面向21世纪课程教材
Textbook Series for 21st Century

高等学校经济管理学科数学基础

主编　范培华　胡显佑

概率论与数理统计

第二版

龙永红　主编

高等教育出版社
HIGHER EDUCATION PRESS

面向21世纪课程教材
Textbook Series for 21st Century

高等学校经济管理学科数学基础

主编　范培华　胡显佑

概率论与数理统计

第二版

龙永红　主编



高等教育出版社
HIGHER EDUCATION PRESS

内容简介

本书第一版是教育部“高等教育面向21世纪教学内容和课程体系改革计划”的研究成果，是面向21世纪的高等院校经济学学科门类和管理学学科门类的数学基础课教材之一。第二版是在高等教育出版社“高等教育百门精品课程教材建设计划”支持下对第一版的修订。

全书以经济与管理类学生易于接受的方式介绍了随机事件与概率、随机变量的分布与数字特征、随机向量、参数估计与假设检验、方差分析以及回归分析等概率论与数理统计的基本内容。与以往教材相比，其主要特点在于：强调概念和内容的直观引入及知识间的联系；强调随机性思维的培养和应用能力的培养；强调有关概念、方法与经济管理学科的联系，并适应现代经济、金融与管理学的发展的需要。

书中每章配有A，B两组习题和参考答案，其中B组习题是为满足读者的较高要求而配备的。本书可作为经济管理类本科教材，也适合于考研学生备考之用。

图书在版编目(CIP)数据

概率论与数理统计/龙永红主编. —2版. —北京：高等教育出版社，2004.4
（高等学校经济管理学科数学基础/范培华，胡显佑主编）
ISBN 7-04-013787-9

Ⅰ.概… Ⅱ.龙… Ⅲ.①概率论-高等学校-教材②数理统计-高等学校-教材 Ⅳ.O21

中国版本图书馆CIP数据核字(2004)第003224号

出版发行	高等教育出版社	购书热线	010-64054588
社　　址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010-82028899		http://www.hep.com.cn
经　　销	新华书店北京发行所		
印　　刷	北京外文印刷厂		
		版　　次	2001年5月第1版
开　　本	787×960　1/16		2004年4月第2版
印　　张	18.5	印　　次	2004年8月第2次印刷
字　　数	350 000	定　　价	19.60元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

版权所有　侵权必究

第二版前言

这套教材从第一版发行至今，已被多所高等院校选为经济管理类专业的数学基础课程的教材。不少同仁来信表示鼓励，指出教材中值得探讨的问题，并对如何修改提出了宝贵的意见。在此我们对关心和支持这套教材的广大同仁表示衷心的感谢。

2002 年 12 月高等教育出版社启动了“高等教育百门精品课程教材建设计划”，这套教材的建设被纳入计划之中。在该项目的支持下，我们对这套教材的第一版进行了修订。修订工作主要包括以下几个方面的内容：

1. 订正了原教材中的疏漏以及排版印刷方面的错误；

2. 调整了一部分例题和习题，使其与相应的内容之间搭配得更加合理；

3. 调整了一些命题的条件或结论，使其阐释得更加精确；

4.《概率论与数理统计》中数理统计部分做了较大的调整和修改，以适合经济管理类专业的需要。

参加这套教材修订工作的主要为原编写人员。此外，需要说明的是，《概率论与数理统计》中数理统计部分由中国人民大学刘刚同志、北京大学陈奇志同志修改。

根据我国高等教育从精英教育向大众化教育转变以及现代教育技术手段在教学中广泛运用的现状，我们对这套教材进行了立体化教学设计，各主教材分别配备了典型例题分析与习题（与主教材同期推出），配套的电子教案也将于近期出版。希望能更好地满足高校教师课堂教学和学生自主学习的需要，对提高教学质量起到辅助作用。

在修订过程中，我们广泛地搜集了读者对原教材的意见和建议。希望通过此次修订，这套教材能在第一版的基础上更加完善。欢迎大家继续批评和指正。

编者

2003 年 9 月

修订说明

本书第一版作为教育部“高等教育面向21世纪教学内容和课程体系改革计划”的研究成果，其主要特点在于：强调概念和内容的直观引入及知识间的联系以加强学生对概率论与数理统计的理论和方法的理解；强调随机性思维方式和经济管理专业学生应用概率统计的能力的培养，避免成为数学系教材的简写本；强调有关概念、方法与经济管理学科的联系，并适应现代经济、金融与管理学的发展的需要。第一版自出版以来受到广泛的关注，为使其特点更加突出、内容与体系更加完善，在高等教育出版社“高等教育百门精品课程教材建设计划”的支持下，我们对第一版实施了本次修订。

