

 大学数学系列丛书

概率论与数理统计

(英文版)

Probability and Statistics

桂文豪 王立春 孔令臣 编著

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$



清华大学出版社
<http://www.tup.com.cn>



北京交通大学出版社
<http://www.bjtup.com.cn>



责任编辑：龙嫚嫚
封面设计：七星博纳

Probability and Statistics

ISBN 978-7-5121-3652-6



9 787512 136526 >

定价：42.00元

大学数学系列丛书

概率论与数理统计

(英文版)

Probability and Statistics

桂文豪 王立春 孔令臣 编著

清华大学出版社

北京交通大学出版社

· 北京 ·

内 容 简 介

本书根据编者多年的双语教学经验编写,介绍了概率论与数理统计的基本概念、原理、计算方法,以及实际应用。在编写过程中,吸取了国内外优秀教材的优点,注重理论与实践相结合,系统性强,图例丰富,突出统计思想,着力培养学生分析问题和解决实际问题的能力。

本书主要内容包括概率与随机事件、随机变量及其分布、多维随机变量及其分布、随机变量的数字特征、大数定律和中心极限定理、参数估计、假设检验、线性回归分析。每章中精选了实用性强的例题和习题。

本书可作为高等院校理工科各专业本科生的概率论与数理统计课程双语教材,也可供工程技术人员、科技工作者参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

概率论与数理统计:英文/桂文豪,王立春,孔令臣编著. —北京:北京交通大学出版社:清华大学出版社,2018.8

(大学数学系列丛书)

ISBN 978-7-5121-3652-6

I. ①概… II. ①桂… ②王… ③孔… III. ①概率论-高等学校-教材-英文 ②数理统计-高等学校-教材-英文 IV. ①O21

中国版本图书馆CIP数据核字(2018)第171839号

概率论与数理统计

GAILÜLUN YU SHULI TONGJI

责任编辑:龙嫚嫚

出版发行:清华大学出版社 邮编:100084 电话:010-62776969 <http://www.tup.com.cn>
北京交通大学出版社 邮编:100044 电话:010-51686414 <http://www.bjtup.com.cn>

印刷者:艺堂印刷(天津)有限公司

经 销:全国新华书店

开 本:185 mm×260 mm 印张:17.5 字数:437千字

版 次:2018年8月第1版 2018年8月第1次印刷

书 号:ISBN 978-7-5121-3652-6/O·171

印 数:1~2 000册 定价:42.00元

本书如有质量问题,请向北京交通大学出版社质监局反映。对您的意见和批评,我们表示欢迎和感谢。
投诉电话:010-51686043, 51686008; 传真:010-62225406; E-mail: press@bjtu.edu.cn。

前 言

为了适应信息化和全球化的社会发展趋势,培养具有国际竞争力的人才,进行双语教学是实现这一目标的有效手段。双语教学不仅有利于学生外语水平的提高,也有利于他们专业知识的学习。概率论与数理统计是理工科相关专业的一门重要公共基础课,在科学的各个领域都有其广泛的应用,该课程学习的好坏直接影响到后续专业的学习以及整个专业知识的掌握。本书正是在此背景下,根据编者多年的双语教学经验编写而成。

本书结合国际班学生的特点,基于国际班学生现有的知识结构和学习需求,着力培养学生国际化视野,力求知识结构严谨、体系完整,紧跟统计学科发展的最新动态,主要特色有:在概念阐述和理论推导中,提供丰富的图解,通过图形的直观认识,学生能更好地掌握概率统计理论。在例题讲解中,不是简单地给出答案,而是注重分析过程,并对题目涉及的知识点进行认真点评,对某些题目进行多角度解析,启发和引导学生深入思考,促进学生更好地掌握解题的方法和技巧,指导学生加深对理论的理解和掌握。在习题方面,强调概率统计与微积分、几何代数、经济学、工程学等其他学科之间的横向联系,丰富学生学的内容,提高学生解决实际问题的本领,扩大课程的影响力。

本书注重归纳概率统计知识点的共性和总结概率统计规律,增加章节之间的内在联系,使学生对知识的掌握更加系统和灵活,体系更加完整,有效提高解决实际问题的技能。本书在取材和写作上,重视内容的新颖性和应用性、英文表达的规范性和准确性、统计思维的严谨性和启发性,便于教学与自学,有利于双语教学的进一步发展。

在编写过程中,本书吸取了国内外优秀教材和讲义资料的优点,是作者在本科教学改革中一种新的探索和尝试,由于水平有限,肯定有不妥或不完善的地方,敬请广大读者给予批评指正,欢迎提出宝贵意见和建议。作者联系方式:桂文豪,北京市海淀区北京交通大学理学院,邮箱 whgui@bjtu.edu.cn,邮编 100044。

编 者

2018 年 7 月

目 录

Chapter 1 Introduction to Probability	1
1.1 Random Experiments	2
1.2 Sample Space	2
1.3 Relations and Operations between Events	3
1.4 The Definition of Probability	7
1.5 Equally Likely Outcomes Model	10
1.6 Conditional Probability	17
1.7 Total Probability and Bayes' Theorem	21
1.8 Independent Events	24
Exercise 1	26
Chapter 2 Random Variables and Distributions	31
2.1 Random Variables	32
2.2 Cumulative Distribution Function	33
2.3 Discrete Distributions	35
2.4 Some Common Discrete Distributions	35
2.5 Continuous Distributions	41
2.6 Some Useful Continuous Distributions	43
2.7 Functions of a Random Variable	50
Exercise 2	56
Chapter 3 Multivariate Probability Distributions	61
3.1 Bivariate Distributions	62
3.2 Marginal Distributions	70
3.3 Conditional Distributions	73
3.4 Independent Random Variables	81

3.5	Functions of Two or More Random Variables	86
	Exercise 3	107
Chapter 4	Characteristics of Random Variables	115
4.1	The Expectation of a Random Variable	116
4.2	Variance	125
4.3	The Characteristics of some Common Distributions	127
4.4	Chebyshev's Inequality	132
4.5	Covariance and Correlation Coefficient	134
4.6	Moment and Covariance Matrix	142
	Exercise 4	146
Chapter 5	Large Random Samples	151
5.1	The Law of Large Numbers	152
5.2	The Central Limit Theorem	154
	Exercise 5	158
Chapter 6	Estimation	161
6.1	Population and Sample	162
6.2	Moment Estimation	165
6.3	Maximum Likelihood Estimation	167
6.4	Properties of Estimators	172
6.5	Three Important Distributions	174
6.6	Confidence Intervals	184
	Exercise 6	193
Chapter 7	Hypothesis Testing	199
7.1	Basics of Hypothesis Testing	200
7.2	Hypothesis Tests for a Population Mean	202
7.3	Testing Differences between Means	209
7.4	Hypothesis Tests for One or Two Variances	212
7.5	Goodness of Fit Tests	216
	Exercise 7	221
Chapter 8	Linear Regression	227
8.1	Linear Regression Model	228

8.2	Least Squares Estimation	229
8.3	Properties of Linear Regression Estimators	232
8.4	Inferences Concerning the Slope	236
8.5	Regression Validity	238
8.6	Confidence Interval for Mean Response	239
8.7	Inference for Prediction	241
	Exercise 8	244
Appendix A	Binomial Probability Distribution	247
Appendix B	Poisson Cumulative Distribution	250
Appendix C	Standard Normal Table	253
Appendix D	t-distribution Upper Quantiles $t_{\alpha}(n)$	255
Appendix E	χ^2-distribution Upper Quantiles $\chi_{\alpha}^2(n)$	258
Appendix F	F-distribution Upper Quantiles $F_{\alpha}(n_1, n_2)$	261
Appendix G	Some Common Probability Distributions	268
Bibliography		270

Chapter 1

Introduction to Probability

◆ 1.1 Random Experiments ◆

Two types of phenomena exist in nature and human society. One is called deterministic phenomenon. For example, the sun rises every morning and sets every evening. Free fall occurs whenever an object is acted upon by gravity alone. Deterministic phenomenon makes us believe that things have their own way of being in the world. The other one is nondeterministic or random phenomenon whose outcome is determined by chance. Examples include tossing a coin, rolling a die, knowing one's score before taking the exam, etc.

The occurrence of random phenomenon urges people to try to understand the reason of this kind of phenomenon and the possibility of occurrence so as to make sound and reasonable decision according to the situation.

Although the occurrence of random phenomena is outwardly random and accidental, through a large number of observations and experiments, people can often find the intrinsic regularity behind the randomness. This intrinsic regularity, called statistical regularity, is observed in a large number of repeated observations or experiments. Probability and statistics is a subject to study and reveal the statistical regularity of random phenomena.

Definition 1.1.1 A random experiment is real or hypothetical process satisfying the following conditions:

- (1) The experiment can be repeated indefinitely under essentially the same conditions.
- (2) The possible outcomes may not be unique and can be identified ahead of time.
- (3) The outcome is uncertain before the experiment.

Example 1.1.1 Tossing a coin and observing whether it lands heads or tails.

Example 1.1.2 Tossing a coin twice and observing the sequence of heads and tails.

Example 1.1.3 Rolling a die and observing the number of spots.

Example 1.1.4 Measuring the lifetime of a light bulb.

◆ 1.2 Sample Space ◆

Definition 1.2.1 The collection of all possible outcomes of an experiment is called the **sample space** of the experiment, denoted by Ω or S . The elements in sample space are called sample points or basic events.

Definition 1.2.2 An **event** is a well-defined subset of the sample space of the experiment, denoted by A, B, C , etc. The sample space S is guaranteed to happen and called the certain event. The empty set $\emptyset$ that cannot happen is called the impossible event.

Example 1.2.1 Rolling a die. The sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

The event A is that an even number is obtained, and it can be represented as the subset

$$A = \{2, 4, 6\}.$$

The event B that a number greater than 2 is obtained is defined by the subset

$$B = \{3, 4, 5, 6\}.$$

Example 1.2.2 Tossing a coin twice and observing the sequence of heads (H) and tails (T). The sample space is

$$S = \{HH, HT, TH, TT\}.$$

The event A is that the first toss results in a head, and it can be represented as the subset

$$A = \{HH, HT\}.$$

◆ 1.3 Relations and Operations between Events ◆

Suppose that A is an event. We say that the event A “occurs” if the outcome of the experiment is contained in A . Since the events are subsets of a sample space, they obey the rules of set relations and operations.

Definition 1.3.1 It is said that a set A is contained in another set B if every element of the set A also belongs to the set B . For events, $A \subset B$ means that if A occurs then so does B . See Figure 1.1.

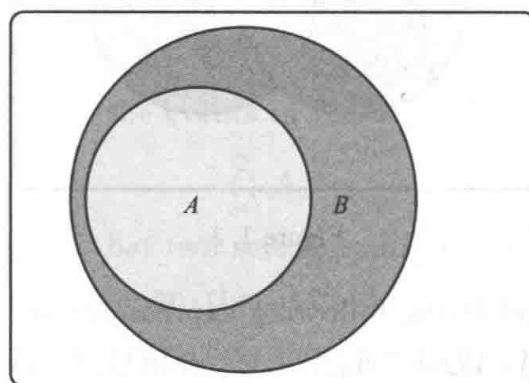


Figure 1.1

Theorem 1.3.1 Let A, B and C be events, then $A \subset S$. If $A \subset B$ and $B \subset A$, then $A = B$. If $A \subset B$ and $B \subset C$, then $A \subset C$.

Example 1.3.1 Rolling a die. Suppose that A is the event that an even number is obtained and C is the event that a number greater than 1 is obtained.

$$A = \{2, 4, 6\}, C = \{2, 3, 4, 5, 6\}, A \subset C.$$

Definition 1.3.2 The **complement** of a set A is defined to be the set that contains all elements of the sample space S that do not belong to A . The notation for the complement of A is A^c or $\bar{A}$. See Figure 1.2.

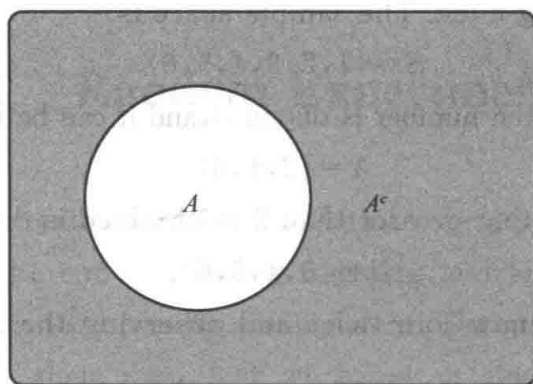


Figure 1.2

Example 1.3.2 Rolling a die. Suppose A is the event that an even number is obtained, then $A = \{2, 4, 6\}$ and $A^c = \{1, 3, 5\}$.

Theorem 1.3.2 Let A be an event, then

$$(A^c)^c = A, \emptyset^c = S, S^c = \emptyset.$$

Definition 1.3.3 If A and B are any two sets, the **union** of A and B is defined to be the set that contains all outcomes that belong to A alone, to B alone, or to both A and B . The notation for the union of A and B is $A \cup B$. For events, $A \cup B$ means that A occurs, or B occurs, or both occur. See Figure 1.3.

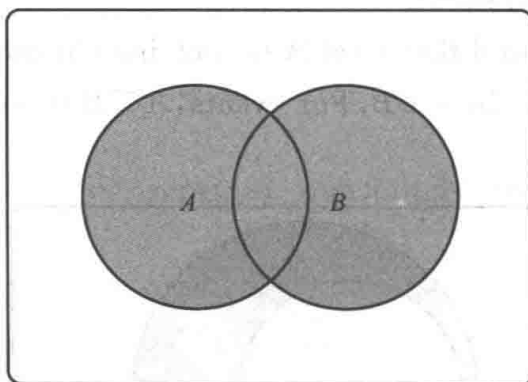


Figure 1.3

Clearly, for all sets A and B , the following expressions are valid:

$$A \cup B = B \cup A, A \cup A = A, A \cup A^c = S, A \cup \emptyset = A, A \cup S = S.$$

Furthermore, if $A \subset B$, then $A \cup B = B$.

Remark 1.3.1 The concept of union can be extended to more than two sets.

Definition 1.3.4 The union of n sets $A_1, \dots, A_n$ is defined to be the set that contains all outcomes that belong to at least one of these n sets. The notation for this union is either of the following:

$$A_1 \cup A_2 \cup \dots \cup A_n \text{ or } \bigcup_{i=1}^n A_i.$$

The union of an infinite sequence of sets $A_1, A_2, \dots$ is the set that contains all outcomes that belong to at least one of the events in the sequence.

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

Definition 1.3.5 If A and B are any two sets, the **intersection** of A and B is defined to be the set that contains all outcomes that belong both to A and to B . The notation for the intersection of A and B is $A \cap B$ or AB . See Figure 1.4.

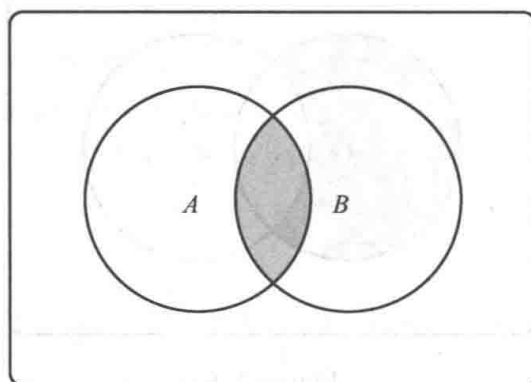


Figure 1.4

For all events A and B , the following expressions hold:

$$A \cap B = B \cap A, A \cap A = A, A \cap A^c = \emptyset, A \cap \emptyset = \emptyset, A \cap S = A.$$

Furthermore, if $A \subset B$, then $A \cap B = A$.

Remark 1.3.2 The concept of intersection can also be extended to more than two sets.

Definition 1.3.6 The intersection of n sets $A_1, \dots, A_n$ is defined to be the set that contains all outcomes that are common to all these n sets. The notation for this intersection is either of the following:

$$A_1 \cap A_2 \cap \dots \cap A_n \text{ or } \bigcap_{i=1}^n A_i.$$

The intersection of an infinite sequence of sets $A_1, A_2, \dots$ is the set that contains all outcomes that are common to all the events in the sequence.

$$\bigcap_{i=1}^{\infty} A_i.$$

Definition 1.3.7 It is said that two sets A and B are **disjoint**, or **mutually exclusive**, if A and B have no outcomes in common, that is, if $A \cap B = AB = \emptyset$. See Figure 1.5. For n sets $A_1, \dots, A_n$, if any two of them are disjoint, then they are called the pairwise disjoint events.

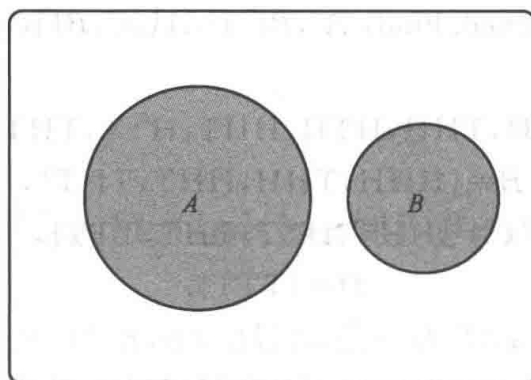


Figure 1.5

Definition 1.3.8 The **difference** of two sets A and B , which we write as $A-B$, is the set formed by all elements that belong to A , but do not belong to B . See Figure 1.6.

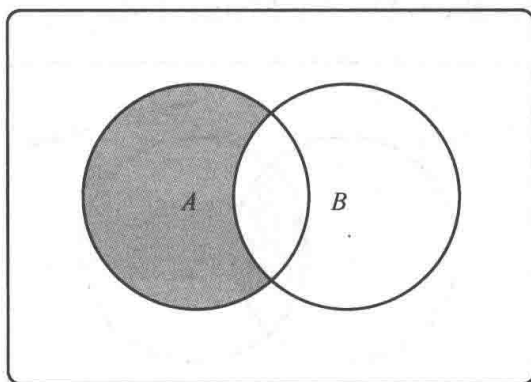


Figure 1.6

Clearly, $A-B = AB^c = A-AB$. But $A-B$ is not same as $B-A$.

Theorem 1.3.3 For events $A, B, C, A_1, A_2, \dots$, the operations satisfy the following properties:

(1) Associative Properties.

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C),$$

$$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C).$$

(2) Distributive Properties.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

(3) De Morgan's Laws.

$$(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c,$$

$$(\cap A_i)^c = \cup A_i^c, (\cup A_i)^c = \cap A_i^c.$$

Example 1.3.3 Suppose that a coin is tossed three times. Then the sample space S contains the following eight possible outcomes:

$$S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}$$

In this notation, H indicates a head and T indicates a tail. Let A be the event that at least one head is obtained in the three tosses; let B be the event that a head is obtained on the second toss; let C be the event that a tail is obtained on the third toss; and let D be the event that no heads are obtained. Find: $A^c, B \cap D, A \cup C, B \cap C, (B \cup C)^c, A \cap (B \cup C)$.

Solution:

$$A = \{HHH, THH, HTH, HHT, HTT, THT, TTH\},$$

$$B = \{HHH, THH, HHT, THT\},$$

$$C = \{HHT, HTT, THT, TTT\},$$

$$D = \{TTT\}.$$

Therefore, $A^c = D = \{TTT\}$; $B \cap D = \emptyset$, $A \cup C = S$, $B \cap C = \{HHT, THT\}$, $(B \cup C)^c = \{HTH, TTH\}$; $A \cap (B \cup C) = \{HHH, THH, HHT, HTT, THT\}$.

Example 1.3.4 Express each of the following events in terms of the events A, B and

C as well as the operations of complementation, union and intersection. In each case draw the corresponding Venn diagram.

(1) Exactly one of the events A, B, C occurs;

(2) At least one of the events A, B, C occurs;

(3) At least two of the events A, B, C occurs;

(4) At most one of the events A, B, C occurs.

Solution: (1) $AB^cC^c \cup A^cBC^c \cup A^cB^cC$. (2) $A \cup B \cup C$. (3) $AB \cup AC \cup BC$. (4) $AB^cC^c \cup A^cBC^c \cup A^cB^cC \cup A^cB^cC^c \cup A^cB^cC^c \cup B^cC^c$ or $A^cB^c \cup A^cC^c \cup B^cC^c$. The corresponding Venn diagrams are shown in Figure 1.7.

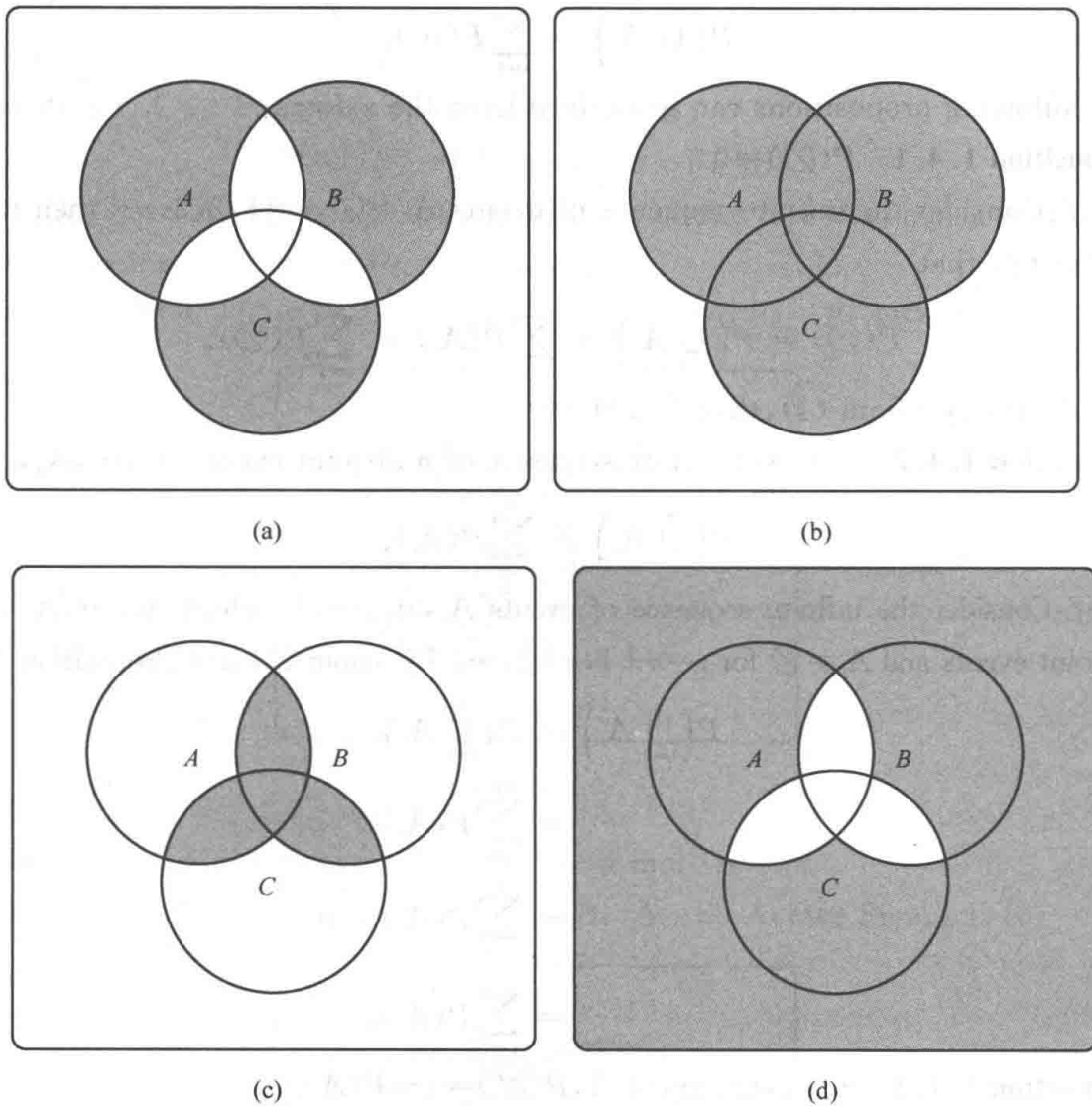


Figure 1.7

◆ 1.4 The Definition of Probability ◆

All possible outcomes of an experiment are known ahead of time. But only knowing what random events may occur is not enough when we study random phenomena. We also need to determine the possibility of the events. The measurement and characterization of

the possibility of occurrence of the event is called the probability.

We begin with the mathematical definition of probability and then present some useful results that follow easily from the definition.

Definition 1.4.1 In a given experiment, for any event A in the sample space S , a real number $P(A)$ is assigned. $P(A)$ indicates the probability that A will occur if the set function $P(\cdot)$ satisfies the following three specific axioms:

- (1) Nonnegativity: For every event A , $P(A) \geq 0$.
- (2) Normality: $P(S) = 1$.
- (3) Countable additivity: For every infinite sequence of disjoint events $A_1, A_2, \dots$,

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

The following propositions can be derived from the axioms:

Proposition 1.4.1 $P(\emptyset) = 0$.

Proof: Consider the infinite sequence of events $A_i = \emptyset, i = 1, 2, 3, \dots$, then it follows from axiom (3) that

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

Since $P(\emptyset) \geq 0$ by axiom (1), then $P(\emptyset) = 0$.

Proposition 1.4.2 For every finite sequence of n disjoint events $A_1, \dots, A_n$,

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

Proof: Consider the infinite sequence of events $A_1, A_2, \dots$, in which $A_1, \dots, A_n$ are the n given disjoint events and $A_i = \emptyset$ for $i = n+1, n+2, \dots$. By axiom (3) and Proposition 1.4.1,

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^n P(A_i) + 0 \\ &= \sum_{i=1}^n P(A_i). \end{aligned}$$

Proposition 1.4.3 For every event A , $P(A^c) = 1 - P(A)$.

Proof: A and A^c are disjoint events and $A \cup A^c = S$, $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$, therefore $P(A^c) = 1 - P(A)$.

Proposition 1.4.4 If $A \subset B$, then $P(B - A) = P(B) - P(A)$, $P(A) \leq P(B)$.

Proof: As illustrated in the Figure 1.8, the event B may be treated as the union of the two disjoint events,

$$B = A \cup (B \cap A^c) = A \cup (B - A),$$

where A and $B - A$ are disjoint, therefore,

$$P(B) = P(A) + P(B - A) \geq P(A).$$

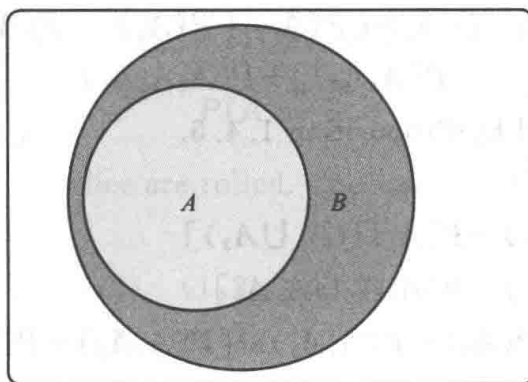


Figure 1.8

Proposition 1.4.5 For every event A , $0 \leq P(A) \leq 1$.

Proof: $A \subset S$, by Proposition 1.4.4 and axiom (2), $P(A) \leq P(S) = 1$.

Proposition 1.4.6 For every two events A and B ,

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

Proof: We have $A \cup B = A \cup BA^c$, where A and BA^c are disjoint, see Figure 1.9.

Hence, by Proposition 1.4.2 and Proposition 1.4.4,

$$P(A \cup B) = P(A) + P(BA^c) = P(A) + P(B) - P(AB).$$

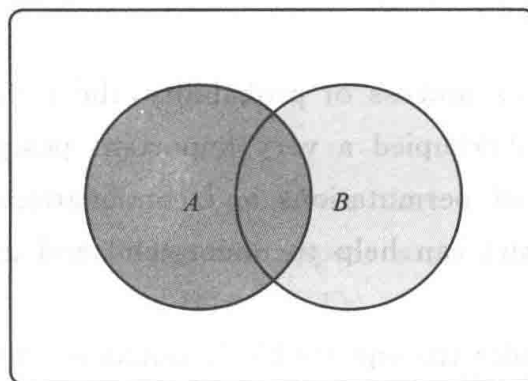


Figure 1.9

Proposition 1.4.6 can be extended to three or more events.

Proposition 1.4.7 For every three events A_1, A_2 and A_3 , see Figure 1.10,

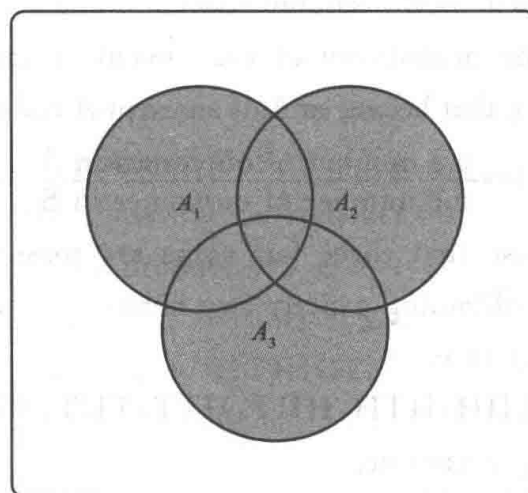


Figure 1.10

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_2A_3) + P(A_1A_3)] + P(A_1A_2A_3).$$

Proof: It can be derived by Proposition 1.4.6.

$$\begin{aligned} & P(A_1 \cup A_2 \cup A_3) \\ &= P(A_1) + P(A_2 \cup A_3) - P[A_1 \cap (A_2 \cup A_3)] \\ &= P(A_1) + P(A_2 \cup A_3) - P[A_1A_2 \cup A_1A_3] \\ &= P(A_1) + P(A_2) + P(A_3) - P(A_2A_3) - [P(A_1A_2) + P(A_1A_3) - P(A_1A_2A_3)] \\ &= P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_2A_3) + P(A_1A_3)] + P(A_1A_2A_3). \end{aligned}$$

Proposition 1.4.8 For every n events $A_1, \dots, A_n$,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_iA_j) + \sum_{1 \leq i < j < k \leq n} P(A_iA_jA_k) - \dots + (-1)^{n+1} P(A_1A_2 \dots A_n).$$

Proof: The conclusion can be proved by mathematical induction.

◆ 1.5 Equally Likely Outcomes Model ◆

Early in the development process of probability, the equally likely outcomes model (classical probability model) occupied a very important position. Its calculation mainly depends on the knowledge of permutations and combinations. Although its formula is simple, it is very intuitive and can help to understand and grasp the basic concepts of probability theory.

In this section, we consider the equally likely outcomes model. The experiments have the two features:

- (1) The sample space S contains only a finite number of points $s_1, \dots, s_n$.
- (2) The probability assigned to each of the outcomes $s_1, \dots, s_n$ is $\frac{1}{n}$.

The sample space is called simple sample space.

Proposition 1.5.1 The probability of each event A can be found by adding the probabilities of all outcomes s_i that belong to A . If an event A contains exactly m outcomes, then

$$P(A) = \frac{\text{the number of outcomes in } A}{\text{the number of outcomes in } S} = \frac{m}{n}.$$

Example 1.5.1 Suppose that three fair coins are tossed simultaneously. We shall determine the probability of obtaining exactly two heads.

Solution: The sample space is

$$S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}.$$

The probabilities of the basic events are

$$P(\{HHH\}) = \dots = P(\{TTT\}) = \frac{1}{8}.$$

$A = \{\text{obtaining exactly two heads}\} = \{HHT, HTH, THH\}$, therefore,

$$P(A) = \frac{3}{8}.$$

Example 1.5.2 Two fair dice are rolled. The sample space S comprises the following 36 outcomes:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

S is a simple sample space and that the probability of each of these outcomes is $\frac{1}{36}$.

(1) What is the probability that the sum of the two numbers is 6?

(2) What is the probability that the sum of the two numbers that appear is larger than 2?

Solution: (1) Let A be the sum of the two numbers is 6. A comprises the following 5 outcomes.

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

Therefore, $P(A) = \frac{5}{36}$.

(2) Let B be the sum of the two numbers that appear is larger than 2, then B^c represents the sum of the two numbers that appear is not larger than 2. That is,

$$B^c = \{(1,1)\}.$$

Therefore, $P(B) = 1 - P(B^c) = 1 - \frac{1}{36} = \frac{35}{36}$.

Remark 1.5.1 When using the basic formula in Proposition 1.5.1 to calculate the probability of a random event, we need analyze the event content, determine the sample points contained in the sample space, count the number of basic events in the event considered, and then by the formula to calculate the probability.

In these calculations, some permutation and combination formulas are often used. For the convenience of the readers, we present some basic formulas for counting methods.

Proposition 1.5.2 (Multiplication Rule) Suppose that an experiment has k steps ($k \geq 2$), that the i -th step can have n_i possible outcomes ($i = 1, \dots, k$), then the total number of elements in S will be equal to the product $n_1 \cdot n_2 \cdot \dots \cdot n_k$.

Example 1.5.3 Suppose that we toss six coins. Each outcome in S will consist of a sequence of six heads and tails, such as "HTHHHT". There are two possible outcomes for each of the six coins, the number of sample points in S is $2^6 = 64$. The event that observing exactly one head contains 6 sample points. Therefore,

$$P(\text{exactly one head}) = \frac{6}{64} = \frac{3}{32}.$$

Proposition 1.5.3 (Permutations) Select k of the elements each time **without replacement** from a set of n elements. Let each outcome consist of the k elements in the order selected. Each such outcome is called a permutation, the number of distinct such permutations is

$$P_n^k = n \times (n-1) \times \cdots \times (n-k+1).$$

When $k=n$, the number of possible permutations will be the number

$$P_n^n = n! = n \times (n-1) \times \cdots \times 2 \times 1.$$

$0!$ is defined by the relation

$$0! = 1.$$

Example 1.5.4 Suppose that a club consists of 25 members and that a president and a secretary are to be chosen from the membership. We shall determine the total possible number of ways in which these two positions can be filled.

Solution: $P_{25}^2 = 25 \times 24 = 600.$

Example 1.5.5 Suppose that six different books are to be arranged on a shelf. The number of possible permutations of the books is

$$6! = 720.$$

Proposition 1.5.4 (Combinations) Consider a set with n elements. Each subset of size k chosen from this set is called a combination, the number of distinct such combinations is

$$C_n^k = \frac{P_n^k}{k!} = \frac{n!}{k!(n-k)!}.$$

Example 1.5.6 Consider the set $\{a, b, c, d\}$ containing the four different letters. We want to count the number of distinct subsets of size two.

$$\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}.$$

$$C_4^2 = \frac{4!}{2! \times 2!} = 6.$$

This is different from counting permutations because $\{a, b\}$ and $\{b, a\}$ are the same subset.

Example 1.5.7 (1) Suppose that a committee composed of 8 people is to be selected from a group of 20 people. The number of different groups of people that might be on the committee is

$$C_{20}^8 = \frac{20!}{8! \times 12!} = 125970$$

(2) Suppose that the 8 people in the committee each get a different job to perform on the committee. The number of ways is

$$P_{20}^8 = C_{20}^8 \times 8! = 5079110400.$$

Remark 1.5.2 The number C_n^k is also denoted by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = C_n^k.$$

It is also named as binomial coefficient and used in the following result. For all numbers x and y and each positive integer n ,

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$



Proposition 1.5.5 The number of different arrangements of n objects consisting of k similar objects of one type and $n-k$ similar objects of a second type is $C_n^k = \binom{n}{k}$.

Example 1.5.8 Suppose that a fair coin is to be tossed 10 times, and it is desired to determine the probability of obtaining exactly 3 heads.

Solution: The total possible number of different sequences of 10 heads and tails is 2^{10} , and it may be assumed that each of these sequences is equally probable. The number of these sequences that contain exactly 3 heads will be equal to the number of different arrangements that can be

formed with 3 heads and 7 tails, $\binom{10}{3}$. Therefore, $P = \frac{\binom{10}{3}}{2^{10}} = 0.1172$.

Example 1.5.9 A box contains 4 white balls and 2 black balls. 2 balls are drawn with replacement. (1) What is the probability of getting all white balls? (2) If 2 balls are drawn without replacement, what is the probability of getting all white balls?

Solution: (1) For the first time, there are 6 balls to be selected. It is sampling with replacement, there are also 6 balls to be selected for the second time. By multiplication rule, there are 6×6 ways in the sample space. Each of these possible ways is equally probable. For the event that getting all white balls, there are 4×4 ways. Therefore,

$$P = \frac{4 \times 4}{6 \times 6} = \frac{4}{9}.$$

(2) For the first time, there are 6 balls to be selected. It is sampling without replacement, there are only 5 balls to be selected for the second time. By multiplication rule, there are 6×5 ways in the sample space. For the event that getting all white balls, there are 4×3 ways. Therefore,

$$P = \frac{4 \times 3}{6 \times 5} = \frac{2}{5}.$$

Another method, suppose we don't distinguish between balls of the same color and don't care about the order in which the balls are selected, then there are C_6^2 ways to select 2 balls from 6 balls. And there are C_4^2 ways to select 2 white balls from 4 white balls. Therefore,

$$P = \frac{C_4^2}{C_6^2} = \frac{2}{5}.$$

Example 1.5.10 Suppose that there are 25 persons in a room, and each person's birthday is uniformly distributed over the 365 days (ignore leap year) of the year and independent of the other persons' birthday. What is the probability that no two persons share a birthday?

Solution: By multiplication rule, there are 365^{25} ways for 25 persons' birthdays. No two persons share a birthday, the first person has a probability of 365 days as being his birthday. The second person has 364 different days that could be his and not shared with the first person. The next person has only 363 open days. Therefore,

$$P = \frac{365 \times 364 \times \cdots \times (365 - 25 + 1)}{365^{25}}.$$

Example 1.5.11 Suppose that a deck of 52 cards containing four aces is shuffled

thoroughly and the cards are then distributed among four players so that each player receives 13 cards. We shall determine the probability that each player will receive one ace. See Figure 1. 11.

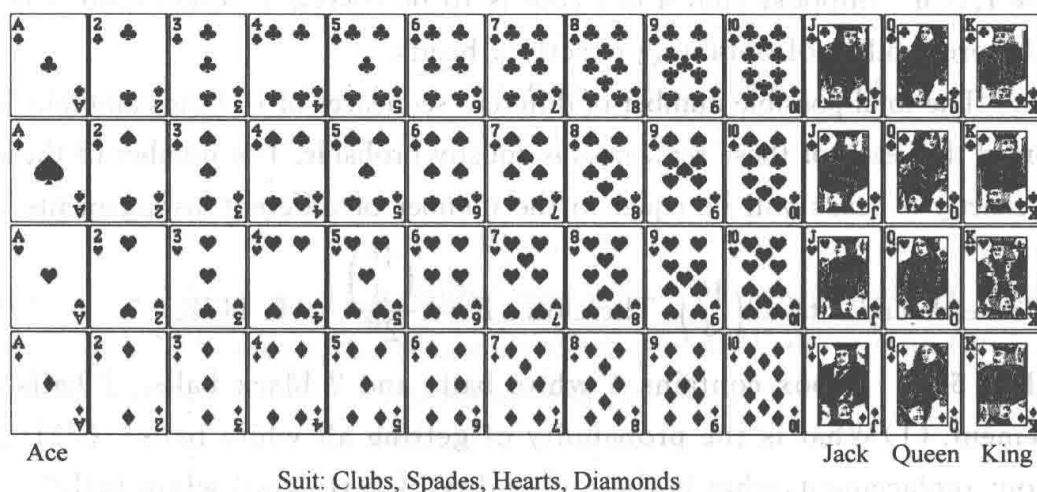


Figure 1. 11

Solution:

Method 1: Using unordered samples, the number of possible different combinations of the four positions occupied by the four aces is $\binom{52}{4}$. If each player is to receive one ace, then there are 13 possible positions for the ace that the first player is to receive, 13 other possible positions for the ace that the second player is to receive, and so on. Therefore, exactly 13^4 combinations will lead to the desired result. Hence,

$$P = \frac{13^4}{\binom{52}{4}} = 0.1055.$$

Method 2: Using ordered samples, we distinguish outcomes with the same cards in different orders, there are 13^4 ways to choose the four positions for the four aces, $4!$ ways to arrange the four aces in these four positions, $48!$ ways to arrange the remaining 48 cards in the 48 remaining positions.

$$P = \frac{4! \times 13^4 \times 48!}{52!} = 0.1055.$$

Method 3: Consider the four players. For the first player, there are $\binom{52}{13}$ ways to receive the 13 cards. For the second player, there are $\binom{39}{13}$ ways to receive the 13 cards from the remaining 39 cards, and so on. The sample space contains $\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13} \times \binom{13}{13}$ elements. If each player is to receive one ace, $4!$ ways to distribute the four aces to the four players. There are $\binom{48}{12}$ ways to choose 12 cards and distribute them to the first player to ensure that he has 13 cards. There are $\binom{36}{12}$ ways to choose 12 cards and distribute them to

the second player, and so on. Therefore,

$$P = \frac{4! \times \binom{48}{12} \times \binom{36}{12} \times \binom{24}{12} \times \binom{12}{12}}{\binom{52}{13} \times \binom{39}{13} \times \binom{26}{13} \times \binom{13}{13}} = 0.1055.$$

Example 1.5.12 Suppose that a person types n letters, types the corresponding addresses on n envelopes, and then places the n letters in the n envelopes in a random manner. Determine the probability that at least one letter will be placed in the correct envelope.

Solution: Let A_i be the event that letter i is placed in the correct envelope ($i=1, \dots, n$), we shall determine the value $P\left(\bigcup_{i=1}^n A_i\right)$.

$$\begin{aligned} \sum_{i=1}^n P(A_i) &= n \times \frac{1}{n} = 1, \\ \sum_{1 \leq i < j \leq n} P(A_i A_j) &= \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2!}, \\ \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) &= \binom{n}{3} \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}, \\ &\vdots \end{aligned}$$

Therefore, by Proposition 1.4.8,

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \\ &\quad \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) - \dots + (-1)^{n+1} P(A_1 A_2 \dots A_n) \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n+1} \frac{1}{n!}. \end{aligned}$$

Example 1.5.13 Suppose that 7 guests check their hats when they arrive at a restaurant, and that these hats are returned to them in a random order when they leave. Determine the probability that (1) no guest will receive the proper hat, (2) exactly 2 guests will receive their own hats.

Solution: (1) Let A_i be the event that guest i receives his own hat ($i=1, \dots, n$), we shall determine the value

$$\begin{aligned} P(n) &= P(A_1^c \cap A_2^c \dots \cap A_n^c) \\ &= P\left[\left(\bigcup_{i=1}^n A_i\right)^c\right] \\ &= 1 - P\left(\bigcup_{i=1}^n A_i\right) \\ &= 1 - \left[1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n+1} \frac{1}{n!}\right] \\ &= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^n \frac{1}{n!}. \end{aligned}$$

Then, $n=7$, the probability that no guest will receive the proper hat is

$$P(7) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} = \frac{103}{280} = 0.3679.$$

(2) Let X be the number of guests who receive their own hats.

$$\begin{aligned} P(X=2) &= C_7^2 \times \frac{1}{7} \times \frac{1}{6} \times P(5) \\ &= \frac{1}{2} \times \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} \right) \\ &= \frac{11}{60} = 0.1833. \end{aligned}$$

In fact, we can compute all the probabilities $P(X=k), k=0, 1, \dots, 7$.

$$\begin{aligned} P(X=0) &= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \frac{1}{7!} = \frac{103}{280}, \\ P(X=1) &= C_7^1 \times \frac{1}{7} \times \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} \right) = \frac{53}{144}, \\ P(X=2) &= C_7^2 \times \frac{1}{7} \times \frac{1}{6} \times \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} \right) = \frac{11}{60}, \\ P(X=3) &= C_7^3 \times \frac{1}{7} \times \frac{1}{6} \times \frac{1}{5} \times \left(\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \right) = \frac{1}{16}, \\ P(X=4) &= C_7^4 \times \frac{1}{7} \times \frac{1}{6} \times \frac{1}{5} \times \frac{1}{4} \times \left(\frac{1}{2!} - \frac{1}{3!} \right) = \frac{1}{72}, \\ P(X=5) &= C_7^5 \times \frac{1}{7} \times \frac{1}{6} \times \frac{1}{5} \times \frac{1}{4} \times \frac{1}{3} \times \left(\frac{1}{2!} \right) = \frac{1}{240}, \\ P(X=6) &= 0, \\ P(X=7) &= \frac{1}{7!} = \frac{1}{5040}. \end{aligned}$$

In this section, we consider the equally likely outcomes model (classical probability model). The sample space considered is finite. For the experiments which have infinite outcomes, the original definition of classical model is not enough and can be extended. One such model is geometric probability model.

In general, we may think of geometric probabilities as non-negative quantities (not exceeding 1) being assigned to subregions of a given domain subject to certain rules. The probability of an event is based on a ratio of geometric measures such as length, area or volume.

Proposition 1.5.6 Suppose the sample space of an experiment is a given domain $\Omega \in \mathbb{R}^n$. The outcomes are points in Ω and are equally likely to occur. Let A be any subregion of Ω . Then the probability that basic events fall into A is

$$P(A) = \frac{S_A}{S_\Omega},$$

where S_A, S_Ω are the geometric measures of A and Ω , respectively.

Example 1.5.14 Two friends agree to meet up in a station. They arrive to the station uniformly between 12:00pm and 13:00pm. Suppose that after arriving randomly, each waits 30 minutes for the other person before leaving. What is the chance the two will meet?

Solution: Let x denote the time when one person arrives, and y denote the time when

the other arrives, we can use the notation (x, y) to denote the time in hours, after 12:00pm, that each person arrives, where $0 \leq x \leq 1, 0 \leq y \leq 1$. The sample space Ω is a square with side length 1. Any coordinate in the square represents a time that the pair arrives.

What is the set of coordinates for which the two will meet? We want the two coordinates to be within 30 minutes or $\frac{1}{2}$ hours of each other. That is, we want all coordinates such that $|x - y| \leq \frac{1}{2}$. We draw the lines $y = x + \frac{1}{2}$ and $y = x - \frac{1}{2}$ and shade the area in between these two lines to denote the event A . See Figure 1.12. Therefore,

$$P(A) = \frac{S_A}{S_\Omega} = \frac{\frac{3}{4}}{1} = \frac{3}{4}.$$

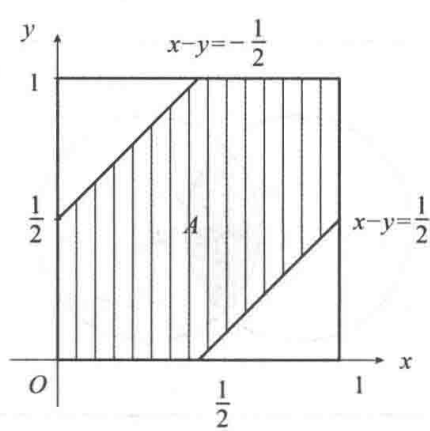


Figure 1.12

Remark 1.5.3 What is the chance the two will arrive at the same time? The set is a line $x = y$ which has no area. The probability that the two will arrive at the same time is zero. But clearly they can arrive at the same time, for example, both arrive at 12:30pm, etc. It implies that zero probability does not mean impossible.

◆ 1.6 Conditional Probability ◆

Conditional probability is an important concept in probability theory and has wide applications in practice.

Example 1.6.1 In a group of 100 people, the numbers who do or do not smoke and do or do not have cancer are reported as shown in Table 1.1.

Table 1.1

	Smoke	Not Smoke	Total
Cancer	20	5	25
No Cancer	30	45	75
Total	50	50	100

If a people is selected at random and it is found that he smokes, what is the probability that he has cancer?

Solution: The sample space contains 100 people. Since additional information is given, the sample space is reduced. There are only 50 elements left. The probability is $\frac{20}{50} = 0.4$.

Definition 1.6.1 Suppose that we learn that an event B has occurred and that we wish to compute the probability of another event A taking into account that we know that B has occurred. See Figure 1.13. The new probability of A is called the conditional probability of the event A given that the event B has occurred and is denoted as $P(A|B)$. If $P(B) > 0$, we compute this probability as

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

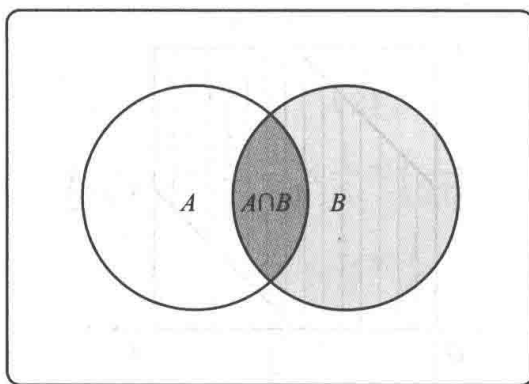


Figure 1.13

Remark 1.6.1 Unconditional probability $P(A)$ can be regarded as the conditional probability of the event A given that the certain event or sample space Ω has occurred, since $P(A|\Omega) = \frac{P(A\Omega)}{P(\Omega)} = P(A)$.

Example 1.6.2 You know only that your colleague has three children. You ask him if he is lucky enough to have at least one daughter and he says yes. What is the probability that all three children are girls?

Solution: Let A be the event that all three children are girls and let B be the event that he has at least one daughter. We shall determine $P(A|B)$.

$$A \subset B, A \cap B = A, P(AB) = P(A) = \frac{1}{8}.$$

$$P(B) = 1 - P(B^c) = 1 - \frac{1}{8} = \frac{7}{8}.$$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{8}}{\frac{7}{8}} = \frac{1}{7}.$$

Example 1.6.3 Suppose that two dice were rolled and it was observed that the sum T of the two numbers was odd. We shall determine the probability that T was less than 8.

Solution: If we let A be the event that $T < 8$ and let B be the event that T is odd, then $A \cap B$ is the event that T is 3, 5 or 7.

$$P(AB) = P(T=3) + P(T=5) + P(T=7)$$

$$= \frac{2}{36} + \frac{4}{36} + \frac{6}{36}$$

$$= \frac{1}{3}.$$

$$P(B) = P(T=3) + P(T=5) + \dots + P(T=11)$$

$$= \frac{2}{36} + \frac{4}{36} + \dots + \frac{2}{36}$$

$$= \frac{1}{2}.$$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Example 1.6.4 Suppose that two dice are to be rolled repeatedly and the sum T of the two numbers is to be observed for each roll. We shall determine the probability p that the value $T=7$ will be observed before the value $T=8$ is observed.

Solution:

Method 1: We could assume that the sample space S contains all sequences of outcomes that terminate as soon as either the sum $T=7$ or the sum $T=8$ is obtained. Then we could find the sum of the probabilities of all the sequences that terminate when the value $T=7$ is obtained. For each roll, we have

$$P(T=7) = \frac{1}{6}, P(T=8) = \frac{5}{36}, P(T \neq 7 \text{ and } T \neq 8) = 1 - \frac{11}{36} = \frac{25}{36}.$$

Let p_i be the probability that the first 7 occurs on the i -th position and no 7 or 8 occurs before the i -th position.

$$p_1 = \frac{1}{6}, p_2 = \frac{25}{36} \times \frac{1}{6}, p_3 = \left(\frac{25}{36}\right)^2 \times \frac{1}{6}, \dots$$

Therefore,

$$p = p_1 + p_2 + \dots = \frac{1}{6} \times \frac{1}{1 - \frac{25}{36}} = \frac{6}{11}.$$

Method 2: If we repeat the experiment until either the sum $T=7$ or the sum $T=8$ is obtained, the effect is to restrict the outcome of the experiment to one of these two values. When we get either 7 or 8, we get the answer to the question 'Is 7 observed before 8?' Once we get either 7 or 8, we know the result of the run. Any rolls we make subsequently will not affect the outcome, so we might ignore them.

Hence, the problem can be restated as follows: given that the outcome of the experiment is either $T=7$ or $T=8$, determine the probability p that the outcome is actually $T=7$.

$$p = P(T=7 | T=7 \text{ or } T=8) = \frac{P(T=7)}{P(T=7 \text{ or } T=8)} = \frac{\frac{1}{6}}{\frac{5}{36} + \frac{1}{6}} = \frac{6}{11}.$$

Proposition 1.6.1 Similar to probability, conditional probability satisfies the three axioms: if $P(B) > 0$,

- (1) Nonnegativity: for every event A , $P(A|B) \geq 0$;
- (2) Normality: $P(S|B) = 1$,
- (3) Countable additivity: for every infinite sequence of disjoint events $A_1, A_2, \dots$,

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

In addition, conditional probability also holds the main properties of probability such as

$$P(A|B) + P(A^c|B) = 1.$$

Theorem 1.6.1 (The Multiplication Rule for Conditional Probabilities) Let A and B be events. If $P(B) > 0$, then

$$P(A \cap B) = P(B) \cdot P(A|B).$$

If $P(A) > 0$, then

$$P(A \cap B) = P(A) \cdot P(B|A).$$

Proof: It follows from the definition of conditional probability.

Example 1.6.5 Consider a bag of marbles, containing 10 red, 20 blue, and 15 green marbles. Suppose that two marbles are drawn without replacement. What is the probability that both marbles drawn are red?

Solution: Let A be the event that the first marble is red, and let B be the event that the second marble is red.

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{10}{45} \times \frac{9}{44} = \frac{1}{22}.$$

Theorem 1.6.2 Suppose that $A_1, A_2, \dots, A_n$ are events such that $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Then

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

Proof: It follows from the definition of conditional probability and mathematical induction.

Example 1.6.6 Three cards are dealt successively at random and without replacement from a standard deck of 52 playing cards. What is the probability of receiving in order a king, a queen, and a jack?

Solution: Let $A_i (i=1, 2, 3)$ denote the event of receiving in order a king, a queen, and a jack. Then

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \\ &= \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} \\ &= \frac{64}{132600}. \end{aligned}$$

◆ 1.7 Total Probability and Bayes' Theorem ◆

Total probability and Bayes' theorem are two important formulas in probability and statistics. They decompose the complex and comprehensive random events into several mutually exclusive simple events. The whole problem is transformed into local problems. One kind of problem is to obtain the desired result by calculating the probability of these simple events, and it is called total probability. Another type of problem is to discuss the probability of the occurrence of each simple event under the condition that the result is given, which is known as Bayes' theorem.

Definition 1.7.1 Let S denote the sample space of some experiments, and consider n events $B_1, \dots, B_n$ in S such that

(1) $B_1, \dots, B_n$ are disjoint;

(2) $\bigcup_{i=1}^n B_i = S$.

Then these events are called a **partition** of S .

Theorem 1.7.1 (Total Probability) Suppose that the events $B_1, \dots, B_n$ form a partition of the space S and $P(B_i) > 0$ for $i = 1, \dots, n$. Then, for every event A in S , see

Figure 1.14, $P(A) = \sum_{i=1}^n P(B_i)P(A|B_i)$.

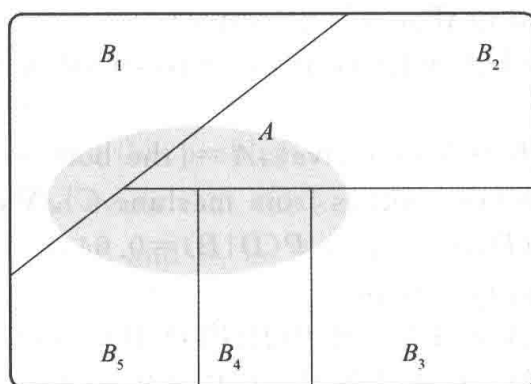


Figure 1.14

Proof:

$$\begin{aligned}
 P(A) &= P(A \cap S) \\
 &= P\left[A \cap \left(\bigcup_{i=1}^n B_i\right)\right] \\
 &= P\left[\bigcup_{i=1}^n (A \cap B_i)\right] \\
 &= \sum_{i=1}^n P(A \cap B_i) \\
 &= \sum_{i=1}^n P(B_i)P(A|B_i).
 \end{aligned}$$

Theorem 1.7.2 (Bayes' Theorem) Suppose that the events $B_1, \dots, B_n$ form a partition of

the space S and $P(B_i) > 0$ for $i=1, \dots, n$. And let A be an event such that $P(A) > 0$, then,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}, i = 1, \dots, n.$$

Proof:

$$\begin{aligned} P(B_i|A) &= \frac{P(AB_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}, i = 1, \dots, n. \end{aligned}$$

Remark 1.7.1 In the case where B and B^c form a partition of S , the simplified versions of total probability and Bayes' Theorem are as shown below:

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c),$$

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

Example 1.7.1 A factory production line is manufacturing bolts using three machines A, B and C . Of the total output, machine A is responsible for 25%, machine B for 35% and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C . A bolt is chosen at random from the factory production line.

(1) What is the probability that it is defective?

(2) If we learn that the bolt selected is defective, what is the probability that it came from machine A ?

Solution: Let $D = \{\text{the bolt is defective}\}$, $A = \{\text{the bolt is from machine } A\}$, $B = \{\text{the bolt is from machine } B\}$, $C = \{\text{the bolt is from machine } C\}$. We know that $P(A) = 0.25$, $P(B) = 0.35$, $P(C) = 0.4$, $P(D|A) = 0.05$, $P(D|B) = 0.04$, $P(D|C) = 0.02$.

(1) By the total probability formula,

$$\begin{aligned} P(D) &= P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) \\ &= 0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02 \\ &= 0.0345. \end{aligned}$$

(2) By the Bayes' formula,

$$\begin{aligned} P(A|D) &= \frac{P(AD)}{P(D)} \\ &= \frac{P(A)P(D|A)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} \\ &= \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02} \\ &= 0.3623. \end{aligned}$$

Example 1.7.2 One bag contains 4 white balls and 3 black balls, and the other bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. (1) What is the probability that a ball now drawn from the second bag is black? (2) If a ball drawn from the second bag is black, what is the

probability that the ball drawn from the first bag is white?

Solution: Let A be the event that a ball drawn from the second bag is black, B be the event that a white ball is drawn from the first bag and B^c be the event that a black ball is drawn from the first bag. Then $P(B) = \frac{4}{7}$, $P(B^c) = \frac{3}{7}$, $P(AB) = \frac{5}{9}$, $P(AB^c) = \frac{6}{9}$.

(1) By the total probability formula,

$$\begin{aligned} P(A) &= P(B)P(A|B) + P(B^c)P(A|B^c) \\ &= \frac{4}{7} \times \frac{5}{9} + \frac{3}{7} \times \frac{6}{9} \\ &= \frac{38}{63}. \end{aligned}$$

(2) We need find $P(B|A)$. By the Bayes' formula,

$$\begin{aligned} P(B|A) &= \frac{P(BA)}{P(A)} \\ &= \frac{P(B)P(A|B)}{P(A)} \\ &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} \\ &= \frac{\frac{4}{7} \times \frac{5}{9}}{\frac{38}{63}} \\ &= \frac{10}{19}. \end{aligned}$$

Example 1.7.3 Suppose that two dice are to be rolled repeatedly and the sum T of the two numbers is to be observed for each roll. We shall determine the probability p that the value $T=7$ will be observed before the value $T=8$ is observed.

Solution: We have discussed this problem in Example 1.6.4. Now we want to propose another solution. Let B_1 be the event that the sum for the first roll is 7, B_2 be the event that the sum for the first roll is 8 and B_3 be the event that the sum for the first roll is a value other than 7 or 8. Let A be the event that the value $T=7$ will be observed before the value $T=8$ is observed. By the law of total probability,

$$\begin{aligned} P(A) &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ &= \frac{1}{6} \times 1 + \frac{5}{36} \times 0 + \frac{25}{36} \times P(A). \end{aligned}$$

$$\text{Thus, } P(A) = \frac{6}{11}.$$

Example 1.7.4 Approximately 1% of women aged 40 ~ 50 have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without breast cancer has a 10% chance of a false positive result. What is the probability a woman has breast cancer given that she just had a positive test?

Solution: Let B be the event that the woman has breast cancer and A be the event that the test is positive. We know $P(B) = 0.01$, $P(A|B) = 0.90$, $P(A|B^c) = 0.10$. We need find $P(B|A)$. By the Bayes' formula,

$$\begin{aligned}
 P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} \\
 &= \frac{0.01 \times 0.90}{0.01 \times 0.90 + (1 - 0.01) \times 0.10} \\
 &= \frac{1}{12}.
 \end{aligned}$$

Remark 1.7.2 This answer is somewhat surprising and interesting. A positive test only means the woman has a $\frac{1}{12} \approx 8.3\%$ chance of breast cancer, rather than 90%. But it makes sense: the test gives a false positive 10% of the time, so there will be a lot of false positives if the population size is large enough. Most of the positive test results will be incorrect.

◆ 1.8 Independent Events ◆

In an experiment, the events usually have some relations, that is, the occurrence of one event can affect that of another one. If the outcome of one event does not affect the outcome of another, the events are independent.

Definition 1.8.1 Two events A and B are independent if

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

Theorem 1.8.1 If two events A and B are independent, and $P(B) > 0$, then $P(A|B) = P(A)$.

Proof: It can be proved by the definition of conditional probability.

Example 1.8.1 Suppose that a balanced die is rolled twice. Find the probability of getting at least one number which is not less than 5.

Solution: Let A be the event that the number on the first roll is not less than 5 and B be the event that the number on the second roll is not less than 5. A and B are physically unrelated and independent.

$$\begin{aligned}
 P(A \cup B) &= P(A) + P(B) - P(AB) \\
 &= P(A) + P(B) - P(A)P(B) \\
 &= \frac{1}{3} + \frac{1}{3} - \frac{1}{3} \times \frac{1}{3} \\
 &= \frac{5}{9}.
 \end{aligned}$$

Example 1.8.2 Suppose that a balanced die is rolled. Let A be the event that an even number is obtained, and let B be the event that one of the numbers 1, 2, 3 or 4 is obtained. Show that the events A and B are independent.

Proof: The sample space $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{1, 2, 3, 4\}$, $A \cap B = \{2, 4\}$.

$$P(A) = \frac{1}{2}, P(B) = \frac{2}{3}, P(AB) = \frac{1}{3}$$

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

It follows that the events A and B are independent events, even though the occurrence of each event depends on the same roll of a die.

Theorem 1.8.2 If two events A and B are independent, then the events A and B^c , A^c and B , A^c and B^c are also independent.

Proof:

$$\begin{aligned} P(AB^c) &= P(A) - P(AB) \\ &= P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] \\ &= P(A)P(B^c). \end{aligned}$$

Therefore, A and B^c are independent. The proofs for the remaining cases are similar.

Theorem 1.8.3 Suppose that A and B are disjoint events for an experiment, each with positive probability. Then A and B are dependent.

Proof: A and B are disjoint, $A \cap B = \emptyset$, $P(AB) = 0$, $P(A)P(B) > 0$, so $P(AB) \neq P(A)P(B)$. A and B are dependent.

Definition 1.8.2 Three events A , B and C are mutually independent if the following four relations are satisfied:

$$\begin{aligned} P(AB) &= P(A)P(B), \\ P(AC) &= P(A)P(C), \\ P(BC) &= P(B)P(C), \\ P(ABC) &= P(A)P(B)P(C). \end{aligned}$$

Three events A , B and C are pairwise independent if the first three relations above are satisfied.

Remark 1.8.1 Mutually and pairwise independence are two different concepts. Mutually independence implies pairwise independence. But the converse is not true.

Example 1.8.3 Suppose that a fair coin is tossed twice. The sample space $S = \{HH, HT, TH, TT\}$. Consider the following three events:

$$\begin{aligned} A &= \{\text{Head on first toss}\} = \{HH, HT\}, \\ B &= \{\text{Head on second toss}\} = \{HH, TH\}, \\ C &= \{\text{Both tosses the same}\} = \{HH, TT\}. \end{aligned}$$

Then $AB = AC = BC = ABC = \{HH\}$. Hence,

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

$$P(AB) = P(AC) = P(BC) = P(ABC) = \frac{1}{4},$$

$$P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(B)P(C),$$

$$P(ABC) \neq P(A)P(B)P(C).$$

So A , B and C are pairwise independent, but not mutually independent.

Definition 1.8.3 The n events $A_1, \dots, A_n$ are independent (or mutually independent) if, for every subset $A_{i_1}, \dots, A_{i_k}$ of k of these events ($k=2, 3, \dots, n$),

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}).$$

◆ Exercise 1 ◆

1.1 Suppose that one card is to be selected from a deck of 20 cards that contains 10 red cards numbered from 1 to 10 and 10 blue cards numbered from 1 to 10. Let A be the event that a card with an even number is selected, let B be the event that a blue card is selected, and let C be the event that a card with a number less than 5 is selected. Describe the sample space S and describe each of the following events both in words and as subsets of S :

(1) $A \cap B \cap C$; (2) $B \cap C^c$; (3) $A \cup B \cup C$; (4) $A \cap (B \cup C)$; (5) $A^c \cap B^c \cap C^c$.

1.2 Two dice are rolled. Let A be the event that the sum of the two dice is odd, let B be the event that at least one of the dice has a 1 showing, and let C be the event that the sum is 5. Describe the events $A \cap B$, $A \cup B$, $B \cap C$, $A \cap B^c$ and $A \cap B \cap C$.

1.3 Consider the set $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and its subsets $A = \{1, 3, 5, 7, 9\}$, $B = \{2, 4, 6, 8\}$, $C = \{1, 2, 3, 4\}$, $D = \{7, 8\}$. Describe the following sets:

(1) B^c ; (2) $A \cap D$; (3) $A \cup B$; (4) $(C \cup D)^c$; (5) $A \cap B \cup D$.

1.4 Suppose A, B, C are three events, and $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(AB) = P(BC) = 0$, $P(AC) = \frac{1}{8}$, determine the probability that at least one of the three events will occur.

1.5 An urn contains 5 white balls and 3 red balls. Draw 4 balls without replacement. Find the probability that 3 white balls are drawn.

1.6 In a group of 20 persons, 5 are females. If 3 persons are picked out of 20 at random, what is the probability that they all are females?

1.7 Consider two events A and B with $P(A) = 0.6$ and $P(B) = 0.7$. Determine the maximum and minimum possible values of $P(A \cap B)$ and the conditions under which each of these values is attained.

1.8 Ten persons in a room are wearing badges marked 1 through 10. Three persons are chosen at random and the numbers of their badges are recorded. Find the probability:

(1) the smallest number is 5; (2) the largest number is 5.

1.9 Choose randomly an integer from $1 \sim 2000$. Find the probability:

(1) the selected number is divisible by 6; (2) the selected number is divisible by 8; (3) the selected number is divisible by 6 and 8.

1.10 A closet contains n pairs of shoes. If $2r$ shoes are chosen at random (where $2r <$



n), what is the probability that (1) the chosen shoes will contain no matching pair?
(2) there will be exactly one complete pair?

1.11 If event A is drawing a queen from a deck of 52 cards and event B is drawing a king from the remaining cards, are events A and B dependent or independent?

1.12 Bag A contains 9 red marbles and 3 green marbles. Bag B contains 9 black marbles and 6 orange marbles. Find the probability of selecting one green marble from bag A and one black marble from bag B .

1.13 Consider two events A and B such that $P(A) = \frac{1}{3}$ and $P(B) = \frac{1}{2}$. Determine the value of $P(B \cap A^c)$ for each of the following conditions: (1) A and B are disjoint; (2) $A \subset B$; (3) $P(A \cap B) = \frac{1}{8}$.

1.14 If 50 percent of the families in a certain city subscribe to the morning newspaper, 65 percent of the families subscribe to the afternoon newspaper, and 85 percent of the families subscribe to at least one of the two newspapers, what percentage of the families subscribe to both newspapers?

1.15 Suppose that A , B and C are three independent events such that $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$, and $P(C) = \frac{1}{2}$.

(1) Determine the probability that none of these three events will occur.

(2) Determine the probability that exactly one of these three events will occur.

1.16 A box contains 100 items of which 4 are defective. Two items are chosen at random from the box. What is the probability of selecting

(1) 2 defectives if the first item is not replaced;

(2) 2 defectives if the first item is put back before choosing the second item;

(3) 1 defective and 1 non-defective if the first item is not replaced?

1.17 The probability that a student passes mathematics is $\frac{2}{3}$ and the probability that he passes english is $\frac{4}{9}$. If the probability that he will pass at least one subject is $\frac{4}{5}$, what is the probability that he will pass both subjects?

1.18 When two balanced dice are rolled, what is the probability that the sum of the numbers on the dice is 6 or 9?

1.19 An elevator has 4 passengers and stops at 7 floors. It is equally likely that a person will get off at any one of the 7 floors. Find the probability that no 2 passengers leave at the same floor.

1.20 The elevator of a four-floor building leaves the first floor with six passengers and stops at all of the remaining three floors. If it is equally likely that a passenger gets off at any of these three floors, what is the probability that, at each stop of the elevator, at

least one passenger departs?

1.21 Six fair dice are rolled. What is the probability that at least two of them show the same face?

1.22 If 10 balls are thrown at random into 20 boxes, what is the probability that no box will receive more than one ball?

1.23 Suppose that we pick three people at random. For each of the following questions, ignore the special case where someone might be born on February 29th, and assume that births are evenly distributed throughout the year.

(1) What is the probability that the first two people share a birthday?

(2) What is the probability that at least two people share a birthday?

1.24 A box contains 24 light bulbs, of which four are defective. If a person selects four bulbs from the box at random, without replacement, what is the probability that all four bulbs will be defective?

1.25 A box contains 24 light bulbs, of which four are defective. If one person selects 10 bulbs from the box in a random manner, and a second person then takes the remaining 14 bulbs, what is the probability that all four defective bulbs will be obtained by the same person?

1.26 A deck of 52 cards contains four aces. If the cards are shuffled and distributed in a random manner to four players so that each player receives 13 cards, what is the probability that all four aces will be received by the same player?

1.27 Suppose that five guests check their hats when they arrive at a restaurant, and that these hats are returned to them in a random order when they leave. Determine the probability that

(1) no guest will receive the proper hat.

(2) at least two of the guests receive their own hats.

1.28 A box contains 4 red balls and 8 blue balls. One ball is selected at random and its color is observed. The ball is then returned to the box and one additional ball of the same color is also put into the box. A second ball is then selected at random, its color is observed, and it is returned to the box together with one additional ball of the same color. Each time another ball is selected, the process is repeated. If four balls are selected, what is the probability that the first three balls will be red and the fourth ball will be blue?

1.29 A box contains three cards. One card is red on both sides, one card is green on both sides, and one card is red on one side and green on the other. One card is selected from the box at random, and the color on one side is observed. If this side is green, what is the probability that the other side of the card is also green?

1.30 What is the probability that the total of two dice will be greater than 8, given that the first die is a 6?

1.31 Given that 5% of men and 0.25% of women are color-blind. A person is randomly selected. Assume males and females to be in equal numbers.

- (1) What is the probability that the selected person is color-blind?
- (2) If the selected person is color-blind, what is the probability that he is male?

1.32 A machine produces defective parts with three different probabilities depending on its state of repair. If the machine is in good working order, it produces defective parts with probability 0.02. If it is wearing down, it produces defective parts with probability 0.1. If it needs maintenance, it produces defective parts with probability 0.3. The probability that the machine is in good working order is 0.8, the probability that it is wearing down is 0.1, and the probability that it needs maintenance is 0.1. Compute the probability that a randomly selected part will be defective.

1.33 Suppose that a medical test has a 92% chance of detecting a disease if the person has it and a 94% chance of correctly indicating that the disease is absent if the person really does not have the disease. Suppose that 10% of the population has the disease. (1) What is the probability that a randomly chosen person will test positive? (2) What is the probability of having the disease given a positive test?

1.34 Suppose that A and B are events such that $P(A) = 0.3$, $P(B) = 0.4$, $P(A|B) = 0.5$. Evaluate

- (1) $P(AB)$, $P(A \cup B)$.
- (2) $P(B|A)$.
- (3) $P(B|A \cup B)$.
- (4) $P(A^c \cup B^c | A \cup B)$.

1.35 Suppose that A , B and C are events such that A and B are independent, $P(A \cap B \cap C) = 0.04$, $P(C|A \cap B) = 0.25$, and $P(B) = 4P(A)$. Evaluate $P(A \cup B)$.

1.36 Suppose $0 < P(A) < 1$, $0 < P(B) < 1$, $P(A|B) + P(A^c|B^c) = 1$. Are events A and B independent?

1.37 Suppose that 13 cards are selected at random from a regular deck of 52 playing cards.

- (1) If it is known that at least one ace has been selected, what is the probability that at least two aces have been selected?
- (2) If it is known that the ace of hearts has been selected, what is the probability that at least two aces have been selected?

1.38 A biased coin (with probability of obtaining a head equal to $p > 0$) is tossed repeatedly and independently until the first head is observed. Compute the probability that the first head appears at an even numbered toss.

1.39 An engineering company advertises a job in three papers, A , B and C . It is known that these papers attract undergraduate engineering readerships in the proportions $2 : 3 : 1$. The probabilities that an engineering undergraduate sees and replies to the job advertisement in these papers are 0.002, 0.001 and 0.005 respectively. Assume that the undergraduate sees only one job advertisement.

- (1) If the engineering company receives only one reply to its advertisements, calculate

the probability that the applicant has seen the job advertised in paper A.

(2) If the company receives two replies, what is the probability that both applicants saw the job advertised in paper A?

1.40 Two players A and B are competing at a trivia quiz game involving a series of questions. On any individual question, the probabilities that A and B give the correct answer are p_A and p_B respectively, for all questions, with outcomes for different questions being independent. The game finishes when a player wins by answering a question correctly. Compute the probability that A wins if

(1) A answers the first question.

(2) B answers the first question.

Chapter 2

Random Variables and Distributions

Example 2.1.5 Consider an experiment where a coin is tossed and the spinner is spun only if the result of the coin toss is head. If the result is tail, $X = -1$, otherwise $X =$ the angle value of the spinner which are real numbers from the interval $[0, 360)$.

and continuous random variables here.

that are neither discrete nor continuous. See Example 2.1.5. We only introduce discrete take an interval of values. See Example 2.1.4. There also exists some random variables infinite. See Example 2.1.1, 2.1.2 and 2.1.3. Another type is continuous and they can on their values. One type is discrete and the values they can take are finite or countable

Definition 2.1.2 A general random variable is classified into two main types based on their values. One type is discrete and the values they can take are finite or countable

Example 2.1.4 A random variable X is said to be continuous if it takes values in an interval of real numbers. For example, the lifetime of a light bulb is a continuous random variable. It can take any real value in the interval $[0, \infty)$.

Example 2.1.3 Let X be the number of heads in n tosses of a fair coin. Then X is a discrete random variable. It can take only integer values from 0 to n .

Example 2.1.2 A random variable X is said to be discrete if it takes values in a countable set. For example, the number of heads in n tosses of a fair coin is a discrete random variable. It can take only integer values from 0 to n .

Example 2.1.1 A random variable X is said to be continuous if it takes values in an interval of real numbers. For example, the lifetime of a light bulb is a continuous random variable. It can take any real value in the interval $[0, \infty)$.

Example 2.1.1 A random variable X is said to be continuous if it takes values in an interval of real numbers. For example, the lifetime of a light bulb is a continuous random variable. It can take any real value in the interval $[0, \infty)$.

Example 2.1.1 A random variable X is said to be continuous if it takes values in an interval of real numbers. For example, the lifetime of a light bulb is a continuous random variable. It can take any real value in the interval $[0, \infty)$.

◆ 2.1 Random Variables ◆

In order to further study the random phenomena, the concept of random variable is introduced in this chapter. The purpose is to be able to use advanced mathematical tools such as calculus to fully reveal the statistical law of random phenomena.

For a random experiment, compared to its outcomes, we tend to be more concerned with a function of the outcomes. For example, rolling two dice, we often care about the sum of the two dice, and do not care about each face of the dice.

Definition 2.1.1 Let S be the sample space for an experiment. If for any $\omega \in S$, there exists a unique real number $X(\omega) \in \mathbb{R}$, then the real-valued function X is called a **random variable**.

$$S \xrightarrow{X} \mathbb{R}.$$

Remark 2.1.1 Random variables are usually denoted by capital letters $X, Y, Z, \dots$ or little greek letters $\xi, \eta, \zeta, \dots$. The values of a random variable depend on the outcomes of the experiment. The range of the values are known but the specific value cannot be predicted with certainty before the experiment is run. Random variables can be used to describe random events.

