long as there exists a family of independent events $\{A_t | t \in T\}$ in $(\Omega, \mathcal{F}, P)$, then $(\Omega, \mathcal{F}, P)$ contains a family of independent sub-tests $\{(\Omega, \mathcal{F}_t) | t \in T\}$ such that $A_t \in \mathcal{F}_t$ ($t \in T$) and a probability subspace $(\Omega, \otimes_{t\in T}\mathcal{F}_t, P)$, therefore $(\Omega, \mathcal{F}, P)$ comes from a big experiment like as (2.9.48).

---

[22]This is the principe of independent experiment. It will be abstracted as the Axiom F_5.

(2) How do we extend the total probability formula?

$P(A|H_i)\,(i \in I)$ in the total probability formula (2.9.43) is some values of probability measure $P(A|\sigma(\mathcal{H}))$. Since A may be any event in $\mathcal{F}$, the total probability theorem can be stated as that the probability measure $P(\mathcal{F})$ is uniquely determined by $P(\sigma(\mathcal{H}))$ and $P(\mathcal{F}|\sigma(\mathcal{H}))$.

In order to declare the great potentialities of the total probability formula, people must find a general formula of total probability which is set up for any sub-test $(\Omega, \mathcal{F}_2)$. It is from (2.9.16) and (2.9.21) that $P(\mathcal{F})$ is uniquely determined by $P(\mathcal{F}_2)$ and $P(\mathcal{F}|\mathcal{F}_2)$, so we believe that there exists a general formula of total probability.

Following discussion has a great inspiration, but not rigorous. Let us use the finest partition $\Omega = \{< \omega > \,|\, \omega \in \Omega\}$ instead of the partition $\mathcal{H} = \{H_i | i \in I\}$, and we write down informally an extension of (2.9.43) as follows:

$$P(A) \sim \sum_{\omega \in \Omega} P(A|\omega) P(< \omega >) \qquad (*)$$

This informal work brings two problems:

① The sample points may not be events. Even if to be events, they may not belong to $\mathcal{F}$ either. At that time $P(< \omega >)$ has no meaning. Even we yield to the point, $< \omega > \in \mathcal{F}$, the set $\{\, \omega \mid P(< \omega >) = 0\}$ may have a positive probability, moreover the value often is one. For example, general K's probability spaces are like this.

② For the sample points mentioned in ①, the conditional probabilities $P(A|\omega)$ has no meaning.

A possible approach to solve those problems is that we use sub-test $(\Omega, \mathcal{F}_2)$ instead of the finest partition Ω, and introduce "the infinitesimal event" $d\omega$ of $\mathcal{F}_2$ and use $P(d\omega)$ instead of $P(< \omega >)$, and believe $P(A|\omega)$ to be a point function $g(\omega)$ which determine uniquely $P(A|\mathcal{F}_2)$. If the three affairs can be realized, then according to the thinking method of differential and integral calculus, we have the hope to obtain

$$P(A) \sim \int_{\Omega} g(\omega) P(d\,\omega) \qquad (**)$$

Indeed the expression $(**)$ is a provable mathematical equality. In fact, it is Radon-Nikodym theorem in the measure theory. Therefore we obtain three gains from informal discussion:

① The expression $(**)$ is a provable mathematical equality;

② The conditional probability may be determined by a point function $g(\omega)$;

③ The expression $(**)$ may be a tool to determine $g(\omega)$.

The three gains just are the source to generate the contents of Section 2.10.

(3) In order to display the probability contents of next section, let us to introduce

Radon-Nikodym theorem. *Let $(\Omega, \mathcal{F}_2, P)$ is a probability space. If $\nu(A), A \in \mathcal{F}_2$ is a set function which satisfies Axiom D_3 and is absolutely continuous with respect to $P(\mathcal{F}_2)$ (i.e. $P(A) = 0$ implying $\nu(A) = 0$), then there exists a finite value $\mathcal{F}_2$-measurable point function $g(\omega), \omega \in \Omega$ such that*

$$\nu(A) = \int_A g(\omega) P(d\,\omega) \qquad (2.9.49)$$

for every $A \in \mathcal{F}_2$. The $g(\omega)$ is unique in the sense of without count zero probability set.

The proof can be found in the reference[16]. In order to avoid unfamiliar terms and symbols, original measure space is confined to probability space.

2.10 Point functions on random test (continued)

In the same way with probability measure, people hope various conditional probabilities can be determined by the point functions. Finding those point functions displays the restriction to be a mathematical operation with intension.

2.10.1 Conditional density $p(A|\mathcal{F}_2)(\omega)$

The conditional probability $P(A|\mathcal{F}_2)$ of the event A given the sub-test $(\Omega, \mathcal{F}_2)$ is a family of conditional probability values. It is difficult to point out the relations among those values from the point of view of set function. But there are close relations among those values indeed, and those relations may be represented by the point functions (see the remark (2) in Subsection 2.9.7).

Definition 2.10.1. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces. Suppose $A \in \mathcal{F}$ and $(\Omega, \mathcal{F}_2)$ is a sub-test. The point function $g(\omega)$, $\omega \in \Omega$ is called a conditional density of the conditional probability $P(A|\mathcal{F}_2)$, if $g(\omega)$ has following properties:*

(1) $g(\omega)$ is a $\mathcal{F}_2$-measurable function;

(2) $P(A|\mathcal{F}_2)$ is determined by $g(\omega)$ using the formula

$$P(A|B) = \frac{1}{P(B)} \int_B g(\omega) P(d\,\omega) \qquad (B \in \mathcal{F}_2^*) \qquad (2.10.1)$$

From now on the conditional density of the conditional probability $P(A|\mathcal{F}_2)$ is denoted by the special symbol $p(A|\mathcal{F}_2)(\omega)$, or $p(A|\mathcal{F}_2)$ customarily, and may also be called as the conditional density of A given sub-test $(\Omega, \mathcal{F}_2)$ (or given σ-subfield $\mathcal{F}_2$).

Theorem 2.10.1. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces. Supposing $P(A|\mathcal{F}_2)$ is the conditional probability of the event A given the sub-test $(\Omega, \mathcal{F}_2)$, then there exists the conditional density $p(A|\mathcal{F}_2)$ of $P(A|\mathcal{F}_2)$. It is unique in the sense of without count zero probability set in $\mathcal{F}_2$.*

Proof : The conditions of the theorem implies that $(\Omega, \mathcal{F}_2, P)$ is a probability space. For any fixed $A \in \mathcal{F}$, it is easy to verify the set function $P(AB), B \in \mathcal{F}_2$ to satisfy Axiom D_3, and to be absolutely continuous with respect to $P(\mathcal{F}_2)$. From Radon-Nikodym theorem we get that there is a $\mathcal{F}_2$-measurable function $g(\omega), \omega \in \Omega$ such that

$$P(AB) = \int_B g(\omega) P(d\,\omega) \qquad (B \in \mathcal{F}_2) \tag{2.10.2}$$

and $g(\omega), \omega \in \Omega$ is unique in the sense of without count zero probability set in $\mathcal{F}_2$. If $B \in \mathcal{F}_2^*$ (i.e. $P(B) > 0$), both side of (2.10.2) are divided by $P(B)$, then the (2.10.1) is obtained. So we have proved the existence and uniqueness of the conditional density, and it is $g(\omega)$. ∎

Corollary. *Supposing $\mathcal{F}_2 = \sigma(\mathcal{H})$, where $\mathcal{H} = \{H_i \mid i \in I\}$ is a partition of Ω and $H_i \in \mathcal{F}$, then the conditional density of $P(A|\mathcal{F}_2)$ is*

$$p(A|\mathcal{F}_2)(\omega) = \sum_{i \geq 1} P(A|H_i) \chi_{H_i}(\omega) \tag{2.10.3}$$

That is to say

$$p(A|\mathcal{F}_2)(\omega) = \begin{cases} P(A|H_1), & \omega \in H_1 \\ \cdots & \cdots \\ P(A|H_i), & \omega \in H_i \\ \cdots & \cdots \end{cases} \tag{2.10.4}$$

where $\chi_{H_i}(\omega)$ is the indicator function of set H_i and we agree upon that $P(A|H_i) = 0$ when $P(H_i) = 0$ $(i \in I)$.

Proof : Since $H_i \in \mathcal{F}_2$ $(i \in I)$, the function defined by (2.10.3) is $\mathcal{F}_2$-measurable. In order to complete the corollary's proof, it is only need to prove that for any $B \in \mathcal{F}_2$, the $p(A|\mathcal{F}_2)(\omega)$ satisfies (2.10.2).

Now substituting (2.10.3) into the right of (2.10.2), since $B = \bigcup_{i \in I} BH_i$

and all BH_i $(i \in I)$ are disjoint each other, we have

$$\begin{aligned}\int_B p(A|\mathcal{F}_2)(\omega)P(d\omega) &= \sum_{i\in I}\int_{BH_i} p(A|\mathcal{F}_2)(\omega)P(d\omega)\\ &= \sum_{i\in I}\int_{BH_i} P(A|H_i)P(d\omega)\\ &= \sum_{i\in I} P(A|H_i)P(BH_i)\end{aligned}$$

It is from the construction of $\mathcal{F}_2$ that either $H_i \subset B$ or $BH_i = \emptyset$. So

$$\begin{aligned}\int_B p(A|\mathcal{F}_2)(\omega)P(d\omega) &= \sum_{i\in I}\frac{P(AH_i)}{P(H_i)}P(BH_i)\\ &= \sum_{i\in I,\, H_i\subset B} P(AH_i)\\ &= P(AB)\end{aligned}$$

■

Example 1. Let $(\Omega_5, \mathcal{F}_5, P)$ is the probability spaces of classical type, where $(\Omega_5, \mathcal{F}_5)$ is the random test of Example 5 in Section 2.4, $P(\mathcal{F}_5)$ is a probability measure of classical type. Let $(\Omega_5, \mathcal{F}_6)$ and $(\Omega_5, \mathcal{F}_7)$ separately are the random tests of Examples 6 and 7 in Section 2.4 . At that time $(\Omega_5, \mathcal{F}_6)$ and $(\Omega_5, \mathcal{F}_7)$ are sub-tests of $(\Omega_5, \mathcal{F}_5)$.

Let $A = \{1, 2, 3\}$ $(A \in \mathcal{F}_5$, but $A \notin \mathcal{F}_6$, $A \notin \mathcal{F}_7)$. Using Corollary of Theorem 2.10.1 we get that the conditional densities of A given $\mathcal{F}_5$, $\mathcal{F}_6$ and $\mathcal{F}_7$ separately are

$$p(A|\mathcal{F}_5):\ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6\\ 1 & 1 & 1 & 0 & 0 & 0\end{pmatrix} \tag{2.10.5}$$

$$p(A|\mathcal{F}_6):\ \begin{pmatrix} 1 & 3 & 5 & 2 & 4 & 6\\ \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix} \tag{2.10.6}$$

$$p(A|\mathcal{F}_7):\ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6\\ \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & \frac{3}{4} & 0 & 0\end{pmatrix} \tag{2.10.7}$$

For the degenerate sub-test $(\Omega_5, \mathcal{F}_0)$, the conditional densities of A given

$\mathcal{F}_0$ is

$$p(A|\mathcal{F}_0): \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (2.10.8)$$

■

Let us point out five points about the new concept.

(1) It has a lot of convenience to use the conditional densities $p(A|\mathcal{F}_2)$ instead of the conditional probability $P(A|\mathcal{F}_2)$. Therefore the conditional densities plays an important role in the modern probability theory.

So far, the probability theory does not introduce our conditional probability $P(A|\mathcal{F}_2)$, so it calls our conditional densities $p(A|\mathcal{F}_2)$ as "the conditional probability" and write as "$P(A|\mathcal{F}_2)$".

Obviously, (2.10.5), (2.10.6), (2.10.7) and (2.10.8) are not the distribution law, neither there is the equality $p(A|\mathcal{F}_2)(\omega) = P(A| <\omega>)$ except (2.10.5), so the original name and symbol are not proper and bring forth ambiguity. In order to press close to essence of the concept, we have used new name and symbol. I hope that they are popular with.

(2) Different with conditional probability, the conditional density has no obvious intuitive illustration. A popular and informal intuitive illustration is that $p(A|\mathcal{F}_2)(\omega)$ is the conditional probability value of the event A, given $<\omega>$ under any performance of the sub-test $(\Omega, \mathcal{F}_2)$[23].

In fact, "given $<\omega>$ under any performance of the sub-test $(\Omega, \mathcal{F}_2)$" implies condition $<\omega> \in \mathcal{F}_2$. At that time, if $P(<\omega>) > 0$, then taking $B = <\omega>$ in (2.10.1) we get

$$P(A\,|<\omega>) = p(A|\mathcal{F}_2)(\omega) \quad (2.10.9)$$

So the intuitive illustration is correct (see (2.10.5)); If $P(<\omega>) = 0$, then the left of (2.10.9) has no meaning, the value of the right is immaterial (since $p(A|\mathcal{F}_2)(\omega)$ may have no meaning on a zero probability set), so we may take over the intuitive illustration.

Please must note that if $<\omega> \notin \mathcal{F}_2$, then the intuitive illustration loses efficacy. In fact, if $<\omega> \notin \mathcal{F}_2$, $<\omega> \in \mathcal{F}$ and $P(<\omega>) > 0$, then

$$P(A\,|<\omega>) \neq p(A|\mathcal{F}_2)(\omega) \quad (2.10.10)$$

in general case (see (2.10.6) – (2.10.8)).

[23]This may be the first expression in written form of the intuitive illustration. The cause may be that there is not the concept of sub-test. In fact, if the intuitive illustration is stated by "$p(A|\mathcal{F}_2)(\omega)$ is the conditional probability value of the event A, given $<\omega>$", then the illustration becomes incorrect (see (2.10.10) and (2.10.11)).

In addition, "under any performance of the sub-test $(\Omega, \mathcal{F}_2)$" in the intuitive illustration must be emphasized. The reason is that if $(\Omega, \mathcal{F}_2)$ and $(\Omega, \mathcal{F}_3)$ are different sub-tests, then the set

$$\{\, \omega \mid p(A|\mathcal{F}_2)(\omega) \neq p(A|\mathcal{F}_3)(\omega) \,\} \tag{2.10.11}$$

often has a positive probability, even probability value with 1 (see (2.10.5) — (2.10.8)).

(3) Suppose $(\Omega, \mathcal{F}_2)$ and $(\Omega, \mathcal{F}_3)$ are sub-tests and $\mathcal{F}_3 \subset \mathcal{F}_2$. From the definition of conditional probability we have

$$P(A|\mathcal{F}_3) = P(A|\mathcal{F}_2) \restriction \mathcal{F}_3 \tag{2.10.12}$$

But their conditional densities $p(A|\mathcal{F}_2)(\omega)$ and $p(A|\mathcal{F}_3)(\omega)$ are completely different functions (for example, taking values, measurability, and so on). The fact declares the restriction to be a mathematical operation with the intension.

(4) When A is an argument, we get a family of conditional probabilities

$$\{\, P(A|\mathcal{F}_2) \mid A \in \mathcal{F}\} \tag{2.10.13}$$

and a family of conditional densities

$$\{\, p(A|\mathcal{F}_2)(\omega) \mid A \in \mathcal{F}\} \tag{2.10.14}$$

For the point functions in the family (2.10.14), we can discussion their values, positive or negative, four fundamental rules, and so on. Using the knowledge of integration from Definition 2.10.1 we deduct

① For any $A \in \mathcal{F}$, we have

$$p(A|\mathcal{F}_2)(\omega) \geq 0 \qquad (P(\mathcal{F}_2) - a.s.) \tag{2.10.15}$$

②

$$p(\Omega|\mathcal{F}_2)(\omega) = 1 \qquad (P(\mathcal{F}_2) - a.s.) \tag{2.10.16}$$

③ For disjoint events A_i $(i \in I)$ in $\mathcal{F}$, we have

$$p(\bigcup_{i \in I} A_i|\mathcal{F}_2)(\omega) = \sum_{i \in I} p(A_i|\mathcal{F}_2)(\omega) \qquad (P(\mathcal{F}_2) - a.s.) \tag{2.10.17}$$

where "$P(\mathcal{F}_2) - a.s$" represents that the equalities and inequalities above may be incorrect or have no meaning on a zero probability set in $\mathcal{F}_2$. For writing convenience, "$P(\mathcal{F}_2) - a.s$" often is omitted.

It should be point out that (2.10.15) – (2.10.17) with the corresponding expressions of Axioms $E_1 - E_3$ are very similar in appearance, but they are different essentially. Here the $p(A|\mathcal{F}_2), p(\Omega|\mathcal{F}_2), p(A_i|\mathcal{F}_2)$, etc. are only the symbols of the function's corresponding laws. Different A, Ω, A_i represent they to be different functions. The A, Ω, A_i can not be considered as arguments mistakenly.

(5) When all various sub-tests may be performed, we get a bigger family of conditional probabilities

$$\{ P(A|\mathcal{F}_2) \mid A \in \mathcal{F},\ \mathcal{F}_2 \subset \mathcal{F},\ \mathcal{F}_2 \text{ is an event } \sigma\text{-field}\} \qquad (2.10.18)$$

and a bigger family of conditional densities

$$\{ p(A|\mathcal{F}_2)(\omega) \mid A \in \mathcal{F},\ \mathcal{F}_2 \subset \mathcal{F},\ \mathcal{F}_2 \text{ is an event } \sigma\text{-field}\} \qquad (2.10.19)$$

The point functions in the family (2.10.19) have many important properties except (2.10.15) – (2.10.17). They play very important roles in modern probability theory. Since those functions are the random variables in Chapter 3, they will be deeply studied after axiomatization of probability theory.

2.10.2 Conditional density-probability $p(A, \omega|\mathcal{F}_2)$

The conclude of Theorem 2.10.1 can also be written as

Theorem 2.10.2. *Let $(\Omega, \mathcal{F}, P)$ is a probability spaces. If $(\Omega, \mathcal{F}_1)$, $(\Omega, \mathcal{F}_2)$ is two sub-tests, then the conditional probability (a family of probabilities measures) $P(\mathcal{F}_1|\mathcal{F}_2)$ and the family of conditional densities*

$$\{ p(A|\mathcal{F}_2)(\omega) \mid A \in \mathcal{F} \} \qquad (2.10.20)$$

are determined mutually and satisfy

$$P(A|B) = \frac{1}{P(B)} \int_B p(A|\mathcal{F}_2)(\omega) P(d\omega) \quad (A \in \mathcal{F}_1,\ B \in \mathcal{F}_2^*) \qquad (2.10.21)$$

Calling again attention to that $p(A|\mathcal{F}_2)$ in $p(A|\mathcal{F}_2)(\omega)$ only is a symbol of corresponding law. The family of functions (2.10.20) can not be be considered as a function of two variables $p(A|\mathcal{F}_2)(\omega)$, $(A, \omega) \in \mathcal{F}_1 \times \Omega$. In fact, since $p(A|\mathcal{F}_2)(\omega)$ is a $\mathcal{F}_2$-measurable function, so its exact analytic expression is

$$p(A|\mathcal{F}_2)(\omega), \qquad \omega \in \Omega \setminus N_A \qquad (2.10.22)$$

where $N_A \in \mathcal{F}_2$ and $P(N_A) = 0$. It follows that (2.10.20) can only generate a set-point function of two variables

$$g(A, \omega) = p(A|\mathcal{F}_2)(\omega), \quad A \in \mathcal{F}_1,\ \omega \in \Omega \setminus \bigcup_{A \in \mathcal{F}_1} N_A \qquad (2.10.23)$$

General speaking, $\mathcal{F}_1$ implies uncountable events. An uncountable union of sets $\bigcup_{A\in\mathcal{F}_1} N_A$ may not be event, even it is an event, its probability may not be zero.

But we hope (2.10.20) can be a set-point function with domain $\mathcal{F}_1 \times \Omega$ and determine uniquely $P(\mathcal{F}_1|\mathcal{F}_2)$ indeed, because this set-point function would bring a lot of convenience for study $P(\mathcal{F}_1|\mathcal{F}_2)$.