本次修订由中国人民大学数学系龙永红教授提出整体方案，并负责第1～3章的修订；成世学教授负责第4章的修订；刘刚老师负责原书第5～7章(修订后的第5、6章)的修订；北京大学光华管理学院陈奇志博士负责原书第8章(修订后的第7章)的修订；取消原书第9章的内容。全书由龙永红教授统纂定稿。具体修订内容如下：

1. 概率论部分(第1～3章)主要订正原书中的一些疏漏和印刷错误，调整了个别习题。

2. 将第4章不作教学要求的内容(主要是一些常用统计量密度函数的推导)列入附录中，使教学要求更加突出，同时对部分习题作了调整。

3. 考虑到参数估计与假设检验间的联系，将原书第5、6章合并为新版的第5章，将双正态总体的置信区间在正文中删去，只在附录的一览表中列出。

4. 对参数估计部分的例题和习题作了较大的调整，对假设检验部分进行了改写，使内容更加简练，重点更加突出。

5. 对方差分析和回归分析的内容重新作了处理，略去了一些繁琐的推导细节，增加统计软件SPSS的输出结果实例，突出思想、方法和应用。此外，增加了一些经济学实例。

6. 考虑到教学学时的限制，取消了原书中主成分分析与典型相关分析的内容。

本次修订吸取了部分使用本教材第一版教学的教师的意见，我们在此表示衷心的感谢，也期待广大读者进一步提出批评和建议。

龙永红

2003年10月1日

第一版前言

1996年原国家教委开始组织实施“高等教育面向21世纪教学内容和课程体系改革计划”,其中子项目经济学门类数学基础课研究和管理学门类数学基础课研究分别由中国人民大学和北京大学承担。考虑到这两大学科门类数学基础课程的共同点,教育部又将这两个子项目整合为“经济管理学类专业数学基础课程设置与教学内容改革研究”,集中力量合作研究,并成立了以魏权龄教授和范培华教授为项目主持人的课题组。两年多来,课题组对国内外高等院校同类专业数学基础课程的现状进行了调查研究,编写了教学大纲,组织了多次有关课程体系、课程内容的研讨会。其中,于1997年7月在长春召开的中国数量经济学会年会上,全国40余所院校的教师就经济管理类专业的数学基础课、数量经济分析课程的体系、课程设置、内容等进行了深入的讨论;1998年4月,教育部在京召开了管理类专业面向21世纪教学内容和课程体系改革的研讨会上,初步确定了数学基础课应包括微积分、线性代数和概率统计三门课程,共16学分。其中“微积分”8学分,“线性代数”3学分,“概率统计”5学分。

在调查研究和充分讨论的基础上,课题组拟定了《经济管理学科数学基础教学大纲》(草案),并邀请北京地区部分高校就该大纲进行了讨论。

受教育部委托,北京大学光华管理学院和中国人民大学信息学院共同承担了编写经济管理学科数学基础系列教材的任务。整套教材分为《微积分》、《线性代数》和《概率论与数理统计》3个分册,由魏权龄教授任编写组顾问,范培华教授、胡显佑教授任主编。这套教材的《微积分》分册由朱来义教授主编,参加编写的有朱来义、吴岚、范培华和严守权;《线性代数》分册由卢刚副教授主编,参加编写的有卢刚、胡显佑、崔兆鸣;《概率论与数理统计》分册由龙永红副教授主编,参加编写的有龙永红、张贻兰、成世学、王明进。

根据高等教育面向21世纪教学内容和课程体系改革总体目标的要求,我们在编写这套教材时,主要考虑了下述问题:

1. 为适应我国在21世纪社会主义建设和经济发展的需要,培养“厚基础、宽口径、高素质”的人才,基础课,特别是数学基础课不应削弱,而应适当加强。

2. 考虑到目前绝大多数综合性大学、工科院校都设立了经济或管理学科的有关专业,但各校、各专业方向对数学基础的要求有一定的差异,这套教材应照顾到多数院校教学的实际情况,便于教师和学生使用。

3. 作为一门数学基础课的教材,我们首先注意保持数学学科本身的科学

性、系统性，但在引入一些概念时尽可能采用学生易于接受的方式叙述，对个别冗长、繁琐的推理则略去，而更突出有关理论、方法的应用和经济数学模型的介绍。

4. 作为经济管理学科各专业的数学基础教材，我们注意了专业后继课程的需要，并考虑学生继续深造的需要，教材的各章均配备了A,B两组习题。一般，达到A组习题的水平，就已经符合本课程的基本要求。B组习题是为数学基础要求较高的专业或学生准备的。各章中打有“*”号(或小字排版)的内容是为对数学基础要求较高的院校或专业编写的，可以作为选学内容或学生自学用。