Example 2.1.1 A fair coin is tossed three times. Let X stand for the real-valued function defined on sample space S that counts the number of heads in each outcome. For example, if $s \in S$ is the sequence HHT, then $X(s) = 2$. The possible values for the function X are 0, 1, 2, 3. X is a random variable.

Example 2.1.2 A die is rolled. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Let $X = 1$ if the outcome is even and $X = 0$ if the outcome is odd. Let $Y = 1$ if the outcome is six and $Y = 0$ otherwise. X and Y are two random variables defined on the same sample space.

Example 2.1.3 Let X be the number of vehicles that pass through an intersection in a one-hour interval. X is a random variable and takes $\{0, 1, 2, \dots\}$.

Example 2.1.4 Measure the lifetime of a light bulb. The lifetime is a random variable. It can take all nonnegative values.

Remark 2.1.2 In general, random variables are classified into two main types based on their values. One type is discrete and the values they can take are finite or countable infinite. See Example 2.1.1, 2.1.2 and 2.1.3. Another type is continuous and they can take an interval of values. See Example 2.1.4. There also exists some random variables that are neither discrete nor continuous. See Example 2.1.5. We only introduce discrete and continuous random variables here.

Example 2.1.5 Consider an experiment where a coin is tossed and the spinner is spun only if the result of the coin toss is head. If the result is tail, $X = -1$, otherwise $X =$ the angle value of the spinner which are real numbers from the interval $[0, 360)$.

◆ 2.2 Cumulative Distribution Function ◆

Random variable X is a real-valued function of sample point ω . If A is a subset of real numbers, then $X \in A$ represents the following random event

$$\{\omega | X(\omega) \in A\}.$$

Specially, $\{X \leq a\}$, $\{X > a\}$ and $\{a < X < b\}$ are events. We are interested in the collection of all probabilities of the form $P(X \in A)$. Since

$$\begin{aligned} \{a < X \leq b\} &= \{X \leq b\} - \{X \leq a\}, \\ \{X > a\} &= S - \{X \leq a\}. \end{aligned}$$

It is sufficient to know the probability of $\{X \leq x\}$ for any real number x .

Definition 2.2.1 The distribution function or cumulative distribution function (abbreviated as CDF) $F(x)$ of a random variable X is the function

$$F(x) = P(X \leq x), -\infty < x < +\infty.$$

If X is thought as a random point on the real line, then $F(x)$ is the probability that X falls into the interval $(-\infty, x]$. The cumulative distribution function is defined as above for every random variable X , regardless of whether X is discrete, continuous or mixed.

Theorem 2.2.1 Cumulative distribution function $F(x)$ has the following basic properties:

(1) Nondecreasing. The function $F(x)$ is monotone nondecreasing function as x increases. That is, for any real numbers $x_1 < x_2$, $F(x_1) \leq F(x_2)$.

(2) Bounded. For any real number x , $0 \leq F(x) \leq 1$, and

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

(3) Continuity from the right. $F(x)$ is always continuous from the right, that is, for any real number x_0 ,

$$\lim_{x \rightarrow x_0^+} F(x) = F(x_0) \text{ or } F(x_0^+) = F(x_0).$$

Proof: Property (1) follows directly from Proposition 1.4.4. Since $F(x)$ is the probability that X falls into the interval $(-\infty, x]$, so $0 \leq F(x) \leq 1$. $F(x)$ is monotone nondecreasing, for any integers m and n , we have

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{m \rightarrow -\infty} F(m), \quad \lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow +\infty} F(n).$$

By countable additivity,

$$\begin{aligned} 1 &= P(-\infty < X < +\infty) \\ &= P\left(\bigcup_{i=-\infty}^{+\infty} \{i-1 < X \leq i\}\right) \\ &= \sum_{i=-\infty}^{+\infty} P(i-1 < X \leq i) \\ &= \lim_{n \rightarrow +\infty} \left[\lim_{m \rightarrow -\infty} \sum_{i=m}^n P(i-1 < X \leq i) \right] \\ &= \lim_{n \rightarrow +\infty} F(n) - \lim_{m \rightarrow -\infty} F(m), \end{aligned}$$

therefore, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.

Since $F(x)$ is monotone nondecreasing and bounded, for any real number x_0 , the right limit $F(x_0^+)$ exists. To prove the continuity from the right, we consider a decreasing sequence $x_1 > x_2 > \cdots > x_n > \cdots > x_0$,

$$\begin{aligned} F(x_1) - F(x_0) &= P(x_0 < X \leq x_1) \\ &= P\left[\bigcup_{i=1}^{+\infty} (x_{i+1} < X \leq x_i)\right] \\ &= \sum_{i=1}^{+\infty} P(x_{i+1} < X \leq x_i) \\ &= \sum_{i=1}^{+\infty} [F(x_i) - F(x_{i+1})] \\ &= \lim_{n \rightarrow +\infty} [F(x_1) - F(x_n)] \\ &= F(x_1) - \lim_{n \rightarrow +\infty} F(x_n). \end{aligned}$$

Therefore, $F(x_0) = \lim_{n \rightarrow +\infty} F(x_n) = F(x_0^+)$.

The above properties are important characteristics of cumulative distribution function. Further, it can be proved any function satisfying the three properties is a distribution function. The three properties are necessary and sufficient conditions to determine whether a function is a CDF.

It is convenient to express random events using CDF.

Proposition 2.2.1 For every real numbers $a < b$, the following formulas are valid:

$$\begin{aligned} P(X > a) &= 1 - F(a), \\ P(a < X \leq b) &= F(b) - F(a), \\ P(X < a) &= F(a^-), \\ P(X = a) &= F(a) - F(a^-). \end{aligned}$$

Example 2.2.1 X is a random variable for which the CDF is as follows:

$$F(x) = A + B \arctan x, \quad -\infty < x < +\infty.$$

Find (1) constants A and B , (2) $P\left(\frac{\sqrt{3}}{3} < X < 1\right)$.

Solution: By the properties of the CDF,

$$0 = \lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} (A + B \arctan x) = A - \frac{\pi}{2} B.$$

$$1 = \lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} (A + B \arctan x) = A + \frac{\pi}{2} B.$$

Therefore, $A = \frac{1}{2}$, $B = \frac{1}{\pi}$. The CDF is $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$.

$$\begin{aligned} P\left(\frac{\sqrt{3}}{3} < X < 1\right) &= F(1^-) - F\left(\frac{\sqrt{3}}{3}\right) \\ &= \lim_{x \rightarrow 1^-} \left(\frac{1}{2} + \frac{1}{\pi} \arctan x\right) - \left(\frac{1}{2} + \frac{1}{\pi} \arctan \frac{\sqrt{3}}{3}\right) \\ &= \frac{1}{12}. \end{aligned}$$

◆ 2.3 Discrete Distributions ◆

Definition 2.3.1 A random variable X has a discrete distribution or X is a discrete random variable if it can take only a finite number of different values $x_1, \dots, x_k$ or, at most, an infinite sequence of different values $x_1, x_2, \dots$

Definition 2.3.2 If a random variable X has a discrete distribution, the probability function (abbreviated as PF) of X is defined as the function $f(\cdot)$ such that for every real number x ,

$$f(x) = P(X=x) \geq 0.$$

It can also be represented as the following:

X	x_1	x_2	$\dots$	x_n	$\dots$
$f(x)$	p_1	p_2	$\dots$	p_n	$\dots$

The PF of X fully reflects all the values of X can take and their corresponding probabilities. The PF has the following propositions.

Proposition 2.3.1 Let X be a discrete random variable with the PF $f(x)$. The sequence $x_1, x_2, \dots$ includes all the possible values of X , then

$$(1) p_k = P(X=x_k) \geq 0, k=1, 2, \dots;$$

$$(2) \sum_{k=1}^{+\infty} p_k = 1.$$

Proposition 2.3.2 If X has a discrete distribution with the PF $f(x)$, its CDF can be determined from the relation

$$F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} P(X = x_i).$$

$F(x)$ will have a jump of magnitude $f(x_i)$ at each possible value x_i of X , and $F(x)$ will be constant between every pair of successive jumps. $F(x)$ and $f(x)$ can be determined from each other.

◆ 2.4 Some Common Discrete Distributions ◆

In the following, we will introduce some common discrete distributions.

Definition 2.4.1 A random variable X that takes only two values 0 and 1 with $P(X=1)=p$ has the **Bernoulli distribution** with parameter $0 < p < 1$.

$$P(X=1)=p, P(X=0)=1-p.$$

Bernoulli distribution is also called 0-1 distribution. We denote it as $X \sim B(1, p)$.

Example 2.4.1 Let X have the Bernoulli distribution with parameter p , find its CDF.

Solution: $P(X=1)=p, P(X=0)=1-p$.

(1) For $x < 0, F(x) = P(\emptyset) = 0$.

(2) For $0 \leq x < 1, F(x) = P(X \leq x) = P(X=0) = 1-p$.

(3) For $x \geq 1, F(x) = P(X=0 \text{ or } X=1) = P(X=0) + P(X=1) = 1$.

In summary, $F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$. See Figure 2.1.

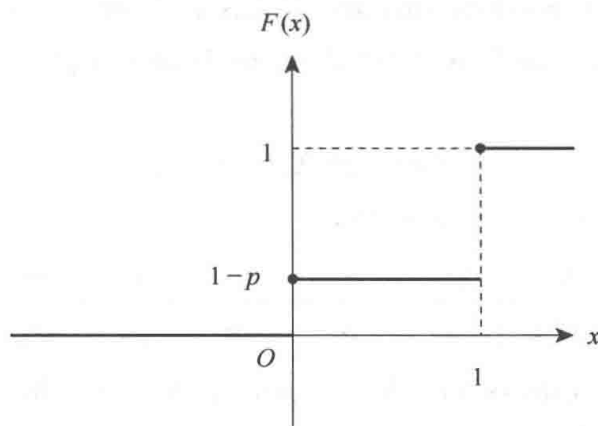


Figure 2.1

Definition 2.4.2 Consider an experiment that consists of observing n independent repetitions(trials) with only two possible results for each trial. For convenience, call the two possible results “success” and “failure”. Let p be the probability of success on each trial. The number X of successes will follow the **binomial distribution** with parameters n and p , denoted by $X \sim B(n, p)$ or $\text{Binomial}(n, p)$.

Proposition 2.4.1 Suppose $X \sim B(n, p)$. The PF of X will be as follows:

$$f(x) = P(X=x) = C_n^x p^x (1-p)^{n-x}, x=0, 1, \dots, n.$$

The binomial distribution is the sum of n independent Bernoulli random variables with same parameter p . $X = \sum_{i=1}^n X_i \sim B(n, p)$, where $X_i \sim B(1, p)$.

Note: not every sum of Bernoulli random variables has a binomial distribution. The Bernoulli random variables must be mutually independent, and they must all have the same parameter.

The PF of X can be understood as follows: we want exactly x successes (p^x) and $n-x$ failures $(1-p)^{n-x}$. Moreover, the x successes can occur anywhere among the n trials, and there are C_n^x different ways of distributing x successes in a sequence of n trials.

If $n=1$, the binomial distribution reduces to the Bernoulli distribution.

Theorem 2.4.1 Suppose $X \sim B(n, p)$, if $(n+1)p$ is an integer, the most likely number of X is $k_0 = (n+1)p$ or $(n+1)p-1$; otherwise the most likely number of X is $[(n+1)p]$, the greatest integer less than $(n+1)p$. The least likely number of X is 0 or n .

Proof: Consider the ratio of two consecutive probabilities, for $k=1, 2, \dots, n$,

$$\begin{aligned}\frac{f(k)}{f(k-1)} &= \frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} \\ &= \frac{(n-k+1)p}{k(1-p)} \\ &= 1 + \frac{(n+1)p - k}{k(1-p)}.\end{aligned}$$

So $f(k) \geq f(k-1)$ if and only if $k \leq (n+1)p$. $f(k)$ is increasing until it reaches a maximum value at the point when k reaches the largest integer less than or equal to $(n+1)p$, and then decreases. In the case that $(n+1)p$ is an integer, if $k_0 = (n+1)p$, we get

$$\frac{f(k_0)}{f(k_0-1)} = 1, \text{ both } (n+1)p \text{ and } (n+1)p-1 \text{ are maxima.}$$

Since $f(k)$ is increasing and then decreasing, the least likely number of X is either 0 or n . It is 0 if $p > 0.5$ and n if $p < 0.5$, if $p = 0.5$, then both are equally likely.

Example 2.4.2 If the probability that John will hit a target on any shot is 0.7. He shoots 10 times, find:

- (1) the probability that he hits the target exactly three times;
- (2) the probability that he hits the target at least three times;
- (3) the most likely times he hits the target.

Solution: Let X be the times that he hits the target, then $X \sim B(10, 0.7)$.

$$(1) P(X=3) = C_{10}^3 \times 0.7^3 \times 0.3^7 \approx 0.009.$$

(2) The probability that he hits the target at least three times is given by

$$\begin{aligned}P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - C_{10}^0 \times 0.7^0 \times 0.3^{10} - C_{10}^1 \times 0.7^1 \times 0.3^9 - C_{10}^2 \times 0.7^2 \times 0.3^8 \\ &\approx 0.9984.\end{aligned}$$

(3) Since $(n+1)p = 11 \times 0.7 = 7.7$, $k_0 = [7.7] = 7$. The most likely times he hits the target is 7. See Figure 2.2.

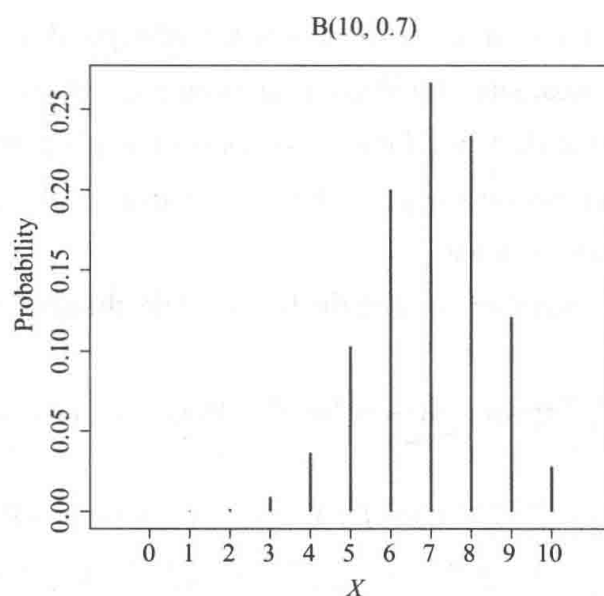


Figure 2.2

Definition 2.4.3 A random variable X has the **Poisson distribution** with parameter $\lambda > 0$, denoted as $X \sim P(\lambda)$ or $\text{Poisson}(\lambda)$, if the PF of X is as follows:

$$f(x) = P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}, x=0, 1, 2, \dots.$$

Many phenomena in the life and social sciences can be modeled using the Poisson distribution. It is useful in counting the number of occurrences in a given unit of measurement. Examples include the arrivals of customers entering a bank on a given day, the arrivals of calls at a switchboard in a day, the number of misprints per page of a book, and so forth. In addition, Poisson distribution is a limiting case of a binomial distribution when the number of trials gets very large and the probability of success is small.

Theorem 2.4.2 For each $n \geq 1$, let $X_n \sim B(n, p_n)$, If $\lim_{n \rightarrow +\infty} np_n = \lambda$, then X can be approximated by $X \sim P(\lambda)$. That is, for $k \geq 0$,

$$\lim_{n \rightarrow +\infty} P(X_n = k) = \lim_{n \rightarrow +\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Proof: Let $np_n = \lambda_n$, then

$$\begin{aligned} P(X_n = k) &= C_n^k p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{\lambda_n^k}{k!} \left(1-\frac{1}{n}\right) \cdots \left(1-\frac{k-1}{n}\right) \left(1-\frac{\lambda_n}{n}\right)^{n-k}. \end{aligned}$$

The theorem is then a consequence of the following limits:

$$\begin{aligned} \lim_{n \rightarrow +\infty} \lambda_n^k &= \lim_{n \rightarrow +\infty} (np_n)^k = \lambda^k, \\ \lim_{n \rightarrow +\infty} \left(1-\frac{1}{n}\right) \cdots \left(1-\frac{k-1}{n}\right) &= 1, \\ \lim_{n \rightarrow +\infty} \left(1-\frac{\lambda_n}{n}\right)^{n-k} &= \lim_{n \rightarrow +\infty} \left(1-\frac{\lambda_n}{n}\right)^{-\frac{n}{\lambda_n} \times \frac{n-k}{n} \times (-\lambda_n)} = e^{-\lambda}. \end{aligned}$$

The theorem states that for binomial distribution $X \sim B(n, p)$, if n is a large number and p is relatively small, X can be approximated by Poisson distribution with parameter $\lambda = np$.

Example 2.4.3 Suppose that in a large population the proportion of people who have a certain disease is 0.01. Determine the probability that in a random group of 200 people at least four people will have the disease.

Solution: Let X be the number of people having the disease among the 200 people. $X \sim B(200, 0.01)$.

This distribution can be approximated by the Poisson distribution with $\lambda = np = 2$. $X \sim P(2)$. See Figure 2.3.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) \\ &= 1 - \frac{e^{-2} \times 2^0}{0!} - \frac{e^{-2} \times 2^1}{1!} - \frac{e^{-2} \times 2^2}{2!} - \frac{e^{-2} \times 2^3}{3!} \\ &\approx 0.1428. \end{aligned}$$

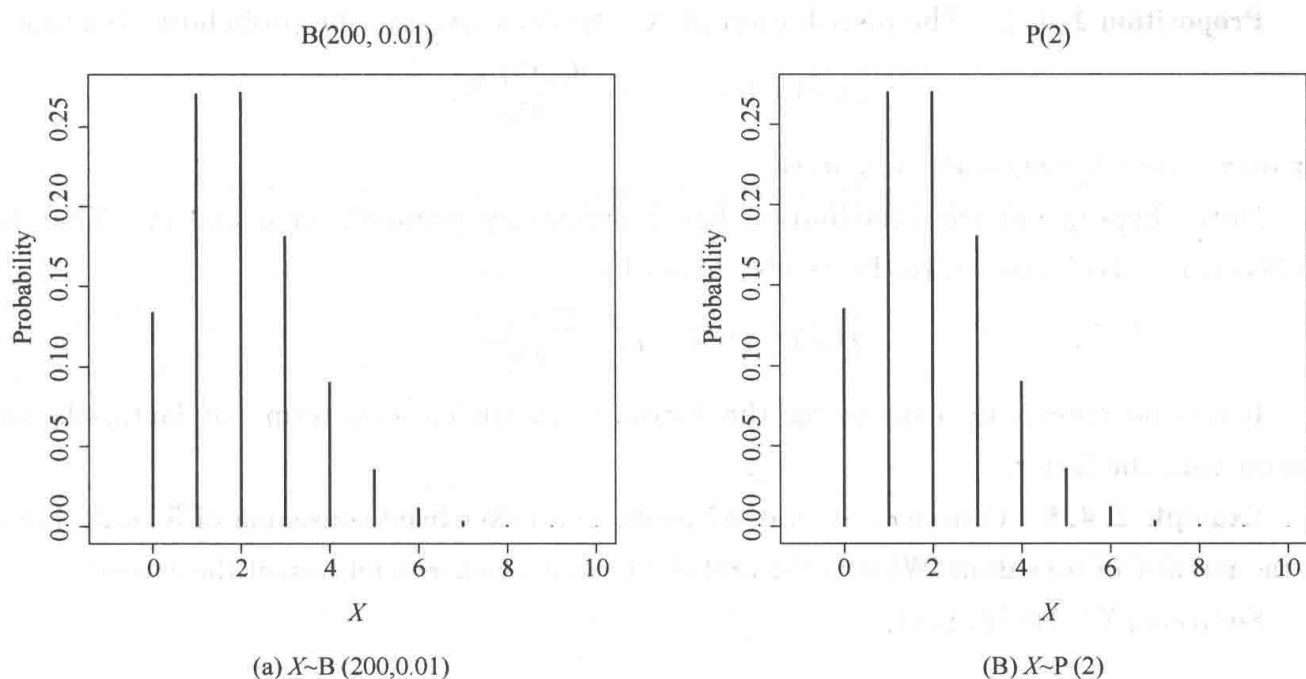


Figure 2.3

Example 2.4.4 There is a 0.998 probability that a randomly selected 30-year-old male lives through the year (based on data from the U. S. Department of Health and Human Services). A fidelity life insurance company charges \$1,200 for insuring that the male will live through the year. If the male does not survive the year, the policy pays out \$200,000 as a death benefit. Assume that the company sells 2500 such policies to 30-year-old males, so it collects \$3,000,000 in policy payments. Find the probability that the company makes a profit not less than \$1,000,000 from the 2500 policies.

Solution: Let X be the number of deaths among the 30-year-old males. $X \sim B(2500, 0.002)$. This distribution can be approximated by the Poisson distribution with $\lambda = 2500 \times 0.002 = 5$.

$$P(\text{the company makes a profit not less than } \$1,000,000)$$

$$= P(3000000 - 200000X \geq 1000000)$$

$$= P(X \leq 10)$$

$$\approx \sum_{k=0}^{10} \frac{5^k}{k!} e^{-5}$$

$$\approx 0.9863.$$

Definition 2.4.4 Suppose that the finite population to be sampled has N elements. For each trial of the experiment, there are two possible outcomes, success or failure. There are exactly m successes in the population. A sample of size n is obtained from the population of size N without replacement. Let X denote the number of successes that are obtained, then X follows the **Hypergeometric distribution**, denoted as $X \sim H(N, m, n)$.

Let $X_i = 1$ if the i -th draw is success and $X_i = 0$ if not. Then each X_i has a Bernoulli distribution, but $X_1, \dots, X_n$ are not independent in general. $X = \sum_{i=1}^n X_i$ is not binomial.

Proposition 2.4.2 The distribution of $X \sim H(N, m, n)$ has the probability function

$$f(x) = P(X=x) = \frac{C_m^x C_{N-m}^{n-x}}{C_N^n},$$

for $\max\{0, n-N+m\} \leq x \leq \min\{n, m\}$.

Note: hypergeometric distribution has a symmetry property in n and m . That is, $H(N, m, n) = H(N, n, m)$. Its PF is also given by

$$f(x) = P(X=x) = \frac{C_n^x C_{N-n}^{m-x}}{C_N^m}.$$

It can be shown by expressing the binomial coefficients in terms of factorials and rearranging the latter.

Example 2.4.5 Consider a deck of 52 cards. In a poker hand consisting of 5 cards, Let X be the number of aces dealt. What is the probability that a poker hand has all the 4 aces?

Solution: $X \sim H(52, 4, 5)$.

$$P(X=4) = \frac{C_4^4 C_{48}^1}{C_{52}^5} = 0.00001846.$$

Or, we can use $X \sim H(52, 5, 4)$.

$$P(X=4) = \frac{C_5^4 C_{47}^0}{C_{52}^4} = 0.00001846.$$

Definition 2.4.5 Suppose that an infinite sequence of Bernoulli trials with probability of success p are available. Let X denote the total number of trials required to produce r successes, then X follows the **negative binomial distribution**, denoted as $X \sim NB(r, p)$.

Remark 2.4.1 There are a couple variations of the negative binomial distribution. If someone talks about a negative binomial random variable, find out how they are defining it. For example, some text books define X as the failures that occur before the r -th success.

Proposition 2.4.3 The distribution of $X \sim NB(r, p)$ has the PF

$$f(x) = P(X=x) = C_{x-1}^{r-1} p^r (1-p)^{x-r},$$

for $x = r, r+1, \dots$.

Particularly, if $r=1$, $X \sim NB(1, p)$, $f(x) = p(1-p)^{x-1}$, $x = 1, 2, \dots$. It deals with the number of trials required for a single success. In this case, we say that $X \sim \text{Geo}(p)$ has the **geometric distribution** with parameter p .

Example 2.4.6 Tony is a high school basketball player. His probability of making a free throw is 0.80. During the season, (1) what is the probability that Tony makes his third free throw on his fifth shot? (2) what is the probability that Tony makes his first free throw on his fifth shot?

Solution: (1) This is an example of a negative binomial random variable. Let X be the number of shots needed to make the third free throw.

$$X \sim NB(3, 0.8),$$

$$P(X=5) = C_4^2 \times 0.8^3 \times (1-0.8)^2 = 0.1229.$$

(2) This is an example of a geometric distribution, which is a special case of a negative binomial distribution. Let Y be the number of shots needed to make the first free throw.



$$Y \sim \text{Geo}(0.8),$$

$$P(Y=5) = 0.8 \times (1-0.8)^4 = 0.00128.$$

2.5 Continuous Distributions

Definition 2.5.1 X is a continuous random variable if there exists a nonnegative function $f(x)$, defined on the real line, such that for every real number x , the CDF of X can be expressed by

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt.$$

The function $f(x)$ is called the probability density function (abbreviated as PDF) of X .

Every PDF $f(x)$ has the following two properties:

- (1) $f(x) \geq 0$;
- (2) $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Continuous random variable X can be uniquely determined by its PDF $f(x)$. $f(x)$ is not necessarily to be continuous while $F(x)$ must be continuous. There are at most countable discontinuity points of $f(x)$. If $f(x)$ is continuous at x , then,

$$F'(x) = f(x).$$

If X has a continuous distribution, for any real value a , $P(X=a) = F(a) - F(a^-) = 0$. Continuous distributions assign probability 0 to individual values. That is, probability 0 does not mean impossible.

See Figure 2.4. If $a \leq b$, then

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b).$$

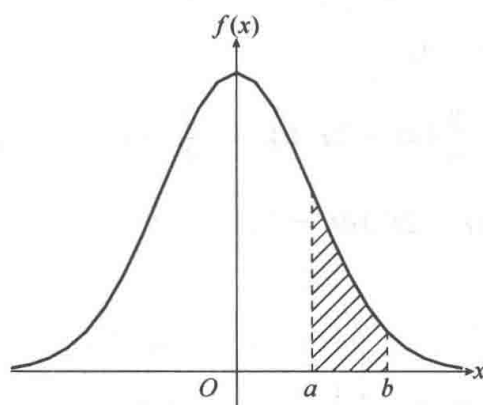


Figure 2.4

For any interval G ,

$$P(X \in G) = \int_G f(x) dx.$$

Example 2.5.1 Suppose that the PDF of a certain random variable X has the following form:

$$f(x) = \begin{cases} c(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

(1) Determine the value of c ; (2) Find $P(X > 1)$; (3) Find the CDF of X .

Solution: (1) For every PDF $f(x)$, it must be true that

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} f(x) dx \\ &= \int_0^2 c(4x - 2x^2) dx \\ &= c \left(2x^2 - \frac{2}{3}x^3 \right) \Big|_0^2 \\ &= \frac{8}{3}c. \end{aligned}$$

Hence $c = \frac{3}{8}$. The PDF is

$$f(x) = \begin{cases} \frac{3}{8}(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} (2) \quad P(X > 1) &= \int_1^{+\infty} f(x) dx \\ &= \int_1^2 f(x) dx + \int_2^{+\infty} f(x) dx \\ &= \int_1^2 \frac{3}{8}(4x - 2x^2) dx \\ &= \frac{3}{8} \left(2x^2 - \frac{2}{3}x^3 \right) \Big|_1^2 \\ &= \frac{1}{2}. \end{aligned}$$

(3) The CDF of X is $F(x) = \int_{-\infty}^x f(t) dt$.

$$\text{If } x \leq 0, F(x) = \int_{-\infty}^x 0 dt = 0.$$

$$\text{If } 0 < x < 2, F(x) = \int_0^x \frac{3}{8}(4t - 2t^2) dt = \frac{1}{4}(3x^2 - x^3).$$

$$\text{If } x \geq 2, F(x) = \int_0^2 \frac{3}{8}(4t - 2t^2) dt = 1.$$

The CDF $F(x)$ is

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{4}(3x^2 - x^3), & 0 < x < 2 \\ 1, & x \geq 2 \end{cases}$$

Example 2.5.2 The lifetime X of a certain electrical component is a continuous random variable with the following PDF:

$$f(x) = \begin{cases} 0, & x \leq 100 \\ \frac{100}{x^2}, & x > 100 \end{cases}$$

One system has 5 identical components and the components operate independently. Find the probability that exactly 2 components fail in less than 150 hours.

Solution: Let A be the event that a component fails in less than 150 hours, then

$$p = P(A) = P(X \leq 150) = \int_{-\infty}^{150} f(x) dx = \int_{100}^{150} \frac{100}{x^2} dx = \frac{1}{3}.$$

Let Y count the number of failing components in less than 150 hours, then Y is a binomial random variable. $Y \sim B\left(5, \frac{1}{3}\right)$. The probability that exactly 2 components fail in less than 150 hours is given as

$$P(Y=2) = C_5^2 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^3 = 0.3292.$$

Definition 2.5.2 When the CDF of a random variable X is continuous and one-to-one over the whole set of possible values of X , the inverse $F^{-1}(p)$ of $F(x)$ exists and equals the quantile function of X .

$$P[X \leq F^{-1}(p)] = p.$$

◆ 2.6 Some Useful Continuous Distributions ◆

Definition 2.6.1 Let a and b be two given real numbers such that $a < b$. Let X be a random variable such that it is known that $a \leq X \leq b$ and, for every subinterval of $[a, b]$, the probability that X will belong to that subinterval is proportional to the length of that subinterval. We then say that the random variable X has the **uniform distribution** on the interval $[a, b]$, denoted by $X \sim U[a, b]$. Its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}.$$

The CDF of $X \sim U[a, b]$ is

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}.$$

See Figure 2.5.

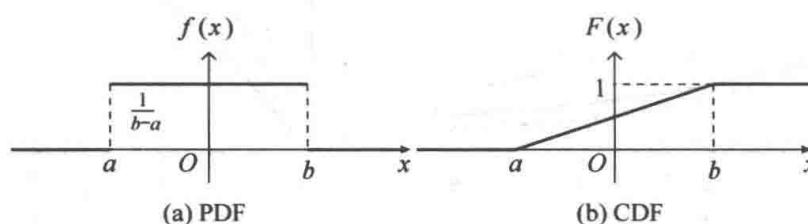


Figure 2.5

Example 2.6.1 Consider the equation $4x^2 + 4\xi x + (\xi + 2) = 0$, where ξ follows the uniform distribution on the interval $[-3, 6]$. Calculate the probability that the equation has real roots.

Solution: The PDF of ξ is

$$f(x) = \begin{cases} \frac{1}{9}, & -3 \leq x \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

Let A be the event that the equation has real roots, then

$$\begin{aligned} P(A) &= P[(4\xi)^2 - 4 \times 4 \times (\xi + 2) \geq 0] \\ &= P[(\xi + 1)(\xi - 2) \geq 0] \\ &= P(\xi \leq -1 \text{ or } \xi \geq 2) \\ &= P(\xi \leq -1) + P(\xi \geq 2) \\ &= \int_{-3}^{-1} \frac{1}{9} dx + \int_2^6 \frac{1}{9} dx \\ &= \frac{2}{3}. \end{aligned}$$

Definition 2.6.2 Continuous random variable X is said to follow the **exponential distribution** if its PDF is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

where $\lambda > 0$ is parameter. It is denoted by $X \sim E(\lambda)$.

The CDF of $X \sim E(\lambda)$ is

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

See Figure 2.6. The exponential distribution is widely used in reliability applications. The exponential distribution is used to model lifetime data. Light bulbs often have an exponential time to failure. The time between customers coming into a store or bank is often exponential.

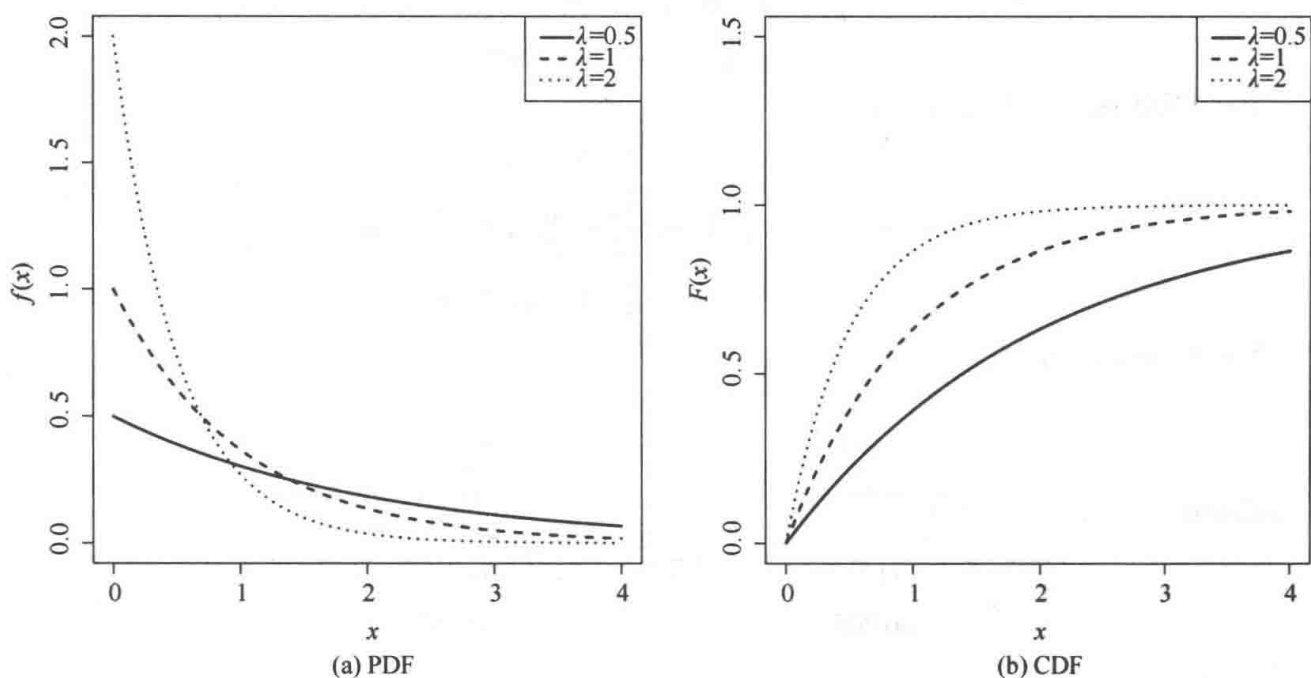


Figure 2.6

Proposition 2.6.1 Suppose $X \sim E(\lambda)$, then for any $s, t > 0$,

$$P(X > s+t | X > s) = P(X > t).$$

Proof:

$$P(X > s+t | X > s) = \frac{P(X > s+t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

This is known as the memoryless property. Let X be the waiting time, the probability that you'll wait more than $s+t$ minutes given that you've already been waiting s minutes is equal to the probability that you'll wait t minutes. This means that the probability that you'll wait t more minutes given that you've already been waiting s minutes is the same as the probability that you'd wait for more than t minutes from the beginning.

Example 2.6.2 Suppose that the amount of time one spends in a bank is exponentially distributed with parameter $\lambda = \frac{1}{10}$.

(1) What is the probability that a customer will spend more than 15 minutes in the bank?

(2) What is the probability that a customer will spend more than 15 minutes in the bank given that he is still in the bank after 10 minutes?

Solution:

$$(1) \quad P(X > 15) = e^{-15\lambda} = e^{-\frac{3}{2}} = 0.2231.$$

$$(2) \quad P(X > 15 | X > 10) = P(X > 5) = e^{-5\lambda} = e^{-\frac{1}{2}} = 0.6065.$$

Definition 2.6.3 A random variable X follows the **normal distribution** if the PDF of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

where $-\infty < \mu < +\infty$ and $\sigma > 0$ are two parameters. We write $X \sim N(\mu, \sigma^2)$.

Proposition 2.6.2 The area under the normal curve is 1.

Proof: $f(x) \geq 0$, we need to verify that $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. Let $y = \frac{x-\mu}{\sigma}$, then

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

It suffices to prove

$$I = \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}.$$

Notice that

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-\frac{r^2}{2}} r dr d\theta \quad (\text{set } x = r \cos \theta, y = r \sin \theta) \\ &= 2\pi. \end{aligned}$$

Therefore, $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$

Proposition 2.6.3 The normal PDF $f(x)$ have the following features:

(1) the curve is bell shaped and symmetric about the vertical line $x=\mu$, that is, for $h>0$, $P(\mu-h < X \leq \mu) = P(\mu < X \leq \mu+h).$

(2) If $x=\mu$, $f(x)$ achieves its maximum $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}.$

(3) The transition points are where the graph changes from cupping upward to cupping downward(or visa versa). The transition points occur at $x=\mu \pm \sigma.$

(4) x axis is an asymptote of the curve. The curve approaches arbitrarily close to the axis, but never reaches it.

(5) Changes in μ , without changes in σ , result in moving the curve to the right or left, depending upon whether the new value of μ was larger or smaller than the previous value, but does not change the shape of the curve. σ specifies the shape of the curve. See Figure 2.7.

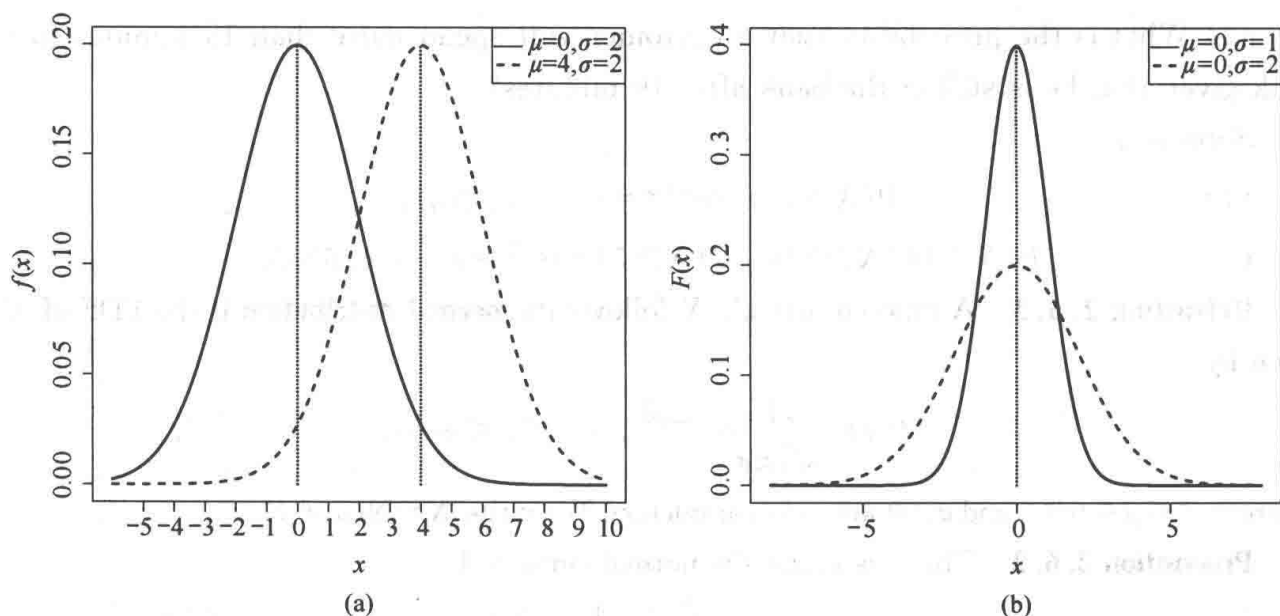


Figure 2.7

The normal distribution is the most important distribution. It describes well the distribution of random variables that arise in the natural and social sciences. Also, it is important for the central limit theorem, the approximation of other distributions such as the binomial, etc.

Specially, if $\mu=0, \sigma=1, X \sim N(0, 1)$ is called the standard normal distribution. The PDF of X is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < +\infty.$$

See Figure 2.8, its CDF is given by

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

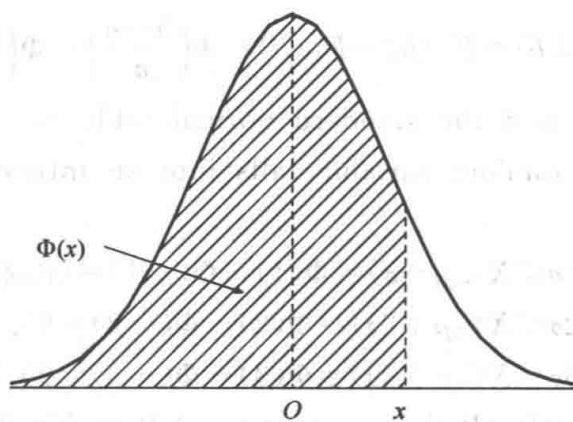


Figure 2.8

By symmetry, for any x ,

$$\Phi(-x) = 1 - \Phi(x)$$

For $x \geq 0$, the value of $\Phi(x)$ is given in the standard normal table in Appendix C. For example, $\Phi(1.02) = 0.8461$, $\Phi(-1.02) = 1 - \Phi(1.02) = 1 - 0.8461 = 0.1539$.

Lemma 2.6.1 If $X \sim N(\mu, \sigma^2)$, then

(1) $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$, where a and b are constants ($a \neq 0$);

(2) $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$.

Proof: (1) If $a > 0$, the CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

Take the derivative with respect to y ,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{a}.$$

If $a < 0$, the CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right).$$

Take the derivative with respect to y ,

$$f_Y(y) = -f_X\left(\frac{y-b}{a}\right) \frac{1}{a}.$$

Therefore,

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{y-b}{a} - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{[y - (a\mu + b)]^2}{2a^2\sigma^2}}.$$

Thus, $Y \sim N(a\mu + b, a^2\sigma^2)$.

(2) Set $a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma}$ in (1), the conclusion follows.

A value from any normal distribution can be transformed into its corresponding value on the standard normal distribution. Let $X \sim N(\mu, \sigma^2)$, then

$$F_X(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

For two constants $a < b$,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Using the formulas above and the standard normal table, we can easily determine the probability that a normal random variable falls into an interval. For instance, suppose $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} P(\mu - \sigma < X \leq \mu + \sigma) &= \Phi(1) - \Phi(-1) = 68.26\%, \\ P(\mu - 2\sigma < X \leq \mu + 2\sigma) &= \Phi(2) - \Phi(-2) = 95.44\%, \\ P(\mu - 3\sigma < X \leq \mu + 3\sigma) &= \Phi(3) - \Phi(-3) = 99.74\%. \end{aligned}$$

It indicates that approximately all the observations fall within 3σ of μ . It is called “three sigma rule”.

Example 2.6.3 Suppose $X \sim N(2, 9)$, find $P(1 \leq X < 5)$.

Solution:

$$\begin{aligned} P(1 \leq X < 5) &= P\left(\frac{1-2}{3} \leq \frac{X-2}{3} < \frac{5-2}{3}\right) \\ &= \Phi\left(\frac{5-2}{3}\right) - \Phi\left(\frac{1-2}{3}\right) \\ &= \Phi(1) - \Phi\left(-\frac{1}{3}\right) \\ &= \Phi(1) + \Phi\left(\frac{1}{3}\right) - 1 \\ &= 0.8413 + 0.6293 - 1 \\ &= 0.4706. \end{aligned}$$

Example 2.6.4 Scores on a standardized college entrance examination (CEE) are normally distributed with mean 510 and standard deviation 60. A selective university considers for admission only applicants with CEE scores over 650. Find percentage of all individuals who took the CEE meet the university’s CEE requirement for consideration for admission.

Solution: Let X denote the score made on the CEE by a randomly selected individual. Then $X \sim N(510, 60)$. The probability that X lie in a particular interval is the same as the proportion of all exam scores that lie in that interval. Thus the solution to the problem is $P(X > 650)$, expressed as a percentage.

$$\begin{aligned} P(X > 650) &= P\left(\frac{X-510}{60} > \frac{650-510}{60}\right) \\ &= P(Z > 2.33) \\ &= 1 - \Phi(2.33) \\ &= 1 - 0.9901 \\ &= 0.0099. \end{aligned}$$

The proportion of all CEE scores that exceed 650 is 0.0099, hence 0.99% or about 1% meet the requirement.

Suppose X is a random variable, $F(x)$ is its CDF. For any real x , $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$. Sometimes, we need consider the inverse problem. That is, if $P(X >$

$x) = \alpha$ is given, determine the value of x .

Definition 2.6.4 Suppose the CDF of X is $F(x)$, $0 < \alpha < 1$, then x_α is called upper α quantile or α quantile if

$$P(X > x_\alpha) = \alpha.$$

Note: some textbooks define the lower quantile using the left tail.

For the standard normal distribution $Z \sim N(0, 1)$, the α quantile is denoted by z_α .

$$P(Z > z_\alpha) = \int_{z_\alpha}^{+\infty} \varphi(x) dx = \alpha,$$

$$\Phi(z_\alpha) = 1 - \alpha.$$

$z_\alpha = \Phi^{-1}(1 - \alpha)$ can be found in the standard normal table. By symmetry, $z_{1-\alpha} = -z_\alpha$. For example, $z_{0.05} = 1.645$, $z_{0.95} = -z_{0.05} = -1.645$.

Definition 2.6.5 X follows the **gamma distribution** $X \sim \text{Gamma}(\alpha, \beta)$ if its PDF is given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

where $\alpha > 0, \beta > 0$ are two parameters, see Figure 2.9. $\Gamma(\alpha)$ is a gamma function defined by

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

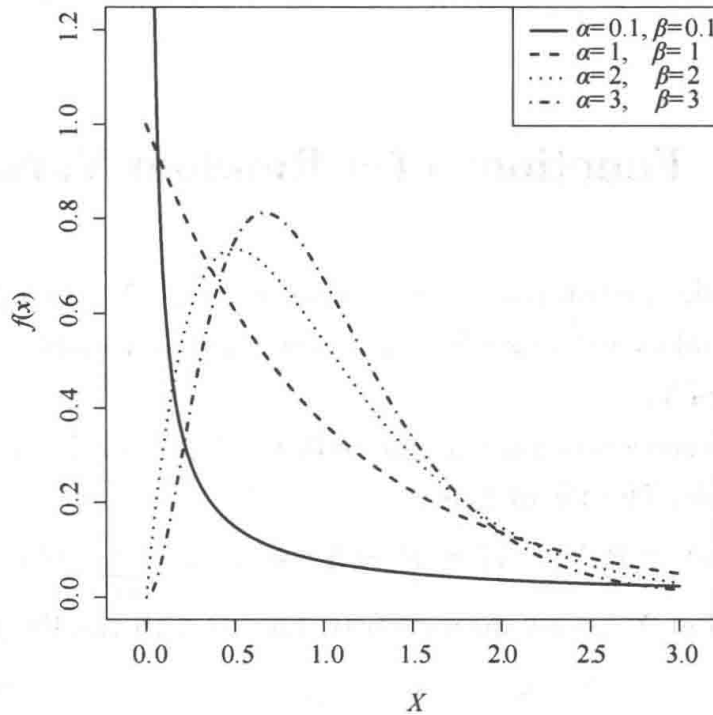


Figure 2.9

The gamma function satisfies:

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha),$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \Gamma(1) = 1.$$

If n is an integer, $\Gamma(n) = (n-1)!$.

Remark 2.6.1 If $\alpha=1$, the PDF of $X \sim \text{Gamma}(\alpha, \beta)$ reduces to

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

It is identical to the PDF of exponential distribution $X \sim E(\frac{1}{\beta})$. That means that exponential distribution is a special case of gamma distribution.

If $\alpha=n$, the PDF of $X \sim \text{Gamma}(\alpha, \beta)$ reduces to

$$f(x) = \begin{cases} \frac{1}{\beta^n (n-1)!} x^{n-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

It is called Erlang distribution, an important distribution widely used in queueing systems and stochastic processes.

If $\alpha = \frac{n}{2}, \beta = 2$, the PDF of $X \sim \text{Gamma}(\alpha, \beta)$ reduces to

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

It is the χ^2 -distribution with degrees of freedom n , an important distribution in mathematical statistics.

◆ 2.7 Functions of a Random Variable ◆

We have studied the distribution of a random variable X . Consider the function $Y = r(X)$, if X takes x , Y takes $y = r(x)$. Y is another random variable. In practice, often we want the distribution of Y .

Suppose X is a discrete random variable with the PF $f(x)$, $Y = r(X)$, then Y is also a discrete random variable. The PF of Y is

$$g(y) = P(Y = y) = P[r(X) = y] = \sum_{x: r(x)=y} f(x).$$

Example 2.7.1 Let X have a discrete distribution with the PF $f(x)$:

X	1	2	3	4	5	6	7	8	9
$f(x)$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Find the PF $g(y)$ of $Y = |X - 5|$.

Solution: The possible values of Y are 0, 1, 2, 3 and 4.

$$g(0) = P(Y=0) = P(|X-5|=0) = P(X=5) = \frac{1}{9},$$

$$g(1) = P(Y=1) = P(|X-5|=1) = P(X=4 \text{ or } X=6) = \frac{2}{9},$$

$$g(2) = P(Y=2) = P(|X-5|=2) = P(X=3 \text{ or } X=7) = \frac{2}{9},$$

$$g(3) = P(Y=3) = P(|X-5|=3) = P(X=2 \text{ or } X=8) = \frac{2}{9},$$

$$g(4) = P(Y=4) = P(|X-5|=4) = P(X=1 \text{ or } X=9) = \frac{2}{9}.$$

Thus, the PF $g(y)$ of $Y=|X-5|$ is given as follows:

Y	0	1	2	3	4
$g(y)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

Example 2.7.2 Suppose the PF of X is

$$P\left(X=k \frac{\pi}{2}\right) = pq^k; k=0, 1, 2, \dots; p+q=1; 0 < p < 1.$$

Find the PF $g(y)$ of $Y=\sin X$.

Solution: Y can take $0, 1, -1$.

$$\begin{aligned} P(Y=0) &= P\left(\bigcup_{m=0}^{+\infty} \left\{X = 2m \cdot \frac{\pi}{2}\right\}\right) \\ &= \sum_{m=0}^{+\infty} pq^{2m} \\ &= \frac{p}{1-q^2}, \end{aligned}$$

$$\begin{aligned} P(Y=1) &= P\left(\bigcup_{m=0}^{+\infty} \left\{X = 2m\pi + \frac{\pi}{2}\right\}\right) \\ &= P\left(\bigcup_{m=0}^{+\infty} \left\{X = (4m+1) \cdot \frac{\pi}{2}\right\}\right) \\ &= \sum_{m=0}^{+\infty} pq^{4m+1} \\ &= \frac{pq}{1-q^4}, \end{aligned}$$

$$\begin{aligned} P(Y=-1) &= P\left(\bigcup_{m=0}^{+\infty} \left\{X = 2m\pi + \frac{3\pi}{2}\right\}\right) \\ &= P\left(\bigcup_{m=0}^{+\infty} \left\{X = (4m+3) \cdot \frac{\pi}{2}\right\}\right) \\ &= \sum_{m=0}^{+\infty} pq^{4m+3} \\ &= \frac{pq^3}{1-q^4}. \end{aligned}$$

Thus, the PF $g(y)$ of $Y=\sin X$ is given as follows:

Y	-1	0	1
$g(y)$	$\frac{pq^3}{1-q^4}$	$\frac{p}{1-q^2}$	$\frac{pq}{1-q^4}$

Suppose X is a continuous random variable with the PDF $f_X(x)$, $Y=r(X)$, suppose Y is also a continuous random variable, we want to determine the PDF of Y .

One way to proceed is by direct calculation. The CDF $G(y)$ of Y can be derived as

$$G(y) = P(Y \leq y) = P(r(X) \leq y) = \int_{x: r(x) \leq y} f(x) dx.$$

Its PDF $g(y)$ can be obtained from the relation

$$g(y) = \frac{dG(y)}{dy} = G'(y).$$

This relation is satisfied at every point y at which $G(y)$ is differentiable.

Example 2.7.3 Suppose that X has the uniform distribution on the interval $[-1, 1]$,

$$f(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the PDF of the random variable $Y=X^2$.

Solution: Since $Y=X^2$, $-1 \leq X \leq 1$, then Y must belong to the interval $0 \leq Y \leq 1$.

(1) If $y < 0$, $G(y) = P(Y \leq y) = P(\emptyset) = 0$.

(2) If $y > 1$, $G(y) = P(Y \leq y) = P(\Omega) = 1$.

(3) If $0 \leq y \leq 1$,

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(x) dx \\ &= \sqrt{y}. \end{aligned}$$

It follows that the PDF of Y is

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Example 2.7.4 Suppose that X is a continuous random variable with the PDF

$$f(x) = \begin{cases} \frac{2x}{\pi^2}, & 0 < x < \pi \\ 0, & \text{otherwise} \end{cases}.$$

Determine the PDF of $Y = \sin X$.

Solution: The CDF of Y is $G(y) = P(Y \leq y) = P(\sin X \leq y)$. See Figure 2.10.

(1) If $y \leq 0$, $G(y) = 0$.

(2) If $0 < y < 1$,

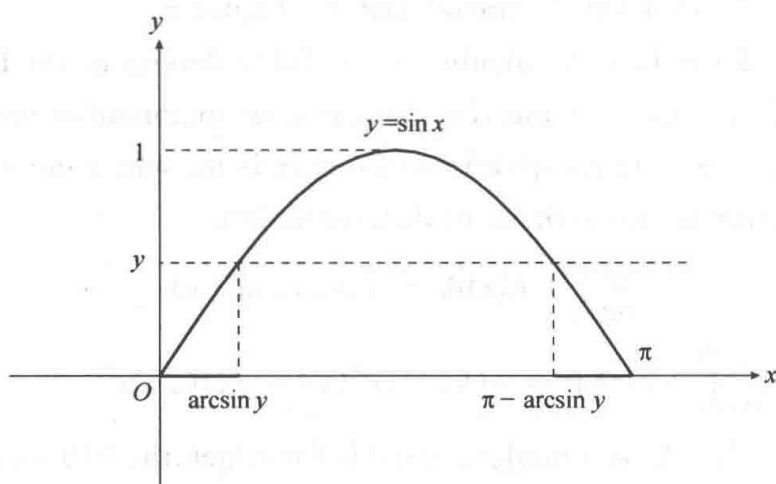


Figure 2.10

$$\begin{aligned}
 G(y) &= P(\sin X \leq y) \\
 &= P(0 < X < \arcsin y) + P(\pi - \arcsin y < X < \pi) \\
 &= \int_0^{\arcsin y} \frac{2x}{\pi^2} dx + \int_{\pi - \arcsin y}^{\pi} \frac{2x}{\pi^2} dx \\
 &= \frac{2\arcsin y}{\pi}.
 \end{aligned}$$

(3) If $1 \leq y$, $G(y) = 1$.

The PDF of Y is $g(y) = G'(y) = \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$.

Example 2.7.5 Suppose that $X \sim N(0, 1)$, its PDF is given as follows

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Determine the PDF of the random variable $Y = X^2$.

Solution: Since $Y = X^2$, $-\infty < X < +\infty$, then $Y \geq 0$.

(1) If $y < 0$, $G(y) = P(Y \leq y) = P(\emptyset) = 0$.

(2) If $y \geq 0$,

$$\begin{aligned}
 G(y) &= P(Y \leq y) \\
 &= P(X^2 \leq y) \\
 &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
 &= \int_{-\sqrt{y}}^{\sqrt{y}} \varphi(x) dx.
 \end{aligned}$$

It follows that the PDF $g(y)$ of Y is

$$g(y) = G'(y) = \frac{1}{2\sqrt{y}} [\varphi(\sqrt{y}) + \varphi(-\sqrt{y})] = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}.$$

Thus, the PDF $g(y)$ of Y is

$$g(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

It shows that $Y \sim \chi^2(1)$ that will be introduced in Chapter 6.

Remark 2.7.1 Some facts in calculus are useful to determine the PDF $g(y)$ when the CDF $G(y)$ has no closed form. Using the chain rule in combination with the fundamental theorem of calculus we may find derivatives of integrals for which one or the other limit of integration is a function of the variable of differentiation.

$$\frac{d}{dx} \int_a^{h(x)} f(t) dt = f(h(x))h'(x).$$

$$\frac{d}{dx} \int_{s(x)}^{h(x)} f(t) dt = f(h(x))h'(x) - f(s(x))s'(x).$$

Theorem 2.7.1 Let X be a random variable for which the PDF is $f(x)$ and for which $P(a < X < b) = 1$. Let $Y = r(X)$, and suppose that $r(x)$ is differentiable and one-to-one (strictly monotonic) for $a < x < b$. Let (α, β) be the image of the interval (a, b) under the function r . Let $s(y)$ be the inverse function of $r(x)$, then the PDF $g(y)$ of Y is

$$g(y) = \begin{cases} f(s(y)) \cdot \left| \frac{ds(y)}{dy} \right|, & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases}.$$

Proof: Since $r(x)$ is differentiable and one-to-one for $a < x < b$, we have $r'(x) > 0$ or $r'(x) < 0$. If $r'(x) > 0$, $r(x)$ is increasing and its inverse function $s(y)$ is also increasing.

If $y \leq \alpha$, $G(y) = P(Y \leq y) = P(\emptyset) = 0$. If $y \geq \beta$, $G(y) = P(Y \leq y) = P(\Omega) = 1$.

If $\alpha < y < \beta$,

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P[r(X) \leq y] \\ &= P[X \leq s(y)] \\ &= F(s(y)). \end{aligned}$$

Take the derivative with respect to y ,

$$g(y) = \begin{cases} f(s(y)) \cdot \frac{ds(y)}{dy}, & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases}.$$

Similarly, if $r'(x) < 0$, $s'(y) < 0$, we can obtain that

$$g(y) = \begin{cases} f(s(y)) \cdot \left[-\frac{ds(y)}{dy} \right], & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, the PDF of Y is

$$g(y) = \begin{cases} f(s(y)) \cdot \left| \frac{ds(y)}{dy} \right|, & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases}.$$

Example 2.7.6 Suppose that X has the PDF

$$f(x) = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the PDF of the random variable $Y = X^2$.

Solution: $y=r(x)=x^2$ is one-to-one for $0<x<1$ and has its inverse function $s(y)=y^{\frac{1}{2}}, 0<y<1$.

$$g(y)=\begin{cases} 3(y^{\frac{1}{2}})^2 \cdot \left| \frac{1}{2}y^{-\frac{1}{2}} \right| = \frac{3}{2}y^{\frac{1}{2}}, & 0<y<1 \\ 0, & \text{otherwise} \end{cases}$$

Example 2.7.7 Suppose $X \sim N(\mu, \sigma^2)$, determine the PDF of $Y=e^X$.

Solution: The PDF of X is

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<+\infty.$$

$y=r(x)=e^x$ is one-to-one for $-\infty<x<+\infty$ and has its inverse function $s(y)=\ln y, y>0$. Therefore,

$$g(y)=\begin{cases} f(\ln y) \cdot \left| \frac{ds(y)}{dy} \right|, & y>0 \\ 0, & \text{otherwise} \end{cases}$$

$$=\begin{cases} \frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, & y>0 \\ 0, & \text{otherwise} \end{cases}$$

This distribution is called lognormal distribution.

Theorem 2.7.2 Let X be a random variable for which the PDF is $f(x)$ and for which $P(a<X<b)=1$. Let $Y=r(X)$, Partition the range of X into n subintervals $I_1, I_2, \dots, I_n$ such that $r(x)=r_i(x) (i=1, 2, \dots, n)$ is differentiable and one-to-one in each subinterval. Let (α, β) be the image of the interval (a, b) under the function r . Let $s_i(y)$ be the inverse function of $r_i(x) (i=1, 2, \dots, n)$, then the PDF $g(y)$ of Y is

$$g(y)=\begin{cases} \sum_{i=1}^n f(s_i(y)) \cdot \left| \frac{ds_i(y)}{dy} \right|, & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases}$$

Example 2.7.8 Suppose that X is a continuous random variable with the PDF

$$f(x)=\begin{cases} \frac{2x}{\pi^2}, & 0<x<\pi \\ 0, & \text{otherwise} \end{cases}$$

Determine the PDF of $Y=\sin X$.

Solution: We have done this problem by direct calculation. Now we present another method.

$y=\sin x$ is strictly monotonic increasing and differentiable in $(0, \frac{\pi}{2}]$ and has its inverse function $s_1(y)=\arcsin y$. $y=\sin x$ is strictly monotonic decreasing and differentiable in $(\frac{\pi}{2}, \pi)$ and has its inverse function $s_2(y)=\pi - \arcsin y$. If $0<x<\pi$, then $y=\sin x \in (0, 1)$.

Therefore, if $0<y<1$,

$$\begin{aligned}
 g(y) &= f(s_1(y)) \cdot |s_1'(y)| + f(s_2(y)) \cdot |s_2'(y)| \\
 &= \frac{2}{\pi^2} \arcsin y \cdot \frac{1}{\sqrt{1-y^2}} + \frac{2}{\pi^2} (\pi - \arcsin y) \cdot \frac{1}{\sqrt{1-y^2}} \\
 &= \frac{2}{\pi \sqrt{1-y^2}}.
 \end{aligned}$$

The PDF of Y is $g(y) = \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$

◆ Exercise 2 ◆

2.1 An urn contains 4 balls marked with numbers 0, 1, 1, 2, respectively. Draw 2 balls with replacement. Let X be the product of the numbers of the balls selected. Find the PF of X .

2.2 Suppose that a random variable X has a discrete distribution with the following PF:

$$f(x) = \begin{cases} cx, & x=1, 2, \dots, 5 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the value of the constant c .

2.3 Suppose that a random variable X has a discrete distribution with the following PF:

$$f(x) = \begin{cases} \frac{c}{2^x}, & x=0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}.$$

Find the value of the constant c .

2.4 Suppose that two balanced dice are rolled, and let X denote the absolute value of the difference between the two numbers that appear. Determine and sketch the PF of X .

2.5 A civil engineer is studying a left-turn lane that is long enough to hold seven cars. Let X be the number of cars in the lane at the end of a randomly chosen red light. The engineer believes that the probability that $X=x$ is proportional to $(x+1)(8-x)$ for $x=0, \dots, 7$ (the possible values of X).

(1) Find the PF of X .

(2) Find the probability that X will be at least 5.

2.6 Suppose a discrete random variable X has the PF given by

X	3	5	7	8	9	10	12
$P(X=x)$	0.08	0.10	0.16	0.25	0.20	0.03	0.18

What is the probability that: (1) $X > 9$; (2) $4 < X < 11$; (3) X is 7.

2.7 Suppose that a box contains seven red balls and three blue balls. If five balls are selected at random, without replacement. Determine the PF of the number of red balls that will be obtained.

2.8 Suppose that a random variable X has the binomial distribution with parameters $n=15$ and $p=0.5$. Find $P(X < 6)$.

2.9 If 10 percent of the balls in a certain box are red, and if 20 balls are selected from the box at random, with replacement. What is the probability that more than three red balls will be obtained?

2.10 The probability of winning at a certain game is 0.3. If you play the game 10 times, what is the probability that you win at most once?

2.11 Suppose that X follows the Poisson distribution. $P(X=1) = P(X=2)$. Find $P(X=4)$.

2.12 The number of calls coming per minute into a hotel reservation center is Poisson random variable with $\lambda=3$.

(1) Find the probability that no calls come in a given one-minute period.

(2) Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two-minute period.

2.13 Suppose that a part has a one in a hundred chance of failing and we sample from 100 independent parts. What is the probability that none of the 100 parts fails?

2.14 The number of E-mails received in a day has a Poisson distribution with parameter λ . On that day, the probability of one E-mail being spammed is denoted by p . Show that the number of spam E-mails received that day also follows a Poisson distribution.

2.15 A crate contains 50 light bulbs of which 5 are defective and 45 are not. A quality control inspector randomly samples 4 bulbs without replacement. Let X be the number of defective bulbs selected. Find the PF of the discrete random variable X .

2.16 Products produced by a machine has a 5% defective rate.

(1) What is the probability that the first defective occurs in the sixth item inspected?

(2) What is the probability that the first defective occurs in the first six inspections?

2.17 An oil company conducts a geological study that indicates that an exploratory oil well should have a 30% chance of striking oil.

(1) What is the probability that the first strike comes on the third well drilled?

(2) What is the probability that the third strike comes on the seventh well drilled?

2.18 Suppose that the PDF of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{1}{36}(9-x^2), & -3 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

Sketch this PDF and determine the values of the following probabilities: (1) $P(X < 0)$;

(2) $P(-1 \leq X \leq 1)$; (3) $P(X > 2)$.

2.19 Suppose that the PDF of a random variable X is as follows:

$$f(x) = \begin{cases} cx^2, & 1 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}.$$

(1) Find the value of the constant c and sketch the PDF.

(2) Find the value of $P(X > \frac{3}{2})$, $P(X = \frac{3}{2})$.

2.20 Suppose that the PDF of a random variable X is as follows:

$$f(x) = \begin{cases} \frac{1}{8}x, & 0 \leq x \leq 4 \\ 0, & \text{otherwise} \end{cases}.$$

(1) Find the value of a such that $P(X \leq a) = \frac{1}{4}$.

(2) Find the value of b such that $P(X \geq b) = \frac{1}{2}$.

2.21 Suppose that the PDF of a random variable X is as follows:

$$f(x) = \begin{cases} ce^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

(1) Find the value of the constant c and sketch the PDF.

(2) Find the value of $P(X < 2)$.

2.22 Suppose that a random variable X can take only the values $-2, 0, 1, 4$, and that the probabilities of these values are as follows: $P(X=2)=0.4$, $P(X=0)=0.1$, $P(X=1)=0.3$, $P(X=4)=0.2$. Sketch the CDF of X .

2.23 Suppose that the CDF of a random variable X is as follows:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{9}x^2, & 0 < x \leq 3 \\ 1, & x \geq 3 \end{cases}.$$

Find and sketch the PDF of X .

2.24 Suppose that X has the PDF

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Find and sketch the CDF of X .

2.25 Suppose that K follows the uniform distribution over the interval $(0, 5)$. Find the probability that the following equation has real roots.

$$4x^2 + 4Kx + K + 2 = 0.$$

2.26 Suppose that X follows a discrete distribution with the PF

X	-2	-1	0	1	2	4
P	0.2	0.1	0.3	0.1	0.2	0.1

Find the probability that the following equation has real roots.

$$3a^2 + 2aX + X + 1 = 0.$$

2.27 Suppose that the amount of waiting time X (unit: minute) a customer spends in a bank is exponentially distributed with the following PDF

$$f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

The customer decides to leave if the waiting time is not less than 10 minutes. He usually visits the bank 5 times per month. Let Y be the times that he is not served within a certain month. Find $P(Y \geq 1)$.

2.28 Suppose that $X \sim N(160, \sigma^2)$. If $P(120 < X < 200) \geq 0.80$, what is the largest value σ can take?

2.29 Suppose X , the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10. The instructor wants to give 15% of the class an A. What cutoff should the instructor use to determine who gets an A?

2.30 A random variable X has a normal distribution with $\sigma = 10$. If the probability that the random variable will take on a value less than 82.5 is 0.8212, what is the probability that it will take on a value greater than 58.3?

2.31 An electronic device has probabilities of 0.1, 0.001, 0.2 of failing subject to voltage of less than 220 V, 220 ~ 240 V, larger than 240 V, respectively. Suppose the voltage $X \sim N(220, 25^2)$.

(1) Find the probability that the electronic device will fail.

(2) If it failed, what is the probability that it had received the voltage of 220 ~ 240 V.

2.32 A certain type of electronic component has a lifetime X (in hours) with PDF given by

$$f(x) = \begin{cases} \frac{c}{x^2}, & x > 1000 \\ 0, & x \leq 1000 \end{cases}.$$

(1) Find the constant c .

(2) Find $P(X \leq 1800 | X > 1500)$.

(3) Consider 3 independent electronic components of this type. What is the probability that only one component failed for the first 1500 hours?

2.33 Suppose the PDF of X is

$$f(x) = \begin{cases} Ax, & 1 < x < 2 \\ B, & 2 < x < 3 \\ 0, & \text{otherwise} \end{cases}.$$

And $P(1 < X < 2) = 2P(2 < X < 3)$. Find:

(1) constants A and B ;

(2) the CDF of X .

2.34 The PF of X is

X	-2	0	2	3
P	0.2	0.2	0.5	0.1

Find the PF of $Y = X^2 + 1$.

2.35 Suppose that X is exponentially distributed with the following PDF

$$f(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Find the PDF of (1) $Y = \frac{1}{X}$. (2) $Y = X^2$. (3) $Y = \sqrt{X}$.

2.36 Suppose that the PDF of X is

$$f(x) = \begin{cases} x^3 e^{-x^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Find the PDF of (1) $Y = 2X + 4$, (2) $Y = X^2$, (3) $Y = \ln X$.

2.37 Suppose $X \sim U(-2, 5)$, find the PDF of $Y = X^2$.

2.38 Suppose $\ln X \sim N(1, 2^2)$, find $P\left(\frac{1}{2} < X < 2\right)$.

2.39 Suppose that the PDF of a random variable X is as follows:

$$f(x) = \begin{cases} (c^2 + c)x, & 1 \leq x \leq \sqrt{\frac{3}{2}} \\ 0, & \text{otherwise} \end{cases}.$$

(1) Determine $c > 0$ that renders $f(x)$ a valid density function.

(2) Find $P(X < 1.1)$.

(3) Find $P(X = 1.1)$.

2.40 Suppose that X is exponentially distributed with the following PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

where $\lambda > 0$. Let $Y = \min\{2, X\}$. Find the CDF of Y .

Chapter 3

Multivariate Probability Distributions



$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{(x_i, y_j) \leq (x, y)} p(x_i, y_j)$$

◆ 3.1 Bivariate Distributions ◆

So far we have been only dealing with the probability distributions of single random variables. However, in some random phenomena, we need consider several random variables simultaneously. For example, the components of wind velocity in studies of atmospheric turbulence, the horizontal distance and vertical distance between the bullet point and the target.

In such studies, it is not only interesting to study the behavior of each variable individually, but also to investigate the relationship between them.

Definition 3.1.1 Suppose E is a random experiment with sample space $S = \{e\}$. $X_1 = X_1(e), X_2 = X_2(e), \dots, X_n = X_n(e)$ are random variables defined on S . $(X_1, X_2, \dots, X_n)$ is called an n -dimensional random variable or n -dimensional random vector.

We will focus on two dimensional random variable and most of the results can be extended to the n -dimensional case.

Similar to the CDF of univariate random variable, we define the CDF of bivariate random variable as follows.

Definition 3.1.2 The joint distribution function or joint cumulative distribution function (joint CDF) of two random variables (discrete or continuous) X and Y is defined by

$$F(x, y) = P(X \leq x, Y \leq y),$$

for $-\infty < x < +\infty, -\infty < y < +\infty$. Similarly, the joint distribution function of n random variables $(X_1, X_2, \dots, X_n)$ is defined by

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

for $-\infty < x_i < +\infty, i = 1, 2, \dots, n$.

(X, Y) can be considered as the coordinates of a random point in the xy -plane. $F(x, y)$ is defined as the probability that the random point will fall within the semi-infinite rectangle with upper limits defined by $X = x, Y = y$. See Figure 3.1.

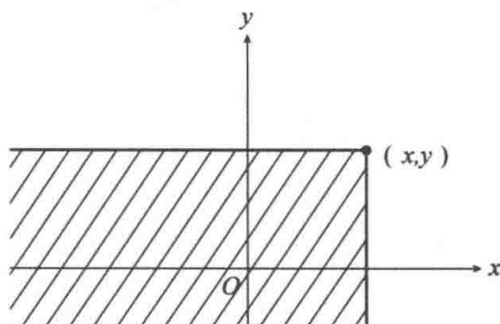


Figure 3.1

The probability that the random point will fall within the rectangle $\{(X,Y) | a < X \leq b, c < Y \leq d\}$ is given by

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c),$$

for given numbers $a < b$ and $c < d$. See Figure 3.2.

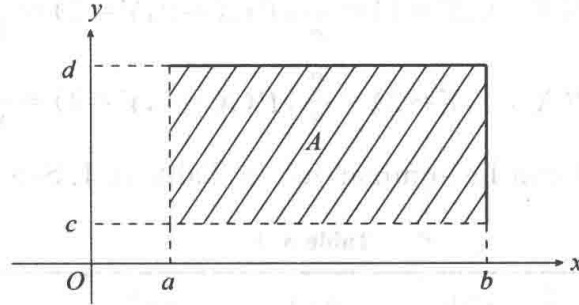


Figure 3.2

This bivariate distribution function has properties analogous to those of the univariate distribution function.

Proposition 3.1.1 The joint CDF $F(x, y)$ of (X, Y) has the following properties:

(1) $F(x, y)$ is non-decreasing in both x and y . That is, for fixed y , if $x_1 < x_2$, then $F(x_1, y) \leq F(x_2, y)$; for fixed x , if $y_1 < y_2$, then $F(x, y_1) \leq F(x, y_2)$.

(2) $0 \leq F(x, y) \leq 1$, and for fixed x , $F(x, -\infty) = 0$; for fixed y , $F(-\infty, y) = 0$; $F(-\infty, -\infty) = 0$; $F(+\infty, +\infty) = 1$.

(3) $F(x, y)$ is right continuous at x and y , that is,

$$F(x, y) = F(x^+, y), F(x, y) = F(x, y^+).$$

(4) For any real numbers $a < b$ and $c < d$,

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0.$$

Definition 3.1.3 Let X and Y be random variables, and consider the ordered pair (X, Y) . If there are only finitely or at most countably many different possible values (x, y) for the pair (X, Y) , then we say that X and Y have a discrete joint distribution.

Definition 3.1.4 The joint probability function (abbreviated as joint PF) of X and Y is defined as the function f such that for every point (x, y) in the xy -plane,

$$f(x, y) = P(X = x, Y = y).$$

Proposition 3.1.2 Let X and Y have a discrete joint distribution, $f(x, y) \geq 0$. If (x, y) is not one of the possible values of the pair (X, Y) , then $f(x, y) = 0$. Also,

$$\sum_{\text{All } (x, y)} f(x, y) = 1.$$

Additionally, for each set A of ordered pairs,

$$P[(X, Y) \in A] = \sum_{(x, y) \in A} f(x, y).$$

The joint cumulative distribution function of (X, Y) can be written as

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x, y_j \leq y} f(x_i, y_j),$$

where (x_i, y_j) are all possible values of (X, Y) , $i, j = 1, 2, \dots$.



Example 3.1.1 A fair coin is tossed three times independently. Let X denote the number of heads on the first toss and Y denote the total number of heads. Find the joint probability function of X and Y .

Solution: X takes 0, 1; Y takes 0, 1, 2, 3. And

$$P(X=0, Y=0) = \frac{1}{8}, P(X=0, Y=1) = \frac{2}{8}, P(X=0, Y=2) = \frac{1}{8}$$

$$P(X=1, Y=1) = \frac{1}{8}, P(X=1, Y=2) = \frac{2}{8}, P(X=1, Y=3) = \frac{1}{8}$$

The joint PF of (X, Y) can be summarized in Table 3.1. See Figure 3.3.

Table 3.1

$f(x, y)$	$y=0$	$y=1$	$y=2$	$y=3$
$x=0$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
$x=1$	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

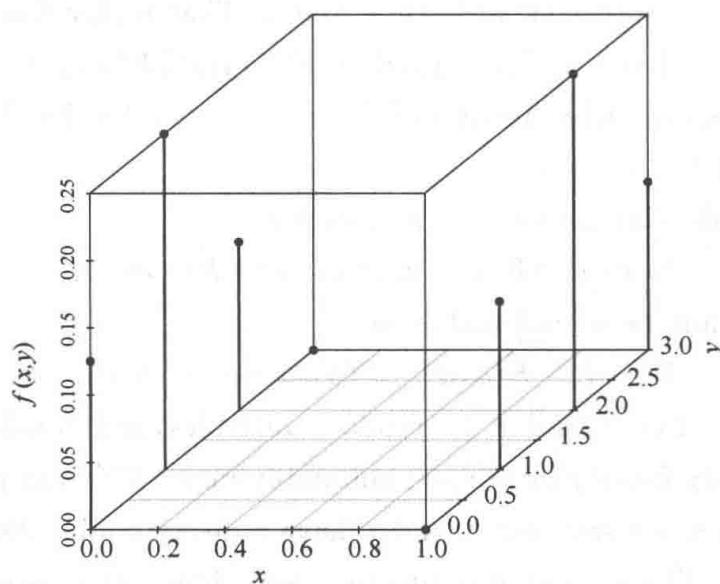


Figure 3.3

Example 3.1.2 The joint PF of (X, Y) is as specified in Table 3.2. See Figure 3.4. Determine $P(X \geq 2, Y \geq 2)$ and $P(X=1)$.