Definition 2.10.2. *A set-point function $g(A,\omega)$, $(A,\omega) \in \mathcal{F}_1 \times \Omega$ is called a conditional density-probability of $P(\mathcal{F}_1|\mathcal{F}_2)$, if it has three properties as follows:*

(1) For any fixed A, $g(A,\omega)$ is a $\mathcal{F}_2$-measurable function with respect to ω.

(2) For any fixed ω, $g(A,\omega)$ is a probability measure on $(\Omega,\mathcal{F}_1)$ with respect to A.

(3) $g(A,\omega)$ uniquely determines $P(\mathcal{F}_1|\mathcal{F}_2)$ by the formula,

$$P(A|B) = \frac{1}{P(B)}\int_B g(A,\omega)P(d\omega) \qquad (A \in \mathcal{F}_1, B \in \mathcal{F}_2^*) \qquad (2.10.24)$$

Afterwards we use special symbol $p(A,\omega|\mathcal{F}_2)$ to represent $g(A,\omega)$, $(A,\omega) \in \mathcal{F}_1 \times \Omega$, and also call $p(A,\omega|\mathcal{F}_2)$ as conditional density-probability of the sub-test $(\Omega,\mathcal{F}_1)$ given the sub-test $(\Omega,\mathcal{F}_2)$(or $\mathcal{F}_1$ given $\mathcal{F}_2$)[24]. Different with the conditional density, the conditional density-probability may exist or may not exist.

Theorem 2.10.3. *Let $(\Omega,\mathcal{F},P)$ is a probability spaces. If N_A in (2.10.22) satisfy*

$$\bigcup_{A\in\mathcal{F}_1} N_A \in \mathcal{F}_2\,, \quad P(\bigcup_{A\in\mathcal{F}_1} N_A) = 0 \qquad (2.10.25)$$

then there exists the conditional density-probability $p(A,\omega|\mathcal{F}_2)$ of $P(\mathcal{F}_1|\mathcal{F}_2)$.

Proof : by using (2.10.25), the measurable version (2.10.22) can be reformed as

$$p(A|\mathcal{F}_2)(\omega), \qquad \omega \in \Omega \setminus N \qquad (2.10.26)$$

where $N = \bigcup_{A\in\mathcal{F}_1} N_A$. Therefore we can construct a set-point function of two variables as follows

$$g(A,\omega) = \begin{cases} p(A|\mathcal{F}_2)(\omega) & (A,\omega) \in \mathcal{F}_1 \times \Omega \setminus N \\ \text{assigning any way} & (A,\omega) \in \mathcal{F}_1 \times N \end{cases}$$

[24]The $p(A,\omega|\mathcal{F}_2)$ is called the regular conditional probability in modern probability theory. Since the name which press close with essential of the concept makes the thinking with easy and smooth, we use a new name.

Obviously $g(A, \omega)$ has the property (1) in Definition 2.10.2. The property (3) is got by Theorem 2.10.2. When $\omega \in \Omega \setminus N$, the express (2.10.15) – (2.10.17) become the property (2); When $\omega \in N$, we may demand that the assigning values satisfy the property (2). So the proof is complete. ■

It is easy to point out that if $\mathcal{F}_1$ is a finite set, then the (2.10.25) is set up. People have proved that if the Ω is a separable complete metric space, $\mathcal{F}$ is the Borel σ-field on Ω in probability space $(\Omega, \mathcal{F}, P)$, then for any conditional probability $P(\mathcal{F}_1|\mathcal{F}_2)$, its conditional density-probability $p(A, \omega|\mathcal{F}_2)$ exists always. A kind of special cases of the conclude is

Theorem 2.10.4. *Let $(\Omega, \mathcal{F}, P)$ is discrete type, n or countable dimension Kolmogorov probability spaces. Then for any conditional probability $P(\mathcal{F}_1|\mathcal{F}_2)$, its conditional density-probability $p(A, \omega|\mathcal{F}_2)$, $(A, \omega) \in \mathcal{F}_1 \times \Omega$ exists always.*

The proof can be found in reference [22].

Until now, we have completed the task that the set-argument B in (2.9.21) and (2.9.22) is transformed to the point-argument ω. In the later half of this section we will discuss the problem to transform the set-argument A in (2.9.21) and (2.9.23) to the point-argument ω. Similarly to probability measure, this problem may be solved only in several kind of typical probability spaces.

2.10.3 Point functions on the probability space of discrete type

In this subsection we suppose $(\Omega, \mathcal{F}, P)$ is a probability space of discrete type. Without loss of generality, let

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n, \cdots\} \equiv \{\, \omega_i \,|\, i \in I \,\} \tag{2.10.27}$$

and $P(\mathcal{F})$ is determined by a distribution law $f(\omega), \omega \in \Omega$, where

$$f : \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots \\ f_1 & f_2 & \cdots & f_n & \cdots \end{pmatrix} \tag{2.10.28}$$

We agree upon that our discussion starts from this distribution law f, not from a probability measure $P(\mathcal{F})$. Thought a probability measure may generate a lot of distribution laws sometimes (see Theorem 2.8.1), the concludes here do not change[25].

(1) **The conditional distribution law $f(\omega|B)$, $\omega \in \Omega$**

[25]If $(\Omega, \mathcal{F}, P)$ is standardize beforehand, then the problem of uniqueness does not appeared.

Definition 2.10.3. *Let $(\Omega, \mathcal{F}, P)$ is a probability space of discrete type and (2.10.28) is a distribution law of $P(\mathcal{F})$. If $B \in \mathcal{F}$, $P(B) > 0$, then we call the point function*

$$f(\omega_n|B) = \frac{f_n \chi_B(\omega_n)}{P(B)}, \quad \omega_n \in \Omega \tag{2.10.29}$$

or written as

$$f(\omega|B): \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots \\ \frac{f_1\chi_B(\omega_1)}{P(B)} & \frac{f_2\chi_B(\omega_2)}{P(B)} & \cdots & \frac{f_n\chi_B(\omega_n)}{P(B)} & \cdots \end{pmatrix} \tag{2.10.30}$$

as a conditional distribution law of $f(\omega)$ given B, where $\chi_B(\omega)$ is the indicator function of B and

$$P(B) = \sum_{n:\ \omega_n \in B} f_n \tag{2.10.31}$$

Theorem 2.10.5. *Let $(\Omega, \mathcal{F}, P)$ is a probability space of discrete type and (2.10.28) is a distribution law of $P(\mathcal{F})$. Then for any sub-test $(\Omega, \mathcal{F}_1)$, the conditional probability (measure) $P(\mathcal{F}_1|B)$ is determined by the conditional distribution law $f(\Omega|B)$ using formula*

$$P(A|B) = \sum_{\omega \in A} f(\omega|B) \qquad (A \in \mathcal{F}_1) \tag{2.10.32}$$

Proof : We substitute (2.10.29) in the right of (2.10.32) and obtain

$$\begin{aligned} \sum_{\omega \in A} f(\omega|B) &= \frac{\sum_{n:\ \omega_n \in A} f_n \chi_B(\omega_n)}{P(B)} \\ &= \frac{\sum_{n:\ \omega_n \in AB} f_n}{P(B)} \\ &= \frac{P(AB)}{P(B)} \end{aligned}$$

So (2.10.32) is got and the proof is completed. ■

(2) **The conditional density** $p(A|\mathcal{F}_2)(\omega)$, $\omega \in \Omega$

Suppose $(\Omega, \mathcal{F}_2)$ is a sub-test. It is from Theorems 2.5.3 and 2.5.6 that there are the atomic events H_k^* $(k \geq 1)$ in $\mathcal{F}_2$ such that

$$\mathcal{H} = \{ H_k^* \,|\, k \geq 1 \} \tag{2.10.33}$$

is a partition of Ω and

$$\mathcal{F}_2 = \sigma(\mathcal{H}) \tag{2.10.34}$$

At that time $(\mathcal{H}, \mathcal{F}_2)$ is a standard test. Now we rename all H_i^* with $P(H_i^*) > 0$ as H_i $(i \geq 1)$, and let $N = \bigcup_{j:\, P(H_j^*)=0} H_j^*$. Suppose

$$H_i = \{\omega_{i1}, \omega_{i2}, \cdots\} \tag{2.10.35}$$

$$N = \{\omega_{N1}, \omega_{N2}, \cdots\} \tag{2.10.36}$$

then

$$\Omega = \bigcup_{i\geq 1} H_i \bigcup N = \{\omega_{ij},\, \omega_{Nj} \,|\, i,\, j \geq 1\} \tag{2.10.37}$$

At that time we can rewrite the contribution law (2.10.28) as

$$f: \begin{pmatrix} H_1 & H_2 & \cdots & H_i & \cdots & N \\ \omega_{11}\omega_{12}\cdots & \omega_{21}\omega_{22}\cdots & \cdots & \omega_{i1}\omega_{i2}\cdots & \cdots & \omega_{N1}\omega_{N2}\cdots \\ f_{11}f_{12}\cdots & f_{21}f_{22}\cdots & \cdots & f_{i1}f_{i2}\cdots & \cdots & 0\ \ 0\ \ \cdots \end{pmatrix} \tag{2.10.38}$$

where

$$f_{ij} = P(<\omega_{ij}>); \quad \sum_{j\geq 1} f_{ij} = P(H_i) \ \ (i \geq 1); \quad \sum_{i\geq 1, j\geq 1} f_{ij} = 1 \tag{2.10.39}$$

Theorem 2.10.6. *Let $(\Omega, \mathcal{F}, P)$ is a probability space of discrete type. If $(\Omega, \mathcal{F}_2)$ is a sub-test with the form (2.10.34), then for any $A \in \mathcal{F}$, the conditional density of $P(A|\mathcal{F}_2)$ is*

$$p(A|\mathcal{F}_2)(\omega) = \sum_{i\geq 1} P(A|H_i)\chi_{H_i}(\omega) \qquad (\omega \in \Omega) \tag{2.10.40}$$

That is

$$p(A|\mathcal{F}_2): \begin{pmatrix} H_1 & H_2 & \cdots & H_i & \cdots & N \\ \omega_{11}\omega_{12}\cdots & \omega_{21}\omega_{22}\cdots & \cdots & \omega_{i1}\omega_{i2}\cdots & \cdots & \omega_{N1}\omega_{N2}\cdots \\ P(A|H_1) & P(A|H_2) & \cdots & P(A|H_i) & \cdots & 0\ \ 0\ \ \cdots \end{pmatrix} \tag{2.10.41}$$

where

$$P(A|H_i) = \frac{\sum_{j\,:\,\omega_{ij}\in B} f_{ij}}{\sum_{j\geq 1} f_{ij}} \qquad (i \geq 1) \tag{2.10.42}$$

Proof : Since H_n, $N \in \mathcal{F}_2$ $(n \geq 1)$, the function defined by (2.10.40) is $\mathcal{F}_2$-measurable. Using the (2.10.28) at present the (2.10.1) becomes

$$P(A|B) = \frac{1}{P(B)} \sum_{n\,:\,\omega_n \in B} p(A|\mathcal{F}_2)(\omega_n) f_n \qquad (B \in \mathcal{F}_2) \qquad (2.10.43)$$

In order to complete the proof, it is only need to verify that the point function defined by (2.10.40) satisfies (2.10.43).

Now substituting (2.10.40) into the summation expression of the (2.10.43), we have

$$\begin{aligned}
\sum_{n\,:\,\omega_n \in B} p(A|\mathcal{F}_2)(\omega_n) f_n &= \sum_{n\,:\,\omega_n \in B} [\sum_{i \geq 1} P(A|H_i)\chi_{H_i}(\omega_n)]\, f_n \\
&= \sum_{i \geq 1} [\sum_{n\,:\,\omega_n \in B} \chi_{H_i}(\omega_n) f_n]\, P(A|H_i) \\
&= \sum_{i \geq 1} [\sum_{n\,:\,\omega_n \in BH_i} f_n]\, P(A|H_i) \\
&= \sum_{i \geq 1} P(BH_i) P(A|H_i) \\
&= \sum_{i \geq 1} P(AH_i) \frac{P(BH_i)}{P(H_i)}
\end{aligned}$$

It is from the atomicity of H_i that either $BH_i = H_i$ or $BH_i = \emptyset$. So

$$\begin{aligned}
\sum_{n\,:\,\omega_n \in B} p(A|\mathcal{F}_2)(\omega_n) f_n &= \sum_{H_i \subset B} P(AH_i) \\
&= \sum_{i \geq 1} P(ABH_i) \\
&= P(AB)
\end{aligned}$$

We have proved that the function defined by (2.10.40) satisfies (2.10.43). ■

(3)**The conditional density-probability** $p(A, \omega|\mathcal{F}_2)$, $(A, \omega) \in \mathcal{F}_1 \times \Omega$

Theorem 2.10.7. *Let $(\Omega, \mathcal{F}, P)$ is a probability space of discrete type. If $(\Omega, \mathcal{F}_1)$ and $(\Omega, \mathcal{F}_2)$ are two sub-tests and $(\Omega, \mathcal{F}_2)$ has the form (2.10.34), then $P(\mathcal{F}_1|\mathcal{F}_2)$ has its conditional density-probability $p(A, \omega|\mathcal{F}_2)$, it is* [26]

$$p(A, \omega|\mathcal{F}_2) = \sum_{i \geq 1} P(A|H_i)\chi_{H_i}(\omega) \qquad (A \in \mathcal{F}_1,\ \omega \in \Omega) \qquad (2.10.44)$$

[26] Comparing (2.10.40) and (2.10.44) we get that for any $A \in \mathcal{F}_1$, $\omega \in \Omega$, we have $p(A|\mathcal{F}_2)(\omega) = p(A, \omega|\mathcal{F}_2)$.

Its tabulation form is

$$\left(\begin{array}{c|ccccccc}
 & H_1 & H_2 & \cdots & H_i & \cdots & N & \\
A \in \mathcal{F}_1 & \omega_{11}\omega_{12}\cdots & \omega_{21}\omega_{22}\cdots & \cdots & \omega_{i1}\omega_{i2}\cdots & \cdots & \omega_{N1}\omega_{N2}\cdots & \\
\hline
A_1 & P(A_1|H_1) & P(A_1|H_2) & \cdots & P(A_1|H_i) & \cdots & 0 \quad 0 & \cdots \\
A_2 & P(A_2|H_1) & P(A_2|H_2) & \cdots & P(A_2|H_i) & \cdots & 0 \quad 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\cdots & \cdots \\
A & P(A|H_1) & P(A|H_2) & \cdots & P(A|H_i) & \cdots & 0 \quad 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\cdots & \cdots
\end{array}\right) \tag{2.10.45}$$

where

$$P(A|H_i) = \frac{\sum_{j\,:\,\omega_{ij}\in B} f_{ij}}{\sum_{j\geq 1} f_{ij}} \qquad (i \geq 1) \tag{2.10.46}$$

Proof : The $g(A,\omega)$ denotes the right of (2.10.44). Since $H_n, N \in \mathcal{F}_2$ $(n \geq 1)$, so for any fix A, $g(A,\omega)$ is $\mathcal{F}_2$-measurable. When $\omega \notin N$, it is from Theorem 2.10.6 that $g(A,\omega), A \in \mathcal{F}_1$ is a probability measure; When $\omega \in N$, we do not consider it or by assigning any way values make $g(A,\omega), A \in \mathcal{F}_1$ is a probability measure. In short, $g(A,\omega)$ satisfies the properties ① and ② of Definition 2.10.2. Remarking for any fixed $A \in \mathcal{F}_1$, we have $g(A,\omega) = p(A|\mathcal{F}_2)(\omega)$. It is also from Theorem 2.10.6 that $g(A,\omega)$ satisfies the property ③ of Definition 2.10.2. So $g(A,\omega)$ is the conditional density-probability of $P(\mathcal{F}_1|\mathcal{F}_2)$. The proof is complete. ■

(4) **The conditional density-distribute law** $f(\omega^*,\omega|\mathcal{F}_2)$, $(\omega^*,\omega) \in \Omega\times\Omega$

In this paragraph we show all conditional density-probability $p(A,\omega|\mathcal{F}_2)$, $(A,\omega) \in \mathcal{F}_1 \times \Omega$ ($\mathcal{F}_1$ is any σ-subfield of $\mathcal{F}$) are determined by a point function of two variables.

Definition 2.10.4. *Let $(\Omega,\mathcal{F},P)$ is a probability space of discrete type. Suppose $(\Omega,\mathcal{F}_2)$ is a sub-test with the form (2.10.34), and $P(\mathcal{F})$ has the distribution law (2.10.38). Then we call the point function of two variables*

$$f(\omega^*,\omega|\mathcal{F}_2) = \begin{cases} \dfrac{f_{ij}\chi_{H_i}(\omega)}{\sum_{k\geq 1} f_{ik}} & \omega^* = \omega_{ij},\ \omega \in \Omega\,(i \geq 1, j \geq 1) \\[2ex] 0 & \omega^* \in N,\ \omega \in \Omega \end{cases} \tag{2.10.47}$$

as a conditional density-distribution law of $f(\omega)$ given sub-test $(\Omega,\mathcal{F}_2)$ (or given $\mathcal{F}_2$).

Note. *The tableau format of* $f(\omega^*, \omega|\mathcal{F}_2)$ *is*

$$
f: \left(\begin{array}{c|cccccc}
 & H_1 & H_2 & \cdots & N & & \\
(\omega^*, \omega) & \omega_{11}\omega_{12}\cdots & \omega_{21}\omega_{22}\cdots & \cdots & \omega_{N1}\omega_{N2}\cdots & & \\
\hline
\omega_{11} & & 0 & \cdots & 0 & 0 & \cdots \\
\omega_{12} & \dfrac{f_{1j}}{\sum_{k\geq 1} f_{1k}} & 0 & \cdots & 0 & 0 & \cdots \\
\vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\
\omega_{21} & 0 & & \cdots & 0 & 0 & \cdots \\
\omega_{22} & 0 & \dfrac{f_{2j}}{\sum_{k\geq 1} f_{2k}} & \cdots & 0 & 0 & \cdots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
\cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \cdots
\end{array}\right) \qquad (2.10.48)
$$

Theorem 2.10.8. *The conditional density-distribution law* $f(\omega^*, \omega|\mathcal{F}_2)$ *has the following properties:*

① *For any fixed* ω^*, *it is a* $\mathcal{F}_2$*-measurable function with respect to* ω.

② *For any fixed* ω, *it is a distribution law on* $(\Omega, \mathcal{F})$ *with respect to* ω^*.

③ *For any sub-test* $(\Omega, \mathcal{F}_1)$, *the conditional density-probability* $p(A, \omega|\mathcal{F}_2), (A, \omega) \in \mathcal{F}_1 \times \Omega$ *is determined by it using formula below*

$$p(A, \omega|\mathcal{F}_2) = \sum_{\omega^* \in A} f(\omega^*, \omega|\mathcal{F}_2) \qquad (A \in \mathcal{F}_1, \omega \in \Omega) \qquad (2.10.49)$$

Proof : It is from (2.10.34) and (2.10.47) that $f(\omega^*, \omega|\mathcal{F}_2)$ has the property ①. From (2.10.48) immediately we get the property ② except $\omega \in N$. Since $P(N) = 0$, the ω in N may not be considered, if it is need, then we can make it to have the property ② by assignment values any way.

It is from the footnote 26 and Theorem 2.10.6 that the property ③ is held. ∎

In the probability space of discrete type, the conditional distribution law, the conditional density, the conditional density-probability(measure) and the conditional density- distribute law all have simple and clear expression. They provide reference for research of the other probability spaces.

2.10.4 Point functions on n-dimension Kolmogorov's probability space

For various kind of conditional probabilities, there always exist correspondent point functions determining them in the finite dimension K's probability space . At that time those point functions are the popular function on Euclidean space. The abstract integral in Subsections 2.10.1 and 2.10.2

is transformed to Lebesgue-Stieltjes integral (afterwards it is called as L-S integral). Under very wide conditions L-S integral are the series or B. Riemenn integral. So probability theory may be introduced on the basic of the preliminary differentiation and integration.

In this subsection, $(\mathbf{R}^n, \mathcal{B}^n, P)$ is a n-dimension K's probability space, $F(x_1, x_2, \cdots, x_n)$ is the distribution function of $P(\mathcal{B}^n)$, it is existent and unique.