1999年12月，由教育部高教司聘请了有关专家对教材的初稿进行了审定。参加审稿会的有：北京航空航天大学李心灿教授、清华大学胡金德教授、南开大学周概容教授、(以下以姓氏笔画为序)湖南财经学院苏醒教授、北方交通大学季文铎教授、中央财经大学单立波教授、华侨大学龚德恩教授、中南财经大学彭勇行教授。他们对教材初稿提出了许多中肯的建议和具体的修改意见，这对于完善教材是非常有益的，在此向参加审定会的各位教授表示诚挚的谢意。

在各次研讨会上，全国各高校的许多同行都对这一项目和教材提出了极有价值的建议。在此向有关院校的老师们表示衷心感谢。在教材编写过程中，我们得到了教育部高教司的大力支持，得到高等教育出版社有关部门的协助，在此一并致谢。

范培华　胡显佑

2000年3月

目　录

第 1 章　随机事件与概率 …… 1
§1.1　随机事件 …… 1
§1.2　随机事件的概率 …… 8
§1.3　古典概型与几何概型 …… 12
§1.4　条件概率 …… 17
§1.5　事件的独立性 …… 24
习题一 …… 29
第 2 章　随机变量的分布与数字特征 …… 34
§2.1　随机变量及其分布 …… 34
§2.2　随机变量的数字特征 …… 42
§2.3　常用的离散型分布 …… 51
§2.4　常用的连续型分布 …… 57
§2.5　随机变量函数的分布 …… 63
习题二 …… 67
第 3 章　随机向量 …… 72
§3.1　随机向量的分布 …… 72
§3.2　条件分布与随机变量的独立性 …… 80
§3.3　随机向量的函数的分布与数学期望 …… 89
§3.4　随机向量的数字特征 …… 97
§3.5　大数定律与中心极限定理 …… 106
习题三 …… 112
第 4 章　数理统计的基础知识 …… 119
§4.1　总体与样本 …… 119
§4.2　统计量 …… 123
§4.3　常用的统计分布 …… 125
§4.4　抽样分布 …… 132
附录 4.1　三个常用统计分布密度函数的推导 …… 137
附录 4.2　正态总体抽样分布的基础性定理 …… 139
附录 4.3　一般总体枢轴量 T_n 的极限分布 …… 141
习题四 …… 142
第 5 章　参数估计与假设检验 …… 145
§5.1　点估计概述 …… 145
§5.2　参数的最大似然估计与矩估计 …… 149

§5.3 置信区间 …… 155
§5.4 假设检验概述 …… 163
§5.5 单正态总体的参数假设检验 …… 169
§5.6 双正态总体的参数假设检验 …… 173
§5.7 一般总体的参数假设检验 …… 177
§5.8 拟合优度 χ^2 检验与独立性检验 …… 182
附录 5.1 正态总体参数的置信区间一览表 …… 187
附录 5.2 单正态总体参数假设检验一览表 …… 188
附录 5.3 双正态总体参数假设检验一览表 …… 189
习题五 …… 192
第 6 章 方差分析 …… 196
§6.1 方差分析概述 …… 196
§6.2 单因素方差分析 …… 198
§6.3 双因素方差分析 …… 203
习题六 …… 211
第 7 章 回归分析 …… 215
§7.1 一元线性回归模型及其参数估计 …… 215
§7.2 一元线性回归模型的检验 …… 222
§7.3 一元线性回归的残差分析 …… 229
§7.4 一元线性回归的预测和控制 …… 231
§7.5 一元非线性问题的线性化 …… 236
§7.6 多元线性回归分析 …… 241
附录 7.1 命题 7.1 的证明 …… 248
习题七 …… 249
习题参考答案 …… 252
常用统计分布表 …… 267
附表 1 泊松分布概率值表 …… 267
附表 2 标准正态分布函数值表 …… 270
附表 3 χ^2 分布上侧分位数表 …… 273
附表 4 F 分布上侧分位数表 …… 276
附表 5 t 分布上侧分位数表 …… 286
主要参考文献 …… 287

第 1 章

随机事件与概率

概率论是研究随机现象的规律性的数学学科.为了对随机现象的有关问题作出明确的数学阐述,像其他数学学科一样,概率论具有自己的严格的概念体系和严密的逻辑结构.本章重点介绍概率论的两个最基本的概念:随机事件及其概率.主要内容包括:随机事件和随机事件的概率的定义、古典概型与几何概型、条件概率、乘法公式、全概率公式与贝叶斯公式以及事件的独立性等.这些内容是进一步学习概率论的基础.