Table 3.2

	$y=1$	$y=2$	$y=3$	$y=4$
$x=1$	0.1	0	0.1	0
$x=2$	0.3	0	0.1	0.2
$x=3$	0	0.2	0	0

Solution:

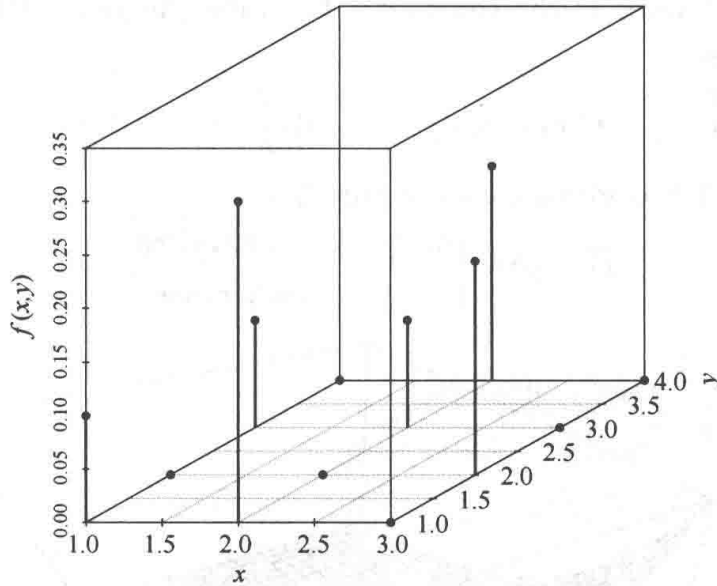


Figure 3.4

$$P(X \geq 2, Y \geq 2) = f(2,2) + f(2,3) + f(2,4) + f(3,2) + f(3,3) + f(3,4) = 0.5.$$

$$P(X = 1) = \sum_{y=1}^4 f(1, y) = 0.2.$$

Definition 3.1.5 Two random variables X and Y have a continuous as joint distribution if there exists a nonnegative function $f(x, y)$ defined over the entire xy -plane such that for any real values x and y ,

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv.$$

The function f is called the joint probability density function (abbreviated as joint PDF) of X and Y .

Proposition 3.1.3 A joint PDF of (X, Y) satisfies the following properties:

- (1) $f(x, y) \geq 0$.
- (2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = F(+\infty, +\infty) = 1$.
- (3) Suppose G is a region in xy -plane, then

$$P[(X, Y) \in G] = \iint_G f(x, y) dx dy.$$

- (4) The joint PDF can be derived from the joint CDF by using the relations

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

at every point (x, y) at which these second-order derivatives exist.

In geometry, $Z = f(x, y)$ represents a surface in space. $P[(X, Y) \in G]$ equals the volume under density surface above G .

Example 3.1.3 Suppose that the joint PDF of (X, Y) is specified as follows:

$$f(x, y) = \begin{cases} ce^{-(2x+3y)}, & x \geq 0, y \geq 0 \\ 0 & , \text{otherwise} \end{cases}.$$

(1) Determine the value of the constant c . (2) Find the joint CDF of (X, Y) .

Solution: (1) Since

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \int_0^{+\infty} dx \int_0^{+\infty} c e^{-(2x+3y)} dy = \frac{c}{6},$$

thus, $c=6$. The joint PDF is given as, see Figure 3.5,

$$f(x, y) = \begin{cases} 6e^{-(2x+3y)}, & x \geq 0, y \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

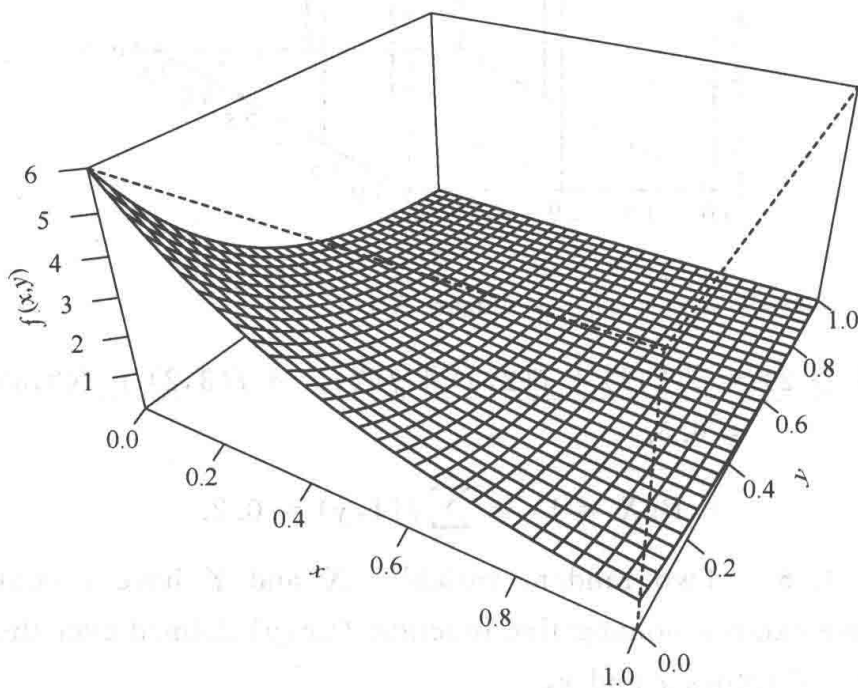


Figure 3.5

(2) The support S of (X, Y) is sketched in Figure 3.6. $f(x, y)$ takes different forms in various regions, so the joint CDF should be calculated by regions. If $x \geq 0, y \geq 0$,

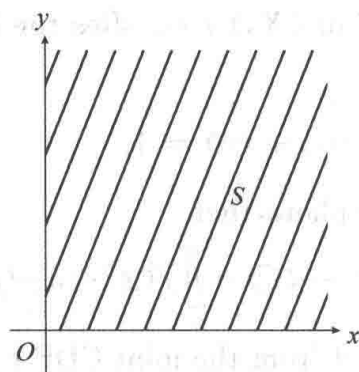


Figure 3.6

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \\ &= \iint_S f(u, v) du dv \\ &= \int_0^x \left[\int_0^y 6e^{-(2u+3v)} dv \right] du \\ &= (1 - e^{-2x})(1 - e^{-3y}). \end{aligned}$$

If $x < 0$ or $y < 0$, $f(x, y) = 0$, then

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = 0.$$

Therefore, the CDF of (X, Y) is, see Figure 3.7,

$$F(x, y) = \begin{cases} (1 - e^{-2x})(1 - e^{-3y}), & x \geq 0, y \geq 0 \\ 0 & , \text{otherwise} \end{cases}$$

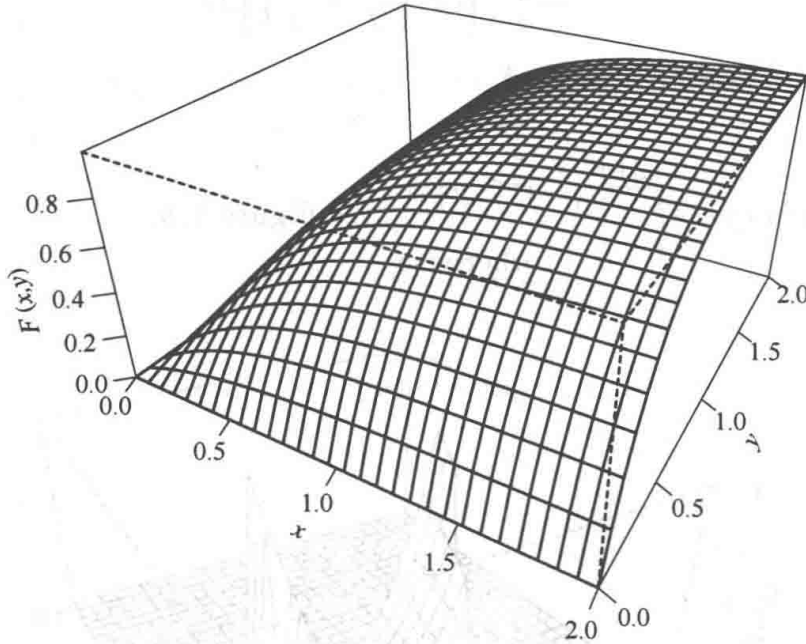


Figure 3.7

Note: the plots of joint PDF $f(x, y)$ and joint CDF $F(x, y)$ are depicted just for illustration. They are not required or necessary for the homework problems. But the support of (X, Y) should be sketched.

Example 3.1.4 Suppose that the joint PDF of X and Y is specified as follows:

$$f(x, y) = \begin{cases} cx^2y, & x^2 \leq y \leq 1 \\ 0 & , \text{otherwise} \end{cases}$$

(1) Determine the value of the constant c . (2) Determine the value of $P(X \geq Y)$.

Solution: (1) The support S of (X, Y) is sketched in Figure 3.8. For $f(x, y) = 0$ outside S , it follows that

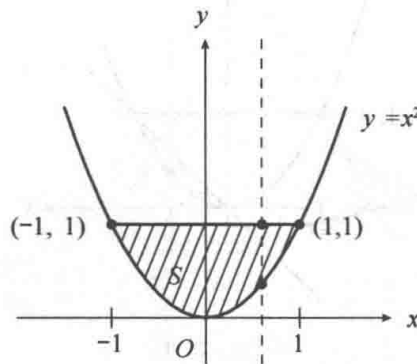


Figure 3.8

$$\begin{aligned}
 1 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy \\
 &= \iint_S f(x, y) dx dy \\
 &= \int_{-1}^1 \left(\int_{x^2}^1 cx^2 y dy \right) dx \\
 &= \int_{-1}^1 \left[cx^2 \left(\frac{1}{2} - \frac{x^4}{2} \right) \right] dx \\
 &= \frac{4}{21} c.
 \end{aligned}$$

Thus, $c = \frac{21}{4}$. $f(x, y) = \begin{cases} \frac{21x^2y}{4}, & x^2 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$. See Figure 3. 9.

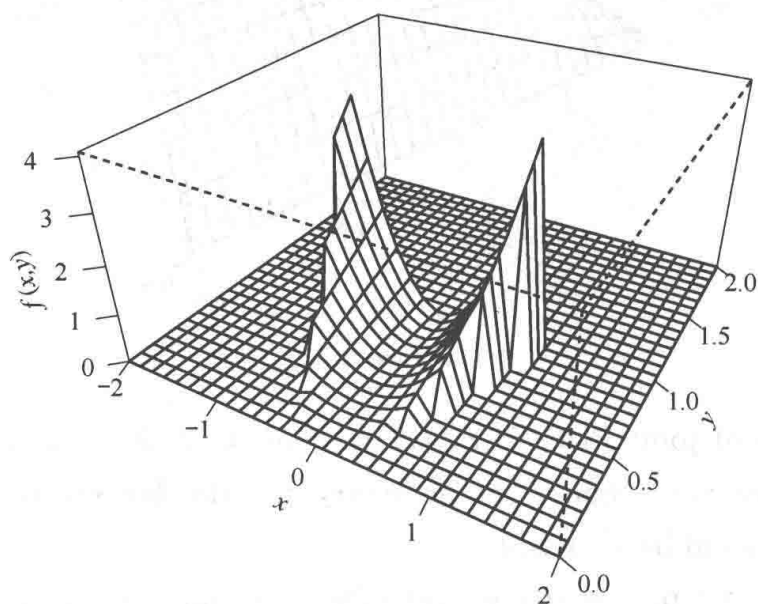


Figure 3. 9

(2) $P(X \geq Y) = P[(X, Y) \in S_0]$, where S_0 is shown in Figure 3. 10.

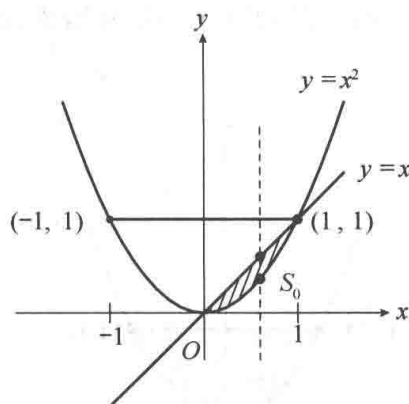


Figure 3. 10

$$\begin{aligned}
 P(X \geq Y) &= \iint_{S_0} f(x, y) dx dy \\
 &= \int_0^1 \left(\int_{x^2}^x \frac{21x^2 y}{4} dy \right) dx \\
 &= \int_0^1 \left[\frac{21}{8} (x^4 - x^6) \right] dx \\
 &= \frac{3}{20}.
 \end{aligned}$$

Definition 3.1.6 (X, Y) follows bivariate uniform distribution in a bounded region G if the joint PDF is given as

$$f(x, y) = \begin{cases} \frac{1}{S_G}, & (x, y) \in G \\ 0, & \text{otherwise} \end{cases},$$

where S_G is the area of G .

If (X, Y) follows bivariate uniform distribution in a bounded region G , then for any subregion D in G ,

$$P[(X, Y) \in D] = \iint_D f(x, y) dx dy = \iint_D \frac{1}{S_G} dx dy = \frac{S_D}{S_G},$$

where S_D is the area of D . It shows that the probability that the random point (X, Y) lies in a subregion depends only on its area and has no relation with its location and shape if (X, Y) follows bivariate uniform distribution.

Definition 3.1.7 When the joint PDF of two random variables X and Y is of the form

$$\begin{aligned}
 f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\}, \\
 &\quad -\infty < x, y < +\infty.
 \end{aligned}$$

It is said that X and Y have the bivariate normal distribution $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Bivariate normal distribution is an important continuous distribution. The shape of its joint PDF looks like a hat, see Figure 3.11.

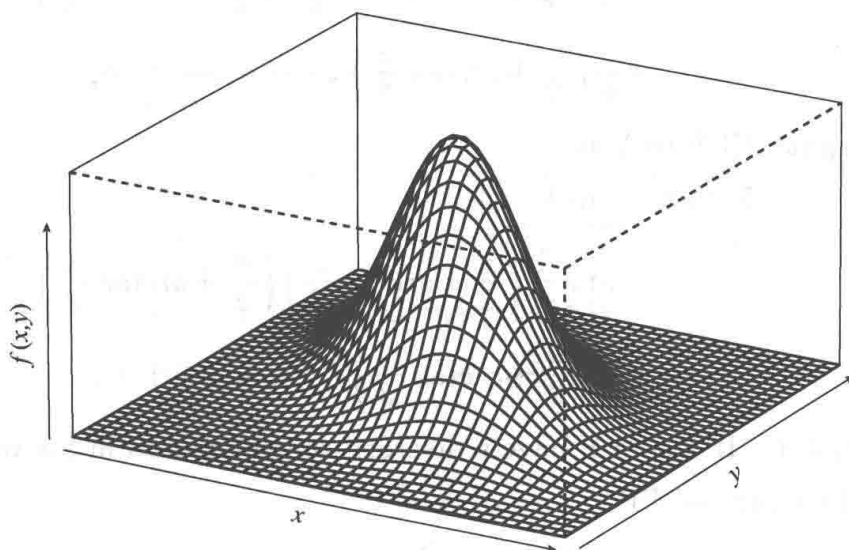


Figure 3.11

◆ 3.2 Marginal Distributions ◆

Let X and Y have a joint CDF $F(x, y)$. The CDF $F_X(x)$ of the single random variable X can be derived from the joint CDF $F(x, y)$ as

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < +\infty) = F(x, +\infty) = \lim_{y \rightarrow +\infty} F(x, y).$$

Similarly, the CDF $F_Y(y)$ of Y equals

$$F_Y(y) = P(Y \leq y) = P(x < +\infty, Y \leq y) = F(+\infty, y) = \lim_{x \rightarrow +\infty} F(x, y).$$

$F_X(x)$ and $F_Y(y)$ are also called marginal CDFs of (X, Y) . Since we're considering more than one variable at a time now, a subscript in the function is used for clarification.

Example 3.2.1 Let X and Y have a joint CDF $F(x, y) = A(B + \arctan \frac{x}{2})(C + \arctan \frac{y}{3})$, $-\infty < x, y < +\infty$. (1) Determine the constants A, B, C . (2) Find the marginal CDFs $F_X(x)$ and $F_Y(y)$.

Solution: (1) By the properties of $F(x, y)$,

$$1 = F(+\infty, +\infty) = A\left(B + \frac{\pi}{2}\right)\left(C + \frac{\pi}{2}\right),$$

$$0 = F(x, -\infty) = A\left(B + \arctan \frac{x}{2}\right)\left(C - \frac{\pi}{2}\right),$$

$$0 = F(-\infty, y) = A\left(B - \frac{\pi}{2}\right)\left(C + \arctan \frac{y}{3}\right).$$

Solve the equations, $A = \frac{1}{\pi^2}$, $B = \frac{\pi}{2}$, $C = \frac{\pi}{2}$.

(2) The marginal CDF of X is

$$\begin{aligned} F_X(x) &= \lim_{y \rightarrow +\infty} F(x, y) \\ &= \lim_{y \rightarrow +\infty} \frac{1}{\pi^2} \left(\frac{\pi}{2} + \arctan \frac{x}{2} \right) \left(\frac{\pi}{2} + \arctan \frac{y}{3} \right) \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \frac{x}{2} \right), \quad -\infty < x < +\infty. \end{aligned}$$

Similarly, the marginal CDF of Y is

$$\begin{aligned} F_Y(y) &= \lim_{x \rightarrow +\infty} F(x, y) \\ &= \lim_{x \rightarrow +\infty} \frac{1}{\pi^2} \left(\frac{\pi}{2} + \arctan \frac{x}{2} \right) \left(\frac{\pi}{2} + \arctan \frac{y}{3} \right) \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \frac{y}{3} \right), \quad -\infty < y < +\infty. \end{aligned}$$

Proposition 3.2.1 If X and Y have a discrete joint distribution for which the joint PF is $f(x, y)$, then the marginal PF $f_X(x)$ of X is

$$f_X(x) = \sum_{\text{All } y} f(x, y).$$

Similarly, the marginal PF $f_Y(y)$ of Y is

$$f_Y(y) = \sum_{\text{All } x} f(x, y).$$

Example 3.2.2 The joint PF $f(x, y)$ of X and Y is as specified in the following Table 3.3. Find the two marginal PFs.

Table 3.3

	$y=1$	$y=2$	$y=3$	$y=4$
$x=1$	0.1	0	0.1	0
$x=2$	0.3	0	0.1	0.2
$x=3$	0	0.2	0	0

Solution: The marginal PF $f_X(x)$ of X is

$$f_X(x=1) = 0.1 + 0 + 0.1 + 0 = 0.2,$$

$$f_X(x=2) = 0.3 + 0 + 0.1 + 0.2 = 0.6,$$

$$f_X(x=3) = 0 + 0.2 + 0 + 0 = 0.2.$$

See Table 3.4.

Table 3.4

x	$x=1$	$x=2$	$x=3$
$f_X(x)$	0.2	0.6	0.2

The marginal PF $f_Y(y)$ of Y is

$$f_Y(y=1) = 0.1 + 0.3 + 0 = 0.4,$$

$$f_Y(y=2) = 0 + 0 + 0.2 = 0.2,$$

$$f_Y(y=3) = 0.1 + 0.1 + 0 = 0.2,$$

$$f_Y(y=4) = 0 + 0.2 + 0 = 0.2.$$

See Table 3.5.

Table 3.5

y	$y=1$	$y=2$	$y=3$	$y=4$
$f_Y(y)$	0.4	0.2	0.2	0.2

Suppose (X, Y) is a bivariate continuous random variable with PDF $f(x, y)$, then

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(u, y) dy \right] du.$$

Take derivative with respect to x to get the marginal PDF of X ,

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

Similarly, the marginal PDF of Y is

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Example 3.2.3 Suppose that (X, Y) is uniformly distributed over the bounded

region D enclosed by $y=x^2$ and line $y=x$, see Figure 3. 12. Determine the joint PDF of (X,Y) and the marginal PDFs $f_X(x)$ and $f_Y(y)$.

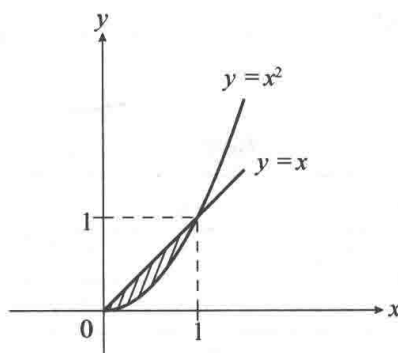


Figure 3. 12

Solution: The area of region D is

$$S_D = \int_0^1 (x - x^2) dx = \frac{1}{6}.$$

The joint PDF of (X,Y) is $f(x,y) = \begin{cases} 6, & x^2 \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$. We determine first the marginal PDF $f_X(x)$ of X .

(1) If $x < 0$ or $x > 1$, $f_X(x) = 0$.

(2) If $0 \leq x \leq 1$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x,y) dy \\ &= \int_{x^2}^x 6 dy \\ &= 6(x - x^2). \end{aligned}$$

$$\text{In summary, } f_X(x) = \begin{cases} 6(x - x^2), & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

We determine the marginal PDF $f_Y(y)$ of Y .

(1) If $y < 0$ or $y > 1$, $f_Y(y) = 0$.

(2) If $0 \leq y \leq 1$,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y) dx \\ &= \int_y^{\sqrt{y}} 6 dx \\ &= 6(\sqrt{y} - y). \end{aligned}$$

$$\text{In summary, } f_Y(y) = \begin{cases} 6(\sqrt{y} - y), & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Example 3. 2. 4 Suppose X and Y have the bivariate normal distribution $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Determine the marginal PDFs $f_X(x)$ and $f_Y(y)$.

Solution: The joint PDF of (X,Y) is of the form

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

$$-\infty < x, y < +\infty.$$

Since

$$\begin{aligned} & -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right] \\ &= -\frac{1}{2(1-\rho^2)}\left[(1-\rho^2)\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)^2\right] \\ &= -\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{(x-\mu_1)^2}{2\sigma_1^2}. \end{aligned}$$

The marginal PDF of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)^2\right\} dy. \end{aligned}$$

Set $u = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)$, then

$$f_X(x) = \frac{1}{2\pi\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \frac{1}{2\pi\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \sqrt{2\pi} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

where $-\infty < x < +\infty$. It shows that $X \sim N(\mu_1, \sigma_1^2)$. In the same way, the marginal PDF of

Y is $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$, $-\infty < y < +\infty$. $Y \sim N(\mu_2, \sigma_2^2)$.

Remark 3.2.1 The marginal distributions of bivariate normal distribution are two univariate normal distributions, both of them have no relation with the parameter ρ . In general, given the two marginal distributions, the joint distribution can not be determined. Further, even if the two marginal distributions are normal distributions, their joint distribution can be nonnormal. For example, consider the joint PDF of (X, Y) ,

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} (1 + \sin x \sin y), \quad -\infty < x, y < +\infty.$$

(X, Y) does not follow bivariate normal distribution while $X \sim N(0, 1)$ and $Y \sim N(0, 1)$.

3.3 Conditional Distributions

Sometimes we need consider the behavior of one random variable given that another variable takes a particular value. We generalize the concept of conditional probability to conditional distributions.

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

Definition 3.3.1 Let X and Y have a discrete joint distribution with joint PF $f(x, y) =$

$P(X=x, Y=y)$. Let $f_Y(y) = P(Y=y)$ denote the marginal PF of Y . For each y such that $f_Y(y) = P(Y=y) > 0$, define

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Then $f_{X|Y}(x|y)$ is called the conditional PF of X given Y . The discrete distribution whose PF is $f_{X|Y}(\cdot|y)$ is called the conditional distribution of X given that $Y=y$.

Similarly, if x is a given value of X such that $f_X(x) = P(X=x) > 0$, then

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

is the conditional PF of Y given that $X=x$.

Example 3.3.1 The joint PF of (X, Y) is given in Table 3.6. Determine

- (1) the conditional PF of X given that $Y=1$;
- (2) the conditional PF of Y given that $X=2$.

Table 3.6

	$y=1$	$y=2$	$y=3$	$y=4$	Total
$x=1$	0.1	0	0.1	0	0.2
$x=2$	0.3	0	0.1	0.2	0.6
$x=3$	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1

Solution: (1) From Table 3.6, $f_Y(1) = P(Y=1) = 0.4$. The conditional probability $f_{X|Y}(x|y=1)$ that X will take a particular value x is

$$f_{X|Y}(x|y=1) = \frac{f(x, y=1)}{f_Y(y=1)} = \frac{f(x, y=1)}{0.4}.$$

$$f_{X|Y}(x=1|y=1) = \frac{1}{4}, f_{X|Y}(x=2|y=1) = \frac{3}{4}, f_{X|Y}(x=3|y=1) = 0.$$

The conditional PF of X given that $Y=1$ is shown in Table 3.7.

Table 3.7

$x y=1$	$x=1$	$x=2$	$x=3$
$f_{X Y}(x y=1)$	$\frac{1}{4}$	$\frac{3}{4}$	0

(2) From Table 3.6, $f_X(2) = P(X=2) = 0.6$. The conditional probability $f_{Y|X}(y|x=2)$ that Y will take a particular value y is

$$f_{Y|X}(y|x=2) = \frac{f(x=2, y)}{f_X(x=2)} = \frac{f(x=2, y)}{0.6}.$$

$$f_{Y|X}(y=1|x=2) = \frac{1}{2}, f_{Y|X}(y=2|x=2) = 0,$$

$$f_{Y|X}(y=3|x=2) = \frac{1}{6}, f_{Y|X}(y=4|x=2) = \frac{1}{3}.$$

The conditional PF of Y given that $X=2$ is shown in Table 3.8.

Table 3.8

$y x=2$	$y=1$	$y=2$	$y=3$	$y=4$
$f_{Y X}(y x=2)$	$\frac{1}{2}$	0	$\frac{1}{6}$	$\frac{1}{3}$

Example 3.3.2 A shooter shoots a target until getting two hits. In each shot, the probability of hitting the target is p . Let X be the number of shots to get the first hit and Y be the number of total shots. Determine the joint PF of (X, Y) and its conditional distributions.

Solution: (1) X takes $1, 2, \dots$. Y takes $2, 3, \dots$, and $X < Y$. The joint PF of (X, Y) is

$$P(X=m, Y=n) = p^2 q^{n-2},$$

where $m=1, 2, \dots, n-1; n=2, 3, \dots; q=1-p$.

(2) The marginal PF of X is

$$\begin{aligned} P(X=m) &= \sum_{n=m+1}^{+\infty} P(X=m, Y=n) \\ &= \sum_{n=m+1}^{+\infty} p^2 q^{n-2} \\ &= pq^{m-1}, \end{aligned}$$

where $m=1, 2, \dots$.

The marginal PF of Y is

$$\begin{aligned} P(Y=n) &= \sum_{m=1}^{n-1} P(X=m, Y=n) \\ &= \sum_{m=1}^{n-1} p^2 q^{n-2} \\ &= (n-1)p^2 q^{n-2}, \end{aligned}$$

where $n=1, 2, \dots$.

(3) The conditional PF of X given that $Y=n$ is

$$P(X=m | Y=n) = \frac{P(X=m, Y=n)}{P(Y=n)} = \frac{p^2 q^{n-2}}{(n-1)p^2 q^{n-2}} = \frac{1}{n-1},$$

where $m=1, 2, \dots, n-1$.

The conditional PF of Y given that $X=m$ is

$$P(Y=n | X=m) = \frac{P(X=m, Y=n)}{P(X=m)} = \frac{p^2 q^{n-2}}{pq^{m-1}} = pq^{n-m-1},$$

where $n=m+1, m+2, \dots$.

For continuous variables, the conditional density function plays the analogous role to the conditional probability function.

Definition 3.3.2 Let X and Y have a continuous joint distribution with the joint PDF $f(x, y)$ and respective marginals $f_X(x)$ and $f_Y(y)$. Let y be a value such that $f_Y(y) > 0$. Then the conditional PDF $f_{X|Y}(x | y)$ of X given that $Y=y$ is defined as follows:

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}, -\infty < x < +\infty.$$

For values of y such that $f_Y(y) = 0$, we are free to define $f_{X|Y}(x|y)$ however we wish, so long as $f_{X|Y}(x|y)$ is a PDF as a function of x . The conditional PDF $f_{Y|X}(y|x)$ of Y given that $X=x$, can be defined in the same way.

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, -\infty < y < +\infty.$$

The conditional PDF is in fact a valid probability density function, as it is nonnegative and integrates to 1. For a given y ,

$$\int_{-\infty}^{+\infty} f_{X|Y}(x|y) dx = \int_{-\infty}^{+\infty} \frac{f(x,y)}{f_Y(y)} dx = \frac{1}{f_Y(y)} \int_{-\infty}^{+\infty} f(x,y) dx = \frac{f_Y(y)}{f_Y(y)} = 1.$$

Geometrically, the conditional PDF of X given $Y=y$ is a one-dimensional “slice” of the two-dimensional joint PDF along the line $Y=y$, and renormalized so that the resulting curve integrates to 1.

When working with conditional PDF $f_{X|Y}(x|y)$, the conditioning variable y is treated as fixed. The conditional PDF is a function of its argument x . When working with conditional PDF $f_{Y|X}(y|x)$, the conditioning variable x is treated as fixed. The conditional PDF is a function of its argument y .

The conditional PDFs are used to compute conditional probabilities and behave just like densities. For an interval $C \in \mathbb{R}$,

$$P(X \in C | Y = y) = \int_C f_{X|Y}(x|y) dx.$$

Example 3.3.3 Suppose that X and Y have a joint continuous distribution with the joint PDF

$$f(x,y) = \begin{cases} e^{-x}, & 0 \leq y \leq x \\ 0, & \text{otherwise} \end{cases}.$$

Find (1) $f_{X|Y}(x|y)$; (2) $f_{Y|X}(y|x)$; (3) $P(X < 2 | Y = 1)$.

Solution: (1) Determine the marginal PDF $f_X(x)$ of X . See Figure 3.13(a).

① If $x < 0$, $f_X(x) = 0$.

② If $x \geq 0$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x,y) dy \\ &= \int_0^x e^{-x} dy \\ &= xe^{-x}. \end{aligned}$$

In summary, $f_X(x) = \begin{cases} xe^{-x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$.

(2) Determine the marginal PDF $f_Y(y)$ of Y . See Figure 3.13(b).

① If $y < 0$, $f_Y(y) = 0$.

② If $y \geq 0$,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \int_y^{+\infty} e^{-x} dx = e^{-y}.$$

In summary, $f_Y(y) = \begin{cases} e^{-y}, & y \geq 0 \\ 0, & \text{otherwise} \end{cases}$. Therefore,

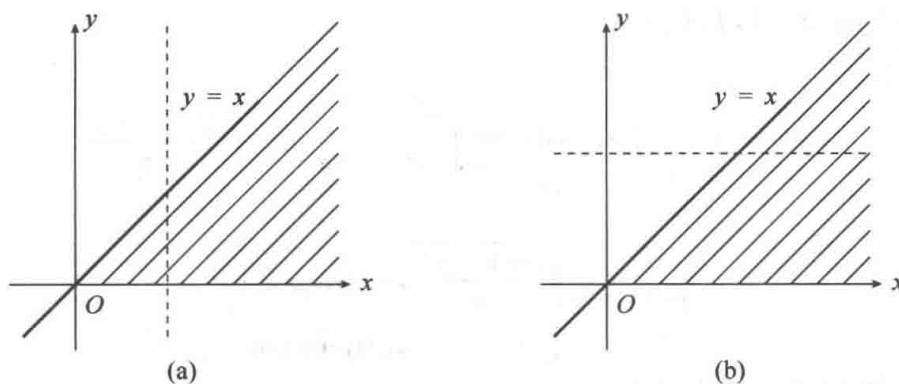


Figure 3.13

$$(3) \quad f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} e^{y-x}, & 0 \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} \frac{1}{x}, & 0 \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$$

Given that $Y=1$, $f_{X|Y}(x|1) = \begin{cases} e^{1-x}, & 1 \leq x \\ 0, & \text{otherwise} \end{cases}$,

$$P(X < 2 | Y = 1) = \int_{-\infty}^2 f_{X|Y}(x|1) dx = \int_1^2 e^{1-x} dx = 1 - e^{-1}.$$

Example 3.3.4 Suppose (X, Y) is uniformly distributed over the unit circle $\{(x, y) | x^2 + y^2 \leq 1\}$, determine

(1) conditional PDF $f_{X|Y}(x|y)$;

(2) $P(X > \frac{1}{2} | Y = \frac{1}{2})$;

(3) $P(X > \frac{1}{2} | Y \geq \frac{1}{2})$.

Solution: (1) The joint PDF of (X, Y) is

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine the marginal PDF $f_Y(y)$ of Y . See Figure 3.14.

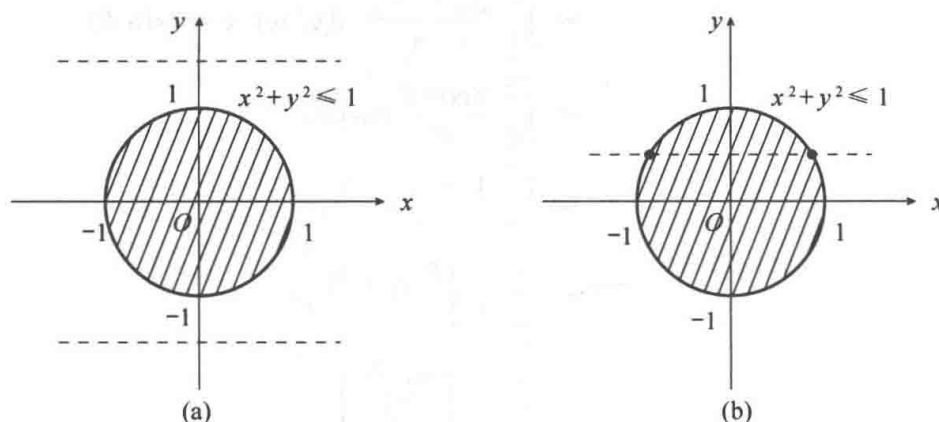


Figure 3.14

① If $y < -1$ or $y > 1$, $f_Y(y) = 0$.

② If $-1 \leq y \leq 1$,

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi} dx = \frac{2\sqrt{1-y^2}}{\pi}.$$

Therefore,

$$f_Y(y) = \begin{cases} \frac{2\sqrt{1-y^2}}{\pi}, & -1 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

The conditional PDF $f_{X|Y}(x|y)$ is

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \begin{cases} \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}}, & x^2 + y^2 \leq 1 \\ \frac{1}{2\sqrt{1-y^2}}, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

(2) Since $P(Y = \frac{1}{2}) = 0$, the conditional PDF should be used. When $Y = \frac{1}{2}$, the conditional PDF in (1) is reduced to

$$f_{X|Y=\frac{1}{2}}(x|y=\frac{1}{2}) = \begin{cases} \frac{\sqrt{3}}{3}, & -\frac{\sqrt{3}}{2} \leq x \leq \frac{\sqrt{3}}{2} \\ 0, & \text{otherwise} \end{cases}.$$

Therefore,

$$P(X > \frac{1}{2} | Y = \frac{1}{2}) = \int_{\frac{1}{2}}^{+\infty} f_{X|Y=\frac{1}{2}}(x|y=\frac{1}{2}) dx = \int_{\frac{1}{2}}^{\frac{\sqrt{3}}{2}} \frac{\sqrt{3}}{3} dx = \frac{3-\sqrt{3}}{6}.$$

(3) Calculate $P(X > \frac{1}{2} | Y \geq \frac{1}{2})$. Since $P(Y \geq \frac{1}{2}) > 0$, the conditional probability formula can be used.

$$P(X > \frac{1}{2} | Y \geq \frac{1}{2}) = \frac{P(X > \frac{1}{2}, Y \geq \frac{1}{2})}{P(Y \geq \frac{1}{2})}.$$

The denominator is

$$\begin{aligned} P(Y \geq \frac{1}{2}) &= \int_{\frac{1}{2}}^{+\infty} f_Y(y) dy \\ &= \int_{\frac{1}{2}}^1 \frac{2\sqrt{1-y^2}}{\pi} dy \quad (\text{let } y = \sin \theta) \\ &= \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \frac{2\cos \theta}{\pi} \cos \theta d\theta \\ &= \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \frac{1 + \cos 2\theta}{\pi} d\theta \\ &= \frac{1}{3} + \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \frac{\cos 2\theta}{\pi} d\theta \\ &= \frac{1}{3} + \left[\frac{\sin 2\theta}{2\pi} \right]_{\theta=\frac{\pi}{6}}^{\theta=\frac{\pi}{2}} \\ &= \frac{1}{3} - \frac{\sqrt{3}}{4\pi}. \end{aligned}$$

The numerator is (see Figure 3. 15)

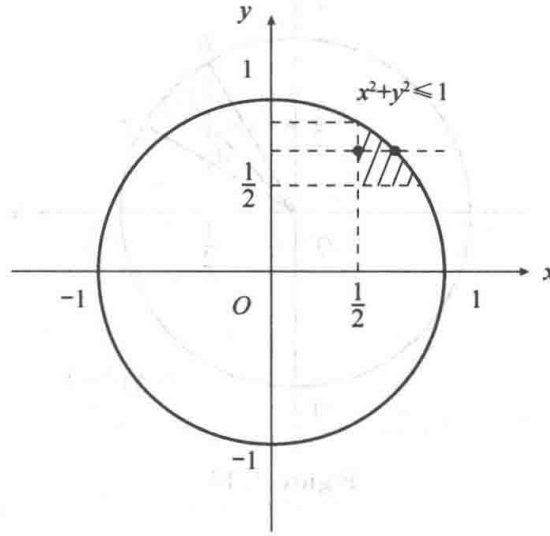


Figure 3. 15

$$\begin{aligned}
 P\left(X > \frac{1}{2}, Y \geq \frac{1}{2}\right) &= \int_{\frac{1}{2}}^{+\infty} \int_{\frac{1}{2}}^{+\infty} f(x, y) dx dy \\
 &= \iint_{\text{shaded region}} \frac{1}{\pi} dx dy \\
 &= \frac{1}{\pi} \int_{\frac{1}{2}}^{\frac{\sqrt{3}}{2}} \left(\int_{\frac{1}{2}}^{\sqrt{1-y^2}} dx \right) dy \\
 &= \frac{1}{\pi} \int_{\frac{1}{2}}^{\frac{\sqrt{3}}{2}} \left(\sqrt{1-y^2} - \frac{1}{2} \right) dy \\
 &= \frac{1}{\pi} \left(\int_{\frac{1}{2}}^{\frac{\sqrt{3}}{2}} \sqrt{1-y^2} dy - \frac{\sqrt{3}-1}{4} \right) (\text{let } y = \sin \theta) \\
 &= \frac{1}{\pi} \left(\int_{\frac{\pi}{6}}^{\frac{\pi}{3}} \cos^2 \theta d\theta - \frac{\sqrt{3}-1}{4} \right) \\
 &= \frac{1}{\pi} \left(\int_{\frac{\pi}{6}}^{\frac{\pi}{3}} \frac{1 + \cos 2\theta}{2} d\theta - \frac{\sqrt{3}-1}{4} \right) \\
 &= \frac{1}{\pi} \left[\frac{\pi}{12} + \left(\frac{\sin 2\theta}{4} \right)_{\theta=\frac{\pi}{6}}^{\theta=\frac{\pi}{3}} - \frac{\sqrt{3}-1}{4} \right] \\
 &= \frac{1}{\pi} \left(\frac{\pi}{12} - \frac{\sqrt{3}-1}{4} \right).
 \end{aligned}$$

In addition, the numerator can also be evaluated using an elementary method. See Figure 3. 16.

$$\begin{aligned}
 P\left(X > \frac{1}{2}, Y \geq \frac{1}{2}\right) &= \int_{\frac{1}{2}}^{+\infty} \int_{\frac{1}{2}}^{+\infty} f(x, y) dx dy \\
 &= \iint_{BAC} \frac{1}{\pi} dx dy = \frac{1}{\pi} S_{BAC},
 \end{aligned}$$

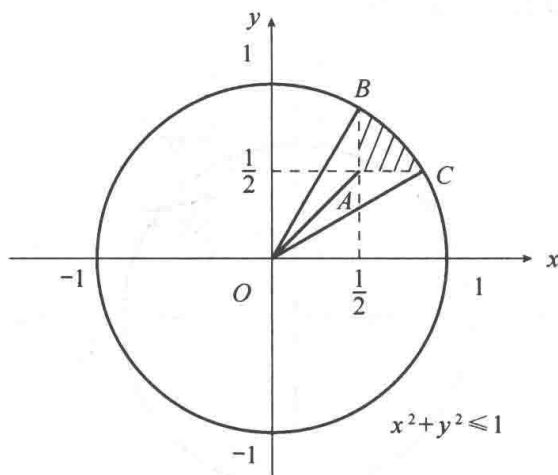


Figure 3.16

where S_{BAC} equals the area difference of the sector OBC and two triangles OAB and OAC .

Since $BOy = COx = \arcsin \frac{1}{2} = \frac{\pi}{6}$, the sector OBC occupies $\frac{1}{3}$ of the circle in the first quadrant.

$$S_{\text{sector } OBC} = \pi \times \frac{1}{4} \times \frac{1}{3} = \frac{\pi}{12}.$$

For the triangle OAB , its base $|AB| = \frac{\sqrt{3}}{2} - \frac{1}{2} = \frac{\sqrt{3}-1}{2}$, its height is $\frac{1}{2}$, so its area equals

$$S_{\triangle OAB} = \frac{1}{2} \times \frac{\sqrt{3}-1}{2} \times \frac{1}{2} = \frac{\sqrt{3}-1}{8}.$$

Similarly, $S_{\triangle OAC} = \frac{\sqrt{3}-1}{8}$. Therefore

$$S_{BAC} = \frac{\pi}{12} - \frac{\sqrt{3}-1}{8} - \frac{\sqrt{3}-1}{8} = \frac{\pi}{12} - \frac{\sqrt{3}-1}{4}.$$

$$P\left(X > \frac{1}{2}, Y \geq \frac{1}{2}\right) = \frac{1}{\pi} S_{BAC} = \frac{1}{\pi} \left(\frac{\pi}{12} - \frac{\sqrt{3}-1}{4} \right).$$

Therefore,

$$P\left(X > \frac{1}{2} \mid Y \geq \frac{1}{2}\right) = \frac{P\left(X > \frac{1}{2}, Y \geq \frac{1}{2}\right)}{P\left(Y \geq \frac{1}{2}\right)} = \frac{\frac{1}{\pi} \left(\frac{\pi}{12} - \frac{\sqrt{3}-1}{4} \right)}{\frac{1}{3} - \frac{\sqrt{3}}{4\pi}} = \frac{\pi+3-3\sqrt{3}}{4\pi-3\sqrt{3}}.$$

Example 3.3.5 Suppose X and Y have the bivariate normal distribution $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Determine its conditional PDFs $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

Solution: The joint PDF of (X, Y) is of the form

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}, -\infty < x, y < +\infty,$$

and the marginal PDF of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, -\infty < y < +\infty.$$

For any y , $f_Y(y) > 0$,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2\right\}, \end{aligned}$$

It shows that $X|Y=y \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2))$.

In the same way, $Y|X=x \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2))$.

Remark 3.3.1 The joint PDF can determine its two marginal PDFs and conditional PDFs. Conversely, two marginal PDFs do not determine the joint PDF, neither the two conditional PDFs do. However, one conditional PDF and its corresponding marginal PDF can determine the joint PDF since $f(x,y) = f_X(x)f_{Y|X}(y|x)$.

◆ 3.4 Independent Random Variables ◆

The independence of two random variables comes from the independence of two events. Suppose X and Y are two random variables, $\{X \leq x\}$ and $\{Y \leq y\}$ are two events, the two events are independent, that is,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

or it can be rewritten as

$$F(x,y) = F_X(x)F_Y(y).$$

Definition 3.4.1 Suppose $F(x,y)$, $F_X(x)$ and $F_Y(y)$ are the joint CDF and marginal CDFs of X and Y . If for all real numbers x and y , $F(x,y) = F_X(x)F_Y(y)$, then it is said that X and Y are independent.

Specially, we can use PF and PDF to describe the independence for discrete and continuous random variables respectively.

If X and Y are discrete random variables, X and Y are independent if for all possible values (x_i, y_j) ,

$$P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j).$$

If X and Y are continuous random variables, X and Y are independent if

$$f(x,y) = f_X(x)f_Y(y)$$

holds almost everywhere in the xy -plane.

Example 3.4.1 The joint PF of (X,Y) is given in Table 3.9. Determine X and Y are independent or not.

Table 3.9

	$y=1$	$y=2$	$y=3$	$y=4$	Total
$x=1$	0.1	0	0.1	0	0.2
$x=2$	0.3	0	0.1	0.2	0.6
$x=3$	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1

Solution: Since

$$0 = P(X=1, Y=2) \neq P(X=1)P(Y=2) = 0.2 \times 0.2.$$

Thus, X and Y are not independent.

Example 3.4.2 The joint PF of (X, Y) is given in Table 3.10. Verify that X and Y are independent.

Table 3.10

	$x=1$	$x=2$	$x=3$	Total
$y=1$	0.16	0.16	0.08	0.4
$y=2$	0.24	0.24	0.12	0.6
Total	0.4	0.4	0.2	1

Solution: Since $0.16 = 0.4 \times 0.4$, $0.16 = 0.4 \times 0.4$, $0.08 = 0.2 \times 0.4$, $0.24 = 0.4 \times 0.6$, $0.24 = 0.4 \times 0.6$, $0.12 = 0.2 \times 0.6$, $P(X=x, Y=y) = P(X=x)P(Y=y)$, X and Y are independent.

Example 3.4.3 Suppose that the joint PDF of X and Y is specified as $f(x, y) = \begin{cases} \frac{21}{4}x^2y, & x^2 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$. Determine X and Y are independent or not.

Solution: The support set S of points (x, y) for which $f(x, y) > 0$ is sketched in Figure 3.17.

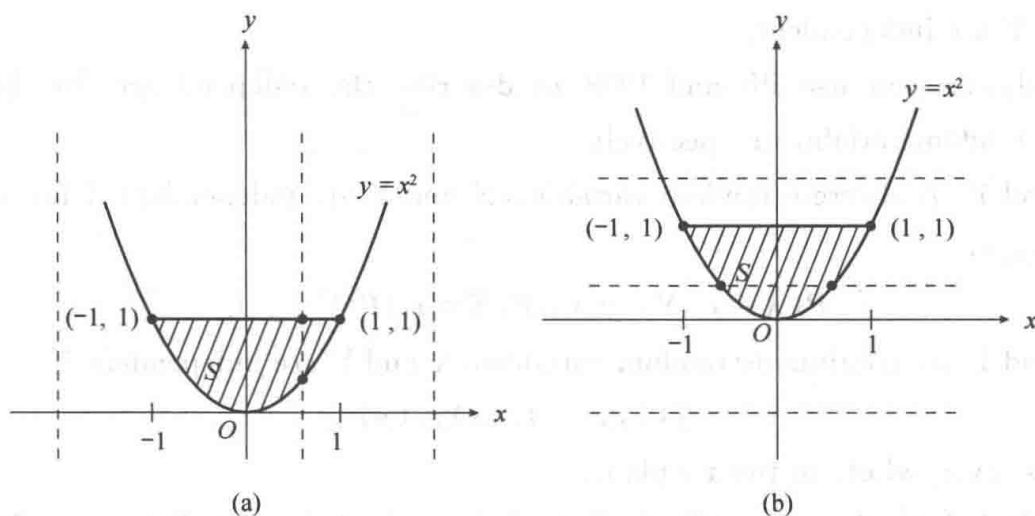


Figure 3.17

(1) We determine first the marginal PDF $f_X(x)$ of X .

① If $x < -1$ or $x > 1$, $f_X(x) = 0$.

② If $-1 \leq x \leq 1$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \int_{x^2}^1 \frac{21}{4} x^2 y dy \\ &= \frac{21}{8} x^2 (1 - x^4). \end{aligned}$$

In summary, $f_X(x) = \begin{cases} \frac{21}{8} x^2 (1 - x^4), & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$.

(2) We determine the marginal PDF $f_Y(y)$ of Y .

① If $y < 0$ or $y > 1$, $f_Y(y) = 0$.

② If $0 \leq y \leq 1$,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y dx \\ &= \frac{7}{2} y^{\frac{5}{2}}. \end{aligned}$$

In summary, $f_Y(y) = \begin{cases} \frac{7}{2} y^{\frac{5}{2}}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$.

Over the support set S , $\frac{21}{4} x^2 y \neq \frac{21}{8} x^2 (1 - x^4) \times \frac{7}{2} y^{\frac{5}{2}}$, $f(x, y) \neq f_X(x) f_Y(y)$. X and Y

are not independent.

Example 3.4.4 Suppose that the joint PDF $f(x, y)$ of X and Y is as follows: $f(x, y) =$

$$\begin{cases} 2e^{-x-2y}, & x \geq 0, y \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Determine X and Y are independent or not.

Solution: (1) Find the marginal PDF $f_X(x)$ of X . See Figure 3.18(a).

① If $x < 0$, $f_X(x) = 0$.

② If $x \geq 0$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \int_0^{+\infty} 2e^{-x-2y} dy \\ &= e^{-x}. \end{aligned}$$

In summary, $f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$.

(2) Find the marginal PDF $f_Y(y)$ of Y . See Figure 3.18(b).

① If $y < 0$, $f_Y(y) = 0$.

② If $y \geq 0$,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \\ &= \int_0^{+\infty} 2e^{-x-2y} dx \\ &= 2e^{-2y}. \end{aligned}$$

In summary, $f_Y(y) = \begin{cases} 2e^{-2y}, & y \geq 0 \\ 0, & \text{otherwise} \end{cases}$.

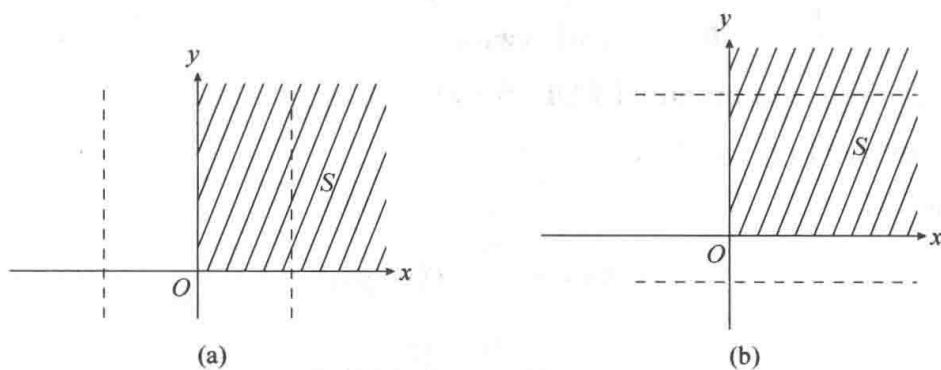


Figure 3.18

(3) $f(x, y) = f_X(x)f_Y(y)$, X and Y are independent.

Example 3.4.5 A man and a woman agree to meet at a certain location. If the man arrives at a time uniformly distributed between 8:00 am and 10:00 am and the woman independently arrives at a time uniformly distributed between 7:00 am and 9:00 am. Find the probability that the time difference between the two arrivals is not larger than 10 minutes.

Solution: Let X be the time the man arrives and Y be the time the woman arrives. The

PDF of X is $f_X(x) = \begin{cases} \frac{1}{2}, & 8 \leq x \leq 10 \\ 0, & \text{otherwise} \end{cases}$. The PDF of Y is $f_Y(y) = \begin{cases} \frac{1}{2}, & 7 \leq y \leq 9 \\ 0, & \text{otherwise} \end{cases}$. Since X and

Y are independent, the joint PDF of (X, Y) is

$$f(x, y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{4}, & 8 \leq x \leq 10, 7 \leq y \leq 9 \\ 0, & \text{otherwise} \end{cases}.$$

The probability that the time difference between the two arrivals is not larger than 10 minutes is

$$\begin{aligned} P(|X - Y| \leq \frac{1}{6}) &= \iint_{|x-y| \leq \frac{1}{6}} f(x, y) dx dy \\ &= \frac{1}{4} S_G, \end{aligned}$$

where $S_G = \frac{1}{2} \times \left(\frac{7}{6}\right)^2 - \frac{1}{2} \times \left(\frac{5}{6}\right)^2 = \frac{1}{3}$, see Figure 3. 19. Therefore, $P(|X-Y| \leq \frac{1}{6}) = \frac{1}{12}$.

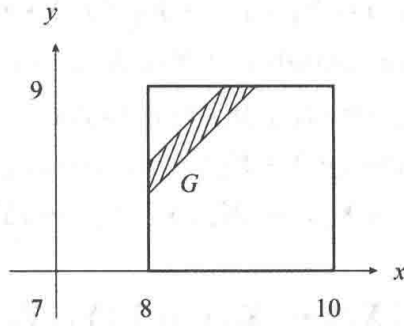


Figure 3. 19

Example 3. 4. 6 Suppose X and Y have the bivariate normal distribution $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Show that X and Y are independent if and only if $\rho=0$.

Solution: The joint PDF of (X, Y) is of the form

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}, -\infty < x, y < +\infty,$$

and the two marginal PDFs are

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, -\infty < x < +\infty,$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, -\infty < y < +\infty.$$

If $\rho=0$, the joint PDF of (X, Y) is

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\} \\ = f_X(x)f_Y(y).$$

It shows that X and Y are independent. On the other hand, given that X and Y are independent, then for any real x and y ,

$$f(x, y) = f_X(x)f_Y(y).$$

Specially, $f(\mu_1, \mu_2) = f_X(\mu_1)f_Y(\mu_2)$, that is,

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \frac{1}{\sqrt{2\pi}\sigma_2}.$$

Therefore, $\rho=0$.

Statistical independence plays an important role in practice. In the case of independence, conditional distribution and marginal distribution are identical. The joint cumulative probability distribution function is the product of the marginal distribution functions. It allows transformation of a multivariate problem into a univariate problem.

The concepts regarding two random variables can be easily extended to several random variables.

Definition 3.4.2 Random variables $X_1, X_2, \dots, X_n$ are independent if for any real $x_1, x_2, \dots, x_n$,

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n).$$

Definition 3.4.3 Random variables $(X_1, X_2, \dots, X_m)$ and $(Y_1, Y_2, \dots, Y_n)$ are independent if for any real $x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n$,

$$F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = F_X(x_1, x_2, \dots, x_m) \cdot F_Y(y_1, y_2, \dots, y_n),$$

where F, F_X, F_Y are CDFs of $(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n), (X_1, X_2, \dots, X_m)$ and $(Y_1, Y_2, \dots, Y_n)$ respectively.

Proposition 3.4.1 If $(X_1, X_2, \dots, X_m)$ and $(Y_1, Y_2, \dots, Y_n)$ are independent, then $X_i (i=1, 2, \dots, m)$ and $Y_j (j=1, 2, \dots, n)$ are independent. In addition, if h, g are two continuous functions, then $h(X_1, X_2, \dots, X_m)$ and $g(Y_1, Y_2, \dots, Y_n)$ are also independent.

◆ 3.5 Functions of Two or More Random Variables ◆

In Chapter 2, we have discussed the functions of a single random variable. Now we want to study the functions of two or more random variables.

Example 3.5.1 The joint PF of X and Y is as specified in Table 3.11. Find the PF of $Z=X+Y$.

Table 3.11

	$y=1$	$y=2$	$y=3$	$y=4$
$x=1$	0.1	0	0.1	0
$x=2$	0.3	0	0.1	0.2
$x=3$	0	0.2	0	0

Solution: $Z=X+Y$ can take 2, 3, ..., 7.

$$P(Z=2) = P(X+Y=2) = P(X=1, Y=1) = 0.1$$

$$\begin{aligned} P(Z=3) &= P(X+Y=3) \\ &= P(X=1, Y=2) + P(X=2, Y=1) \\ &= 0.3 \end{aligned}$$

⋮

The PF of $Z=X+Y$ is given in Table 3.12.

Table 3.12

Z	2	3	4	5	6	7
$P(Z=z)$	0.1	0.3	0.1	0.3	0.2	0

Example 3.5.2 Suppose X and Y are independent, the PFs are $P(X=k)=p(k), k=0,1,2,\dots, P(Y=r)=q(r), r=0,1,2,\dots$. Determine the PF of $Z=X+Y$.

Solution: The PF of Z is

$$\begin{aligned} P(Z=i) &= P(X+Y=i) \\ &= P\left(\bigcup_{k=0}^i \{X=k, Y=i-k\}\right) \\ &= \sum_{k=0}^i P(X=k, Y=i-k) \\ &= \sum_{k=0}^i P(X=k)P(Y=i-k) \\ &= \sum_{k=0}^i p(k)q(i-k), i=0,1,2,\dots \end{aligned}$$

The above formula is called discrete convolution. Specially, if $X \sim B(m, p)$ and $Y \sim B(n, p)$ are independent, then for $i=0,1,\dots, m+n$,

$$\begin{aligned} P(Z=i) &= \sum_{k=0}^i C_m^k p^k (1-p)^{m-k} C_n^{i-k} p^{i-k} (1-p)^{n-i+k} \\ &= p^i (1-p)^{m+n-i} \sum_{k=0}^i C_m^k C_n^{i-k} \\ &= C_{m+n}^i p^i (1-p)^{m+n-i}. \end{aligned}$$

Therefore, $Z = X + Y \sim B(m+n, p)$. Binomial random variables satisfy the additive properties.

Proposition 3.5.1 Assume that $X_1, \dots, X_n$ are independent identically distributed (abbreviated as IID) random variables having the Bernoulli distribution $B(1, p)$ with the parameter p . Let $Y = X_1 + X_2 + \dots + X_n$, then Y has the binomial distribution $B(n, p)$.

Suppose the joint PDF of continuous random variables X and Y is $f(x, y)$, we would like to determine the PDF of $Z = g(X, Y)$. The CDF of Z can be found as

$$F_Z(z) = P(Z \leq z) = P[g(X, Y) \leq z] = \iint_{g(x, y) \leq z} f(x, y) dx dy,$$

then the PDF follows by taking the derivative.

Theorem 3.5.1 Let X and Y be two continuous random variables with the joint PDF $f(x, y)$ and let $Z = X + Y$. The distribution of Z is called the convolution of the distributions of X and Y . The PDF of Z is

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx = \int_{-\infty}^{+\infty} f(z-y, y) dy.$$

Proof: The CDF $F_Z(z)$ of Z is

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(X+Y \leq z) \\ &= \iint_{x+y \leq z} f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f(x, y) dx \right] dy. \end{aligned}$$

The plot for $\{(x, y) | x + y \leq z\}$ is shown in Figure 3. 20. Take the derivative with respect to z on both sides,

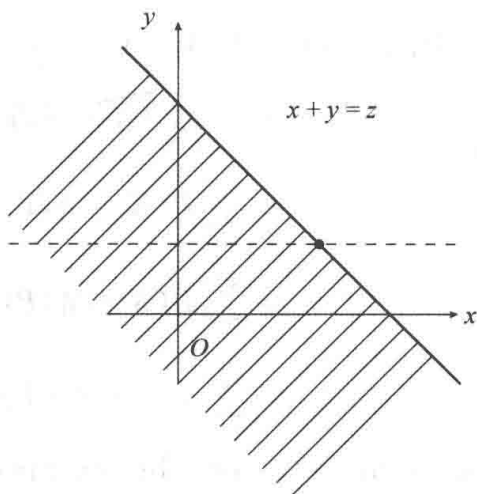


Figure 3. 20

$$\begin{aligned} f_Z(z) &= F_Z'(z) \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f(x, y) dx \right]'_z dy \\ &= \int_{-\infty}^{+\infty} f(z-y, y) dy \end{aligned}$$

Since X and Y are symmetric, the PDF of Z can be rewritten as

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx.$$

Specially, if X and Y are independent, the PDF of Z is

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy.$$

Remark 3. 5. 1 The above formula is called continuous convolution. In the proof, we have used a calculus fact about variable upper limit integral.

$$\frac{d}{dx} \int_a^{h(x)} f(t) dt = f[h(x)] \times h'(x),$$

where a is a constant. More generally,

$$\frac{d}{dx} \int_{g(x)}^{h(x)} f(t) dt = f[h(x)] \times h'(x) - f[g(x)] \times g'(x).$$

Example 3. 5. 3 Suppose that X and Y are independent random variables with common distribution $N(0, 1)$, find the PDF of $Z = X + Y$.

Solution: The PDF of X is $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty < x < +\infty$. The PDF of Y is $f_Y(y) =$

$\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, $-\infty < y < +\infty$. Using the convolution formula, we have

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{2}x - \frac{z}{\sqrt{2}})^2}{2}} dx \quad \left(\text{let } t = \sqrt{2}x - \frac{z}{\sqrt{2}} \right) \\
&= \frac{1}{\sqrt{2\pi} \sqrt{2}} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\
&= \frac{1}{\sqrt{2\pi} \sqrt{2}} e^{-\frac{z^2}{4}}, \quad -\infty < z < +\infty.
\end{aligned}$$

That is, $Z \sim N(0, 2)$.

Remark 3.5.2 In general, if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then $Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. This result can be extended to n independent normal random variables. That is, if $X_i \sim N(\mu_i, \sigma_i^2)$ ($i=1, 2, \dots, n$) are independent, then,

$$Z = X_1 + X_2 + \dots + X_n \sim N(\mu_1 + \mu_2 + \dots + \mu_n, \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2).$$

Normal random variables satisfy the additive property. More generally, if $X_i \sim N(\mu_i, \sigma_i^2)$ ($i=1, 2, \dots, n$) are independent, $c_1, c_2, \dots, c_n$ are constants and not all zeros, then,

$$Z = \sum_{i=1}^n c_i X_i \sim N\left(\sum_{i=1}^n c_i \mu_i, \sum_{i=1}^n c_i^2 \sigma_i^2\right).$$

Definition 3.5.1 Let $X_1, \dots, X_n$ be random variables. The average of these n random variables, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, is called their sample mean.

Theorem 3.5.2 Suppose that the random variables $X_1, \dots, X_n$ form a random sample from the normal distribution $N(\mu, \sigma^2)$, then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Example 3.5.4 Suppose that a random sample of size n is to be taken from the normal distribution $N(\mu, \sigma^2=9)$, determine the minimum value of n for which $P(|\bar{X} - \mu| \leq 1) \geq 0.95$.

$$\text{Solution: } \bar{X} \sim N\left(\mu, \frac{9}{n}\right), Z = \frac{\bar{X} - \mu}{\sqrt{\frac{9}{n}}} = \frac{n^{\frac{1}{2}}}{3} (\bar{X} - \mu) \sim N(0, 1).$$

$$P(|\bar{X} - \mu| \leq 1) = P\left(|Z| \leq \frac{n^{\frac{1}{2}}}{3}\right) \geq 0.95.$$

$$\text{That is, } P\left(Z > \frac{n^{\frac{1}{2}}}{3}\right) = 1 - \Phi\left(\frac{n^{\frac{1}{2}}}{3}\right) \leq 0.025,$$

$$\Phi\left(\frac{n^{\frac{1}{2}}}{3}\right) \geq 0.975,$$

$$\frac{n^{\frac{1}{2}}}{3} \geq \Phi^{-1}(0.975) = 1.96,$$

$$n \geq 34.6.$$

See Figure 3. 21. The sample size must be at least 35.

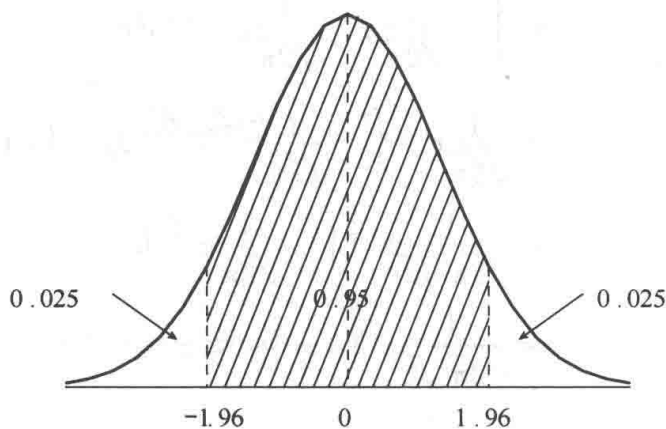


Figure 3. 21

Example 3. 5. 5 Suppose that X and Y are independent exponential random variables with common distribution having the PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the PDF of $Z = X + Y$.

Solution: The joint PDF of (X, Y) is

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Using the convolution formula, we have

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx,$$

where

$$f_X(x) f_Y(z-x) = \begin{cases} \lambda^2 e^{-\lambda z}, & x > 0, z-x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Consider the support: $\{(x, z) | x > 0 \text{ and } z-x > 0\}$. See Figure 3. 22.

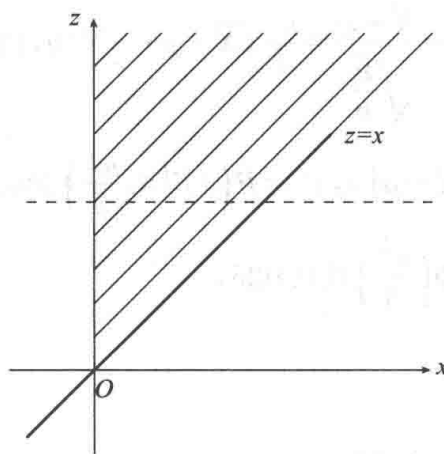


Figure 3. 22

If $z \leq 0, f_Z(z) = 0$.

If $z > 0$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\ &= \int_0^z \lambda^2 e^{-\lambda x} dx \\ &= \lambda^2 z e^{-\lambda z}. \end{aligned}$$

In summary,

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0 \\ 0, & \text{otherwise} \end{cases}$$

This density function does not belong to exponential distribution. The sum $Z = X + Y$ does not follow exponential distribution. In fact, $Z = X + Y$ follows the gamma distribution with parameters $\alpha = 2$ and $\beta = \frac{1}{\lambda}$. In general, if $X_i \sim E(\lambda) (i = 1, 2, \dots, n)$ are independent, then $X_1 + X_2 + \dots + X_n \sim \text{Gamma}\left(n, \frac{1}{\lambda}\right)$.

Theorem 3. 5. 3 Suppose that $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$ are independent random variables, then $Z = X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$.

Solution: The PDF of $X \sim \text{Gamma}(\alpha_1, \beta)$ is given by

$$f_X(x) = \begin{cases} \frac{1}{\beta^{\alpha_1} \Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

The PDF of $Y \sim \text{Gamma}(\alpha_2, \beta)$ is given by

$$f_Y(y) = \begin{cases} \frac{1}{\beta^{\alpha_2} \Gamma(\alpha_2)} y^{\alpha_2-1} e^{-\frac{y}{\beta}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

The joint PDF of (X, Y) is

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1-1} y^{\alpha_2-1} e^{-\frac{x+y}{\beta}}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Using the convolution formula, we have

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx,$$

where

$$f_X(x) f_Y(z-x) = \begin{cases} \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1-1} (z-x)^{\alpha_2-1} e^{-\frac{z}{\beta}}, & x > 0, z-x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Consider the support: $\{(x, z) | x > 0, z-x > 0\}$. See Figure 3. 23.

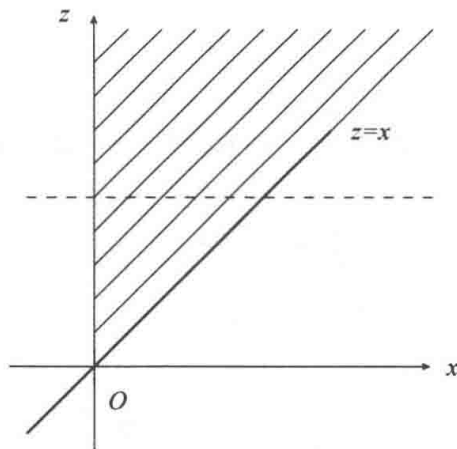


Figure 3.23

If $z \leq 0$, $f_Z(z) = 0$.

If $z > 0$,

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\
 &= \int_0^z \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1-1} (z-x)^{\alpha_2-1} e^{-\frac{z}{\beta}} dx \\
 &= \frac{e^{-\frac{z}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^z x^{\alpha_1-1} (z-x)^{\alpha_2-1} dx \quad (\text{let } t = \frac{x}{z}) \\
 &= \frac{e^{-\frac{z}{\beta}} z^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^1 t^{\alpha_1-1} (1-t)^{\alpha_2-1} dt \\
 &= \frac{e^{-\frac{z}{\beta}} z^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} B(\alpha_1, \alpha_2) \\
 &= \frac{e^{-\frac{z}{\beta}} z^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1 + \alpha_2)},
 \end{aligned}$$

where $B(s, t) = \int_0^1 x^{s-1} (1-x)^{t-1} dx$ is the beta function ($s, t > 0$). In summary,

$$f_Z(z) = \begin{cases} \frac{e^{-\frac{z}{\beta}} z^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1 + \alpha_2)}, & z > 0 \\ 0, & \text{otherwise} \end{cases}$$

$Z = X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$. That is, if the second parameter β is fixed, the gamma distribution satisfies the additive property with respect to the first parameter.

If $\alpha = \frac{n}{2}$, $\beta = 2$, the PDF of $X \sim \text{Gamma}(\alpha, \beta)$ reduces to

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

It is the χ^2 distribution with degrees of freedom n , an important distribution in mathematical statistics. χ^2 distribution satisfies the additive property. If $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ are independent, then $X + Y \sim \chi^2(m+n)$.

Theorem 3.5.4 Let X and Y be two continuous random variables with the joint PDF $f(x, y)$ and let $Z = X - Y$, the PDF of Z is

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z + y, y) dy.$$

Theorem 3.5.5 Suppose that $X_1, \dots, X_n$ form a random sample of size n from a distribution for which the PDF is $f(x)$ and the CDF is $F(x)$. The largest value Y_n and the smallest value Y_1 in the random sample are defined as follows:

$$Y_n = \max\{X_1, \dots, X_n\}, \quad Y_1 = \min\{X_1, \dots, X_n\}.$$

Then the PDF of Y_n is

$$g_n(y) = n[F(y)]^{n-1} f(y), \quad -\infty < y < +\infty,$$

the PDF of Y_1 is

$$g_1(y) = n[1 - F(y)]^{n-1} f(y), \quad -\infty < y < +\infty.$$

Proof: Consider Y_n first. Let G_n stand for its CDF, and let g_n be its PDF,

$$\begin{aligned} G_n(y) &= P(Y_n \leq y) = P(X_1 \leq y, \dots, X_n \leq y) \\ &= P(X_1 \leq y) \cdot \dots \cdot P(X_n \leq y) \\ &= [F(y)]^n, \end{aligned}$$

$$g_n(y) = \frac{dG_n(y)}{dy} = n[F(y)]^{n-1} f(y), \quad -\infty < y < +\infty.$$

Consider Y_1 with the CDF G_1 and the PDF g_1 ,

$$\begin{aligned} G_1(y) &= P(Y_1 \leq y) = 1 - P(Y_1 > y) \\ &= 1 - P(X_1 > y, \dots, X_n > y) \\ &= 1 - P(X_1 > y) \cdot \dots \cdot P(X_n > y) \\ &= 1 - [1 - F(y)]^n, \end{aligned}$$

$$g_1(y) = \frac{dG_1(y)}{dy} = n[1 - F(y)]^{n-1} f(y), \quad -\infty < y < +\infty.$$

Example 3.5.6 If $X_i \sim E(\lambda)$ ($i = 1, 2, \dots, n$) are independent, find the distribution of $Z = \min\{X_1, X_2, \dots, X_n\}$.

Solution: The PDF and CDF of $X \sim E(\lambda)$ are

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \quad F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

The CDF of Z is

$$\begin{aligned} F_Z(x) &= P(\min\{X_1, \dots, X_n\} \leq x) \\ &= 1 - P(\min\{X_1, \dots, X_n\} > x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - P(X_1 > x) \cdot \dots \cdot P(X_n > x) \\ &= 1 - [1 - P(X_1 \leq x)] \cdot \dots \cdot [1 - P(X_n \leq x)] \\ &= 1 - [1 - F(x)]^n \\ &= \begin{cases} 1 - e^{-n\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}. \end{aligned}$$

The PDF of Z is $f_Z(x) = F'_Z(x) = \begin{cases} n\lambda e^{-n\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$. That is, Z also follows the exponential distribution, $Z = \min\{X_1, \dots, X_n\} \sim E(n\lambda)$.

Definition 3.5.2 Suppose that $X_1, \dots, X_n$ form a random sample of size n from a distribution X for which the PDF is $f(x)$ and the CDF is $F(x)$. The ordered sample values

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

are called the order statistics. Since $F(x)$ is continuous, then with probability 1 the order statistics of the sample take distinct values (and conversely).

Theorem 3.5.6 Suppose that $X_1, \dots, X_n$ form a random sample of size n from a distribution X for which the PDF is $f(x)$ and the CDF is $F(x)$. The PDF of the i -th order statistic $X_{(i)}$ is given by

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x),$$

where $i=1, 2, \dots, n$.

Proof:

Method 1: First we consider the event that $X_{(i)}$ falls in the interval $(x, x+\epsilon]$,

$$P(x < X_{(i)} \leq x+\epsilon) = F_{X_{(i)}}(x+\epsilon) - F_{X_{(i)}}(x),$$

where $F_{X_{(i)}}(x)$ is the CDF of $X_{(i)}$. On the other hand, this event is equivalent to the following event: there are $i-1$ out of $X_1, \dots, X_n$ are less than or equal to x ; 1 out of $X_1, \dots, X_n$ are between x and $x+\epsilon$; the others are larger than $x+\epsilon$. See Figure 3.24.

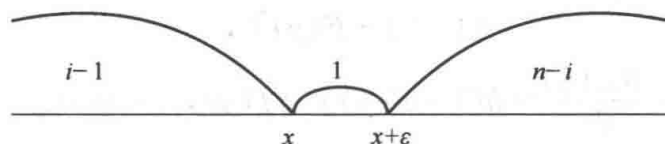


Figure 3.24

$$\begin{aligned} P(x < X_{(i)} \leq x+\epsilon) &= C_n^{i-1} C_{n-i+1}^1 C_{n-i}^{n-i} [P(X \leq x)]^{i-1} P(x < X \leq x+\epsilon) [P(X > x+\epsilon)]^{n-i} \\ &\approx \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x+\epsilon)]^{n-i} f(x)\epsilon. \end{aligned}$$

Thus,

$$\frac{F_{X_{(i)}}(x+\epsilon) - F_{X_{(i)}}(x)}{\epsilon} \approx \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x).$$

Let $\epsilon \rightarrow 0$, we have

$$f_{X_{(i)}}(x) = F'_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x).$$

Method 2: The event $\{X_{(i)} \leq x\}$ occurs if and only if at least i out of $X_1, \dots, X_n$ are less than or equal to x . The CDF of $X_{(i)}$ is

$$F_{X_{(i)}}(x) = P(X_{(i)} \leq x)$$

$$\begin{aligned}
 &= \sum_{k=i}^n C_n^k [P(X \leq x)]^k [P(X > x)]^{n-k} \\
 &= \sum_{k=i}^n C_n^k [F(x)]^k [1 - F(x)]^{n-k}.
 \end{aligned}$$

Therefore, the PDF of $X_{(i)}$ is

$$\begin{aligned}
 f_{X_{(i)}}(x) &= F'_{X_{(i)}}(x) \\
 &= \frac{d\left\{ \sum_{k=i}^n C_n^k [F(x)]^k [1 - F(x)]^{n-k} \right\}}{dx} \\
 &= \sum_{k=i}^n C_n^k \{ k[F(x)]^{k-1} [1 - F(x)]^{n-k} f(x) - (n-k)[F(x)]^k [1 - F(x)]^{n-k-1} f(x) \} \\
 &= \sum_{k=i}^n k C_n^k [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x) - \sum_{k=i}^n (n-k) C_n^k [F(x)]^k [1 - F(x)]^{n-k-1} f(x) \\
 &= \sum_{k=i}^n k C_n^k [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x) - \sum_{k=i}^{n-1} (n-k) C_n^k [F(x)]^k [1 - F(x)]^{n-k-1} f(x) \\
 &= \sum_{k=i}^n n C_{n-1}^{k-1} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x) - \sum_{k=i}^{n-1} n C_{n-1}^k [F(x)]^k [1 - F(x)]^{n-k-1} f(x) \\
 &= \sum_{i=i-1}^{n-1} n C_{n-1}^i [F(x)]^i [1 - F(x)]^{n-i-1} f(x) - \sum_{k=i}^{n-1} n C_{n-1}^k [F(x)]^k [1 - F(x)]^{n-k-1} f(x) \\
 &= n C_{n-1}^{i-1} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x) \\
 &= \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x).
 \end{aligned}$$

Note: let $i=n$, we obtain the PDF of maximum $X_{(n)}$. Let $i=1$, we obtain the PDF of minimum $X_{(1)}$. We also discussed these two statistics in Theorem 3.5.5.