(1) **The conditional distribution function** $F(x_1, x_2, \cdots, x_n|B)$

We know that for any fixed $B \in \mathcal{B}^n$, $P(B) > 0$, the conditional probability $P(\mathcal{B}^n|B)$ is a probability measure (see Corollary of Lemma 2.9.1). It is from Theorem 2.8.3 that $P(\mathcal{B}^n|B)$ and its distribution function $\widetilde{F}(x_1, x_2, \cdots, x_n)$ are uniquely determined each other. In order to reflect $\widetilde{F}$ to depend on B, $\widetilde{F}$ is reformed as $F(x_1, x_2, \cdots, x_n|B)$. We have

Definition 2.10.5. *We call the distribution function of* $P(\mathcal{B}^n|B)$

$$F(x_1, x_2, \cdots, x_n|B) = P(\prod_{i=1}^{n}(-\infty, x_i] \,|\, B) \quad (x_1, x_2, \cdots, x_n) \in \mathbf{R}^n \tag{2.10.50}$$

as the conditional distribution function of F *given* B, *where* $B \in \mathcal{B}^{n*}$, *i.e.* $B \in \mathcal{B}^n$ *and* $P(B) > 0$.

Theorem 2.10.9. *Let* $(\mathbf{R}^n, \mathcal{B}^n, P)$ *is a* n*-dimension K's probability space. Then for any* $B \in \mathcal{B}^{n*}$ *and sub-test* $(\mathbf{R}^n, \mathcal{B}_1)$, *the conditional probability* $P(\mathcal{B}_1|B)$ *is uniquely determined by* $F(x_1, x_2, \cdots, x_n|B)$ *using formula below*

$$P(A|B) = \int_A d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n|B) \quad (B \in \mathcal{B}_1) \tag{2.10.51}$$

Proof : It is from Corollary of Theorem 2.8.3 that $P(\mathcal{B}^n|B)$ is determined by $F(x_1, x_2, \cdots, x_n|B)$ using formula

$$P(A|B) = \int_A d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n|B) \quad (B \in \mathcal{B}^n)$$

Remarking $P(\mathcal{B}_1|B) = P(\mathcal{B}^n|B) \upharpoonright \mathcal{B}_1$, so the equation above implies (2.10.51). ■

Note that $F(x_1, x_2, \cdots, x_n|B)$ is uniquely determined by $F(x_1, x_2, \cdots, x_n)$ using formula

$$F(x_1, x_2, \cdots, x_n|B) = \frac{1}{P(B)} \int_{B\Pi(\mathbf{x})} d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n) \quad (B \in \mathcal{B}^n) \tag{2.10.52}$$

where $\Pi(\mathbf{x}) = \prod_{i=1}^{n}(-\infty, x_i]$. In fact, we have

$$F(x_1, x_2, \cdots, x_n|B) = P(\prod_{i=1}^{n}(-\infty, x_i] \,|\, B) = \frac{P(B\Pi(\mathbf{x}))}{P(B)}$$
$$= \frac{1}{P(B)} \int_{B\Pi(\mathbf{x})} d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n)$$

(2) **The conditional density** $p(A|\mathcal{B}_2)(y_1, y_2, \cdots, y_n)$

Using L–S integral, Theorem 2.10.1 may be stated in the K's probability space as follows:

Theorem 2.10.10. $(\mathbf{R}^n, \mathcal{B}^n, P)$ *and sub-test* $(\mathbf{R}^n, \mathcal{B}_2)$ *are given. Then for any* $A \in \mathcal{B}^n$, *there exists the conditional density* $p(A|\mathcal{B}_2)(y_1, y_2, \cdots, y_n)$. *That is that* $p(A|\mathcal{B}_2)$ *has the properties as follows:*

① *It is a* $\mathcal{B}_2$*-measurable with respect to argument* $(y_1, y_2, \cdots, y_n)$;

② *For any* $B \in \mathcal{B}_2^*$, *we have*

$$P(A|B) = \frac{1}{P(B)} \int_B p(A|\mathcal{B}_2)(y_1, y_2, \cdots, y_n) d_{y_1} d_{y_2} \cdots d_{y_n} F(y_1, y_2, \cdots, y_n) \tag{2.10.53}$$

and $p(A|\mathcal{B}_2)$ *to satisfy* ① *and* ② *is unique, where* $F(y_1, y_2, \cdots, y_n)$ *is the distribution of* $P(\mathcal{B}^n)$.

(3) **The conditional density-probability** $p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$

Using Theorem 2.10.4, the Theorem 2.10.10 has may be enhanced as

Theorem 2.10.11. $(\mathbf{R}^n, \mathcal{B}^n, P)$ *is given. Then for any sub-tests* $(\mathbf{R}^n, \mathcal{B}_1)$ *and* $(\mathbf{R}^n, \mathcal{B}_2)$, *there exists the conditional density-probability* $p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$ *of the conditional probability* $P(\mathcal{B}_1|\mathcal{B}_2)$. *That is that* $p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$ *has three properties as follows:*

① *For fixed* A, *it is a* $\mathcal{B}_2$*-measurable function with respect to argument* $(y_1, y_2, \cdots, y_n)$.

② *For fixed* $(y_1, y_2, \cdots, y_n)$, *it is a probability measure on* $(\mathbf{R}^n, \mathcal{B}_1)$ *with respect to argument* A.

③ *For any* $A \in \mathcal{B}_1$ *and* $B \in \mathcal{B}_2^*$, *we have*

$$P(A|B) = \frac{1}{P(B)} \int_B p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2) d_{y_1} d_{y_2} \cdots d_{y_n} F(y_1, y_2, \cdots, y_n) \tag{2.10.54}$$

where $F(y_1, y_2, \cdots, y_n)$ *is the distribution function of* $P(\mathcal{B}^n)$.

(4) **The conditional density-distribution function** $F(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$

Taking $\mathcal{B}_1 = \mathcal{B}^n$ in Theorem 2.10.11, we obtain a probability measure $p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2), A \in \mathcal{B}^n$. If we denote its distribution function by useing $F(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$ and remark $(y_1, y_2, \cdots, y_n)$ to be arguments, then we can introduce the following concept.

Definition 2.10.6. *Let $p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$ is a conditional density-probability, Then we call*

$$F(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n|\mathcal{B}_2) = p\,(\prod_{i=1}^{n}(-\infty, x_i]; y_1, y_2, \cdots, y_n|\mathcal{B}_2) \tag{2.10.55}$$

as the conditional density-distribution function of F given sub-test $(\mathbf{R}^n, \mathcal{B}_2)$.

Theorem 2.10.12. *The conditional density-distribution function $F(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n|\mathcal{B}_2)$ has following properties:*

① *For fixed $(x_1, x_2, \cdots, x_n)$, it is a $\mathcal{B}_2$-measurable function with respect to argument $(y_1, y_2, \cdots, y_n)$.*

② *For fixed $(y_1, y_2, \cdots, y_n)$, it is a distribution function with respect to argument $(x_1, x_2, \cdots, x_n)$.*

③ *For any sub-test $(\mathbf{R}^n, \mathcal{B}_1)$, using the formula*

$$\begin{aligned} & p(A; y_1, y_2, \cdots, y_n|\mathcal{B}_2) \\ = & \frac{1}{P(B)} \int_A d_{x_1} d_{x_2} \cdots d_{x_n} F(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n) \\ & (A \in \mathcal{B}_1,\ (y_1, y_2, \cdots, y_n) \in \mathbf{R}^n\ ) \end{aligned} \tag{2.10.56}$$

it uniquely determines the conditional density-probability of $\mathcal{B}_1$ given $\mathcal{B}_2$.

Proof : The properties ① and ② follow from (2.10.55). It is from Corollary of Theorem 2.8.3 that the (2.10.56) is true. ■

2.10.5 Remark

Let $B = \Omega$ in (2.10.1). We get

$$P(A) = \int_{\Omega} p(A|\mathcal{F}_2)(\omega) P(d\,\omega) \qquad (B \in \mathcal{F}_2^*) \tag{2.10.57}$$

The (2.10.57) is the total probability formula in general case.

In fact, if $\mathcal{F}_2$ is $\sigma(\mathcal{H})$ in Theorem 2.9.9, then Corollary of Theorem 2.10.1 declares (2.10.57) becomes common formula of total probability. The (2.10.57) may be illustrated as the extension of (2.9.43) by using the finest partition of Ω instead of the partition $\mathcal{H}$.

2.11 The sixth group of axioms: the group of axioms of probability modelling

It is declared in Section 2.1 that the probability theoretical system may be deducted by using the first – fifth groups of axioms. Therefore those five groups of axioms constitute axiomatic system of formalization.

The probability theory is a subject whose applicability is very forceful. Just as the state in Subsection 2.4.4, the application of probability theory requires people to abstract at first the concerning random phenomena as the research object in theoretical system (for example, the realization of modelling work in (2.7.4), and the work in Section 3.1), afterwards we deduct the research objects and obtain various concludes, finally by those concludes represent the statistical knowledge and statistical laws in the concerning random phenomena.

In order to standardize the modelling work (2.7.4), the work to establish the random tests is discussed in Section 2.4, and the work to establish the probability measure is discussed in this section. We first introduce a group of axioms. This group abstracts the six axioms from the three long-tested principles (the principle of equal likelihood, the frequency illustration and the principle of independent experiment). By using them we can establish common probability spaces, for example, classical type, discrete type, geometric type, independent produce type and K's type probability spaces.

This group of axioms is open. It is our hope that it will be replenished and corrected by better and more effective axioms. The intuitive background of this group has be implied in Section 1.3.

2.11.1 The group of axioms of probability modelling

Suppose $(\Omega, \mathcal{F})$ is a random test, A, $A_i \in \mathcal{F}\,(i = 1, 2, \cdots, n;\ n \geq 2)$ and $\mathcal{A} = \{A_1, A_2, \cdots, A_n\}$ is a partition of A. It is from the partition that $A_i \neq \emptyset\ (i = 1, 2, \cdots, n)$.

Axiom F_1 (Axiom of equal likelihood). *Suppose $P(A)$ is know, and any sub-event of A has not be assigned with probability. If we believe all events $A_1, A_2, \cdots, A_n$ have same occurrent possibility, then they are assigned with probability*

$$P(A_1) = P(A_2) = \cdots = P(A_n) = \frac{1}{n}P(A) \tag{2.11.1}$$

We call this way as equally likely treatment of A with the partition $\mathcal{A}$.

Axiom F_2 (Axiom of assigning values). *Suppose $P(A)$ is know, and any sub-event of A has not be assigned with probability. If we reasonably* *(subjectively or objectively) believe A_i to have probability value $P(A_i)$ $(i = 1, 2, \cdots, n)$ such that*

$$P(A_1) + P(A_2) + \cdots + P(A_n) = P(A) \tag{2.11.2}$$

then we believe "the reason" to be reasonable and acceptable. We call this way as the treatment of assigning values of A with the partition $\mathcal{A}$, and the treatment of assigning values in brief.

Afterwards, every element of $\mathcal{A}$ is called an used partition block, or an used block in brief. Obviously, the equally likely treatment is a special treatment of assigning values.

Axiom F_3 (Axiom of order). *When the axiom of assigning values is used many times, we must arrange the treatments of assigning values in several layers. At 0th layer we can only assign a value to some event A (may be Ω, at that time $P(\Omega) = 1$), at $(k+1)$th layer we can only assign values to some used blocks of kth layers (suppose these blocks are $D_{k_1}, D_{k_2}, \cdots, D_{k_r}$)$(k = 1, 2, 3, \cdots)$. In addition, we suppose that there is positive integral M such that*

$$k_r \leq M \ (k = 1, 2, 3, \cdots); \quad \lim_{k \to \infty} \sum_{i=k_1}^{k_r} P(D_{k_i}) = 0 \tag{2.11.3}$$

The used block D is called terminal if for some k, it is a used block of kth layer, but not one of $(k+1)$th layer. The (2.11.3) ensures that there are only finite or countable terminal blocks, and the assigned probability values obey Axiom D_5

Axiom F_4 (Axiom of geometrical assignment). *If random test is Borel's $(\mathbf{R}^n, \mathcal{B}^n)$, then when the axiom of order is used, we reduce from (2.11.3) to the condition that the assigning value of every used block of kth layer tend to zero as $k \to \infty$.*

Axiom F_5 (Axiom of independent experiments) *Suppose the experiment $\mathcal{E}_j$ generates the probability space $(\Omega_j, \mathcal{F}_j, P_j)$ $(j \in J,\ J$ is any index set). If for any $j \in J$, the results of the experiment $\mathcal{E}_j$ (i.e. the events in the $\mathcal{F}_j$ and their probability) are not effected of any performance of the other experiments, then the joint experiment generates the independent produce probability space $(\prod_{j \in J} \Omega_j, \prod_{j \in J} \mathcal{F}_j, \prod_{j \in J} P_j)$.*

Particularly, suppose the experiment $\mathcal{E}$ generates the probability space $(\Omega, \mathcal{F}, P)$. The $\mathcal{E}^n$ and $\mathcal{E}^\infty$ denote a new experiment generated by $\mathcal{E}$ duplicating independently n times or countable times, respectively, where "independently duplicating" means to satisfy the conditions in Axiom F_5. Then $\mathcal{E}^n$ and $\mathcal{E}^\infty$ generate the probability spaces $(\Omega^n, \mathcal{F}^n, P^n)$ and $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$, respectively.

Axiom F_6 (Axiom of frequency) *Suppose the experiment $\mathcal{E}$ generates the probability space $(\Omega, \mathcal{F}, P)$. If $\mathcal{E}$ is duplicated independently countable* *times, then for any $A \in \mathcal{F}$, in $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$ we have*

$$P^\infty\{\omega \mid \lim_{n\to\infty} \frac{\nu_{nA}(\omega)}{n} = P(A)\} = 1^{27} \tag{2.11.4}$$

where $\omega = (\omega_1, \omega_2, \cdots, \omega_n, \cdots) \in \Omega^\infty$, $\nu_{nA}(\omega)$ is the number of the symbol i satisfies $\omega_i \in A$ $(i = 1, 2, \cdots, n)$.

Intuitively, $\nu_{nA}(\omega)$ is the number of times the event A occurs during front n performance of the experiment $\mathcal{E}$. Therefore $\frac{\nu_{nA}(\omega)}{n}$ is called the frequency of the event A at sample point ω. Axiom F_6 declares that as $n \to \infty$, the limit value of the sequence of frequency $\frac{\nu_{nA}(\omega)}{n}$ exists and equals to the probability of A. Theorem 2.7.8 ensures Axiom F_6 is consistent with front five groups of axioms. Similarly, Corollary of Theorem 2.9.6 guarantees Axiom F_5 to have same consistency.

2.11.2 Modelling of probability space of classical type

Theorem 2.11.1. *Let $(\Omega, \mathcal{F})$ is a standard test of n type, $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$. If people received the equally likely treatment of Ω with the partition $\Omega = \{\{\omega_i\} \mid i = 1, 2, \cdots, n\}$, then Axiom F_1 assigns a classical probability measure $P(\mathcal{F})$ to the test $(\Omega, \mathcal{F})$. Therefore we obtain a probability space of classical type $(\Omega, \mathcal{F}, P)$.*

Proof : We get (2.7.10) by Axiom F_1. The conclude follows from Theorem 2.7.2 and its Corollary. ■

For a standard test of n type in reality, whether we receive the equally likely treatment of Ω with the finest partition would be verified by practice. Usually, people can receive the equally likely treatment for Examples 1, 5 and 8 in Section 2.4. According Mendel's law of genetics, for Example 3 in Section 2.4 we can treat with the equal likelihood, but for Example 4 we can not.

2.11.3 Modelling of probability space of discrete type

The random test $(\Omega, \mathcal{F})$ is give. Suppose the axiom of order is performed, and the sure event Ω is the used block in 0th layer. All terminal used blocks are denoted by $H_1, H_2, \cdots, H_n, \cdots$, and let $H_0 = \Omega \setminus \bigcup_{i\geq 1} H_i$ (it may be

[27] Usually this expresion is also written as

$$\lim_{n\to\infty} \frac{\nu_{nA}(\omega)}{n} = P(A) \qquad (P^\infty - a.s.)$$

empty set). Therefore the $\mathcal{H}$,

$$\mathcal{H} = \{ H_1, H_2, \cdots, H_n, \cdots H_0 \} \tag{2.11.5}$$

is a partition of Ω. Let the f_i denotes the probability value assigned to the event H_i by Axiom F_3. It follows from (2.11.3) that $\sum_{i \geq 1} f_i = P(\Omega) = 1$. That is to say that the axiom of order assigns a distribution law f,

$$f : \begin{pmatrix} H_1 & H_2 & \cdots & H_n & \cdots & H_0 \\ f_1 & f_2 & \cdots & f_n & \cdots & 0 \end{pmatrix} \tag{2.11.6}$$

on the standard test $(\mathcal{H}, \sigma(\mathcal{H}))$ of discrete type. It is from Theorem 2.8.1 that the distribution law f determines a probability measure $P(\mathcal{H})$ on $(\mathcal{H}, \sigma(\mathcal{H}))$. Therefore we have proved

Lemma 2.11.1. *Let $(\Omega, \mathcal{F})$ is a random test. If the axiom of order is performed as above, then we obtain a sub-test $(\Omega, \sigma(\mathcal{H}))$ and a probability measure P on it, i.e. a standard probability space $(\mathcal{H}, \sigma(\mathcal{H}), P)$ of discrete type.*

Example 1. *Suppose a coin has no defect, the desk is smooth, throw is at random. Try to establish a probability measure on the test $(\Omega_9, \mathcal{F}_9)$ of Example 9 in Section 2.4.*

Solution. Since Ω_9 is the sure event, $P(\Omega_9) = 1$. Under supposed conditions people can receive the equally likely treatment of Ω_9 with the partition $\{\omega_1, A_1\}$ and obtain

$$P(\{\omega_1\}) = P(A_1) = \frac{1}{2}P(\Omega) = \frac{1}{2} \tag{2.11.7}$$

where $A_1 = \{\omega_2, \omega_3, \cdots, \omega_\infty)$. Only A_1 is treated in the second layer. People can receive the equally likely treatment of A_1 with the partition $\{\omega_2, A_2\}$ and obtain

$$P(\{\omega_2\}) = P(A_2) = \frac{1}{2}P(A_1) = \frac{1}{4} \tag{2.11.8}$$

where $A_2 = \{\omega_3, \omega_4, \cdots, \omega_\infty)$. Going on like this, only A_{n-1} is treated in the nth layer. People can receive the equally likely treatment of A_{n-1} with the partition $\{\omega_n, A_n\}$ and obtain

$$P(\{\omega_n\}) = P(A_n) = \frac{1}{2}P(A_{n-1}) = \frac{1}{2^n} \tag{2.11.9}$$

where $A_n = \{\omega_{n+1}, \omega_{n+2}, \cdots, \omega_\infty)$.

We can treat any positive integral layer By the mathematical induction. Finally we get the terminal used blocks $\{\omega_i\}\,(i = 1, 2, \cdots, n, \cdots)$ (noting that $\{\omega_\infty\}$ is not included). It is from (2.11.3) and Lemma 2.11.1 that the axiom of order assigns the distribution law

$$f : \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots & \omega_\infty \\ \frac{1}{2} & \frac{1}{4} & \cdots & \frac{1}{2^n} & \cdots & 0 \end{pmatrix} \tag{2.11.10}$$

on the test $(\Omega_9, \mathcal{F}_9)$.

The $P_9(\mathcal{F}_9)$ denotes probability measure generated by f. So we have established requisite probability space $(\Omega_9, \mathcal{F}_9, P_9)$ of this example. ■

Theorem 2.11.2. *Let $(\Omega, \mathcal{F}, P)$ is any probability space of discrete type. If $(\Omega, \mathcal{F})$ is known, then we can establish the probability measure $P(\mathcal{F})$ by the axiom of order.*

Proof : Without loss of generality, let

$$\Omega = \{\omega_1,\ \omega_2,\ \cdots,\ \omega_n,\ \cdots\} \tag{2.11.11}$$

and the distribution law of $P(\mathcal{F})$ is

$$f : \begin{pmatrix} \omega_1 & \omega_2 & \cdots & \omega_n & \cdots \\ f_1 & f_2 & \cdots & f_n & \cdots \end{pmatrix} \tag{2.11.12}$$

Suppose first $(\Omega, \mathcal{F})$ is standard. Then we can construct $P(\mathcal{F})$ by the manner as follows. In the first layer, the Ω is treated by assigning values with the partition $\{\omega_1, A_1\}$ and obtain

$$P(\{\omega_1\}) = f_1 \qquad P(A_1) = 1 - f_1 \tag{2.11.13}$$

where $A_1 = \{\omega_2, \omega_3, \cdots)$.