§1.1 随机事件

一、随机现象

在自然界和人类社会生活中普遍存在着两类现象,一类是在一定条件下必然出现的现象,称为**确定性现象**.例如,一物体从高度为 h(米)处垂直下落,则必然在$\sqrt{\frac{2h}{g}}$秒后落到地面,其中 $g=9.8$(米/秒^2)为重力加速度.我们已学过的数学学科均以确定性现象为研究对象,并为研究其量的规律性提供工具和方法.另一类则是我们事先无法准确预知其结果的现象,称为**随机现象**.例如,投掷一枚硬币,我们不能事先预知将出现正面还是反面.在实际中,我们经常要面对和处理随机现象,比如,明天是否会下雨?某种股票明天价格是多少?电视机价格是否会在近期下调?这些问题往往事先不能得到明确的答案,却与我们的切身利益密切相关.概率论以随机现象为研究对象.

二、随机现象的统计规律性

由于随机现象的结果事先不能预知,初看起来,随机现象毫无规律可言.然而人们发现同一随机现象在大量重复出现时,其每种可能的结果出现的频率却具有稳定性,从而表明随机现象也有其固有的量的规律性.人们把随机现象在大量重复出现时所表现出来的量的规律性称为随机现象的**统计规律性**.

为了对随机现象的统计规律性进行研究，人们往往要对随机现象进行观察，我们把对随机现象的观察称为**随机试验**，简称为**试验**. 例如，某射手对固定目标进行射击，观察其是否射中；观察某地区夏季暴雨次数；观察某电话交换台每日收到的呼叫次数等均为随机试验. 一般地，一个随机试验要求满足下列特点：

(1) 可重复性　试验原则上可在相同条件下重复进行；

(2) 可观察性　试验结果是可观察的，所有可能的结果是明确的；

(3) 随机性　每次试验将要出现的结果是不确定的，事先无法准确预知.

历史上，研究随机现象统计规律性的最著名的试验是投掷硬币的试验. 我们知道，投掷一枚均匀硬币时，事先无法准确预知将出现正面还是反面. 但是，当人们重复投掷上千次时，却发现出现正面和反面的次数大致相等，即各自占总试验次数的比例(即频率)大致等于 0.5，而且随着试验次数的增加，这一比例会更加稳定地靠近 0.5，表 1.1 列出了历史上一些试验的记录.

表 1.1　历史上投掷硬币试验的记录

试验者	投掷次数(n)	正面次数(r_n)	正面频率$\left(\frac{r_n}{n}\right)$
De Morgan	2 048	1 061	0.518 1
Buffon	4 040	2 048	0.506 9
Pearson K	12 000	6 019	0.501 6
Pearson K	24 000	12 012	0.500 5

三、样本空间

正如前面指出的，一个随机试验将要出现的结果是不确定的，但其所有可能结果是明确的. 我们把随机试验的每一个可能结果称为一个**样本点**，因而一个随机试验的所有样本点也是明确的，它们的全体，称为**样本空间**，通常用 Ω 表示. Ω 中的点，即样本点，用 ω 表示.

例 1.1　在投掷一枚硬币观察其出现正面还是反面的试验中，有两个样本点：正面、反面. 样本空间为

$$\Omega=\{\text{正面},\text{反面}\},$$

记 $\omega_1=$“正面”，$\omega_2=$“反面”，则样本空间可表示为

$$\Omega=\{\omega_1,\omega_2\}.$$

例 1.2　在投掷一枚骰子，观察其出现的点数的试验中，有 6 个样本点：1 点，2 点，…，6 点. 样本空间为

$$\Omega=\{1\text{ 点},2\text{ 点},\cdots,6\text{ 点}\},$$

或干脆将样本点分别简记为 1，2，…，6，相应地，样本空间记为

$$\Omega=\{1,2,\cdots,6\}.$$

例 1.3 观察某电话交换台在一天内收到的呼叫次数，其样本点有可数无穷多个：i 次，$i=0,1,2,\cdots$，样本空间为

$$\Omega=\{0\text{ 次},1\text{ 次},2\text{ 次},\cdots\},$$

或简记为 $\Omega=\{0,1,2,\cdots\}$.