Example 3.5.7 Suppose that $X_1, \dots, X_n$ form a random sample of size n from a distribution $X \sim U[0, 1]$. Determine the PDF of $X_{(i)}$.

Solution: $X \sim U[0, 1]$. Its PDF and CDF are

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}, \quad F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}.$$

The PDF of $X_{(i)}$ is

$$f_{X_{(i)}} = \begin{cases} \frac{n!}{(i-1)!(n-i)!} x^{i-1} [1-x]^{n-i}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

This is an example of the beta distribution $X_{(i)} \sim \text{Beta}(i, n-i+1)$.

Theorem 3.5.7 Let X_1, X_2 have a continuous joint distribution for which the joint PDF is $f(x_1, x_2)$. Assume that there is a subset S of $\mathbb{R}^2$ such that $P[(X_1, X_2) \in S] = 1$. Define two new random variables Y_1, Y_2 as follows:

$$Y_1 = r_1(X_1, X_2),$$

$$Y_2 = r_2(X_1, X_2),$$

where we assume that the functions r_1, r_2 define a one-to-one differentiable transformation of S onto a subset T of $\mathbb{R}^2$. Let the inverse of this transformation be given as follows:

$$X_1 = s_1(Y_1, Y_2),$$

$$X_2 = s_2(Y_1, Y_2).$$

Then the joint PDF $g(y_1, y_2)$ of (Y_1, Y_2) is

$$g(y_1, y_2) = \begin{cases} f(s_1, s_2) \times |J|, & (y_1, y_2) \in T \\ 0, & \text{otherwise} \end{cases}.$$

where J is the determinant

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \frac{\partial s_1}{\partial y_2} \\ \frac{\partial s_2}{\partial y_1} & \frac{\partial s_2}{\partial y_2} \end{bmatrix}$$

Example 3.5.8 Suppose that two random variables X_1 and X_2 have a continuous joint distribution for which the joint PDF is as follows:

$$f(x_1, x_2) = \begin{cases} 4x_1x_2, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the joint PDF of two new random variables $Y_1 = \frac{X_1}{X_2}$ and $Y_2 = X_1X_2$ and the marginal PDFs of Y_1 and Y_2 .

Solution:

$$\begin{cases} Y_1 = r_1(X_1, X_2) = \frac{X_1}{X_2} \\ Y_2 = r_2(X_1, X_2) = X_1X_2 \end{cases}.$$

$$S = \{(x_1, x_2) \mid 0 < x_1 < 1, 0 < x_2 < 1\}.$$

r_1, r_2 define a one-to-one differentiable transformation of S onto a subset T of $\mathbb{R}_2$. See Figure 3.25. The inverse functions can be found as

$$\begin{cases} X_1 = s_1(Y_1, Y_2) = \sqrt{Y_1Y_2} \\ X_2 = s_2(Y_1, Y_2) = \sqrt{\frac{Y_2}{Y_1}} \end{cases}.$$

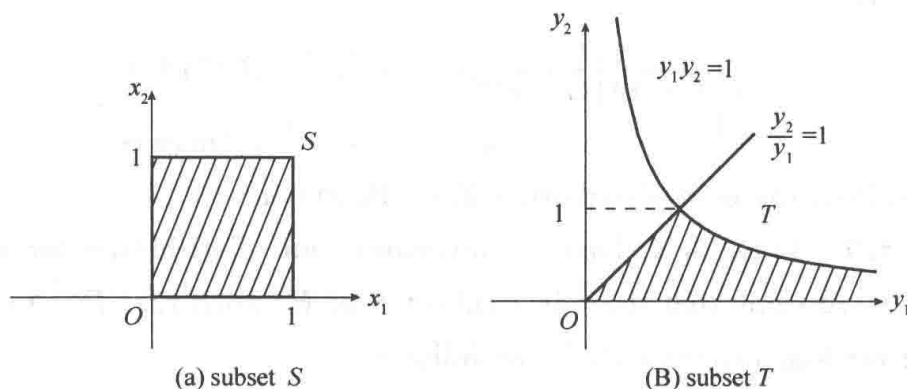


Figure 3.25

Therefore, $Y_1 > 0, Y_2 > 0, 0 < \sqrt{Y_1 Y_2} < 1, 0 < \sqrt{\frac{Y_2}{Y_1}} < 1$. and

$$T = \left\{ (y_1, y_2) \mid y_1 > 0, y_2 > 0, 0 < y_1 y_2 < 1, 0 < \frac{y_2}{y_1} < 1 \right\}.$$

$$J = \begin{vmatrix} \frac{\partial s_1}{\partial y_1} & \frac{\partial s_1}{\partial y_2} \\ \frac{\partial s_2}{\partial y_1} & \frac{\partial s_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2\sqrt{\frac{y_2}{y_1}}} & \frac{1}{2\sqrt{\frac{y_1}{y_2}}} \\ -\frac{1}{2\sqrt{\frac{y_2}{y_1^3}}} & \frac{1}{2\sqrt{\frac{1}{y_1 y_2}}} \end{vmatrix} = \frac{1}{2y_1}, |J| = \frac{1}{2y_1}.$$

Therefore,

$$g(y_1, y_2) = \begin{cases} f(s_1, s_2) \times |J|, & (y_1, y_2) \in T \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} 4\sqrt{y_1 y_2} \times \sqrt{\frac{y_2}{y_1}} \times \frac{1}{2y_1}, & (y_1, y_2) \in T \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{2y_2}{y_1}, & (y_1, y_2) \in T \\ 0, & \text{otherwise} \end{cases}.$$

Find the marginal PDF $g_1(y_1)$ of $Y_1 = \frac{X_1}{X_2}$. See Figure 3. 26.

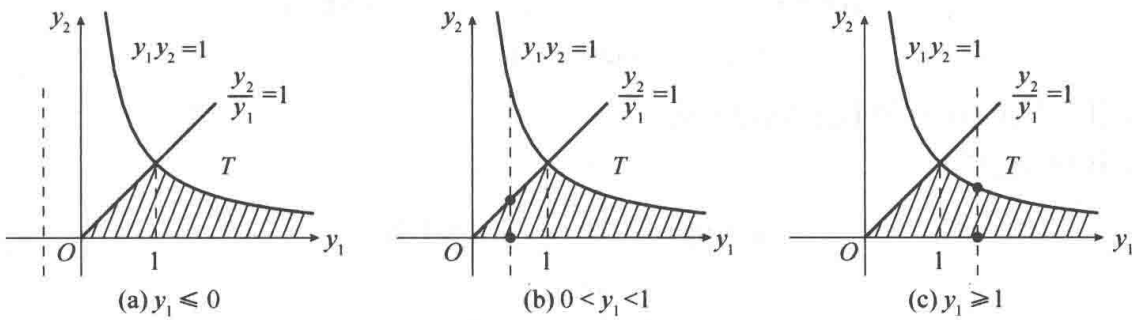


Figure 3. 26

(1) If $y_1 \leq 0, g_1(y_1) = 0$.

(2) If $0 < y_1 < 1$,

$$g_1(y_1) = \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2$$

$$= \int_0^{y_1} \frac{2y_2}{y_1} dy_2$$

$$= \frac{1}{y_1} [y_2^2]_{y_2=0}^{y_2=y_1}$$

$$= y_1.$$

(3) If $y_1 \geq 1$,

$$\begin{aligned}
 g_1(y_1) &= \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2 \\
 &= \int_0^{\frac{1}{y_1}} \frac{2y_2}{y_1} dy_2 \\
 &= \frac{1}{y_1} [y_2^2]_{y_2=0}^{y_2=\frac{1}{y_1}} \\
 &= \frac{1}{y_1^3}.
 \end{aligned}$$

In summary,

$$g_1(y_1) = \begin{cases} 0, & y_1 \leq 0 \\ y_1, & 0 < y_1 < 1 \\ \frac{1}{y_1^3}, & y_1 \geq 1 \end{cases}.$$

Find the marginal PDF $g_2(y_2)$ of $Y_2 = X_1 X_2$. See Figure 3. 27

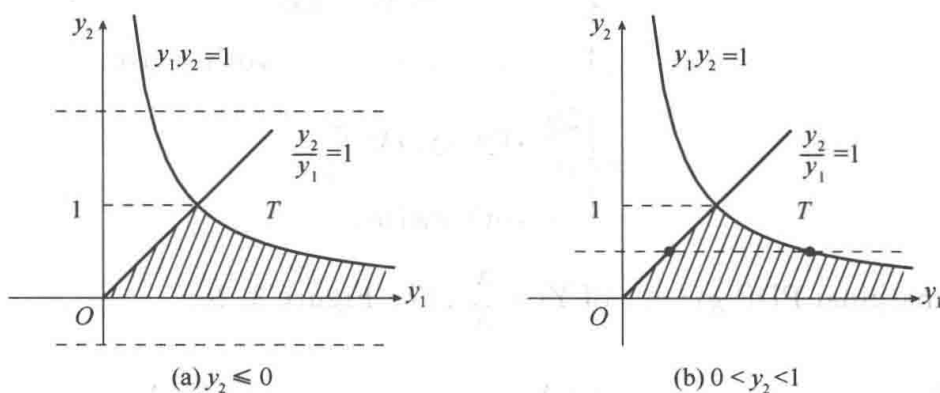


Figure 3. 27

- (1) If $y_2 \leq 0$ or $y_2 \geq 1$, $g_2(y_2) = 0$.
- (2) If $0 < y_2 < 1$,

$$\begin{aligned}
 g_2(y_2) &= \int_{-\infty}^{+\infty} g(y_1, y_2) dy_1 \\
 &= \int_{y_2}^{\frac{1}{y_2}} \frac{2y_2}{y_1} dy_1 \\
 &= 2y_2 [\ln y_1]_{y_1=y_2}^{y_1=\frac{1}{y_2}} \\
 &= -4y_2 \ln y_2.
 \end{aligned}$$

In summary,

$$g_2(y_2) = \begin{cases} 0, & y_2 \leq 0 \text{ or } y_2 \geq 1 \\ -4y_2 \ln y_2, & 0 < y_2 < 1 \end{cases}.$$

Example 3. 5. 9 Assume $X, Y \sim N(0, 1)$ and they are independent. Let $U = X + Y, V = X - Y$. Find the joint and marginal distributions of (U, V) .

Solution: The joint PDF of (X, Y) is

$$f(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi} e^{-\frac{x^2 + y^2}{2}}, x, y \in \mathbb{R}.$$

$$\begin{cases} U = r_1(X, Y) = X + Y \\ V = r_2(X, Y) = X - Y \end{cases}$$

r_1, r_2 define a one-to-one differentiable transformation of $S = \mathbb{R}^2$ onto $T = \mathbb{R}^2$. The inverse functions can be found as

$$\begin{cases} X = s_1(U, V) = \frac{U+V}{2} \\ Y = s_2(U, V) = \frac{U-V}{2} \end{cases}$$

$$J = \begin{vmatrix} \frac{\partial s_1}{\partial u} & \frac{\partial s_1}{\partial v} \\ \frac{\partial s_2}{\partial u} & \frac{\partial s_2}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}, |J| = \frac{1}{2}.$$

Therefore, the joint PDF of (U, V) is

$$g(u, v) = f(s_1, s_2) \times |J| = \frac{1}{4\pi} e^{-\frac{u^2+v^2}{4}}, u, v \in \mathbb{R}.$$

$(U, V) \sim N(0, 0, 2, 2, 0), U \sim N(0, 2), V \sim N(0, 2).$

Example 3.5.10 Assume $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$, and they are independent. Let $U = X + Y, V = \frac{X}{X+Y}$. Find the joint and marginal distributions of (U, V) .

Solution: The PDF of $X \sim \text{Gamma}(\alpha_1, \beta)$ is given by

$$f_X(x) = \begin{cases} \frac{1}{\beta^{\alpha_1} \Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

The PDF of $Y \sim \text{Gamma}(\alpha_2, \beta)$ is given by

$$f_Y(y) = \begin{cases} \frac{1}{\beta^{\alpha_2} \Gamma(\alpha_2)} y^{\alpha_2-1} e^{-\frac{y}{\beta}}, & y > 0 \\ 0, & y \leq 0 \end{cases}.$$

The joint PDF of (X, Y) is

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1-1} y^{\alpha_2-1} e^{-\frac{x+y}{\beta}}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}.$$

$$\begin{cases} U = r_1(X, Y) = X + Y \\ V = r_2(X, Y) = \frac{X}{X+Y} \end{cases}.$$

$$S = \{(x, y) \mid x > 0, y > 0\}.$$

r_1, r_2 define a one-to-one differentiable transformation of S onto a subset T of $\mathbb{R}^2$. See Figure 3.28. The inverse functions can be found as

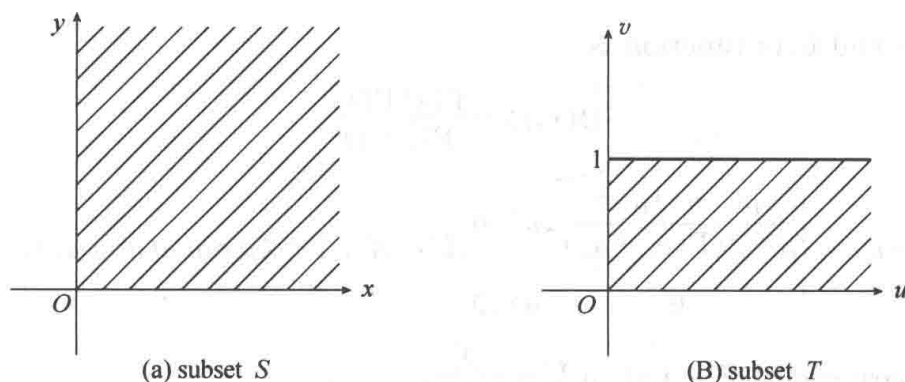


Figure 3.28

$$\begin{cases} X=s_1(U,V)=UV \\ Y=s_2(U,V)=U(1-V) \end{cases}$$

Therefore, $UV>0, U(1-V)>0$, and $T=\{(u,v)|u>0, 1>v>0\}$.

$$J=\begin{vmatrix} \frac{\partial s_1}{\partial u} & \frac{\partial s_1}{\partial v} \\ \frac{\partial s_2}{\partial u} & \frac{\partial s_2}{\partial v} \end{vmatrix}=\begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix}=-u, |J|=u.$$

Therefore, the joint PDF of (U,V) is

$$\begin{aligned} g(u,v) &= \begin{cases} f(s_1, s_2) \times |J| & , (u,v) \in T \\ 0 & , \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} (uv)^{\alpha_1-1} [u(1-v)]^{\alpha_2-1} e^{-\frac{u}{\beta}} \times u, & (u,v) \in T \\ 0 & , \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} v^{\alpha_1-1} (1-v)^{\alpha_2-1}, & (u,v) \in T \\ 0 & , \text{otherwise} \end{cases} \end{aligned}$$

Find the marginal PDF $g_1(u)$ of $U=X+Y$.

(1) If $u \leq 0$, $g_1(u)=0$.

(2) If $u > 0$,

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{+\infty} g(u,v) dv \\ &= \int_0^1 \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} v^{\alpha_1-1} (1-v)^{\alpha_2-1} dv \\ &= \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^1 v^{\alpha_1-1} (1-v)^{\alpha_2-1} dv \\ &= \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} B(\alpha_1, \alpha_2) \\ &= \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1 + \alpha_2)}, \end{aligned}$$

where $B(s,t)=\int_0^1 x^{s-1} (1-x)^{t-1} dx$ is the beta function ($s,t>0$). The relationship between gamma function and beta function is

$$B(s,t)=\frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}.$$

In summary, $g_1(u)=\begin{cases} \frac{u^{\alpha_1+\alpha_2-1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1 + \alpha_2)}, & u>0 \\ 0 & , u\leq 0 \end{cases}$. $U=X+Y \sim \text{Gamma}(\alpha_1 + \alpha_2)$.

Find the marginal PDF $g_2(v)$ of $V=\frac{X}{X+Y}$.

(1) If $v \leq 0$ or $v \geq 1$, $g_2(v) = 0$.

(2) If $0 < v < 1$,

$$\begin{aligned} g_2(v) &= \int_{-\infty}^{+\infty} g(u, v) du \\ &= \int_0^{+\infty} \frac{u^{\alpha_1 + \alpha_2 - 1} e^{-\frac{u}{\beta}}}{\beta^{\alpha_1 + \alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} v^{\alpha_1 - 1} (1 - v)^{\alpha_2 - 1} du \\ &= \frac{v^{\alpha_1 - 1} (1 - v)^{\alpha_2 - 1}}{\beta^{\alpha_1 + \alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} \int_0^{+\infty} u^{\alpha_1 + \alpha_2 - 1} e^{-\frac{u}{\beta}} du \\ &= \frac{v^{\alpha_1 - 1} (1 - v)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} \end{aligned}$$

In summary, $g_2(v) = \begin{cases} \frac{v^{\alpha_1 - 1} (1 - v)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)}, & 0 < v < 1 \\ 0, & \text{otherwise} \end{cases}$. $V = \frac{X}{X + Y}$ follows beta distribution with

parameters α_1 and α_2 . Notice that $g(u, v) = g_1(u) \cdot g_2(v)$, U and V are independent.

Example 3.5.11 Suppose X and Y are independent, their PDFs are

$$\begin{aligned} f_X(x) &= \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}, \\ f_Y(y) &= \begin{cases} e^{-y}, & y > 0 \\ 0, & y \leq 0 \end{cases}. \end{aligned}$$

Determine the PDF of $Z = 2X + Y$.

Solution:

Method 1: Let $W = 2X$, $w = 2x$ is an increasing function, its inverse function is $x = \frac{w}{2}$, $\frac{dx}{dw} = \frac{1}{2}$.

Since $0 \leq X \leq 1$, $0 \leq W \leq 2$. By Theorem 2.7.1,

$$\begin{aligned} W \sim f_W(w) &= \begin{cases} f_X\left(\frac{w}{2}\right) \times \left|\frac{1}{2}\right|, & 0 \leq w \leq 2 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{2}, & 0 \leq w \leq 2 \\ 0, & \text{otherwise} \end{cases}. \end{aligned}$$

By the convolution formula in Theorem 3.5.1, the PDF of $Z = W + Y$ is $f_Z(z) =$

$$\int_{-\infty}^{+\infty} f(w, z - w) dw = \int_{-\infty}^{+\infty} f_W(w) f_Y(z - w) dw.$$

$$f_W(w) f_Y(z - w) = \begin{cases} \frac{1}{2} e^{-(z - w)}, & 0 \leq w \leq 2, z - w > 0 \\ 0, & \text{otherwise} \end{cases},$$

where the region $\{(w, z) | 0 \leq w \leq 2, z - w > 0\}$ is sketched in Figure 3.29.

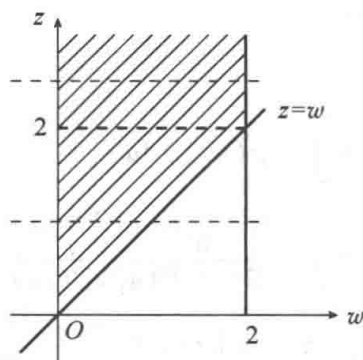


Figure 3.29

(1) If $z \leq 0$, $f_Z(z) = 0$.

(2) If $0 < z \leq 2$, $f_Z(z) = \int_0^z \frac{1}{2} e^{-(z-w)} dw = \frac{1 - e^{-z}}{2}$.

(3) If $z > 2$, $f_Z(z) = \int_0^2 \frac{1}{2} e^{-(z-w)} dw = \frac{e^{2-z} - e^{-z}}{2}$.

Therefore,

$$f_Z(z) = \begin{cases} \frac{e^{2-z} - e^{-z}}{2}, & z > 2 \\ \frac{1 - e^{-z}}{2}, & 0 < z \leq 2 \\ 0, & z \leq 0 \end{cases}$$

Method 2: $Z = 2X + Y$ can be written as $Z = X + Y + X$, we can find the joint PDF of $U = X + Y$ and $V = X$, then apply the convolution formula. The joint PDF of X and Y is

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} e^{-y}, & 0 \leq x \leq 1, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Let $S = \{(x,y) | 0 \leq x \leq 1, y > 0\}$, $T = \{(u,v) | u > v, 0 \leq v \leq 1\}$, then $U = X + Y$ and $V = X$ define a one-to-one transformation from S onto T , its inverse transformation is given by $X = V, Y = U - V$. The Jacobian determinant is

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1, |J| = 1.$$

Thus, the joint PDF of U and V is

$$f_{U,V}(u,v) = \begin{cases} e^{v-u}, & u > v, 0 \leq v \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Now we determine the PDF of $Z = 2X + Y = U + V$, $f_Z(z) = \int_{-\infty}^{+\infty} f_{U,V}(u, z-u) du$, where

$$f_{U,V}(u, z-u) = \begin{cases} e^{z-2u}, & 2u > z, 0 \leq z-u \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The region $\{(u,z) | 2u > z, 0 \leq z-u \leq 1\}$ is sketched in Figure 3.30.

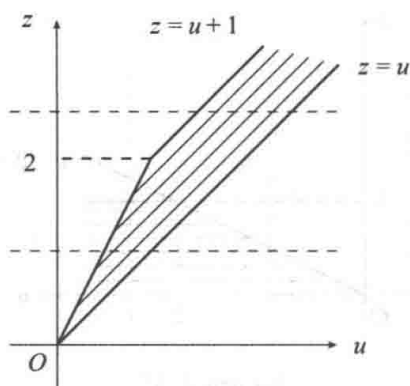


Figure 3.30

(1) If $z \leq 0$, $f_Z(z) = 0$.

(2) If $0 < z \leq 2$, $f_Z(z) = \int_{\frac{z}{2}}^z e^{z-2u} du = \frac{1 - e^{-z}}{2}$.

(3) If $z > 2$, $f_Z(z) = \int_{z-1}^z e^{z-2u} du = \frac{e^{2-z} - e^{-z}}{2}$.

Therefore, the PDF of $Z = 2X + Y$ is

$$f_Z(z) = \begin{cases} \frac{e^{2-z} - e^{-z}}{2}, & z > 2 \\ \frac{1 - e^{-z}}{2}, & 0 < z \leq 2 \\ 0, & z \leq 0 \end{cases}$$

Method 3: First we determine the joint PDF of $U = 2X + Y$ and $V = X$, then find the marginal PDF of U .

The joint PDF of X and Y is

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} e^{-y}, & 0 \leq x \leq 1, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Let

$$S = \{(x,y) | 0 \leq x \leq 1, y > 0\},$$

$$T = \{(u,v) | 0 \leq v \leq 1, u - 2v > 0\} = \{(u,v) | u > 2v, 0 \leq v \leq 1\},$$

then $U = 2X + Y$ and $V = X$ define a one-to-one transformation from S onto T , its inverse transformation is $X = V, Y = U - 2V$.

The Jacobian determinant is

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -2 \end{vmatrix} = -1, |J| = 1.$$

Thus, the joint PDF of U and V is

$$f_{U,V}(u,v) = \begin{cases} e^{2v-u}, & u > 2v, 0 \leq v \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The region $\{(u,v) | u > 2v, 0 \leq v \leq 1\}$ is sketched in Figure 3.31.

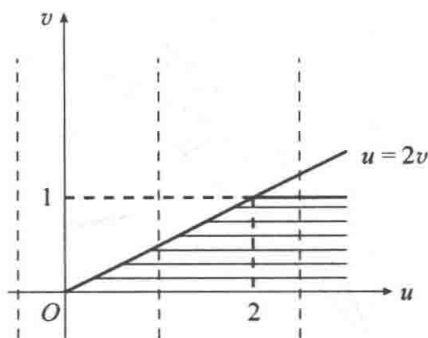


Figure 3.31

Next we determine the marginal PDF of $U=2X+Y$, $f_U(u) = \int_{-\infty}^{+\infty} f_{U,V}(u,v) dv$.

(1) If $u \leq 0$, $f_U(u) = 0$.

(2) If $0 < u \leq 2$, $f_U(u) = \int_0^{\frac{u}{2}} e^{2v-u} dv = \frac{1-e^{-u}}{2}$.

(3) If $u > 2$, $f_U(u) = \int_0^1 e^{2v-u} dv = \frac{e^{2-u} - e^{-u}}{2}$.

Thus, the PDF of $U=2X+Y$ is

$$f_U(u) = \begin{cases} \frac{e^{2-u} - e^{-u}}{2}, & u > 2 \\ \frac{1-e^{-u}}{2}, & 0 < u \leq 2 \\ 0, & u \leq 0 \end{cases}$$

Method 4: First we determine the CDF $F_Z(z)$ of $Z=2X+Y$.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(2X+Y \leq z) \\ &= \iint_{2x+y \leq z} f_{X,Y}(x,y) dx dy. \end{aligned}$$

Notice that the joint PDF of X and Y is

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} e^{-y}, & 0 \leq x \leq 1, y > 0 \\ 0, & \text{otherwise} \end{cases}.$$

(1) If $z \leq 0$, $F_Z(z) = P(\emptyset) = 0$. See Figure 3.32.

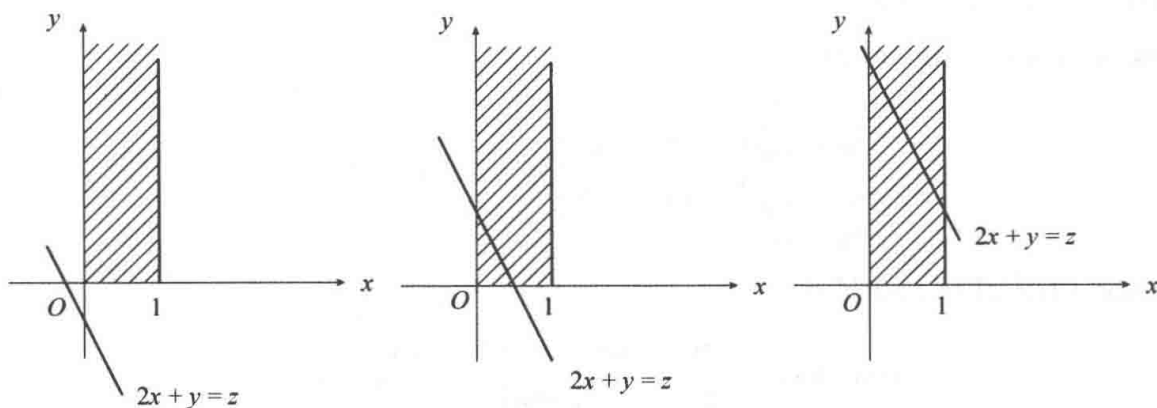


Figure 3.32

(2) If $0 < z \leq 2$,

$$\begin{aligned}
 F_Z(z) &= \iint_{2x+y \leq z} f(x,y) dx dy \\
 &= \int_0^{\frac{z}{2}} \left(\int_0^{z-2x} e^{-y} dy \right) dx \\
 &= \int_0^{\frac{z}{2}} (1 - e^{2x-z}) dx \\
 &= \frac{1}{2} (e^{-z} + z - 1).
 \end{aligned}$$

(3) If $z > 2$,

$$\begin{aligned}
 F_Z(z) &= \iint_{2x+y \leq z} f(x,y) dx dy \\
 &= \int_0^1 \left(\int_0^{z-2x} e^{-y} dy \right) dx \\
 &= \int_0^1 (1 - e^{2x-z}) dx \\
 &= \frac{1}{2} (e^{-z} - e^{2-z} + 2).
 \end{aligned}$$

Therefore, the CDF of $Z=2X+Y$ is

$$F_Z(z) = \begin{cases} 0 & , z \leq 0 \\ \frac{1}{2} (e^{-z} + z - 1) & , 0 < z \leq 2 \\ \frac{1}{2} (e^{-z} - e^{2-z} + 2) & , z > 2 \end{cases}$$

The PDF of $Z=2X+Y$ is

$$f_Z(z) = \begin{cases} \frac{e^{2-z} - e^{-z}}{2} & , z > 2 \\ \frac{1 - e^{-z}}{2} & , 0 < z \leq 2 \\ 0 & , z \leq 0 \end{cases}$$

Example 3.5.12 Suppose (X,Y) follows uniform distribution over $D = \{(x,y) \mid y \geq 0, x^2 + y^2 \leq 1\}$, see Figure 3.33. U and V are defined as

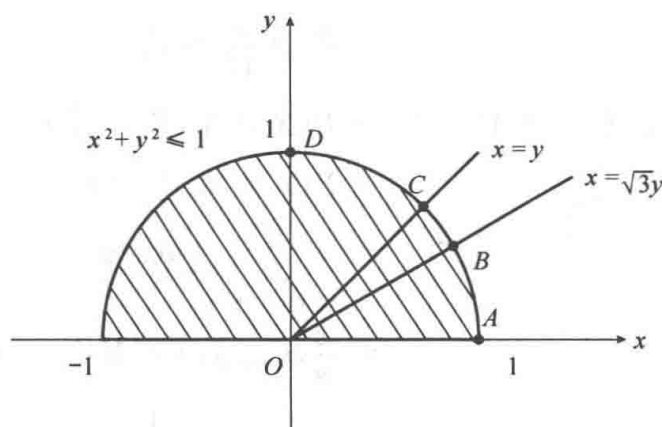


Figure 3.33

$$U = \begin{cases} 0, X < 0 \\ 1, 0 \leq X < Y \\ 2, X \geq Y \end{cases}, \quad V = \begin{cases} 0, X \geq \sqrt{3}Y \\ 1, X < \sqrt{3}Y \end{cases}.$$

Determine the joint PF of (U, V) .

Solution: The joint PDF of (X, Y) is $f(x, y) = \begin{cases} \frac{2}{\pi}, y \geq 0, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$.

$$P(U = 0, V = 0) = P(X < 0, X \geq \sqrt{3}Y) = P(\emptyset) = 0,$$

$$P(U = 0, V = 1) = P(X < 0, X < \sqrt{3}Y)$$

$$= P(X < 0)$$

$$= \iint_{x < 0 \cap D} \frac{2}{\pi} dx dy$$

$$= \frac{2}{\pi} \times \frac{\pi}{4} = \frac{1}{2},$$

$$P(U = 1, V = 0) = P(0 \leq X < Y, X \geq \sqrt{3}Y) = P(\emptyset) = 0,$$

$$P(U = 1, V = 1) = P(0 \leq X < Y, X < \sqrt{3}Y)$$

$$= P(0 \leq X < Y)$$

$$= \iint_{0 \leq x < y \cap D} \frac{2}{\pi} dx dy$$

$$= \frac{2}{\pi} S_{0 \leq x < y \cap D} (\angle COD = \frac{\pi}{4})$$

$$= \frac{2}{\pi} \times \frac{\pi}{8} = \frac{1}{4},$$

$$P(U = 2, V = 0) = P(X \geq Y, X \geq \sqrt{3}Y)$$

$$= P(X \geq \sqrt{3}Y)$$

$$= \iint_{x \geq \sqrt{3}y \cap D} \frac{2}{\pi} dx dy$$

$$= \frac{2}{\pi} S_{x \geq \sqrt{3}y \cap D}$$

$$= \frac{2}{\pi} S_{\text{sector } AOB} (\angle AOB = \frac{\pi}{6})$$

$$= \frac{2}{\pi} \times \frac{\pi}{12} = \frac{1}{6},$$

$$P(U = 2, V = 1) = P(X \geq Y, X < \sqrt{3}Y)$$

$$= P(Y \leq X < \sqrt{3}Y)$$

$$= \iint_{y \leq x < \sqrt{3}y \cap D} \frac{2}{\pi} dx dy$$

$$= \frac{2}{\pi} S_{\text{sector } BOC} (\angle AOB = \frac{\pi}{6}, \angle AOC = \frac{\pi}{4})$$

$$= \frac{2}{\pi} \times \frac{\pi}{24} = \frac{1}{12}.$$

Therefore, the joint PF of (U, V) is

	$V=0$	$V=1$
$U=0$	0	$\frac{1}{2}$
$U=1$	0	$\frac{1}{4}$
$U=2$	$\frac{1}{6}$	$\frac{1}{12}$

◆ Exercise 3 ◆

3.1 Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let X_i equal 1 if the i -th ball selected is white and let it equal 0 otherwise. Find the joint PF of (X_1, X_2) .

3.2 Two balanced dice are rolled. Let X be the larger of the two values shown on the dice, and let Y be the absolute value of the difference of the two values shown. Give the joint PF of X and Y .

3.3 Toss an unbiased coin four times. Let X be the number of heads in the first three tosses and Y be the number of heads in the last two tosses. Find the joint PF of (X, Y) .

3.4 A bag contains 5 chocolate-chip muffins, 3 blueberry muffins and 2 lemon muffins. Choose three muffins from this bag at random. Let X be the number of lemon muffins chosen and Y be the number of blueberry muffins chosen. Find the joint PF of (X, Y) .

3.5 Suppose that 2 batteries are randomly chosen without replacement from the following group of 12 batteries: 3 new, 4 used, 5 defective. Let X denote the number of new batteries chosen. Let Y denote the number of used batteries chosen. Find the joint PF of (X, Y) .

3.6 Suppose the joint PDF of two random variables X and Y is given by

$$f(x, y) = \begin{cases} Ce^{-(x+y)}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find:

- (1) the constant C ;
- (2) the joint CDF $F(x, y)$;
- (3) $P(X > 1)$;
- (4) $P(X \geq Y)$;
- (5) $P(X + 2Y \leq 1)$.

3.7 Let the joint PDF of two random variable X and Y be given by

$$f(x, y) = \begin{cases} Ce^{-2(x+y)}, & x > 0, y > 0 \\ 0 & , \text{otherwise} \end{cases}.$$

Find:

- (1) the constant C ;
- (2) the joint CDF $F(x, y)$;
- (3) the marginal CDFs and PDFs of X and Y ;
- (4) the probability that (X, Y) falls into the region $\{(x, y) | x + y < 1\}$.

3.8 Suppose the joint PDF of two random variable X and Y is

$$f(x, y) = \begin{cases} ae^{-y}, & 0 < x < y \\ 0 & , \text{otherwise} \end{cases}.$$

Find:

- (1) the constant a ;
- (2) the marginal PDFs of X and Y ;
- (3) $P(X + Y \leq 1)$.

3.9 Suppose that (X, Y) follows the uniform distribution over the region

$$D = \{(x, y) | 0 \leq y \leq 1 - x^2\}.$$

- (1) What is the joint PDF of (X, Y) ?
- (2) What are the marginal PDFs of X and Y ?
- (3) What is the conditional PDF of Y when $X = -0.5$?
- (4) Find $P[(X, Y) \in B]$, where $B = \{(x, y) | y \geq x^2\}$.

3.10 The joint PF of (X, Y) is

$$P(X = n, Y = m) = \frac{e^{-14} \times 7.14^m \times 6.86^{n-m}}{m!(n-m)!}, m = 0, 1, \dots, n; n = 0, 1, 2, \dots.$$

Find:

- (1) the marginal PFs;
- (2) the conditional PFs.

3.11 Let the joint PDF of two random variable X and Y be given by

$$f(x, y) = \begin{cases} 4xy, & 0 \leq x < 1, 0 \leq y < 1 \\ 0 & , \text{otherwise} \end{cases}.$$

Find:

- (1) the conditional PDFs $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$;
- (2) the joint CDF $F(x, y)$.

3.12 Let the joint PDF of two random variable X and Y be given by

$$f(x, y) = \begin{cases} Cxe^{-y}, & 0 < x < y < +\infty \\ 0 & , \text{otherwise} \end{cases}.$$

- (1) Find the constant C .
- (2) Are X and Y independent? why?
- (3) Find $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

(4) Find $P(X < 1 | Y < 2)$ and $P(X < 1 | Y = 2)$.

3.13 The PDF of X is

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3.1)$$

Y is uniformly distributed on $(0, X)$. Find:

- (1) $f_{Y|X}(y|x)$;
- (2) the joint PDF of (X, Y) ;
- (3) the PDF of Y .

3.14 The joint PDF of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{2x^2 y}, & 1 \leq x, \frac{1}{x} \leq y \leq x \\ 0, & \text{otherwise} \end{cases}$$

Are X and Y independent?

3.15 X and Y are two independent random variables. $X \sim U(0, 1)$. The PDF of Y is given by

$$f(y) = \begin{cases} \frac{1}{2} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

- (1) Determine the joint PDF of (X, Y) .
- (2) Find the probability that the following equation has real roots.

$$aX^2 + 2aX + Y^2 = 0.$$

3.16 Let a needle of length L be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance d from each other, with $d > L$. What is the probability that the needle will intersect one of these lines?

3.17 Assume that U is uniform on $[0, 1]$ and that $F(x)$ is the CDF of a continuous random variable X . Show that $F^{-1}(U)$ has the same distribution as X .

3.18 Suppose $X \sim P(\lambda_1)$ and $Y \sim P(\lambda_2)$ are independent. Find the probability function of $Z = X + Y$.

3.19 Suppose the joint PDF of two random variables X and Y is given by

$$f(x, y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the PDF of $Z = X + Y$.

3.20 Suppose the joint PDF of two random variables X and Y is given by

$$f(x, y) = \begin{cases} 3x, & 0 < x < 1, 0 < y < x \\ 0, & \text{otherwise} \end{cases}$$

Find the PDF of $Z = X - Y$.

3.21 Suppose that X_1 and X_2 are IID. random variables and that the PDF of each of them is as follows:

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the PDF of $Y = X_1 - X_2$.

3.22 X and Y are two independent random variables. Their PDF are given by

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}, f_Y(y) = \begin{cases} e^{-y}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Determine the PDF of $Z = X + Y$.

3.23 Suppose the joint PDF of two random variables X and Y is given by

$$f(x, y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Determine the PDF of $Z = |X - Y|$.

3.24 The joint PF of (X, Y) is given in the following table.

	$Y = -1$	$Y = 1$	$Y = 2$
$X = -1$	$\frac{5}{20}$	$\frac{2}{20}$	$\frac{6}{20}$
$X = 2$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{1}{20}$

Find the PF of

$$(1) Z_1 = X + Y; (2) Z_2 = XY; (3) Z_3 = \frac{X}{Y}; (4) Z_4 = \max\{X, Y\}.$$

3.25 Suppose that (X, Y) follows the uniform distribution on region $D = \{(x, y) | 0 < x < a, 0 < y < a\}$.

$$(1) \text{ Find the PDF of } Z = \frac{X}{Y}.$$

$$(2) \text{ Find the PDF of } M = \max\{X, Y\}.$$

3.26 Suppose that X_1, X_2, X_3, X_4 are IID and

$$P(X_i = 0) = 0.6, P(X_i = 1) = 0.4, i = 1, 2, 3, 4.$$

$$(1) \text{ Find the PF of } X = \begin{vmatrix} X_1 & X_2 \\ X_3 & X_4 \end{vmatrix}.$$

(2) Find the probability that the following system has exactly one zero solution $x_1 = x_2 = 0$.

$$\begin{cases} X_1 x_1 + X_2 x_2 = 0 \\ X_3 x_1 + X_4 x_2 = 0 \end{cases}$$

3.27 Suppose that a point in the xy -plane is chosen at random from the interior of a circle for which the equation is $x^2 + y^2 = 1$; and suppose that the probability that the point will belong to each region inside the circle is proportional to the area of that region. Let Z denote a random variable representing the distance from the center of the circle to the point. Find and sketch the CDF of Z .



3.28 The joint PF of (X, Y) is given in the table below.

	$Y=0$	$Y=1$	$Y=2$	$Y=3$
$X=0$	0.08	0.07	0.04	0
$X=1$	0.06	0.15	0.05	0.04
$X=2$	0.05	0.04	0.10	0.06
$X=3$	0	0.03	0.04	0.07
$X=4$	0	0.01	0.05	0.06

Determine each of the following probabilities:

- (1) $P(X=2)$; (2) $P(Y \geq 2)$; (3) $P(X \leq 2, Y \leq 2)$; (4) $P(X=Y)$.

3.29 Suppose that X and Y have a discrete joint distribution for which the joint PF is defined as follows:

$$f(x, y) = \begin{cases} c|x+y|, & x = -2, -1, 0, 1, 2 \text{ and } y = -2, -1, 0, 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

Determine:

- (1) the value of the constant c ;
 (2) $P(X=0, Y=-2)$;
 (3) $P(X=1)$;
 (4) $P(|X-Y| \leq 1)$.

3.30 Suppose that X and Y have a continuous joint distribution for which the joint PDF is defined as follows:

$$f(x, y) = \begin{cases} cy^2, & 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine:

- (1) the value of the constant c ;
 (2) $P(X+Y > 2)$;
 (3) $P(Y < \frac{1}{2})$;
 (4) $P(X \leq 1)$;
 (5) $P(X=3Y)$.

3.31 Suppose that the joint PDF of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} c(x^2 + y), & 0 \leq y \leq 1 - x^2 \\ 0, & \text{otherwise} \end{cases}$$

Determine:

- (1) the value of the constant c ;
 (2) $P(0 \leq X \leq \frac{1}{2})$;
 (3) $P(Y \leq X+1)$;

(4) $P(Y=X^2)$.

3.32 Suppose that X and Y are random variables such that (X, Y) must belong to the rectangle in the xy -plane containing all points (x, y) for which $0 \leq x \leq 3$ and $0 \leq y \leq 4$. Suppose also that the joint CDF of X and Y at every point (x, y) in this rectangle is specified as follows:

$$F(x, y) = \frac{1}{156}xy(x^2 + y).$$

Determine:

- (1) $P(1 \leq X \leq 2 \text{ and } 1 \leq Y \leq 2)$;
- (2) $P(2 \leq X \leq 4 \text{ and } 2 \leq Y \leq 4)$;
- (3) the CDF of Y ;
- (4) the joint PDF of X and Y ;
- (5) $P(Y \leq X)$.

3.33 Suppose that X and Y have a discrete joint distribution for which the joint PF is defined as follows:

$$f(x, y) = \begin{cases} \frac{1}{30}(x+y), & x=0, 1, 2 \text{ and } y=0, 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$$

Are X and Y independent?

3.34 Suppose that the joint PDF of X and Y is as follows:

$$f(x, y) = \begin{cases} \frac{15}{4}x^2, & 0 \leq y \leq 1-x^2 \\ 0, & \text{otherwise} \end{cases}$$

Are X and Y independent?

3.35 Suppose that the joint PDF of X and Y is as follows:

$$f(x, y) = \begin{cases} 2xe^{-y}, & 0 \leq x \leq 1 \text{ and } 0 < y < +\infty \\ 0, & \text{otherwise} \end{cases}$$

Are X and Y independent?

3.36 Suppose that a point (X, Y) is chosen at random from the circle S defined as follows:

$$S = \{(x, y) | x^2 + y^2 \leq 1\}.$$

(1) Determine the joint PDF of X and Y , the marginal PDF of X , and the marginal PDF of Y .

(2) Are X and Y independent?

3.37 Suppose that a point (X, Y) is chosen at random from the disk S defined as follows:

$$S = \{(x, y) | (x-1)^2 + (y+2)^2 \leq 9\}.$$

Determine:

- (1) the conditional PDF of Y for every given value of X ;
- (2) $P(Y > 0 | X = 2)$.

3.38 Suppose that the joint PDF of two random variables X and Y is as follows:

$$f(x, y) = \begin{cases} \frac{3}{16}(4 - 2x - y), & x > 0, y > 0 \text{ and } 2x + y < 4 \\ 0, & \text{otherwise} \end{cases}$$

Determine:

- (1) the conditional PDF of Y for every given value of X ;
- (2) $P(Y \geq 2 | X = 0.5)$.

3.39 Suppose that $X_1, \dots, X_n$ form a random sample of size n from the uniform distribution on the interval $[0, 1]$ and that $Y_n = \max\{X_1, \dots, X_n\}$. Find the smallest value of n such that $P(Y_n \geq 0.99) \geq 0.95$.

3.40 Suppose that the random variables X and Y have the following joint PDF:

$$f(x, y) = \begin{cases} 8xy, & 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Also, let $U = \frac{X}{Y}$ and $V = Y$.

- (1) Determine the joint PDF of U and V .
- (2) Are X and Y independent?
- (3) Are U and V independent?

Chapter 4

Characteristics of Random Variables

◆ 4.1 The Expectation of a Random Variable ◆

The CDF of a random variable completely describes its statistical regularity. Sometimes we need know the characteristics of random variables. Summaries of the distribution, such as the average value, or expected value, can be useful for giving us an idea of where we expect the variable to be without trying to describe the entire distribution.

Example 4.1.1 There are N students in a class, the students took part in an exam. n_i students obtained x_i points, $i=1, 2, \dots, k$. $\sum_{i=1}^k n_i = N$. Find the average score of the class.

Solution: Let $\bar{x}$ be the average score,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k x_i \frac{n_i}{N}.$$

Let X be the score of any student selected, X is a random variable and

$$P(X=x_i) = \frac{n_i}{N}, \quad i=1, 2, \dots, k,$$

so the average score $\bar{x}$ can be rewritten as

$$\bar{x} = \sum_{i=1}^k x_i P(X=x_i)$$

The average value of X equals the sum of the probability of each possible outcome of the experiment multiplied by the value itself.

Definition 4.1.1 Let X be a discrete random variable whose PF is

$$P(X=x_k) = p_k, k=1, 2, \dots.$$

If $\sum_{k=1}^{+\infty} |x_k| p_k < +\infty$, then the mean, expectation, or expected value of X is said to exist and is defined to be

$$E(X) = \sum_{k=1}^{+\infty} x_k p_k.$$

Remark 4.1.1 We not only require that $\sum_{k=1}^{+\infty} x_k p_k$ converges, but also converges absolutely since a rearrangement of non-absolutely-convergent countably-infinite sum can have a different sum. Absolute convergence guarantees the uniqueness of expectation.

Example 4.1.2 Suppose a random variable X has the following PF:

X	-2	0	1	4
$P(X=x)$	0.1	0.4	0.3	0.2

Determine its expectation.

Solution: $E(X) = -2 \times 0.1 + 0 \times 0.4 + 1 \times 0.3 + 4 \times 0.2 = 0.9$.

Example 4.1.3 Suppose a random variable X has the following PF:

$$p_k = P[X = (-1)^k k] = \frac{1}{k(k+1)}, k = 1, 2, \dots$$

Does the expectation of X exist?

Solution: $\sum_{k=1}^{+\infty} |x_k| p_k = \sum_{k=1}^{+\infty} |(-1)^k k| \cdot \frac{1}{k(k+1)} = \sum_{k=1}^{+\infty} \frac{1}{k+1} = +\infty$. The expectation of X does not exist.

Definition 4.1.2 Let X be a continuous random variable whose PDF is $f(x)$, if $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$, then the mean, expectation, or expected value of X is said to exist and is defined to be

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

Example 4.1.4 Suppose a random variable X has the following PDF:

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Determine its expectation.

$$\text{Solution: } E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^1 x(2x) dx = \frac{2}{3}.$$

Example 4.1.5 Suppose that a random variable X has a continuous distribution for which the PDF is as follows:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty.$$

This distribution is called the Cauchy distribution.

$$\int_{-\infty}^{+\infty} |x| f(x) dx = 2 \int_0^{+\infty} x \frac{1}{\pi(1+x^2)} dx = \frac{\ln(1+x^2)}{\pi} \Big|_0^{+\infty} = +\infty.$$

Hence, the expectation of X does not exist.

Theorem 4.1.1 Suppose X is a random variable, Y is a function of X , $Y = r(X)$,

(1) If X is discrete with PF $P(X = x_k) = p_k, k = 1, 2, \dots$. And $\sum_{k=1}^{+\infty} r(x_k) p_k$ converges absolutely, then

$$E(Y) = E[r(X)] = \sum_{k=1}^{+\infty} r(x_k) p_k = \sum_{\text{All } y} y g(y),$$

where $g(y)$ is the PF of Y .

(2) If X is continuous with PDF $f(x)$ and $\int_{-\infty}^{+\infty} r(x) f(x) dx$ converges absolutely, then

$$E(Y) = \int_{-\infty}^{+\infty} r(x) f(x) dx = \int_{-\infty}^{+\infty} y g(y) dy,$$

where $g(y)$ is the PDF of Y .

Example 4.1.6 Suppose a random variable X has the following PF:

X	-2	0	2
$f(x)$	0.1	0.4	0.5

Let $Y = X^2$, Find $E(Y)$.

Solution:

Method 1: $E(Y) = E(X^2) = (-2)^2 \times 0.1 + 0^2 \times 0.4 + 2^2 \times 0.5 = 2.4$.

Method 2: Find the PF $g(y)$ of Y .

Y	4	0
$g(y)$	0.6	0.4

$$E(Y) = 4 \times 0.6 + 0 \times 0.4 = 2.4.$$

Example 4.1.7 Suppose a random variable X has the following PDF:

$$f(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Let $Y = X^2$, find $E(Y)$.

Solution:

$$\text{Method 1: } E(Y) = E(X^2) = \int_{-1}^1 x^2 \frac{1}{2} dx = \frac{1}{3}.$$

Method 2: The CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y).$$

$$\text{If } y \leq 0, F_Y(y) = P(\emptyset) = 0.$$

$$\text{If } y \geq 1, F_Y(y) = P(\Omega) = 1.$$

$$\text{If } 0 < y < 1,$$

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx \\ &= \sqrt{y}. \end{aligned}$$

The PDF $g(y)$ of Y is

$$g(y) = F'_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

$$\text{Therefore, } E(Y) = \int_{-\infty}^{+\infty} yg(y) dy = \int_0^1 y \frac{1}{2\sqrt{y}} dy = \frac{1}{3}.$$

Theorem 4.1.2 Suppose that $X_1, \dots, X_n$ are random variables with the joint PDF $f(x_1, \dots, x_n)$, and suppose that $Y = r(X_1, \dots, X_n)$. Then $E(Y)$ can be determined directly from the relation

$$E(Y) = \int \cdots \int r(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

If the mean exists. Similarly, if $X_1, \dots, X_n$ have a discrete joint distribution with PF $f(x_1, \dots, x_n)$,

$$E(Y) = \sum \cdots \sum r(x_1, \dots, x_n) f(x_1, \dots, x_n)$$

Example 4.1.8 Suppose that a point (X, Y) is chosen at random from the square S containing all points (x, y) such that $0 \leq x \leq 1$ and $0 \leq y \leq 1$. See Figure 4.1. Determine the expected value of $Z = X^2 + Y^2$.

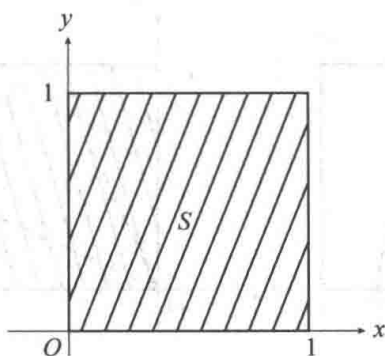


Figure 4.1

Solution:

Method 1: The PDF of (X, Y) is

$$f(x, y) = \begin{cases} 1, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(Z) &= E(X^2 + Y^2) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) f(x, y) dx dy \\ &= \int_0^1 \int_0^1 (x^2 + y^2) dx dy \\ &= \frac{2}{3}. \end{aligned}$$

Method 2: The marginal PDF of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3}.$$

Similarly, $E(Y^2) = \frac{1}{3}$, therefore, $E(Z) = E(X^2) + E(Y^2) = \frac{2}{3}$.

Method 3: We would like to use the PDF of Z to calculate its expectation. First, we want to determine the CDF of Z .

$$F_Z(z) = P(Z \leq z) = P(X^2 + Y^2 \leq z)$$

If $z \leq 0$, $F_Z(z) = P(\emptyset) = 0$.

If $0 < z < 1$, see Figure 4.2(a),

$$\begin{aligned}
 F_Z(z) &= P(X^2 + Y^2 \leq z) \\
 &= \iint_{x^2 + y^2 \leq z} f(x, y) dx dy \\
 &= \iint_{D_1} 1 dx dy \\
 &= S_{D_1} \\
 &= \frac{\pi z}{4}.
 \end{aligned}$$

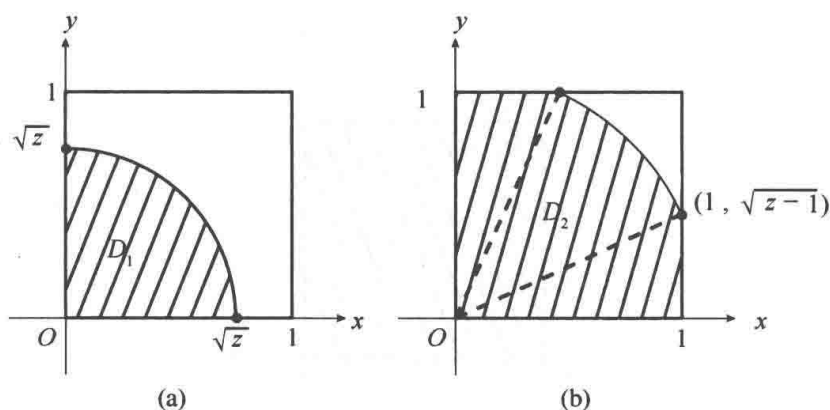


Figure 4.2

If $1 \leq z < 2$, see Figure 4.2(b)

$$\begin{aligned}
 F_Z(z) &= P(X^2 + Y^2 \leq z) \\
 &= \iint_{x^2 + y^2 \leq z} f(x, y) dx dy \\
 &= \iint_{D_2} 1 dx dy \\
 &= S_{D_2} \\
 &= \sqrt{z-1} + \frac{1}{2}z \left(\frac{\pi}{2} - 2\arctan \sqrt{z-1} \right).
 \end{aligned}$$

If $z \geq 2$, $F_Z(z) = P(\Omega) = 1$. In summary, the CDF of Z is

$$F_Z(z) = \begin{cases} 0 & , z \leq 0 \\ \frac{\pi z}{4} & , 0 < z < 1 \\ \sqrt{z-1} + \frac{1}{2}z \left(\frac{\pi}{2} - 2\arctan \sqrt{z-1} \right) & , 1 \leq z < 2 \\ 1 & , z \geq 2 \end{cases}.$$

The PDF of Z is

$$f_Z(z) = F'_Z(z) = \begin{cases} 0 & , z \leq 0 \text{ or } z \geq 2 \\ \frac{\pi}{4} & , 0 < z < 1 \\ \frac{1}{4}(\pi - 4\arctan \sqrt{z-1}) & , 1 \leq z < 2 \end{cases}.$$

Therefore, the expectation of Z is

$$\begin{aligned}
 E(Z) &= \int_{-\infty}^{+\infty} z f_Z(z) dz \\
 &= \int_0^1 \frac{\pi}{4} z dz + \int_1^2 \frac{z}{4} (\pi - 4 \arctan \sqrt{z-1}) dz \\
 &= \int_0^1 \frac{\pi}{4} z dz - \int_1^2 z \arctan \sqrt{z-1} dz \\
 &= \frac{\pi}{2} - \int_1^2 \arctan \sqrt{z-1} d\left(\frac{z^2}{2}\right) \\
 &= \frac{\pi}{2} - \left[\frac{z^2}{2} \arctan \sqrt{z-1} \Big|_{z=1}^{z=2} - \int_1^2 \frac{z}{4\sqrt{z-1}} dz \right] \\
 &= \int_1^2 \frac{z}{4\sqrt{z-1}} dz \quad (\text{let } t = \sqrt{z-1}) \\
 &= \int_0^1 \frac{1+t^2}{2} dt \\
 &= \frac{2}{3}.
 \end{aligned}$$

Example 4.1.9 Suppose $X, Y \sim U(0, 1)$ are independent, find $E(|X - Y|)$.

Solution:

Method 1: The PDFs of X and Y are

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}, \quad f_Y(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Since X and Y are independent, the joint PDF of (X, Y) is

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

The region $\{(x, y) | 0 < x < 1, 0 < y < 1\}$ is shown in Figure 4.3. Therefore,

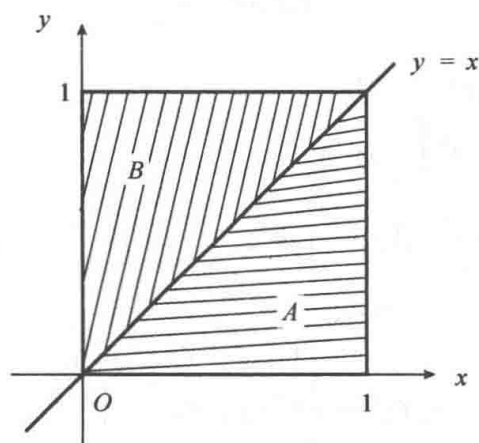


Figure 4.3

$$\begin{aligned}
 E(|X - Y|) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y| f(x, y) dx dy \\
 &= \int_0^1 \int_0^1 |x - y| dx dy \\
 &= \iint_A |x - y| dx dy + \iint_B |x - y| dx dy \\
 &= \int_0^1 \left[\int_0^x (x - y) dy \right] dx + \int_0^1 \left[\int_0^y (y - x) dx \right] dy \\
 &= 2 \int_0^1 \left[\int_0^x (x - y) dy \right] dx \\
 &= 2 \int_0^1 \left(x^2 - \frac{x^2}{2} \right) dx \\
 &= \frac{1}{3}.
 \end{aligned}$$

Method 2: Find the PDF of $Z = X - Y$, $f_Z(z) = \int_{-\infty}^{+\infty} f(z + y, y) dy$, where

$$f(z + y, y) = \begin{cases} 1, & 0 < z + y < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

The region $\{(y, z) | 0 < z + y < 1, 0 < y < 1\}$ is shown in Figure 4. 4.

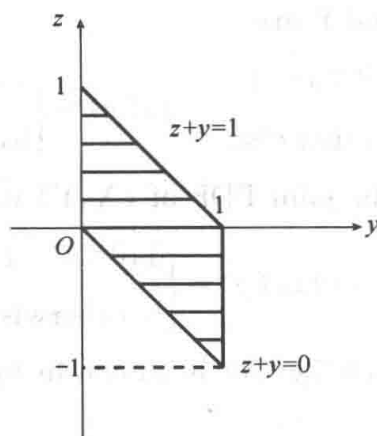


Figure 4. 4

- (1) If $z \leq -1$ or $z \geq 1$, $f_Z(z) = 0$.
- (2) If $-1 < z < 0$, $f_Z(z) = \int_{-z}^1 dy = 1 + z$.
- (3) If $0 \leq z < 1$, $f_Z(z) = \int_0^{1-z} dy = 1 - z$.

In summary,

$$f_Z(z) = \begin{cases} 1 + z, & -1 < z < 0 \\ 1 - z, & 0 \leq z < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore,

$$\begin{aligned}
 E(|X-Y|) &= E|Z| \\
 &= \int_{-\infty}^{+\infty} |z| f_Z(z) dz \\
 &= \int_{-1}^0 (-z)(1+z) dz + \int_0^1 z(1-z) dz \\
 &= \left[-\frac{z^2}{2} - \frac{z^3}{3} \right]_{z=-1}^{z=0} + \left[\frac{z^2}{2} - \frac{z^3}{3} \right]_{z=0}^{z=1} \\
 &= \frac{1}{3}.
 \end{aligned}$$

Example 4.1.10 Suppose the yearly demand X (unit: ton) in the international market for a certain export commodity made in China is a random variable value, it follows the uniform distribution $U[2000, 4000]$. If one ton of this product is sold, thirty thousand dollars will be earned. If the product is not sold and backlogged in the warehouse, the maintenance cost is ten thousand per ton. How many products should be organized in order to maximize the profit?

Solution: Let y be the number of products ready for export, then $2000 \leq y \leq 4000$. Let Z be the profit (unit: ten thousand dollars), then

$$Z = g(X) = \begin{cases} 3y, & X \geq y \\ 3X - (y - X), & X < y \end{cases} = \begin{cases} 3y, & X \geq y \\ 4X - y, & X < y \end{cases}.$$

The PDF of X is

$$f_X(x) = \begin{cases} \frac{1}{2000}, & 2000 \leq X \leq 4000 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, the expectation of Z is

$$\begin{aligned}
 E(Z) &= E[g(X)] \\
 &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx \\
 &= \frac{1}{2000} \int_{2000}^{4000} g(x) dx \\
 &= \frac{1}{2000} \int_{2000}^y (4x - y) dx + \frac{1}{2000} \int_y^{4000} 3y dx \\
 &= -\frac{1}{1000} (y^2 - 7000y + 4 \times 10^6).
 \end{aligned}$$

$E(Z)$ is a function of y . Let $\frac{d}{dy} E(Z) = \frac{1}{500} (-y + 3500) = 0, y = 3500$. $\frac{d^2}{dy^2} E(Z) = -\frac{1}{500} < 0$, therefore, $E(Z)$ achieves its maximum at $y = 3500$.

Theorem 4.1.3 If $Y = aX + b$, where a and b are finite constants, then

$$E(Y) = aE(X) + b.$$

Note: in general, $E[g(X)] \neq g[E(X)]$, for example, $E[X^2] \neq [E(X)]^2$.

Theorem 4.1.4 If there exists a constant a such that $P(X \geq a) = 1$, then $E(X) \geq a$. If

there exists a constant b such that $P(X \leq b) = 1$, then $E(X) \leq b$.

Theorem 4.1.5 If $X_1, \dots, X_n$ are n random variables such that each expectation $E(X_i)$ is finite ($i=1, \dots, n$), then

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

Note: it is true even if $X_1, \dots, X_n$ are dependent.

Theorem 4.1.6 If $X_1, \dots, X_n$ are n independent random variables such that each expectation $E(X_i)$ is finite ($i=1, \dots, n$), then

$$E(X_1 X_2 \cdots X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n).$$

Proof: We only prove it for the continuous case. The discrete case is similar.

$$\begin{aligned} E(X_1 X_2 \cdots X_n) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_1 \cdots x_n f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_1 \cdots x_n [f_{X_1}(x_1) \cdots f_{X_n}(x_n)] dx_1 \cdots dx_n \\ &= \int_{-\infty}^{+\infty} x_1 f_{X_1}(x_1) dx_1 \cdots \int_{-\infty}^{+\infty} x_n f_{X_n}(x_n) dx_n \\ &= E(X_1) E(X_2) \cdots E(X_n). \end{aligned}$$

Example 4.1.11 Suppose n balls (numbered $1 \sim n$) are randomly placed into n boxes and one box contains one ball. If a ball and its corresponding box have the same number, it is called a match. Let X be the number of matches, find $E(X)$.

Solution: Let $X_i = \begin{cases} 1, & \text{the } i\text{-th box is placed into the } i\text{-th box,} \\ 0, & \text{otherwise} \end{cases}$, then $X = \sum_{i=1}^n X_i$.

Here $X_i (i=1, 2, \dots, n)$ are not independent but have identical distributions.

$$P(X_i=1) = \frac{1}{n}, \quad P(X_i=0) = 1 - \frac{1}{n}.$$

Therefore, $E(X_i) = \frac{1}{n}$, $E(X) = \sum_{i=1}^n E(X_i) = 1$.

Example 4.1.12 Twenty persons board a train. They are equally likely to get off at one of the next 10 stations and choose where to get off independently. The train stops only if someone wants to get off. Find the expected number of stops the train will make.

Solution: Let $X_i = \begin{cases} 1, & \text{at least one person gets off at station } i \\ 0, & \text{nobody gets off at station } i \end{cases}$, then $X = \sum_{i=1}^{10} X_i$. Here

$X_i (i=1, 2, \dots, 10)$ are not independent but have identical distributions.

$$P(X_i=0) = \left(1 - \frac{1}{10}\right)^{20}, \quad P(X_i=1) = 1 - \left(1 - \frac{1}{10}\right)^{20}.$$

Therefore, $E(X_i) = 1 - \left(1 - \frac{1}{10}\right)^{20}$.

$$E(X) = \sum_{i=1}^{10} E(X_i) = 10 \times \left[1 - \left(1 - \frac{1}{10}\right)^{20}\right] = 8.784.$$

◆ 4.2 Variance ◆

Although the mean of a distribution is a useful summary, it does not convey very much information about the distribution. For example, we not only consider the mean lifetime of an electronic component, but also want to know the deviation between its lifetime and its mean lifetime.

Definition 4.2.1 Let X be a random variable with finite mean $\mu = E(X)$. The variance of X is defined as follows:

$$\sigma_X^2 = V(X) = E[(X - \mu)^2].$$

If X has infinite mean or if the mean of X does not exist, we say that $V(X)$ does not exist. The standard deviation σ_X of X is the nonnegative square root of $V(X)$ if the variance exists.

The variance describes the deviation between the values of X and its expectation. If its values are more concentrated, the variance is smaller. If its values are more scattered, the variance is larger. The variance measures the dispersion of a random variable.

Theorem 4.2.1 For every random variable X ,

$$V(X) = E(X^2) - [E(X)]^2.$$

Proof: Using the definition of variance and properties of expectation,

$$\begin{aligned} V(X) &= E\{[X - E(X)]^2\} \\ &= E\{X^2 + 2XE(X) + [E(X)]^2\} \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned}$$

Example 4.2.1 Suppose a random variable X has the following PF:

X	-2	0	2
P	0.1	0.4	0.5

Find its variance.

Solution:

$$\begin{aligned} E(X) &= -2 \times 0.1 + 0 \times 0.4 + 2 \times 0.5 = 0.8, \\ E(X^2) &= (-2)^2 \times 0.1 + 0^2 \times 0.4 + 2^2 \times 0.5 = 2.4, \\ V(X) &= E(X^2) - [E(X)]^2 = 2.4 - 0.8^2 = 1.76. \end{aligned}$$

Example 4.2.2 Suppose a random variable X has the following PDF:

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine its variance.

Solution:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x(2x)dx = \frac{2}{3},$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 x^2 (2x) dx = \frac{1}{2},$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Theorem 4.2.2 For each $X, V(X) \geq 0$. If X is a bounded random variable, then $V(X)$ must exist and be finite.

Theorem 4.2.3 $V(X) = 0$ if and only if there exists a constant $c = E(X)$ such that $P(X=c) = 1$.

Theorem 4.2.4 If X and Y are random variables such that $V(X) < +\infty$ and $V(Y) < +\infty$, let a, b and c be constants, then

$$V(X+Y) = V(X) + V(Y) + 2[E(XY) - E(X)E(Y)],$$

$$V(aX+bY+c) = a^2V(X) + b^2V(Y) + 2ab[E(XY) - E(X)E(Y)].$$

Theorem 4.2.5 If $X_1, \dots, X_n$ are independent random variables with finite means, and if $a_1, \dots, a_n$ and b are arbitrary constants, then

$$V[a_1X_1 + \dots + a_nX_n + b] = a_1^2V(X_1) + \dots + a_n^2V(X_n).$$

Example 4.2.3 A box contains n cards numbered $1 \sim n$. Draw k cards with replacement from this box. Let X be the sum of the numbers of the cards selected. Determine $V(X)$.

Solution: Let X_i be the card number for the i -th draw, $i=1, 2, \dots, k$, then $X = \sum_{i=1}^k X_i$. Since the sampling is with replacement, X_i ($i=1, 2, \dots, k$) are independent and have identical distributions.

$$P(X_i=j) = \frac{1}{n}, \quad j=1, 2, \dots, n, \quad i=1, 2, \dots, k.$$

Therefore,

$$E(X_i) = \sum_{j=1}^n jP(X_i=j) = \frac{n+1}{2}, \quad i=1, 2, \dots, k.$$

$$E(X_i^2) = \sum_{j=1}^n j^2 P(X_i=j) = \frac{(n+1)(2n+1)}{6}, \quad i=1, 2, \dots, k.$$

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = \frac{n^2-1}{12}.$$

Since X_i are independent, $V(X) = \sum_{i=1}^k V(X_i) = \frac{k(n^2-1)}{12}$.

Example 4.2.4 Suppose n balls (numbered $1 \sim n$) are randomly placed into n boxes and one box contains one ball. If a ball and its corresponding box have the same number, it is called a match. Let X be the number of matches, find $V(X)$.

Solution: Let $X_i = \begin{cases} 1, & \text{the } i\text{-th ball is placed into the } i\text{-th box} \\ 0, & \text{otherwise} \end{cases}$, then $X = \sum_{i=1}^n X_i$. Here

X_i ($i=1, 2, \dots, n$) are not independent but have identical distributions.

$$P(X_i=1) = \frac{1}{n}, \quad P(X_i=0) = 1 - \frac{1}{n}.$$

We have showed that $E(X) = \sum_{i=1}^n E(X_i) = 1$ in example 4.1.11.

$$\begin{aligned}
 E(X^2) &= E\left[\left(\sum_{i=1}^n X_i\right)^2\right] \\
 &= E\left(\sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} X_i X_j\right) \\
 &= \sum_{i=1}^n E(X_i^2) + \sum_{1 \leq i \neq j \leq n} E(X_i X_j) \\
 &= \sum_{i=1}^n \left(1^2 \times \frac{1}{n}\right) + \sum_{1 \leq i \neq j \leq n} 1 \times 1 \times P(X_i = 1, X_j = 1) \\
 &= 1 + \sum_{1 \leq i \neq j \leq n} \left(\frac{1}{n} \times \frac{1}{n-1}\right) \\
 &= 2.
 \end{aligned}$$

Therefore, $V(X) = E(X^2) - [E(X)]^2 = 1$.

◆ 4.3 The Characteristics of some Common Distributions ◆

Theorem 4.3.1 A random variable X has the Bernoulli distribution with parameter p ($0 \leq p \leq 1$) if X can take only the values 0 and 1 and the probabilities are $P(X=1)=p, P(X=0)=1-p$, then

$$\begin{aligned}
 E(X) &= 1 \times p + 0 \times (1-p) = p, \\
 E(X^2) &= 1^2 \times p + 0^2 \times (1-p) = p, \\
 V(X) &= E(X^2) - [E(X)]^2 = p - p^2 = p(1-p).
 \end{aligned}$$

If the random variables in a finite or infinite sequence $X_1, X_2, \dots$ are IID, and if each random variable X_i has the Bernoulli distribution with parameter p , then it is said that $X_1, X_2, \dots$ are Bernoulli trials with parameter p .

Theorem 4.3.2 If the random variables $X_1, \dots, X_n$ form n Bernoulli trials with parameter p , and if $X = X_1 + \dots + X_n$, then X has the binomial distribution with parameters n and p . The PF is as follows:

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Then $E(X) = np, V(X) = np(1-p)$.

Proof:

Method 1:

$$\begin{aligned}
 E(X) &= \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
 &= np \sum_{t=0}^{n-1} \frac{(n-1)!}{t!(n-1-t)!} p^t (1-p)^{n-1-t} \\
 &= np [p + (1-p)]^{n-1} \\
 &= np.
 \end{aligned}$$

$$\begin{aligned}
 E(X^2) &= \sum_{k=0}^n k^2 C_n^k p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + np \\
 &= n(n-1)p^2 \sum_{t=0}^{n-2} \frac{(n-2)!}{t!(n-2-t)!} p^t (1-p)^{n-2-t} + np \\
 &= n(n-1)p^2 [p + (1-p)]^{n-2} + np \\
 &= n(n-1)p^2 + np.
 \end{aligned}$$

$$V(X) = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Method 2: $X = X_1 + X_2 + \cdots + X_n$, where $X_i \sim B(1, p)$ are independent ($i=1, 2, \dots, n$).

Using the properties of the mean and variance, we have

$$E(X) = \sum_{i=1}^n E(X_i) = np,$$

$$V(X) = \sum_{i=1}^n V(X_i) = np(1-p).$$

Theorem 4.3.3 If $X \sim P(\lambda)$, then $E(X) = V(X) = \lambda$.

Proof: The PF of X is $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, k=0, 1, \dots$,

$$E(X) = \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{t=0}^{+\infty} \frac{\lambda^t}{t!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

$$\begin{aligned}
 E(X^2) &= \sum_{k=0}^{+\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
 &= \sum_{k=0}^{+\infty} [k(k-1) + k] \frac{\lambda^k}{k!} e^{-\lambda} \\
 &= \sum_{k=2}^{+\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \lambda \\
 &= \lambda^2 e^{-\lambda} \sum_{t=0}^{+\infty} \frac{\lambda^t}{t!} + \lambda \\
 &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda \\
 &= \lambda^2 + \lambda.
 \end{aligned}$$

$$V(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Theorem 4.3.4 Let $X \sim H(N, m, n)$ with strictly positive parameters N, m and n . Then

$$E(X) = \frac{nm}{N},$$

$$V(X) = \frac{nm}{N} \left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right].$$

Proof:

Method 1: We derive a recursive argument. Using the convention that $C'_t = 0$ if $t < 0$ or $t > s$, for $r \geq 1$,

$$\begin{aligned}
 E(X^r) &= \sum_{k=0}^n k^r \frac{C_m^k C_{N-m}^{n-k}}{C_N^n} \\
 &= \sum_{k=1}^n k^r \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}.
 \end{aligned}$$

Observe that $kC_m^k = mC_{m-1}^{k-1}$ and $nC_N^n = NC_{N-1}^{n-1}$,

$$\begin{aligned}
 E(X^r) &= \sum_{k=1}^n k^{r-1} \frac{nm C_{m-1}^{k-1} C_{N-m}^{n-k}}{NC_{N-1}^{n-1}} \\
 &= \frac{nm}{N} \sum_{k=1}^n k^{r-1} \frac{C_{m-1}^{k-1} C_{N-m}^{n-k}}{C_{N-1}^{n-1}} \\
 &= \frac{nm}{N} \sum_{i=1}^{n-1} (i+1)^{r-1} \frac{C_{m-1}^i C_{(N-1)-(m-1)}^{n-1-i}}{C_{N-1}^{n-1}}
 \end{aligned}$$

Let $Y \sim H(N-1, m-1, n-1)$, then

$$E(X^r) = \frac{nm}{N} E[(Y+1)^{r-1}].$$

Therefore, we have

$$E(X) = \frac{nm}{N},$$

$$E(X^2) = \frac{nm}{N} E[(Y+1)] = \frac{nm}{N} (EY+1) = \frac{nm}{n} \left[\frac{(n-1)(m-1)}{N-1} + 1 \right].$$

Consequently,

$$V(X) = E(X^2) - [E(X)]^2 = \frac{nm}{N} \left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right].$$

Method 2: Let $X_i = 1$ if the i -th draw is success and $X_i = 0$ if not. Then each X_i has a Bernoulli distribution $X_i \sim B\left(1, \frac{m}{N}\right)$, $i = 1, 2, \dots, n$. $X = \sum_{i=1}^n X_i$. Here $X_1, \dots, X_n$ are not independent in general.