In the n-th layer, the A_{n-1} is treated by assigning values with the partition $\{\omega_n, A_n\}$ and obtain

$$P(\{\omega_n\}) = f_n \qquad P(A_n) = 1 - \sum_{i=1}^{n} f_i \tag{2.11.14}$$

where $A_n = \{\omega_{n+1}, \omega_{n+2}, \cdots)$. We can treat any positive integral layer by the mathematical induction. The final result is to get the distribution law (2.11.12). Therefore we get the probability measure $P(\mathcal{F})$.

Afterwards suppose $(\Omega, \mathcal{F})$ is not standard. Then we can standardize it by using Theorems 2.5.3 or 2.5.6. Using proved conclude we finish the proof. ■

2.11.4 Modelling of probability space of geometrical type

Let $(D, \mathcal{B}^n(D))$ is a Borel test on D, and μ is the Lebesgue measure. Suppose $0 < \mu(D) < \infty$.

It is from the theory about Lebesgue measure that there exist subsets D_0 and $D_1 = D \setminus D_0$ of D such that $\mu(D_0) = \mu(D_1)$. If people believe that the possibility of occurrence of D_0 and D_1 are equally likely, then we can perform the equally likely treatment of D with the partition $\{D_0, D_1\}$ and obtain

$$P(D_{i_1}) = \frac{1}{2}P(D) = \frac{1}{2} = \frac{\mu(D_{i_1})}{\mu(D)} \qquad (i_1 = 0 \text{ or } 1) \tag{2.11.15}$$

This is the treatment of the first layer in the axiom of order.

In the second layer, we treat two used blocks D_0 and D_1 in same time. Every used block D_{i_1} is treated similarly to D. We get four used blocks $D_{i_1 i_2}(i_1, i_2 = 0 \text{ or } 1)$ and

$$P(D_{i_1 i_2}) = \frac{1}{2}P(D_{i_1}) = \frac{1}{4} = \frac{\mu(D_{i_1 i_2})}{\mu(D)} \qquad (i_1, i_2 = 0 \text{ or } 1) \tag{2.11.16}$$

Suppose the treatment of front $(n-1)$ layers are finished. In the nth layer, we treat 2^{n-1} used blocks $D_{i_1 i_2 \cdots i_{n-1}}(i_1, i_2, \cdots, i_{n-1} = 0 \text{ or } 1)$ in same time. Every used block $D_{i_1 i_2 \cdots i_{n-1}}$ is treated similarly to D. We get 2^n used blocks $D_{i_1 i_2 \cdots i_n}(i_1, i_2, \cdots, i_n = 0 \text{ or } 1)$ and

$$P(D_{i_1 i_2 \cdots i_n}) = \frac{1}{2}P(D_{i_1 i_2 \cdots i_{n-1}}) = \frac{1}{2^n} = \frac{\mu(D_{i_1 i_2 \cdots i_n})}{\mu(D)}$$
$$(i_1, i_2, \cdots, i_n = 0 \text{ or } 1) \tag{2.11.17}$$

Going on like this, using the mathematical induction, Axioms F_3 and F_4 guarantees that we may treat for any positive integral layer.

The $\mathcal{D}$ represents the set of all used blocks after performance of the axiom of order. It is

$$\mathcal{D} = \{D_{i_1 i_2 \cdots i_n} \,|\, n \geq 1;\ i_1, i_2, \cdots, i_n = 0 \text{ or } 1\} \tag{2.11.18}$$

As is known to all, for selected suitably used blocks (for example, if D is a bounded set, then it is only needed that the diameters of all used blocks in the nth layer tend to zero as $n \to \infty$), we have

$$\mathcal{B}^n(D) = \sigma(\mathcal{D}), \quad P(B) = \frac{\mu(B)}{\mu(D)} \quad (B \in \mathcal{B}^n(D)) \tag{2.11.19}$$

Therefore we have proved

Theorem 2.11.3. *Let* $(D, \mathcal{B}^n(D))$ *is a Borel test on* D*, and* $\mu(\mathcal{B}^n(D))$ *is the Lebesgue measure. Suppose* $0 < \mu(D) < \infty$*, then we can establish the geometrical probability measure* $P(\mathcal{B}^n(D))$ *by Axioms* F_1*,* F_3 *and* F_4*, and get a probability space of geometrical type* $(D,\ \mathcal{B}^n(D), P)$*.*

2.11.5 Modelling of n-dimension Kolmogorov's probability space

Theorem 2.11.4. *Let* $(\mathbf{R}^n, \mathcal{B}^n, P)$ *is any* n*-dimension K's probability space. Then we can establish the probability measure* $P(\mathcal{B}^n)$ *by Axioms* $F_1 - F_4$*.*

Proof : We only prove the case of $n = 1$. The proof of $n \neq 1$ is similar unless overelaborate symbols.

Let $F(x)$ is the distribution function of $P(\mathcal{B})$. At first, we suppose that $F(x)$ is a continuous function. At that time, there exists a real number a_1 such that $F(a_1) = \frac{1}{2}$. Let

$$D_0 = (-\infty, a_1], \qquad D_1 = (a_1, +\infty) \tag{2.11.20}$$

If people believe that the possibility of occurrence of D_0 and D_1 are equally likely, then we can perform the equally likely treatment of the sure event $\mathbf{R}$ with the partition $\{D_0, D_1\}$ and obtain

$$P(D_{i_1}) = \frac{1}{2} = F(\beta_{i_1}) - F(\alpha_{i_1}) \qquad (i_1 = 0 \ \textit{or} \ 1) \tag{2.11.21}$$

where $(\alpha_{i_1}, \beta_{i_1}]$ represents the used block D_{i_1}. This is the treatment of the first layer in the axiom of order.

In the second layer, we treat two used blocks D_0 and D_1 in same time. It is from the continuity of $F(x)$ that there are a_{10} and a_{11} such that

$$F(a_{10}) = \frac{1}{4}, \qquad F(a_{11}) = \frac{3}{4} \tag{2.11.22}$$

Let

$$\begin{cases} D_{00} = (-\infty, a_{10}], & D_{01} = (a_{10}, a_1], \\ D_{10} = (a_1, a_{11}], & D_{11} = (a_{11}, +\infty) \end{cases} \tag{2.11.23}$$

If people believe that the possibility of occurrence of D_{i_10} and D_{i_11} ($i_1 = 0$ or 1) are equally likely, then we can perform the equally likely treatment of D_{i_1} with the partition $\{D_{i_10}, D_{i_11}\}$ and obtain

$$P(D_{i_1i_2}) = \frac{1}{2}P(D_{i_1}) = \frac{1}{4} = F(\beta_{i_1i_2}) - F(\alpha_{i_1i_2}) \qquad (i_1,\ i_2 = 0 \ \textit{or} \ 1) \tag{2.11.24}$$

where $(\alpha_{i_1 i_2}, \beta_{i_1 i_2}]$ represents the used block $D_{i_1 i_2}$. This is the treatment of the second layer in the axiom of order.

Suppose the treatment of front $(n-1)$ layers are finished. The used blocks of the $(n-1)$th layer are

$$D_{i_1 i_2 \cdots i_{n-1}} = (\alpha_{i_1 i_2 \cdots i_{n-1}}, \beta_{i_1 i_2 \cdots i_{n-1}}] \quad (i_1, i_2, \cdots, i_{n-1} = 0 \textit{ or } 1) \tag{2.11.25}$$

In the nth layer, we treat 2^{n-1} used blocks $D_{i_1 i_2 \cdots i_{n-1}}$ $(i_1, i_2, \cdots, i_{n-1} = 0 \textit{ or } 1)$ in same time. Every used block $D_{i_1 i_2 \cdots i_{n-1}}$ is treated similarly to D. We get 2^n used blocks $D_{i_1 i_2 \cdots i_n}$ $(i_1, i_2, \cdots, i_n = 0 \textit{ or } 1)$ and

$$P(D_{i_1 i_2 \cdots i_n}) = \frac{1}{2} P(D_{i_1 i_2 \cdots i_{n-1}}) = \frac{1}{2^n} = \frac{\mu(D_{i_1 i_2 \cdots i_n})}{\mu(D)} \quad (i_1, i_2, \cdots, i_n = 0 \textit{ or } 1) \tag{2.11.26}$$

where

$$D_{i_1 i_2 \cdots i_n} = (\alpha_{i_1 i_2 \cdots i_n}, \beta_{i_1 i_2 \cdots i_n}] \quad (i_1, i_2, \cdots, i_n = 0 \textit{ or } 1) \tag{2.11.27}$$

Using the mathematical induction, Axioms F_3 and F_4 guarantees that we may treat for any positive integral layer. Let $\mathcal{D}$ denotes the set which consists of all used blocks, i.e.

$$\mathcal{D} = \{D_{i_1 i_2 \cdots i_n} \mid n \geq 1;\ i_1, i_2, \cdots, i_n = 0 \textit{ or } 1\} \tag{2.11.28}$$

As is known to all, $\mathcal{B} = \sigma(\mathcal{D})$, $P(\mathcal{D})$ assigned above uniquely extended to a probability measure $P(\mathcal{B})$. We would like to prove the distribution function of $P(\mathcal{B})$ to be $F(x)$. i.e.

$$F(x) = P(\,(-\infty, x]\,), \qquad x \in \mathbf{R} \tag{2.11.29}$$

In fact, if x is the end point of an used block, the equation above is the result of equally likely treat and probability addition. For general x, this equation may be deducted from the continuity of $F(x)$ and the continuity from below and above of probability measure.

Afterwards suppose $F(x)$ is a general distribution function. It is from the Lebuegue's resolution theory that

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x) \tag{2.11.30}$$

where $F_1(x)$ is a distribution function of discrete type[28], and $F_2(x)$ is a continuous distribution function, and $\alpha_1, \alpha_2 \geq 0$, $\alpha_1 + \alpha_2 = 1$.

[28] Let Ω^* denotes a set which consists of the discontinuous points of $F(x)$. As is known to all, $F_1(x)$ and some distribution law on Ω^* are uniquely determined each other.

Therefore we may assign the $F_1(x)$ of discrete type on $(\mathbf{R}, \mathcal{B})$ by using Theorem 2.11.2, and the continuous $F_1(x)$ on $(\mathbf{R}, \mathcal{B})$ by using proved conclude. So we have assigned the distribution function $F(x)$ on $(\mathbf{R}, \mathcal{B})$ and obtain the probability measure $P(\mathcal{B})$. ■

2.11.6 Modelling of independent product probability space

Axiom F_5 has given the modelling of independent product probability space. This class of spaces has been used widely in probability theory. The theorem below illustrates why the probability spaces of classical type appear very much.

Theorem 2.11.5. *Suppose the experiment $\mathcal{E}_i$ generates the probability space of classical type $(\Omega^i, \mathcal{F}^i, P^i)\,(i = 1, 2, \cdots, n)$. Suppose All experiment $\mathcal{E}_i\,(i = 1, 2, \cdots, n)$ are performed independently (i.e. obeying the condition of Axiom F_5) and they take form a joint experiment $\mathcal{E}^*$. Then the $(\prod_{i=1}^n \Omega^i, \prod_{i=1}^n \mathcal{F}^i, \prod_{i=1}^n P^i)$ generated by $\mathcal{E}^*$ is a probability space of classical type.*

Proof : We may suppose Ω^i imply r_i elements. Under this supposition we may write down

$$\prod_{i=1}^n \Omega^i = \{(\omega^1, \omega^2, \cdots, \omega^n) \,|\, \omega^i \in \Omega^i,\ i = 1, 2, \cdots, n\} \tag{2.11.31}$$

$$\prod_{i=1}^n \mathcal{F}^i = \{A^1 \times A^2 \times \cdots \times A^n \,|\, A^i \in \mathcal{F}^i,\ i = 1, 2, \cdots, n\} \tag{2.11.32}$$

and $\prod_{i=1}^n \Omega^i$ implies $r_1 r_2 \cdots r_n$ sample points. Writing down product probability measure $\prod_{i=1}^n P^i$ as P in brief, we have

$$\begin{aligned} P(A^1 \times A^2 \times \cdots \times A^n) &= P^1(A^1)P^2(A^2)\cdots P^n(A^n) \\ &= \prod_{i=1}^n \frac{\text{the number of sample points in } A^i \text{ } of \text{ } \Omega^i}{r_i} \\ &= \frac{\text{the number of sample points in } A^1 \times A^2 \times \cdots \times A^n \text{ } of \text{ } \prod_{i=1}^n \Omega^i}{r_1 r_2 \cdots r_n} \end{aligned} \tag{2.11.33}$$

So we have proved the $(\prod_{i=1}^n \Omega^i,\ \prod_{i=1}^n \mathcal{F}^i,\ \prod_{i=1}^n P^i)$ to be a probability space of classical type. ■

2.11.7 Importance of the axiom of frequency

People can not duplicate an experiment $\mathcal{E}$ infinitely in reality. Therefore the equation

$$\lim_{n\to\infty} \frac{\nu_{nA}(\omega)}{n} = P(A) \qquad (P^\infty - a.s.) \tag{2.11.34}$$

in Axiom F_6 can not be used to find the exact value of $P(A)$, and only the approximate value. Even so, people still may model according the manner in Subsection 1.3.5. Mendel's work is one of the most glorious modelling models.

The importance of Axiom F_6 locates

(1) The (2.11.34) is an important tool to understand the abstract concept "probability".

(2) The (2.11.34) is an important tool to verify whether effective is the theoretic system of probability theory.

(3) The (2.11.34) is the base of mathematical statistics. The statistical method has become an important scientific method to discover, establish and verify other scientific theories and scientific laws.

The principle of equal likelihood and the frequency illustration are two great mainstays to support the whole building of probability theory. Therefore the NAS does not regard them as corollaries of front five groups of axioms, furthermore regard them as the parallel axioms, and modelling foundation.

Chapter 3

Introduction of Random Variables

The NAS first introduces the causation space, afterwards introduces two classes of "graphes"—the random tests and the probability spaces, and the tools to study them—the joint test and conditional probability. Therefore standardized probability theory turn to study various "graphes" in the causation space, specially various random tests and probability spaces. The basic task of study can be summarized as

- (1) Finding the random tests in the applications, and to study the relations among the different tests (including various sub-tests).
- (2) Assigning a probability measure to given test $(\Omega, \mathcal{F})$ according the requirements of application problems, people get probability space $(\Omega, \mathcal{F}, P)$ accepted popularly. In many situations only a part of knowledge of $P(\mathcal{F})$ is obtained.
- (3) Under the condition of $(\Omega, \mathcal{F}, P)$ to be given, to find the statistical knowledge implied in the probability space, specially the statistical laws in the knowledge.
- (4) To study the relations among various probability spaces, particularly the statistical knowledge and statistical laws in the relations.

Usually, mathematical statistics lays special emphasis on study the basic tasks (1) and (2); probability theory[1] lays special emphasis on study the

[1]The term "probability theory" can be used in broad and narrow sense. Here it is used in narrow sense. In other place of this book it always is used in broad sense, i.e. the mathematical statistics also is a part of probability theory.

basic tasks (3) and (4).

If we restrict the research objects into Borel tests and Kolmogorov probability spaces, then the basic task can be worked out in the concrete

- ① Finding Borel tests and their various sub-tests needed in applications, and to study the relations among them.
- ② To find a distribution function on given Borel test (i.e. assigning a probability measure to given Borel test) according the requirements of application problems, and to obtain a Kolmogorov probability space required by application problems. In a lot of situations we can only get a part of knowledge of the distribution function.
- ③ To study the statistical knowledge implied in Kolmogorov probability space, specially the statistical laws in the knowledge.
- ④ To study the relations among various Kolmogorov probability spaces, particularly the statistical knowledge and statistical laws in the relations.

The random variables or the families of the random variables are one kind of point functions on probability space $(\Omega, \mathcal{F}, P)$. Because there are K's probability spaces in their range, we can get a lot of statistical knowledge and statistical laws of $(\Omega, \mathcal{F}, P)$ from ones of K's probability space. Customarily, those knowledge and laws are called as the statistical knowledge and statistical laws of the random variable (or the family). That is to say that **the basic tasks (1)–(4) can be transformed to ①–④ in fixed meaning. Therefore the random variables become an important tool to study general probability spaces.**

The very interesting affair is that **even if $(\Omega, \mathcal{F}, P)$ can not written down in the concrete, we may still obtain the statistical knowledge and statistical laws of the random variable (or the family) on $(\Omega, \mathcal{F}, P)$ by using K's probability space.**

On another side, the random phenomena concerned in application often come from some abstract probability space $(\Omega, \mathcal{F}, P)$. In many situations, we can only ensure $(\Omega, \mathcal{F}, P)$ to be exist, but can not write down it in the concrete (for example, the weather forecast experiment of Section 3.1).

How do we study the random phenomena in this situation? For this purpose people introduce the quantified indexes depicting the random phenomena concerned, and turn to study the random phenomena involving only the indexes. For example, the quantified indexes "1, 2, 3, 4, 5, 6" are introduced in the experiment of throw a dice; the temperature, the humidity, ..., are introduced in the experiment of weather forecast, and some new quantified indexes are worked out according requirement. We can ensure (see Section 3.1) that **the quantified indexes are random variables**; the statistical knowledge and statistical laws of random phenomena involving the indexes are ones of random variables. Therefore **the random variables provide an effective method to study quantified indexes, and promote the quantified process in modern science, particularly the social science.**

Being due to both side above, the random variables have become main research objects after the axiomatization of probability theory.

The purpose of this book is the establishment of the NAS, so we can only introduce the random variables in brief. In Section 3.1 we demonstration the certainty for the random variables to become main research objects of probability theory. In Section 3.2 we introduce the concept of random variables and the tools to study them: the distribution function and digital characters. In Sections 3.3 and 3.4 we discuss the relations among the random variables, and introduce the concept of family of random variables and the tools to study them: various distribution functions and digital characters, conditional probabilities and conditional mathematical expectations.

3.1 Intuitive background of random variables

3.1.1 Intuitive background: the situation of probability space to be known

Four players, A, B, C, D play a game. The banker D prepares to play the experiment $\mathcal{E}_5$ of Example 5 in Section 2.4: throw once a material well-distributed dice. A, B, C are gamblers. The gamble rules promised by them and banker, respectively, are (let $a \neq 0$)

	the number to show	1	2	3	4	5	6
ξ	the money wined by A	a	-a	a	-a	a	-a

(3.1.1)

	the number to show	1	2	3	4	5	6
η	the money wined by B	-a	-a	-a	-a	2a	2a

(3.1.2)

	the number to show	1	2	3	4	5	6
ζ	the money wined by C	-3a	-a	-3a	-a	0	2a

(3.1.3)

The values of the money wined by the banker D denoted by θ, then

$$\vartheta = -(\xi + \eta + \zeta) \tag{3.1.4}$$

θ also may be tabulated as

	the number to show	1	2	3	4	5	6
θ	the money wined by D	a	a	a	a	a	a

(3.1.5)

What the four players concern is that will I win or not? and how much will I win? Suppose temporarily they do not concern other things, then they concern only the number for the dice to show and each gamble rule of themselves. Let us focus our analysis on ξ, i.e. examine the thing concerned by gambler A.

(1) The base of gamble to exist is the probability space $(\Omega_5, \mathcal{F}_5, P_5)$.

The premise for the gamble to play is that the banker performs the experiment $\mathcal{E}_5$. The $\mathcal{E}_5$ generates the probability space $(\Omega_5, \mathcal{F}_5, P_5)$, where $(\Omega_5, \mathcal{F}_5)$ is the random test defined by Example 5 in Section 2.4, and $P_5(\mathcal{F}_5)$ is a probability measure of classical type.

(2) ξ is a point function defined on $(\Omega_5, \mathcal{F}_5, P_5)$.