例 1.4 观察一个新灯泡的寿命，其样本点也有无穷多个(且不可数!)：t 小时，$0\leqslant t<+\infty$，样本空间为

$$\Omega=\{t\text{ 小时}\mid 0\leqslant t<+\infty\},$$

或简记为

$$\Omega=\{t\mid 0\leqslant t<+\infty\}=[0,+\infty).$$

四、随机事件

在随机试验中，人们除了关心试验的结果本身外，往往还关心试验的结果是否具备某一指定的可观察的特征，概率论中将这一可观察的特征称为一个**事件**. 例如，投掷一枚骰子，我们也许会关心出现的点数是否为偶数，“点数为偶数”就是一个事件. 同样，“点数小于 7”也是一个事件. 但这两个事件有着根本的区别，前者在随机试验中可能发生也可能不发生，这样的事件称为**随机事件**. 后者在试验中是必然发生的，这样的事件称为**必然事件**. 与必然事件完全对立的是，在试验中一定不发生的事件，称为**不可能事件**. 比如在上述试验中“点数不小于 7”是不可能事件. 虽然必然事件与不可能事件是完全对立的，但它们有一个共同的特点，那就是在试验之前我们能够准确预知其是否发生，因而均不是随机事件，通常称之为**确定性事件**. 概率论研究的是随机事件，但为方便起见常常将必然事件和不可能事件视为随机事件的极端情形，并将随机事件简称为**事件**，通常记作 $A,B,\cdots$.

例 1.5 在投掷一枚骰子的试验中，分别记

“点数是 6”为 A，

“点数小于 5”为 B，

“点数小于 5 的偶数”为 C，

则 A,B,C 均为事件，但事件 A 的结构最为简单，它对应于一个惟一的可能结果，即样本点，这样的事件称为**基本事件**①. 在本例中，共有 6 个基本事件(对应于 6 个样本点)：“点数为 1”，“点数为 2”，$\cdots$，“点数为 6”. 基本事件的称谓缘于相对其他事件而言，它们是最基本的，其他事件均可由它们复合而成，而它们自身又不能再分解成其他事件. 事件 B 和 C 均不是基本事件，它们分别可以由一些基本事件复合而成. 比如事件 C 可由“点数为 2”和“点数为 4”两个基本事件复合而成.

① 在一般的讨论中，基本事件并不一定是事件，在本书的讨论中，我们均假设基本事件是事件

五、事件的集合表示

根据定义,样本空间 Ω 是随机试验的所有可能结果——即样本点 ω 的全体,因而样本空间实际上是所有样本点构成的集合,相应的每一样本点是该集合中的元素.而一个事件是由具有该事件所要求的特征的那些可能结果所构成,所以一个事件对应于 Ω 中具有相应特征的样本点(元素)构成的集合,它是 Ω 的一个子集.于是任何一个事件,可以用 Ω 的某一子集来表示,通常用符号 A, B,…来记.某事件发生,就是属于该集合的某一样本点在试验中出现.如果记 ω 为试验中出现的样本点,那么事件 A 当且仅当 $\omega \in A$ 时发生.

例 1.6 在例 1.5 中,样本空间为 $\Omega=\{1,2,3,4,5,6\}$,事件 B 和 C 则可分别表示为

$$B=\{1,2,3,4\},$$
$$C=\{2,4\}.$$

由于样本空间 Ω 包含所有可能结果,试验结果必是其中之一,所以样本空间作为一个事件是必然发生的,即为必然事件,今后用 Ω 表示必然事件.空集 $\varnothing$ 作为 Ω 的子集不含有任何样本点,不管试验的结果是什么,$\varnothing$作为一个事件总不会发生,因而是不可能事件.今后用 $\varnothing$来表示不可能事件.

六、事件间的关系与运算

在一个随机试验中,一般有很多随机事件,为了通过对简单事件的研究来掌握复杂事件,我们需要研究事件之间的关系和事件之间的一些运算.前面引进的事件的集合表示为这一任务提供极大的便利.

1. 事件的包含

如果事件 A 发生必然导致 B 发生,即属于 A 的每一个样本点一定也属于 B,则称**事件 B 包含事件 A**,或称事件 A **包含于事件 B**,或称 A 是 B 的**子事件**,记作

$$B \supset A \text{ 或 } A \subset B.$$

显然,事件 $A \subset B$ 的含义与集合论中的含义是一致的.对任意事件 A,易知

$$\varnothing \subset A \subset \Omega.$$

2. 事件的相等

如果事件 A 包含事件 B,事件 B 也包含事件 A,则称事件 A 与 B **相等**(或**等价**),记作 $A=B$.易见,相等的两个事件总是同时发生或同时不发生.更直接的表述是:$A=B$ 是指 A 与 B 所含的样本点完全相同,这等同于集合的相等.