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n P(X_i = 1) = \frac{nm}{n},$$

$$\begin{aligned}
 V(X) &= V\left(\sum_{i=1}^n X_i\right) \\
 &= \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\
 &= n \times \frac{m}{N} \left(1 - \frac{m}{N}\right) + 2 \sum_{i < j} [E(X_i X_j) - E(X_i)E(X_j)] \\
 &= n \times \frac{m}{N} \left(1 - \frac{m}{N}\right) + 2 \sum_{i < j} \left[\frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2} \right] \\
 &= n \times \frac{m}{N} \left(1 - \frac{m}{N}\right) + n(n-1) \left[\frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2} \right] \\
 &= \frac{nm}{N} \left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right].
 \end{aligned}$$

Theorem 4.3.5 If X has the negative binomial distribution with parameters r and p , the mean and the variance of X must be

$$E(X) = \frac{r}{p}, \quad V(X) = \frac{r(1-p)}{p^2}.$$

Proof: The PF of X is

$$f(x) = P(X = x) = C_{x-1}^{-1} p^r (1-p)^{x-r}, x = r, r+1, \dots$$

$$\begin{aligned} E(X) &= \sum_{x=r}^{+\infty} x P(X = x) \\ &= \sum_{x=r}^{+\infty} x C_{x-1}^{-1} p^r (1-p)^{x-r} \\ &= \sum_{x=r}^{+\infty} r C_x^r p^r (1-p)^{x-r} \\ &= \frac{r}{p} \sum_{x=r}^{+\infty} C_x^r p^{r+1} (1-p)^{x-r} \\ &= \frac{r}{p} \sum_{y=r+1}^{+\infty} C_{y-1}^r p^{r+1} (1-p)^{y-1-r} \\ &= \frac{r}{p}. \end{aligned}$$

The last equality holds since $Y \sim \text{NB}(r+1, p)$. Similarly,

$$\begin{aligned} E(X^2) &= \sum_{x=r}^{+\infty} x^2 P(X = x) \\ &= \sum_{x=r}^{+\infty} (x+1)x P(X = x) - \sum_{x=r}^{+\infty} x P(X = x) \\ &= \sum_{x=r}^{+\infty} (x+1)x C_{x-1}^{-1} p^r (1-p)^{x-r} - \frac{r}{p} \\ &= \sum_{x=r}^{+\infty} (r+1)r C_{x+1}^{r+1} p^r (1-p)^{x-r} - \frac{r}{p} \\ &= \frac{(r+1)r}{p^2} \sum_{x=r}^{+\infty} C_{x+1}^{r+1} p^{r+2} (1-p)^{x-r} - \frac{r}{p} \\ &= \frac{(r+1)r}{p^2} \sum_{y=r+2}^{+\infty} C_{y-1}^{r+1} p^{r+2} (1-p)^{y-2-r} - \frac{r}{p} \\ &= \frac{(r+1)r}{p^2} - \frac{r}{p}. \end{aligned}$$

The last equality holds since $Y \sim \text{NB}(r+2, p)$. Therefore,

$$V(X) = \frac{(r+1)r}{p^2} - \frac{r}{p} - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}.$$

Theorem 4.3.6 Suppose $X \sim U[a, b]$, then $E(X) = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12}$.

Proof: X follows the uniform distribution on the interval $[a, b]$. Its probability density function is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}.$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^2 + ab + a^2}{3}.$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Theorem 4.3.7 Suppose $X \sim E(\lambda)$, then $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$.

Proof: The PDF of X is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt = \frac{1}{\lambda}.$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \int_0^{+\infty} t^2 e^{-t} dt = \frac{2}{\lambda^2}.$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

Theorem 4.3.8 Suppose $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$, $V(X) = \sigma^2$.

Proof: The PDF of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

where $-\infty < \mu < +\infty$ and $\sigma > 0$ are two parameters.

Let $Y = \frac{X-\mu}{\sigma}$, then $Y \sim N(0, 1)$. The PDF of Y is

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad -\infty < y < +\infty.$$

$$E(Y) = \int_{-\infty}^{+\infty} y \varphi(y) dy = \int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 0.$$

$$V(Y) = E(Y^2) - [E(Y)]^2$$

$$= \int_{-\infty}^{+\infty} y^2 \varphi(y) dy$$

$$= \int_{-\infty}^{+\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} y^2 e^{-\frac{y^2}{2}} dy \quad (\text{let } t = \frac{y^2}{2})$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{+\infty} t^{\frac{1}{2}} e^{-t} dt$$

$$= \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$$= \frac{2}{\sqrt{\pi}} \times \frac{1}{2} \times \sqrt{\pi}$$

$$= 1.$$

Therefore, $E(X) = E(\sigma Y + \mu) = \mu$, $V(X) = V(\sigma Y + \mu) = \sigma^2$.

The two parameters of normal distribution are its mean and variance respectively.

Normal distribution is completely determined by its mean and variance.

Theorem 4.3.9 Suppose X follows the gamma distribution $X \sim \text{Gamma}(\alpha, \beta)$, then $E(X) = \alpha\beta, V(X) = \alpha\beta^2$.

Proof: The PDF of X is given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0 & , x \leq 0 \end{cases}, \quad \alpha, \beta > 0.$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \int_0^{+\infty} x \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^\alpha e^{-\frac{x}{\beta}} dx \quad (\text{let } t = \frac{x}{\beta}) \\ &= \frac{\beta}{\Gamma(\alpha)} \int_0^{+\infty} t^\alpha e^{-t} dt \\ &= \frac{\beta}{\Gamma(\alpha)} \Gamma(\alpha + 1) \\ &= \alpha\beta. \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\ &= \int_0^{+\infty} x^2 \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^{\alpha+1} e^{-\frac{x}{\beta}} dx \quad (\text{let } t = \frac{x}{\beta}) \\ &= \frac{\beta^2}{\Gamma(\alpha)} \int_0^{+\infty} t^{\alpha+1} e^{-t} dt \\ &= \frac{\beta^2}{\Gamma(\alpha)} \Gamma(\alpha + 2) \\ &= \beta^2 (\alpha + 1) \alpha. \end{aligned}$$

$$V(X) = E(X^2) - [E(X)]^2 = \beta^2 (\alpha + 1) \alpha - (\alpha\beta)^2 = \alpha\beta^2.$$

Remark 4.3.1 If $\alpha = \frac{n}{2}, \beta = 2$, the PDF of $X \sim \text{Gamma}(\alpha, \beta)$ reduces to

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0 & , x \leq 0 \end{cases}.$$

It is the χ^2 -distribution with degrees of freedom n , an important distribution in mathematical statistics. If $X \sim \chi^2(n)$, then $E(X) = n, V(X) = 2n$.

◆ 4.4 Chebyshev's Inequality ◆

Theorem 4.4.1 Let X be a random variable with mean $E(X)$, and variance $V(X) =$

σ^2 , then for any $\epsilon > 0$,

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

Other equivalent forms can be written for this inequality, by simple manipulation:

$$P(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}, P(|X - \mu| \geq n\sigma) \leq \frac{1}{n^2}.$$

Proof: We just prove the continuous case and leave the discontinuous case as an exercise. Suppose the PDF of X is $f(x)$, then

$$\begin{aligned} P(|X - \mu| \geq \epsilon) &= \int_{|x - \mu| \geq \epsilon} f(x) dx \\ &\leq \int_{|x - \mu| \geq \epsilon} \frac{|x - \mu|^2}{\epsilon^2} f(x) dx \\ &\leq \frac{1}{\epsilon^2} \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \\ &= \frac{\sigma^2}{\epsilon^2}. \end{aligned}$$

Remark 4.4.1 Chebyshev's inequality provides a method to approximate the probabilities of $\{|X - \mu| < \epsilon\}$ or $\{|X - \mu| \geq \epsilon\}$ when the distribution of X is unknown.

Example 4.4.1 The probability that a radish seed will germinate is $\frac{1}{6}$. Estimate the probability that of 600 randomly selected seeds, the absolute value of the difference between the fraction of seeds that will germinate and $\frac{1}{6}$ is not larger than 0.02.

Solution: Let X be the number of seeds that will germinate of the 600 selected seeds, then $X \sim B(600, \frac{1}{6})$.

$$E(X) = 600 \times \frac{1}{6} = 100, V(X) = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{250}{3}.$$

By Chebyshev's inequality,

$$\begin{aligned} P\left(\left|\frac{X}{600} - \frac{1}{6}\right| \leq 0.02\right) &= P\left(\left|\frac{X - 100}{600}\right| \leq 0.02\right) \\ &= P(|X - 100| \leq 12) \\ &\geq 1 - \frac{V(X)}{12^2} \\ &= 1 - \frac{250/3}{144} \\ &= 0.4213. \end{aligned}$$

Remark 4.4.2 The true probability $P(|X - 100| \leq 12) = P(88 \leq X \leq 112) = \sum_{k=88}^{112} C_{600}^k \left(\frac{1}{6}\right)^k \left(1 - \frac{1}{6}\right)^{600-k} = 0.8294$. Since Chebyshev's inequality makes no assumptions on the shape of the distribution, it is something of a drawback. Although true, it is far too general to be of practical use, and is therefore mainly of theoretical interest. The

probability considered above is much higher than the one provided by Chebyshev. Chebyshev's result is true, but very conservative, when compared with the actual value.

◆ 4.5 Covariance and Correlation Coefficient ◆

Definition 4.5.1 Let X and Y be random variables having finite means. Let $E(X) = \mu_X$ and $E(Y) = \mu_Y$. The covariance of X and Y is defined as

$$\sigma_{XY} = \text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E(X)E(Y).$$

Specially, if $X = Y$, $\text{Cov}(X, X) = V(X)$.

The covariance between two random variables is a measure of the nature of the association between the two. The covariance only describes the linear relationship between the two random variables.

Theorem 4.5.1 If X and Y are random variables such that $V(X) < +\infty$ and $V(Y) < +\infty$, let a, b and c be constants, then

$$V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X, Y),$$

$$V(aX+bY+c) = a^2V(X) + b^2V(Y) + 2ab\text{Cov}(X, Y).$$

Proposition 4.5.1 The covariance has the following properties: for random variables X, Y, Z and constants a, b ,

- (1) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- (2) $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$;
- (3) $\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

Definition 4.5.2 Let X and Y be random variables with finite variances $V(X)$ and $V(Y)$, respectively. Then the correlation coefficient of X and Y is defined as follows:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}.$$

Remark 4.5.1 Let $X^* = \frac{X-E(X)}{\sqrt{V(X)}}$ and $Y^* = \frac{Y-E(Y)}{\sqrt{V(Y)}}$, then $E(X^*) = E(Y^*) = 0$, $V(X^*) = V(Y^*) = 1$, X^* and Y^* are called standardized variables of X and Y respectively. $\rho_{XY} = \text{Cov}(X^*, Y^*)$. ρ_{XY} is free of units of X and Y .

Theorem 4.5.2 The correlation coefficient has the following properties:

- (1) $-1 \leq \rho_{XY} \leq 1$;
- (2) $|\rho_{XY}| = 1$ if and only if there exist constants a, b such that $P(Y = a + bX) = 1$, and $\rho_{XY} = 1$ for $b > 0$, $\rho_{XY} = -1$ for $b < 0$.

Proof: (1) Let $X^* = \frac{X-E(X)}{\sqrt{V(X)}}$ and $Y^* = \frac{Y-E(Y)}{\sqrt{V(Y)}}$, since $V(X^* \pm Y^*) = V(X^*) + V(Y^*) \pm 2\text{Cov}(X^*, Y^*) = 2 \pm 2\rho_{XY} \geq 0$, therefore, $1 \pm \rho_{XY} \geq 0$, $-1 \leq \rho_{XY} \leq 1$.

(2) Necessity. If $\rho_{XY} = 1$, then $V(X^* - Y^*) = 2 - 2\rho_{XY} = 0$. Since $E(X^* - Y^*) = E(X^*) - E(Y^*) = 0$, using the property of variance, we have

$$P(X^* - Y^* = 0) = 1.$$

$$P\left[\frac{X - E(X)}{\sqrt{V(X)}} - \frac{Y - E(Y)}{\sqrt{V(Y)}} = 0\right] = 1.$$

Equivalently,

$$P\left[Y = \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}X + E(Y) - \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}E(X)\right] = 1.$$

Take $b = \frac{\sqrt{V(Y)}}{\sqrt{V(X)}} > 0$, $a = E(Y) - \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}E(X)$. Then

$$P(Y = a + bX) = 1.$$

Similarly, if $\rho_{XY} = -1$,

$$P\left[\frac{X - E(X)}{\sqrt{V(X)}} + \frac{Y - E(Y)}{\sqrt{V(Y)}} = 0\right] = 1.$$

Take $b = -\frac{\sqrt{V(Y)}}{\sqrt{V(X)}} < 0$, $a = E(Y) + \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}E(X)$. Then

$$P(Y = a + bX) = 1.$$

Sufficiency. If there exist constants a and $b \neq 0$, $P(Y = a + bX) = 1$, then $V(Y) = V(a + bX) = b^2 V(X)$, $\text{Cov}(X, Y) = \text{Cov}(X, a + bX) = bV(X)$, therefore,

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)} \sqrt{V(Y)}} = \frac{b}{|b|}.$$

That is $|\rho_{XY}| = 1$, and if $b > 0$, $\rho_{XY} = 1$; if $b < 0$, $\rho_{XY} = -1$.

Correlation coefficient near 1 means that X and Y are typically big together or small together. Correlation coefficient near -1 means that when X is big, Y is small (and vice versa).

Example 4.5.1 Let X and Y be two random variables having the following joint PF $f(x, y)$.

$f(x, y)$	$x=0$	$x=1$	$x=2$	$f_Y(y)$
$y=0$	$\frac{1}{6}$	$\frac{2}{9}$	$\frac{1}{36}$	$\frac{5}{12}$
$y=1$	$\frac{1}{3}$	$\frac{1}{6}$	0	$\frac{1}{2}$
$y=2$	$\frac{1}{12}$	0	0	$\frac{1}{12}$
$f_X(x)$	$\frac{7}{12}$	$\frac{7}{18}$	$\frac{1}{36}$	1

(1) Find the covariance of X and Y .

(2) Find the correlation coefficient of X and Y .

Solution:

$$\begin{aligned} E(XY) &= \sum_x \sum_y (xy) f(x, y) \\ &= (0 \times 0) \times \frac{1}{6} + (0 \times 1) \times \frac{1}{3} + (0 \times 2) \times \frac{1}{12} + \\ &\quad (1 \times 0) \times \frac{2}{9} + (1 \times 1) \times \frac{1}{6} + (1 \times 2) \times 0 + \\ &\quad (2 \times 0) \times \frac{1}{36} + (2 \times 1) \times 0 + (2 \times 2) \times 0 \\ &= \frac{1}{6}. \end{aligned}$$

The marginal PF of X is as follows,

x	0	1	2
$f_X(x) = P(X=x)$	$\frac{7}{12}$	$\frac{7}{18}$	$\frac{1}{36}$

$$\begin{aligned}
 E(X) &= \sum_x x f_X(x) \\
 &= 0 \times \frac{7}{12} + 1 \times \frac{7}{18} + 2 \times \frac{1}{36} \\
 &= \frac{4}{9}.
 \end{aligned}$$

$$\begin{aligned}
 E(X^2) &= \sum_x x^2 f_X(x) \\
 &= 0^2 \times \frac{7}{12} + 1^2 \times \frac{7}{18} + 2^2 \times \frac{1}{36} \\
 &= \frac{1}{2}.
 \end{aligned}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{4}{9}\right)^2 = \frac{49}{162}.$$

The marginal PF of Y is as follows,

y	0	1	2
$f_Y(y) = P(Y=y)$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

$$\begin{aligned}
 E(Y) &= \sum_y y f_Y(y) \\
 &= 0 \times \frac{5}{12} + 1 \times \frac{1}{2} + 2 \times \frac{1}{12} \\
 &= \frac{2}{3}.
 \end{aligned}$$

$$\begin{aligned}
 E(Y^2) &= \sum_y y^2 f_Y(y) \\
 &= 0^2 \times \frac{5}{12} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{12} \\
 &= \frac{5}{6}.
 \end{aligned}$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{5}{6} - \left(\frac{2}{3}\right)^2 = \frac{7}{18}.$$

The covariance is

$$\begin{aligned}
 \sigma_{XY} &= \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \\
 &= \frac{1}{6} - \frac{4}{9} \times \frac{2}{3} \\
 &= -\frac{7}{54}.
 \end{aligned}$$

The correlation coefficient is

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{-\frac{7}{54}}{\sqrt{\frac{49}{162}} \times \sqrt{\frac{7}{18}}} = -\frac{\sqrt{7}}{7}.$$

Example 4.5.2 Let X and Y be two random variables having the joint PDF

$$f(x, y) = \begin{cases} 2, & 0 < x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Find the correlation coefficient between X and Y .

Solution: see Figure 4.5.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dx dy \\ &= \int_0^1 \left(\int_0^y xy \times 2 dx \right) dy + 0 \\ &= \int_0^1 (y \times x^2 \big|_{x=0}^{x=y}) dy \\ &= \int_0^1 y^3 dy \\ &= \frac{y^4}{4} \bigg|_{y=0}^{y=1} \\ &= \frac{1}{4}. \end{aligned}$$

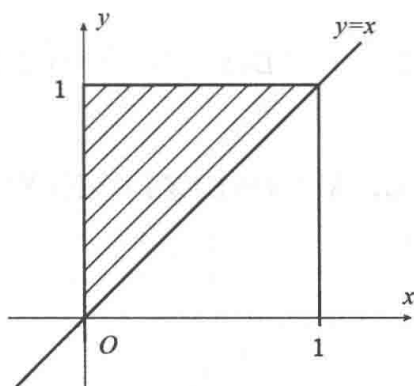


Figure 4.5

The marginal PDF of X is as follows.

(1) If $x \leq 0$ or $x \geq 1$, $f_X(x) = 0$.

(2) If $0 < x < 1$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ &= \int_x^1 2 dy \\ &= 2(1 - x). \end{aligned}$$

In summary,

$$f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 x \times 2(1-x) dx = \frac{1}{3}.$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 x^2 \times 2(1-x) dx = \frac{1}{6}.$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}.$$

The marginal PDF of Y is as follows.

(1) If $y \leq 0$ or $y \geq 1$, $f_Y(y) = 0$.

(2) If $0 < y < 1$,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \\ &= \int_0^y 2 dx \\ &= 2y. \end{aligned}$$

In summary,

$$f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \times 2y dy = \frac{2y^3}{3} \Big|_{y=0}^{y=1} = \frac{2}{3}.$$

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \int_0^1 y^2 \times 2y dy = \frac{y^4}{2} \Big|_{y=0}^{y=1} = \frac{1}{2}.$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

The covariance is

$$\begin{aligned} \sigma_{XY} &= \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \\ &= \frac{1}{4} - \frac{1}{3} \times \frac{2}{3} \\ &= \frac{1}{36}. \end{aligned}$$

The correlation coefficient is

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\frac{1}{36}}{\sqrt{\frac{1}{18}} \times \sqrt{\frac{1}{18}}} = \frac{1}{2}.$$

Note: in Example 4.5.2, we compute $E(XY) = \frac{1}{4}$ using the joint PDF of (X, Y) .

Here we propose another method to compute $E(XY)$. We want to find the PDF of $U = XY$ and use this PDF to compute $E(XY)$.

The joint PDF of (X, Y) is

$$f(x, y) = \begin{cases} 2, & 0 < x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

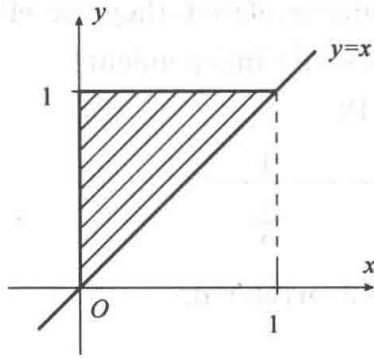
Let

$$\begin{cases} U = r_1(X, Y) = XY \\ V = r_2(X, Y) = X/Y \end{cases}$$

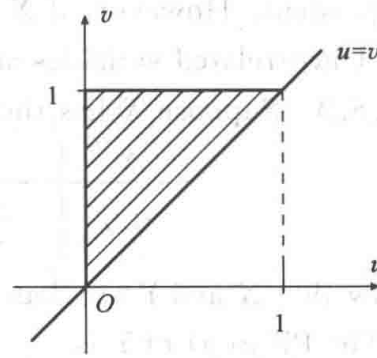
r_1, r_2 define a one-to-one differentiable transformation of $S = \{(x, y) | 0 < x \leq y < 1\}$ onto a subset T of $\mathbb{R}^2$. See Figure 4.6. The inverse functions can be found as

$$\begin{cases} X = s_1(U, V) = \sqrt{UV} \\ Y = s_2(U, V) = \sqrt{U/V} \end{cases}$$

Therefore, $U > 0, V > 0, 0 < \sqrt{UV} \leq \sqrt{U/V} < 1$, and $T = \{(u, v) | 0 < u < v \leq 1\}$.



(a) subset S



(b) subset T

Figure 4.6

$$J = \begin{vmatrix} \frac{\partial s_1}{\partial u} & \frac{\partial s_1}{\partial v} \\ \frac{\partial s_2}{\partial u} & \frac{\partial s_2}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2}u^{-\frac{1}{2}}v^{\frac{1}{2}} & \frac{1}{2}u^{\frac{1}{2}}v^{-\frac{1}{2}} \\ \frac{1}{2}u^{-\frac{1}{2}}v^{-\frac{1}{2}} & -\frac{1}{2}u^{\frac{1}{2}}v^{-\frac{3}{2}} \end{vmatrix} = -\frac{1}{2v}, |J| = \frac{1}{2v}.$$

Therefore, the joint PDF of (U, V) is

$$\begin{aligned} g(u, v) &= \begin{cases} f(s_1, s_2) \times |J|, & (u, v) \in T \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{v}, & (u, v) \in T \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Find the PDF $g_1(u)$ of $U = XY$.

(1) If $u \leq 0$ or $u \geq 1$, $g_1(u) = 0$.

(2) If $0 < u < 1$,

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{+\infty} g(u, v) dv \\ &= \int_u^1 \frac{1}{v} dv \\ &= [\ln v]_{v=u}^{v=1} \\ &= -\ln u. \end{aligned}$$

The PDF of $U = XY$ is

$$g_1(u) = \begin{cases} -\ln u, & 0 < u < 1 \\ 0, & \text{otherwise} \end{cases}$$

Therefore,

$$E(XY) = E(U) = \int_0^1 u \times (-\ln u) du = \frac{1}{4}.$$

Definition 4.5.3 X and Y are positive correlated if $\text{Cov}(X, Y) > 0$. X and Y are negative correlated if $\text{Cov}(X, Y) < 0$. X and Y are uncorrelated if $\text{Cov}(X, Y) = 0$.

Remark 4.5.2 If X and Y are uncorrelated, it means that X and Y have no linear relationship. It is possible that there exists other relationships between the two variables.

By Theorem 4.1.6, if X and Y are independent, then $E(XY) = E(X)E(Y)$, $\text{Cov}(X, Y) = 0$. That is, independence implies uncorrelation. Equivalently, if $\text{Cov}(X, Y) \neq 0$ or $\rho_{XY} \neq 0$, then X and Y are dependent. However, if X and Y are uncorrelated, they are either dependent or independent. Uncorrelated variables are not necessarily independent!

Example 4.5.3 Suppose X has the following PF

X	-1	0	1
$f(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

and $Y = X^2$. Show that X and Y are dependent but uncorrelated.

Solution: The PF $g(y)$ of Y is

Y	1	0
$g(y)$	$\frac{2}{3}$	$\frac{1}{3}$

Since $P(Y=1, X=1) = P(X^2=1, X=1) = P(X=1) = \frac{1}{3}$, $P(Y=1) = \frac{2}{3}$, $P(X=1) = \frac{1}{3}$, $P(Y=1, X=1) \neq P(Y=1) \times P(X=1)$, X and Y are dependent.

On the other hand, $E(X) = (-1) \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$, $E(XY) = E(X^3) = (-1)^3 \times \frac{1}{3} + 0^3 \times \frac{1}{3} + 1^3 \times \frac{1}{3} = 0$, $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$. X and Y are uncorrelated.

Example 4.5.4 Suppose (X, Y) is uniformly distributed over the unit circle $\{(x, y) | x^2 + y^2 \leq 1\}$, show that X and Y are dependent but uncorrelated.

Solution: The joint PDF of (X, Y) is

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

The support set is shown in Figure 4.7.

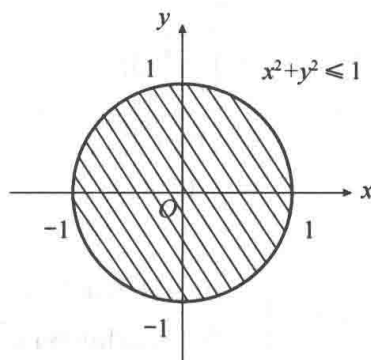


Figure 4.7

$$\begin{aligned}
 E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dx dy \\
 &= \iint_{x^2+y^2 \leq 1} xy \frac{1}{\pi} dx dy \\
 &= \int_0^{2\pi} \left[\int_0^1 (r \cos \theta) (r \sin \theta) \frac{1}{\pi} r dr \right] d\theta \\
 &= \frac{1}{8\pi} \int_0^{2\pi} \sin(2\theta) d\theta \\
 &= 0. \\
 E(X) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y) dx dy \\
 &= \iint_{x^2+y^2 \leq 1} x \frac{1}{\pi} dx dy \\
 &= \int_{-1}^1 \left(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{x}{\pi} dx \right) dy \\
 &= 0.
 \end{aligned}$$

Similarly, $E(Y) = 0$, therefore, $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$, X and Y are uncorrelated.

Determine the marginal PDF $f_X(x)$ of X :

- (1) If $x < -1$ or $x > 1$, $f_X(x) = 0$.
- (2) If $-1 \leq x \leq 1$,

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi}.$$

In summary,

$$f_X(x) = \begin{cases} \frac{2\sqrt{1-x^2}}{\pi}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Similarly, the PDF of Y is

$$f_Y(y) = \begin{cases} \frac{2\sqrt{1-y^2}}{\pi}, & -1 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Since $f(x, y) \neq f_X(x)f_Y(y)$, X and Y are dependent.

Example 4.5.5 Suppose X and Y have the bivariate normal distribution $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Determine ρ_{XY} .

Solution: The joint PDF of (X, Y) is of the form

$$\begin{aligned}
 f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \times \right. \\
 &\quad \left. \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\}, \\
 &\quad -\infty < x, y < +\infty.
 \end{aligned}$$

And the marginal PDF of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, \quad -\infty < y < +\infty.$$

For any $y, f_Y(y) > 0$,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2\right\}. \end{aligned}$$

That is $X|Y=y \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right)$. $\frac{(X|Y=y) - \left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]}{\sigma_1\sqrt{1-\rho^2}} \sim$

$N(0,1)$. We can compute

$$\begin{aligned} \int_{-\infty}^{+\infty} (x - \mu_1) f_{X|Y}(x|y) dx &= \int_{-\infty}^{+\infty} [z\sigma_1(1-\rho^2) + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)] \varphi(z) dz \\ &= \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2). \end{aligned}$$

Then we can apply it to obtain

$$\begin{aligned} \text{Cov}(X,Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_1)(y - \mu_2) f(x,y) dx dy \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} (x - \mu_1) f_{X|Y}(x|y) dx \right] (y - \mu_2) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)^2 f_Y(y) dy \\ &= \rho\sigma_1\sigma_2. \end{aligned}$$

$$\text{Therefore, } \rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho.$$

The parameter ρ is exactly the correlation coefficient between X and Y . In previous Chapter, we know the fact: suppose X and Y have the bivariate normal distribution $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Then X and Y are independent if and only if $\rho=0$. So we obtain the following property.

Proposition 4.5.2 Suppose X and Y have the bivariate normal distribution $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Then X and Y are independent if and only if they are uncorrelated.

◆ 4.6 Moment and Covariance Matrix ◆

Definition 4.6.1 Suppose X is a random variable with finite $E(X^k)$ ($k=1,2,\dots$), then $E(X^k)$ is the k -th moment about the origin of X . If $E\{[X-E(X)]^k\}$ ($k=2,3,\dots$) exists, $E[X-E(X)]^k$ is called the k -th central moment of X . If $E\{[X-E(X)]^k[Y-E(Y)]^l\}$ exists, it is called the (k,l) -th joint central moment of X and Y .

$E(Y)]^l\}$ ($k, l=1, 2, \dots$) exists, $E\{[X-E(X)]^k[Y-E(Y)]^l\}$ is called the mixed central moment of order $k+l$ of X and Y .

The mean $E(X)$ is the first moment about the origin of X . The variance $V(X)$ is the second central moment of X . The covariance $\text{Cov}(X, Y)$ is the second bivariate mixed central moment of X and Y .

Example 4.6.1 Suppose $Z \sim N(0, 1)$, find $E(Z^n)$.

Solution: The PDF of Z is $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $-\infty < z < +\infty$.

If n is odd, $E(Z^n) = \int_{-\infty}^{+\infty} z^n \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0$.

If n is even,

$$\begin{aligned} E(Z^n) &= \int_{-\infty}^{+\infty} z^n \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} z^n e^{-\frac{z^2}{2}} dz \quad \left(\text{let } t = \frac{z^2}{2} \right) \\ &= \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \int_0^{+\infty} t^{\frac{n-1}{2}} e^{-t} dt \\ &= \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) \\ &= \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \times \frac{n-1}{2} \times \frac{n-3}{2} \times \dots \times \frac{1}{2} \times \Gamma\left(\frac{1}{2}\right) \\ &= (n-1) \times (n-3) \times \dots \times 1. \end{aligned}$$

Specially, $E(Z^4) = 3$.

Definition 4.6.2 Suppose the second central moments of (X_1, X_2) exist, they are denoted as

$$\begin{aligned} c_{11} &= E[X_1 - E(X_1)]^2 = V(X_1), \\ c_{12} &= E\{[X_1 - E(X_1)][X_2 - E(X_2)]\} = \text{Cov}(X_1, X_2), \\ c_{21} &= E\{[X_2 - E(X_2)][X_1 - E(X_1)]\} = \text{Cov}(X_2, X_1), \\ c_{22} &= E[X_2 - E(X_2)]^2 = V(X_2). \end{aligned}$$

Then

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

is defined as the covariance matrix of (X_1, X_2) .

Similarly, the covariance matrix of $(X_1, X_2, \dots, X_n)$ is defined as

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix},$$

where $c_{ij} = \text{Cov}(X_i, X_j)$ ($i, j = 1, 2, \dots, n$). Since $c_{ij} = c_{ji}$, $i \neq j$ ($i, j = 1, 2, \dots, n$), the

covariance matrix is symmetric.

Using covariance matrix, we can rewrite the PDF of bivariate normal distribution. Let $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then the covariance matrix of (X_1, X_2) is

$$\mathbf{C} = \begin{bmatrix} V(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & V(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Note that if $|\mathbf{C}| \neq 0$, the inverse of $\mathbf{C}$ exists, (X, Y) is said to be non-degenerate. If $|\mathbf{C}| = 0$, (X, Y) is degenerate and has no density function. For non-degenerate case, the PDF of (X_1, X_2) is

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\} \\ &= \frac{1}{(2\pi)^{\frac{2}{2}} (|\mathbf{C}|)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{X}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X}-\boldsymbol{\mu}) \right\}, \end{aligned}$$

where $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $|\mathbf{C}|$ and $\mathbf{C}^{-1}$ are the determinant and inverse of $\mathbf{C}$ respectively, $(\mathbf{X}-\boldsymbol{\mu})^T$ is the transpose of $(\mathbf{X}-\boldsymbol{\mu})$.

The above result can be easily extended to n -dimensional normal random vector $(X_1, X_2, \dots, X_n)$, the PDF is

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} (|\mathbf{C}|)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{X}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X}-\boldsymbol{\mu}) \right\},$$

where $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$, $|\mathbf{C}|$ and $\mathbf{C}^{-1}$ are the determinant and inverse of the covariance matrix $\mathbf{C}$ respectively, $(\mathbf{X}-\boldsymbol{\mu})^T$ is the transpose of $(\mathbf{X}-\boldsymbol{\mu})$.

n -dimensional normal random vector $(X_1, X_2, \dots, X_n)$ has the following properties:

(1) Given an n -dimensional normal random vector $(X_1, \dots, X_n)$, all of its components $X_i (i=1, \dots, n)$ are normal random variables; Conversely, if $X_i (i=1, \dots, n)$ is normal random variables and independent, then $(X_1, \dots, X_n)$ is an n -dimensional normal random vector.

(2) $(X_1, \dots, X_n)$ follows n -dimensional normal random vector if and only if any of its linear combinations $a_1 X_1 + \dots + a_n X_n$ follows normal distribution ($a_1, \dots, a_n$ are not all zeros).

(3) If $(X_1, \dots, X_n)$ follows n -dimensional normal random vector, $Y_1, Y_2, \dots, Y_m$ are linear functions of $X_i (i=1, \dots, n)$ (coefficients are not all zeros), then $(Y_1, Y_2, \dots, Y_m)$ follows m -dimensional normal distribution.

Note: this property (3) shows the invariance of normal distributions under linear transformations. The m -dimensional normal distribution above is possibly degenerate and has no density function.

Example 4.6.2 Suppose $X \sim N(0, 1)$, $Y_1 = X$, $Y_2 = X$, consider $(Y_1, Y_2) = (X, X)$, the covariance matrix is

$$C = \begin{bmatrix} V(Y_1) & \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_2, Y_1) & V(Y_2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Clearly, $|C| = 0$, $(Y_1, Y_2) = (X, X)$ is a degenerate normal random variable and has no density function.

How to obtain a non-degenerate normal distribution after linear transformation? We have the following result.

If $\mathbf{X} = (X_1, \dots, X_n)$ follows n -dimensional non-degenerate normal distribution $N_n(\boldsymbol{\mu}, C)$, matrix $A_{m \times n}$ is full row rank $r(A) = m$, then $\mathbf{Y} = A\mathbf{X}$ follows m -dimensional non-degenerate normal distribution.

$$\mathbf{Y} \sim N_m(A\boldsymbol{\mu}, ACA^T)$$

(4) Suppose $(X_1, \dots, X_n)$ follows n -dimensional normal distribution, then $X_1, \dots, X_n$ are independent if and only if they are uncorrelated.

Example 4.6.3 Suppose $(X, Y) \sim N\left(1, 0, 9, 16, -\frac{1}{2}\right)$, $Z = \frac{X}{3} + \frac{Y}{2}$, are X and Z independent?

Solution: $X \sim N(1, 9)$, $Y \sim N(0, 16)$, $\rho_{XY} = -\frac{1}{2}$.

$$\begin{aligned} \text{Cov}(X, Z) &= \text{Cov}\left(X, \frac{X}{3} + \frac{Y}{2}\right) \\ &= \frac{1}{3}V(X) + \frac{1}{2}\text{Cov}(X, Y) \\ &= 3 + \frac{1}{2} \times \left(-\frac{1}{2}\right) \times \sqrt{9} \times \sqrt{16} \\ &= 0. \end{aligned}$$

That is, X, Z are uncorrelated. On the other hand, $\begin{bmatrix} X \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$, matrix $A =$

$\begin{bmatrix} 1 & 0 \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix}$ is full row rank, $|A| \neq 0$, $r(A) = 2$, therefore, (X, Z) follows two dimensional non-degenerate normal distribution.

X and Z are uncorrelated if and only if X and Z are independent, so X and Z are independent.

Theorem 4.6.1 If $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then

$$a_1X + a_2Y + b \sim N(a_1\mu_1 + a_2\mu_2 + b, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\rho\sigma_1\sigma_2).$$

Example 4.6.4 If $(X_1, X_2) \sim N(66.8, 70, 2^2, 2^2, 0.68)$, find $P(X_1 > X_2)$.

Solution: Since X_1 and X_2 have a bivariate normal distribution, it follows that the distribution $X_1 - X_2$ will be the normal distribution.

$$E(X_1 - X_2) = 66.8 - 70 = -3.2.$$

$$V(X_1 - X_2) = V(X_1) + V(X_2) - 2\text{Cov}(X_1, X_2) = 2.56.$$

So $X_1 - X_2 \sim N(-3.2, 2.56)$,

$$\begin{aligned} P(X_1 > X_2) &= P(X_1 - X_2 > 0) \\ &= P\left[\frac{X_1 - X_2 - (-3.2)}{\sqrt{2.56}} > \frac{0 - (-3.2)}{\sqrt{2.56}}\right] \\ &= P(Z > 2) \\ &= 1 - \Phi(2) \\ &= 0.0227. \end{aligned}$$

◆ Exercise 4 ◆

4.1 Suppose a discrete random variable X has the PF given by

X	3	5	7	8	9	10	12	16
$P(X=x)$	0.08	0.10	0.16	0.25	0.20	0.03	0.13	0.05

Find (1) $E(X)$; (2) $V(X)$; (3) $E(X^2 - 4X)$.

4.2 Let X be a continuous random variable with the following PDF

$$f(x) = \begin{cases} x, & 0 < x \leq 1 \\ 2-x, & 1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Determine (1) $E(X)$; (2) $V(X)$.

4.3 The PDF of X is $f(x) = \begin{cases} a+bx^2, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$. If $E(X) = 3/5$, find a and b .

4.4 Let X be a continuous random variable with the following PDF

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Let $Y = X + e^{-2X}$. Determine (1) $E(Y)$; (2) $V(Y)$.

4.5 Let X and Y be two random variables having the following joint PF.

	$Y=1$	$Y=2$	$Y=3$
$X=-1$	0	$\frac{1}{15}$	$\frac{3}{15}$
$X=0$	$\frac{2}{15}$	$\frac{5}{15}$	$\frac{4}{15}$

Determine (1) $E(X)$, $V(X)$; (2) $E(Y)$, $V(Y)$; (3) $E(XY)$, $V(XY)$.

4.6 Suppose that X and Y have a continuous joint distribution for which the joint PDF is as follows:

$$f(x, y) = \begin{cases} 12y^2, & 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $E(XY)$ and $V(XY)$.

4.7 Suppose that X and Y have a continuous joint distribution for which the joint

PDF is as follows:

$$f(x, y) = \begin{cases} 3x, & 0 < x < 1, 0 < y < x \\ 0, & \text{otherwise} \end{cases}$$

Determine (1) $E(X), V(X)$; (2) $E(Y), V(Y)$.

4.8 Suppose that a point is chosen at random on a stick of unit length and that the stick is broken into two pieces at that point. Find the expected value of the length of the longer piece.

4.9 Suppose that the random variables $X_1, \dots, X_n$ form a random sample of size n from the uniform distribution on the interval $[0, 1]$. Let $Y_1 = \min\{X_1, \dots, X_n\}$, and let $Y_n = \max\{X_1, \dots, X_n\}$. Find $E(Y_1)$ and $E(Y_n)$.

4.10 Suppose that $X_1, \dots, X_n$ form a random sample of size n from a continuous distribution with the following PDF:

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Let $Y_n = \max\{X_1, \dots, X_n\}$. Evaluate $E(Y_n)$.

4.11 Suppose that X and Y are IID and $X \sim B(1, 0.4)$. Find $E(\max\{X, Y\})$ and $E(\min\{X, Y\})$.

4.12 A person has n keys and one lock. After every failed attempt to unlock, the person may pick a new key and discard the old one. What is the expected number of attempts needed to open the lock?

4.13 A fair die is rolled 20 times. Find the expected sum of the 20 rolls.

4.14 Suppose that two persons A and B each randomly, and independently, choose 3 of 10 objects. Find the expected number of objects (1) chosen by both A and B; (2) not chosen by either A or B; (3) chosen by exactly one of A or B.

4.15 Six couples go to dinner together. The waiter seats the men randomly on one side of the table and the women randomly on the other side of the table. Find the expected value and variance of the number of couples who are seated across from each other.

4.16 A total of n balls, numbered 1 through n , are put into n urns, also numbered 1 through n in such a way that ball i is equally likely to go into any of the urns $1, 2, \dots, i$. Find the expected number of empty urns.

4.17 Suppose that three random variables X_1, X_2, X_3 form a random sample from the uniform distribution on $[0, 1]$. Determine $E[(X_1 - 2X_2 + X_3)^2]$.

4.18 Suppose that the random variable X has the uniform distribution on the interval $[0, 1]$, that the random variable Y has the uniform distribution on the interval $[5, 9]$, and that X and Y are independent. Suppose also that a rectangle is to be constructed for which the lengths of two adjacent sides are X and Y . Determine the expected value of the area of the rectangle.

4.19 Suppose that a fair coin is tossed repeatedly until a head is obtained for the first time. (1) What is the expected number of tosses that will be required? (2) What is the expected number of tails that will be obtained before the first head is obtained?

4.20 Suppose that X and Y are independent random variables for which $X \sim N(0, 4)$, $Y \sim U(0, 4)$. Find the values of $V(2X+3Y)$ and $V(2X-3Y)$.

4.21 Suppose $X \sim B(n, p)$. $E(X) = 2.4$, $V(X) = 1.44$. Find n and p .

4.22 Suppose $X \sim U(a, b)$. $E(X) = 3$, $V(X) = \frac{1}{3}$. Find $P(1 < X < 3)$.

4.23 If X_1, X_2, X_3, X_4 are (pairwise) uncorrelated random variables each having mean 0 and variance 1, compute the correlations of (1) $X_1 + X_2$ and $X_2 + X_3$; (2) $X_1 + X_2$ and $X_3 + X_4$.

4.24 Let X and Y have a continuous distribution with the joint PDF

$$f(x, y) = \begin{cases} x+y, & 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Compute the covariance $\text{Cov}(X, Y)$.

4.25 The random variables X and Y have a joint PDF given by

$$f(x, y) = \begin{cases} \frac{2e^{-2x}}{x}, & 0 < x < +\infty \text{ and } 0 \leq y < x \\ 0, & \text{otherwise} \end{cases}.$$

Compute the covariance $\text{Cov}(X, Y)$.

4.26 Suppose that X and Y have a continuous joint distribution for which the joint PDF is as follows:

$$f(x, y) = \begin{cases} \frac{1}{3}(x+y), & 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the value of $V(2X-3Y+8)$.

4.27 Suppose that X and Y are random variables such that $V(X) = 9$, $V(Y) = 4$, and $\rho_{XY} = -\frac{1}{6}$. Determine $V(X+Y)$ and $V(X-3Y+4)$.

4.28 Suppose that an automobile dealer pays an amount X (in thousands of dollars) for a used car and then sells it for an amount Y . Suppose that the random variables X and Y have the following joint PDF:

$$f(x, y) = \begin{cases} \frac{1}{36}x, & 0 < x < y < 6 \\ 0, & \text{otherwise} \end{cases}.$$

Determine the dealer's expected gain from the sale.

4.29 Four independent random variables X_1, X_2, X_3 and X_4 each normal distributed with a mean of 0 and a standard deviation of 4 are combined to form a new random variable

$$Y = X_1^2 + X_2^2 + X_3^2 + X_4^2.$$

(1) What is the expected value of Y ?

(2) What is the variance of Y ?

4.30 Let X and Y be random variables with $E(X) = 2$, $V(X) = 1$, $E(Y) = 3$, $V(Y) = 4$. What are the smallest and largest possible values of $V(X+Y)$?

4.31 Your company must make a sealed bid for a construction project. If you succeed in

winning the contract (by having the lowest bid), then you plan to pay another firm \$100,000 to do the work. If you believe that the maximum bid (in thousands of dollars) of the other participating companies can be modeled as being the value of a random variable that is uniformly distributed on (70,140), how much should you bid to maximize your expected profit?

4.32 Two independent random variables X and Y have the variances $V(X)=12$ and $V(Y)=18$. Let $W=-4X+2Y$ and $Z=3X-6Y$.

- (1) Find the variances of W and Z .
- (2) Find the correlation coefficient of W and Z .

4.33 Find the mean and variance of the distribution whose PDF is given as

$$f(x) = \frac{1}{\sqrt{128\pi}} e^{-\frac{(x+7)^2}{128}}.$$

4.34 Find the mean and variance of the distribution whose PDF is given as

$$f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2+2x-1}.$$

4.35 Let $\mathbf{X}$ be a 2×1 random vector and denote its components by X_1 and X_2 . The covariance matrix of $\mathbf{X}$ is

$$V(\mathbf{X}) = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}.$$

Compute the variance of the random variable $Y=3X_1+4X_2$.

4.36 In n independent Bernoulli trials, each with probability of success p , let X be the number of successes and Y be the number of failures. Calculate $E(XY)$ and $\text{Cov}(X,Y)$.

4.37 Use Chebyshev's inequality to find an upper bound on the probability that the number of tails that come up when a fair coin is tossed n times deviates from the mean by more than $5\sqrt{n}$.

4.38 The joint PDF of (X,Y) is given by

$$f(x,y) = \begin{cases} \frac{1}{4}(1-x^3y+xy^3), & -1 < x < 1, -1 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Show that X and Y are uncorrelated and dependent.

4.39 The joint PF of (X,Y) is given by

	$X=-2$	$X=-1$	$X=1$	$X=2$
$Y=1$	0	$\frac{1}{4}$	$\frac{1}{4}$	0
$Y=4$	$\frac{1}{4}$	0	0	$\frac{1}{4}$

Show that X and Y are uncorrelated and dependent.

4.40 Suppose A and B are two events. $P(A)>0, P(B)>0$, let

$$X = \begin{cases} 1, & A \text{ occurs} \\ -1, & A \text{ does not occur} \end{cases}, \quad Y = \begin{cases} 1, & B \text{ occurs} \\ -1, & B \text{ does not occur} \end{cases}.$$

Show that if $\rho_{XY}=0$, then X and Y are independent.

Chapter 5

Large Random Samples



Exercise 5.1. Suppose the P.D.F. of X is $f(x)$, then $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same distribution as X . Then $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ has the P.D.F. $f_n(x) = \frac{1}{\sigma\sqrt{2\pi n}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right)^2}$. Show that as $n \rightarrow \infty$, $f_n(x) \rightarrow \delta(x - \mu)$, where $\delta(x - \mu)$ is the Dirac delta function. (Hint: Use the fact that $\int_{-\infty}^{\infty} \delta(x - \mu) dx = 1$.)

Exercise 5.2. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same distribution as X . Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that as $n \rightarrow \infty$, $\frac{S_n^2}{n} \rightarrow \sigma^2$ in probability. (Hint: Use Chebyshev's inequality.)

Exercise 5.3. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same distribution as X . Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that as $n \rightarrow \infty$, $\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \rightarrow 0$ in probability. (Hint: Use Chebyshev's inequality.)

Exercise 5.4. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same distribution as X . Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that as $n \rightarrow \infty$, $\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \rightarrow 0$ in probability. (Hint: Use Chebyshev's inequality.)

Exercise 5.5. Let $X_1, X_2, \dots, X_n$ be independent random variables with the same distribution as X . Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that as $n \rightarrow \infty$, $\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \rightarrow 0$ in probability. (Hint: Use Chebyshev's inequality.)

◆ 5.1 The Law of Large Numbers ◆

Suppose a fair coin is tossed. $P(\text{Heads}) = \frac{1}{2}$, on each toss. We also assume that tosses of the coin are independent, so that whether we obtain head or tail on any one toss is not influenced by the results on other tosses.

Let $X_i = 1$ or 0 ($i = 1, 2, \dots$) according as the i -th toss results in head or tail, respectively. Let $S_n = X_1 + X_2 + \dots + X_n$ be the number of heads seen in the first n tosses. And let $\bar{X}_n = \frac{S_n}{n}$ be the proportion of heads in the first n tosses. From Figure 5.1, we can see that for large enough n , $\bar{X}_n$ converges to $\frac{1}{2}$. The probability of a certain outcome is as the frequency with which that outcome occurs in the long run, when the experiment is repeated a large number of times.

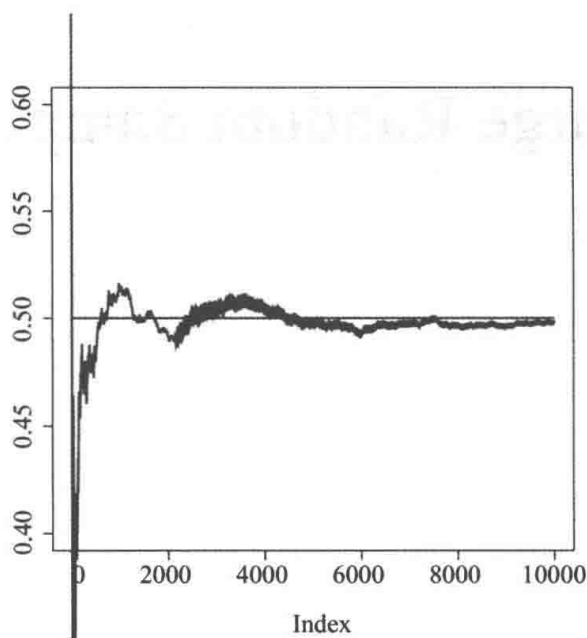


Figure 5.1

Theorem 5.1.1 (Chebyshev's Inequality) Let X be a random variable for which $V(X)$ exists, then for every number $t > 0$,

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}.$$

Proof: We only prove the continuous case and leave the discontinuous case as an exercise. Suppose the PDF of X is $f(x)$, then

$$\begin{aligned} P(|X - \mu| \geq t) &= \int_{|x - \mu| \geq t} f(x) dx \\ &\leq \int_{|x - \mu| \geq t} \frac{|x - \mu|^2}{t^2} f(x) dx \\ &\leq \frac{1}{t^2} \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \\ &= \frac{V(X)}{t^2}. \end{aligned}$$

Theorem 5.1.2 Let $X_1, \dots, X_n$ be a random sample from a distribution with mean μ and variance σ^2 . Let $\bar{X}$ be the sample mean. Then for any $\epsilon > 0$,

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - \mu| < \epsilon) = 1.$$

Proof: Since

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n},$$

by Chebyshev's inequality, for any $\epsilon > 0$, we have

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2},$$

or

$$P(|\bar{X} - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}.$$

Therefore,

$$\lim_{n \rightarrow +\infty} P(|\bar{X} - \mu| < \epsilon) = 1.$$

The above result is referred as the law of large numbers. As the sample size n increases, the sample mean converges to the population mean.

Example 5.1.1 Suppose that a random sample is to be taken from a distribution for which the value of the mean μ is not known, but the standard deviation $\sigma \leq 2$ units. We shall determine how large the sample size such that

$$P(|\bar{X} - \mu| < 1) \geq 0.99.$$

Solution: From Chebyshev's inequality,

$$P(|\bar{X} - \mu| \geq 1) \leq \frac{\sigma^2}{n} \leq \frac{4}{n},$$

it follows that n must be chosen so that $\frac{4}{n} \leq 0.01, n \geq 400$.

Example 5.1.2 Suppose we want to calculate the definite integral

$$\int_a^b g(x) dx.$$

Let $X_1, X_2, \dots$ be independent random variables uniformly distributed on the interval $[a, b]$ and write $f(x)$ for their common density. Then, by the law of large numbers, for n large we have that

$$\frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n} \rightarrow E[g(X_1)] = \int_a^b g(x) f(x) dx = \frac{1}{b-a} \int_a^b g(x) dx.$$

Therefore, $\int_a^b g(x) dx \approx (b-a) \times \frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n}$, which leads to a procedure for estimating the integral as follows: simulate uniform random variables $X_1, X_2, \dots, X_n$ on the interval $[a, b]$; compute $g(X_1), g(X_2), \dots, g(X_n)$; average these values and multiply by $b-a$ to estimate the integral.

This method is called Monte Carlo simulation, using the law of large numbers as its

theoretical basis. This approach has been widely applied in practice.

Theorem 5.1.3 Let n_A be the number of the event A in n Bernoulli trials, and $p(0 < p < 1)$ be the probability of the event A on any one Bernoulli trial, then for any $\epsilon > 0$,

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right) = 1.$$

Proof: For $i = 1, 2, \dots$, let $X_i \sim B(1, p)$, then $E(X_i) = p, V(X_i) = p(1 - p)$. Since $n_A = \sum_{i=1}^n X_i \sim B(n, p)$, by the law of large numbers, for any $\epsilon > 0$,

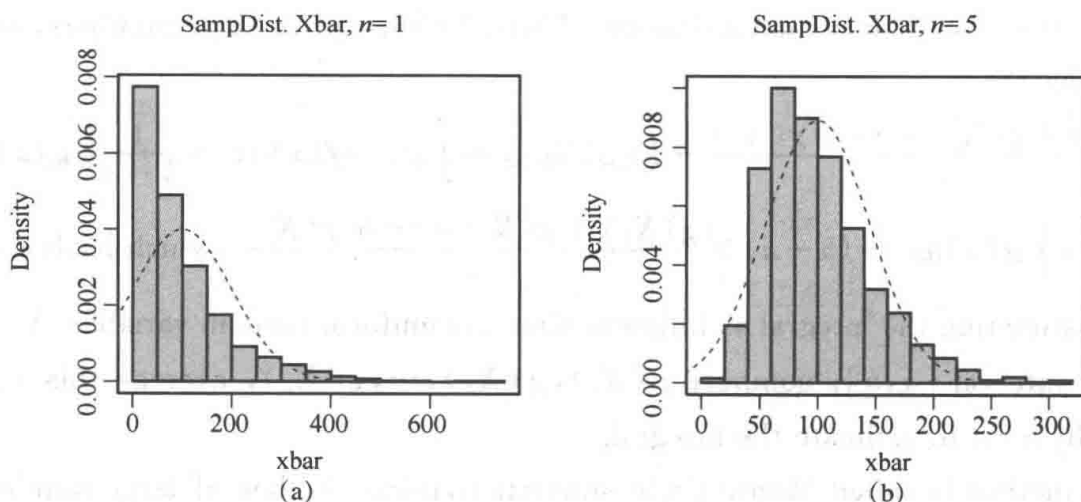
$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right) = \lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| < \epsilon\right) = 1.$$

It shows that the relative frequency $\frac{n_A}{n}$ of successes converges to the probability p of successes with probability 1.

◆ 5.2 The Central Limit Theorem ◆

If we sample from normal distribution $N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ follows exactly the normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ regardless of the sample size n . How about the population is not normal?

Let's check whether this fact holds by testing it in a case where our population is not normally distributed. Let's take our samples from a population whose values are distributed exponentially with the rate parameter $\lambda = \frac{1}{100}$ (the mean of an exponential distribution is 100). Now let us plot the sampling distribution of the sample mean. We take 1000 samples of size $n=1, 5, 10$ and 30 each from this exponentially distributed population. As shown in Figure 5.2, it turns out that the sampling distribution of the sample mean will be normal, as long as the sample size n is large enough.



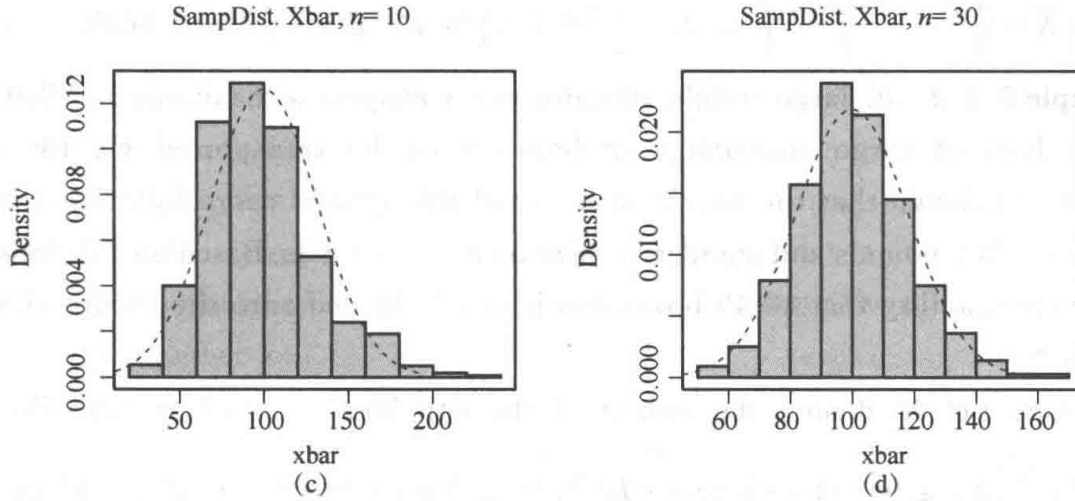


Figure 5. 2

Theorem 5. 2. 1 (Central Limit Theorem) If the random variables $X_1, \dots, X_n$ form a random sample of size n from a given distribution with mean μ and variance σ^2 ($0 < \sigma^2 < +\infty$), then for each fixed number x ,

$$\lim_{n \rightarrow +\infty} P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x).$$

If a large random sample (usually $n \geq 30$) is taken from any distribution, then the distribution of the random variable $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ will be approximately the standard normal

distribution. Or, equivalently, the distribution of the sum $\sum_{i=1}^n X_i$ will be approximately the normal distribution regardless of the shape of the population distribution.

To summarize: the sample mean $\bar{X} \sim N(\mu, \sigma^2/n)$ or $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

Similarly, the central limit theorem states that the sum $X_1 + X_2 + \dots + X_n$ follows approximately the normal distribution, $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$ or $\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$.

Example 5. 2. 1 Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval $[0, 1]$. We shall approximate the value of $P\left(\left|\bar{X} - \frac{1}{2}\right| \leq 0.1\right)$.

Solution: $X_i \sim U[0, 1]$, $\mu = E(X_i) = \frac{1}{2}$, $\sigma^2 = V(X_i) = \frac{1}{12}$. By the central limit theorem, we have

$$\bar{X} \sim N\left(\frac{1}{2}, \frac{1}{144}\right).$$

Thus, $P\left(\left|\bar{X}-\frac{1}{2}\right|\leq 0.1\right)=P\left(12\left|\bar{X}-\frac{1}{2}\right|\leq 1.2\right)\approx P(|Z|\leq 1.2)=0.7698$.

Example 5.2.2 A large freight elevator can transport a maximum of 9800 pounds. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean $\mu=205$ pounds and standard deviation $\sigma=15$ pounds. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

Solution: Let X_i denote the weight of the i -th box, $i=1,2,\dots,49$. We want to estimate $P\left(\sum_{i=1}^{49} X_i \leq 9800\right)$, where $\mu=E(X_k)=205$ pounds, $\sigma=\sqrt{V(X_k)}=15$ pounds, and $n=49$.

We can therefore apply the central limit theorem to estimate

$$\begin{aligned} P\left(\sum_{i=1}^{49} X_i \leq 9800\right) &= P\left(\frac{\sum_{i=1}^{49} X_i - n\mu}{\sqrt{n}\sigma} \leq \frac{9800 - 49 \times 205}{\sqrt{49} \times 15}\right) \\ &\approx P(Z \leq -2.33) \\ &= 1 - \Phi(2.33) \\ &= 0.0099. \end{aligned}$$

Example 5.2.3 One thousand measurements are recorded to several decimal places. Each of these numbers is rounded off to the nearest integer. The sum of the original numbers is approximated by the sum of those integers. Assume that the errors made in rounding off are independent identically distributed random variables with a uniform distribution over the interval $(-0.5, 0.5)$. Compute approximately the probability that the sum of the integers is within ten units of the true sum.

Solution: Let $X_1, X_2, \dots, X_{1000}$ denote the measurements, and $Y_1, Y_2, \dots, Y_{1000}$ be the corresponding nearest integers after rounding off. Then $U_i = X_i - Y_i$ for $i=1,2,\dots,1000$, where $U_1, U_2, \dots, U_{1000}$ are independent identically distributed uniform random variables on the interval $(-0.5, 0.5)$. We then have that $E(U_i) = \mu = 0$, $V(U_i) = \sigma^2 = \frac{1}{12}$, for all i . By the central limit theorem, $\sum_{i=1}^{1000} U_i \sim N\left(0, \frac{1000}{12}\right)$. Thus,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{1000} U_i\right| \leq 10\right) &= P\left(\frac{-10-0}{\sqrt{\frac{1000}{12}}} \leq \frac{\sum_{i=1}^{1000} U_i - 0}{\sqrt{\frac{1000}{12}}} \leq \frac{10-0}{\sqrt{\frac{1000}{12}}}\right) \\ &\approx P(-1.1 \leq Z \leq 1.1) \\ &= 2\Phi(1.1) - 1 \\ &= 0.7286. \end{aligned}$$

The usefulness of Theorem 5.2.1 is that for large enough n , the distribution of the

sample mean $\bar{X}$ or sum $\sum_{i=1}^n X_i$ approaches normality regardless of the shape of the distribution of the individual X_i . It holds for any distribution with mean μ and variance σ^2 ($0 < \sigma^2 < +\infty$). We apply Theorem 5.2.1 to Bernoulli trials $X_i \sim B(1, p)$, $i=1, 2, \dots, n$ and obtain the following particular result.

Theorem 5.2.2 (De Moivre-Laplace Theorem) Consider n independent Bernoulli trials, each having probability $0 < p < 1$ of success, let S_n denote the number of successes, then for any real number x ,

$$\lim_{n \rightarrow +\infty} P\left[\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right] = \Phi(x).$$

The De Moivre-Laplace theorem, which is a special case of the central limit theorem, states that the normal distribution may be used as an approximation to the binomial distribution under certain conditions. See Figure 5.3. According to the De Moivre-Laplace theorem, as n grows larger, the shape of the discrete distribution converges to the continuous symmetric curve of the normal distribution.

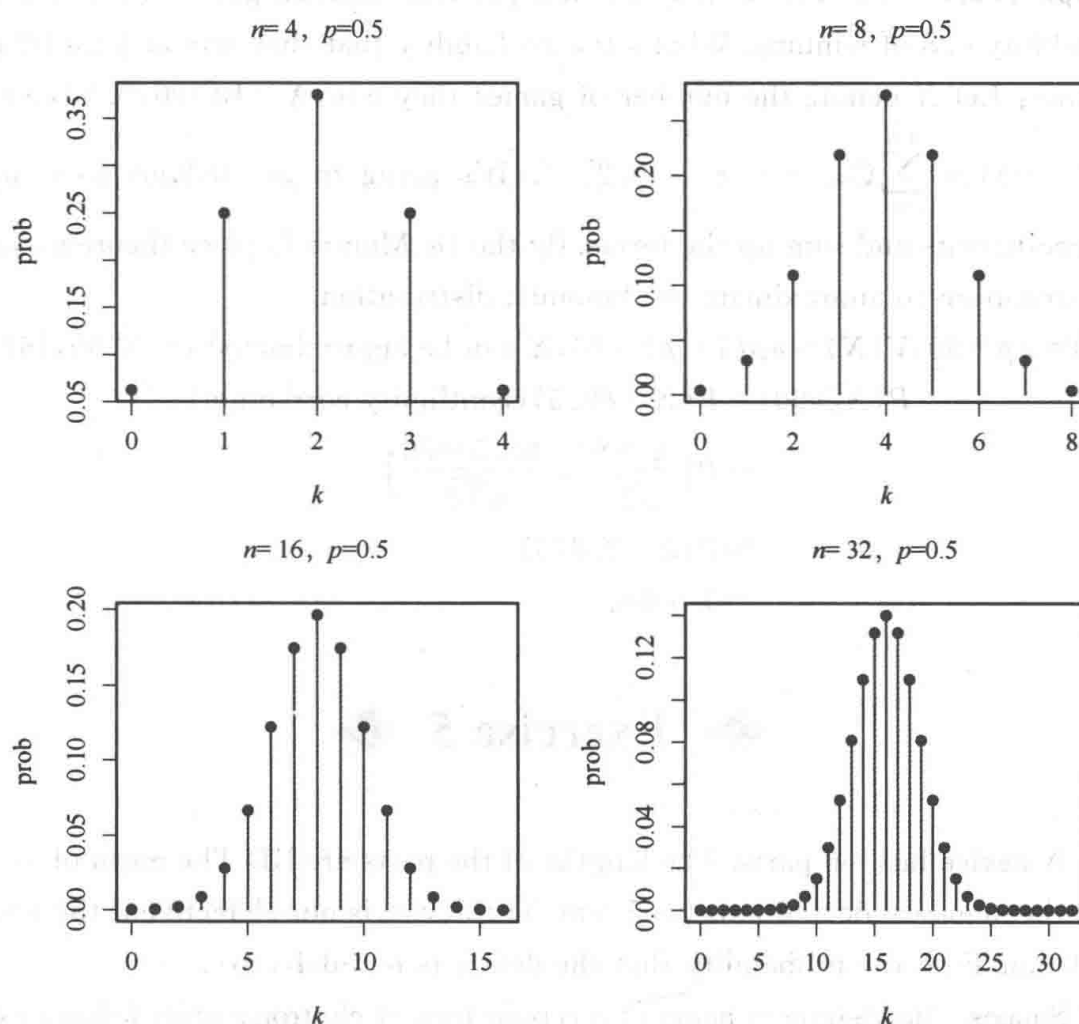


Figure 5.3

Since the binomial distribution is discrete and the normal distribution is continuous, correction for continuity is needed when doing the approximation. For two integers $k_1 < k_2$,

we use the following correction criterion

$$P(k_1 \leq X \leq k_2) \approx P(k_1 - 0.5 < X < k_2 + 0.5).$$

Example 5.2.4 Suppose that a coin is tossed 20 times and that all tosses are independent. What is the probability of obtaining exactly 10 heads?

Solution: Let X denote the number of heads obtained. $X \sim B(20, \frac{1}{2})$.

$$P(X = 10) = C_{20}^{10} \times \left(\frac{1}{2}\right)^{10} \times \left(1 - \frac{1}{2}\right)^{20-10} = 0.1762.$$

On the other hand, $E(X) = np = 10$, $V(X) = np(1-p) = 5$, X can be approximated by $N(10, 5)$.

$$\begin{aligned} P(X=10) &= P(9.5 < X < 10.5) \\ &= P\left(\frac{9.5-10}{\sqrt{5}} < \frac{X-10}{\sqrt{5}} < \frac{10.5-10}{\sqrt{5}}\right) \\ &\approx P(-0.2236 < Z < 0.2236) \\ &= 0.177. \end{aligned}$$

Example 5.2.5 The Braves play 100 independent baseball games, each of which they have probability 0.8 of winning. What's the probability that they win at least 90 games?

Solution: Let X denote the number of games they win, $X \sim B(100, 0.8)$, we want to find $P(X \geq 90) = \sum_{i=90}^{100} C_{100}^i \times 0.8^i \times 0.2^{100-i}$. It's going to get tedious to compute the binomial coefficients and sum up the terms. By the De Moivre-Laplace theorem, we can use normal distribution to approximate the binomial distribution.

$E(X) = np = 80$, $V(X) = np(1-p) = 16$, X can be approximated by $N(80, 16)$.

$$\begin{aligned} P(X \geq 90) &= P(X > 89.5) \text{ (continuity correction)} \\ &= P\left(\frac{X-80}{\sqrt{16}} > \frac{89.5-80}{\sqrt{16}}\right) \\ &\approx P(Z > 2.375) \\ &= 0.0088. \end{aligned}$$

◆ Exercise 5 ◆

5.1 A device has ten parts. The lengths of the parts are IID The mean of each part is 2 mm and the standard deviation is 0.05 mm. The device is non-defective if the total length is 20 ± 0.1 mm. Find the probability that the device is non-defective.

5.2 Suppose the lifetime in hours of a certain type of electronic units follows exponential distribution with $\lambda = 0.1$. Let T be the total lifetime of 30 electronic units of this type. Determine $P(T \geq 350)$.

5.3 Suppose that 75 percent of the people in a certain metropolitan area live in the city and 25 percent of the people live in the suburbs. If 1200 people attending a certain

concert represent a random sample from the metropolitan area, what is the probability that the number of people from the suburbs attending the concert will be fewer than 270?

5.4 Suppose that the distribution of the number of defects on any given bolt of cloth is the Poisson distribution with mean 5, and the number of defects on each bolt is counted for a random sample of 125 bolts. Determine the probability that the average number of defects per bolt in the sample will be less than 5.5.

5.5 Let X be the length of a pregnancy in days. Suppose that it has approximately a normal distribution with a mean of 266 days and a standard deviation of 16 days.

(1) What is the probability that a randomly selected pregnancy lasts more than 274 days?

(2) Suppose that we have a random sample of $n = 25$ pregnant women. What is the probability that the sample mean pregnancy length is more than 274 days?

5.6 Suppose that a random sample of size n is to be taken from a distribution for which the mean is μ and the standard deviation is 3. Use the central limit theorem to determine approximately the smallest value of n for which the following relation will be satisfied:

$$P(|\bar{X} - \mu| < 0.3) \geq 0.95.$$

5.7 Suppose that the proportion of defective items in a large manufactured lot is 0.1. What is the smallest random sample of items that must be taken from the lot in order for the probability to be at least 0.99 that the proportion of defective items in the sample will be less than 0.13?

5.8 A local bakery randomly sells three types of cupcakes. Accordingly, the cupcake's price in dollars is a random variable with the following PF.

X	1	1.2	1.5
P	0.3	0.2	0.5

On a given day, the bakery sold 300 cupcakes.

(1) Find the probability that its earning is at least 400 dollars.

(2) Find the probability that the bakery sold at least 60 cupcakes at the price of 1.2 dollars.

5.9 A physicist makes 25 independent measurements of the specific gravity of a certain body. He knows that the limitations of his equipment are such that the standard deviation of each measurement is σ units.

(1) By using the Chebyshev's inequality, find a lower bound for the probability that the average of his measurements will differ from the actual specific gravity of the body by less than $\sigma/4$ units.

(2) By using the central limit theorem, find an approximate value for the probability in part (1).

5.10 Let $X_1, \dots, X_{30}$ be independent random variables each having a discrete

distribution with PF

$$f(x) = \begin{cases} \frac{1}{4}, & x = 0 \text{ or } 2 \\ \frac{1}{2}, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Use the central limit theorem and the correction for continuity to approximate the probability that $X_1 + \dots + X_{30}$ is at most 33.

5.11 Suppose that a pair of balanced dice are rolled 120 times, and let X denote the number of rolls on which the sum of the two numbers is 7. Use the central limit theorem to determine a value of k such that $P(|X - 20| \leq k)$ is approximately 0.95.

5.12 Let X be normally distributed with standard deviation σ . Determine $P(|X - \mu| \geq 2\sigma)$. Compare with the Chebyshev's inequality.

5.13 Let X be a random variable with $E(X) = 0$ and $V(X) = 1$. What integer value k will assure that $P(|X| \geq k) \leq 0.01$?

5.14 A certain drug is to be rated either effective or ineffective. Suppose lab results indicate that 75% of the time the drug increases the lifespan of a patient by 5 years (effective) and 25% of the time the drug causes a complication which decreases the lifespan of a patient by 1 year (ineffective). As part of a study you administer the drug to 10000 patients. What is the probability that the lifespans of those in the study will be increased by an average of 3.55 years or more?

Chapter 6

Estimation

Probability theory has been introduced in previous chapters. From now on, we focus on statistics. Statistics is based on probability and plays a vital role in almost every field like industry, commerce, economics, biology and education etc.

In probability, the cumulative distribution function of a random variable and the values of its parameters are usually given. It is not always true in practice. Maybe the cumulative distribution function is unknown, or the cumulative distribution function is given with unknown parameters. The cumulative distribution function and its parameters should be inferred based on some observations. Statistics is the science that deals with the collection, classification, analysis, and interpretation of data, and that, by use of mathematical theories of probability, makes decisions.

Some basic concepts in statistics and parameter estimation will be covered in this chapter.

◆ 6.1 Population and Sample ◆

A population consists of all subjects that are being studied. For example, we want to know the mean lifetime μ of air conditioners in a factory. The population includes the lifetimes of all air conditioners produced by this factory. Typically it is impossible or unfeasible to survey/measure the entire population due to time and cost.

Instead, we could take a subset of this population called a sample $X_1, X_2, \dots, X_n$, and use this sample of size n to draw inferences about the population X under study, given some conditions. Thus we could measure the mean lifetime $\bar{X}$ of air conditioners in a sample of the population which we call a statistic and use this to draw inferences about the parameter of interest in the population. A measurable characteristic of a population, such as a mean μ or standard deviation σ , is called a parameter; but a measurable characteristic of a sample is called a statistic.

Each X_i is drawn from the population X and $(X_1, X_2, \dots, X_n)$ is a n -dimensional random vector. The specific values of a sample are called observations, denoted by $(x_1, x_2, \dots, x_n)$.

The main purpose of a sample is to estimate or infer some measurements of a population as accurate as possible. There will be some uncertainty and inaccuracy involved in making conclusions about the population based upon a sample. We have fewer members in our sample than our population, therefore we have lost some information.

There are many ways to select a sample and the study of this is called sampling theory. A commonly used method is called simple random sampling (SRS). In SRS each member of the population has an equal probability of being included in the sample, hence the individuals in the sample are independent and representative.

In this textbook, we consider simple random sample. If the population X follows a CDF $F(x)$, then the joint CDF of the sample $(X_1, X_2, \dots, X_n)$ is given by

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i).$$

If X is discrete with PF $p(x)$, then the joint PF of the sample $(X_1, X_2, \dots, X_n)$ is given by

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n p(x_i).$$

If X is continuous with PDF $f(x)$, then the joint PDF of the sample $(X_1, X_2, \dots, X_n)$ is given by

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

Example 6.1.1 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from $X \sim N(\mu, \sigma^2)$, determine its joint PDF.

Solution: The PDF of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

Thus, the joint PDF of $(X_1, X_2, \dots, X_n)$ is

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}},$$

for $-\infty < x_i < +\infty, i=1, 2, \dots, n$.

Statistical inference is based on sample, but its observations are irregular and impractical to be used to solve the problem under study. It is necessary to summarize the observations and extract and purify the information. Some appropriate functions of sample should be proposed to meet the needs of the problems.

Definition 6.1.1 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from X , a function of sample $g(X_1, X_2, \dots, X_n)$ is called statistic if it does not contain any unknown parameters.

Note: the statistic $g(X_1, X_2, \dots, X_n)$ is a random variable. $g(x_1, x_2, \dots, x_n)$ is its specific observation.

Example 6.1.2 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from $X \sim N(\mu, \sigma^2)$, where μ is known but σ^2 is unknown, then $\min\{X_1, X_2, \dots, X_n\}$, $\frac{(X_1 + X_2)}{2}$ and $\frac{(X_1 + X_2 + \dots + X_n)}{n} - \mu$ are statistics. $\frac{(X_1 + X_2)^2}{\sigma^2}$ is not a statistic.

Definition 6.1.2 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from X , several common useful statistics are defined as follows.

- Sample mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

- Sample variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$.
- Sample standard deviation: $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$.
- Sample j -th raw moment: $A_j = \frac{1}{n} \sum_{i=1}^n X_i^j, j = 1, 2, \dots$.
- Sample j -th central moment: $B_j = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^j, j = 1, 2, \dots$.

The above statistics are numeric characteristics of sample. The corresponding parts for population are as follows.

- Population mean: $E(X) = \mu$.
- Population variance: $V(X) = E\{[X - E(X)]^2\} = \sigma^2$.
- Population standard deviation: $\sigma = \sqrt{V(X)}$.
- Population j -th raw moment $E(X^j), j = 1, 2, \dots$.
- Population j -th central moment $E\{[X - E(X)]^j\}, j = 1, 2, \dots$.

Definition 6.1.3 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from $X, (x_1, x_2, \dots, x_n)$ is a specific observations, reorder them so that $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, let $x_{(k)}$ be the observation of $X_{(k)}$, then $X_{(k)}$ is called the k -th ordered statistic. The first order statistic (or smallest order statistic) is the minimum of the sample, that is, $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$. The n -th order statistic (or largest order statistic) is the maximum, that is, $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$. The statistical range is defined as $X_{(n)} - X_{(1)}$.

Definition 6.1.4 Suppose $(X_1, X_2, \dots, X_n)$ is a sample drawn from $X, (x_1, x_2, \dots, x_n)$ is a specific observations, reorder them so that $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, the empirical distribution function is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x} = \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, \\ 1, & x \geq x_{(n)} \end{cases}$$

where $I_{X_i \leq x}$ is the indicator function.

Note: the empirical distribution function $F_n(x)$ describes the frequency of the event $\{X \leq x\}$ and the cumulative distribution function $F(x) = P(X \leq x)$, according to the law of large numbers, as n is large enough, $F_n(x)$ converges to $F(x)$.

Example 6.1.3 Suppose $-0.02, -0.38, -0.23, 2.15, 1.57, -1.73, 0.28, 0.73, -0.14, -1.20$ is a set of observations, the empirical distribution function $F_n(x)$ is sketched in Figure 6.1.