In appearance, the value ξ of money wined by the A is an indefinite value. ξ takes the value according the chances such that if odd point occurs, then A wins the a Yuan; if even point occurs, then A loses the a Yuan.

The great important meaning of the sample space is that due to its existence people discover ξ to be a real value function defined on Ω. In fact, (3.1.1) is its tableau format, its analytic expression is

$$\xi(\omega) = \begin{cases} a, & \omega = 1, 3, 5 \\ -a, & \omega = 2, 4, 6 \end{cases} \tag{3.1.6}$$

(3) The sub-test $(\Omega_5, \sigma(\xi))$ generated by ξ.

The general knowledge declares that the A does not concern all things. The A only concerns four events, they are

$$\{\ \emptyset,\ \{1,3,5\},\ \{2,4,6\},\ \Omega_5\ \} \tag{3.1.7}$$

where Ω_5 and $\emptyset$ represent whether the experiment $\mathcal{E}_5$ performs; the other two events represent, respectively

$$\begin{aligned} \text{“ to win the a Yuan ”} &= \{1, 3, 5\} \\ &= \{\omega \mid \xi(\omega) = a\} \stackrel{def}{=} \{\xi = a\} \end{aligned} \tag{3.1.8}$$

$$\begin{aligned} \text{“ to lose the a Yuan ”} &= \{2, 4, 6\} \\ &= \{\omega \mid \xi(\omega) = -a\} \stackrel{def}{=} \{\xi = -a\} \end{aligned} \tag{3.1.9}$$

It follows that an event has many representations. The left "literal expression" reflects the real content of the event; the other expressions reflect it to be the element in $\mathcal{F}_5$. The expression of the right end is simpler and closer to the "literal expression" than the other two because it has hidden the sample points from readers.

For convenience of extension we restate the general knowledge above as that the gambler A (correctly to say, ξ) concerns the set of bud events to be

$$\mathcal{A} = \{ \ (\xi = a), \ (\xi = -a) \ \} \tag{3.1.10}$$

The all events concerned by him constitute the event σ-field $\sigma(\mathcal{A})$ which is usually denoted by $\sigma(\xi)$. Therefore

$$\sigma(\xi) = \sigma(\mathcal{A}) = \{ \emptyset, \ (\xi = a), \ (\xi = -a), \ \Omega_5 \} \tag{3.1.11}$$

Afterwards $\sigma(\xi)$ is called the event σ-field of ξ. $(\Omega_5, \sigma(\xi))$ is called the sub-test generated by ξ.

(4) The representation probability space and distribution law of ξ.

$(\Omega_5, \sigma(\xi), P_5)$ is called the representation probability space of ξ. From Theorem 2.8.1 we know the probability measure $P_5(\sigma(\xi))$ and the distribution law

$$\begin{pmatrix} (\xi = -a) & (\xi = a) \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \text{ written in brif } \begin{pmatrix} -a & a \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{3.1.12}$$

uniquely determined each other. (3.1.12) is called the distribution law of ξ.

The synthesis of (1)–(4) gets that **the gambler A (correctly to say, ξ) does not concern all statistical knowledge of $(\Omega_5, \mathcal{F}_5, P_5)$. The A only concerns the representation probability space $(\Omega_5, \sigma(\xi), P_5)$.**

Similarly, the gamblers B and C, respectively, only concern representation probability spaces $(\Omega_5, \sigma(\eta), P_5)$ and $(\Omega_5, \sigma(\zeta), P_5)$ of themselves, where $\sigma(\eta)$ and $\sigma(\zeta)$, respectively, are the event σ-fields generated by the bud sets

$$\mathcal{A}_2 = \{ \ (\eta = -a), \ (\eta = 2a) \ \} \tag{3.1.13}$$

$$\mathcal{A}_3 = \{ \ (\zeta = -3a), \ (\zeta = -a), \ (\zeta = 0), \ (\zeta = 2a) \ \} \tag{3.1.14}$$

$P_5(\sigma(\eta))$ and $P_5(\sigma(\zeta))$ are determined separately by the distribution laws

$$\begin{pmatrix} -a & 2a \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \tag{3.1.15}$$

$$\begin{pmatrix} -3a & -a & 0 & 2a \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \tag{3.1.16}$$

Similarly again, the banker D only concerns the representation probability space $(\Omega_5, \sigma(\vartheta), P_5)$ of θ, where $\sigma(\vartheta) = \{\emptyset, \Omega\}$, $P(\sigma(\vartheta))$ is determined by the distribution law

$$\begin{pmatrix} a \\ 1 \end{pmatrix} \tag{3.1.17}$$

It is of interest to note that $(\Omega_5, \sigma(\vartheta), P_5)$ is the degenerate probability space. This reflects the fact in gamble that the banker D does not interest for the number of points of the dice to occur. He only concerns whether the experiment $\mathcal{E}_5$ to be performed. He must win the a Yuan provided performance.

3.1.2 Importance of the distribution law

For simplicity and convenience of the statement, we generalized this kind of variables ξ, η in advance.

Stipulation 1[2]. Let $(\Omega, \mathcal{F}, P)$ is a probability space and $X(\omega), \omega \in \Omega$ is a real value function. The X is called a discrete random variable if X only takes finite or countable values $x_1, x_2, \cdots$, and can generate its representation probability space $(\Omega, \sigma(X), P)$, where

$$\sigma(X) = \sigma\{(X = x_1), (X = x_2), \cdots\} \tag{3.1.18}$$

$P(\sigma(X))$ is determined by the distribution law

$$f: \begin{pmatrix} x_1 & x_2 & \cdots \\ f_1 & f_2 & \cdots \end{pmatrix} \tag{3.1.19}$$

Stipulation 2. We call the value $\sum_{n\geq 1} x_n f_n$ the mathematical expec-

---

[2]Since any concept and symbol can not defined in this section (see the footnote 6 in Section 2.3), so the term "stipulation" is used. In fact, the stipulation can only give the concept's heuristic definition which is not rigorous.

tation of X, write it as EX. We call the value $\sum_{n\geq1}(x_n - EX)^2 f_n$ the variance of X, write it as DX or $var(X)$.

We can construct many values by using the distribution law (3.1.19), and generally call them the digital characters of X. EX and DX are two of the most important digital characters. Particularly, if X has special distribution law

$$f: \quad \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \tag{3.1.20}$$

then the mathematical expectation becomes

$$EX = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{3.1.21}$$

This is the average value of n numbers $x_1, x_2, \cdots, x_n$ in general knowledge. This fact is the source of the mathematical expectation to be used widely.

The distribution law is very important. An embodiment of importance is that even if $(\Omega, \mathcal{F}, P)$ is not known, only the distribution law is known, then the representation probability space $(\Omega, \sigma(X), P)$ of X can be written down by using (3.1.19).

In fact, from the first line of the distribution law (3.1.19) we get that the bud set of X is

$$\mathcal{A} = \{\, (X = x_1),\ (X = x_2), \cdots \} \tag{3.1.22}$$

Therefore the event σ-field $\sigma(X)$ of X is known, the probability measure $P(\sigma(X))$ is determined by the distribution law (3.1.19), Ω is the domain of X (if it can not be get, we can select anyway a sample space by using Lemma 2.4.1). So we have written down the representation probability space $(\Omega, \sigma(X), P)$.

The importance of distribution law locates at that **the distribution law (3.1.19) implies all statistical knowledge of X, and generates the digital characters which are new tools to study X.**

It is easy to point out that for the problem of the gamble of four players in Subsection 3.1.1, if the a takes various values, then ξ defined by (3.1.1) is various, but those different ξ have same representation probability space. Therefore **the statistical knowledge implies by ξ is more abundant than its representation probability space. The additional part comes from ξ's taking values. The reason is those values to give the events in $\sigma(\xi)$ more realistic contents (see (3.1.8) and (3.1.9)).**

How do we represent this part of statistical knowledge? Only the gamblers can help us to solve this question. The clever gamblers may ask a sequence of problems. Are the game rules fair? If they are not fair, who benefits from this game? How mach money does he expect to win in each throw? How much is the difference between the real wined money and the expected one in each throw? and so on.

The clever gamblers answer that the gamble rules ξ and η are fair since the expectation of wined money is zero; the gamble rules ζ and θ are no fair since the expectation of wined money, respectively, are $-a$ and a. So if $a > 0$, the gamble rule ζ does not benefit to the gambler C; the gamble rule ϑ does benefit to the banker D. Where the expectation of wined money is the average values (3.1.21) calculated by using (3.1.1)—(3.1.4).

The probability theory is more clever than the gamblers. It extends the average concept generated in special situation (3.1.20) to general situation (3.1.19), and introduce the concept of mathematical expectation. By using stipulation 2 and (3.1.12), (3.1.15)—(3.1.17) we easily calculate out

$$E\xi = E\eta = 0, \quad E\zeta = -a, \quad E\vartheta = a \tag{3.1.23}$$

This is identical with the experience of gamblers.

Until now, **the probability theory has found the most convenient and simplest tools to represent the statistical knowledge of discrete random variables — the distribution law and the digital characters.**

3.1.3 Intuitive background: the situation of probability space to be not known

We represent the experiment of weather forecast by the $\mathcal{E}$. The weather station declares that the highest temperature is T, the lowest temperature is T_1 every day, and the other weather forecasts relating to the rain, snow, cloudy, clear, and so on. People may ask that how do we understand those indexes of the weather forecast. We would discuss them for the highest temperature to be as example as follows. For exact statement, we may believe T to be the atmospheric temperature of noon 12:00 tomorrow.

In appearance, T is not a certain quantity. We can not forecast its value since it takes different value when atmospheric conditions are different, and atmospheric conditions change randomly. It follows that **people at beginning consider them to be a kind of variable which takes values randomly**, so is called the random variable in brief.

The NAS provides the strong foundation to study the random variables. Let $\mathcal{F}$ is the set which consists of all events generated by the experiment $\mathcal{E}$. From the third group of axioms we know $\mathcal{F}$ to be an event σ-field. From Lemma 2.4.1 we get that there is a living random test $(\Omega, \mathcal{F})$. Furthermore by (2.7.4), we can assign the probability measure $P(\mathcal{F})$ to $(\Omega, \mathcal{F})$. In a word, the experiment $\mathcal{E}$ generates the probability space $(\Omega, \mathcal{F}, P)$.

Now, the difficulty to study is that although there exists $(\Omega, \mathcal{F}, P)$, but it is not able to be written down concretely. The method to overcome the difficulty is that **giving up the hope to obtain all statistical knowledge of $(\Omega, \mathcal{F}, P)$, and we turn to study the statistical knowledge relating only to the index T.** The concrete method is as follows.

(1) The base of performance of weather forecast is the existence of probability space $(\Omega, \mathcal{F}, P)$.

Although we can not write down $(\Omega, \mathcal{F}, P)$, but it still has clear intuitive meaning. The sample point ω represents a kind of weather conditions. **If ω pseudo-occurs, i.e. under this kind of weather conditions, all indexes describing the weather (for example, the temperature, the humidity, the rain , the snow, the cloudy, the clear and so on) take a certain value or a certain state.** The randomness of the weather forecast represents as that all sample points all have the chance to pseudo-occur, and some sample points pseudo-occur with more chances than the others.

The reason for the term "pseudo-occur" to be used is that $\{\omega\}$ usually is not event; even if it is an event but it is with zero probability. If people hope use term "occur" instead of "pseudo-occur", the sentence above should be reformed as follows. The randomness of weather forecast represents as that some events (they are the set of some sample point, i.e. a group of weather conditions) occur with larger probability than the others.

(2) T is a point function defined on $(\Omega, \mathcal{F}, P)$.

Having the sample space Ω, people can say that the index T (i.e. the highest temperature) is a real value point function defined on Ω, i.e.

$$T = T(\omega), \qquad \omega \in \Omega \tag{3.1.24}$$

In the history of the development of probability theory, it is a great progress that the understanding of random variable rose from " the variable to take various values randomly " to " the function defined on the sample space ", because the rising enables the rigorous mathematical theory of the probability to be built at the measure theory.

(3) T generates the sub-test $(\Omega, \mathcal{T})$.

Intuitively, " the temperature T is higher than a, lower than b or equal to b" and "the temperature T is lower than x or equal to x" and so on all are the events in $\mathcal{F}$. It is interesting that although the sample space Ω can not be written out, but the set of sample points forming those events still can be exactly represented as

$$\text{" the temperature } T \text{ is higher than } a, \text{ lower than } b$$
$$\text{or equal to } b\text{"} = \{\,\omega \mid a < T(\omega) \le b\,\} \stackrel{def}{=} \{\,a < T \le b\,\} \qquad (3.1.25)$$

$$\text{"the temperature } T \text{ is lower than } x \text{ or equal to } x\text{"}$$
$$= \{\,\omega \mid T(\omega) \le x\,\} \stackrel{def}{=} \{\,T \le x\,\} \qquad (3.1.26)$$

Let $\sigma(T)$ denotes the event σ-field generated by the event set

$$\mathcal{A} = \{\,(a < T \le b] \mid -\infty < a < b < +\infty\,\} \qquad (3.1.27)$$

The general knowledge supports the inference that $\sigma(T)$ implies all events relating to T, so $\mathcal{A}$ may become the set of bud events of T.

Afterwards, $\sigma(T)$ is called the the event σ-field of T, $(\Omega, \sigma(T))$ is called the sub-test generated by T.

The conclude with great meaning is that **the structure of** $(\Omega, \sigma(T))$ **can be represented by Borel test** $(\mathbf{R}, \mathcal{B})$. In fact, comparing (3.1.27) and (2.5.20), by using Theorem 2.4.3 we get

$$\sigma(T) = \{\,(T \in B) \mid B \in \mathcal{B}\,\} \qquad (3.1.28)$$

where $(T \in B) = \{\,\omega \mid T(\omega) \in \mathcal{B}\,\}$.

(4) T's representation probability space and the value probability space.

$(\Omega, \sigma(T), P)$ is called the representation probability space of T. Therefore when we study the statistical knowledge of T, we need not to concern $(\Omega, \mathcal{F}, P)$, and we need only to study its subspace $(\Omega, \sigma(T), P)$, where the events of $\sigma(T)$ can be listed by (3.1.28).

Using (3.1.28), we introduce on $(\mathbf{R}, \mathcal{B})$ the set function

$$P_T(B) = P(\,T \in B\,), \quad B \in \mathcal{B} \qquad (3.1.29)$$

Using the knowledge of the measure transformation in the measure theory we easily prove $P_T(\mathcal{B})$ to be probability measure[16]. From Theorem 2.8.2 we obtain that there is the distribution function $F_T(x), x \in \mathbf{R}$ such that $P_T(\mathcal{B})$ and $F_T(\mathbf{R})$ are uniquely determined each other by using the formulas

$$F_T(x) = P(\,T \le x\,) \qquad (\,x \in \mathbf{R}\,) \qquad (3.1.30)$$

$$P_T(B) = \int_B dF_T(x) \qquad (\,B \in \mathcal{B}\,) \qquad (3.1.31)$$

Afterwards, we call Kolmogorov probability space $(\mathbf{R}, \mathcal{B}, P_T)$ as the value probability space of T. The $P_T(\mathcal{B})$ is called the probability distribution of T, and $F_T(x)$ is called the distribution function of T.

3.1.4 Importance of the distribution function

The importance of the distribution function is that knowing the distribution function $F_T(x)$, we can find the probability distribution $P_T(\mathcal{B})$ by (3.1.31), so obtain the value probability space $(\mathbf{R}, \mathcal{B}, P_T)$ of T.

The importance of the value probability space $(\mathbf{R}, \mathcal{B}, P_T)$ is that

- ① It determines the representation probability space $(\Omega, \sigma(T), P)$ by (3.1.28) and (3.1.29);
- ② **The value space $(\mathbf{R}, \mathcal{B}, P_T)$ implies all statistical knowledge of T.**

We easily verify that T and $kT\,(k \neq 0, 1)$ have same representation probability space, but different value probability spaces. Therefore T implies more statistical knowledge than the representation probability space. **The additional part of statistical knowledge comes from T's taking value, since those values give the events in $\sigma(T)$ more realistic contents (see (3.1.25) and (3.1.26))**. This is causes which there are different value probability spaces.

In order to represent the additional part of statistical knowledge we need the help of who forecasts the weather. In reality the used weather forecast is

$$\text{The highest atmospheric temperature is } \overline{T} \tag{3.1.32}$$

where $\overline{T}$ is a certain value. But according the discussion above, the exact forecast is

$$\begin{aligned}&\text{The highest atmospheric temperature is a random}\\ &\qquad\text{variable } T \text{ with distribution function } F_T(x)\end{aligned} \tag{3.1.33}$$

In the later forecast all statistical knowledge are included in fact. For example, we can calculate the probability of the events with the form like (3.1.25) and (3.1.26). Supposing we get $P(\,26.5 < T \leq 27.5\,) = 0.9$, the weather station will forecast the highest atmospheric temperature to be in the interval $(26.5, 27.5]$ with probability 90%.

Two methods of the forecast both are rational. The first suits to popular. The second suits to weather researchers. The general knowledge tell us that $\overline{T}$ is the average of T, $|T - \overline{T}|$ is the biased difference generated by the average. In order to understand exactly the average and the biased difference, we need to extend the stipulation 2 as follows.

Stipulation 3. Let T is a random variable, $F_T(x)$ is its distribution function. Then the integral

$$\int_{-\infty}^{+\infty} x dF_T(x)$$

is called the mathematical expectation of T, written as ET; the integral

$$\int_{-\infty}^{+\infty} (x - ET)^2 dF_T(x)$$

is called the variance of T, written as DT or *var(X)*; and $\sqrt{DT}$ is called the standard deviation or the square root difference.

Therefore, $\overline{T}$ in the first forecast method is the mathematical expectation of T; the biased difference $|\overline{T} - T|$ generated by the forecast is closely related $\sqrt{DT}$. As people's scientific knowledge level raises, we may design the third method of the forecast:

$$\text{The highest atmospheric temperature} \\ \text{will be } \overline{T}\text{, its standard deviation will be } \sqrt{DT} \tag{3.1.34}$$

3.1.5 Universality of the random variables in application

Let X denotes the index involving "the rain, snow, cloudy and clear" (i.e. there is no difference among the storm, the drizzle and so on) in weather forecast. Similarly to the discussion getting (3.1.24) we obtain X to be a function $X(\omega), \omega \in \Omega$ which is defined on Ω and its values are the states. If we quantify the states, for example, use the quantitative table

state	rain	snow	cloudy	clear
corresponding value	1	2	3	4

(3.1.35)

then X becomes the discrete random variable defined on $(\Omega, \mathcal{F}, P)$ with distribution law

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ f_1 & f_2 & f_3 & f_4 \end{pmatrix} \tag{3.1.36}$$

Therefore the problem to forecast "the rain, snow, cloudy and clear" can be summed up to find the distribution law (3.1.36). If (3.1.36) is known, the weather station would declare the weather forecast "It will be raining with the probability f_1 tomorrow".

By the way we would point out that if we artificially consider the taking value of X in $\mathbf{R}$, i.e. consider that X may take all real numbers and stipulate the probability of the event

$$\{\omega \,|\, X(\omega) \neq 1, 2, 3, 4\} \tag{3.1.37}$$

to be zero (if X is in the stipulation 1, then $\{\omega \,|\, \omega \neq x_n, n = 1, 2, \cdots\}$ is instead of (3.1.37)), then the discrete random variable becomes a special case of the random variable in Subsection 3.1.3.

3.1.6 Universality of the random variables in theory

The random variable $X(\omega)$ generates the representation probability space $(\Omega, \sigma(X), P)$. The reverse problem is that supposing $(\Omega, \mathcal{F}_0, P)$ is a probability subspace of $(\Omega, \mathcal{F}, P)$, we would like to ask whether there is a random variable $X(\omega)$, $\omega \in \Omega$ such that its representation probability space is the $(\Omega, \mathcal{F}_0, P)$.

It may be proved that if $\mathcal{F}_0$ is a separable event σ-field[3], then the problem has the positive answer[24][16].