3. 事件的并(或和)

"事件 A 与 B 至少有一个发生"这一事件称作事件 A 与 B 的**并**(或**和**),记

作 $A \cup B$ 或 $A+B$. 显然,事件 $A \cup B$ 是由 A 和 B 的样本点共同构成的事件,这与集合的并集的含义是一致的.

例 1.7 在投掷一枚骰子的试验中,记

$$A = \text{“点数为奇数”},$$

$$B = \text{“点数小于 5”},$$

则 $A \cup B = \{1,2,3,4,5\}$.

4. 事件的交(或积)

“事件 A 和 B 都发生”这一事件称为事件 A 与 B 的**交(或积)**,记作 $A \cap B$ (或 AB). 显然,$A \cap B$ 实际上是由 A 和 B 的公共样本点所构成,这与集合的交的含义一致.

在例 1.7 中,事件 A 与 B 的交为

$$A \cap B = \{1,3\}.$$

5. 事件的差

“事件 A 发生而 B 不发生”这一事件称为事件 A 与 B 的**差**,记作 $A-B$.

在例 1.7 中,事件 A 与 B 的差为

$$A - B = \{5\}.$$

6. 互不相容事件

若事件 A 与 B 不可能同时发生,也就是说,AB 是不可能事件,即 $AB = \varnothing$,则称事件 A 与 B 是**互不相容事件**.

比如,在投掷一枚骰子的试验中,“点数小于 3”和“点数大于 4”这两个事件是互不相容事件.

7. 对立事件

“事件 A 不发生”,这一事件称为事件 A 的**对立事件**,记作 $\overline{A}$,易见,$\overline{A} = \Omega - A$,且$(\overline{\overline{A}}) = A$. 根据定义,在一次试验中,如果 A 发生,则 $\overline{A}$ 一定不发生,如果 A 不发生,则 $\overline{A}$ 一定发生,也就是说 A 与 $\overline{A}$ 一定也只能发生其中之一,因而有

$$A\overline{A} = \varnothing, A \cup \overline{A} = \Omega$$

例如,在投掷一枚骰子的试验中记 A 为事件“点数为偶数”. 则 $\overline{A}$ 为事件“点数为奇数”.

8. 有限个或可数个事件的并与交

设有 n 个事件 $A_1, A_2, \cdots, A_n$,则称“$A_1, A_2, \cdots, A_n$ 至少有一个发生”这一事件为事件 $A_1, A_2, \cdots, A_n$ 的并,记作 $A_1 \cup A_2 \cup \cdots \cup A_n$ 或 $\bigcup_{i=1}^{n} A_i$. 称“$A_1, A_2, \cdots, A_n$ 都发生”这一事件为事件 $A_1, A_2, \cdots, A_n$ 的交,记作 $A_1 A_2 \cdots A_n$ 或 $\bigcap_{i=1}^{n} A_i$.

设有可数个事件 $A_1, A_2, \cdots, A_n, \cdots$,则称“$A_1, A_2, \cdots, A_n, \cdots$ 至少有一个发

生"这一事件为事件 $A_1, A_2, \cdots, A_n, \cdots$ 的并，记作 $\bigcup_{i=1}^{\infty} A_i$，称"$A_1, A_2, \cdots, A_n, \cdots$ 都发生"这一事件为事件 $A_1, A_2, \cdots, A_n, \cdots$ 的交，记作 $\bigcap_{i=1}^{\infty} A_i$.

9. 完备事件组

设 $A_1, A_2, \cdots, A_n, \cdots$ 是有限或可数个事件，如果其满足：

(1) $A_iA_j = \varnothing, i \neq j, i, j = 1, 2, \cdots$；

(2) $\bigcup_i A_i = \Omega$，

则称 $A_1, A_2, \cdots, A_n, \cdots$ 是一个**完备事件组**.

显然，A 与 $\overline{A}$ 构成一个完备事件组.

10. 事件的关系与运算的文氏图

上述关于事件的各种关系与运算可直观地用图形(文氏图)来表示(见图 1.1).