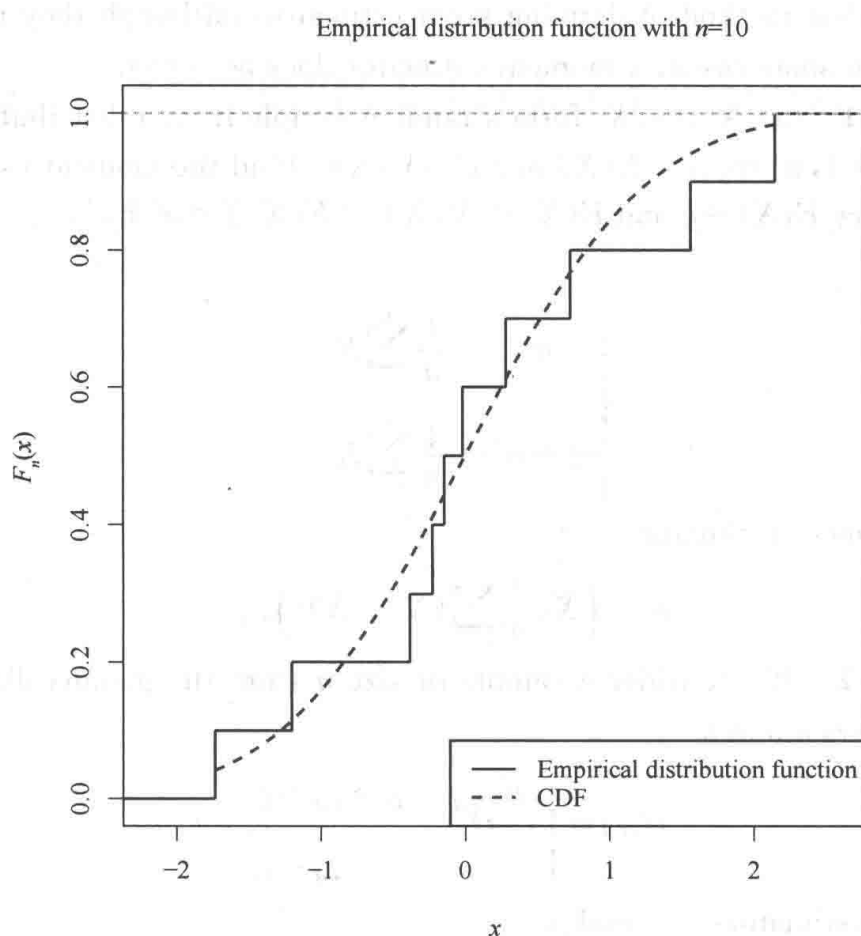


Figure 6.1

◆ 6.2 Moment Estimation ◆

In practice, the distribution function $F(x, \theta)$ of population X is usually known with one or more parameters θ which need to be estimated from a sample $(X_1, X_2, \dots, X_n)$. $(x_1, x_2, \dots, x_n)$ is a set of specific observations. We want to construct an appropriate statistic $\theta(X_1, X_2, \dots, X_n)$ and use its observation $\hat{\theta}(x_1, x_2, \dots, x_n)$ to estimate the value of θ . $\hat{\theta}(X_1, X_2, \dots, X_n)$ is called an estimator and $\hat{\theta}(x_1, x_2, \dots, x_n)$ is called an estimate. We will introduce some methods to find the estimator.

Definition 6.2.1 Assume that $X_1, \dots, X_n$ form a random sample from a distribution X with a k -dimensional parameter θ . We consider the following moments.

(1) Sample moments: $m_j = \frac{1}{n} \sum_{i=1}^n X_i^j$.

(2) Population moments: $\mu_j(\theta) = E(X^j) = E(X_i^j)$.

For $j=1, 2, \dots, k$. Set up the k equations $m_j = \mu_j(\theta)$ and solve for θ . The solution $\hat{\theta}$ is the moment estimator of θ .

This method of deriving estimators is called the method of moments. An important statistical principle, the substitution principle, is used in this method. The method of

moments is the oldest method of deriving point estimators, although they may not be the best estimators. In some cases, a moment estimator does not exist.

Example 6.2.1 Let $X_1, \dots, X_n$ form a random sample from a distribution X with the parameter $\theta = (\mu, \sigma^2)$, where $\mu = E(X)$ and $\sigma^2 = V(X)$. Find the moment estimator of θ .

Solution: Since $E(X) = \mu$ and $E(X^2) = V(X) + [E(X)]^2 = \sigma^2 + \mu^2$,
let

$$\begin{cases} \mu &= \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 + \mu^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}.$$

We obtain the moment estimator

$$\hat{\theta} = \left(\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right).$$

Example 6.2.2 We consider a sample of size n from the gamma distribution with unknown parameters α and β .

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0 \\ 0 & , x \leq 0 \end{cases}.$$

Find the moment estimators of α and β .

Solution:

$$\mu_1 = E(X) = \frac{\alpha}{\beta}, \mu_2 = E(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

The method of moments says to replace the right-hand sides of these equations by the sample moments and then solve for α and β .

Let $\mu_1 = m_1, \mu_2 = m_2$, then

$$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}, \hat{\beta} = \frac{m_1}{m_2 - m_1^2}.$$

Example 6.2.3 Let $X_1, \dots, X_n$ form a random sample from the uniform distribution $X \sim U(a, b)$, where $a < b$ are unknown parameters. Find the moment estimators of a and b .

Solution: Since $\mu_1 = E(X) = \frac{a+b}{2}, \mu_2 = E(X^2) = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2$, let

$$\begin{cases} \frac{a+b}{2} &= \frac{1}{n} \sum_{i=1}^n X_i \\ \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}.$$

Solve these equations and obtain the moment estimators

$$\begin{cases} \hat{a} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \\ \hat{b} = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \end{cases}.$$

◆ 6.3 Maximum Likelihood Estimation ◆

Suppose we have a random sample $X_1, \dots, X_n$ whose assumed probability distribution X depends on some unknown parameter θ . Our goal here will be to find a point estimator $\hat{\theta}(X_1, \dots, X_n)$, such that $\hat{\theta}(x_1, \dots, x_n)$ is a good point estimate of θ , where $x_1, \dots, x_n$ are the observed values of the random samples.

The basic idea behind the method of maximum likelihood estimation is as follows: maximum likelihood estimation chooses the estimate of θ as the value for the parameter that makes the observed data most probable.

Definition 6.3.1 When the joint PDF or the joint PF

$$f_n(x; \theta) = f(x_1; \theta) \times \dots \times f(x_n; \theta)$$

of the observations in a random sample is regarded as a function of θ for given values of $x_1, \dots, x_n$, it is called the likelihood function and denoted by

$$L(\theta; \mathbf{x}) = f(x_1; \theta) \times \dots \times f(x_n; \theta).$$

Note: the expressions for the likelihood function and the joint PDF or the joint PF are identical. For the likelihood function, the parameter θ is a variable and the observations are fixed. For the joint PDF or the joint PF, the parameter θ is fixed and $x_1, \dots, x_n$ are variables.

Definition 6.3.2 For each possible observed vector $\mathbf{x} = (x_1, \dots, x_n)$, let

$$\hat{\theta}(\mathbf{x}) = \arg \max_{\theta} L(\theta; \mathbf{x}).$$

Denote a value of $\theta \in \Omega$ for which the likelihood function is a maximum, and let $\hat{\theta}(\mathbf{X})$ be the estimator of θ defined in this way. The estimator $\hat{\theta}$ is called a maximum likelihood estimator of θ . After $\mathbf{X} = \mathbf{x}$ is observed, the value $\hat{\theta}(\mathbf{x})$ is called a maximum likelihood estimate of θ . The set Ω of all possible values of a parameter/parameters is called a parameter space.

Note: rather than maximizing $L(\theta; \mathbf{x})$ which can be quite tedious, we often use the fact that the logarithm is an increasing function so it will be equivalent to maximize the log-likelihood:

$$l(\theta) = \sum_{i=1}^n \ln f(x_i; \theta).$$

Example 6.3.1 Choose a sample of size 3 from the exponential distribution with parameter $\lambda > 0$, the observed data is $(x_1, x_2, x_3) = (3, 1.5, 2.1)$. Find the maximum likelihood estimate of λ .

Solution:

$$X \sim f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

The likelihood function is

$$L(\lambda) = f(x_1)f(x_2)f(x_3) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)} = \lambda^3 e^{-6.6\lambda}.$$

Since the logarithm is an increasing function, the value of θ that maximizes the likelihood function $L(\lambda)$ will be the same as the value of λ that maximizes

$$l(\lambda) = \ln L(\lambda) = 3 \ln \lambda - 6.6\lambda.$$

The log-likelihood function $l(\lambda)$ is sketched in Figure 6.2. Taking the derivative, setting the derivative to 0, and solving for λ yields

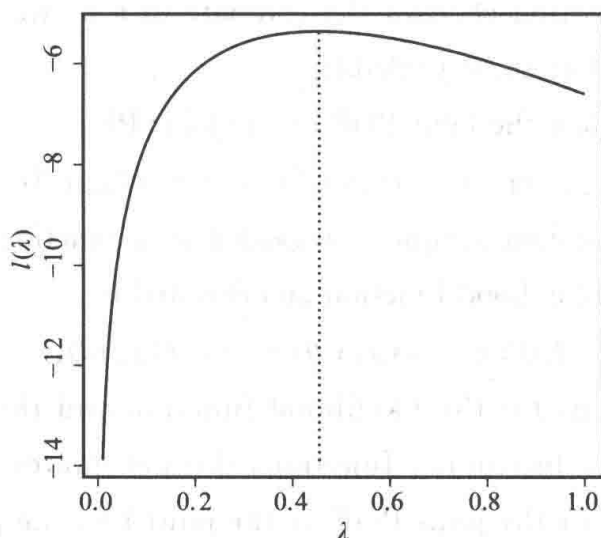


Figure 6.2

$$\frac{dl(\lambda)}{d\lambda} = \frac{3}{\lambda} - 6.6, \quad \frac{d^2l(\lambda)}{d\lambda^2} = -\frac{3}{\lambda^2} < 0, \quad \frac{dl(\lambda)}{d\lambda} = 0, \quad \lambda = \frac{3}{6.6} = 0.455.$$

The maximum likelihood estimate is then 0.455.

Example 6.3.2 Suppose that X is a discrete random variable with the following PF: where $0 < \theta < 1$ is a parameter.

X	0	1	2	3
P	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{(1-\theta)}{3}$

The following 10 independent observations were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1). What is the maximum likelihood estimate of θ ?

Solution: The likelihood function is

$$L(\theta) = \prod_{i=1}^{10} P(X_i = x_i) = \left(\frac{2\theta}{3}\right)^2 \left(\frac{\theta}{3}\right)^3 \left[\frac{2(1-\theta)}{3}\right]^3 \left(\frac{1-\theta}{3}\right)^2.$$

The likelihood function $L(\theta)$ is not easy to maximize. Let us consider the log-likelihood function

$$l(\theta) = \ln L(\theta) = 5\ln 2 - 10\ln 3 + 5\ln \theta + 5\ln(1-\theta).$$

The log-likelihood function $l(\theta)$ is sketched in Figure 6.3. Taking the derivative, setting the derivative to 0, and solving for θ yields

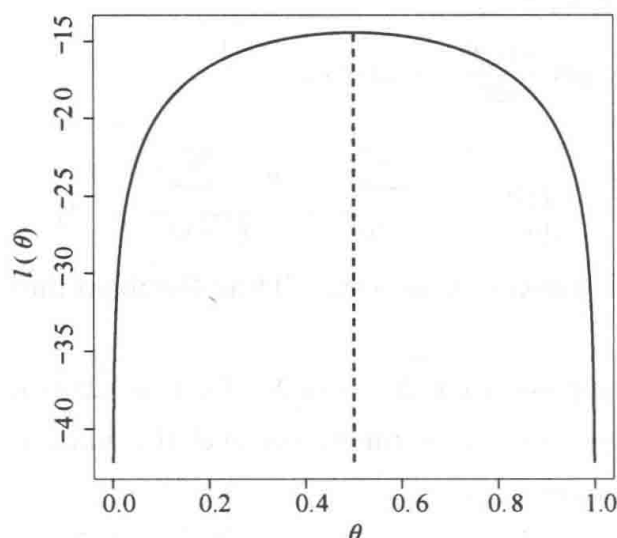


Figure 6.3

$$\frac{dl(\theta)}{d\theta} = \frac{5}{\theta} - \frac{5}{1-\theta}, \quad \frac{d^2l(\theta)}{d\theta^2} = -\frac{5}{\theta^2} - \frac{5}{(1-\theta)^2} < 0,$$

$$\frac{dl(\theta)}{d\theta} = 0, \quad \theta = \frac{1}{2}.$$

The maximum likelihood estimate is then $\frac{1}{2}$.

Note: the method of moment estimate is $\frac{5}{12}$ (try it!), which is different from maximum likelihood estimation.

Example 6.3.3 Suppose that the random variables $X_1, \dots, X_n$ form a random sample from the Bernoulli distribution with parameter θ , which is unknown ($0 \leq \theta \leq 1$). Find the maximum likelihood estimator of θ .

Solution: For all observed values $x_1, \dots, x_n$, where each x_i is either 0 or 1, the likelihood function is

$$L(\theta) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}.$$

Let us consider the log-likelihood function

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n x_i \ln \theta + \left(n - \sum_{i=1}^n x_i\right) \ln(1-\theta).$$

$$\frac{dl(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta}.$$

① If $\sum_{i=1}^n x_i = 0$, $\frac{dl(\theta)}{d\theta} = -\frac{n}{1-\theta} < 0$, then $l(\theta)$ is a decreasing function of θ , and hence

$l(\theta)$ achieves its maximum at $\theta = 0 = \bar{x}$.

② If $\sum_{i=1}^n x_i = n$, $\frac{dl(\theta)}{d\theta} = \frac{n}{\theta} > 0$, then $l(\theta)$ is a increasing function of θ , and hence $l(\theta)$ achieves its maximum at $\theta = 1 = \bar{x}$.

③ If $\sum_{i=1}^n x_i \neq \{0, n\}$, set $\frac{dl(\theta)}{d\theta} = 0, \theta = \bar{x}$,

$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{\sum_{i=1}^n x_i}{\theta^2} - \frac{n - \sum_{i=1}^n x_i}{(1 - \theta)^2} < 0,$$

Then $l(\theta)$ achieves its maximum at $\theta = \bar{x}$: Thus the maximum likelihood estimator of θ is $\bar{X}$.

Example 6.3.4 Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution for which the mean μ is unknown and the variance σ^2 is known. Find the maximum likelihood estimator of μ .

Solution: For all observations $x_1, \dots, x_n$, the likelihood function is

$$L(\mu) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right].$$

$L(\mu)$ will be maximized by the value of μ that minimizes

$$Q(\mu) = \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2$$

We see that $Q(\mu)$ is a quadratic in μ with positive coefficient on μ^2 . It follows that $Q(\mu)$ will be minimized where its derivative is 0.

$$\frac{dQ(\mu)}{d\mu} = -2 \sum_{i=1}^n x_i + 2n\mu = 0.$$

$$\mu = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

The maximum likelihood estimator of μ is $\hat{\mu} = \bar{X}$.

Example 6.3.5 Suppose again that $X_1, \dots, X_n$ form a random sample from a normal distribution for which the mean μ is unknown and the variance σ^2 is also unknown. Find the maximum likelihood estimator of $\theta = (\mu, \sigma^2)$.

Solution: For all observations $x_1, \dots, x_n$, the likelihood function is

$$L(\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right].$$

Let us consider the log-likelihood function

$$l(\theta) = \ln L(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

We shall find the value of $\theta = (\mu, \sigma^2)$ for which $l(\theta)$ is maximum.

$$\begin{cases} \frac{\partial l(\theta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial l(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases}$$

Solve these two equations, we have $\mu = \bar{x}, \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Thus, the maximum likelihood estimator of θ is $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2) = (\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2)$.

Example 6.3.6 Suppose that $X_1, \dots, X_n$ form a random sample from the uniform distribution on the interval $[0, \theta]$, where $\theta > 0$. Find the maximum likelihood estimator of θ .

Solution: The PDF of the uniform distribution on the interval $[0, \theta]$ is

$$X \sim f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{otherwise} \end{cases}$$

For all observations $x_1, \dots, x_n$, the likelihood function is

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 \leq x_i \leq \theta (i=1, \dots, n) \\ 0, & \text{otherwise} \end{cases}$$

The likelihood function $L(\theta)$ is sketched in Figure 6.4. The maximum likelihood estimator of θ must be a value of θ for which $x_i \leq \theta (i=1, \dots, n)$ and that maximizes $\frac{1}{\theta^n}$ among all such values.

Since $\frac{1}{\theta^n}$ is a decreasing function of θ , the estimator will be the smallest value of θ such that $\theta \geq x_i$ for $i=1, \dots, n$. Since this value is $\theta = \max\{x_1, \dots, x_n\}$, the maximum likelihood estimator of θ is $\hat{\theta} = \max\{X_1, \dots, X_n\}$.

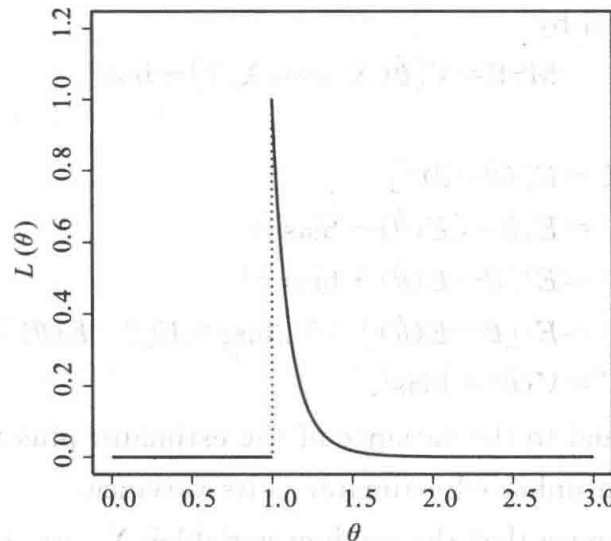


Figure 6.4

Theorem 6.3.1 If $\hat{\theta}$ is the maximum likelihood estimator of θ and if g is a one-to-one function, then $g(\hat{\theta})$ is the maximum likelihood estimator of $g(\theta)$.

Note: in other words, we simply substitute in $g(\theta)$ the value $\hat{\theta}$ instead of θ and obtain the maximum likelihood estimator for the new parameter $g(\theta)$. This is called the invariance property of the maximum likelihood estimator.

Example 6.3.7 Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution

$N(\mu, \sigma^2)$ for which both the mean μ and the variance σ^2 are unknown. It was found that the maximum likelihood estimator of $\theta = (\mu, \sigma^2)$ is

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2) = \left(\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right).$$

From the invariance property, we can conclude that the maximum likelihood estimator of σ

is $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$, also, the maximum likelihood estimator of μ^2 is $\hat{\mu}^2 = \bar{X}^2$.

◆ 6.4 Properties of Estimators ◆

Definition 6.4.1 An estimator $\hat{\theta}(X_1, \dots, X_n)$ is an unbiased estimator of the parameter θ if $E[\hat{\theta}(X_1, \dots, X_n)] = \theta$. An estimator that is not unbiased is called a biased estimator. The bias is the difference

$$\text{bias} = E[\hat{\theta}(X_1, \dots, X_n)] - \theta.$$

The mean squared error (MSE) of an estimator can be used to assess its quality and is defined by

$$\text{MSE} = E\{[\hat{\theta}(X_1, \dots, X_n) - \theta]^2\}.$$

Note: the MSE is a measure of the goodness of an estimator. The estimators with smaller MSE are generally preferred to those with larger.

Proposition 6.4.1 Suppose $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ is an estimator of the parameter θ , its MSE can also be expressed by

$$\text{MSE} = V[\hat{\theta}(X_1, \dots, X_n)] + \text{bias}^2.$$

Proof:

$$\begin{aligned} \text{MSE} &= E[(\hat{\theta} - \theta)^2] \\ &= E\{\hat{\theta} - [E(\hat{\theta}) - \text{bias}]\}^2 \\ &= E\{[\hat{\theta} - E(\hat{\theta}) + \text{bias}]^2\} \\ &= E\{[\hat{\theta} - E(\hat{\theta})]^2\} + 2\text{bias} \times E[\hat{\theta} - E(\hat{\theta})] + \text{bias}^2 \\ &= V(\hat{\theta}) + \text{bias}^2. \end{aligned}$$

Thus, the MSE is equal to the variance of the estimator plus the square of the bias. In particular, the MSE for an unbiased estimator is its variance.

Example 6.4.1 Suppose that the random variables $X_1, \dots, X_n$ form a random sample from the Bernoulli distribution with parameter θ , which is unknown. It was found that the maximum likelihood estimator of θ is the sample mean $\bar{X}$. Since

$$E(\bar{X}) = \frac{1}{n} [E(X_1) + \dots + E(X_n)] = \frac{1}{n} (\theta + \dots + \theta) = \theta.$$

Thus, $\bar{X}$ is an unbiased estimator for θ .

Example 6.4.2 Let $(X_1, \dots, X_n)$ be a random sample from a distribution X . Assume that the variance σ^2 and mean μ of the distribution are finite. Show the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

is an unbiased estimator of the variance σ^2 .

Proof:

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \{V(X_i) + [E(X_i)]^2\} - n\{V(\bar{X}) + [E(\bar{X})]^2\}\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] \\ &= \sigma^2. \end{aligned}$$

Example 6.4.3 Sample (X_1, X_2, X_3) is chosen from $X \sim f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$.

Find the maximum likelihood estimator $\hat{\lambda}$ of λ , is it unbiased?

Solution: The likelihood function is

$$L(\lambda) = f(x_1)f(x_2)f(x_3) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)}.$$

Let us consider the log-likelihood function

$$\ln L(\lambda) = 3\ln \lambda - (x_1 + x_2 + x_3)\lambda.$$

Set $\frac{d\ln L(\lambda)}{d\lambda} = \frac{3}{\lambda} - (x_1 + x_2 + x_3) = 0$, solve it and obtain $\lambda = \frac{3}{x_1 + x_2 + x_3}$. Thus, the maximum likelihood estimator of λ is

$$\hat{\lambda} = \frac{3}{X_1 + X_2 + X_3}.$$

Noting that $X_1, X_2, X_3 \sim E\left(\frac{1}{\lambda}\right) = \text{Gamma}\left(1, \frac{1}{\lambda}\right)$, using the additive property of the gamma distribution (see Theorem 3.5.3), we have $T = X_1 + X_2 + X_3 \sim \text{Gamma}\left(3, \frac{1}{\lambda}\right)$. Thus, the PDF of T is

$$\begin{aligned} T \sim f(t) &= \begin{cases} \frac{\lambda^3}{2} t^2 e^{-\lambda t}, & t > 0 \\ 0, & t \leq 0 \end{cases} \\ E(\hat{\lambda}) &= E\left(\frac{3}{T}\right) = \int_0^{+\infty} \frac{3}{t} \times \frac{\lambda^3}{2} t^2 e^{-\lambda t} dt = \frac{3\lambda}{2}. \end{aligned}$$

Therefore, $\hat{\lambda}$ is a biased estimator of λ . Its bias is $\frac{3\lambda}{2} - \lambda = \frac{\lambda}{2}$.

Noting that $E\left(\frac{2\hat{\lambda}}{3}\right) = \lambda$, then, $\frac{2\hat{\lambda}}{3} = \frac{2}{X_1 + X_2 + X_3} = \frac{2}{T}$ is an unbiased estimator of λ .

In the following, we want to compare the MSEs of the two estimators $\hat{\lambda}_1 = \frac{3}{T}$ and $\hat{\lambda}_2 = \frac{2}{T}$. Since

$$\begin{aligned} E\left(\frac{1}{T}\right) &= \int_0^{+\infty} \frac{1}{t} \times \frac{\lambda^3}{2} t^2 e^{-\lambda t} dt = \frac{\lambda}{2}, \\ E\left(\frac{1}{T^2}\right) &= \int_0^{+\infty} \frac{1}{t^2} \times \frac{\lambda^3}{2} t^2 e^{-\lambda t} dt = \frac{\lambda^2}{2}, \\ V\left(\frac{1}{T}\right) &= E\left(\frac{1}{T^2}\right) - \left[E\left(\frac{1}{T}\right)\right]^2 = \frac{\lambda^2}{4}. \end{aligned}$$

Thus,

$$\begin{aligned} \text{MSE}\left(\frac{3}{T}\right) &= V\left(\frac{3}{T}\right) + \text{bias}^2 = \frac{9\lambda^2}{4} + \left(\frac{\lambda}{2}\right)^2 = \frac{5\lambda^2}{2}, \\ \text{MSE}\left(\frac{2}{T}\right) &= V\left(\frac{2}{T}\right) = \frac{4\lambda^2}{4} = \lambda^2. \end{aligned}$$

Definition 6.4.2 If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of a parameter θ , we say that $\hat{\theta}_1$ is relatively more efficient if

$$V(\hat{\theta}_1) < V(\hat{\theta}_2).$$

Example 6.4.4 Suppose $(X_1, \dots, X_n)$ ($n > 2$) is a random sample from a Poisson distribution with parameter λ . Let $\hat{\lambda}_1 = \bar{X}$ and $\hat{\lambda}_2 = \frac{X_1 + X_2}{2}$ be two unbiased estimators of λ . Which one is more efficient?

Solution:

$$\begin{aligned} E(\hat{\lambda}_1) &= E(\bar{X}) = \lambda, \\ V(\hat{\lambda}_1) &= V(\bar{X}) = \frac{\lambda}{n}, \\ E(\hat{\lambda}_2) &= E\left(\frac{X_1 + X_2}{2}\right) = \lambda, \\ V(\hat{\lambda}_2) &= V\left(\frac{X_1 + X_2}{2}\right) = \frac{\lambda}{2}. \end{aligned}$$

Since $V(\hat{\lambda}_1) < V(\hat{\lambda}_2)$, $\hat{\lambda}_1$ is more efficient.

Definition 6.4.3 The estimator $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ of a parameter θ is said to be consistent estimator if for any positive $\epsilon > 0$,

$$\lim_{n \rightarrow +\infty} P(|\hat{\theta} - \theta| \leq \epsilon) = 1$$

or

$$\lim_{n \rightarrow +\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

The consistency means that having a sufficiently large number of observations n , it is possible to find the value of θ with arbitrary precision. Under some conditions, the maximum likelihood estimator is consistent.

◆ 6.5 Three Important Distributions ◆

As a special random variable, a statistic is a function of a sample. Its distribution is

called sampling distribution. Sampling distribution is important in statistics because it allows analytical considerations to be based on the sampling distribution of a statistic, rather than on the joint probability distribution of all the individual sample values.

The sampling distribution depends on the underlying distribution of the population, the statistic being considered and the sample size used. In general, it is complicated to determine the sampling distribution of a given statistic. But for population is normal, sampling distributions have been widely investigated. We will introduce three useful sampling distributions in this section.

Theorem 6.5.1 Suppose that $Z \sim N(0,1)$, its PDF is given as follows

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Then, if $X = Z^2$, we say that X follows the chi-square distribution with 1 degree of freedom. $X \sim \chi^2(1)$.

Proof: Since $X = Z^2$, $-\infty < Z < +\infty$, then $X \geq 0$.

(1) If $x \leq 0$, $F_X(x) = P(X \leq x) = P(\emptyset) = 0$.

(2) If $x > 0$,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(Z^2 \leq x) \\ &= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\ &= \int_{-\sqrt{x}}^{\sqrt{x}} \varphi(z) dz. \end{aligned}$$

Therefore,

$$f_X(x) = F'_X(x) = \frac{1}{2\sqrt{x}} [\varphi(\sqrt{x}) + \varphi(-\sqrt{x})] = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}.$$

Thus, the PDF $f_X(x)$ of X is

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

This is the PDF of Gamma $\left(\frac{1}{2}, 2\right)$ and it is called the chi-square distribution with 1 degree of freedom. $X \sim \chi^2(1)$.

Note: recall that X follows gamma distribution $X \sim \text{Gamma}(\alpha, \beta)$ if its PDF is given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

where $\alpha > 0, \beta > 0$ are two parameters, $\Gamma(\alpha)$ is a gamma function defined by

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

Its mean and variance are $E(X) = \alpha\beta, V(X) = \alpha\beta^2$.

Definition 6.5.1 Let $Z_1, \dots, Z_n$ be independent random variables with $Z_i \sim N(0,1)$, then

the statistic $Z_1^2 + \cdots + Z_n^2$ follows the chi-square distribution with n degrees of freedom.

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n).$$

Note: for $i=1, \cdots, n, Z_i^2 \sim \chi^2(1) = \text{Gamma}\left(\frac{1}{2}, 2\right)$. Applying the additive property of the gamma distribution(see Theorem 3.5.3), we have

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n) = \text{Gamma}\left(\frac{n}{2}, 2\right).$$

Thus, the PDF of $X \sim \chi^2(n)$ is

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

For various values of the degrees of freedom, the PDF of $X \sim \chi^2(n)$ is depicted in Figure 6.5.

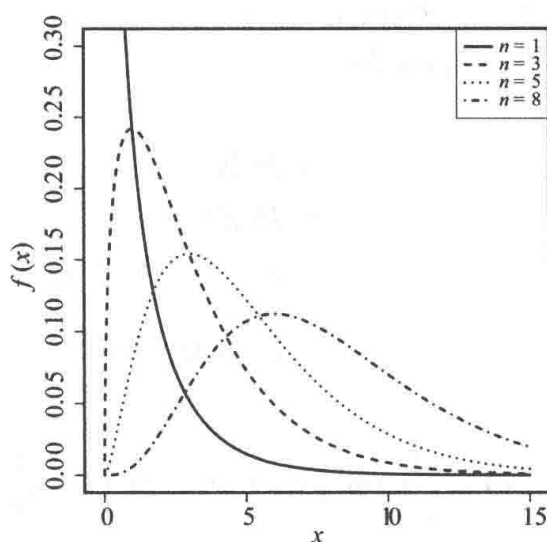


Figure 6.5

Theorem 6.5.2 Suppose $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$ and X, Y are independent, then $X + Y \sim \chi^2(n_1 + n_2)$.

Proof: It follows the additive property of the gamma distribution(see Theorem 3.5.3).

Theorem 6.5.3 Suppose $X \sim \chi^2(n)$, then $E(X) = n, V(X) = 2n$.

Proof: (1) Since $X \sim \chi^2(n)$, X can be written as $X = Z_1^2 + \cdots + Z_n^2$, where $Z_i \sim N(0, 1)$ ($i=1, 2, \cdots, n$) are independent.

$$E(Z_i^2) = V(Z_i) = 1.$$

$$V(Z_i^2) = E(Z_i^4) - [E(Z_i^2)]^2 = 3 - 1 = 2.$$

The above equation holds since we have the following fact: if $Z \sim N(0, 1)$ and n is even, then $E(Z^n) = (n-1) \times (n-3) \times \cdots \times 1$, see Example 4.6.1. Thus,

$$E(X) = E(Z_1^2 + \cdots + Z_n^2) = \sum_{i=1}^n E(Z_i^2) = n.$$

$$V(X) = V(Z_1^2 + \cdots + Z_n^2) = \sum_{i=1}^n V(Z_i^2) = 2n.$$

(2) Notice that the chi-square distribution is a special case of gamma distribution.

$\chi^2(n) = \text{Gamma}\left(\frac{n}{2}, 2\right)$, for $\text{Gamma}(\alpha, \beta)$, its mean and variance are $E(X) = \alpha\beta, V(X) = \alpha\beta^2$. Thus,

$$E(X) = \frac{n}{2} \times 2 = n, V(X) = \frac{n}{2} \times 2^2 = 2n.$$

Note: in general, the shape of the chi-square distribution is right skewed. But as the degrees of freedom n is large enough, $X = Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n)$ converges to $N(n, 2n)$ because of the central limit theorem.

Definition 6.5.2 Suppose $X \sim N(0, 1)$ and $Y \sim \chi^2(n)$ are independent, then the ratio

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

follows the t -distribution with degrees of freedom equal to n , denoted by $T \sim t(n)$.

Theorem 6.5.4 Let $T \sim t(n)$, then its PDF is given by

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty.$$

Proof: Given $Y = y$, the conditional distribution of T is normal,

$$f_{T|Y}(t|y) = \frac{1}{\sqrt{2\pi} \sqrt{n/y}} e^{-\frac{t^2}{2 \times \frac{n}{y}}}, \quad -\infty < t < +\infty, y > 0.$$

The joint PDF of (T, Y) is

$$\begin{aligned} f(t, y) &= f_{T|Y}(t|y) \times f_Y(y) \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\frac{n}{y}}} e^{-\frac{t^2}{2 \times \frac{n}{y}}} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \\ &= \frac{1}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n+1}{2}}} y^{\frac{n+1}{2}-1} e^{-\frac{y}{2} \left(1 + \frac{t^2}{n}\right)}, \end{aligned}$$

for $-\infty < t < +\infty, y > 0$. Integrate the joint PDF of (T, Y) with respect to y to obtain the PDF of T .

$$\begin{aligned} f_T(t) &= \int_0^{+\infty} f(t, y) dy \\ &= \int_0^{+\infty} \frac{1}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n+1}{2}}} y^{\frac{n+1}{2}-1} e^{-\frac{y}{2} \left(1 + \frac{t^2}{n}\right)} dy \\ &= \frac{1}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n+1}{2}}} \int_0^{+\infty} y^{\frac{n+1}{2}-1} e^{-\frac{y}{2} \left(1 + \frac{t^2}{n}\right)} dy \\ &= \frac{1}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right) 2^{\frac{n+1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1 + \frac{t^2}{n}}\right)^{\frac{n+1}{2}} \end{aligned}$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < +\infty.$$

Note: (1) The t-distribution with 1 degree of freedom is known as the Cauchy distribution with the PDF

$$f(t) = \frac{1}{\pi(1+t^2)}, -\infty < t < +\infty.$$

(2) As the degrees of freedom n goes to infinity, by the Stirling's approximation, the t-distribution converges to $N(0, 1)$. For various values of the degrees of freedom, the PDF of $T \sim t(n)$ is depicted in Figure 6. 6.

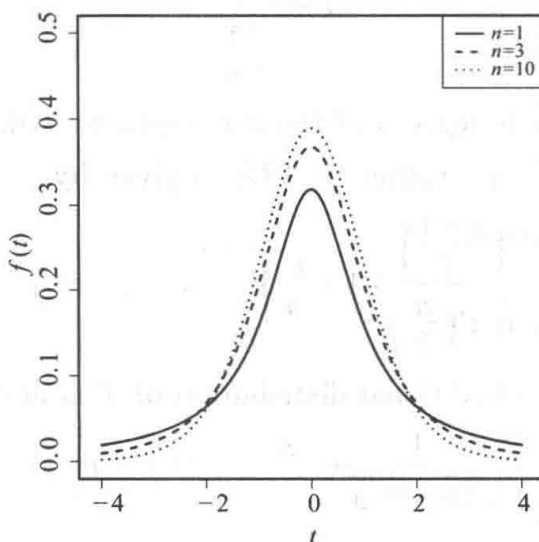


Figure 6. 6

Theorem 6. 5. 5 Let $T \sim t(n)$, then $E(T) = 0$, for $n > 1$; $V(T) = \frac{n}{n-2}$, for $n > 2$.

Proof: $T = \frac{X}{\sqrt{\frac{Y}{n}}}$, where $X \sim N(0, 1)$ and $Y \sim \chi^2(n)$ are independent. $E(X) = 0$,

$$E(X^2) = 1.$$

By independence, $E(T) = \sqrt{n}E(X)E(Y^{-\frac{1}{2}})$. If $n > 1$, $E(Y^{-\frac{1}{2}}) = \int_0^{+\infty} y^{-\frac{1}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy =$

$$\frac{\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})} < +\infty. \text{ Therefore, } E(T) = 0.$$

By independence, $E(T^2) = \sqrt{n}E(X^2)E\left(\frac{1}{Y}\right)$. If $n > 2$, $E\left(\frac{1}{Y}\right) = \int_0^{+\infty} y^{-1} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy =$

$$\frac{\Gamma(\frac{n-1}{2})}{2\Gamma(\frac{n}{2})} = \frac{1}{n-2}. \text{ Therefore, } V(T) = E(T^2) = \frac{n}{n-2}.$$

Definition 6.5.3 Suppose $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ are independent, then the ratio

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

follows the F -distribution with degrees of freedom equal to (m, n) , denoted by $F \sim F(m, n)$.

Note: if $F \sim F(m, n)$, then $\frac{1}{F} \sim F(n, m)$.

Theorem 6.5.6 Let $F \sim F(m, n)$, then its PDF is given by

$$f(u) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} \left(1 + \frac{m}{n}u\right)^{-\frac{m+n}{2}}, & u > 0 \\ 0, & u \leq 0 \end{cases}$$

Proof: For convenience, let $U = F = \frac{X}{Y}$. Given $Y = y$, the conditional distribution of U

is $U|Y=y \sim \frac{n}{my}X$,

$$f_{U|Y}(u|y) = f_X\left(\frac{myu}{n}\right) \frac{my}{n} = \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} \left(\frac{myu}{n}\right)^{\frac{m}{2}-1} e^{-\frac{myu}{2n}} \frac{my}{n}, u > 0, y > 0.$$

The joint PDF of (U, Y) is

$$\begin{aligned} f(u, y) &= f_{U|Y}(u|y) \times f_Y(y) \\ &= \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} \left(\frac{myu}{n}\right)^{\frac{m}{2}-1} e^{-\frac{myu}{2n}} \frac{my}{n} \times \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \\ &= \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} y^{\frac{m+n}{2}-1} e^{-\left(\frac{mu+n}{2n}\right)y}, \end{aligned}$$

for $u > 0, y > 0$. Integrate the joint PDF of (U, Y) with respect to y to obtain the PDF of U .

$$\begin{aligned} f_U(u) &= \int_0^{+\infty} f(u, y) dy \\ &= \int_0^{+\infty} \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} y^{\frac{m+n}{2}-1} e^{-\left(\frac{mu+n}{2n}\right)y} dy \\ &= \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} \int_0^{+\infty} y^{\frac{m+n}{2}-1} e^{-\left(\frac{mu+n}{2n}\right)y} dy \\ &= \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\left(\frac{mu+n}{2n}\right)^{\frac{m+n}{2}}} \end{aligned}$$

$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} u^{\frac{m}{2}-1} \left(1 + \frac{m}{n}u\right)^{-\frac{m+n}{2}}, u > 0.$$

For various values of the degrees of freedom, the PDF of $F \sim F(m, n)$ is depicted in Figure 6. 7.

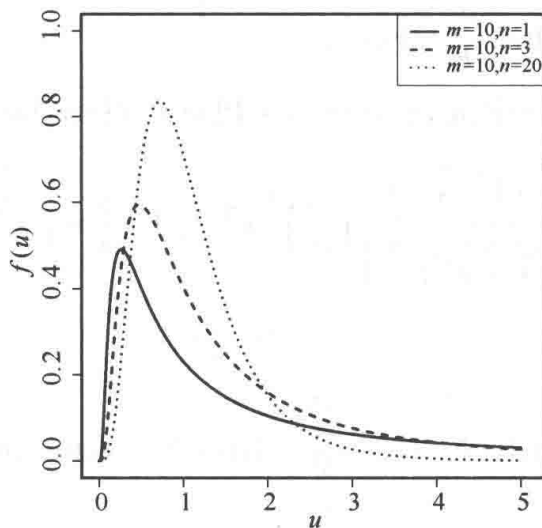


Figure 6. 7

Theorem 6. 5. 7 Let $F \sim F(m, n)$, then $E(F) = \frac{n}{n-2}$, for $n > 2$; $V(F) = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$, for $n > 4$.

Proof: $F = \frac{\frac{X}{m}}{\frac{Y}{n}}$, where $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ are independent. $E(X) = m$, $E(X^2) = V(X) + [E(X)]^2 = 2m + m^2$.

By independence, $E(F) = \frac{n}{m} E(X) E(Y^{-1})$. If $n > 2$,

$$E\left(\frac{1}{Y}\right) = \int_0^{+\infty} y^{-1} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy = \frac{\Gamma\left(\frac{n-2}{2}\right)}{2\Gamma\left(\frac{n}{2}\right)} = \frac{1}{n-2}.$$

Therefore, $E(F) = \frac{n}{m} \times m \times \frac{1}{n-2} = \frac{n}{n-2}$.

By independence, $E(F^2) = \frac{n^2}{m^2} E(X^2) E(Y^{-2})$. If $n > 4$,

$$E(Y^{-2}) = \int_0^{+\infty} y^{-2} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dy = \frac{\Gamma\left(\frac{n-4}{2}\right)}{4\Gamma\left(\frac{n}{2}\right)} = \frac{1}{(n-2)(n-4)}.$$

Therefore, $E(F^2) = \frac{n^2}{m^2} \times (2m + m^2) \times \frac{1}{(n-2)(n-4)} = \frac{n^2(2+m)}{m(n-2)(n-4)}$.

$$V(F) = E(F^2) - [E(F)]^2$$

$$\begin{aligned}
&= \frac{n^2(2+m)}{m(n-2)(n-4)} - \left(\frac{n}{n-2}\right)^2 \\
&= \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}.
\end{aligned}$$

Theorem 6.5.8 Let $T \sim t(n)$, then $T^2 \sim F(1, n)$.

Proof: $T = \frac{X}{\sqrt{\frac{Y}{n}}}$, where $X \sim N(0, 1)$ and $Y \sim \chi^2(n)$ are independent. $X^2 \sim \chi^2(1)$.

Therefore,

$$T^2 = \frac{\frac{X^2}{1}}{\frac{Y}{n}} \sim F(1, n)$$

Suppose X is a random variable, $F(x) = P(X \leq x)$ is its CDF. For any real x , $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$. Sometimes, we need consider the inverse problem. That is, if $P(X > x) = \alpha$ is given, determine the value of x .

Definition 6.5.4 Suppose the CDF of X is $F(x)$, $0 < \alpha < 1$, then x_α is called upper α quantile or α quantile if

$$P(X > x_\alpha) = \alpha.$$

Note: some textbooks define the lower quantile using the left tail.

- For the standard normal distribution $Z \sim N(0, 1)$, the α quantile is denoted by z_α .

$$\begin{aligned}
P(Z > z_\alpha) &= \int_{z_\alpha}^{+\infty} \varphi(x) dx = \alpha. \\
\Phi(z_\alpha) &= 1 - \alpha.
\end{aligned}$$

$z_\alpha = \Phi^{-1}(1 - \alpha)$ can be found in the standard normal table. By symmetry, $z_{1-\alpha} = -z_\alpha$. For example, $z_{0.05} = 1.645$, $z_{0.95} = -z_{0.05} = -1.645$.

- For the chi-square distribution $X \sim \chi^2(n)$, the α quantile is denoted by $\chi_\alpha^2(n)$.

$$P(X > \chi_\alpha^2(n)) = \alpha.$$

$\chi_\alpha^2(n)$ can be found in the χ^2 table. For example, $\chi_{0.90}^2(20) = 12.443$. If $n > 30$, $\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$.

- For the t-distribution $T \sim t(n)$, the α quantile is denoted by $t_\alpha(n)$.

$$P(T > t_\alpha(n)) = \alpha.$$

$t_\alpha(n)$ can be found in the t table. By symmetry, $t_{1-\alpha}(n) = -t_\alpha(n)$. For example, $t_{0.05}(4) = 2.1318$, $t_{0.95}(4) = -t_{0.05}(4) = -2.1318$. If $n > 30$, $t_\alpha(n) \approx z_\alpha$.

- For the F-distribution $F \sim F(m, n)$, the α quantile is denoted by $F_\alpha(m, n)$.

$$P(F > F_\alpha(m, n)) = \alpha.$$

$F_\alpha(m, n)$ can be found in the F table. By the fact that if $F \sim F(m, n)$, then $\frac{1}{F} \sim F(n, m)$,

thus

$$F_{1-\alpha}(m, n) = \frac{1}{F_\alpha(n, m)}.$$

For example, $F_{0.95}(7, 4) = \frac{1}{F_{0.05}(4, 7)} = \frac{1}{4.12} = 0.2427$.

Theorem 6.5.9 Suppose $(X_1, \dots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ is the sample mean and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is the sample variance. Then

(1) $\bar{X}$ and S^2 are independent;

(2) $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$;

(3) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;

(4) $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$.

Proof: The joint PDF of $(X_1, \dots, X_n)$ is

$$f(x_1, \dots, x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n x_i^2 - 2n\bar{x}\mu + n\mu^2}{2\sigma^2}}.$$

Transform the random variables $X_i (i=1, \dots, n)$ to

$$(Y_1, \dots, Y_n)^T = A(X_1, \dots, X_n)^T,$$

where A is an orthogonal matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2 \times 1}} & -\frac{1}{\sqrt{2 \times 1}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{3 \times 2}} & \frac{1}{\sqrt{3 \times 2}} & -\frac{2}{\sqrt{3 \times 2}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n \times (n-1)}} & \frac{1}{\sqrt{n \times (n-1)}} & \frac{1}{\sqrt{n \times (n-1)}} & \dots & -\frac{n-1}{\sqrt{n \times (n-1)}} \end{bmatrix}.$$

The Jacobian of the transformation can be shown to be 1. Notice that

$$\bar{X} = \frac{1}{\sqrt{n}} Y_1,$$

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2.$$

The joint PDF of $(Y_1, \dots, Y_n)$ is

$$\begin{aligned} f(y_1, \dots, y_n) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n y_i^2 - 2\sqrt{n}y_1\mu + n\mu^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=2}^n y_i^2 + (y_1 - \sqrt{n}\mu)^2}{2\sigma^2}}. \end{aligned}$$

Therefore, $Y_i (i=1, \dots, n)$ are independent and $Y_i \sim N(0, \sigma^2) (i=2, \dots, n)$, $Y \sim N(\sqrt{n}\mu, \sigma^2)$.

Thus, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$, the statement (2) is proved.

Since $(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=2}^n Y_i^2$, (1) is proved.

Since $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^n \left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2(n-1)$, (3) is proved.

Since $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ are independent, by the definition of

t-distribution,

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

Thus, (4) is proved.

Theorem 6.5.10 Suppose $(X_1, X_2, \dots, X_{n_1})$ and $(Y_1, Y_2, \dots, Y_{n_2})$ are independent random samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. $\bar{X}, \bar{Y}, S_1^2, S_2^2$ are the sample means and sample variances of the two samples. Then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2).$$

Proof:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

By Theorem 6.5.9, $\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1)$, $\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$, and they are independent. Applying the additive property of the chi-square distribution,

$$\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

By Theorem 6.5.9, $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ and $\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2}$ are independent. Thus,

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\frac{\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2}}{n_1 + n_2 - 2}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2).$$

Theorem 6.5.11 Suppose $(X_1, X_2, \dots, X_{n_1})$ and $(Y_1, Y_2, \dots, Y_{n_2})$ are independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. S_1^2 and S_2^2 are the sample variances. Then

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 2).$$

Proof: By Theorem 6.5.9, $\frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$, $\frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$, and they are independent. By the definition of the F-distribution,

$$\frac{\frac{(n_1 - 1)S_1^2}{\sigma_1^2}}{\frac{(n_2 - 1)S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 1).$$

That is,

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 1).$$

◆ 6.6 Confidence Intervals ◆

In section 1 and 2, we discussed point estimates for parameters from the sample. They provide one single value about the parameter. But a point estimate provides no information about the precision and reliability of estimation. In practice, one needs to calculate and report an entire interval of plausible values, that is a confidence interval(CI).

Definition 6.6.1 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim F(x; \theta)$, where θ is unknown. For pre-specified $0 < \alpha < 1$, an interval estimate gives an interval such that the true value is falling into the interval with a given probability.

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha.$$

$(\hat{\theta}_L, \hat{\theta}_U) = (\hat{\theta}_L(X_1, \dots, X_n), \hat{\theta}_U(X_1, \dots, X_n))$ is called a $1 - \alpha$ confidence interval. Statistic $\hat{\theta}_L$ is the confidence lower limit, $\hat{\theta}_U$ is the confidence upper limit, and $1 - \alpha$ is the confidence level.

The most frequently used confidence levels $1 - \alpha$ are 90%, 95% and 99%. The higher the confidence level, the more strongly we believe that the true value of the parameter being estimated lies within the interval.

Example 6.6.1 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim N(\mu, \sigma_0^2)$, where μ

is unknown and σ_0^2 is known. Find a $1-\alpha$ confidence interval of μ .

Solution: The key is the sampling distribution of the point estimate. Since $\bar{X} \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$, then

$$\frac{\bar{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0,1).$$

$\frac{\bar{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}}$ contains the unknown parameter μ but its distribution does not depend on μ . For

confidence level $1-\alpha$,

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha,$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ upper quantile of $N(0,1)$, see Figure 6.8. Therefore,

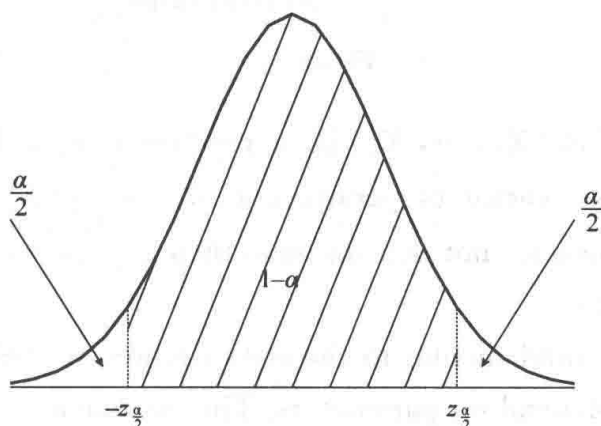


Figure 6.8

$$P\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = 1-\alpha.$$

$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)$ is a $1-\alpha$ confidence interval of μ .

It is NOT the probability of the parameter being inside the interval since the parameter is not random. The meaning of confidence interval is as follows: if we select K simple random samples of size n , $(x_{1k}, x_{2k}, \dots, x_{nk})$, $k=1, \dots, K$, and we calculate the $1-\alpha$ confidence interval from each sample, then about $K \times (1-\alpha)$ of those K intervals would contain the true parameter. For example, suppose $(X_1, \dots, X_{36})$ is a sample from $N(20, 16)$, We can repeatedly choose 100 samples of size 36 and build 95% confidence intervals for each sample. See Figure 6.9. The confidence intervals that do not contain the population mean $\mu=20$ are marked with thicker lines.

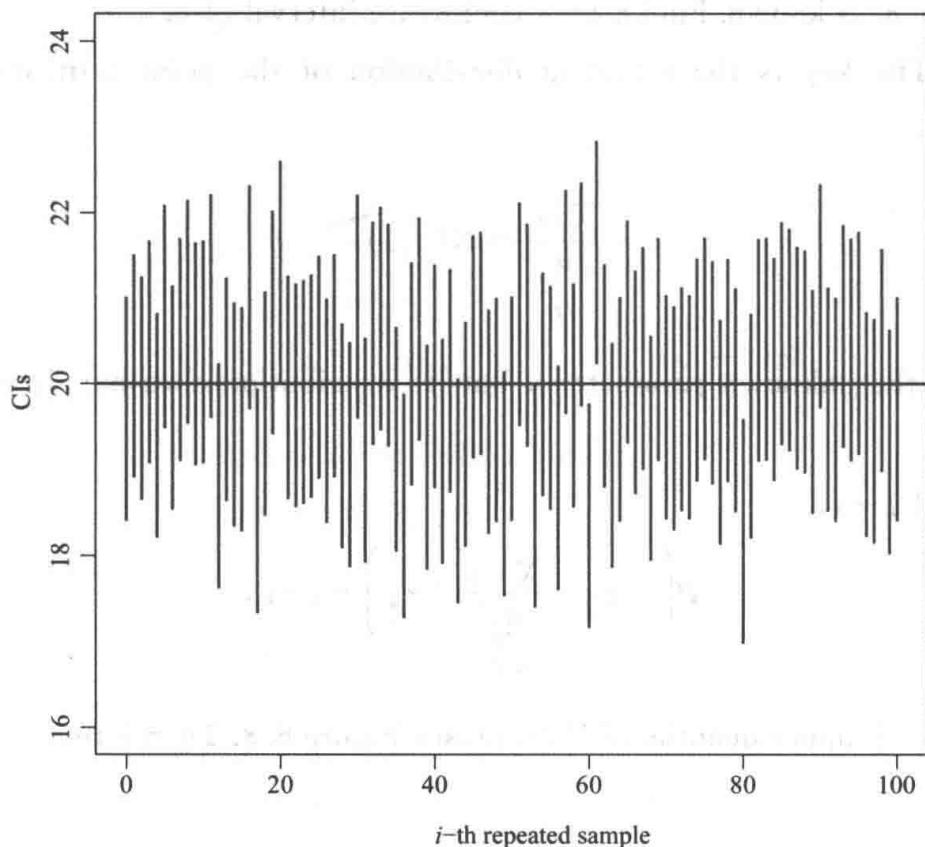


Figure 6.9

Definition 6.6.2 Let $(X_1, \dots, X_n)$ be a random sample from a distribution that depends on a parameter (or vector of parameters) θ . Let $g(X_1, \dots, X_n, \theta)$ be a random variable whose distribution does not depend on θ , then $g(X_1, \dots, X_n, \theta)$ is called a pivotal quantity (or simply a pivot).

Pivotal quantities are fundamental to the construction of confidence intervals, as they allow the statistic to not depend on parameters. The confidence intervals have exactly the stated confidence level, as opposed to so-called “asymptotic” confidence intervals which only have approximately the stated confidence level. The confidence interval can be constructed as follows:

(1) Find a pivotal quantity $g(X_1, \dots, X_n, \theta)$.

(2) For a specified confidence level, find confidence upper limits and confidence lower limit on the pivotal quantity, that is, numbers a and b such that

$$P(a < g(X_1, \dots, X_n, \theta) < b) = 1 - \alpha.$$

(3) Solve the inequalities; the confidence interval is

$$\hat{\theta}_L < \theta < \hat{\theta}_U.$$

The constants a and b are called critical values. They are obtained from a table for the distribution of the pivotal quantity or from a computer program. In general, they are not unique and thus the confidence interval is not unique. A convenient choice is to use equal-tailed critical values as this often leads to the shortest confidence interval.

Proposition 6.6.1 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim N(\mu, \sigma^2)$, if σ^2

$=\sigma_0^2$ is known, then a $1-\alpha$ confidence interval of μ is

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}} \right).$$

We derived this interval in Example 6.6.1.

Example 6.6.2 The cost of the textbooks for statistics courses is normally distributed, and the standard deviation is \$100. A sample of fifty-five textbooks is selected and the mean cost is \$124. Construct the 95% confidence interval for the mean cost of all statistics textbooks.

Solution: We know the population standard deviation $\sigma = \$100$. Let μ be the mean cost of all statistics textbooks.

$n = 55, \bar{x} = 124, 1-\alpha = 95\%, z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$. The confidence interval for μ is

$$\begin{aligned} \left(\bar{x} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}} \right) &= \left(124 - 1.96 \times \frac{100}{\sqrt{55}}, 124 + 1.96 \times \frac{100}{\sqrt{55}} \right) \\ &\approx (97.6, 150.4). \end{aligned}$$

We are 95% confident that the mean cost is between \$97.6 and \$150.4.

Proposition 6.6.2 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim N(\mu, \sigma^2)$, if σ^2 is unknown, then a $1-\alpha$ confidence interval of μ is

$$\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right).$$

Proof: If the population standard deviation σ is unknown, we can not apply the procedures above. We can not continue to use the standard normal distribution or Z-table. Since the sample variance S^2 is an unbiased estimator of σ^2 , we use the following pivotal quantity

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

For confidence level $1-\alpha$,

$$P\left(-t_{\frac{\alpha}{2}}(n-1) < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha,$$

where $t_{\frac{\alpha}{2}}(n-1)$ is the $\frac{\alpha}{2}$ upper quantile of $t(n-1)$. Therefore,

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

Example 6.6.3 A sample of 23 first-year students at a large university has a mean SAT math score of 525 with a standard deviation of 88. Construct the 90% confidence interval for the mean SAT math score for the entire first-year class.

Solution: We do not know the population standard deviation σ . Let μ be the mean SAT math score of all first-year students at this university. $n = 23$, sample mean $\bar{x} = 525$, sample standard deviation $s = 88$, $1-\alpha = 90\%$, $DF = n-1 = 22$, $t_{\frac{\alpha}{2}}(22) = 1.7171$. The

confidence interval for μ is

$$\begin{aligned} & \left(\bar{x} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) < \mu < \bar{x} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right) \\ & = \left(525 - \frac{88}{\sqrt{23}} \times 1.7171, 525 + \frac{88}{\sqrt{23}} \times 1.7171 \right) \\ & \approx (493.5, 556.5). \end{aligned}$$

Proposition 6.6.3 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim N(\mu, \sigma^2)$, if μ is known, then a $1-\alpha$ confidence interval of σ^2 is

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right).$$

Proof: If the population mean μ is known, we want to find a confidence interval of σ^2 . The following pivotal quantity can be used.

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n).$$

For confidence level $1-\alpha$,

$$P\left[\chi_{1-\frac{\alpha}{2}}^2(n) < \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 < \chi_{\frac{\alpha}{2}}^2(n)\right] = 1-\alpha,$$

where $\chi_{\frac{\alpha}{2}}^2(n)$ and $\chi_{1-\frac{\alpha}{2}}^2(n)$ are the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ upper quantiles of $\chi^2(n)$, respectively. See Figure 6.10. Therefore,

$$P\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)}\right] = 1-\alpha.$$

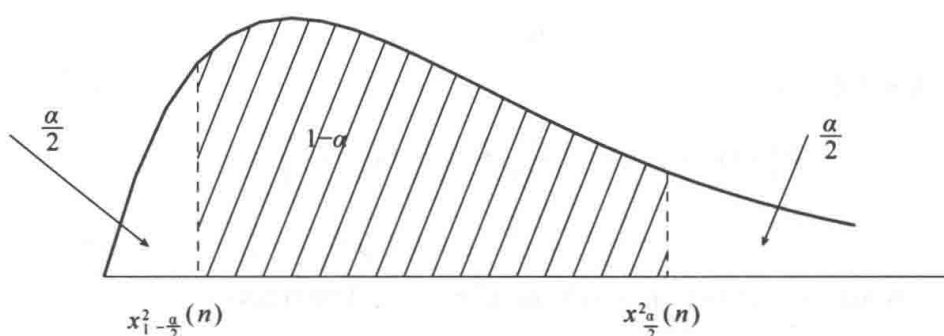


Figure 6.10

Proposition 6.6.4 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim N(\mu, \sigma^2)$, if μ is unknown, then a $1-\alpha$ confidence interval of σ^2 is

$$\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right).$$

Proof: If the population mean μ is unknown, we want to find a confidence interval of σ^2 . The following pivotal quantity can be used.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

For confidence level $1-\alpha$,

$$P\left[\chi_{1-\frac{\alpha}{2}}^2(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi_{\frac{\alpha}{2}}^2(n-1)\right] = 1-\alpha,$$

where $\chi_{\frac{\alpha}{2}}^2(n-1)$ and $\chi_{1-\frac{\alpha}{2}}^2(n-1)$ are the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ upper quantiles of $\chi^2(n-1)$, respectively. Therefore,

$$P\left[\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}\right] = 1-\alpha.$$

Example 6.6.4 The lifetimes of six electronic units are measured. The observations are 5.2, 5.5, 5.4, 5.8, 6.1, 6.3 (in 1000 hours). Assuming the lifetimes are normally distributed, construct a 95% confidence interval for the population standard deviation.

Solution: $n=6$, sample variance $s^2=0.182$, $\alpha=0.05$, $\chi_{\frac{\alpha}{2}}^2(n-1)=\chi_{0.025}^2(5)=12.832$, $\chi_{1-\frac{\alpha}{2}}^2(n-1)=\chi_{0.975}^2(5)=0.831$, a 95% confidence interval for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}\right) = \left(\frac{5 \times 0.182}{12.832}, \frac{5 \times 0.182}{0.831}\right) \approx (0.071, 1.095).$$

Thus, a 95% confidence interval for the population standard deviation σ is

$$\left(\sqrt{\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}}\right) \approx (0.266, 1.046).$$

In practice, we often want to estimate the difference between two population means and the ratio between two population variances. Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let $\bar{X}$ and S_1^2 be sample mean and sample variance from X , let $\bar{Y}$ and S_2^2 be sample mean and sample variance from Y . The confidence level is set to be $1-\alpha$.

Proposition 6.6.5 Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let $\bar{X}$ and S_1^2 be sample mean and sample variance from X , let $\bar{Y}$ and S_2^2 be sample mean and sample variance from Y , if σ_1^2 and σ_2^2 are known, then a $1-\alpha$ confidence interval of $\mu_1 - \mu_2$ is

$$\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right).$$

Proof: If σ_1^2 and σ_2^2 are known, we use the following pivotal quantity

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1).$$

For confidence level $1-\alpha$,

$$P\left[-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < z_{\frac{\alpha}{2}}\right] = 1-\alpha,$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ upper quantile of $N(0,1)$. Therefore,

$$P\left[\bar{X}-\bar{Y}-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}<\mu_1-\mu_2<\bar{X}-\bar{Y}+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right]=1-\alpha.$$

Proposition 6.6.6 Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let $\bar{X}$ and S_1^2 be sample mean and sample variance from X , let $\bar{Y}$ and S_2^2 be sample mean and sample variance from Y , if σ_1^2 and σ_2^2 are unknown, then a $1-\alpha$ confidence interval of $\mu_1 - \mu_2$ is

$$\left(\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(m+n-2)S_p\sqrt{\frac{1}{m}+\frac{1}{n}}, \bar{X}-\bar{Y}+t_{\frac{\alpha}{2}}(m+n-2)S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right),$$

where $S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$.

Proof: If σ_1^2 and σ_2^2 are unknown, but $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we use the following pivotal quantity

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2).$$

For confidence level $1-\alpha$,

$$P\left[-t_{\frac{\alpha}{2}}(m+n-2) < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}} < t_{\frac{\alpha}{2}}(m+n-2)\right] = 1-\alpha,$$

where $t_{\frac{\alpha}{2}}(m+n-2)$ is the $\frac{\alpha}{2}$ upper quantile of $t(m+n-2)$. Let

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

be the pooled estimator of σ^2 . Then,

$$P\left[\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(m+n-2)S_p\sqrt{\frac{1}{m}+\frac{1}{n}}<\mu_1-\mu_2<\bar{X}-\bar{Y}+t_{\frac{\alpha}{2}}(m+n-2)S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right]=1-\alpha.$$

Example 6.6.5 Comparison of species diversity index(SDI) in two stations; downstream station(D) and upstream station(U). Assume that populations in two locations(D and U) are approximately normal with equal variances. For 12 monthly samples collected in D the mean value for the SDI is $\bar{x}_1 = 3.11$ and standard deviation is $s_1 = 0.771$. For 10 monthly samples from U gives mean $\bar{x}_2 = 2.04$ and standard deviation $s_2 = 0.448$. Find 90% confidence interval for $\mu_D - \mu_U$.

Solution: Our point estimate of $\mu_D - \mu_U$ is $\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07$. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ are unknown but equal. The pooled estimate s_p^2 of the common variance σ^2 is

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1-1+n_2-1} = \frac{11 \times 0.771^2 + 9 \times 0.448^2}{12+10-2} \approx 0.417.$$

Taking the square root, we obtain $s_p = \sqrt{0.417} \approx 0.646$. The confidence level = 90%. $\alpha = 0.1$. $t_{\frac{\alpha}{2}}(20) = t_{0.05}(20) = 1.7247$ for $DF = n_1 + n_2 - 2 = 20$. Therefore, the 90% confidence interval for $\mu_D - \mu_U$ is

$$\left(1.07 - 1.7247 \times 0.646 \times \sqrt{\frac{1}{12} + \frac{1}{10}}, 1.07 + 1.7247 \times 0.646 \times \sqrt{\frac{1}{12} + \frac{1}{10}}\right) \\ \approx (0.593, 1.547).$$

Proposition 6.6.7 Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let S_1^2 be the sample variance from X , let S_2^2 be the sample variance from Y , if μ_1 and μ_2 are unknown, then a $1-\alpha$ confidence interval of $\frac{\sigma_1^2}{\sigma_2^2}$ is

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)}, \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(m-1, n-1)} \right).$$

Proof: We use the following pivotal quantity

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(m-1, n-1).$$

For confidence level $1-\alpha$,

$$P\left[F_{1-\frac{\alpha}{2}}(m-1, n-1) < \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} < F_{\frac{\alpha}{2}}(m-1, n-1)\right] = 1-\alpha,$$

where $F_{\frac{\alpha}{2}}(m-1, n-1)$ and $F_{1-\frac{\alpha}{2}}(m-1, n-1)$ are the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ upper quantiles of $F(m-1, n-1)$, respectively. Therefore,

$$P\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(m-1, n-1)}\right] = 1-\alpha.$$

Example 6.6.6 Among 11 patients in a certain study, the standard deviation of the property of interest is 5.8. In another group of 4 patients, the standard deviation is 3.4. We wish to construct a 95 percent confidence interval for the ratio of the variances of these two normal populations.

Solution: $m=11, s_1=5.8, n=4, s_2=3.4, \alpha=0.05, F_{0.025}(10, 3)=14.4, F_{0.975}(10, 3)=\frac{1}{F_{0.025}(3, 10)}=\frac{1}{4.83}=0.207$, a 95 percent confidence interval for the ratio of the variances is

$$\left(\frac{5.8^2}{3.4^2} \times \frac{1}{14.4}, \frac{5.8^2}{3.4^2} \times \frac{1}{0.207}\right) \approx (0.202, 14.058).$$

We have discussed several exact confidence intervals under normality. In practice, the exact distribution of a pivotal quantity under non-normal population is hard to be determined. In this case, the central limit theorem and law of large numbers can be applied.

An asymptotically pivotal quantity can be used to construct, so-called asymptotic or large sample confidence intervals if the asymptotic distribution does not depend on parameters. The method of asymptotic pivotal quantities proceeds just likes the method of exact pivotal quantities.

Example 6.6.7 Suppose $(X_1, \dots, X_n) (n \geq 30)$ is a sample from $X \sim B(1, p)$. Find a $1-\alpha$ confidence interval of p .

Solution: By the Laplace central limit theorem, as n is large enough,

$$\bar{X} \simeq N\left(p, \frac{p(1-p)}{n}\right).$$

Thus,

$$\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \simeq N(0, 1).$$

$\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is an asymptotic pivotal quantity. For confidence interval $1-\alpha$,

$$P\left[-z_{\frac{\alpha}{2}} < \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\frac{\alpha}{2}}\right] \approx 1-\alpha.$$

The event inside the brackets is equivalent to

$$(n+z_{\frac{\alpha}{2}}^2)p^2 - (2n\bar{X}+z_{\frac{\alpha}{2}}^2)p + n\bar{X}^2 < 0.$$

Let $a=n+z_{\frac{\alpha}{2}}^2$, $b=-(2n\bar{X}+z_{\frac{\alpha}{2}}^2)$, $c=n\bar{X}^2$, the discriminant of this quadratic form is

$$b^2-4ac=4n\bar{X}(1-\bar{X})z_{\frac{\alpha}{2}}^2+z_{\frac{\alpha}{2}}^4 > 0.$$

Let

$$p_L = \frac{(-b - \sqrt{b^2-4ac})}{2a}, p_U = \frac{(-b + \sqrt{b^2-4ac})}{2a}.$$

See Figure 6.11. Thus,

$$P(p_L < p < p_U) \approx 1-\alpha.$$

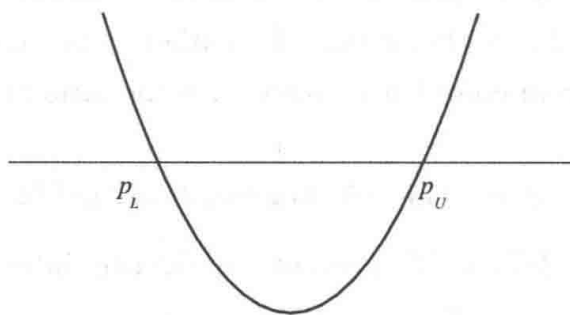


Figure 6.11

A $1-\alpha$ confidence interval of p is (p_L, p_U) .

The confidence intervals discussed so far give both a lower confidence bound and an upper confidence bound for the parameter being estimated. In some circumstances, it can make more sense to express the confidence interval in only one direction—with only a lower confidence limit, or only an upper confidence limit, but not both.

Definition 6.6.3 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim F(x; \theta)$, where θ is unknown. For pre-specified $0 < \alpha < 1$, if there exists a statistic $\hat{\theta}_L = \hat{\theta}_L(X_1, \dots, X_n)$

such that

$$P(\hat{\theta}_L < \theta) = 1 - \alpha.$$

Then $(\hat{\theta}_L, +\infty)$ is called a $1 - \alpha$ one-sided confidence interval of θ . Statistic $\hat{\theta}_L$ is the confidence lower limit, and $1 - \alpha$ is the confidence level.

Definition 6.6.4 Suppose $(X_1, \dots, X_n)$ is a random sample from $X \sim F(x; \theta)$, where θ is unknown. For pre-specified $0 < \alpha < 1$, if there exists a statistic $\theta_U = \hat{\theta}_U(X_1, \dots, X_n)$ such that

$$P(\theta < \hat{\theta}_U) = 1 - \alpha.$$

Then $(-\infty, \hat{\theta}_U)$ is called a $1 - \alpha$ one-sided confidence interval of θ . Statistic $\hat{\theta}_U$ is the confidence upper limit, and $1 - \alpha$ is the confidence level.

Example 6.6.8 The lifetimes of six electronic units are measured. The observations are 5.2, 5.5, 5.4, 5.8, 6.1, 6.3 (in 1000 hours). Assuming the lifetimes are normally distributed, determine a 95% confidence lower limit for the population variance.

Solution: The following pivotal quantity can be used.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

For confidence level $1 - \alpha$,

$$P\left[\frac{(n-1)S^2}{\sigma^2} < \chi_\alpha^2(n-1)\right] = 1 - \alpha,$$

where $\chi_\alpha^2(n-1)$ is the α upper quantile of $\chi^2(n-1)$, see Figure 6.12. Therefore,

$$P\left[\frac{(n-1)S^2}{\chi_\alpha^2(n-1)} < \sigma^2\right] = 1 - \alpha.$$

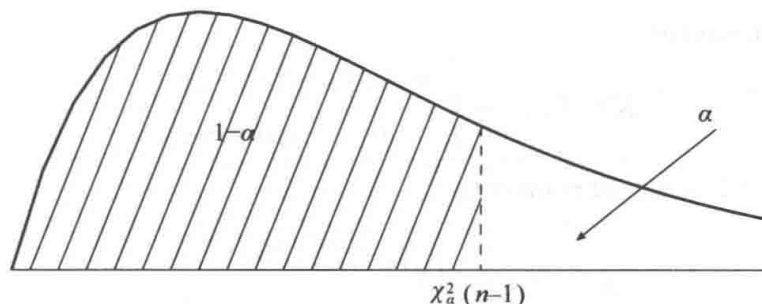


Figure 6.12

$n=6$, sample variance $s^2=0.182$, $\alpha=0.05$, $\chi_\alpha^2(n-1)=\chi_{0.05}^2(5)=11.07$, a 95% one-sided confidence interval for σ^2 is

$$\left(\frac{5 \times 0.182}{11.07}, +\infty\right) \approx (0.0822, +\infty).$$

The 95% one-sided confidence lower limit is 0.0822.

Exercise 6

6.1 Suppose $X \sim N(\mu, \sigma^2)$, where μ is known and σ^2 is unknown. $(X_1, X_2, \dots, X_n)$ is

a sample from this distribution.

(1) Determine the joint PDF of $(X_1, X_2, \dots, X_n)$;

(2) Identify the statistics from the following functions. $\sum_{i=1}^n X_i^2$, $\max_{1 \leq i \leq n} \{X_i\}$,

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}, \sum_{i=1}^n X_i + \mu.$$

6.2 Suppose we have the following sample of $n=15$ observations. 0.22, -0.67, -0.07, 1.65, 0.66, 0.88, 0.74, 0.72, -1.19, 0.14, -1.02, -0.21, -0.54, 0.99, 1.39. Find the empirical CDF and sketch its plot.

6.3 Suppose X follows exponential distribution with the PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$(X_1, X_2, \dots, X_n)$ is a sample from the population X , $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are ordered statistics. Find the PDF of the range $Y = X_{(n)} - X_{(1)}$.

6.4 Make a histogram for the following ages of 35 persons.

67	74	60	90	64	72	83
77	67	85	56	57	73	66
60	80	63	90	78	70	63
79	49	88	68	56	78	71
71	46	53	67	64	65	58

6.5 Suppose $(X_1, \dots, X_n)$ is a sample from population X , determine the moment estimators of the parameters.

$$(1) \quad X \sim f(x) = \begin{cases} (\theta+1)x^\theta, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where $\theta > -1$ is an unknown parameter.

$$(2) \quad X \sim f(x) = \begin{cases} \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases},$$

where $\theta > 0$ is an unknown parameter.

$$(3) \quad X \sim f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}, & x \geq \mu \\ 0, & \text{otherwise} \end{cases},$$

where $\theta > 0$ and μ are unknown parameters.

$$(4) \quad X \sim f(x) = \begin{cases} 1, & \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2} \\ 0, & \text{otherwise} \end{cases},$$

where θ is an unknown parameter.

6.6 Determine the maximum likelihood estimators of the parameters in previous problem.

6.7 Suppose $X \sim B(n, p)$, $0 < p < 1$, n is known. $(X_1, \dots, X_n)$ is a sample from population X , determine the moment estimator and maximum likelihood estimator of the parameter p .

6.8 Suppose $X \sim P(\lambda)$, where $\lambda > 0$ is a parameter. $(X_1, \dots, X_n)$ is a sample from population X . (1) Find the maximum likelihood estimator of λ ; (2) Is the maximum likelihood estimator unbiased?

6.9 Suppose that $X_1, \dots, X_n$ form a random sample from the uniform distribution on the interval $[\theta_1, \theta_2]$, where both θ_1 and θ_2 are unknown ($-\infty < \theta_1 < \theta_2 < +\infty$). Find the maximum likelihood estimators of θ_1 and θ_2 .

6.10 Suppose that $X_1, \dots, X_n$ form a random sample from an exponential distribution for which the value of the parameter λ is unknown. Determine the maximum likelihood estimator of the median of the distribution.

6.11 Suppose that $X_1, \dots, X_n$ form a random sample from a normal distribution for which both the mean and the variance are unknown. Find the maximum likelihood estimator of the 0.95 quantile of the distribution, that is, of the point x_0 such that $P(X \leq x_0) = 0.95$.

6.12 The PDF of X is given by

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x \leq \theta \\ 0, & \text{otherwise} \end{cases},$$

where $\theta > 0$ is an unknown parameter. (X_1, X_2, X_3) is a sample from X . Show that $\frac{4}{3} \max_{1 \leq i \leq 3} \{X_i\}$ and $4 \min_{1 \leq i \leq 3} \{X_i\}$ are unbiased estimators of θ . Which one is more efficient?

6.13 $(X_1, \dots, X_n)$ is a sample from X , determine the value of k such that $\hat{\sigma}^2 = k \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$ is an unbiased estimator of population variance σ^2 .

6.14 Suppose $X \sim N(\mu, \sigma^2)$, (X_1, X_2, X_3) is a sample from population X . Let

$$\hat{\mu}_1 = \frac{1}{5}X_1 + \frac{3}{10}X_2 + \frac{1}{2}X_3,$$

$$\hat{\mu}_2 = \frac{1}{3}X_1 + \frac{1}{4}X_2 + \frac{5}{12}X_3,$$

$$\hat{\mu}_3 = \frac{1}{3}X_1 + \frac{1}{6}X_2 + \frac{1}{2}X_3.$$

Show that the three statistics are unbiased estimators of μ . Which one is more efficient?

6.15 Suppose that X is a random variable whose distribution is completely unknown, but it is known that all the moments $E(X^k)$ ($k=1, 2, \dots$) are finite. Suppose also that $X_1, \dots, X_n$ form a random sample from this distribution. Show that the k -th sample moment $\frac{1}{n} \sum_{i=1}^n X_i^k$ is an unbiased estimator of $E(X^k)$.

6.16 Suppose that $X_1, \dots, X_n$ form a random sample from the geometric distribution

with unknown parameter p . Find a statistic that will be an unbiased estimator of $1/p$.