3.2 Basic conceptions of random variable

Definition 3.2.1. *Let $(\Omega, \mathcal{F}, P)$ is a probability space. Suppose $X(\omega), \omega \in \Omega$ is a real value function. X is called a random variable on $(\Omega, \mathcal{F}, P)$, if for any real number x,*

$$(X \leq x) \stackrel{def}{=} \{\omega \mid X(\omega) \leq x\} \in \mathcal{F} \tag{3.2.1}$$

Definition 3.2.2. *Let $(\Omega, \mathcal{F}, P)$ is a probability space and $(\mathbf{R}, \mathcal{B})$ is a Borel test. Suppose $X(\omega), \omega \in \Omega$ is a function taking values in $\mathbf{R}$. If for any $B \in \mathcal{B}$,*

$$(X \in B) \stackrel{def}{=} \{\omega \mid X(\omega) \in B\} \in \mathcal{F} \tag{3.2.2}$$

then X is called a random variable on $(\Omega, \mathcal{F}, P)$,

As is known to all, the two definitions are equivalent. The purpose of giving two definitions is for ease to understand the random variable from the two respects. The set consisted of all events in (3.2.1) and (3.2.2) are represent, respectively, by $\mathcal{A}$ and $\sigma(X)$, i.e.

$$\mathcal{A} = \{(X \leq x) \mid x \in \mathbf{R})\tag{3.2.3}$$

$$\sigma(X) = \{(X \in B) \mid B \in \mathcal{B})\tag{3.2.4}$$

So $\sigma(X)$ is an event σ-subfield of $\mathcal{F}$, and $\mathcal{A}$ is its bud event set. According the intuitive background in Section 3.1 and the remark in Subsection 2.8.5, we introduce the definition as follows.

Definition 3.2.3. *The $\sigma(X)$ is called the event σ-field of X. It is the set consisted of all events involving X. The $(\Omega, \sigma(X))$ is called the random test generated by X.*

The research of the random variables include two respects of contents:

[3] An event σ-field is called separable if it has a bud set which consists of countable elements.

- (1) To find the statistical knowledge and statistical laws of X under the conditions of $(\Omega, \mathcal{F}, P)$ to be known or unknown.
- (2) To study the relations among the random variables, in particular, the statistical knowledge and statistical laws implied in the relations.

In this section we will discuss the first term of contents. In the later two sections we will discuss the second term of contents.

3.2.1 Representation probability space and the digital characters

Definition 3.2.4. *The $(\Omega, \sigma(X), P)$ is called the representation probability space of X.*

The statistical knowledge of X includes the representation probability space, and are richer than later since the values taken by X assign more realistic contents to the events in $\sigma(X)$. Therefore people construct new quantities by using taken value by X, and express the additional parts of the statistical knowledge by the new quantities. Those new quantities are called the digital characters of X. Here we only introduce two most important digital characters.

Definition 3.2.5. *Let X is a random variable on $(\Omega, \mathcal{F}, P)$. If X is absolutely integral (i. e. $\int_\Omega |X(\omega)|P(d\omega) < +\infty$), then the $\int_\Omega X(\omega)P(d\omega)$ is called the mathematical expectation (or the expectation or the average), written as EX. i.e.*

$$EX = \int_\Omega X(\omega)P(d\omega) \tag{3.2.5}$$

Definition 3.2.6. *Let X is a random variable on $(\Omega, \mathcal{F}, P)$. If there exists the mathematical expectation of $(X - EX)^2$, then $E(X - EX)^2$ is called the variance of X, written as DX or $var(X)$. i.e.*

$$DX = E(X - EX)^2 = \int_\Omega [X(\omega) - EX]^2 P(d\omega) \tag{3.2.6}$$

and $\sqrt{DX}$ is called the standard deviation or the square root variance.

For any $A \in \mathcal{F}$, $\chi_A(\omega)$ is the indicator function of A. By simple integral calculation we obtain

$$E\chi_A = P(A) \qquad D\chi_A = P(A)P(\overline{A}) \tag{3.2.7}$$

It follows that the probability of the event equals to the mathematical expectation of its indicator function.

The expectation and the variance may reflect many statistical laws of the random variable. One classic result of this respect is following theorem.

Theorem 3.2.1(Chebychev theorem)[4]. *Let X is a random variable on $(\Omega, \mathcal{F}, P)$. If EX and DX exist, then for any $\varepsilon > 0$,*

$$P(\,|X - EX| \geq \varepsilon\,) \leq \frac{DX}{\varepsilon^2} \tag{3.2.8}$$

In particular, taking $\varepsilon = 10\sqrt{DX}$ we have

$$P(\,EX - 10\sqrt{DX} < X < EX + 10\sqrt{DX}\,) \geq 0.99 \tag{3.2.9}$$

So we obtain the statistical law that for any random variable X, we can infer that X will occur into the interval $(\,EX - 10\sqrt{DX}, EX + 10\sqrt{DX}\,)$ with the assurance 99%.

3.2.2 Value probability space and the distribution function

Suppose $(\mathbf{R}, \mathcal{B})$ is the Borel test in Definition 3:2.2. Using (3.2.4) we introduce

$$P_X(B) = P(\,X \in B\,) \quad (\,B \in \mathcal{B}\,) \tag{3.2.10}$$

It is easily proved that $P_X(\mathcal{B})$ is a probability measure. From the theorem 2.8.2 we obtain that there is a distribution function $F_X(x), x \in \mathbf{R}$ such that $F_X(\mathbf{R})$ and $P_X(\mathcal{B})$ are uniquely determined each other by the formulas

$$F_X(x) = P_X(\,(-\infty, x]\,) = P(\,X \leq x\,) \quad (\,x \in \mathbf{R}\,) \tag{3.2.11}$$

$$P_X(B) = \int_B dF_X(x) \qquad (\,B \in \mathcal{B}\,) \tag{3.2.12}$$

Definition 3.2.7. *The $(\mathbf{R}, \mathcal{B}, P_X)$ is called the value probability space of X, and $P_X(\mathcal{B})$ is called the probability distribution of X, and $F_X(\mathbf{R})$ is called the distribution function of X.*

The measurable transformation in the measure theory insures that the properties and the expressions of X on $(\Omega, \mathcal{F}, P)$ and the corresponding ones of $(\mathbf{R}, \mathcal{B}, P_X)$ may mutually transform. i.e. **The value probability space includes all statistical knowledge of the random variable.** In particular, (3.2.5) and (3.2.6) become, respectively,

$$EX = \int_{\mathbf{R}} x P_X(dx) = \int_{\mathbf{R}} x\, dF_X(x) \tag{3.2.13}$$

$$DX = \int_{\mathbf{R}} (x - EX)^2 P_X(dx) = \int_{\mathbf{R}} (x - EX)^2 dF_X(x) \tag{3.2.14}$$

[4]This chapter is simple introduction, the proofs of all theorems are omitted.

in the value probability space, where the right is Lebesgue-Stieltjes integral.

Therefore, we have found the most convenient and the simplest tools to study the random variable — the distribution function and the digital characters generated by it. **The distribution function can represent all statistical knowledge and statistical laws of the random variable.**

The distribution function and the value probability space have the very important roles in the theory and application. The reasons are as follows.

- (1) To study the random variables we need not rigidly adhere to writing down the probability space $(\Omega, \mathcal{F}, P)$ defining them. The thing has the great useful value, since the $(\Omega, \mathcal{F}, P)$ can usually not be written down in the application.

  In fact, suppose the distribution function $F_X(x)$ of X is known, then the value probability space $(\mathbf{R}, \mathcal{B}, P_X)$ is known, too. Therefor the event σ-field $\sigma(X)$ may completely be represented by (3.2.4), the probability distribution $P(\sigma(X))$ may be got by (3.2.10). So in the meaning of Lemma 2.4.1 the representation probability space $(\Omega, \sigma(X), P)$ is got. Lastly we can use $(\Omega, \sigma(X), P)$ instead of $(\Omega, \mathcal{F}, P)$ without loss of any statistical knowledge of X.

- (2) It transforms the formulas in the measure theory (for example, ones in Subsection 3.2.1) to ones in the real variable functions. The thing brings much convenience for probability calculation. For example, we calculate the probability, the expectation, the variance usually by using (3.2.12)-(3.2.14).

- (3) The Lebesgue-Stieltjes integral is the series and Riemann integral under very wide conditions. If the random variables are restricted as discrete type and continuous type (they can satisfy the large part of the requirements in the application), the formulas in this section become ones in form of series or Riemann integral. So it need only the primary differentiation and integration that we study the random variables in the value probability space and get many properties of X.

In a word, **we represent the statistical knowledge and statistical laws of X by using the formulas in the value probability space, but the probability meaning of the formulas is illustrated in the representation probability space.**

3.2.3 Remark: probability analytic method

We will extent the contents in this section to a family of random variables. For this purpose let us talk the problem of the research method, not going into detail.

Subsection 3.2.1 is a direct study for the random variable $X(\omega)$ in the probability space $(\Omega, \mathcal{F}, P)$; Subsection 3.2.2 is a indirect study for the random variable $X(\omega)$ by the tools of the distribution function $F_X(x)$ in the value probability space $(\mathbf{R}, \mathcal{B}, P_X)$. So the former is called the probability method, the later is called the analytic method.

Since for extension, we tentatively agree that the discussion about X also is suitable the family of random variable $X_T \stackrel{def}{=} \{ X_t(\omega) \mid t \in T \}$ (in fact, this is not complete right conclude), and for clear intuitive meaning we suppose the index set T is the time $(-\infty, +\infty)$.

(1) **The probability method**

This is the direct study for X_T in the probability space $(\Omega, \mathcal{F}, P)$. Since X_T is the existing quantity in reality with clear intuitive background, so the concepts, the expressions, mathematical calculation and conclusions all have clear intuitive appearance, and are easily understood. For example,

① The representation probability space $(\Omega, \sigma(X_T), P)$ represents all events relating to X_T and their probabilities.

② The expressions (3.2.11), (3.2.5) and (3.2.6) represent, respectively, the distribution function, the expectation and the variance of single random variable in X_T. Regarding them as a whole, we get the families of the distribution functions of X_T, the expectation functions and the variance functions defined on T.

③ According the intuitive background, people concern the moment T_G arriving first the area G by X_T, i.e.

$$T_G(\omega) = \inf\{ t \mid X_t(\omega) \in G, t > 0 \} \tag{3.2.15}$$

T_G is called the hitting times of G by X_T. Similarly,

$$\varepsilon_G(\omega) = \sup\{ t \mid X_t(\omega) \in G \} \tag{3.2.16}$$

is called the last hitting times of G by X_T. It is the last moment for X_T to leave from G.

Obviously, both of the theory and the application require the statistical knowledge of T_G and ε_G.

④ The random variables (for example, the T_G and ε_G) determined by X_T are called the random functional. Try to find the random functionals with the applied value in practice, and to study their statistical knowledge.

⑤ To inquire into the relations between X_t and t. In particular, whether X_t is continuous, differentiate, integrable with respect to t? How do we understand the concept of the continuity, the differentiability and integrability?

The advantages of probability method are the intuition and the appearance. They promote people to put forward continuously new problems, new tasks, to find new method. So they promote the probability theory to be filled with the vitality and the vigor and to have wide applications.

The character of probability method is that in order to theorize attributes with the intuitive appearance and their variation, we need to transform the intuitive to the formulas, the variations to the calculation for X_T. So it is necessary to be familiar with the measure theory.

(2) **The analytic method**

The essential of the analytic method is that the solution is divide into three steps:

- (1) The probability problems are transformed to the analytic problems about distribution laws or distribution functions according the intuitive background.
- (2) The answer of the problems is found by using the analytic method.
- (3) The analytic answers are transformed to the probability illustration and conclusions in the representation probability space according the intuitive background.

The advantage of the analytic method is to study probability theory by using the mature analytic tools. In particular, under the base of elemental differentiation and integration we can just deeply study the contents of probability theory with wide applications.

The distribution function is not the quantity in reality. It is the introduced tool for people to represent the statistical knowledge of random variable. So the two transformations (1) and (3) require people to understand and grasp the intuitive background of probability theory. In history, before the sample space and the measure theory to occur, the random variable is a ambiguous quantity with the difficulty to be defined. The scientists of that time studied the random variables by using the distribution functions only relying the acute intuitive of probability theory, and obtained the great deal of the results, had formed "the analytic probability theory" which is called by modern people.

The modern probability theory studies the random variables by using the probability analytic method. That is to say that the probability theory is studied by the both of analytic method and probability method with supplement and complement each other, the two methods are made most effects. So the probability theory is developed vigorously.

The analytic mathematics also need to develop continuously. The blend of the two methods not only the probability theory can be studied by using the analytic method but also the analytic theory can be studied by using the probability method.

Now there are some works which new results in the analysis are discovered and proved by the probability method. They open up a new area for the analytic mathematics.

3.3 Basic conceptions of random vector

Definition 3.3.1. *Let* $(\Omega, \mathcal{F}, P)$ *is a probability space. Suppose* $\mathbf{X}(\omega) \overset{def}{=} (X_1(\omega), X_2(\omega), \cdots, X_n(\omega)), \omega \in \Omega$ *is a vector value function.* $\mathbf{X}$ *is called a n-dimensional random vector on* $(\Omega, \mathcal{F}, P)$, *if its every component is a random variable. i.e. for any real number* x_i,

$$(X_i \le x_i) \overset{def}{=} \{\omega \,|\, X_i(\omega) \le x_i\} \in \mathcal{F} \quad (i = 1, 2, \cdots, n) \tag{3.3.1}$$

It is easily verified that the condition (3.3.1) is equivalent to the condition represented by the cylinder events: if for any $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$,

$$\begin{aligned}(\mathbf{X} \le \mathbf{x}) &\overset{def}{=} (X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n) \\ &= \{\omega | X_1(\omega) \le x_1, X_2(\omega), \cdots, X_n(\omega) \le x_n\} \in \mathcal{F}\end{aligned} \tag{3.3.2}$$

Definition 3.3.2. *Let* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $(\mathbf{R}^n, \mathcal{B}^n)$ *is a n-dimension Borel test. Suppose* $\mathbf{X}(\omega) \overset{def}{=} (X_1(\omega), X_2(\omega), \cdots, X_n(\omega)), \omega \in \Omega$ *is a function taking values in* $\mathbf{R}^n$. *If for any* $B \in \mathcal{B}^n$,

$$(\mathbf{X} \in B) \overset{def}{=} \{\omega \,|\, X(\omega) \in B\} \in \mathcal{F} \tag{3.3.3}$$

then $\mathbf{X}$ *is called a n-dimensional random vector on* $(\Omega, \mathcal{F}, P)$.

As is known to all, the two definitions are equivalence. The set consisted of all events of (3.3.2) and (3.3.3) are represent, respectively, by $\mathcal{A}_n$ and $\sigma(X)$, i.e.

$$\mathcal{A}_n = \{(\mathbf{X} \le \mathbf{x}) \,|\, \mathbf{x} \in \mathbf{R}^n) \tag{3.3.4}$$

$$\sigma(\mathbf{X}) = \{(\mathbf{X} \in B) \,|\, B \in \mathcal{B}^n) \tag{3.3.5}$$

So $\sigma(\mathbf{X})$ is an event σ-subfield of $\mathcal{F}$, and $\mathcal{A}_n$ is its bud event set. Similarly to situation of the random variable, we introduce the definition as follows.

Definition 3.3.3. *The* $\sigma(\mathbf{X})$ *is called the event* σ*-field of* $\mathbf{X}$. *It is the set consisted of all events involving* $\mathbf{X}$. *The* $(\Omega, \sigma(\mathbf{X}))$ *is called the random test generated by* $\mathbf{X}$.

Now we can illustrate why two definitions are introduced. Comparing Definitions 3.3.2 and 3.2.2 we can guess that regarding $\mathbf{X}$ as a whole the research methods and conclusions in Section 3.2 may probably be extended to the random vector. This idea introduces the contents of Subsections 3.3.1 and 3.3.2. Comparing Definitions 3.3.1 and 3.2.1 we can point out that it is convenient to study the relations among n components $X_1, X_2, \cdots, X_n$ by this definition. This idea introduces the contents of Subsection 3.3.3 – 3.3.7.

3.3.1 Representation probability space and the digital characters

Definition 3.3.4 *The* $(\Omega, \sigma(\mathbf{X}), P)$ *is called the representation probability space of* $\mathbf{X}$.

Definition 3.3.5. *Let* $\mathbf{X}$ *is a random vector on* $(\Omega, \mathcal{F}, P)$. *If the integral* $\int_\Omega \mathbf{X}(\omega)P(d\omega)$ *exists*[5], *then it is called the mathematical expectation (or the expectation vector, or the average), written as* $E\mathbf{X}$. *i.e.*

$$E\mathbf{X} = \int_\Omega \mathbf{X}(\omega)P(d\omega) \tag{3.3.6}$$

Definition 3.3.6. *Let* $\mathbf{X}$ *is a random vector on* $(\Omega, \mathcal{F}, P)$. *If the integral* $\int_\Omega [\mathbf{X}(\omega) - E\mathbf{X}]^T[\mathbf{X}(\omega) - E\mathbf{X}]P(d\omega)$ *exists*[6], *then it is called the covariance matrix of* $\mathbf{X}$, *written as* $Cov(\mathbf{X}, \mathbf{X})$. *i.e.*

$$\begin{aligned} Cov(\mathbf{X}, \mathbf{X}) &= E[\mathbf{X}(\omega) - E\mathbf{X}]^T[\mathbf{X}(\omega) - E\mathbf{X}] \\ &= \int_\Omega [\mathbf{X}(\omega) - E\mathbf{X}]^T[\mathbf{X}(\omega) - E\mathbf{X}]P(d\omega) \end{aligned} \tag{3.3.7}$$

From the integral of the vector function we know that $E\mathbf{X}$ to be a value vector, i.e.

$$E\mathbf{X} = (EX_1, EX_2, \cdots, EX_n) \tag{3.3.8}$$

From the integral of the matrix function we know that Cov($\mathbf{X}$,$\mathbf{X}$) to be a

[5]The vector integral $\int_\Omega \mathbf{X}(\omega)P(d\omega)$ is defined as $(\int_\Omega X_1(\omega)P(d\omega),\ \int_\Omega X_2(\omega)P(d\omega),\ \cdots,\ \int_\Omega X_n(\omega)P(d\omega))$. The existence of the integral means its every component to be absolutely integrable.

[6]The integral of the matrix is defined as the integral of its every components. The existence of the integral means its every component to be absolutely integrable.

$n \times n$ value matrix. It is

$$Cov(\mathbf{X},\mathbf{X}) = E[\mathbf{X}(\omega) - E\mathbf{X}]^T[\mathbf{X}(\omega) - E\mathbf{X}]$$

$$= \begin{pmatrix} Cov(X_1,X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Cov(X_2,X_2) & \cdots & Cov(X_2,X_n) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Cov(X_n,X_n) \end{pmatrix} \quad (3.3.9)$$

where

$$Cov(X_i,X_j) = E(X_i - EX_i)(X_j - EX_j)$$

$$= \int_\Omega [X_i(\omega) - EX_i][X_j(\omega) - EX_j]P(d\omega)$$

$$(i,j = 1,2,\cdots,n) \quad (3.3.10)$$

We call $Cov(X_i,X_j)$ as the covariance of X_i and X_j. If $i = j$, then $Cov(X_i,X_j) = DX_i$.

3.3.2 Value probability space and the distribution function

Suppose $(\mathbf{R}^n,\mathcal{B}^n)$ is the n-dimension Borel test in Definition 3.3.2. Using (3.3.5) we introduce a set function

$$P_X(B) = P(X \in B) \quad (B \in \mathcal{B}^n) \quad (3.3.11)$$

on the $(\mathbf{R}^n,\mathcal{B}^n)$. It is easily proved that $P_X(\mathcal{B}^n)$ is a probability measure. From Theorem 2.8.3 we obtain that there is distribution function of n variables $F_\mathbf{X}(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^n$ such that $F_\mathbf{X}(\mathbf{R}^n)$ and $P_\mathbf{X}(\mathcal{B}^n)$ are uniquely determined each other by the formulas

$$F_\mathbf{X}(\mathbf{x}) = P_\mathbf{X}(\prod_{i=1}^n(-\infty,x_i]) = P(\mathbf{X} \le \mathbf{x}) \quad (\mathbf{x} \in \mathbf{R}^n) \quad (3.3.12)$$

$$P_\mathbf{X}(B) = \int_B dF_\mathbf{X}(\mathbf{x}) \qquad (B \in \mathcal{B}^n) \quad (3.3.13)$$

where $\mathbf{x} = (x_1,x_2,\cdots,x_n)$.