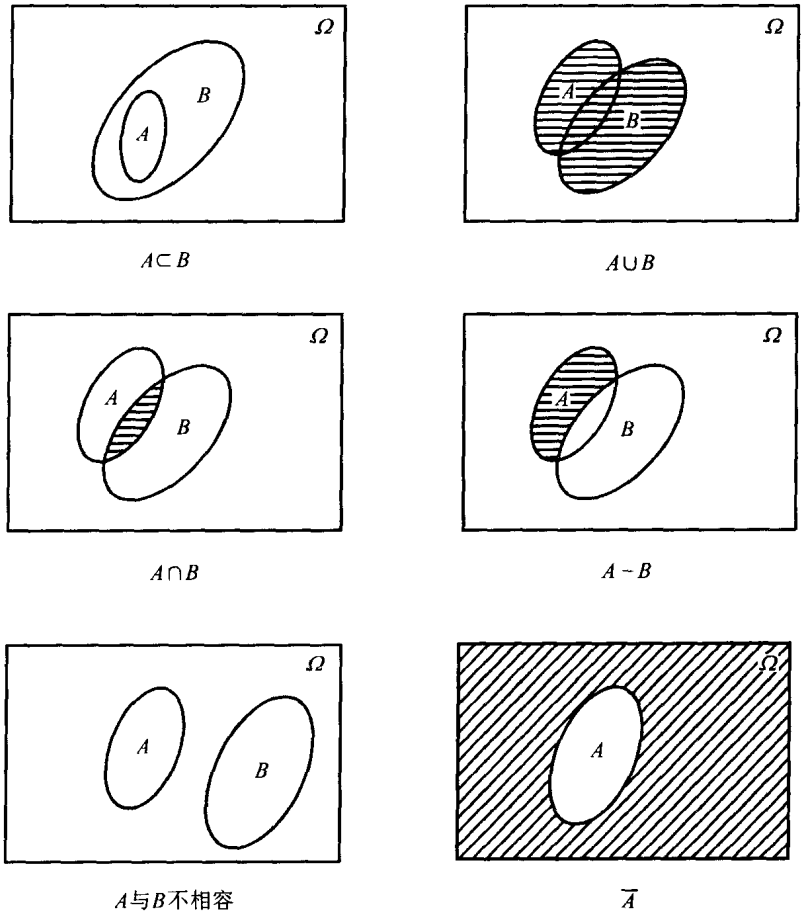


图 1.1　事件的关系与运算的文氏图

例 1.8 考察某一位同学在一次数学考试中的成绩，分别用 A,B,C,D,P,F 表示下列各事件(括号中表示成绩所处的范围)：

A——优秀([90,100])，　D——及格([60,70))，

B——良好([80,90))，　P——通过([60,100])，

C——中等([70,80))，　F——未通过([0,60))，

则 A,B,C,D,F 是两两不相容事件；P 与 F 是互为对立的事件，即有 $\overline{P}=F$；A,B,C,D 均为 P 的子事件，且有 $P=A\cup B\cup C\cup D$.

例 1.9 甲、乙、丙三人各射一次靶，记 A——“甲中靶”，B——“乙中靶”，C——“丙中靶”，则可用上述三个事件的运算来分别表示下列各事件：

(1) “甲未中靶”——$\overline{A}$；

(2) “甲中靶而乙未中靶”——$A\overline{B}$；

(3) “三人中只有丙未中靶”——$AB\overline{C}$；

(4) “三人中恰好有一人中靶”——$A\overline{B}\,\overline{C}+\overline{A}B\overline{C}+\overline{A}\,\overline{B}C$；

(5) “三人中至少有一人中靶”——$A+B+C$；

(6) “三人中至少有一人未中靶”——$\overline{A}+\overline{B}+\overline{C}$；

(7) “三人中恰有两人中靶”——$AB\overline{C}+A\overline{B}C+\overline{A}BC$；

(8) “三人中至少两人中靶”——$AB+AC+BC$；

(9) “三人均未中靶”——$\overline{A}\,\overline{B}\,\overline{C}$；

(10) “三人中至多一人中靶”——$A\overline{B}\,\overline{C}+\overline{A}B\overline{C}+\overline{A}\,\overline{B}C+\overline{A}\,\overline{B}\,\overline{C}$；

(11) “三人中至多两人中靶”——$\overline{ABC}$；

用其他事件的运算来表示一个事件，方法往往不惟一(比如例 1.9 中的(6)和(11)实际上是同一事件)读者应学会用不同方法表达同一事件，在解决具体问题时，特别是进行概率计算时，往往要根据需要选择其中的一种方法. 例 1.9 中许多事件不止一种表达方法，读者可对此进行讨论.