6.17 Suppose that $X_1, \dots, X_n$ form a random sample from the uniform distribution on the interval $[0, \theta]$, where the value of the parameter θ is unknown; and let $Y_n = \max\{X_1, \dots, X_n\}$. Show that $[(n+1)/n]Y_n$ is an unbiased estimator of θ .

6.18 Suppose $X_1, \dots, X_{36}$ form a random sample from $X \sim N(55, 6.3^2)$. Find the probability that the sample mean falls into the interval $(53.8, 56.8)$.

6.19 Suppose $X \sim N(0, 1)$, $(X_1, X_2, \dots, X_n)$ form a random sample from X . Find the distributions of the following statistics:

(1) $X_1 - X_2$;

(2) $\frac{X_1 - X_2}{\sqrt{X_3^2 + X_4^2}}$;

(3) $\frac{\sqrt{n-1} X_n}{\sum_{i=1}^{n-1} X_i^2}$;

(4) $\frac{n-5}{5} \cdot \frac{\sum_{i=1}^5 X_i^2}{\sum_{i=6}^n X_i^2}$.

6.20 Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2 . Find the distribution of $\frac{n(\bar{X} - \mu)^2}{\sigma^2}$.

6.21 Suppose that six random variables $X_1, \dots, X_6$ form a random sample from the standard normal distribution, and let

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2.$$

Determine a value of c such that the random variable cY will have a χ^2 -distribution.

6.22 Suppose that the random variables $X_1, \dots, X_5$ are IID and that each has the standard normal distribution. Determine a constant c such that the random variable $\frac{c(X_1 + X_2)}{(X_3^2 + X_4^2 + X_5^2)^{\frac{1}{2}}}$ will have a t -distribution.

6.23 Suppose that the random variables X_1 and X_2 are independent and that each has the normal distribution with mean μ and variance σ^2 . Determine the value of $P\left[\frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} < 4\right]$.

6.24 If $X \sim U(0, 1)$, show that $-2\ln X \sim \chi^2(2)$.

6.25 $(X_1, X_2, \dots, X_{10})$ is a sample from population $X \sim N(0, 0.3^2)$, compute $P\left(\sum_{i=1}^{10} X_i^2 > 1.44\right)$.

6.26 Suppose that a point (X, Y) is to be chosen at random in the xy -plane, where X and Y are independent random variables and each has the standard normal distribution. If a circle is drawn in the xy -plane with its center at the origin, what is the radius of the

smallest circle that can be chosen in order for there to be probability 0.99 that the point (X, Y) will lie inside the circle?

6.27 Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2 . Assuming that the sample size n is 16, determine the values of the following probabilities:

$$(1) P\left[\frac{1}{2}\sigma^2 \leq \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \leq 2\sigma^2\right];$$

$$(2) P\left[\frac{1}{2}\sigma^2 \leq \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \leq 2\sigma^2\right].$$

6.28 Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2 , and let S^2 denote the sample variance. Determine the smallest values of n for which the following relations are satisfied:

$$(1) P\left(\frac{S^2}{\sigma^2} \leq 1.5\right) \geq 0.95;$$

$$(2) P\left(|S^2 - \sigma^2| \leq \frac{1}{2}\sigma^2\right) \geq 0.8.$$

6.29 Suppose population $X \sim N(\mu, \sigma^2)$, $(X_1, X_2, \dots, X_n)$ is a sample from population X , $\bar{X}$ is the sample mean and S^2 is sample variance. Determine

$$(1) \text{ the value of } k \text{ such that } P(\mu - kS < \bar{X} < \mu + kS) = 0.9;$$

$$(2) \text{ the value of } a \text{ such that } P(S^2 > a) = 0.1 \text{ when } \sigma = 5.$$

6.30 Suppose that a random sample of 16 observations is drawn from the normal distribution with mean μ and standard deviation 12, and that independently another random sample of 25 observations is drawn from the normal distribution with the same mean μ and standard deviation 20. Let $\bar{X}$ and $\bar{Y}$ denote the sample means of the two samples. Evaluate $P(|\bar{X} - \bar{Y}| < 5)$.

6.31 Suppose that a random sample is to be taken from the normal distribution with unknown mean μ and standard deviation 2. Let $\bar{X}$ be its sample mean.

(1) How large a random sample must be taken in order that $E(|\bar{X} - \mu|^2) \leq 0.1$ for every possible value of μ ?

(2) How large a random sample must be taken in order that $E(|\bar{X} - \mu|) \leq 0.1$ for every possible value of μ ?

6.32 Here are the lengths (unit: cm) for 16 different nails of a certain brand:

2.14, 2.10, 2.13, 2.15, 2.13, 2.12, 2.13, 2.10,

2.15, 2.12, 2.14, 2.10, 2.13, 2.11, 2.14, 2.11.

Assume that these numbers are the observed values from a random sample of 16 independent normal random variables with mean μ and variance σ^2 . Find a 90% confidence interval for the mean length μ .

(1) $\sigma = 0.01$ (cm) is known. (2) σ is unknown.

6.33 Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution

with unknown mean μ and known variance σ^2 . How large a random sample must be taken in order that there will be a confidence interval for μ with confidence level 0.95 and length no larger than L ?

6.34 A random sample of 10 light bulbs has a mean lifetime of 1500 hours and the sample standard deviation of the lifetime is 20 hours. Assume that the lifetime follows normal distribution.

(1) Construct a 95% confidence interval estimate of the mean lifetime.

(2) Construct a 95% confidence interval estimate of the population standard deviation.

6.35 Suppose the amount of drying time(unit: h) of a particular paint follows $N(\mu, \sigma^2)$. The following data are 9 observations of a sample.

6.0, 5.7, 5.8, 6.5, 7.0, 6.3, 5.6, 6.1, 5.0.

(1) Construct a 95% confidence interval estimate of the mean drying time.

(2) Construct a 95% confidence interval estimate of the population standard deviation.

6.36 A research team is interested in the difference between serum uric acid levels in patients with and without Down's syndrome. In a large hospital for the treatment of the mentally retarded, a sample of 12 individuals with Down's syndrome yielded a mean of $\bar{x}_1 = 4.5$ mg/100 ml and sample variance $s_1^2 = 4.9$. In a general hospital a sample of 15 normal individuals of the same age and sex were found to have a mean value of $\bar{x}_2 = 3.4$ mg/100 ml and sample variance $s_2^2 = 5.6$. If it is reasonable to assume that the two populations of values are normally distributed with equal variances, find the 95 percent confidence interval for $\mu_1 - \mu_2$.

6.37 Among 11 patients in a certain study, the standard deviation of the property of interest was 5.8. In another group of 4 patients, the standard deviation was 3.4.

(1) Construct a 95% confidence interval for the ratio of the variances of these two populations.

(2) Find the lower limit of an one-sided 95% confidence interval.

6.38 Find the 99% confidence interval for the proportion of heads obtained if 175 heads were counted after 400 tosses of a coin. Is this coin balanced?

6.39 Suppose $X \sim P(\lambda)$, where $\lambda > 0$ is a parameter. $(X_1, \dots, X_n)$ is a sample from population X . Construct a $1-\alpha$ ($0 < \alpha < 1$) confidence interval for λ .

6.40 Suppose the population follows the exponential distribution with its PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases},$$

where $\lambda > 0$ is a parameter. $(X_1, \dots, X_n)$ is a sample from the population X . Find the lower limit of a one-sided $1-\alpha$ ($0 < \alpha < 1$) confidence interval for λ .

Chapter 7

Hypothesis Testing

If the population parameter is μ and the claim value is μ_0 , then some pairs of null and alternative hypotheses are:

$H_0: \mu \geq \mu_0$, $H_a: \mu < \mu_0$

$H_0: \mu \leq \mu_0$, $H_a: \mu > \mu_0$

$H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$

The first two pairs are labeled *right-tailed* and *left-tailed* tests, respectively. The third pair is labeled *two-tailed* test. If you are unsure which pair to use, ask yourself: "What am I trying to prove?"

If you are trying to prove that the population mean is less than a certain value, you would use the first pair. If you are trying to prove that the population mean is greater than a certain value, you would use the second pair. If you are trying to prove that the population mean is not equal to a certain value, you would use the third pair.

Statistical inference is the process of deducing properties of an underlying distribution by analysis of data. It includes deriving estimates and hypothesis testing. This chapter introduces the basics of hypothesis testing: the logic of hypothesis testing, test statistics, and decision making etc.

◆ 7.1 Basics of Hypothesis Testing ◆

In practice, we often want to determine whether or not we can reject some assumptions about a population given information about a sample. For example,

(1) Select 100 light bulbs and record their lifetimes. Based on information from the sample, can you reject the hypothesis that the mean lifetime for the entire population of bulbs is 600 hours?

(2) 40 students took an exam. Based on information from the sample, can you reject the hypothesis that the population mean score is greater than or equal to 80 points?

These are the problems of hypothesis testing.

Definition 7.1.1 A hypothesis is a statement or claim regarding a characteristic of one or more populations. Hypothesis testing (or test of significance) is a procedure, based on a sample evidence and probability, used to test claims regarding a characteristic of one or more populations.

To conduct a hypothesis test, one should state a pair of hypotheses in advance. The null hypothesis (denoted by H_0) is a population hypothesis that contains a statement of equality. The alternative hypothesis (denoted by H_1 or H_a) is the population statement that contains an inequality.

If the corresponding sample value is close to the value stated in the null hypothesis, the sample won't cause us to reject the null hypothesis. If the sample value is far away from the value stated in the null hypothesis, then the sample indicates that the null hypothesis isn't true. We thus reject the null hypothesis in favor of the alternative hypothesis.

If the population parameter is θ and the claim value is θ_0 , then some pairs of null and alternative hypotheses are

$$H_0: \theta \geq \theta_0, \quad H_1: \theta < \theta_0$$

$$H_0: \theta \leq \theta_0, \quad H_1: \theta > \theta_0$$

$$H_0: \theta = \theta_0, \quad H_1: \theta \neq \theta_0$$

The first two pairs are for one-sided test (left-tailed, right-tailed) and the third is for two-sided test (two-tailed).

Looking at the exam example from above, the null and alternative hypotheses are

$$H_0: \mu \geq 80, \quad H_1: \mu < 80,$$

where μ is the population mean exam score.

Once the data is collected and summarized, A decision should be made. Should we reject H_0 and accept H_1 . Or reject H_1 and accept H_0 ?

We would not expect the sample mean $\bar{x}$ to be exactly equal to population mean due to the deviation by chance. We need to know, is the event due to chance alone, or is there some cause? We need to have a method to compare events that easily occur by chance and those that are highly unlikely to occur randomly. If H_0 is true, the sample mean $\bar{x}$ will be close to the population mean. Very large sample mean $\bar{x}$ wouldn't allow us to reject the null hypothesis while sample mean much less than 80 would allow us to reject the null hypothesis.

A test statistic will be used in deciding whether or not to reject the null hypothesis. The test statistic is a value computed from the sample and converted into a standard score under the assumption that the null hypothesis is true. In many scenarios, there are standard ones that are used, like the Z-test, t-test, χ^2 -test, which we will discuss in this chapter.

One way to decide whether to reject the null hypothesis is to determine whether the standardized test statistic falls within a rejection region of the sampling distribution. Here we apply the small probability events principle—unlikely events do not occur.

The critical region(or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis. A critical value separates the rejection region from the non-rejection region.

When conducting a test and making the decision, our conclusion based on the sample will be to either reject the null hypothesis or fail to reject it. There are four possible outcomes, see Table 7. 1.

Table 7. 1

		Our decision	
		H_0	H_1
Reality	H_0	Correct Prob= $1-\alpha$	Type I error Prob= α
	H_1	Type II error Prob= β	Correct Prob= $1-\beta$

It is obvious that such conclusions are sometimes correct and sometimes not. There are two types of errors.

Definition 7. 1. 2 A type I error occurs when rejecting the null hypothesis when it is actually true. The probability of making a type I error is α .

Definition 7. 1. 3 A type II error occurs when failing to reject the null hypothesis, when it is actually false. The probability of making a type II error is β .

When conducting a hypothesis test, we would like the test to have small probability of type II error. And we would also like the probability of a type I error to be small. But as the probability of making one type of error falls, the probability of making the other type

risers. In general, tests are set up so as to minimize the probability of a type I error. The choice of the level of significance depends on the consequences of making type I error. If the consequences are severe, the level of significance should be small. Common used values for the significance level α are 0.05, 0.10 or 0.01. Its default value is 0.05.

Definition 7.1.4 The power of a test is $1 - \beta$, the probability of rejecting H_0 when H_1 is true.

There are two different methods usually used to conduct hypothesis tests. One is known as the traditional method (critical value approach), and the other is known as a p -value approach.

The critical value approach to hypothesis testing is as follows:

- (1) State the null and alternative hypothesis H_0 and H_1 .
- (2) Specify the level of significance α .
- (3) Find the test statistic, compute the test value, determine the critical region.
- (4) Make a decision to reject or fail to reject the null hypothesis. If the test statistic is in the critical region, then we reject the null hypothesis. If the test statistic is not in the critical region, then we fail to reject the null hypothesis.
- (5) Summarize the results in such a way that the original claim is addressed.

Definition 7.1.5 A p -value (or probability value) is the probability of getting a value of the sample test statistic that is at least as extreme as the one found from the sample data, assuming that null hypothesis is true.

H_0 is rejected if p -value is smaller than a pre-specified level. For left-tailed test, p -value is the area to the left of the test statistic. For right-tailed test, p -value is the area to the right of the test statistic. For two-tailed test, p -value is twice the area to the left of a negative test statistic or is twice the area to the right of positive test statistic.

The p -value approach is nearly identical to the critical value approach.

The p -value approach to hypothesis testing is as follows:

- (1) State the null and alternative hypothesis H_0 and H_1 .
- (2) Specify the level of significance α .
- (3) Find the test statistic, compute the test value, find the p -value.
- (4) Make a decision to reject or fail to reject the hypothesis. We reject the null hypothesis if p -value is less than or equal to α . We fail to reject the null hypothesis if the p -value is greater than α .
- (5) Summarize the results in such a way that the original claim is addressed.

◆ 7.2 Hypothesis Tests for a Population Mean ◆

Suppose $X \sim N(\mu, \sigma^2)$, $(X_1, X_2, \dots, X_n)$ is a sample and $x_1, x_2, \dots, x_n$ are its observations. For significant level α , we discuss the following sets of hypotheses:

- (1) $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$ (two-tailed test)
- (2) $H_0: \mu = \mu_0, H_1: \mu > \mu_0$ (right-tailed test)
- (3) $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$
- (4) $H_0: \mu = \mu_0, H_1: \mu < \mu_0$ (left-tailed test)
- (5) $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$

where μ_0 is a constant. Hypotheses (1), (2) and (4) have simple null hypotheses in which all parameters are specified. (3) and (5) have composite null hypotheses.

If population variance σ^2 is known, we discuss the hypothesis test for population mean μ . For case (2),

$$H_0: \mu = \mu_0, H_1: \mu > \mu_0.$$

Since $\bar{X}$ is the maximum likelihood estimator of μ , and $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Under H_0 , the test statistic

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

Suppose H_0 is true, $\bar{X}$ should be close to μ_0 . If $\bar{X} - \mu_0$ is large enough, say $\bar{X} - \mu_0 \geq k$, the small probability event occurs, H_0 should be rejected. Since the significance level is α ,

$$\alpha = P[\bar{X} - \mu_0 \geq k | H_0] = P\left[\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right] = P\left[\bar{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right].$$

Thus, the rejection region is as follows

$$\begin{aligned} W &= \left\{ (x_1, x_2, \dots, x_n) \mid \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha \right\} \\ &= \left\{ (x_1, x_2, \dots, x_n) \mid \bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\}. \end{aligned}$$

For p -value approach,

$$p = P\left[Z \geq \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right] = 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right),$$

where $\bar{x}$ is the observed sample mean.

For case (3),

$$H_0: \mu \leq \mu_0, H_1: \mu > \mu_0.$$

Under H_0 ,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

Thus,

$$\alpha = P\left[\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right] \geq P\left[\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right].$$

The rejection region is the same as that in (2).

$$\begin{aligned} W &= \left\{ (x_1, x_2, \dots, x_n) \mid \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha \right\} \\ &= \left\{ (x_1, x_2, \dots, x_n) \mid \bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\}. \end{aligned}$$

Their p -values are also identical. The same result applies to (4) and (5). In the textbook, we will not differentiate these null hypotheses.

For two-tailed test and left-tailed test, we can discuss the procedures in a similar way. The conclusions are summarized in Table 7.2.

Table 7.2

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
Z-test (σ^2 is known)	$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	$ z \geq z_{\frac{\alpha}{2}}$	$2 \left[1 - \Phi \left(\left \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right \right) \right]$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$z \geq z_\alpha$	$1 - \Phi \left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$z \leq z_{1-\alpha}$	$\Phi \left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)$

Example 7.2.1 All students at a college took a certain standardized test. The standard deviation of the score is 30. A teacher suspects that the mean score is greater than 120, she randomly selects 50 students and finds that their mean score is 128. (1) Conduct an appropriate hypothesis test. Use $\alpha = 0.05$. (2) Compute the power of the test.

Solution: (1) We propose three different methods to conduct an appropriate hypothesis test. Let μ be the mean score of all students at this college.

$$H_0: \mu \leq 120, \quad H_1: \mu > 120.$$

① Since $\sigma^2 = 30^2$ is known, we use the Z-test. The test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

The rejection region is

$$\begin{aligned}
 W &= \{(x_1, \dots, x_n) \mid z \geq z_{0.05}\} \\
 &= \{(x_1, \dots, x_n) \mid \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq 1.645\}.
 \end{aligned}$$

$n=50, \bar{x}=128$, the observation of Z is $z = \frac{128-120}{\frac{30}{\sqrt{50}}} \approx 1.89 \geq 1.645$. The sample data fall

in the rejection region and H_0 should be rejected. The sample data indicate that the mean score is greater than 120.

② We can also adopt the p -value approach to hypothesis testing. The z -score is

$$z = \frac{128-120}{\frac{30}{\sqrt{50}}} \approx 1.89.$$

The p -value is the probability that we obtain the observed value of the test statistic that is more extreme in the direction of the alternative hypothesis, calculated when H_0 is true. See Figure 7.1.

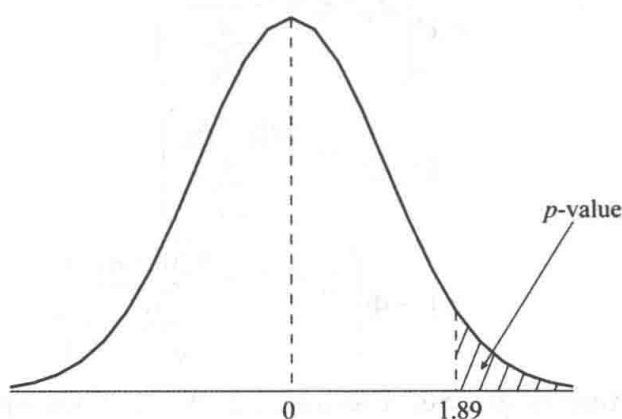


Figure 7.1

$$p = P(Z > 1.89) = 0.0294.$$

$\alpha=0.05, p\text{-value}=0.0294, p\text{-value} < \alpha$, we should reject H_0 . The sample data indicate that the mean score is greater than 120. The conclusion is identical to that obtained by the critical value method.

③ Notice that the test statistic is the same as the pivotal quantity we used to find confidence interval in previous section. In fact, there is a one-to-one close relationship between confidence intervals and hypothesis tests.

For significance level α , the acceptance region for the test is

$$\bar{W} = \{(x_1, \dots, x_n) \mid z < z_\alpha\} = \left\{ (x_1, \dots, x_n) \mid \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < z_\alpha \right\}.$$

It can be rewritten as

$$\bar{W} = \left\{ (x_1, \dots, x_n) \mid \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu_0 \right\}.$$

Let μ_0 take any real value, a $1-\alpha$ confidence interval of μ is obtained:

$$\left(\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, +\infty\right) = \left(128 - 1.645 \times \frac{30}{\sqrt{50}}, +\infty\right) \approx (121.02, +\infty).$$

Since 120 is out of this interval, H_0 should be rejected. The sample data indicate that the mean score is greater than 120.

(2) The power $1-\beta$ of the test (the likelihood of avoiding a type II error) is the probability of correctly rejecting the null hypothesis when it is false. Let's consider

$$H_0: \mu \leq 120, \quad H_1: \mu = \mu_1 > 120.$$

$$1-\beta = P[\text{reject } H_0 | H_1 \text{ is true}]$$

$$= P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_{\alpha} \mid \mu = \mu_1\right)$$

$$= P\left(\frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} \geq z_{\alpha} + \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} \mid \mu = \mu_1\right)$$

$$= P\left(Z \geq z_{\alpha} + \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= 1 - \Phi\left(z_{\alpha} + \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= 1 - \Phi\left(1.645 + \frac{120 - \mu_1}{\frac{30}{\sqrt{50}}}\right).$$

The power is a function of μ_1 . See Figure 7.2. As μ_1 goes up, the power $1-\beta$ of the test rises, see Figure 7.3.

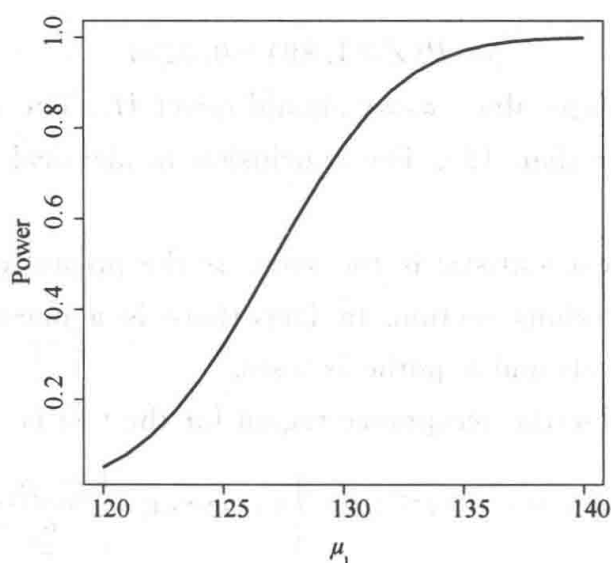


Figure 7.2

In fact, to increase the power of a test, one may do any of the following: ① increase the effect

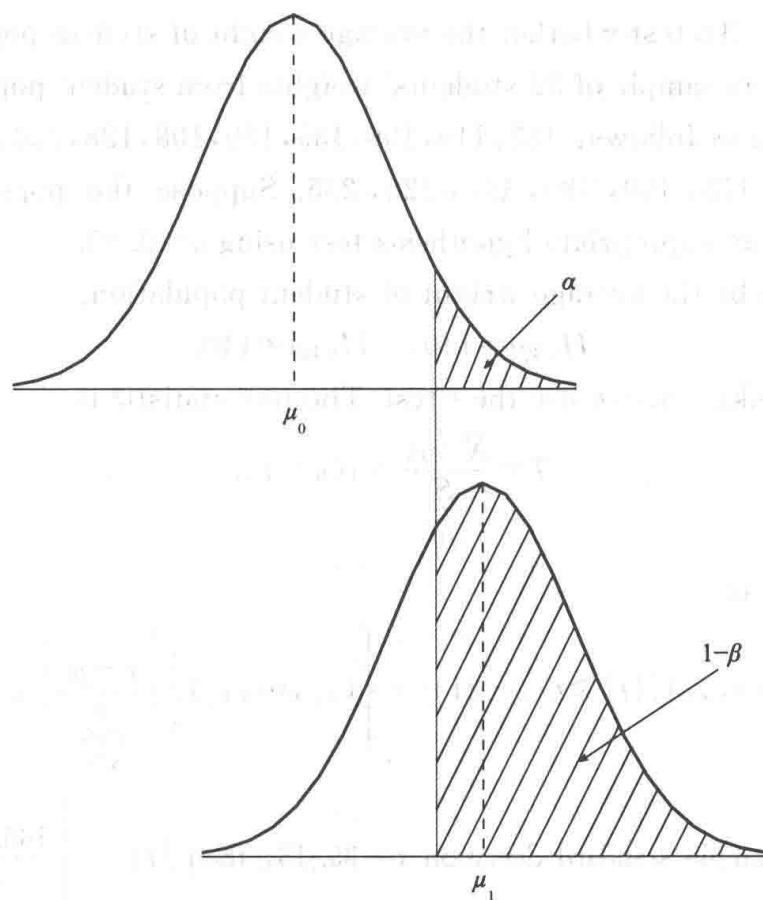


Figure 7.3

size $|\mu_1 - \mu_0|$ (the difference between the null and alternative values); ②increase the sample size n ; ③decrease the variability σ^2 in the sample; ④increase the significance level α of the test.

If the population variance σ^2 is unknown, Z-test can not be applied. Since sample variance S^2 is an unbiased estimator of σ^2 , t -test will be used. The test statistic under the null hypothesis H_0 is

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

The procedures are similar to those in Z-test and summarized in Table 7.3.

Table 7.3

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
t -test (σ^2 is unknown)	$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$	$ t \geq t_{\frac{\alpha}{2}}(n-1)$	$P\left(T \geq \left \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right \right)$
	$\mu \leq \mu_0$	$\mu > \mu_0$		$t \geq t_{\alpha}(n-1)$	$P\left(T \geq \left \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right \right)$
	$\mu \geq \mu_0$	$\mu < \mu_0$		$t \leq t_{1-\alpha}(n-1)$	$P\left(T \leq \left \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right \right)$

Example 7.2.2 To test whether the average weight of student population is different from 140 lb. A random sample of 22 students' weights from student population is selected. The observations are as follows: 135, 119, 106, 135, 180, 108, 128, 160, 143, 175, 170, 205, 195, 185, 182, 150, 175, 190, 180, 195, 220, 235. Suppose the population is normally distributed, conduct an appropriate hypothesis test using $\alpha=0.05$.

Solution: Let μ be the average weight of student population.

$$H_0: \mu=140, \quad H_1: \mu \neq 140.$$

① Since σ^2 is unknown, we use the t -test. The test statistic is

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

The rejection region is

$$W = \{(x_1, \dots, x_n) \mid |t| \geq t_{0.025}(21)\} = \left\{ (x_1, \dots, x_n) \mid \left| \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \right| \geq 2.0796 \right\}.$$

$$n=22, \bar{x}=166.86, \text{ sample standard deviation } s=35.18, \text{ then } |t| = \left| \frac{166.86-140}{\frac{35.18}{\sqrt{22}}} \right| = 3.58 \geq$$

2.0796. The sample data fall in the rejection region and H_0 should be rejected. The sample data indicate that the average weight of student population is statistically significantly different from 140 lb.

② We can also adopt the p -value approach to hypothesis testing. The t -score is

$$t = \frac{166.86-140}{\frac{35.18}{\sqrt{22}}} \approx 3.58.$$

The p -value is the probability that we obtain the observed value of the test statistic that is more extreme in the direction of the alternative hypothesis, calculated when H_0 is true. See Figure 7.4.

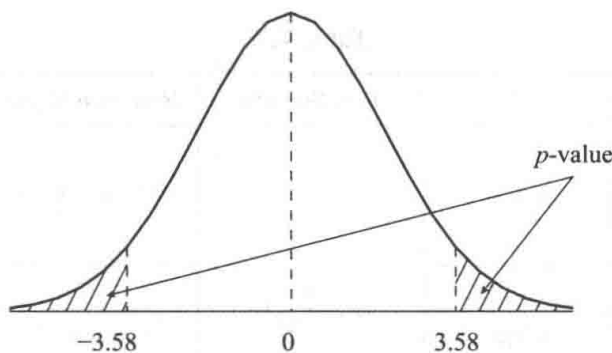


Figure 7.4

$$p = P(|T(21)| \geq 3.58) = 2P(T(21) \geq 3.58) = 0.0018.$$

$\alpha=0.05$, $p\text{-value}=0.0018$, $p\text{-value}<\alpha$, we should reject H_0 . The sample data indicate that the average weight of student population is statistically significantly different from 140 lb. The conclusion is identical to that obtained by the critical value method.

③ For significance level α , the acceptance region for the test is

$$\bar{W} = \{(x_1, \dots, x_n) \mid |t| < t_{\frac{\alpha}{2}}(n-1)\} = \left\{ (x_1, \dots, x_n) \mid \left| \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \right| < t_{\frac{\alpha}{2}}(n-1) \right\}.$$

It can be rewritten as

$$\bar{W} = \left\{ (x_1, \dots, x_n) \mid \bar{x} - t_{\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} < \mu_0 < \bar{x} + t_{\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} \right\}.$$

Let μ_0 take any real value, a $1-\alpha$ confidence interval of μ is obtained:

$$\begin{aligned} & \left(\bar{x} - t_{\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} \right) \\ &= \left(166.86 - 2.0796 \times \frac{35.18}{\sqrt{22}}, 166.86 + 2.0796 \times \frac{35.18}{\sqrt{22}} \right) \\ &\approx (151.26, 182.46). \end{aligned}$$

Since 140 is out of this interval, H_0 should be rejected. The sample data indicate that the average weight of student population is statistically significantly different from 140 lb.

7.3 Testing Differences between Means

Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let $\bar{X}$ and S_1^2 be sample mean and sample variance from X , let $\bar{Y}$ and S_2^2 be sample mean and sample variance from Y . We consider the following hypotheses:

- (1) $H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 \neq 0$.
- (2) $H_0: \mu_1 - \mu_2 \leq 0$, $H_1: \mu_1 - \mu_2 > 0$.
- (3) $H_0: \mu_1 - \mu_2 \geq 0$, $H_1: \mu_1 - \mu_2 < 0$.

These problems will be discussed in two cases.

If σ_1^2 and σ_2^2 are known, $\bar{X} - \bar{Y}$ is an estimator of $\mu_1 - \mu_2$ and

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right).$$

Thus, Z -test can be used. The test statistic under the null hypothesis H_0 is

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1).$$

The procedures are summarized in Table 7.4.

Table 7. 4

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
Z-test (σ_1^2, σ_2^2 are known)	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$	$ z \geq z_{\frac{\alpha}{2}}$	$2 \left[1 - \Phi \left(\left \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right \right) \right]$
	$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$		$z \geq z_{\alpha}$	$1 - \Phi \left(\left \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right \right)$
	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$		$z \leq z_{1-\alpha}$	$\Phi \left(\left \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \right \right)$

If σ_1^2 and σ_2^2 are unknown but equal, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. $\bar{X} - \bar{Y}$ is an estimator of $\mu_1 - \mu_2$ and $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right)$.

Since

$$\frac{(m-1)S_1^2}{\sigma^2} \sim \chi^2(m-1), \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(n-1),$$

thus,

$$\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2).$$

Let

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

be the pooled estimator of σ^2 . Then,

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

Thus, t -test can be used. The test statistic under the null hypothesis H_0 is

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

The procedures are summarized in Table 7. 5.

Table 7. 5

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
t -test ($\sigma_1^2 = \sigma_2^2 = \sigma^2$ are unknown)	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$	$ t \geq t_{\frac{\alpha}{2}}(m+n-2)$	$P \left(T \geq \left \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \right \right)$

continued

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
t -test $(\sigma_1^2 = \sigma_2^2 = \sigma^2$ are unknown)	$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$	$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$	$t \geq t_{\alpha}(m+n-2)$	$P\left(T \geq \left \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \right \right)$
	$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$		$t \leq t_{1-\alpha}(m+n-2)$	$P\left(T \leq \left \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \right \right)$

Example 7.3.1 Students at a college attended a test. The summaries below are of the scores of some students. Do the data provide evidence that the mean score of male students is higher than that of female students at the 0.05 level of significance? Assume the populations to be approximately normal with equal variances.

Group	Count	Mean	Standard Deviation
Male Students	12	86	4
Female Students	10	84	5

Solution: Let μ_1 be the mean score of male students, and μ_2 be the mean score of female students.

$$H_0: \mu_1 - \mu_2 = 0, \quad H_1: \mu_1 - \mu_2 > 0.$$

① Since the population variances σ_1^2 and σ_2^2 are unknown but equal, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we use the t -test. The test statistic under H_0 is

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

The rejection region is

$$W = \left\{ (x_1, \dots, x_n) \mid \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \geq t_{\alpha}(m+n-2) \right\}.$$

$$m=12, n=10, t_{\alpha}(m+n-2) = t_{0.05}(20) = 1.7247, \bar{x} = 86, \bar{y} = 84, S_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}} = \sqrt{\frac{11 \times 16 + 9 \times 25}{20}} = 4.478, \text{ then } t = \frac{86 - 84}{4.478 \times \sqrt{\frac{1}{12} + \frac{1}{10}}} = 1.04 < 1.7247. \text{ The sample data do}$$

not fall in the rejection region and H_0 can not be rejected. The sample data do not rule out the possibility that the mean score of male students is same as that of female students.

② We can also adopt the p -value approach to hypothesis testing. The t -score is

$$t = \frac{86 - 84}{4.478 \times \sqrt{\frac{1}{12} + \frac{1}{10}}} = 1.04.$$

The p -value is the probability that we obtain the observed value of the test statistic that is more extreme in the direction of the alternative hypothesis, calculated when H_0 is true. See Figure 7. 5.

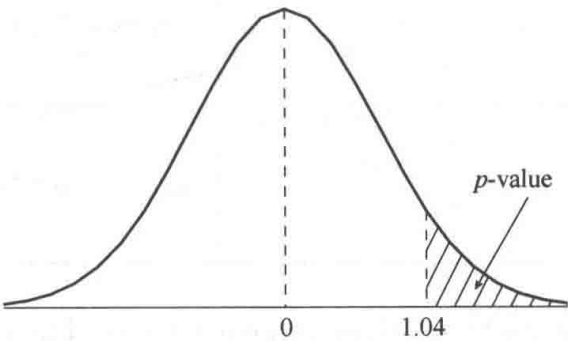


Figure 7. 5

$$p=P(T(20)\geqslant 1.04)=0.1554.$$

$\alpha=0.05$, $p\text{-value}=0.1554$, $p\text{-value}>\alpha$, we do not reject H_0 . The sample data do not rule out the possibility that the mean score of male students is same as that of female students. The conclusion is identical to that obtained by the critical value method.

◆ 7.4 Hypothesis Tests for One or Two Variances ◆

Suppose $X\sim N(\mu,\sigma^2)$, $(X_1,X_2,\cdots,X_n)$ is a sample and $x_1,x_2,\cdots,x_n$ are its observations. For significant level α , we discuss the following three sets of hypotheses.

- (1) $H_0:\sigma^2=\sigma_0^2, \quad H_1:\sigma^2\neq\sigma_0^2$
- (2) $H_0:\sigma^2\leqslant\sigma_0^2, \quad H_1:\sigma^2>\sigma_0^2$
- (3) $H_0:\sigma^2\geqslant\sigma_0^2, \quad H_1:\sigma^2<\sigma_0^2$

where σ_0^2 is a constant.

Suppose the population mean μ is unknown, sample variance S^2 is an unbiased estimator of σ^2 . Here χ^2 -test will be used. The test statistic under the null hypothesis H_0 is

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1).$$

The procedures are summarized in Table 7. 6.

Table 7. 6

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
χ^2 -test	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$	$\chi^2\geqslant\chi_{\frac{\alpha}{2}}^2(n-1)$ or $\chi^2\leqslant\chi_{1-\frac{\alpha}{2}}^2(n-1)$	$2\min\left\{P\left(\chi^2\geqslant\frac{(n-1)s^2}{\sigma_0^2}\right),P\left(\chi^2\leqslant\frac{(n-1)s^2}{\sigma_0^2}\right)\right\}$
	$\sigma^2\leqslant\sigma_0^2$	$\sigma^2>\sigma_0^2$		$\chi^2\geqslant\chi_{\alpha}^2(n-1)$	$P\left(\chi^2\geqslant\frac{(n-1)s^2}{\sigma_0^2}\right)$
	$\sigma^2\geqslant\sigma_0^2$	$\sigma^2<\sigma_0^2$		$\chi^2\leqslant\chi_{1-\alpha}^2(n-1)$	$P\left(\chi^2\leqslant\frac{(n-1)s^2}{\sigma_0^2}\right)$

Example 7.4.1 A manufacturer of helmets for construction workers is concerned about the mean and the variation of the forces that the helmets transmit to wearers when subjected to an external force. The manufacturer has designed the helmets so that the mean force transmitted by the helmets to the workers is 800 pounds (or less) with a standard deviation to be less than 40 pounds. Tests were run on a random sample of $n=40$ helmets, and the sample mean and sample standard deviation were found to be 825 pounds and 48.5 pounds, respectively.

(1) Do the data provide sufficient evidence, at the $\alpha=0.05$ level, to conclude that the population standard deviation exceeds 40 pounds?

(2) Do the data provide sufficient evidence, at the $\alpha=0.05$ level, to conclude that the population standard deviation differs from 40 pounds?

Solution: (1) We want to test the hypotheses:

$$H_0: \sigma^2 \leq 40^2 = 1600, \quad H_1: \sigma^2 > 40^2 = 1600.$$

The test statistic under H_0 is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

The rejection region is

$$W = \left\{ (x_1, \dots, x_n) \mid \frac{(n-1)s^2}{\sigma_0^2} \geq \chi_{\alpha}^2(n-1) \right\}.$$

$n=40, s=48.5, \sigma_0=40, \chi_{\alpha}^2(n-1) = \chi_{0.05}^2(39) = 54.572$, then the value of the test statistic is $\chi^2 = \frac{(40-1) \times 48.5^2}{40^2} = 57.336 > 54.572$. The test statistic falls in the rejection region.

That is, we reject the null hypothesis in favor of the alternative hypothesis. The sample data indicate that the population standard deviation exceeds 40 pounds.

In this case, the p -value is the probability that we would observe a $\chi^2(39)$ random variable more extreme than 57.336, see Figure 7.6.

$$p = P(\chi^2(39) \geq 57.336) = 0.029 < 0.05 = \alpha.$$

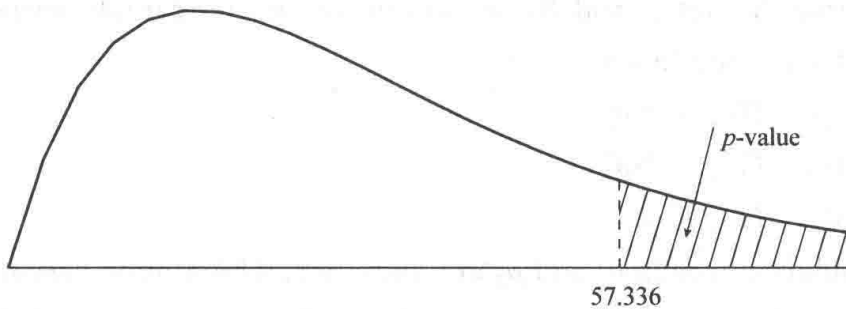


Figure 7.6

We reject the null hypothesis in favor of the alternative hypothesis. The p -value approach yields the same conclusion.

(2) We want to test the hypotheses:

$$H_0: \sigma^2 = 40^2 = 1600, \quad H_1: \sigma^2 \neq 40^2 = 1600.$$

the test statistic under H_0 is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

$n=40, s=48.5, \sigma_0=40, \chi_{\frac{\alpha}{2}}^2(n-1) = \chi_{0.025}^2(39) = 58.119, \chi_{1-\frac{\alpha}{2}}^2(n-1) = \chi_{0.975}^2(39) = 23.654$, the rejection region is

$$\begin{aligned} W &= \left\{ (x_1, \dots, x_n) \mid \frac{(n-1)s^2}{\sigma_0^2} \geq \chi_{\frac{\alpha}{2}}^2(n-1) \text{ or } \frac{(n-1)s^2}{\sigma_0^2} \leq \chi_{1-\frac{\alpha}{2}}^2(n-1) \right\} \\ &= \left\{ (x_1, \dots, x_n) \mid \frac{(n-1)s^2}{\sigma_0^2} \geq 58.119 \text{ or } \frac{(n-1)s^2}{\sigma_0^2} \leq 23.654 \right\}. \end{aligned}$$

The value of the test statistic is $\chi^2 = \frac{(40-1) \times 48.5^2}{40^2} = 57.336$. The test statistic does not fall in the rejection region. That is, we fail to reject the null hypothesis. There is insufficient evidence, at the 0.05 level, to conclude that the population standard deviation differs from 40.

In this case, the p -value is given as follows

$$p = 2 \min\{P(\chi^2(39) \geq 57.336), P(\chi^2(39) < 57.336)\} = 0.058 > 0.05 = \alpha.$$

We fail to reject the null hypothesis in favor of the two-sided alternative hypothesis. The p -value approach yields the same conclusion.

Note: this example illustrates a fact that the conclusion (1) for the one-sided test does not always agree with the conclusion (2) for the two-sided test. If you have reasons to believe that the parameter will differ from the null value in a particular direction, then you should conduct the one-sided test.

When testing two independent samples from two populations, it is helpful to know that the two populations have equal variances or not. In order to test these hypotheses, a test of equality of variances must be conducted.

Suppose $(X_1, \dots, X_m)$ is a random sample from $X \sim N(\mu_1, \sigma_1^2)$, $(Y_1, \dots, Y_n)$ is a random sample from $Y \sim N(\mu_2, \sigma_2^2)$, X and Y are independent. Let $\bar{X}$ and S_1^2 be sample mean and sample variance from X , let $\bar{Y}$ and S_2^2 be sample mean and sample variance from Y . We consider the following hypotheses:

- (1) $H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$.
- (2) $H_0: \sigma_1^2 \leq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2$.
- (3) $H_0: \sigma_1^2 \geq \sigma_2^2, H_1: \sigma_1^2 < \sigma_2^2$.

Here the population means μ_1 and μ_2 are unknown. The sample variances S_1^2 and S_2^2 are unbiased estimators of σ_1^2 and σ_2^2 , respectively. Then F -test will be used. The test statistic under the null hypothesis H_0 is

$$F = \frac{S_1^2}{S_2^2} \sim F(m-1, n-1).$$

The procedures are summarized in Table 7.7.

Table 7.7

Test	H_0	H_1	Test Statistic	Rejection Region	p -value
F-test	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2} \sim F(m-1, n-1)$	$F \geq F_{\frac{\alpha}{2}}(m-1, n-1)$ or $F \leq F_{1-\frac{\alpha}{2}}(m-1, n-1)$	$2\min\left\{P\left(F \geq \frac{s_1^2}{s_2^2}\right), P\left(F \leq \frac{s_1^2}{s_2^2}\right)\right\}$
	$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$		$F \geq F_{\alpha}(m-1, n-1)$	$P\left(F \geq \frac{s_1^2}{s_2^2}\right)$
	$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$		$F \leq F_{1-\alpha}(m-1, n-1)$	$P\left(F \leq \frac{s_1^2}{s_2^2}\right)$

Example 7. 4. 2 Lammers Limos offers limousine service from the city hall in Toledo, Ohio, to Metro Airport in Detroit. Sean Lammers, president of the company, is considering two routes. One is via U. S. 25 and the other via I-75. He wants to study the time it takes to drive to the airport using each route and then compare the results. He collected the following sample data, which is reported in minutes.

U. S. 25	52	67	56	45	70	54	64	
I-75	59	60	61	51	56	63	57	65

- (1) Can we conclude that the variation in the driving times for U. S. 25 is larger than the variation in the driving times for I-75 at the 0.05 significance level?
- (2) Can we conclude that the variation in the driving times for U. S. 25 is different from the variation in the driving times for I-75 at the 0.1 significance level?

Solution: (1) Let σ_1^2 and σ_2^2 be the variances of driving times for U. S. 25 and I-75 respectively, we want to test the hypotheses:

$H_0 : \sigma_1^2 \leq \sigma_2^2, \quad H_1 : \sigma_1^2 > \sigma_2^2.$

The test statistic under H_0 is

$F = \frac{S_1^2}{S_2^2} \sim F(m-1, n-1).$

$m=7, n=8, \alpha=0.05, F_{\alpha}(m-1, n-1)=F_{0.05}(6, 7)=3.87$, the rejection region is

$W = \{(x_1, \dots, x_n) \mid F \geq F_{\alpha}(m-1, n-1)\} = \{(x_1, \dots, x_n) \mid F \geq 3.87\}.$

$s_1^2=80.905$ and $s_2^2=19.143$ are the sample variances of driving times for U. S. 25 and I-75 respectively. The value of the test statistic is $F = \frac{s_1^2}{s_2^2} = 4.23$. It falls in the rejection region. So we can conclude that the variation in the driving times for U. S. 25 is larger than the variation in the driving times for I-75 at the 0.05 significance level.

In this case, the p -value is the probability that we would observe an $F(6, 7)$ random variable more extreme than 4.23, see Figure 7.7.

$p = P[F(6, 7) \geq 4.23] = 0.04 < 0.05 = \alpha.$

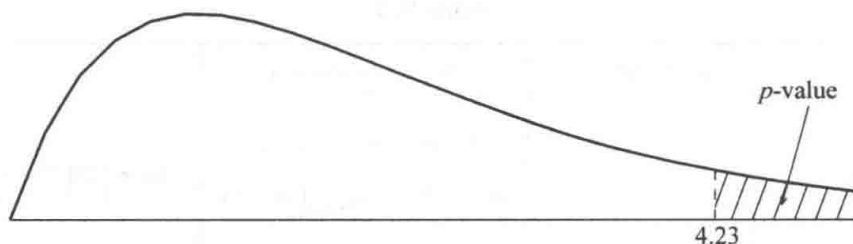


Figure 7.7

We reject the null hypothesis in favor of the alternative hypothesis. The p -value approach yields the same conclusion.

(2) We want to test the hypotheses:

$$H_0: \sigma_1^2 = \sigma_2^2, \quad H_1: \sigma_1^2 \neq \sigma_2^2.$$

The test statistic under H_0 is

$$F = \frac{S_1^2}{S_2^2} \sim F(m-1, n-1).$$

$m=7, n=8, \alpha=0.1, F_{\frac{\alpha}{2}}(m-1, n-1) = F_{0.05}(6, 7) = 3.87, F_{1-\frac{\alpha}{2}}(m-1, n-1) = F_{0.95}(6, 7) = \frac{1}{F_{0.05}(7, 6)} = 0.24$, the rejection region is

$$W = \{(x_1, \dots, x_n) \mid F \geq 3.87 \text{ or } F \leq 0.24\}.$$

$s_1^2 = 80.905$ and $s_2^2 = 19.143$ are the sample variances of driving times for U. S. 25 and I-75 respectively. The value of the test statistic is $F = \frac{s_1^2}{s_2^2} = 4.23$. It falls in the rejection region.

So we can conclude that the variation in the driving times for U. S. 25 is different from the variation in the driving times for I-75 at the 0.1 significance level.

In this case, the p -value is given as follows

$$p = 2\min\{P(F(6, 7) \geq 4.23), P(F(6, 7) \leq 4.23)\} = 0.08 < 0.1 = \alpha.$$

We reject the null hypothesis in favor of the alternative hypothesis. The p -value approach yields the same conclusion.

◆ 7.5 Goodness of Fit Tests ◆

The hypothesis testing problems we have discussed so far are parametric hypothesis tests. We assume that the population form is known and test the claims about the population mean and variance etc. In practice, we attempt to fit a statistical distribution to observed data. We want to know how well the distribution actually reflects the data and test the following hypotheses:

$$H_0: F(x) = F_0(x), \quad H_1: F(x) \neq F_0(x),$$

where $F_0(x)$ is a given distribution. These testing problems are called chi-square goodness of fit tests.

Theorem 7.5.1 (Pearson-Fisher) Suppose a population $X \sim F_0(x; \theta)$, where θ is a unknown parameter vector with r elements. $\hat{\theta}$ is the maximum likelihood estimator of θ . Divide the range $\Omega = (a_0, a_k)$ of X into k non-overlapping bins:

$$A_1 = (a_0, a_1], A_2 = (a_1, a_2], \dots, A_k = (a_{k-1}, a_k).$$

Suppose $(x_1, x_2, \dots, x_n)$ is a sample of size n . Let n_i ($i = 1, \dots, k$) denote the observed frequency in A_i . $\sum_{i=1}^k n_i = n$.

Let p_i ($i = 1, \dots, k$) denote the probabilities hypothesized for k bins. $p_i = F_0(a_i; \theta) - F_0(a_{i-1}; \hat{\theta})$. np_i is theoretical or expected frequency in A_i .

Then, as n goes to infinity,

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \rightarrow \chi^2(k - r - 1).$$

Proof: We provide the proof just for $k = 2$. In this case, $n_1 \sim B(n, p_1)$, $p_1 + p_2 = 1$, $n_1 + n_2 = n$, $n_1 - np_1 = n - n_2 - n(1 - p_2) = np_2 - n_2$. Thus,

$$\chi^2 = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} = \frac{(n_1 - np_1)^2}{np_1 p_2}.$$

By the central limit theorem,

$$\frac{n_1 - np_1}{\sqrt{np_1 p_2}} \rightarrow N(0, 1).$$

Therefore, $\chi^2 \rightarrow \chi^2(1)$. Here we omit the proof for other cases.

Note; if the computed test statistic is large, then the observed and expected values are not close and the distribution is a poor fit to the data. It is always a right tail test.

This test is not valid for small samples. The expected frequency of each bin (category) must be at least five and if some of the counts are less than five, one may need to combine some adjacent bins.

Example 7.5.1 In one of his genetic experiments, Gregor Mendel crossed 556 smooth-yellow male peas with wrinkled-green female peas. For each of the 556 crossings, there are four possible outcomes: smooth-yellow, smooth-green, wrinkled-yellow, and wrinkled-green. Let p_1 be the probability of smooth-yellow, p_2 be the probability of smooth-green, p_3 be the probability of wrinkled-yellow, p_4 be the probability of wrinkled-green. Mendel's theory predicted that

$$p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}.$$

Table 7.8 contains Mendel's data. Do these data support Mendel's theory?

Table 7.8

Type	smooth-yellow	smooth-green	wrinkled-yellow	wrinkled-green
Observed Frequency	315	101	108	32
Expected Values	312.75	104.25	104.25	34.75

Solution: we want to test

$$H_0: p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}.$$

H_1 : at least one p_i is not as specified in H_0 .

The test statistic value is

$$\chi^2 = \frac{\left(315 - 556 \times \frac{9}{16}\right)^2}{556 \times \frac{9}{16}} + \frac{\left(315 - 556 \times \frac{3}{16}\right)^2}{556 \times \frac{3}{16}} + \frac{\left(315 - 556 \times \frac{3}{16}\right)^2}{556 \times \frac{3}{16}} + \frac{\left(315 - 556 \times \frac{1}{16}\right)^2}{556 \times \frac{1}{16}} = 0.47.$$

For significance level $\alpha = 0.05$, $\chi_{0.05}^2(3) = 7.815$, $0.47 < 7.815$, we do not reject the null hypothesis H_0 . There is insufficient evidence, at the 0.05 level, to conclude that the probabilities differ from a 9 : 3 : 3 : 1 ratio.

In this case, the p -value is the probability that we would observe a $\chi^2(3)$ random variable more extreme than 0.47, see Figure 7.8.

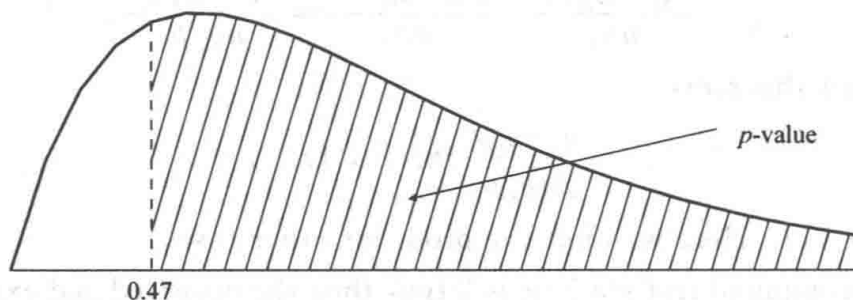


Figure 7.8

$$p = P(\chi^2(3) \geq 0.47) = 0.9254 > 0.05 = \alpha.$$

We fail to reject the null hypothesis in favor of the alternative hypothesis. The p -value approach yields the same conclusion.

Example 7.5.2 Table 7.9 presents count data on the number of storms observed in 63 years in a certain city. Is Poisson distribution a good fit to the data?

Table 7.9

Number of storms per year	0	1	2	3	4	5	6	7	8
Frequency	5	8	14	19	10	4	2	1	1

Solution: Let X be the number of storms per year. We want to test whether a Poisson distribution is a good fit to the data, using Pearson's chi-squared test and the estimated value of the Poisson parameter.

$$H_0: X \sim P(\lambda), \quad H_1: X \text{ does not follow } P(\lambda).$$

Based on the observed data, the maximum likelihood estimator for the Poisson parameter is the mean of the sample data,

$$\hat{\lambda} = \frac{0 \times 5 + 1 \times 8 + \cdots + 8 \times 1}{63} = 2.8571.$$

The calculations are listed in Table 7.10. We combined the adjacent bins whose frequencies are less than five.

Table 7.10

Number	Frequency n_i	Probability p_i	Theoretical Frequency	χ^2 value
0	5	0.0574	3.6184	0.5275
1	8	0.1641	10.3382	0.5288
2	14	0.2344	14.7686	0.0400
3	19	0.2233	14.0651	1.7315
4	10	0.1595	10.0464	0.0002
≥ 5	8	0.1613	10.1634	0.4605

Since the mean λ of the Poisson distribution has been estimated from the data, the number of degrees of freedom is $k - r - 1 = 6 - 1 - 1 = 4$. Using the 0.05 level of significance, the critical value of χ^2 with 4 degrees of freedom is 9.4877. The decision rule is “reject H_0 if $\chi^2 > 9.4877$; otherwise do not reject H_0 ”.

From Table 7.10, $\chi^2 = 0.5275 + 0.5288 + 0.0400 + 1.7315 + 0.0002 + 0.4605 = 3.2885 < 9.4877$. The decision is not to reject H_0 . There is insufficient evidence to conclude that the number of storms per year does not fit a Poisson distribution.

In this case, the p -value is the probability that we would observe a $\chi^2(4)$ random variable more extreme than 3.2885.

$$p = P(\chi^2(4) \geq 3.2885) = 0.5108 > 0.05 = \alpha.$$

We fail to reject the null hypothesis in favor of the alternative hypothesis. The p -value approach yields the same conclusion.

The chi-square goodness of fit test may also be applied to continuous distributions such as normal distribution. In this case, the observed data are grouped into discrete bins so that the chi-square statistic may be calculated. The expected values under the assumed distribution are the probabilities associated with each bin multiplied by the number of observations.

The chi-square test is easily extended to test association of two categorical variables in two-way contingency tables, where the assumed model of independence is evaluated against the observed data.

A two-way table for X and Y with r and c levels, respectively, will have r rows $A_1, A_2, \dots, A_r$ and c columns $B_1, B_2, \dots, B_c$. The table will have $r \times c$ cells, with any one cell entirely determined by the sum of the others (since the total of the observed counts must sum to n observations).

H_0 : X, Y are independent, H_1 : There exists association between the variables.

Let n_{ij} ($i = 1, 2, \dots, r; j = 1, 2, \dots, c$) denote the frequency of event $A_i \cap B_j$. The probability that a randomly selected observation falls into the cell defined by $A_i \cap B_j$ is p_{ij}

and is estimated by $\frac{n_{ij}}{n}$. The probability that a randomly selected observation falls into a row defined by A_i is $p_{i\cdot}$. The probability that a randomly selected observation falls into a column defined B_j is $p_{\cdot j}$.

We'll estimate the p_{ij} assuming that the null hypothesis is true, that is, assuming independence: $p_{ij} = p_{i\cdot} \times p_{\cdot j}$.

Under the assumption of independence H_0 , it is therefore reasonable to estimate the p_{ij} with:

$$\hat{p}_{ij} = \hat{p}_{i\cdot} \times \hat{p}_{\cdot j} = \frac{\sum_{j=1}^c n_{ij}}{n} \times \frac{\sum_{i=1}^r n_{ij}}{n} = \frac{n_{i\cdot} n_{\cdot j}}{n^2}$$

Notice that the fact that the row probabilities add to 1 and the column probabilities add to 1, we need estimate $(r-1) + (c-1)$ parameters to give the marginal probabilities that determine the expected counts, so the degrees of freedom for the chi-square test statistic are

$$rc - [(r-1) + (c-1)] - 1 = (r-1) \times (c-1).$$

The test statistic is

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - n \hat{p}_{ij})^2}{n \hat{p}_{ij}} = \sum_{i=1}^r \sum_{j=1}^c \frac{\left(n_{ij} - \frac{n_{i\cdot} n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot} n_{\cdot j}}{n}}.$$

Example 7.5.3 Is age independent of the desire to ride a bicycle? A random sample of 395 people were surveyed. Each person was asked his interest in riding a bicycle (variable X) and his age (variable Y). The data that resulted from the survey is summarized in Table 7.11. Is there evidence to conclude, at the 0.05 level, that the desire to ride a bicycle depends on age?

Table 7.11

Variable X	Variable Y				Total
	18~24	25~34	35~49	50~64	
Yes	60	54	46	41	201
No	40	44	53	57	194
Total	100	98	99	98	395

Solution: We are interested in conducting a test of independence for testing the null hypothesis:

H_0 : variable X is independent of variable Y , that is, $p_{ij} = p_{i\cdot} \times p_{\cdot j}$.

H_1 : variable X is not independent of variable Y .

Using the 0.05 level of significance, the critical value of χ^2 with $(2-1) \times (4-1) = 3$ degrees of freedom is 7.815. The decision rule is "reject H_0 if $\chi^2 > 7.815$; otherwise do not reject H_0 ".

The expected counts $\frac{n_{i \cdot} \cdot n_{\cdot j}}{n}$ are reported in Table 7.12. The test statistic value is

$$\chi^2 = \frac{(60 - 50.886)^2}{50.886} + \dots + \frac{(57 - 48.132)^2}{48.132} = 8.006 > 7.815.$$

Table 7.12

Variable X	Variable Y				Total
	18~24	25~34	35~49	50~64	
Yes	50.886	49.868	50.377	49.868	201
No	49.114	48.132	48.623	48.132	194
Total	100	98	99	98	395

We reject the null hypothesis H_0 . There is sufficient evidence at the 0.05 level to conclude that the desire to ride a bicycle depends on age.

In this case, the p -value is the probability that we would observe a $\chi^2(3)$ random variable more extreme than 8.006.

$$p = P(\chi^2(3) \geq 8.006) = 0.04589 < 0.05 = \alpha.$$

We reject the null hypothesis H_0 . The p -value approach yields the same conclusion.

Exercise 7

7.1 Suppose $x_1, x_2, \dots, x_n$ form a sample from $N(\mu, 1)$. Consider the following hypothesis testing problem:

$$H_0: \mu = 2, \quad H_1: \mu = 3.$$

The rejection region $W = \{\bar{x} \geq 2.6\}$.

- (1) If $n = 20$, find the probabilities of two types of errors;
- (2) What is the smallest sample size of n such that the probability of type II error satisfies $\beta \leq 0.01$?
- (3) Show that if $n \rightarrow +\infty$, then $\alpha \rightarrow 0, \beta \rightarrow 0$.

7.2 To test whether a coin is fair or not. Flip a coin 10 times. Let X be the number of times that the coin comes up heads. If $|X - 5| > 3$, then reject $H_0: p = 0.5$. What is the significance level of the test?

7.3 Suppose $x_1, x_2, \dots, x_n$ form a sample from uniform distribution $U(0, \theta)$. Consider the following hypothesis testing problem:

$$H_0: \theta \geq 3, \quad H_1: \theta < 3.$$

The rejection region $W = \{x_{(n)} \leq 2.5\}$.

- (1) Find the largest value of α , the probability of type I error.
- (2) What is the smallest value of sample size n such that the value in (1) is no more than 0.05?

7.4 Suppose $x_1, x_2, \dots, x_n$ form a sample from population $X \sim N(\mu, \sigma^2)$, where σ^2 is known, μ takes one of the two values, μ_0 or μ_1 ($\mu_1 > \mu_0$). $\bar{x}$ is the sample mean. Under the significance level α , let's consider

$$H_0: \mu = \mu_0, \quad H_1: \mu = \mu_1.$$

The probability of type II error is $\beta = P\{\bar{X} - \mu_0 < k | \mu = \mu_1\}$. Show that

$$\beta = \Phi\left(z_\alpha \frac{\mu_1 - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) \quad \text{and} \quad n = (z_\alpha + z_\beta)^2 \frac{\sigma^2}{(\mu_1 - \mu_0)^2}.$$

7.5 A simple random sample of 10 people from a certain population has a mean age of 27. The variance is known to be 20. Let $\alpha = 0.05$.

(1) Can we conclude that the mean age of the population is not 30?

(2) Can we conclude that the mean age of the population is less than 30?

7.6 A rental car company claims the mean time to rent a car on their website is 60 seconds with a standard deviation of 30 seconds. A random sample of 36 customers attempted to rent a car on the website. The mean time to rent was 75 seconds. Is this enough evidence to contradict the company's claim?

7.7 A certain type of electronic product is considered to be acceptable if its lifetime is at least 1000 hours. Draw a sample of these products and the sample mean is 950 hours. Assume the lifetime follows normal distribution $N(\mu, \sigma^2)$ and $\sigma = 100$. Are the products acceptable? Use $\alpha = 0.05$.

7.8 Signals are sent from location A to location B. The signal received at location B follows $N(\mu, 0.2^2)$, where μ is the true signal value sent at location A. Suppose the signal is sent five times, and the received values are 8.05, 8.15, 8.2, 8.1, 8.25, can we conclude that the true signal value sent is 8? Use $\alpha = 0.05$.

7.9 When the cutter works properly, the length of each piece of metal-rod produced is a random variable that obeys the normal distribution $X \sim N(\mu, \sigma^2)$, $\mu = 10.5$ cm, $\sigma = 0.15$ cm. From a batch of products, randomly select 15 metal rods, their lengths are as follows:

10.4, 10.6, 10.1, 10.4, 10.5, 10.3, 10.3, 10.2, 10.9, 10.6, 10.8, 10.5, 10.7, 10.2, 10.7.

Does the cutter work properly? Use $\alpha = 0.05$.

7.10 It is regulated that the content of vitamin C in a certain kind of food per 100 g is not less than 21 mg. Assume the determination of vitamin C content obeys normal distribution $N(\mu, \sigma^2)$. Randomly select a sample of size 17 from this batch of food, vitamin C content (unit: mg) are measured and the data are

16, 22, 21, 20, 23, 21, 19, 15, 13, 23, 17, 20, 29, 18, 22, 16, 25

Try to test whether the content of vitamin C in this kind of food is qualified using $\alpha = 0.05$.

7.11 The strength of a steel of a certain type is known to obey normal distribution.

Based on past experience, the average tensile strength is 52.00 (unit: kg/mm²). The formulation of the steelmaking is improved, the new method is used to refine the steel, select 7 steel bars produced and the strengths are measured as follows:

52.45, 48.51, 56.02, 51.53, 49.02, 53.38, 54.04.

Is the mean strength of steel bar produced by the new method improved significantly? Use $\alpha=0.05$.

7.12 A simple random sample of 14 people from a certain population gives body mass indices(BMI). The data are 23, 25, 21, 37, 39, 21, 23, 24, 32, 57, 23, 26, 31, 45. Can we conclude that the BMI is not 35?

7.13 Researchers studying the effects of diet on growth would like to know if a vegetarian diet affects the height of a child. The researchers randomly selected 12 vegetarian children that are 6-year-old. The average height of the children is 42.5 inches with a standard deviation of 3.8 inches. The average height for all 6-year-old children is 45.75 inches. Conduct a hypothesis test to determine whether there is overwhelming evidence at $\alpha=0.05$ that 6-year-old vegetarian children are not the same height as other 6-year-old children. Assume that the heights of 6-year-old vegetarian children are approximately normally distributed.

7.14 A travel writer suspects that hotels have higher room rates in Florida than in Georgia. He carried out a survey of room rates and the results are given below. Suppose the population standard deviations of room rates for Florida and Georgia are $\sigma_1=18$ and $\sigma_2=24$ respectively.

Group	Count	Mean
Florida	14	\$ 112
Georgia	26	\$ 109

Is there evidence that the mean room rate in Florida is higher than the mean room rate in Georgia? Conduct an appropriate hypothesis test using $\alpha=0.05$.

7.15 The amount of a certain trace element in blood is known to vary with a standard deviation of 14.1 ppm (parts per million) for male blood donors and 9.5 ppm for female donors. Random samples of 75 male and 50 female donors yield concentration means of 28 and 33 ppm, respectively. Are the population means of concentrations of the element the same for men and women?

7.16 A grocery has two stores. One store is located on First Street and the other on Main Street and each is run by a different manager. Each manager claims that her store's layout maximizes the amounts customers will purchase on impulse. Both managers surveyed a sample of their customers and asked them how much more they spent than they had planned to, in other words, how much did they spend on impulse? The following table shows the sample data collected from the two stores.

First street: 15.78, 17.73, 10.61, 15.79, 14.22, 13.82, 13.45, 12.86, 10.82, 12.85;

Main street: 15.19, 18.22, 15.38, 15.96, 21.92, 12.87, 12.47, 13.96, 13.79, 13.74, 18.4, 18.57, 17.79, 10.83.

Is there a difference in the mean amounts purchased on impulse at the two stores?

7.17 The following gives observations of the running time (unit: hours) after charging of two types of calculators:

Model A: 5.5, 5.6, 6.3, 4.6, 5.3, 5.0, 6.2, 5.8, 5.1, 5.2, 5.9;

Model B: 3.8, 4.3, 4.2, 4.0, 4.9, 4.5, 5.2, 4.8, 4.5, 3.9, 3.7, 4.6.

Assume the populations to be approximately normal with equal variances. Can you think that the average running time of the calculator model A is longer than model B ($\alpha=0.01$)?

7.18 Select two samples independently from two coal mines. The ash rates are measured (unit: %):

A: 24.3, 20.3, 23.7, 21.3, 17.4;

B: 18.2, 16.9, 20.2, 16.7.

Assume the ash contents obey normal distributions with equal variances. Are there any significant differences in the average ash content of the two coal mines? ($\alpha=0.05$)

7.19 An investigator predicts that dog owners in the country spend more time walking their dogs than do dog owners in the city. The investigator gets a sample of 21 country owners and 23 city owners. The mean number of hours per week that city owners spend walking their dogs is 10.0. The standard deviation of hours spent walking the dog by city owners is 3.0. The mean number of hours per week country owners spent walking their dogs was 15.0. The standard deviation of hours spent walking the dog by owners in the country was 4.0. Do dog owners in the country spend more time walking their dogs than do dog owners in the city? Assume the populations are normal with equal variances.

7.20 From the two branches of a zinc mine, select two samples of size 9 and 8 respectively, the average zinc content and the sample variances are as follows:

East branch: $\bar{x}_1=0.230, s_1^2=0.1337$;

West branch: $\bar{x}_2=0.269, s_2^2=0.1736$.

Assume the contents of zinc of two branches are normal with equal variances. Are the average value of zinc content of two veins the same? ($\alpha=0.05$)

7.21 A factory produces copper wire. The variance of the tensile force must not exceed 16 (kg^2). From the production of copper wire, randomly select 9 copper wires, the tensile forces (unit: kg) are

289, 286, 285, 284, 286, 285, 286, 298, 292.

Assume that the tensile force obeys the normal distribution of $X \sim N(\mu, \sigma^2)$. Does the tensile force of the copper wire produced conform to the standard? ($\alpha=0.05$)

7.22 Select 10 fuses from a batch of fuses. The times needed to melt after the strong

current (unit: second) are

42, 65, 75, 78, 59, 71, 57, 68, 54, 55.

Assume the time needed to melt obeys the normal distribution.

(1) Is the average melting time not less than 65? ($\alpha=0.05$)

(2) Does the variance of the melting time not exceed 80? ($\alpha=0.05$)

7.23 Two lathes produce the same kind of ball. The diameters of the balls obey the normal distribution. Select 8 and 9 products separately from the two lathes. The diameters are measured.

Lathe A: 15.0, 14.5, 15.2, 15.5, 14.8, 15.1, 15.2, 14.8;

Lathe B: 15.2, 15.0, 14.8, 15.2, 15.0, 15.0, 14.8, 15.1, 14.8.

Do the population variances of the ball diameters produced by two lathes have a significant difference? ($\alpha=0.05$)

7.24 A study was performed on patients with pituitary adenomas. The standard deviation of the weights of 12 patients with pituitary adenomas was 21.4 kg. A control group of 5 patients without pituitary adenomas had a standard deviation of the weights of 12.4 kg. We wish to know if the weights of the patients with pituitary adenomas are more variable than the weights of the control group. Conduct a statistical test. ($\alpha=0.05$)

7.25 Select two samples independently from two batches of electronic devices. The resistance values (unit: Ω) are

Batch A: 0.140, 0.138, 0.143, 0.142, 0.144, 0.137;

Batch B: 0.135, 0.140, 0.142, 0.136, 0.138, 0.140.

Assume the resistance values of two batches obey $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively.

(1) Test whether the variances of two populations are equal ($\alpha=0.05$);

(2) Test whether the mean values of two populations are equal ($\alpha=0.05$).

7.26 Suppose that we roll a die 30 times and observe the following table showing the number of times each face ends up on top. Is the die fair?

Face	1	2	3	4	5	6
Count	3	7	5	10	2	3

7.27 Tall cut-leaf tomatoes were crossed with dwarf potato-leaf tomatoes, and $n=1611$ offspring were classified by their phenotypes.

Phenotypes	tall cut-leaf	tall potato-leaf	dwarf cut-leaf	dwarf potato-leaf
Count	926	288	293	104

Genetic theory says that the four phenotypes should occur with relative frequencies 9 : 3 : 3 : 1, and thus are not all equally as likely to be observed. Do the observed data support this theory?

7.28 There are 100 batches of products, each containing 10 products. Some are

defective. Does the number of defective products in a batch follow a binomial distribution? ($\alpha=0.05$)

The number of defective products	0	1	2	3	4	5	6	7	8	9	10
Count	35	40	18	5	1	1	0	0	0	0	0

7.29 A study investigates the association between smoking and asthma among adults observed in a community health clinic. The results obtained from classifying 150 individuals are shown in the table below:

		Ever smoked cigarettes		Total
		Yes	No	
Symptoms of asthma	Yes	20	30	50
	No	22	78	100
Total		42	108	150

Does cigarette smoking affect asthma symptoms? ($\alpha=0.05$)

7.30 Should the proportion of elementary schools students' computer courses be improved? To study this problem, a survey has been conducted and the observations are reported in the following table.

Age	Agree	Disagree	NA
>55	32	28	14
36~55	44	21	17
15~35	47	12	13

Does the factor of age affect the answer to this question? ($\alpha=0.05$)

Chapter 8

Linear Regression

In practice, we often have the following case: there exists some association between variable X and variable Y , but this association differs from a function, the value of Y can not be exactly determined by the value of X . For example, there is a relationship between systolic blood pressure and age. In general, larger values of age are associated with higher systolic blood pressure values. But for the people at the same age, the systolic blood pressure values are different. The systolic blood pressure can not be exactly determined by age.

Regression analysis is a useful tool to investigate how two or more variables are related. Quite often we want to see how a certain variable of interest is affected by one or more variables. In this chapter, we are mainly investigating the linear relationship between two quantitative variables, that is simple linear regression.

◆ 8.1 Linear Regression Model ◆

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. Usually we identify one variable as the outcome of interest. This is often called the response, or dependent variable Y . The other variable is called the explanatory, or independent variable X . For example, one might want to relate the weights of individuals to their heights using a linear regression model, the weight Y is the dependent variable and the height X is the independent variable.

A linear regression model has the following form

$$y = a + bx + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

where ϵ is the random error, x is the independent variable and y is the dependent variable. The slope of the line is b , and a is the intercept.

Definition 8.1.1 Suppose we have data $(x_1, y_1), \dots, (x_n, y_n)$, where y_i is the observation of variable Y when $X = x_i (i = 1, 2, \dots, n)$. The linear regression model can be expressed as

$$\begin{cases} y_i = a + bx_i + \epsilon_i \\ \epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \end{cases}.$$

Based on the data $(x_1, y_1), \dots, (x_n, y_n)$, the estimates $\hat{a}$ and $\hat{b}$ of a and b can be obtained. Then

$$\hat{y} = \hat{a} + \hat{b}x$$

is called linear regression line. $\hat{a}$ is the estimated average value of y when the value of x is zero. $\hat{b}$ is the estimated change in the average value of y as a result of a one-unit change in x . When $x = x_0$, $\hat{y}_0 = \hat{a} + \hat{b}x_0$ is called regression value (or fitted value, predicted value).

Fitting a linear model to observed data does not necessarily imply that one variable causes the other (for example, higher heights do not cause higher weights), but that there is some significant association between the two variables.

A scatterplot is a useful tool in determining the strength of the relationship between two variables. A scatterplot consists of the horizontal axis, the vertical axis and a series of dots. Each dot on the scatterplot represents one observation from a data set. The position of the dot on the scatterplot represents its values. If there appears to be no association between the variables (the scatterplot does not show any increasing or decreasing trends), then fitting a linear regression model to the data probably is inappropriate.

8.2 Least Squares Estimation

The line of best fit is the line that gets closest to all the data points. See Figure 8.1. The most common method for fitting a regression line is the method of least-squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line.

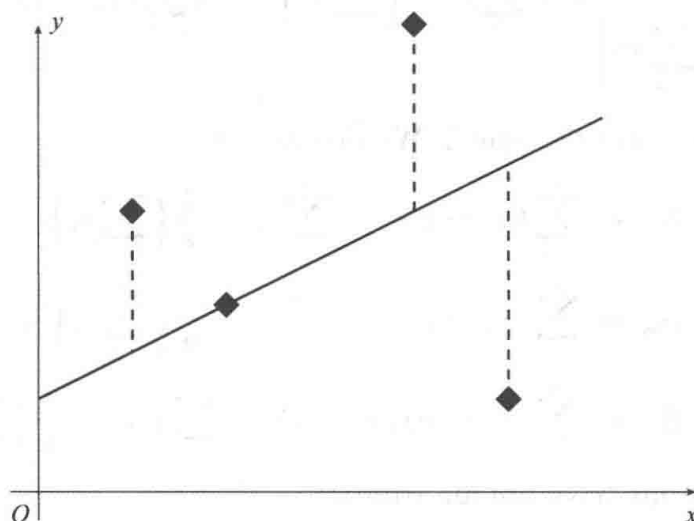


Figure 8.1

Let

$$Q(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Since the deviations $y_i - a - bx_i$ are first squared, then summed, there are no cancellations between positive and negative values.

Definition 8.2.1 A residual is the difference between the observed response y and the predicted response $\hat{y}$. For (x_i, y_i) , its residual is $e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$.

The estimates $\hat{a}$ and $\hat{b}$ satisfy

$$Q(\hat{a}, \hat{b}) = \min\{Q(a, b)\} = \min\left\{\sum_{i=1}^n e_i^2\right\}.$$

So, we would like to minimize the sum of the squared distances of each observed response to its fitted value. That is, we want to minimize the error sum of squares (SSE), where:

$$\text{SSE} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Taking the partial derivatives of $Q(a, b)$ with respect to a, b and setting the partial derivatives to zero,

$$\begin{cases} \frac{\partial Q}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial Q}{\partial b} = -2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0 \end{cases}$$

The two equations are called normal equations. They can be rearranged as

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

The coefficient determinant is

$$\begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = n \sum_{i=1}^n (x_i - \bar{x})^2 \neq 0,$$

since not all the x_i 's are the same. We denote

$$\begin{aligned} S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2, \\ S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2, \\ S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right). \end{aligned}$$

Then the normal equations have unique solution:

$$\begin{cases} \hat{a} = \bar{y} - \hat{b} \bar{x} \\ \hat{b} = \frac{S_{xy}}{S_{xx}} \end{cases}$$

The simple regression line always passes through the mean of the y variable and the mean of the x variable. That is, $\hat{y} - \bar{y} = \hat{b}(x - \bar{x})$.

Example 8.2.1 A real estate agent wishes to examine the relationship between the selling price Y (in \$ 1000s) of a house and its size X (measured in square feet). A random sample of 10 houses is selected. The data are reported in Table 8.1. Determine the linear regression line.