Definition 3.3.7. *The $(\mathbf{R}^n,\mathcal{B}^n,P_X)$ is called the value probability space of $\mathbf{X}$, and $P_\mathbf{X}(\mathcal{B}^n)$ is called the probability distribution of $\mathbf{X}$, and $F_\mathbf{X}(\mathbf{R}^n)$ is called the distribution function of $\mathbf{X}$, or n dimensional joint distribution function of $X_1,X_2,\cdots,X_n$.*

The measurable transformation in the measure theory insures that the properties and the expressions of $\mathbf{X}$ on $(\Omega, \mathcal{F}, P)$ and the corresponding ones of $(\mathbf{R}^n, \mathcal{B}^n, P_X)$ may mutually transform. i.e. **The value probability space includes all statistical knowledge of the random vector.** In particular, the expectation vector (3.3.6) and covariance matrix (3.3.7) become, respectively,

$$EX = \int_{\mathbf{R}^n} \mathbf{x} P_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbf{R}^n} \mathbf{x} dF_{\mathbf{X}}(\mathbf{x}) \tag{3.3.14}$$

$$\begin{aligned} Cov(\mathbf{X}, \mathbf{X}) &= \int_{\mathbf{R}^n} (\mathbf{x} - E\mathbf{X})^T (\mathbf{x} - E\mathbf{X}) P_{\mathbf{X}}(d\mathbf{x}) \\ &= \int_{\mathbf{R}^n} (\mathbf{x} - E\mathbf{X})^T (\mathbf{x} - E\mathbf{X}) dF_{\mathbf{X}}(\mathbf{x}) \end{aligned} \tag{3.3.15}$$

where the right is n dimensional Lebesgue-Stieltjes integral.

Therefore, we have found the most convenient and the simplest tools to study the random vectors—the distribution function and the digital characters generated by it. **The distribution function can represent all statistical knowledge and statistical laws of the random vector.**

The distribution function and the value probability space have the very important roles in the theory and application. The reasons stated for the random variable in Subsection 3.2.2 also is suitable to the random vector.

3.3.3 Random sub-vectors and the marginal distribution functions

The random vector $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is given. We take any positive integers $1 \le i_1 < i_2 < \cdots < i_r \le n, \quad 1 \le r \le n$ and let $\xi \stackrel{def}{=} (X_{i_1}, X_{i_2}, \cdots, X_{i_r})$. We call ξ as r dimensional sub-vector of $\mathbf{X}$.

Definition 3.3.8. *The value probability space $(\mathbf{R}^r, \mathcal{B}^r, P_\xi)$ of ξ is called r dimensional marginal value (probability) space of $\mathbf{X}$. The probability distribution $P_\xi(\mathcal{B}^r)$ of ξ is called r dimensional marginal probability distribution of $\mathbf{X}$. The distribution function $F_\xi(x_{i_1}, x_{i_2}, \cdots, x_{i_r})$ of ξ is called r dimensional marginal distribution function of $\mathbf{X}$.*

For holding simplicity of symbols and without loss of generality, we use $\xi = (X_1, X_2, \cdots, X_r)$ as the representation of the sub-vectors. For any $B \in \mathcal{B}^r$, we construct a new set

$$\begin{aligned} B^* = B \times \prod_{i=r+1}^{n} \mathbf{R} = \{ (x_1, x_2, \cdots, x_n) \,|\, (x_1, \\ x_2, \cdots, x_r) \in B,\ x_{r+1}, x_{r+2}, \cdots, x_n \in \mathbf{R} \} \end{aligned} \tag{3.3.16}$$

Obviously $B^* \in \mathcal{B}^n$, and $(\xi \in \mathcal{B}^r)$ and $(X \in B^*)$ represent the same event. Therefore

$$P_\xi(B) = P(\xi \in B) = P(\mathbf{X} \in B^*) = P_{\mathbf{X}}(B^*) \tag{3.3.17}$$

In particular, taking $B = \prod_{i=1}^r (-\infty, x_i]$, it becomes

Theorem 3.3.1. *The marginal distribution functions of* $\mathbf{X}$ *are determined only by the joint distribution function* $F_{\mathbf{X}}(\mathbf{x})$. *In particular,*

$$F_\xi(x_1, x_2, \cdots, x_r) = \lim_{x_{r+1}, x_{r+2}, \cdots, x_n \to +\infty} F_{\mathbf{X}}(x_1, x_2, \cdots, x_n) \tag{3.3.18}$$

By using the marginal distribution function we may simplify many expressions relating to $F_{\mathbf{X}}(\mathbf{x})$. For example, the expressions of components of (3.3.14) and (3.3.15) are, respectively

$$EX_i = \int_{\mathbf{R}^n} x_i dF_{\mathbf{X}}(\mathbf{x}) \qquad (i = 1, 2, \cdots, n) \tag{3.3.19}$$

$$Cov(X_i, X_j) = \int_{\mathbf{R}^n} (x_i - EX_i)(x_j - EX_j)\, dF_{\mathbf{X}}(\mathbf{x}) \qquad (i, j = 1, 2, \cdots, n) \tag{3.3.20}$$

Now they are be simplified as

$$EX_i = \int_{\mathbf{R}} x_i dF_{X_i}(x_i) \qquad (i = 1, 2, \cdots, n) \tag{3.3.21}$$

$$Cov(X_i, X_j) = \int_{\mathbf{R}^2} (x_i - EX_i)(x_j - EX_j)\, dF_{(X_i, X_j)}(x_i, x_j) \qquad (i, j = 1, 2, \cdots, n) \tag{3.3.22}$$

3.3.4 Independence

Definition 3.3.9. *Suppose* $\mathbf{X_1}, \mathbf{X_2}, \cdots \mathbf{X_m}$ *are* m *random vectors on* $(\Omega, \mathcal{F}, P)$. $\mathbf{X_1}, \mathbf{X_2}, \cdots \mathbf{X_m}$ *are called mutually independent, if the random subtests* $(\Omega, \sigma(\mathbf{X}_i))(i = 1, 2, \cdots, m)$ *are mutually independent.*

It is from Theorems 2.9.3 and 3.3.1 that

Theorem 3.3.2. *A necessary and sufficient condition of the random vectors* $\mathbf{X_1}, \mathbf{X_2}, \cdots \mathbf{X_m}$ *to be mutually independent is*

$$F_{(\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_m})}(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_m}) = \prod_{i=1}^{m} F_{\mathbf{X_i}}(\mathbf{x_i}) \tag{3.3.23}$$

In particular, if $\mathbf{X_1}, \mathbf{X_2}, \cdots \mathbf{X_m}$ all degenerate to the random variables $X_1, X_2, \cdots X_m$, then we obtain the concept of m random variables to be mutually independent.

3.3.5 Various conditional probabilities

The fifth group of axioms introduce four kinds of the conditional probabilities

$$P(\mathcal{F}_1|\mathcal{F}_2), \quad P(C|\mathcal{F}_2), \quad P(\mathcal{F}_1|D), \quad P(C|D) \tag{3.3.24}$$

on $(\Omega, \mathcal{F}, P)$, where $C \in \mathcal{F}$, $D \in \mathcal{F}^*$, $\mathcal{F}_1$ and $\mathcal{F}_2$ are the event σ-subfields of $\mathcal{F}$.

Suppose $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_m)$ are random vectors on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_1 = \sigma(\mathbf{X})$, $\mathcal{F}_2 = \sigma(\mathbf{Y})$. Now we have special interest in the conditional probabilities

$$P(\,\sigma(\mathbf{X})\,|\,\sigma(\mathbf{Y})\,), \quad P(\,C\,|\,\sigma(\mathbf{Y})\,), \quad P(\,\sigma(\mathbf{X})\,|\,D\,) \tag{3.3.25}$$

For simplicity to write, we write them down separately in brief as

$$P(\mathbf{X}|\mathbf{Y}), \quad P(C|\mathbf{Y}), \quad P(\mathbf{X}|D) \tag{3.3.26}$$

and called in brief, respectively the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$; of C given $\mathbf{Y}$; of $\mathbf{X}$ given D.

(1) **The conditional probability distribution and the conditional distribution function of X given** D

In the probability space $(\Omega, \mathcal{F}, P)$, $P(\mathbf{X}|D)$ is a probability measure

$$P(C|D), \quad C \in \sigma(\mathbf{X}) \tag{3.3.27}$$

on the sub-test $(\Omega, \sigma(\mathbf{X}))$.

We turn to the value probability space $(\mathbf{R}^n, \mathcal{B}^n, P_{\mathbf{X}})$. Using (3.3.27) we may introduce a new probability measure $P_{\mathbf{X}}(\mathcal{B}^n|D)$ on $(\mathbf{R}^n, \mathcal{B}^n)$ as follows

$$P_{\mathbf{X}}(A|D) \overset{def}{=} P(\,\mathbf{X} \in A\,|\,D\,), \quad A \in \mathcal{B}^n \tag{3.3.28}$$

For any $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$, let

$$F_{\mathbf{X}}(\,x_1, x_2, \cdots, x_n\,|\,D) \overset{def}{=} P_{\mathbf{X}}(\,\prod_{i=1}^{n}(-\infty, x_i]\,|\,D\,) \tag{3.3.29}$$

We know that it is the distribution function of $P_{\mathbf{X}}(\mathcal{B}^n|D)$ by Theorem 2.8.3.

Definition 3.3.10. $P_{\mathbf{X}}(A|D)$, $A \in \mathcal{B}^n$ *is called the conditional probability distribution of* $\mathbf{X}$ *given* D. $F_{\mathbf{X}}(x_1, x_2, \cdots, x_n|D)$, $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$ *is called the conditional distribution function of* $\mathbf{X}$ *given* D.

(2) **Conditional density of the event** C **given Y**

In the probability space $(\Omega, \mathcal{F}, P)$, $P(C|\mathbf{Y})$ is the conditional probability of C given $\mathbf{Y}$. It is a set function

$$P(C|D), \quad D \in \sigma(\mathbf{Y}) \tag{3.3.30}$$

defined on the sub-test $(\Omega, \sigma(\mathbf{Y}))$. It is from Theorem 2.10.1 that its conditional density $p(C|\sigma(\mathbf{Y}))(\omega)$ exists. We write it down in brief as

$$p(C|\mathbf{Y})(\omega), \quad \omega \in \Omega \tag{3.3.31}$$

We turn to the value probability space $(\mathbf{R}^m, \mathcal{B}^m, P_{\mathbf{Y}})$. Since the function (3.3.31) is a $\sigma(\mathbf{Y})$-measurable with respect to ω, from the knowledge of the measurable transformation we obtain that on $(\mathbf{R}^m, \mathcal{B}^m, P_{\mathbf{Y}})$ there exists a Borel function $\varphi(\mathbf{y})$, $\mathbf{y} \in \mathbf{R}^m$ such that

$$p(C|\mathbf{Y})(\omega) = \varphi(\mathbf{Y}(\omega)) \quad (P_{\mathbf{Y}} - a.s.) \tag{3.3.32}$$

φ is unique in the sense of without count zero probability set.

According the habit of the probability theory, we use the spacial symbol $p(C|\mathbf{Y})(\mathbf{y})$, $\mathbf{y} \in \mathbf{R}^m$ instead of $\varphi(\mathbf{y})$[7].

Definition 3.3.11. *$p(C|\mathbf{Y})$ is called the conditional density of the event C given $\mathbf{Y}$. If its argument is the $\mathbf{y}$, then it is called the conditional density with Kolmogorov's form. If its argument is the ω, then it is called the conditional density with Doob's form.*

(3) **Conditional density–probability, conditional density–probability distribution, conditional density–distribution function of X given the sub-test** $(\Omega, \mathcal{F}_2)$

In the probability space $(\Omega, \mathcal{F}, P)$ the conditional probability $P(\sigma(\mathbf{X})|\mathcal{F}_2)$ may be written in brief as $P(\mathbf{X}|\mathcal{F}_2)$. It is called in brief the conditional probability of $\mathbf{X}$ given $\mathcal{F}_2$, and it is a set function of two arguments

$$P(C|D), \quad C \in \sigma(\mathbf{X}),\ D \in \mathcal{F}_2^* \tag{3.3.33}$$

By using Theorem 2.10.4, for any conditional probability $P(\mathbf{X}|\mathcal{F}_2)$ we know there is its conditional density–probability

$$p(C, \omega|\mathcal{F}_2), \quad (C, \omega) \in \sigma(\mathbf{X}) \times \Omega \tag{3.3.34}$$

Let us turn to the value probability space $(\mathbf{R}^n, \mathcal{B}^n, P_{\mathbf{X}})$. By using (3.3.5) and (3.3.14) we may define a new set-point function

$$p_{\mathbf{X}}(A, \omega|\mathcal{F}_2) \overset{def}{=} p(\mathbf{X}^{-1}(A), \omega|\mathcal{F}_2), \quad (A, \omega) \in \mathcal{B}^n \times \Omega \tag{3.3.35}$$

[7]Using this symbol the (3.3.32) becomes the expression $p(C|\mathbf{Y})(\omega) = p(C|\mathbf{Y})(\mathbf{Y}(\omega))$ with a little orneriness. So in some books the special symbol $p(C|\mathbf{Y} = \mathbf{y})$ is used.

where $\mathbf{X}^{-1}(A) = \{\omega|\mathbf{X}(\omega) \in A\}$ is an element of $\sigma(\mathbf{X})$. It is easily verified that the new set-point function has three properties as follows:

① For fixed A, $p_{\mathbf{X}}$ is $\mathcal{F}_2$-measurable with respect to ω.

② For fixed ω, $p_{\mathbf{X}}$ is a probability measure on $(\mathbf{R}^n, \mathcal{B}^n)$ with respect to A.

③ For any $A \in \mathcal{B}^n$, $D \in \mathcal{F}_2^*$, we have

$$P(\mathbf{X} \in A \,|\, D) = P_{\mathbf{X}}(A|D) = \frac{1}{P(D)} \int_D p_{\mathbf{X}}(A, \omega|\mathcal{F}_2) P(d\omega) \tag{3.3.36}$$

Now for any $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$, $\omega \in \Omega$, let

$$F_{\mathbf{X}|\mathcal{F}_2}(x_1, x_2, \cdots, x_n; \omega) \overset{def}{=} p_{\mathbf{X}}(\prod_{i=1}^{n}(-\infty, x_i], \omega \,|\, \mathcal{F}_2) \tag{3.3.37}$$

Obviously for every fixed ω, it is the distribution function of $p_{\mathbf{X}}(A, \omega|\mathcal{F}_2)$, $A \in \mathcal{B}^n$.

Definition 3.3.12. ① $p(C, \omega|\mathcal{F}_2)$, $(C, \omega) \in \sigma(\mathbf{X}) \times \Omega$ *is called the conditional density-probability of* $\mathbf{X}$ *given sub-test* $(\Omega, \mathcal{F}_2)$ *(or* $\mathcal{F}_2$*).*

② $p_{\mathbf{X}}(A, \omega|\mathcal{F}_2)$, $(A, \omega) \in \mathcal{B}^n \times \Omega$ *is called the conditional density-probability distribution of* $\mathbf{X}$ *given sub-test* $(\Omega, \mathcal{F}_2)$ *(or* $\mathcal{F}_2$*).*

③ $F_{\mathbf{X}|\mathcal{F}_2}(x_1, x_2, \cdots, x_n; \omega)$, $(x_1, x_2, \cdots, x_n; \omega) \in \mathbf{R}^n \times \Omega$ *is called the conditional density-distribution function of* $\mathbf{X}$ *given sub-test* $(\Omega, \mathcal{F}_2)$ *(or* $\mathcal{F}_2$*).*

(4) **Conditional density–probability, conditional density–probability distribution, conditional density–distribution function of X given Y**

Suppose $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_m)$ is a random vector, its value probability space is $(\mathbf{R}^m, \mathcal{B}^m, P_{\mathbf{Y}})$. In the paragraph (3) above taking $\mathcal{F}_2 = \sigma(\mathbf{Y})$, if from beginning (3.3.34) we use $(y_1, y_2, \cdots, y_m)$ instead of ω, then all conclusions are still set up. At that time (3.3.34) becomes

$$p(C; y_1, y_2, \cdots, y_m \,|\, \mathbf{Y}), \quad (C, y_1, y_2, \cdots, y_m) \in \sigma(\mathbf{X}) \times \mathbf{R}^m \tag{3.3.38}$$

and in the sense of without count zero probability set we have

$$p(\,C; Y_1(\omega), Y_2(\omega), \cdots, Y_m(\omega) \,|\, \mathbf{Y}\,) = p(\,C, \omega \,|\mathbf{Y}) \tag{3.3.39}$$

(3.3.35) and (3.3.37) become, respectively

$$p_{\mathbf{X}}(A; y_1, y_2, \cdots, y_m \,|\mathbf{Y}) \overset{def}{=} p(\mathbf{X}^{-1}(A), y_1, y_2, \cdots, y_m \,|\mathbf{Y}),$$
$$(A, y_1, y_2, \cdots, y_m) \in \mathcal{B}^n \times \mathbf{R}^m \tag{3.3.40}$$

$$F_{\mathbf{X}|\mathbf{Y}}(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_m)$$
$$\overset{def}{=} p_{\mathbf{X}}(\prod_{i=1}^{n}(-\infty, x_i], y_1, y_2, \cdots, y_m \,|\mathbf{Y}) \tag{3.3.41}$$

Definition 3.3.13. ① $p(C; y_1, y_2, \cdots, y_m|\mathbf{Y})$, $(C; y_1, y_2, \cdots, y_m) \in \sigma(\mathbf{X}) \times \mathbf{R}^m$ *is called the conditional density-probability of* $\mathbf{X}$ *given* $\mathbf{Y}$ *(or sub-test* $(\Omega, \sigma(\mathbf{Y}))$*).*

② $p_{\mathbf{X}}(A; y_1, y_2, \cdots, y_m|\mathbf{Y})$, $(A; y_1, y_2, \cdots, y_m) \in \mathcal{B}^n \times \mathbf{R}^m$ *is called the conditional density-probability distribution of* $\mathbf{X}$ *given* $\mathbf{Y}$ *(or sub-test* $(\Omega, \sigma(\mathbf{Y}))$*).*

③ $F_{\mathbf{X}|\mathbf{Y}}(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_m), (x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_m) \in \mathbf{R}^n \times \mathbf{R}^m$ *is called the conditional density-distribution function of* $\mathbf{X}$ *given* $\mathbf{Y}$ *(or sub-test* $(\Omega, \sigma(\mathbf{Y}))$*).*

3.3.6 Conditional mathematical expectation of X given sub-test $(\Omega, \mathcal{F}_2)$

Let $(\Omega, \mathcal{F}, P)$ is a probability space, $\mathbf{X}$ is a random vector on it and $(\Omega, \mathcal{F}_2)$ is its sub-test.

Definition 3.3.14. *Suppose* $p(C; \omega|\mathcal{F}_2)$, $(C; \omega) \in \sigma(\mathbf{X}) \times \Omega$ *is the conditional density-probability of* $\mathbf{X}$ *given* $(\Omega, \mathcal{F}_2)$. *If*

$$E(\mathbf{X}|\mathcal{F}_2)(\omega) \overset{def}{=} \int_{\Omega} \mathbf{X}(\omega^*) p(d\omega^*; \omega|\mathcal{F}_2) \tag{3.3.42}$$

exists, then $E(\mathbf{X}|\mathcal{F}_2)(\omega)$ *is called the conditional (mathematical) expectation of* $\mathbf{X}$ *given sub-test* $(\Omega, \mathcal{F}_2)$ *(or* $\mathcal{F}_2$*).*

It is easily proved that when $E\mathbf{X}$ exists, then $E(\mathbf{X}|\mathcal{F}_2)$ must exist. In addition, in the value probability space the corresponding form of (3.3.42) is

$$E(\mathbf{X}|\mathcal{F}_2)(\omega) = \int_{\mathbf{R}^n} \mathbf{x} p_{\mathbf{X}}(d\mathbf{x}; \omega|\mathcal{F}_2) = \int_{\mathbf{R}^n} \mathbf{x} d_{\mathbf{x}} F_{\mathbf{X}|\mathcal{F}_2}(\mathbf{x}; \omega) \tag{3.3.43}$$

Theorem 3.3.3. *The conditional expectation* $E(\mathbf{X}|\mathcal{F}_2)$ *has the properties as follows:*

① *It is a* $\mathcal{F}_2$*-measurable random vector;*

② *For any* $D \in \mathcal{F}_2$, *we have*

$$\int_D E(\mathbf{X}|\mathcal{F}_2)(\omega)P(d\omega) = \int_D \mathbf{X}(\omega)P(d\omega) \tag{3.3.44}$$

Conversely, the function with the properties ① and ② always exists, and is unique in the sense of without count zero probability set.

In modern probability theory the conditional expectation is defined by using the conditions ① and ② in the theorem above. For the display of it to be the extension of mathematical expectation, further people introduce conditional density-probability distribution and the conditional density-distribution function, and prove (3.3.42) and (3.3.43) to be set up[25].

3.3.7 Conditional mathematical expectation of X given Y

When $\mathcal{F}_2 = \sigma(\mathbf{Y})$, we rewrite $E(\mathbf{X}|\sigma(\mathbf{Y}))$ as $E(\mathbf{X}|\mathbf{Y})$. i.e.

$$E(\mathbf{X}|\mathbf{Y})(\omega) = E(\mathbf{X}|\sigma(\mathbf{Y}))(\omega) \tag{3.3.45}$$

Let us turn to the value probability space $(\mathbf{R}^m, \mathcal{B}^m, P_{\mathbf{Y}})$. Since $E(\mathbf{X}|\mathbf{Y})(\omega)$ is a $\sigma(\mathbf{Y})$-measurable function, by using the knowledge about measurable transformation we obtain that there is a Borel function $\psi(\mathbf{y})$, $\mathbf{y} \in \mathbf{R}^m$ such that

$$E(\mathbf{X}|\mathbf{Y})(\omega) = \psi[\mathbf{Y}(\omega)] \quad (P(\sigma(\mathbf{Y})) - a.s.) \tag{3.3.46}$$

According the habit of probability theory, we write $\psi(\mathbf{y})$ as $E(\mathbf{X}|\mathbf{Y})(\mathbf{y})$, $\mathbf{y} \in \mathbf{R}^m$.

Definition 3.3.15. $E(\mathbf{X}|\mathbf{Y})$ *is called the conditional (mathematical) expectation of* $\mathbf{X}$ *given* $\mathbf{Y}$. *If its argument is the* $\mathbf{y}$, *then it is called the conditional expectation with Kolmogorov's form. If its argument is the* ω, *then it is called the conditional expectation with Doob's form*[8].

3.3.8 Remark

The conditional mathematical expectation is one of the most important concepts in the modern probability theory. It plays a key role in the development of the stochastic processes and mathematical statistics. The reasons probably are as follows.

(1) The average value is a popular concept which is used widely and verified for a long time. Its defect is with a narrow suitable range. The

[8] In some books Kolmogorov's form of the conditional expectation is written as $E(\mathbf{X}|\mathbf{Y} = \mathbf{y})$. see the footnote 7.

mathematical expectation and conditional expectation are the extension of the average in general case. They become the important tools reflecting the statistical knowledge and statistical laws of a family of the random variables.

(2) Suppose $(\Omega, \mathcal{F}, P)$ is the probability space generated along the method of (2.7.4). Then $E\mathbf{X}$ may be understood as the average of $\mathbf{X}$ under the condition $\mathcal{C}$. If the condition $\mathcal{C}$ is strengthened by the sub-test $(\Omega, \mathcal{F}_2)$ (see Axiom E_2), then the average of $\mathbf{X}$ varies and becomes the conditional expectation $E(\mathbf{X}|\mathcal{F}_2)$. Therefore the conditional expectation $E(\mathbf{X}|\mathcal{F}_2)$ is the answers of some applicable problems, as well as the important tools to solve other applicable problems.

(3) Generally, after getting the $(\Omega, \mathcal{F}, P)$ generated along the method of (2.7.4), we successively perform the sub-tests $(\Omega, \mathcal{F}_1)$, $(\Omega, \mathcal{F}_2)$, $(\Omega, \mathcal{F}_2) \cdots$. At that time the averages will vary continuously, and obtain a sequence of conditional expectations:

$$E\mathbf{X},\ E[E\mathbf{X}|\mathcal{F}_1],\ E\{E[E\mathbf{X}|\mathcal{F}_1]|\mathcal{F}_2\},\ E\left(E\{E[E\mathbf{X}|\mathcal{F}_1]|\mathcal{F}_2\}|\mathcal{F}_3\right)\ \cdots$$

The research of the sequences like this has great meaning in theory and practice.

(4) The conditional expectation $E(\mathbf{X}|\mathcal{F}_2)$ is a random vector and has many properties. The calculations relating to them have formed a set of mature methods.

(5) For any $C \in \mathcal{F}$, from (3.3.42) we obtain

$$E(\chi_C|\mathcal{F}_2)(\omega) = p(C, \omega|\mathcal{F}_2) = p(C|\mathcal{F}_2)(\omega) \tag{3.3.47}$$

So the conditional density of the event C given $\mathcal{F}_2$ equals to the conditional expectation of its indicator function χ_C given $\mathcal{F}_2$.

3.4 Basic conceptions of broad stochastic process

Suppose J is any infinite index set.

Definition 3.4.1. *Let $(\Omega, \mathcal{F}, P)$ is a probability space. Suppose $X_\alpha(\omega)$ is a random variable on it ($\alpha \in J$). Then the family of the random variables $\{X_\alpha(\omega)|\alpha \in J\}$ is called a broad stochastic process on $(\Omega, \mathcal{F}, P)$*[9]*, and written in brief as $X_\alpha(\omega), \alpha \in J$; or $X_\alpha(\omega)$; or $X_J(\omega)$; or $X(J)$.*

---

[9]This term is made up by us. We emphasize it to be a special family such that it is different from its various subfamily.

Definition 3.4.2. *Let $(\Omega, \mathcal{F}, P)$ is a probability space and $(\mathbf{R}^J, \mathcal{B}^J)$ is a J-dimension Borel test. Suppose $X_J(\omega) \overset{def}{=} \{X_\alpha(\omega)|\alpha \in J)$ is a function defined on Ω taking the values in $\mathbf{R}^J$. If for any $B \in \mathcal{B}^J$, we have*

$$(X_J \in B) \overset{def}{=} \{\omega\,|X_J(\omega) \in B\} \in \mathcal{F} \tag{3.4.1}$$

then $X_J(\omega)$ is called a broad stochastic process on $(\Omega, \mathcal{F}, P)$.

It is from the definition of the J-dimensional Borel test that the two definitions are equivalent. Let

$$\mathcal{A}_J = \{\,(X_J \in A)\,|\,A \text{ is the cylinder event defined by (2.5.34), and } n \geq 1\} \tag{3.4.2}$$

$$\sigma(X_J) = \{\,(X_J \in B)\,|\,B \in \mathcal{B}^J\,\} \tag{3.4.3}$$

Then $\sigma(X_J)$ is an event σ-subfield of $\mathcal{F}$, and $\mathcal{A}_J$ is its bud event set.

Definition 3.4.3. *The $\sigma(X_J)$ is called the event σ-field of X_J. It is the set consisted of all events involving the broad stochastic process X_J. The $(\Omega, \sigma(X_J))$ is called the random test generated by X_J.*

3.4.1 Representation probability space and the digital characters

Definition 3.4.4. *The $(\Omega, \sigma(X_J), P)$ is called the representation probability space of the broad stochastic process X_J.*

Definition 3.4.5. *Let X_J is a broad stochastic process on $(\Omega, \mathcal{F}, P)$. If for any $\alpha \in J$, EX_α exists, then the value function defined on J*

$$EX_J \overset{def}{=} \{\,EX_\alpha\,|\,\alpha \in J,\} \tag{3.4.4}$$

is called the (mathematical) expectation function of the broad stochastic process X_J.

Definition 3.4.6. *Let X_J is a broad stochastic process on $(\Omega, \mathcal{F}, P)$. If for any $\alpha, \beta \in J$, the covariance $Cov(X_\alpha, X_\beta)$ exists, then the value function defined on $J \times J$*

$$Cov(X_J, X_J) \overset{def}{=} \{\,Cov(X_\alpha, X_\beta)\,|\,\alpha, \beta \in J\,\} \tag{3.4.5}$$

is called the covariance function of the broad stochastic process X_J.

The expectation function and the covariance function all may be represented by the integral expression. They are, respectively

$$EX_J = \int_\Omega X_J(\omega)P(d\omega) \overset{def}{=} \{\int_\Omega X_\alpha(\omega)P(d\omega) \mid \alpha \in J\} \tag{3.4.6}$$

$$Cov(X_J, X_J) = \{\int_\Omega [X_\alpha(\omega) - EX_\alpha][X_\beta(\omega) - EX_\beta]P(d\omega) \mid \alpha, \beta \in J\} \tag{3.4.7}$$

3.4.2 Value probability space and the family of finite dimensional distribution functions

Suppose $(\mathbf{R}^J, \mathcal{B}^J)$ is the Borel test in Definition 3.4.2. By using (3.4.3) we may introduce a set function

$$P_{X(J)}(B) = P(X_J \in B), \quad B \in \mathcal{B}^J \tag{3.4.8}$$

It is easily proved that $P_{X(J)}(\mathcal{B}^J)$ is a probability measure. From Theorem 2.8.4 we obtain that there is the family of the finite dimensional distribution functions

$$F = \{F(\alpha_1, x_1; \alpha_2, x_2; \cdots; \alpha_n, x_n) \mid n \geq 1; \alpha_1, \alpha_2, \cdots, \alpha_n \in J\} \tag{3.4.9}$$

such that F and $P_{X(J)}(\mathcal{B}^J)$ are uniquely determined each other by the formulas (2.8.25) and (2.8.26).

Definition 3.4.7. *The $(\mathbf{R}^J, \mathcal{B}^J, P_{X(J)})$ is called the value probability space of the broad stochastic process X_J, and $P_{X(J)}(\mathcal{B}^J)$ is called the probability distribution of X_J, and F is called the family of finite dimensional distribution functions of X_J.*

The measurable transformation in the measure theory insures that the properties and the expressions of X_J on $(\Omega, \mathcal{F}, P)$ and the corresponding ones of $(\mathbf{R}^J, \mathcal{B}^J, P_{X(J)})$ may mutually transform. i.e. **The value probability space includes all statistical knowledge of the broad stochastic process.**

The expressions of the expectation and covariance functions in the value probability space are, respectively

$$EX_J = \int_{\mathbf{R}^J} x_J P_{X(J)}(dx_J) \overset{def}{=} \{\int_{\mathbf{R}^J} x_\alpha P_{X(J)}(dx_J) \mid \alpha \in J\} \tag{3.4.10}$$

$$Cov(X_J, X_J) \overset{def}{=} \{\int_{\mathbf{R}^J} (x_\alpha - EX_\alpha)(x_\beta - EX_\beta)P_{X(J)}(dx_J) \mid \alpha, \beta \in J\} \tag{3.4.11}$$

By using the family of finite dimensional distribution functions F they are simplified as, respectively

$$EX_J = \{ \int_{\mathbf{R}} x d_x F(\alpha, x) \,|\, \alpha \in J \} = \{ EX_\alpha \,|\, \alpha \in J \} \tag{3.4.12}$$

$$\begin{aligned} Cov(X_J, X_J) &= \{ Cov(X_\alpha, X_\beta) \,|\, \alpha, \beta \in J \} \\ &= \{ \int_{\mathbf{R}\times\mathbf{R}} (x - EX_\alpha)(y - EX_\beta) d_x d_y F(\alpha, x; \beta, y) \,|\, \alpha, \beta \in J \} \end{aligned} \tag{3.4.13}$$

where $F(\alpha, x) = P(X_\alpha \leq x)$, $F(\alpha, x; \beta, y) = P(X_\alpha \leq x, X_\beta \leq y)$ are, respectively one dimensional and two dimensional distribution functions.

Therefore the broad stochastic process has found the convenient and the simplest tools to represent its all statistical knowledge — the family of finite dimensional distribution functions and the digital characters generated by it. **The family of finite dimensional distribution function can represent all statistical knowledge and statistic laws of the broad stochastic process.**

The value probability space $(\mathbf{R}^J, \mathcal{B}^J, P_{X(J)})$ and the family of finite dimensional distribution functions on it have great meaning in the study of the broad stochastic process. The reason stated for the random variable in the subsection3.2.2 is also suitable to the broad stochastic process.

The new situation occurring now is that the measure and integration theory on infinite dimensional space $(\mathbf{R}^J, \mathcal{B}^J)$ is abstract. Although the fact brings some difficulties for the broad stochastic process, but it also brings some advantages for the study of infinite dimensional space, since the broad stochastic process provides a kind of the intuitive background materials.

3.4.3 Family of finite dimension random vectors

Definition 3.4.8. *Let X_J is a broad stochastic process. We take anyway n different elements in the index set J and assign the order to them. Supposing they are $\alpha_1, \alpha_2, \cdots, \alpha_n$ after assignment of the order, then $(X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n})$ is called a n-dimensional random vector of the broad stochastic process. For reflecting the order relation to be loose, $(X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n})$ is also rewritten $(\alpha_1, X_{\alpha_1};\ \alpha_2, X_{\alpha_2};\ \cdots; \alpha_n, X_{\alpha_n})$.*

Definition 3.4.9. *We use fV to represent the family of all finite dimensional random vectors. i.e.*

$$fV = \{ (\alpha_1, X_{\alpha_1}; \alpha_2, X_{\alpha_2}; \cdots; \alpha_n, X_{\alpha_n}) \,|\, n \geq 1; \alpha_1, \alpha_2, \cdots, \alpha_n \in J \} \tag{3.4.14}$$

and call the fV the family of finite dimensional random vectors of the broad stochastic process X_J.

By using the family of finite dimensional random vectors we obtain many things which can represent a lot of statistical knowledge and statistical laws of the broad stochastic process. They are

- (1) The n-dimensional random vector $(\alpha_1, X_{\alpha_1}; \alpha_2, X_{\alpha_2}; \cdots; \alpha_n, X_{\alpha_n})$ generates a sub-test $(\Omega, \mathcal{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n})$ and a representation probability subspace $(\Omega, \mathcal{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n}, P)$, where $\mathcal{F}_{\alpha_1,\alpha_2,\cdots,\alpha_n} = \sigma(X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n})$.
- (2) The value probability space and the probability distribution of n-dimensional random vector are, respectively, called the n-dimensional marginal probability space and the n-dimensional marginal probability distribution of the broad stochastic process.
- (3) The distribution function of n-dimensional random vector is called the n-dimensional marginal distribution function of X_J. Obviously, the family of finite dimensional distribution functions F is the family which consists of all marginal distribution functions.
- (4) The digital characters EX_J and $Cov(X_J, X_J)$ are, respectively determined by one dimensional and two dimensional marginal distribution functions.
- (5) The fV may generate many conditional expectations, conditional probabilities and conditional densities in various forms. They reflect the relations among the various random vectors of X_J.

3.4.4 Index set J

If the index set is assigned with the order relation (the total order or the semi-order, etc.), the topological relation, the algebraic relation, then some special kind of the broad stochastic processes are obtain. If the index set is the T which is a subset of the real line $\mathbf{R}$, then the broad stochastic process X_T, or $X_t(\omega), t \in T$ is called the stochastic process[10]. The stochastic process is one kind of the broad stochastic processes with the longest history and the richest statistical knowledge. It is the source of the other kinds of the broad stochastic processes.

In order to strengthen the intuition and appearance, the index set T of the stochastic process is usually illustrated as time. Therefore people can organize the many research objects such as the random sub-tests, the marginal distribution functions, various conditional probabilities and conditional mathematical expectations, and so on, as the well-organized "flow"

[10]In the applications we usually use a kind of its extension form: $\mathbf{X}_T$, or $\mathbf{X}_t(\omega), t \in T$. It is called the stochastic vector process, where $\mathbf{X}_t(\omega)$ is a random vector.

with clear relations according the order of time. For example, one kind of special conditional mathematical expectations "flow" is called the martingale; for example furthermore, one kind of special random sub-tests "flow" is called the event flow or the filtration, and the event flow space or the filtered measurable space are introduced[20][23][26]. The author defined the stochastic process on the event flow space, and brought much convenience for the research of the Markov process and the martingale[20].

In a word, the first task to study the broad stochastic process is how to organize the products generated by the family of finite dimensional random vectors, and to study them deeply.

The mathematics always recognize the infinite by using the finite, so does the broad stochastic process. We first recognize fV primarily, then introduce the concept of the convergence and develop the limit method, finally by using limit method we can deduct out the statistical knowledge and statistical laws of the broad stochastic process or its subfamily from the statistical knowledge of the fV.

This book does not involve this task, but the introduction of the fascinating task above ends this book.

Bibliography

[1] Kolmogorov A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebn. Math. 2, No. 3, Berlin, 1933.

[2] de Finetti B., *Probability, Induction and Statistics, The art of guession.*, John Wiley & Sons Ltd., 1972.

[3] Renyi A., *Probability Theory*, American Elsevier, New York, 1970.

[4] Savage L. J., The Foundations of Statistics, John Wiley, New York, 1954.

[5] Pratt J. W., Raiffa H., Schlaifer R., The foundation of decision under uncertainty: an elementary exposition, J. Amer. Stat. Assoc., 1964, 59: 353–375.

[6] de Groot M., Optimal Statistical Decision, McGraw-Hill, New York, 1970.

[7] Fishburn P. C., Subjective expected untility: a review of normative theories, Theory & Decision, 1981, 13: 139–199.

[8] Fishburn P. C., The axioms of subjective probability, Statistical Science, 1986, 1: 335–358.

[9] Hilbert D., *Grundlagen der Geometrie (12th ed.)*, B. G. Teubner Stuttgart, 1977.

[10] The Department of Mathematics of Eastern China Normal University, The Course of Probability Theory (in Chinese), Higher Education Press, Beijing, 1984.

[11] Zhang Yaoting, The development and the debate of probability concept, see: Mathematics and Culture (in Chinese), Peking University Press, Beijing, 1990, 321–332.

[12] Mises R., Wahrscheinlichkeitsrechnung, Leipzig und Wien, 1931.

[13] Keynes J. M., A Treatise on Probability, MacMilla, Landon, 1921.

[14] Gnedenko B. V., Course in Theory of Probabilities (in Russian), Moscow, 1961.

[15] Sikorski R., Boolean Algebras, Springer–Verlag, Berlin, 1960.

[16] Halmos P. R., Measure Theory, Van Nostrand Company, Inc., New York, 1950.

[17] Hausdorff F., Set Theory (3rd ed.), Chelsea, New York, 1962.

[18] Hanisch H., Hirsch W. M., Renyi A., Measures in denumerable space, Amer. Math. Monthly, 1969, 76: 494–502.

[19] Gabriel J. P., Milasevic P., A remark on probability space, J. of Theoretical Probability, 1991, 4: 191–195.

[20] Xiong Daguo, Theory of Stochastic Processes and Its Applications (in Chinese), National Defence Industry Press, beijing, 1991.

[21] Feller W., An Introduction to Probability Theory and Its Applications, Vol I, John Wiley, New York, 1957.

[22] Loeve M., Probability Theory (4th ed.), Springer Verlag, I, 1977; II, 1978.

[23] Dynkin E. B., Markov Processes, Springer Verlag, Berlin, 1965.

[24] Skorohod A. V., Event σ-algebra on probability space. Similarity and Factorization, Theory Prob. Appl. 1991, 36: 127-137.

[25] Cheng Shihong, Advanced Probability Theory (in Chinese), Peking University Press, Beijing, 1996.

[26] Doob J. L., Classical Potential Theory and Its Probabilistic Counterpart, Springer Verlag, 1984.

[27] Wang Jueyuan, Wang Xianglin, The Chat About Genetics (in Chinese), Jiang Su Science and Technology Press, Nanjing, 1982.

Index