Table 8.1

	1	2	3	4	5	6	7	8	9	10
x	1400	1600	1700	1875	1100	1550	2350	2450	1425	1700
y	245	312	279	308	199	219	405	324	319	255

Solution: The scatterplot (see Figure 8. 2) shows a linear relationship between the two variables. The calculations are reported in Table 8. 2.

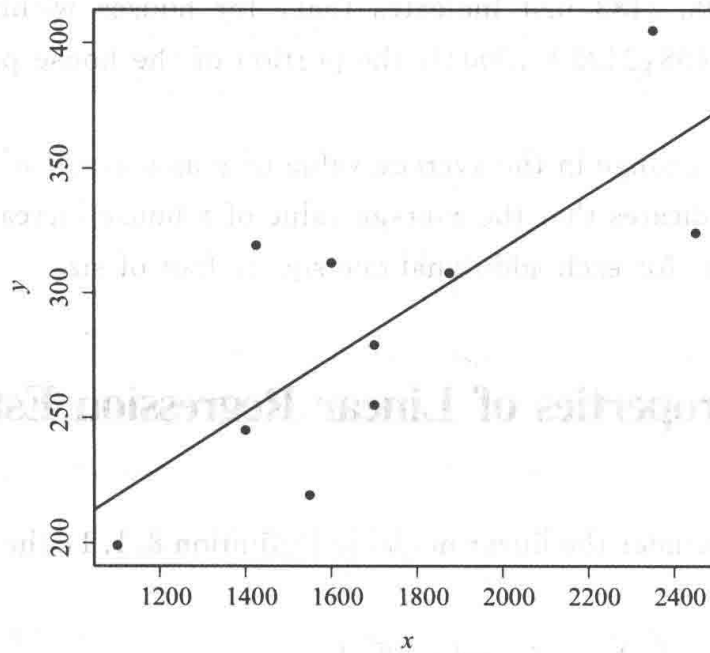


Figure 8. 2

Table 8. 2

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	1400	245	1960000	60025	343000
2	1600	312	2560000	97344	499200
3	1700	279	2890000	77841	474300
4	1875	308	3515625	94864	577500
5	1100	199	1210000	39601	218900
6	1550	219	2402500	47961	339450
7	2350	405	5522500	164025	951750
8	2450	324	6002500	104976	793800
9	1425	319	2030625	101761	454575
10	1700	255	2890000	65025	433500
Sum	17150	2865	30983750	853423	5085975
	$\bar{x} = 1715$	$\bar{y} = 286.5$	$S_{xx} = 1571500$	$S_{yy} = 32600.5$	$S_{xy} = 172500$

Therefore,

$$\begin{cases} \hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{172500}{1571500} = 0.1098 \\ \hat{a} = \bar{y} - \hat{b} \bar{x} = 286.5 - \frac{172500}{1571500} \times 1715 = 98.2483 \end{cases}$$

The regression line is $\hat{y} = 98.2483 + 0.1098x$.

$\hat{a}$ is the estimated average value of y when the value of x is zero. Since no house has 0 square feet, so $\hat{a} = 98.2483$ just indicates that, for houses within the range of sizes observed, \$98248.3 (98.2483×1000) is the portion of the house price not explained by square feet.

$\hat{b}$ is the estimated change in the average value of y as a result of a one unit change in x . Here $\hat{b} = 0.1098$ indicates that the average value of a house increases by $0.1098 \times 1000 = \$109.8$, on average, for each additional one square foot of size.

◆ 8.3 Properties of Linear Regression Estimators ◆

Theorem 8.3.1 Under the linear model in Definition 8.1.1, the following statements are valid.

$$(1) \hat{a} \sim N\left(a, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2\right), \hat{b} \sim N\left(b, \frac{\sigma^2}{S_{xx}}\right).$$

$$(2) \text{Cov}(\hat{a}, \hat{b}) = -\frac{\bar{x}}{S_{xx}}\sigma^2.$$

$$(3) \text{For a given } x_0, \hat{y}_0 = \hat{a} + \hat{b}x_0 \sim N\left(a + bx_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right).$$

Proof: (1) Notice that $\sum_{i=1}^n (x_i - \bar{x}) = 0$, thus,

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] y_i.$$

That is, $\hat{a}$ and $\hat{b}$ are linear combinations of independent $y_i \sim N(a + bx_i, \sigma^2)$, $i = 1, 2, \dots, n$. Therefore, $\hat{a}$ and $\hat{b}$ are normal random variables.

$$E(\hat{b}) = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} E(y_i) = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} (a + bx_i) = b.$$

$$V(\hat{b}) = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S_{xx}} \right)^2 V(y_i) = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S_{xx}} \right)^2 \sigma^2 = \frac{\sigma^2}{S_{xx}}.$$

$$E(\hat{a}) = E(\bar{y}) - E(\hat{b})\bar{x} = a + b\bar{x} - b\bar{x} = a.$$

$$V(\hat{a}) = \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right]^2 V(y_i) = \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2.$$

$$\begin{aligned} (2) \text{Cov}(\hat{a}, \hat{b}) &= \text{Cov}\left(\sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] y_i, \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i\right) \\ &= \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] \frac{x_i - \bar{x}}{S_{xx}} \text{Cov}(y_i, y_i) \\ &= -\frac{\bar{x}}{S_{xx}} \sigma^2. \end{aligned}$$

(3) Since $\hat{y}_0 = \hat{a} + \hat{b}x_0$ is a linear combination of independent $y_i \sim N(a + bx_i, \sigma^2)$, it follows normal distribution.

$$\begin{aligned} E(\hat{y}_0) &= E(\hat{a}) + E(\hat{b})x_0 = a + bx_0. \\ V(\hat{y}_0) &= V(\hat{a}) + V(\hat{b})x_0^2 + 2x_0 \text{Cov}(\hat{a}, \hat{b}) \\ &= \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2 + \frac{x_0^2 \sigma^2}{S_{xx}} - \frac{2x_0 \bar{x}}{S_{xx}} \sigma^2 \\ &= \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right) \sigma^2. \end{aligned}$$

Note: $\hat{a}$, $\hat{b}$ and $\hat{y}_0 = \hat{a} + \hat{b}x_0$ are unbiased estimators of a , b and $E(y_0) = a + bx_0$, respectively. If $\bar{x} = 0$, $\hat{a}$ and $\hat{b}$ are uncorrelated.

Definition 8.3.1 The measure of total variation is denoted by

$$\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2.$$

The sum of squares due to regression model (explained variation) is

$$\text{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

The sum of squares due to error (unexplained variation) is

$$\text{SSE} = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Theorem 8.3.2 Under the linear model in Definition 8.1.1, the total sum of squares can be partitioned into two components, the regression sum of squares and the sum of squares due to error. That is,

$$\text{SST} = \text{SSR} + \text{SSE}.$$

Proof: Since $\hat{a}$ and $\hat{b}$ satisfy the normal equations, we have

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i) &= \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n \epsilon_i = 0, \\ \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)x_i &= \sum_{i=1}^n (y_i - \hat{y}_i)x_i = \sum_{i=1}^n \epsilon_i x_i = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)[\hat{b}(x_i - \bar{x})] \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \end{aligned}$$

Theorem 8.3.3 Under the linear model in Definition 8.1.1, the following statements are valid.

$$(1) E(SSR) = \sigma^2 + b^2 S_{xx}.$$

$$(2) E(SSE) = (n-2)\sigma^2.$$

Proof: (1)

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n [\bar{y} + \hat{b}(x_i - \bar{x}) - \bar{y}]^2 = \hat{b}^2 S_{xx}.$$

Thus,

$$E(SSR) = E(\hat{b}^2 S_{xx}) = \left[\frac{\sigma^2}{S_{xx}} + b^2 \right] S_{xx} = \sigma^2 + b^2 S_{xx}.$$

(2) Notice that $\sum_{i=1}^n (x_i - \bar{x}) = 0$, thus,

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] y_i.$$

By the independence of ϵ_i and y_j ($i \neq j$),

$$\sum_{i=1}^n E(\hat{a}\epsilon_i) = \sum_{i=1}^n E\left\{ \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] y_i \epsilon_i \right\} = \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] \sigma^2 = \sigma^2.$$

$$\sum_{i=1}^n x_i E(\hat{b}\epsilon_i) = \sum_{i=1}^n x_i E\left(\frac{x_i - \bar{x}}{S_{xx}} y_i \epsilon_i \right) = \sum_{i=1}^n \frac{x_i(x_i - \bar{x})}{S_{xx}} \sigma^2 = \sigma^2.$$

Therefore,

$$\begin{aligned} E(SSE) &= E\left[\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right] \\ &= E\left[\sum_{i=1}^n (a + bx_i + \epsilon_i - \hat{a} - \hat{b}x_i)^2 \right] \\ &= E\left\{ \sum_{i=1}^n \left[(a - \hat{a})^2 + x_i^2 (b - \hat{b})^2 + \epsilon_i^2 + 2(a - \hat{a})(b - \hat{b})x_i + 2(a - \hat{a})\epsilon_i + 2(b - \hat{b})x_i\epsilon_i \right] \right\} \\ &= nV(\hat{a}) + V(\hat{b}) \sum_{i=1}^n x_i^2 + n\sigma^2 + 2n\bar{x}\text{Cov}(\hat{a}, \hat{b}) - 2\sigma^2 - 2\sigma^2 \\ &= n\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2 + \frac{\sigma^2}{S_{xx}} \sum_{i=1}^n x_i^2 + n\sigma^2 - 2n\bar{x} \frac{\bar{x}}{S_{xx}} \sigma^2 - 4\sigma^2 \\ &= (n-2)\sigma^2. \end{aligned}$$

Note: $\hat{\sigma}^2 = \frac{SSE}{n-2}$ is an unbiased estimator of σ^2 .

Theorem 8.3.4 Under the linear model in Definition 8.1.1, the following statements are valid.

(1) SSR, SSE and $\bar{y}$ are independent.

(2) $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$.

(3) If $b=0$, $\frac{SSR}{\sigma^2} \sim \chi^2(1)$.

Proof: $y_i \sim N(a + bx_i, \sigma^2)$ ($i=1, 2, \dots, n$) are independent. Construct an orthonormal matrix as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{(n-2)1} & a_{(n-2)2} & \cdots & a_{(n-2)n} \\ \frac{x_1 - \bar{x}}{\sqrt{S_{xx}}} & \frac{x_2 - \bar{x}}{\sqrt{S_{xx}}} & \cdots & \frac{x_n - \bar{x}}{\sqrt{S_{xx}}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{bmatrix}$$

For $i=1, 2, \dots, n-2$,

$$\sum_j a_{ij} = 0, \sum_j a_{ij} x_j = 0, \sum_j a_{ij}^2 = 1.$$

For $1 \leq i < j \leq n-2$,

$$\sum_k a_{ik} a_{jk} = 0.$$

Let $Z = AY = A[y_1 \ y_2 \ \cdots \ y_n]^T$, then Z follows an n -dimensional normal distribution.

$$E(Z) = AE(Y) = (0, \dots, 0, b\sqrt{S_{xx}}, \sqrt{n}(a + b\bar{x}))^T$$

$$V(Z) = AV(Y)A^T = \sigma^2 I_n.$$

So $z_1, z_2, \dots, z_n$ are independent. $z_1, z_2, \dots, z_{n-2} \sim N(0, \sigma^2)$,

$$\begin{aligned} z_{n-1} &= \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sqrt{S_{xx}}} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{S_{xx}}} \\ &= \frac{S_{xy}}{\sqrt{S_{xx}}} \\ &= \sqrt{S_{xx}} \hat{b} \sim N(b\sqrt{S_{xx}}, \sigma^2). \end{aligned}$$

$$z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n y_i = \sqrt{n} \bar{y} \sim N(\sqrt{n}(a + b\bar{x}), \sigma^2).$$

Since $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n [\bar{y} + \hat{b}(x_i - \bar{x}) - \bar{y}]^2 = \hat{b}^2 S_{xx} = z_{n-1}^2.$

$$\sum_{i=1}^n z_i^2 = Z^T Z = Y^T Y = \sum_{i=1}^n y_i^2 = SST + n\bar{y}^2 = SSE + z_{n-1}^2 + z_n^2.$$

Thus, $SSE = z_1^2 + \cdots + z_{n-2}^2$, $SSR = z_{n-1}^2$, $\bar{y} = \frac{z_n}{\sqrt{n}}$ are independent.

$$\frac{SSE}{\sigma^2} = \sum_{i=1}^{n-2} \left(\frac{z_i}{\sigma} \right)^2 \sim \chi^2(n-2)$$

If $b=0$, $z_{n-1} \sim N(b\sqrt{S_{xx}}, \sigma^2) = N(0, \sigma^2)$,

$$\frac{SSR}{\sigma^2} = \left(\frac{z_{n-1}}{\sigma} \right)^2 \sim \chi^2(1).$$

Note: since $SSR = \hat{b}^2 S_{xx}$, thus $\hat{b}$, SSE and $\bar{y}$ are also independent.

◆ 8.4 Inferences Concerning the Slope ◆

The aim of constructing a regression line is to find the relation between the predictor x and the response y . That is, $E(y) = a + bx$. If $b=0$, $E(y)$ does not change as x changes, there is no evidence of a relationship between the two variables. This regression equation is meaningless. If $b \neq 0$, then we conclude that there is evidence of a relationship between the predictor x and the response y .

We want to test

$$H_0: b=0, \quad H_1: b \neq 0.$$

Rejecting H_0 indicates that the regression equation is significant or the equation is practically useful.

Under $H_0: b=0$, $\frac{SSR}{\sigma^2} \sim \chi^2(1)$, $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$, and they are independent as we discussed in previous section. So

$$F = \frac{\frac{SSR}{1}}{\frac{SSE}{n-2}} \sim F(1, n-2).$$

We can use this F -statistic to test the hypotheses. For significance level α , the rejection region is

$$W = \{F \geq F_{\alpha}(1, n-2)\}.$$

Example 8.4.1 In Example 8.2.1, we found the regression line is $\hat{y} = 98.2483 + 0.1098x$. Now we want to test its significance.

$$H_0: b=0, \quad H_1: b \neq 0.$$

Solution:

$$SST = S_{yy} = 32600.5, \quad DF = 10 - 1 = 9.$$

$$SSR = \hat{b}^2 S_{xx} = 0.1098^2 \times 1571500 = 18946.07, \quad DF = 1.$$

$$SSE = SST - SSR = 13654.43, \quad DF = 8.$$

$$F = \frac{\frac{18946.07}{1}}{\frac{13654.43}{8}} = 11.10.$$

The calculations are reported in Table 8.3

Table 8.3

Source	Sum of squares	Degrees of Freedom	Mean Square	F-value	p-value
Model	SSR=18946.07	1	18946.07	11.10	0.01039
Error	SSE=13654.43	8	1706.804	—	—
Total	SST=32600.5	—	—	—	—

At the level $\alpha = 0.05$, $11.10 > F_{0.05}(1, 8) = 5.3177$, we reject H_0 . The regression model is significant.

In this case, the p -value is the probability that we would observe an $F(1, 8)$ random variable more extreme than 11.10,

$$p = P(F(1, 8) \geq 11.10) = 0.01039 < 0.05.$$

The p -value approach yields the same conclusion.

We can also use t -statistic to conduct the test. Since $\hat{b} \sim N\left(b, \frac{\sigma^2}{S_{xx}}\right)$ and $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$ are independent. Under $H_0: b=0$,

$$t = \frac{\frac{\hat{b} - 0}{\sqrt{\frac{\sigma^2}{S_{xx}}}}}{\sqrt{\frac{\frac{SSE}{\sigma^2}}{n-2}}} = \frac{\hat{b} \sqrt{S_{xx}}}{\frac{SSE}{n-2}} \sim t(n-2).$$

For significance level α , the rejection region is

$$W = \{|t| \geq t_{\frac{\alpha}{2}}(n-2)\}.$$

Notice that $t^2 = F$, the t -test and F -test are equivalent.

Example 8.4.2 (revisited) In Example 8.2.1, we found the regression line is $\hat{y} = 98.2483 + 0.1098x$. Now we want to test its significance.

$$H_0: b=0, \quad H_1: b \neq 0.$$

Solution: The value of t -statistic is

$$t = \frac{\hat{b} \sqrt{S_{xx}}}{\sqrt{\frac{SSE}{n-2}}} = \frac{0.1098 \times \sqrt{1571500}}{\sqrt{\frac{13654.43}{8}}} = 3.33.$$

At the level $\alpha = 0.05$, $t_{0.025}(8) = 0.82 < 3.33$, we reject H_0 . The regression model is significant.

In this case, the p -value is

$$p = P[|t(8)| \geq 3.33] = 2P[t(8) \geq 3.33] = 0.01039 < 0.05.$$

The p -value approach yields the same conclusion.

◆ 8.5 Regression Validity ◆

When fitting the data with a regression line, we usually make the following assumptions: The relationship between mean response and predictor is indeed linear. Unexplained variation is normally and independently distributed with constant variance.

To check the assumptions of linearity and constant variance, we may consider the residuals which give useful information about the contribution of individual data points to the overall pattern of scatter.

The residuals versus fitted values plot is the most frequently used plot. It is a scatter plot of residuals $y - \hat{y}$ on the vertical axis and fitted values $\hat{y}$ on the horizontal axis.

Example 8.5.1 (revisited) In Example 8.2.1, we found the regression line is $\hat{y} = 98.2483 + 0.1098x$. Now we want to use residual plots to check for regression validity.

Solution: The residuals are reported in Table 8.4. The residuals versus fitted values plot is sketched in Figure 8.3. Looking at the plot, we have a horizontal band of data of roughly constant width. So the assumptions of linearity and constant variance do seem to hold here.

Table 8.4

	x	y	$\hat{y}$	$y - \hat{y}$
1	1400	245	251.92	-6.92
2	1600	312	273.88	38.12
3	1700	279	284.85	-5.85
4	1875	308	304.06	3.94
5	1100	199	218.99	-19.99
6	1550	219	268.39	-49.39
7	2350	405	356.20	48.80
8	2450	324	367.18	-43.18
9	1425	319	254.67	64.33
10	1700	255	284.85	-29.85

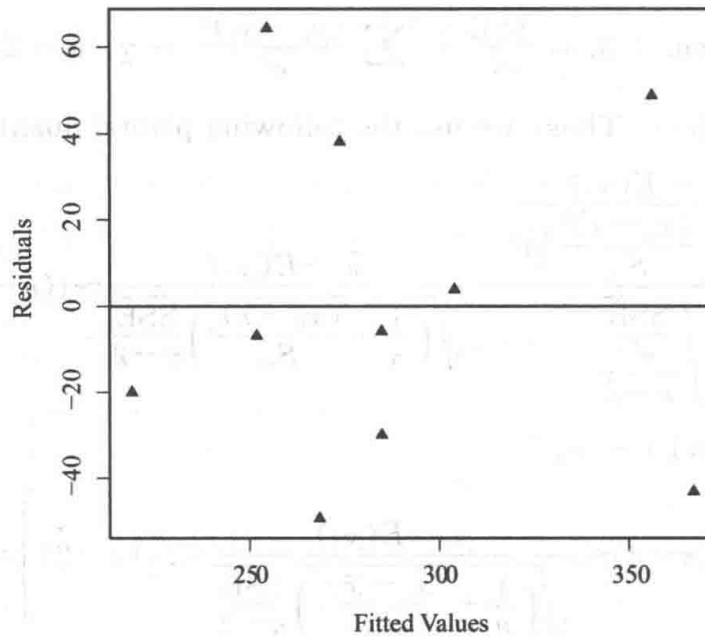


Figure 8.3

In general, if residuals are scattered randomly around 0 with uniform variation, it indicates that the data fit a linear model, have normally distributed residuals for each value of x , and constant variance.

◆ 8.6 Confidence Interval for Mean Response ◆

After the test of significance, the regression equation can be used to estimate and predict some values. In this section, we develop the confidence interval for mean response.

We will use the following notations:

- (1) x_0 = the given value of the independent variable x .
- (2) y_0 = the value of the dependent variable y corresponding to the given x_0 .
- (3) $E(y_0)$ = the mean value of the dependent variable y corresponding to x_0 .
- (4) $\hat{y}_0 = \hat{a} + \hat{b}x_0$.

Where y_0 is a random variable, $E(y_0) = a + bx_0$ is a parameter.

Since $\hat{a}, \hat{b}$ are unbiased estimators of a and b , then $\hat{y}_0$ is the unbiased estimator of $E(y_0)$.

$$E(\hat{y}_0) = E(\hat{a} + \hat{b}x_0) = a + bx_0 = E(y_0).$$

In general, we cannot expect $\hat{y}_0$ to equal $E(y_0)$ exactly. By Theorem 8.3.1, for a given x_0 , $\hat{y} = \hat{a} + \hat{b}x_0 \sim N\left(a + bx_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right)$. Thus,

$$\frac{\hat{y}_0 - E(y_0)}{\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2}} \sim N(0, 1).$$

In addition, by Theorem 8.3.4, $\frac{SSE}{\sigma^2} = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2(n-2)$, SSE and $\hat{y}_0 = \bar{y} + \hat{b}(x_0 - \bar{x})$ are independent. Thus, we use the following pivotal quantity

$$\frac{\frac{\hat{y}_0 - E(y_0)}{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2}}{\sqrt{\frac{\frac{SSE}{\sigma^2}}{n-2}}} = \frac{\hat{y}_0 - E(y_0)}{\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\frac{SSE}{n-2}}} \sim t(n-2).$$

For confidence level $1 - \alpha$,

$$P\left[-t_{\frac{\alpha}{2}}(n-2) < \frac{\hat{y}_0 - E(y_0)}{\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\frac{SSE}{n-2}}} < t_{\frac{\alpha}{2}}(n-2)\right] = 1 - \alpha,$$

where $t_{\frac{\alpha}{2}}(n-2)$ is the $\frac{\alpha}{2}$ upper quantile of $t(n-2)$. Therefore, the $1 - \alpha$ confidence interval for mean response $E(y_0) = a + bx_0$ is

$$\left(\hat{y}_0 - t_{\frac{\alpha}{2}}(n-2)\sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\frac{SSE}{n-2}}, \hat{y}_0 + t_{\frac{\alpha}{2}}(n-2)\sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\frac{SSE}{n-2}}\right).$$

Notice that the width of the $1 - \alpha$ confidence interval for mean response $E(y_0) = a + bx_0$ is

$$2t_{\frac{\alpha}{2}}(n-2)\sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\frac{SSE}{n-2}},$$

which is smallest when $x_0 = \bar{x}$ and the quantity $x_0 - \bar{x} = 0$. In this case, the interval width becomes

$$2t_{\frac{\alpha}{2}}(n-2)\sqrt{\frac{SSE}{n(n-2)}}.$$

This result implies that we can make the most precise estimate of $E(y_0)$ whenever $x_0 = \bar{x}$. The confidence intervals for the mean response will become wider as x_0 deviates more from $\bar{x}$.

Example 8.6.1 (revisited) In Example 8.2.1, we found the regression line is $\hat{y} = 98.2483 + 0.1098x$. Now we want to develop a 95% confidence interval of the mean selling price for all houses with 2000 square feet.

Solution: $x_0 = 2000$, $\hat{y}_0 = 98.2483 + 0.1098 \times 2000 = 317.8483$, $\alpha = 0.05$, $t_{\frac{\alpha}{2}}(n-2) = t_{0.025}(8) = 2.036$, $\bar{x} = 1715$, $S_{xx} = 1571500$, $SSE = 13654.43$, the 95% confidence interval of $E(y_0) = a + bx_0$ is

$$\begin{aligned} & \left(\hat{y}_0 - t_{\frac{\alpha}{2}}(n-2)\sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\frac{SSE}{n-2}}, \hat{y}_0 + t_{\frac{\alpha}{2}}(n-2)\sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\frac{SSE}{n-2}}\right) \\ &= \left(317.8483 - 2.036 \times \sqrt{\left(\frac{1}{10} + \frac{(2000 - 1715)^2}{1571500}\right) \times \frac{13654.43}{8}}, \right. \end{aligned}$$

$$317.8483 + 2.036 \times \sqrt{\left(\frac{1}{10} + \frac{(2000 - 1715)^2}{1571500}\right) \times \frac{13654.43}{8}}$$

$$= (285.0884, 350.6082).$$

If x_0 takes other values, we can obtain the confidence intervals of $E(y_0)$ in the same way. Confidence interval width is smallest at $x_0 = \bar{x} = 1715$. The confidence intervals for the mean response will become wider as x_0 deviates more from $\bar{x} = 1715$. Figure 8.4 shows the confidence band around the regression line $\hat{y} = 98.2483 + 0.1098x$.

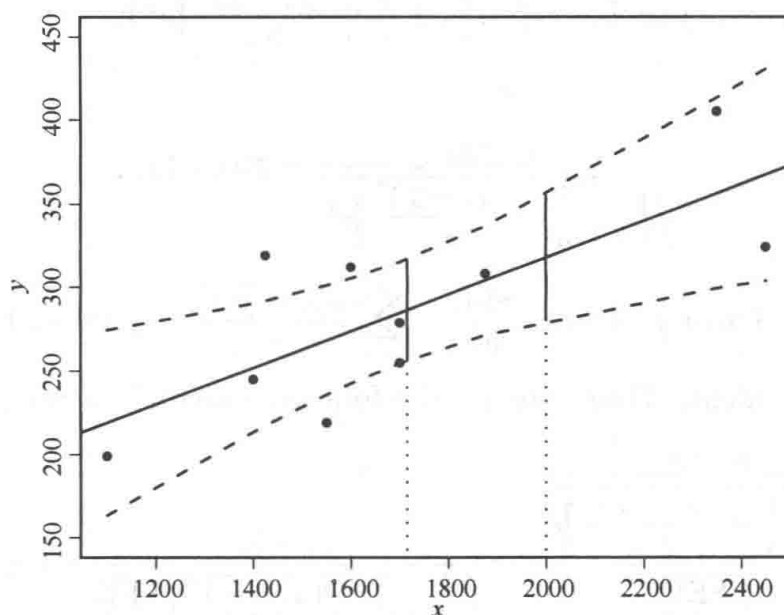


Figure 8.4

◆ 8.7 Inference for Prediction ◆

In previous section, we learned how to calculate a confidence interval for mean response $E(y_0)$, for a given x_0 . Suppose now we're not interested in the average, but instead we want to predict a future value of y_0 for a given x_0 .

A prediction interval is similar to a confidence interval in previous section, except that the prediction interval is developed to cover a random variable y_0 , while the confidence interval is constructed to cover the fixed parameter $E(y_0)$, for a given x_0 .

$$y_0 = a + bx_0 + \epsilon_0 = E(y_0) + \epsilon_0, \quad \epsilon_0 \sim N(0, \sigma^2).$$

$$y_0 \sim N(a + bx_0, \sigma^2).$$

Notice that $E(y_0) = a + bx_0$ can be estimated by $\hat{y}_0 = \hat{a} + \hat{b}x_0$, where $\hat{a}$, $\hat{b}$ are unbiased estimators of a and b .

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = \sum_{i=1}^n \left[\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{S_{xx}} \right] y_i.$$

That is, $\hat{a}$ and $\hat{b}$ are linear combinations of independent $y_i \sim N(a + bx_i, \sigma^2), i=1, 2, \dots, n$. So $\hat{y}_0 = \hat{a} + \hat{b}x_0$ is a linear combination of independent $y_i, i=1, 2, \dots, n$.

By Theorem 8.3.1, $\hat{y}_0 = \hat{a} + \hat{b}x_0 \sim N\left(a + bx_0, \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right)$. It follows that y_0 and $\hat{y}_0$ are independent.

$$y_0 - \hat{y}_0 \sim N\left(0, \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\sigma^2\right).$$

Thus,

$$\frac{y_0 - \hat{y}_0}{\sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\sigma^2}} \sim N(0, 1).$$

In addition, by Theorem 8.3.4, $\frac{SSE}{\sigma^2} = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi^2(n-2)$, SSE and $\hat{y}_0 = \bar{y} + \hat{b}(x_0 - \bar{x})$ are independent. Thus, we use the following pivotal quantity.

$$\frac{\frac{y_0 - \hat{y}_0}{\sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]\sigma^2}}}{\sqrt{\frac{SSE}{\sigma^2} / (n-2)}} = \frac{y_0 - \hat{y}_0}{\sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{SSE}{n-2}}} \sim t(n-2).$$

For confidence level $1 - \alpha$,

$$P\left\{-t_{\frac{\alpha}{2}}(n-2) < \frac{y_0 - \hat{y}_0}{\sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{SSE}{n-2}}} < t_{\frac{\alpha}{2}}(n-2)\right\} = 1 - \alpha,$$

where $t_{\frac{\alpha}{2}}(n-2)$ is the $\frac{\alpha}{2}$ upper quantile of $t(n-2)$. Therefore, the $1 - \alpha$ prediction interval for y_0 is

$$\left(\hat{y}_0 - t_{\frac{\alpha}{2}}(n-2) \sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{SSE}{n-2}}, \hat{y}_0 + t_{\frac{\alpha}{2}}(n-2) \sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{SSE}{n-2}}\right).$$

The formula is very similar, except there is an added 1 in the formula. The prediction interval is wider than the confidence interval for a given x_0 and confidence level. The width of the confidence interval is zero if n is large enough, this is not true for the prediction interval.

The width of the $1 - \alpha$ prediction interval for y_0 is

$$2t_{\frac{\alpha}{2}}(n-2) \sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{SSE}{n-2}},$$

which is smallest when $x_0 = \bar{x}$ and the quantity $x_0 - \bar{x} = 0$. In this case, the interval width becomes

$$2t_{\frac{\alpha}{2}}(n-2)\sqrt{\left(1+\frac{1}{n}\right)\frac{\text{SSE}}{n-2}}.$$

This result implies that we can make the most precise prediction of y_0 whenever $x_0 = \bar{x}$. The prediction intervals for the response will become wider as x_0 deviates more from $\bar{x}$.

Example 8.7.1 (revisited) In Example 8.2.1, we found the regression line is $\hat{y} = 98.2483 + 0.1098x$. Now we want to develop a 95% prediction interval of the selling price for a house with 2000 square feet.

Solution: $x_0 = 2000$, $\hat{y}_0 = 98.2483 + 0.1098 \times 2000 = 317.8483$, $\alpha = 0.05$, $t_{\frac{\alpha}{2}}(n-2) = t_{0.025}(8) = 2.036$, $\bar{x} = 1715$, $S_{xx} = 1571500$, $\text{SSE} = 13654.43$, the 95% prediction interval of y_0 is

$$\begin{aligned} & \left(\hat{y}_0 - t_{\frac{\alpha}{2}}(n-2) \sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{\text{SSE}}{n-2}}, \hat{y}_0 + t_{\frac{\alpha}{2}}(n-2) \sqrt{\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right] \frac{\text{SSE}}{n-2}} \right) \\ &= \left(317.8483 - 2.036 \times \sqrt{\left[1 + \frac{1}{10} + \frac{(2000 - 1715)^2}{1571500}\right] \times \frac{13654.43}{8}}, \right. \\ & \quad \left. 317.8483 + 2.036 \times \sqrt{\left[1 + \frac{1}{10} + \frac{(2000 - 1715)^2}{1571500}\right] \times \frac{13654.43}{8}} \right) \\ &= (227.5797, 408.1169). \end{aligned}$$

If x_0 takes other values, we can obtain the prediction intervals of y_0 in the same way. The width of prediction interval is smallest at $x_0 = \bar{x} = 1715$. The prediction intervals for the response will become wider as x_0 deviates more from $\bar{x} = 1715$. Figure 8.5 shows the prediction band and the confidence band around the regression line $\hat{y} = 98.2483 + 0.1098x$.

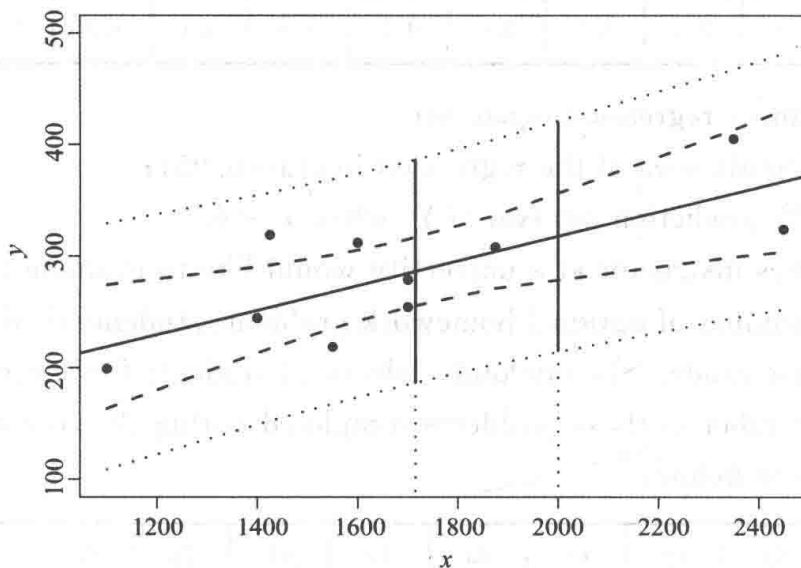


Figure 8.5

Exercise 8

8.1 To investigate the problem of the solubility of a substance in water, the data of dissolution mass and temperature are obtained and shown below

Temperature x_i	0	4	10	15	21	29	36	51	68
Dissolution mass Y_i	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

Assume Y obeys the linear model

$$Y = a + bx + \epsilon, \epsilon \sim N(0, \sigma^2).$$

Determine the estimates of a, b and σ^2 .

8.2 The relationship between the reproduction quantity and the month of a substance is as follows:

Month x_i	2	4	6	8	10
Reproduction quantity Y_i	66	120	210	270	320

Suppose that the linear model is obeyed. $E(Y) = \mu(x) = \beta_0 + \beta_1 x$.

- (1) Find the estimates of β_0 and β_1 .
- (2) Test whether β_1 is equal to zero or not ($\alpha = 0.05$).

8.3 The strength Y of a synthetic material and its tensile multiples x have the following relationship:

x_i	2.0	2.5	2.7	3.5	4.0	4.5	5.2	6.3	7.1	8.0	9.0	10.0
Y_i	1.3	2.5	2.5	2.7	3.5	4.2	5.0	6.4	6.3	7.0	8.0	8.1

- (1) Find the linear regression equation;
- (2) Test the significance of the regression line ($\alpha = 0.05$);
- (3) Find a 95% prediction interval of Y_0 when $x_0 = 6$.

8.4 A statistics instructor at a university would like to examine the relationship (if any) between the number of optional homework problems students do during the semester and their final course grade. She randomly selects 12 students for study and asks them to keep track of the number of these problems completed during the course of the semester. The data are reported below:

Problems	51	58	62	65	68	76	77	78	78	84	85	91
Grade	62	68	66	66	67	72	73	72	78	73	76	75

- (1) For this setting identify the response variable and the predictor variable.
- (2) Find the regression equation.
- (3) Conduct a hypothesis test for significant linear equation.
- (4) Use the regression equation to predict a student's final course grade if 70 optional homework assignments are done.

8.5 The sales of a company (in million dollars) for each year are shown in the table below.

$x(\text{year})$	2005	2006	2007	2008	2009
$y(\text{sales})$	12	19	29	37	45

- (1) Find the least square regression line $y = ax + b$.
- (2) Use the least square regression line to estimate the sales of the company in 2012.

8.6 A dietetics student wants to look at the relationship between calcium intake (mg/day) and knowledge score (out of 50) about calcium in sports science students. The table below shows the data she collected.

Respondent	Knowledge score	Calcium intake	Respondent	Knowledge score	Calcium intake
1	10	450	11	38	940
2	42	1050	12	25	733
3	38	900	13	48	985
4	15	525	14	28	763
5	22	710	15	22	583
6	32	854	16	45	850
7	40	800	17	18	798
8	14	493	18	24	754
9	26	730	19	30	805
10	32	894	20	43	1085

- (1) Find the least square regression line.
- (2) What is the predicted calcium intake for a student with knowledge of calcium score equal to 30 (out of 50)?

8.7 In animal studies, it is sometimes necessary to find out the relationship between the volume and weight of a certain animal. Since the weight of an animal is relatively easy to measure, and the measurement of volume is more difficult, people want to use the weight of an animal to predict the volume. The table below reports the data of volumes

and weights for 18 certain animals. Here, the weight x (unit: kg) is considered as an independent variable and the volume y (unit: dm^3) is the response variable.

x	y	x	y	x	y
10.40	10.20	15.10	14.80	16.50	15.90
10.50	10.40	15.10	15.10	16.70	16.60
11.90	11.60	15.10	14.50	17.10	16.70
12.10	11.90	15.70	15.70	17.10	16.70
13.80	13.50	15.80	15.20	17.80	17.60
15.00	14.50	16.00	15.80	18.40	18.30

- (1) Find the least square regression line.
- (2) Conduct a hypothesis test for the significance of the linear equation.
- (3) What is the predicted volume if an animal's weight is 17.6 kg.
- (4) What is the 95% prediction interval of the volume for an animal with 17.6 kg.

Appendix A Binomial Probability Distribution

$$P(X=k) = C_n^k p^k (1-p)^{n-k}$$

	<i>k</i>	<i>p</i>										
		0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
<i>n</i> =1	0	0.9900	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	0.0100	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000
<i>n</i> =2	0	0.9801	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.0198	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
	2	0.0001	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
<i>n</i> =3	0	0.9703	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.0294	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
	2	0.0003	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
	3	0	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
<i>n</i> =4	0	0.9606	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.0388	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
	2	0.0006	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
	3	0	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
	4	0	0	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
<i>n</i> =5	0	0.9510	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.0480	0.2036	0.3281	0.3915	0.4096	0.3955	0.3601	0.3124	0.2592	0.2059	0.1562
	2	0.0010	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
	3	0	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
	4	0	0	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1562
	5	0	0	0	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0312
<i>n</i> =6	0	0.9415	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.0571	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0937
	2	0.0014	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344

continued

	<i>k</i>	<i>p</i>										
		0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
<i>n</i> = 6	3	0	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
	4	0	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
	5	0	0	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0937
	6	0	0	0	0	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156
<i>n</i> = 7	0	0.9321	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.0659	0.2573	0.3720	0.3960	0.3670	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547
	2	0.0020	0.0406	0.1240	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.2140	0.1641
	3	0	0.0036	0.0230	0.0617	0.1147	0.1730	0.2269	0.2679	0.2903	0.2918	0.2734
	4	0	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734
	5	0	0	0.0002	0.0012	0.0043	0.0115	0.0250	0.0466	0.0774	0.1172	0.1641
	6	0	0	0	0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.0320	0.0547
	7	0	0	0	0	0	0.0001	0.0002	0.0006	0.0016	0.0037	0.0078
<i>n</i> = 8	0	0.9227	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.0746	0.2793	0.3826	0.3847	0.3355	0.2670	0.1977	0.1373	0.0896	0.0548	0.0312
	2	0.0026	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.2090	0.1569	0.1094
	3	0.0001	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2187
	4	0	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734
	5	0	0	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2187
	6	0	0	0	0.0002	0.0011	0.0038	0.0100	0.0217	0.0413	0.0703	0.1094
	7	0	0	0	0	0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0312
	8	0	0	0	0	0	0	0.0001	0.0002	0.0007	0.0017	0.0039
<i>n</i> = 9	0	0.9135	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.0830	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176
	2	0.0034	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703
	3	0.0001	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
	4	0	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461
	5	0	0	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
	6	0	0	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641
	7	0	0	0	0	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
	8	0	0	0	0	0	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
	9	0	0	0	0	0	0	0	0.0001	0.0003	0.0008	0.0020

continued

	<i>k</i>	<i>p</i>										
		0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
<i>n</i> = 10	0	0.9044	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.0914	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
	2	0.0042	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
	3	0.0001	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172
	4	0	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051
	5	0	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461
	6	0	0	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
	7	0	0	0	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172
	8	0	0	0	0	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439
	9	0	0	0	0	0	0	0.0001	0.0005	0.0016	0.0042	0.0098
	10	0	0	0	0	0	0	0	0	0.0001	0.0003	0.0010
<i>n</i> = 11	0	0.8953	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.0995	0.3293	0.3835	0.3248	0.2362	0.1549	0.0932	0.0518	0.0266	0.0125	0.0054
	2	0.0050	0.0867	0.2131	0.2866	0.2953	0.2581	0.1998	0.1395	0.0887	0.0513	0.0269
	3	0.0002	0.0137	0.0710	0.1517	0.2215	0.2581	0.2568	0.2254	0.1774	0.1259	0.0806
	4	0	0.0014	0.0158	0.0536	0.1107	0.1721	0.2201	0.2428	0.2365	0.2060	0.1611
	5	0	0.0001	0.0025	0.0132	0.0388	0.0803	0.1321	0.1830	0.2207	0.2360	0.2256
	6	0	0	0.0003	0.0023	0.0097	0.0268	0.0566	0.0985	0.1471	0.1931	0.2256
	7	0	0	0	0.0003	0.0017	0.0064	0.0173	0.0379	0.0701	0.1128	0.1611
	8	0	0	0	0	0.0002	0.0011	0.0037	0.0102	0.0234	0.0462	0.0806
	9	0	0	0	0	0	0.0001	0.0005	0.0018	0.0052	0.0126	0.0269
	10	0	0	0	0	0	0	0	0.0002	0.0007	0.0021	0.0054
	11	0	0	0	0	0	0	0	0	0	0.0002	0.0005
<i>n</i> = 12	0	0.8864	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.1074	0.3413	0.3766	0.3012	0.2062	0.1267	0.0712	0.0368	0.0174	0.0075	0.0029
	2	0.0060	0.0988	0.2301	0.2924	0.2835	0.2323	0.1678	0.1088	0.0639	0.0339	0.0161
	3	0.0002	0.0173	0.0852	0.1720	0.2362	0.2581	0.2397	0.1954	0.1419	0.0923	0.0537
	4	0	0.0021	0.0213	0.0683	0.1329	0.1936	0.2311	0.2367	0.2128	0.1700	0.1208
	5	0	0.0002	0.0038	0.0193	0.0532	0.1032	0.1585	0.2039	0.2270	0.2225	0.1934
	6	0	0	0.0005	0.0040	0.0155	0.0401	0.0792	0.1281	0.1766	0.2124	0.2256
	7	0	0	0	0.0006	0.0033	0.0115	0.0291	0.0591	0.1009	0.1489	0.1934
	8	0	0	0	0.0001	0.0005	0.0024	0.0078	0.0199	0.0420	0.0762	0.1208
	9	0	0	0	0	0.0001	0.0004	0.0015	0.0048	0.0125	0.0277	0.0537
	10	0	0	0	0	0	0	0.0002	0.0008	0.0025	0.0068	0.0161
	11	0	0	0	0	0	0	0	0.0001	0.0003	0.0010	0.0029
	12	0	0	0	0	0	0	0	0	0	0.0001	0.0002

Appendix B Poisson Cumulative Distribution

$$P(X \leq x) = \sum_{k=0}^x \frac{\lambda^k}{k!} e^{-\lambda}$$

k	λ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0.368
1	0.995	0.982	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.736
2	1	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920
3	1	1	1	0.999	0.998	0.997	0.994	0.991	0.987	0.981
4	1	1	1	1	1	1	0.999	0.999	0.998	0.996
5	1	1	1	1	1	1	1	1	1	0.999
6	1	1	1	1	1	1	1	1	1	1

k	λ									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150	0.135
1	0.699	0.663	0.627	0.592	0.558	0.525	0.493	0.463	0.434	0.406
2	0.900	0.879	0.857	0.833	0.809	0.783	0.757	0.731	0.704	0.677
3	0.974	0.966	0.957	0.946	0.934	0.921	0.907	0.891	0.875	0.857
4	0.995	0.992	0.989	0.986	0.981	0.976	0.970	0.964	0.956	0.947
5	0.999	0.998	0.998	0.997	0.996	0.994	0.992	0.990	0.987	0.983
6	1	1	1	0.999	0.999	0.999	0.998	0.997	0.997	0.995
7	1	1	1	1	1	1	1	0.999	0.999	0.999
8	1	1	1	1	1	1	1	1	1	1

k	λ									
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.122	0.111	0.100	0.091	0.082	0.074	0.067	0.061	0.055	0.050
1	0.380	0.355	0.331	0.308	0.287	0.267	0.249	0.231	0.215	0.199

continued

k	λ									
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
2	0.650	0.623	0.596	0.570	0.544	0.518	0.494	0.469	0.446	0.423
3	0.839	0.819	0.799	0.779	0.758	0.736	0.714	0.692	0.670	0.647
4	0.938	0.928	0.916	0.904	0.891	0.877	0.863	0.848	0.832	0.815
5	0.980	0.975	0.970	0.964	0.958	0.951	0.943	0.935	0.926	0.916
6	0.994	0.993	0.991	0.988	0.986	0.983	0.979	0.976	0.971	0.966
7	0.999	0.998	0.997	0.997	0.996	0.995	0.993	0.992	0.990	0.988
8	1	1	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.996
9	1	1	1	1	1	1	0.999	0.999	0.999	0.999
10	1	1	1	1	1	1	1	1	1	1

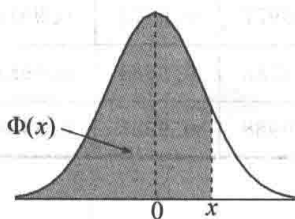
k	λ									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.045	0.041	0.037	0.033	0.030	0.027	0.025	0.022	0.020	0.018
1	0.185	0.171	0.159	0.147	0.136	0.126	0.116	0.107	0.099	0.092
2	0.401	0.380	0.359	0.340	0.321	0.303	0.285	0.269	0.253	0.238
3	0.625	0.603	0.580	0.558	0.537	0.515	0.494	0.473	0.453	0.433
4	0.798	0.781	0.763	0.744	0.725	0.706	0.687	0.668	0.648	0.629
5	0.906	0.895	0.883	0.871	0.858	0.844	0.830	0.816	0.801	0.785
6	0.961	0.955	0.949	0.942	0.935	0.927	0.918	0.909	0.899	0.889
7	0.986	0.983	0.980	0.977	0.973	0.969	0.965	0.960	0.955	0.949
8	0.995	0.994	0.993	0.992	0.990	0.988	0.986	0.984	0.981	0.979
9	0.999	0.998	0.998	0.997	0.997	0.996	0.995	0.994	0.993	0.992
10	1	1	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997
11	1	1	1	1	1	1	1	0.999	0.999	0.999
12	1	1	1	1	1	1	1	1	1	1

k	λ									
	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0.017	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.007	0.007
1	0.085	0.078	0.072	0.066	0.061	0.056	0.052	0.048	0.044	0.040
2	0.224	0.210	0.197	0.185	0.174	0.163	0.152	0.143	0.133	0.125
3	0.414	0.395	0.377	0.359	0.342	0.326	0.310	0.294	0.279	0.265
4	0.609	0.590	0.570	0.551	0.532	0.513	0.495	0.476	0.458	0.440
5	0.769	0.753	0.737	0.720	0.703	0.686	0.668	0.651	0.634	0.616

continued

k	λ									
	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
6	0.879	0.867	0.856	0.844	0.831	0.818	0.805	0.791	0.777	0.762
7	0.943	0.936	0.929	0.921	0.913	0.905	0.896	0.887	0.877	0.867
8	0.976	0.972	0.968	0.964	0.960	0.955	0.950	0.944	0.938	0.932
9	0.990	0.989	0.987	0.985	0.983	0.980	0.978	0.975	0.972	0.968
10	0.997	0.996	0.995	0.994	0.993	0.992	0.991	0.990	0.988	0.986
11	0.999	0.999	0.998	0.998	0.998	0.997	0.997	0.996	0.995	0.995
12	1	1	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998
13	1	1	1	1	1	1	1	1	0.999	0.999
14	1	1	1	1	1	1	1	1	1	1

Appendix C Standard Normal Table

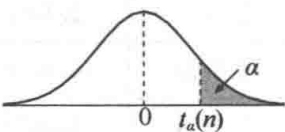


<i>x</i>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

continued

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Appendix D t-distribution Upper Quantiles $t_{\alpha}(n)$



$\alpha \backslash n$	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567
2	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314

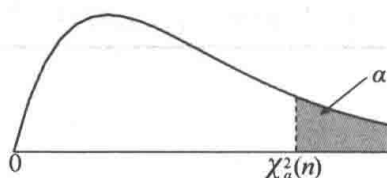
continued

$\alpha \backslash n$	0.20	0.15	0.10	0.05	0.025	0.01	0.005
22	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.8534	1.0541	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.8530	1.0535	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.8526	1.0530	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.8523	1.0525	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.8520	1.0520	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.8517	1.0516	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.8514	1.0512	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.8512	1.0508	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.8509	1.0504	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.8505	1.0497	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.8503	1.0494	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.8501	1.0491	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.8499	1.0488	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.8497	1.0485	1.3006	1.6794	2.0141	2.4121	2.6896
46	0.8495	1.0483	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.8493	1.0480	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.8492	1.0478	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.8490	1.0475	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.8487	1.0471	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.8486	1.0469	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.8485	1.0467	1.2977	1.6741	2.0057	2.3988	2.6718

continued

α n	0.20	0.15	0.10	0.05	0.025	0.01	0.005
54	0.8483	1.0465	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.8482	1.0463	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.8481	1.0461	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.8480	1.0459	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.8479	1.0458	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.8478	1.0456	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603

Appendix E χ^2 -distribution Upper Quantiles $\chi^2_{\alpha}(n)$



$\alpha \backslash n$	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401

continued

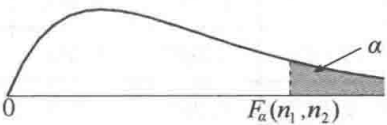
$\alpha \backslash n$	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.893	62.883
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
41	21.421	22.906	25.215	27.326	29.907	52.949	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	30.765	54.090	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	31.625	55.230	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	32.487	56.369	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957	73.166
46	25.041	26.657	29.160	31.439	34.215	58.641	62.830	66.617	71.201	74.437
47	25.775	27.416	29.956	32.268	35.081	59.774	64.001	67.821	72.443	75.704
48	26.511	28.177	30.755	33.098	35.949	60.907	65.171	69.023	73.683	76.969
49	27.249	28.941	31.555	33.930	36.818	62.038	66.339	70.222	74.919	78.231
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
51	28.735	30.475	33.162	35.600	38.560	64.295	68.669	72.616	77.386	80.747
52	29.481	31.246	33.968	36.437	39.433	65.422	69.832	73.810	78.616	82.001
53	30.230	32.018	34.776	37.276	40.308	66.548	70.993	75.002	79.843	83.253



continued

α n	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
54	30.981	32.793	35.586	38.116	41.183	67.673	72.153	76.192	81.069	84.502
55	31.735	33.570	36.398	38.958	42.060	68.796	73.311	77.380	82.292	85.749
56	32.490	34.350	37.212	39.801	42.937	69.919	74.468	78.567	83.513	86.994
57	33.248	35.131	38.027	40.646	43.816	71.040	75.624	79.752	84.733	88.236
58	34.008	35.913	38.844	41.492	44.696	72.160	76.778	80.936	85.950	89.477
59	34.770	36.698	39.662	42.339	45.577	73.279	77.931	82.117	87.166	90.715
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952

Appendix F F-distribution Upper Quantiles $F_{\alpha}(n_1, n_2)$



($\alpha=0.10$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	39.863	8.526	5.538	4.545	4.060	3.776	3.589	3.458	3.360	3.285
2	49.500	9.000	5.462	4.325	3.780	3.463	3.257	3.113	3.006	2.924
3	53.593	9.162	5.391	4.191	3.619	3.289	3.074	2.924	2.813	2.728
4	55.833	9.243	5.343	4.107	3.520	3.181	2.961	2.806	2.693	2.605
5	57.240	9.293	5.309	4.051	3.453	3.108	2.883	2.726	2.611	2.522
6	58.204	9.326	5.285	4.010	3.405	3.055	2.827	2.668	2.551	2.461
7	58.906	9.349	5.266	3.979	3.368	3.014	2.785	2.624	2.505	2.414
8	59.439	9.367	5.252	3.955	3.339	2.983	2.752	2.589	2.469	2.377
9	59.858	9.381	5.240	3.936	3.316	2.958	2.725	2.561	2.440	2.347
10	60.195	9.392	5.230	3.920	3.297	2.937	2.703	2.538	2.416	2.323
12	60.705	9.408	5.216	3.896	3.268	2.905	2.668	2.502	2.379	2.284
14	61.073	9.420	5.205	3.878	3.247	2.881	2.643	2.475	2.351	2.255
16	61.350	9.429	5.196	3.864	3.230	2.863	2.623	2.455	2.329	2.233
18	61.566	9.436	5.190	3.853	3.217	2.848	2.607	2.438	2.312	2.215
20	61.740	9.441	5.184	3.844	3.207	2.836	2.595	2.425	2.298	2.201
25	62.055	9.451	5.175	3.828	3.187	2.815	2.571	2.400	2.272	2.174
30	62.265	9.458	5.168	3.817	3.174	2.800	2.555	2.383	2.255	2.155
40	62.529	9.466	5.160	3.804	3.157	2.781	2.535	2.361	2.232	2.132
60	62.794	9.475	5.151	3.790	3.140	2.762	2.514	2.339	2.208	2.107
120	63.061	9.483	5.143	3.775	3.123	2.742	2.493	2.316	2.184	2.082

continued

($\alpha=0.10$)

$n_2 \backslash n_1$	12	14	16	18	20	25	30	40	60	120
1	3.177	3.102	3.048	3.007	2.975	2.918	2.881	2.835	2.791	2.748
2	2.807	2.726	2.668	2.624	2.589	2.528	2.489	2.440	2.393	2.347
3	2.606	2.522	2.462	2.416	2.380	2.317	2.276	2.226	2.177	2.130
4	2.480	2.395	2.333	2.286	2.249	2.184	2.142	2.091	2.041	1.992
5	2.394	2.307	2.244	2.196	2.158	2.092	2.049	1.997	1.946	1.896
6	2.331	2.243	2.178	2.130	2.091	2.024	1.980	1.927	1.875	1.824
7	2.283	2.193	2.128	2.079	2.040	1.971	1.927	1.873	1.819	1.767
8	2.245	2.154	2.088	2.038	1.999	1.929	1.884	1.829	1.775	1.722
9	2.214	2.122	2.055	2.005	1.965	1.895	1.849	1.793	1.738	1.684
10	2.188	2.095	2.028	1.977	1.937	1.866	1.819	1.763	1.707	1.652
12	2.147	2.054	1.985	1.933	1.892	1.820	1.773	1.715	1.657	1.601
14	2.117	2.022	1.953	1.900	1.859	1.785	1.737	1.678	1.619	1.562
16	2.094	1.998	1.928	1.875	1.833	1.758	1.709	1.649	1.589	1.530
18	2.075	1.978	1.908	1.854	1.811	1.736	1.686	1.625	1.564	1.504
20	2.060	1.962	1.891	1.837	1.794	1.718	1.667	1.605	1.543	1.482
25	2.031	1.933	1.860	1.805	1.761	1.683	1.632	1.568	1.504	1.440
30	2.011	1.912	1.839	1.783	1.738	1.659	1.606	1.541	1.476	1.409
40	1.986	1.885	1.811	1.754	1.708	1.627	1.573	1.506	1.437	1.368
60	1.960	1.857	1.782	1.723	1.677	1.593	1.538	1.467	1.395	1.320
120	1.932	1.828	1.751	1.691	1.643	1.557	1.499	1.425	1.348	1.265

($\alpha=0.05$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	161.448	18.513	10.128	7.709	6.608	5.987	5.591	5.318	5.117	4.965
2	199.500	19.000	9.552	6.944	5.786	5.143	4.737	4.459	4.256	4.103
3	215.707	19.164	9.277	6.591	5.409	4.757	4.347	4.066	3.863	3.708
4	224.583	19.247	9.117	6.388	5.192	4.534	4.120	3.838	3.633	3.478
5	230.162	19.296	9.013	6.256	5.050	4.387	3.972	3.687	3.482	3.326
6	233.986	19.330	8.941	6.163	4.950	4.284	3.866	3.581	3.374	3.217
7	236.768	19.353	8.887	6.094	4.876	4.207	3.787	3.500	3.293	3.135
8	238.883	19.371	8.845	6.041	4.818	4.147	3.726	3.438	3.230	3.072
9	240.543	19.385	8.812	5.999	4.772	4.099	3.677	3.388	3.179	3.020

continued

($\alpha=0.05$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
10	241.882	19.396	8.786	5.964	4.735	4.060	3.637	3.347	3.137	2.978
12	243.906	19.413	8.745	5.912	4.678	4.000	3.575	3.284	3.073	2.913
14	245.364	19.424	8.715	5.873	4.636	3.956	3.529	3.237	3.025	2.865
16	246.464	19.433	8.692	5.844	4.604	3.922	3.494	3.202	2.989	2.828
18	247.323	19.440	8.675	5.821	4.579	3.896	3.467	3.173	2.960	2.798
20	248.013	19.446	8.660	5.803	4.558	3.874	3.445	3.150	2.936	2.774
25	249.260	19.456	8.634	5.769	4.521	3.835	3.404	3.108	2.893	2.730
30	250.095	19.462	8.617	5.746	4.496	3.808	3.376	3.079	2.864	2.700
40	251.143	19.471	8.594	5.717	4.464	3.774	3.340	3.043	2.826	2.661
60	252.196	19.479	8.572	5.688	4.431	3.740	3.304	3.005	2.787	2.621
120	253.253	19.487	8.549	5.658	4.398	3.705	3.267	2.967	2.748	2.580

$n_2 \backslash n_1$	12	14	16	18	20	25	30	40	60	120
1	4.747	4.600	4.494	4.414	4.351	4.242	4.171	4.085	4.001	3.920
2	3.885	3.739	3.634	3.555	3.493	3.385	3.316	3.232	3.150	3.072
3	3.490	3.344	3.239	3.160	3.098	2.991	2.922	2.839	2.758	2.680
4	3.259	3.112	3.007	2.928	2.866	2.759	2.690	2.606	2.525	2.447
5	3.106	2.958	2.852	2.773	2.711	2.603	2.534	2.449	2.368	2.290
6	2.996	2.848	2.741	2.661	2.599	2.490	2.421	2.336	2.254	2.175
7	2.913	2.764	2.657	2.577	2.514	2.405	2.334	2.249	2.167	2.087
8	2.849	2.699	2.591	2.510	2.447	2.337	2.266	2.180	2.097	2.016
9	2.796	2.646	2.538	2.456	2.393	2.282	2.211	2.124	2.040	1.959
10	2.753	2.602	2.494	2.412	2.348	2.236	2.165	2.077	1.993	1.910
12	2.687	2.534	2.425	2.342	2.278	2.165	2.092	2.003	1.917	1.834
14	2.637	2.484	2.373	2.290	2.225	2.111	2.037	1.948	1.860	1.775
16	2.599	2.445	2.333	2.250	2.184	2.069	1.995	1.904	1.815	1.728
18	2.568	2.413	2.302	2.217	2.151	2.035	1.960	1.868	1.778	1.690
20	2.544	2.388	2.276	2.191	2.124	2.007	1.932	1.839	1.748	1.659
25	2.498	2.341	2.227	2.141	2.074	1.955	1.878	1.783	1.690	1.598
30	2.466	2.308	2.194	2.107	2.039	1.919	1.841	1.744	1.649	1.554
40	2.426	2.266	2.151	2.063	1.994	1.872	1.792	1.693	1.594	1.495
60	2.384	2.223	2.106	2.017	1.946	1.822	1.740	1.637	1.534	1.429
120	2.341	2.178	2.059	1.968	1.896	1.768	1.683	1.577	1.467	1.352



continued

($\alpha=0.025$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	647.789	38.506	17.443	12.218	10.007	8.813	8.073	7.571	7.209	6.937
2	799.500	39.000	16.044	10.649	8.434	7.260	6.542	6.059	5.715	5.456
3	864.163	39.165	15.439	9.979	7.764	6.599	5.890	5.416	5.078	4.826
4	899.583	39.248	15.101	9.605	7.388	6.227	5.523	5.053	4.718	4.468
5	921.848	39.298	14.885	9.364	7.146	5.988	5.285	4.817	4.484	4.236
6	937.111	39.331	14.735	9.197	6.978	5.820	5.119	4.652	4.320	4.072
7	948.217	39.355	14.624	9.074	6.853	5.695	4.995	4.529	4.197	3.950
8	956.656	39.373	14.540	8.980	6.757	5.600	4.899	4.433	4.102	3.855
9	963.285	39.387	14.473	8.905	6.681	5.523	4.823	4.357	4.026	3.779
10	968.627	39.398	14.419	8.844	6.619	5.461	4.761	4.295	3.964	3.717
12	976.708	39.415	14.337	8.751	6.525	5.366	4.666	4.200	3.868	3.621
14	982.528	39.427	14.277	8.684	6.456	5.297	4.596	4.130	3.798	3.550
16	986.919	39.435	14.232	8.633	6.403	5.244	4.543	4.076	3.744	3.496
18	990.349	39.442	14.196	8.592	6.362	5.202	4.501	4.034	3.701	3.453
20	993.103	39.448	14.167	8.560	6.329	5.168	4.467	3.999	3.667	3.419
25	998.081	39.458	14.115	8.501	6.268	5.107	4.405	3.937	3.604	3.355
30	1001.414	39.465	14.081	8.461	6.227	5.065	4.362	3.894	3.560	3.311
40	1005.598	39.473	14.037	8.411	6.175	5.012	4.309	3.840	3.505	3.255
60	1009.800	39.481	13.992	8.360	6.123	4.959	4.254	3.784	3.449	3.198
120	1014.020	39.490	13.947	8.309	6.069	4.904	4.199	3.728	3.392	3.140

$n_2 \backslash n_1$	12	14	16	18	20	25	30	40	60	120
1	6.554	6.298	6.115	5.978	5.871	5.686	5.568	5.424	5.286	5.152
2	5.096	4.857	4.687	4.560	4.461	4.291	4.182	4.051	3.925	3.805
3	4.474	4.242	4.077	3.954	3.859	3.694	3.589	3.463	3.343	3.227
4	4.121	3.892	3.729	3.608	3.515	3.353	3.250	3.126	3.008	2.894
5	3.891	3.663	3.502	3.382	3.289	3.129	3.026	2.904	2.786	2.674
6	3.728	3.501	3.341	3.221	3.128	2.969	2.867	2.744	2.627	2.515
7	3.607	3.380	3.219	3.100	3.007	2.848	2.746	2.624	2.507	2.395
8	3.512	3.285	3.125	3.005	2.913	2.753	2.651	2.529	2.412	2.299
9	3.436	3.209	3.049	2.929	2.837	2.677	2.575	2.452	2.334	2.222
10	3.374	3.147	2.986	2.866	2.774	2.613	2.511	2.388	2.270	2.157

continued

($\alpha=0.025$)

$n_2 \backslash n_1$	12	14	16	18	20	25	30	40	60	120
12	3.277	3.050	2.889	2.769	2.676	2.515	2.412	2.288	2.169	2.055
14	3.206	2.979	2.817	2.696	2.603	2.441	2.338	2.213	2.093	1.977
16	3.152	2.923	2.761	2.640	2.547	2.384	2.280	2.154	2.033	1.916
18	3.108	2.879	2.717	2.596	2.501	2.338	2.233	2.107	1.985	1.866
20	3.073	2.844	2.681	2.559	2.464	2.300	2.195	2.068	1.944	1.825
25	3.008	2.778	2.614	2.491	2.396	2.230	2.124	1.994	1.869	1.746
30	2.963	2.732	2.568	2.445	2.349	2.182	2.074	1.943	1.815	1.690
40	2.906	2.674	2.509	2.384	2.287	2.118	2.009	1.875	1.744	1.614
60	2.848	2.614	2.447	2.321	2.223	2.052	1.940	1.803	1.667	1.530
120	2.787	2.552	2.383	2.256	2.156	1.981	1.866	1.724	1.581	1.433

($\alpha=0.01$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	4052.181	98.503	34.116	21.198	16.258	13.745	12.246	11.259	10.561	10.044
2	4999.500	99.000	30.817	18.000	13.274	10.925	9.547	8.649	8.022	7.559
3	5403.352	99.166	29.457	16.694	12.060	9.780	8.451	7.591	6.992	6.552
4	5624.583	99.249	28.710	15.977	11.392	9.148	7.847	7.006	6.422	5.994
5	5763.650	99.299	28.237	15.522	10.967	8.746	7.460	6.632	6.057	5.636
6	5858.986	99.333	27.911	15.207	10.672	8.466	7.191	6.371	5.802	5.386
7	5928.356	99.356	27.672	14.976	10.456	8.260	6.993	6.178	5.613	5.200
8	5981.070	99.374	27.489	14.799	10.289	8.102	6.840	6.029	5.467	5.057
9	6022.473	99.388	27.345	14.659	10.158	7.976	6.719	5.911	5.351	4.942
10	6055.847	99.399	27.229	14.546	10.051	7.874	6.620	5.814	5.257	4.849
12	6106.321	99.416	27.052	14.374	9.888	7.718	6.469	5.667	5.111	4.706
14	6142.674	99.428	26.924	14.249	9.770	7.605	6.359	5.559	5.005	4.601
16	6170.101	99.437	26.827	14.154	9.680	7.519	6.275	5.477	4.924	4.520
18	6191.529	99.444	26.751	14.080	9.610	7.451	6.209	5.412	4.860	4.457
20	6208.730	99.449	26.690	14.020	9.553	7.396	6.155	5.359	4.808	4.405
25	6239.825	99.459	26.579	13.911	9.449	7.296	6.058	5.263	4.713	4.311
30	6260.649	99.466	26.505	13.838	9.379	7.229	5.992	5.198	4.649	4.247
40	6286.782	99.474	26.411	13.745	9.291	7.143	5.908	5.116	4.567	4.165
60	6313.030	99.482	26.316	13.652	9.202	7.057	5.824	5.032	4.483	4.082
120	6339.391	99.491	26.221	13.558	9.112	6.969	5.737	4.946	4.398	3.996

continued

($\alpha=0.01$)

$n_2 \backslash n_1$	12	14	16	18	20	25	30	40	60	120
1	9.330	8.862	8.531	8.285	8.096	7.770	7.562	7.314	7.077	6.851
2	6.927	6.515	6.226	6.013	5.849	5.568	5.390	5.179	4.977	4.787
3	5.953	5.564	5.292	5.092	4.938	4.675	4.510	4.313	4.126	3.949
4	5.412	5.035	4.773	4.579	4.431	4.177	4.018	3.828	3.649	3.480
5	5.064	4.695	4.437	4.248	4.103	3.855	3.699	3.514	3.339	3.174
6	4.821	4.456	4.202	4.015	3.871	3.627	3.473	3.291	3.119	2.956
7	4.640	4.278	4.026	3.841	3.699	3.457	3.304	3.124	2.953	2.792
8	4.499	4.140	3.890	3.705	3.564	3.324	3.173	2.993	2.823	2.663
9	4.388	4.030	3.780	3.597	3.457	3.217	3.067	2.888	2.718	2.559
10	4.296	3.939	3.691	3.508	3.368	3.129	2.979	2.801	2.632	2.472
12	4.155	3.800	3.553	3.371	3.231	2.993	2.843	2.665	2.496	2.336
14	4.052	3.698	3.451	3.269	3.130	2.892	2.742	2.563	2.394	2.234
16	3.972	3.619	3.372	3.190	3.051	2.813	2.663	2.484	2.315	2.154
18	3.909	3.556	3.310	3.128	2.989	2.751	2.600	2.421	2.251	2.089
20	3.858	3.505	3.259	3.077	2.938	2.699	2.549	2.369	2.198	2.035
25	3.765	3.412	3.165	2.983	2.843	2.604	2.453	2.271	2.098	1.932
30	3.701	3.348	3.101	2.919	2.778	2.538	2.386	2.203	2.028	1.860
40	3.619	3.266	3.018	2.835	2.695	2.453	2.299	2.114	1.936	1.763
60	3.535	3.181	2.933	2.749	2.608	2.364	2.208	2.019	1.836	1.656
120	3.449	3.094	2.845	2.660	2.517	2.270	2.111	1.917	1.726	1.533

($\alpha=0.005$)

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	16210.723	198.501	55.552	31.333	22.785	18.635	16.236	14.688	13.614	12.826
2	19999.500	199	49.799	26.284	18.314	14.544	12.404	11.042	10.107	9.427
3	21614.741	199.166	47.467	24.259	16.530	12.917	10.882	9.596	8.717	8.081
4	22499.583	199.250	46.195	23.155	15.556	12.028	10.050	8.805	7.956	7.343
5	23055.798	199.300	45.392	22.456	14.940	11.464	9.522	8.302	7.471	6.872
6	23437.111	199.333	44.838	21.975	14.513	11.073	9.155	7.952	7.134	6.545
7	23714.566	199.357	44.434	21.622	14.200	10.786	8.885	7.694	6.885	6.302
8	23925.406	199.375	44.126	21.352	13.961	10.566	8.678	7.496	6.693	6.116
9	24091.004	199.388	43.882	21.139	13.772	10.391	8.514	7.339	6.541	5.968
10	24224.487	199.400	43.686	20.967	13.618	10.250	8.380	7.211	6.417	5.847

continued

($\alpha=0.005$)

n_2 n_1	1	2	3	4	5	6	7	8	9	10
12	24426.366	199.416	43.387	20.705	13.384	10.034	8.176	7.015	6.227	5.661
14	24571.767	199.428	43.172	20.515	13.215	9.877	8.028	6.872	6.089	5.526
16	24681.467	199.437	43.008	20.371	13.086	9.758	7.915	6.763	5.983	5.422
18	24767.170	199.444	42.880	20.258	12.985	9.664	7.826	6.678	5.899	5.340
20	24835.971	199.450	42.778	20.167	12.903	9.589	7.754	6.608	5.832	5.274
25	24960.340	199.460	42.591	20.002	12.755	9.451	7.623	6.482	5.708	5.153
30	25043.628	199.466	42.466	19.892	12.656	9.358	7.534	6.396	5.625	5.071
40	25148.153	199.475	42.308	19.752	12.530	9.241	7.422	6.288	5.519	4.966
60	25253.137	199.483	42.149	19.611	12.402	9.122	7.309	6.177	5.410	4.859
120	25358.573	199.491	41.989	19.468	12.274	9.001	7.193	6.065	5.300	4.750

n_2 n_1	12	14	16	18	20	25	30	40	60	120
1	11.754	11.060	10.575	10.218	9.944	9.475	9.180	8.828	8.495	8.179
2	8.510	7.922	7.514	7.215	6.986	6.598	6.355	6.066	5.795	5.539
3	7.226	6.680	6.303	6.028	5.818	5.462	5.239	4.976	4.729	4.497
4	6.521	5.998	5.638	5.375	5.174	4.835	4.623	4.374	4.140	3.921
5	6.071	5.562	5.212	4.956	4.762	4.433	4.228	3.986	3.760	3.548
6	5.757	5.257	4.913	4.663	4.472	4.150	3.949	3.713	3.492	3.285
7	5.525	5.031	4.692	4.445	4.257	3.939	3.742	3.509	3.291	3.087
8	5.345	4.857	4.521	4.276	4.090	3.776	3.580	3.350	3.134	2.933
9	5.202	4.717	4.384	4.141	3.956	3.645	3.450	3.222	3.008	2.808
10	5.085	4.603	4.272	4.030	3.847	3.537	3.344	3.117	2.904	2.705
12	4.906	4.428	4.099	3.860	3.678	3.370	3.179	2.953	2.742	2.544
14	4.775	4.299	3.972	3.734	3.553	3.247	3.056	2.831	2.620	2.423
16	4.674	4.200	3.875	3.637	3.457	3.151	2.961	2.737	2.526	2.328
18	4.595	4.122	3.797	3.560	3.380	3.075	2.885	2.661	2.450	2.251
20	4.530	4.059	3.734	3.498	3.318	3.013	2.823	2.598	2.387	2.188
25	4.412	3.942	3.618	3.382	3.203	2.898	2.708	2.482	2.270	2.069
30	4.331	3.862	3.539	3.303	3.123	2.819	2.628	2.401	2.187	1.984
40	4.228	3.760	3.437	3.201	3.022	2.716	2.524	2.296	2.079	1.871
60	4.123	3.655	3.332	3.096	2.916	2.609	2.415	2.184	1.962	1.747
120	4.015	3.547	3.224	2.987	2.806	2.496	2.300	2.064	1.834	1.606

Appendix G Some Common Probability Distributions

Distribution	Parameter	PF/PDF	Mean	Variance
Bernoulli	$0 < p < 1$	$P(X=1) = p, P(X=0) = 1-p$	p	$p(1-p)$
Binomial	$n \geq 1, 0 < p < 1$	$P(X=x) = C_n^x p^x (1-p)^{n-x}, x=0,1,\dots,n$	np	$np(1-p)$
Poisson	$\lambda > 0$	$P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}, x=0,1,2,\dots$	λ	λ
Hypergeometric	N, m, n	$P(X=x) = \frac{C_m^x C_{N-m}^{n-x}}{C_N^n},$ $\max\{0, n-N+m\} \leq x \leq \min\{n, m\}$	$\frac{nm}{N}$	$\frac{nm}{N} \left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right)$
Negative binomial	$r \geq 1, 0 < p < 1$	$P(X=x) = C_{x-1}^{r-1} p^r (1-p)^{x-r}, x=r, r+1, \dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Geometric	$0 < p < 1$	$P(X=x) = p(1-p)^{x-1}, x=1,2,\dots$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Uniform	$a < b$	$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

continued

Distribution	Parameter	PF/PDF	Mean	Variance
Normal	$-\infty < \mu < +\infty, \sigma > 0$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$	μ	σ^2
Gamma	$\alpha > 0, \beta > 0$	$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\alpha\beta$	$\alpha\beta^2$
Erlang	$n \geq 1, \beta > 0$	$f(x) = \begin{cases} \frac{1}{\beta^n (n-1)!} x^{n-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$n\beta$	$n\beta^2$
χ^2	$n \geq 1$	$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	n	$2n$
Beta	$\alpha > 0, \beta > 0$	$f(x) = \begin{cases} \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$
Lognormal	$-\infty < \mu < +\infty, \sigma > 0$	$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$
Cauchy	$-\infty < \mu < +\infty, \sigma > 0$	$f(x) = \frac{1}{\pi\sigma} \frac{1}{1 + \left[\frac{(x-\mu)}{\sigma}\right]^2}, -\infty < x < +\infty$	NA	NA
t	$n \geq 1$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < x < +\infty$	$0, n > 1$	$\frac{n}{n-2}, n > 2$
F	$m, n \geq 1$	$f(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$	$\frac{n}{n-2}, n > 2$	$\frac{n^2 (2m + 2n - 4)}{m(n-2)^2 (n-4)}, n > 4$

Bibliography

- [1] 北京交通大学概率统计课程组. 概率论与数理统计. 北京:科学出版社,2010.
- [2] 茆诗松,程依明,濮晓龙. 概率论与数理统计教程. 2 版. 北京:高等教育出版社,2011.
- [3] 盛骤,谢式千,潘承毅. 概率论与数理统计. 4 版. 北京:高等教育出版社,2008.
- [4] 陈希孺. 概率论与数理统计. 北京:科学出版社,2000.
- [5] MORRIS H D, MARK J S. Probability and statistics. 4th ed. Addison Wesley, 2010.

Images have been losslessly embedded. Information about the original file can be found in PDF attachments. Some stats (more in the PDF attachments):

```
{
  "filename": "MTQ2OTQ5NzEuemlw",
  "filename_decoded": "14694971.zip",
  "filesize": 34632158,
  "md5": "0d6980f2234321ec243aef6f5d84c8ad",
  "header_md5": "59f0ec8bab528c4698ccb7bd22e719bd",
  "sha1": "f7111dbe8acd2e68b75b777b55806746a6df2c90",
  "sha256": "b2a3d18fa8c06869c40509c55988a996833ab0d1589405bbd151e85198915e76",
  "crc32": 2883234231,
  "zip_password": "",
  "uncompressed_size": 45202223,
  "pdg_dir_name": "14694971",
  "pdg_main_pages_found": 270,
  "pdg_main_pages_max": 270,
  "total_pages": 278,
  "total_pixels": 1689136144,
  "pdf_generation_missing_pages": false
}
```