七、随机事件的运算律

我们学过集合的运算律，事件也有相应的运算律，归纳于下：

1. 关于求和运算

(1) $A\cup B=B\cup A$.(交换律)

(2) $(A\cup B)\cup C=A\cup(B\cup C)=A\cup B\cup C$.(结合律)

2. 关于求交运算

(1) $A\cap B=B\cap A$.(交换律)

(2) $(A\cap B)\cap C=A\cap(B\cap C)=A\cap B\cap C$.(结合律)

3. 关于求和与求交运算的混合

(1) $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$.(第一分配律)

(2) $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$.(第二分配律)

4. 关于求对立事件的运算

$$\overline{(\overline{A})}=A.(\text{自反律})$$

5. 关于和及交事件的对立事件

(1) $\overline{A\cup B}=\overline{A}\cap\overline{B}$.(第一对偶律)

(2) $\overline{A\cap B}=\overline{A}\cup\overline{B}$.(第二对偶律)

上述各运算律可以推广到有限个和可数个事件的情形,读者可通过复习集合论的知识自行给出.

§1.2 随机事件的概率

一、概率和频率解释

我们知道,在一次随机试验中,一个随机事件是否会发生,事先不能确定.但是,我们可以问,在一次试验中,事件 A 发生的可能性有多大?比如,投掷一枚均匀硬币,我们不能肯定是否会出现正面.但由于硬币是均匀的,我们有理由认为,出现正面和出现反面的可能性相同,均为$\frac{1}{2}$.但可能性大小究竟意味着什么呢?简单地说,它反映了一次试验中事件 A 出现的机会.然而这种机会在一次试验的实际结果中无法体现.可以想像,如果 A 发生的机会越大,那么在大量重复试验它将出现得越频繁,也就是说出现的频率会越大.在投掷一枚均匀硬币的试验中,如果果真出现正面和反面的机会均等,即可能性均为$\frac{1}{2}$,那么在大量重复试验中,出现正面和反面的频率会接近.事实正是如此,正如前一节指出的,大量重复投掷一枚均匀硬币,出现正面和反面的频率会接近一个稳定值$\frac{1}{2}$.可见频率的稳定值与事件发生的可能性大小存在内在必然的联系.一方面频率的稳定性说明事件发生的可能性大小确实是一种客观存在,另一方面,频率的稳定值对事件发生的可能性大小提供了经验解释.为此,引入下列定义:

定义 1.1 随机事件 A 发生的可能性大小的度量(数值),称为事件 A 发生的**概率**,记作 $P(A)$.

正如前面指出的,一个事件 A 发生的可能性的大小——概率,在经验上表现为大量重复试验中事件 A 发生的频率的稳定值.因而频率的稳定值为概率的含义提供了一种经验上的直观.但频率的稳定值本身并不是概率的本质,不能作为概率的定义.一个事件的概率是由事件本身特征所决定的客观存在,就好比一

根木棒有它的长度一样.频率的稳定值是概率的外在的必然表现,当进行大量重复试验时,频率会接近稳定值,因而,频率可用来作为概率的估计,就好比是测定概率的"尺子",随着试验次数的增加,测定的精度会越来越高.

二、从频率的性质看概率的性质

记一个事件 A 在 n 次重复试验中,发生的次数为 $r_n(A)$,则其发生的频率 $f_n(A)$为

$$f_n(A)=\frac{r_n(A)}{n}.$$

设与试验有关的全体事件的集合为 $\mathscr{F}$,通常称 $\mathscr{F}$ 为**事件域**.随着 A 取遍 $\mathscr{F}$ 中的任意事件,$f_n(A)$便成为定义在 $\mathscr{F}$ 上关于 A 的函数.容易证明,作为一个函数,$f_n(A)$满足下列性质:

(1) $f_n(\Omega)=1$;

(2) 对任意事件 A,有 $f_n(A)\geqslant 0$;

(3) 对任意一组(可数个)两两不相容的事件 $A_1,A_2,\cdots,A_n,\cdots$,有

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

上述各性质的证明是简单的,我们略去.但值得指出的是,$f_n(A)$还满足许多其他性质,比如,比较显然的性质有:$f_n(\varnothing)=0$,$f_n(A)\leqslant 1$.然而这些性质均可由上述三条性质导出,所以上述三条性质是反映频率特征的核心性质.

同频率一样,记事件 A 发生的概率为 $P(A)$,随着 A 取遍任意事件,$P(A)$则可视为定义在全体事件构成的集合,即事件域 $\mathscr{F}$ 上的一个函数.根据概率的频率解释,概率可视为频率的稳定值,从而应具有频率的相应性质,即

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

(2) 对任意事件 A,有 $P(A)\geqslant 0$;

(3) 对任意可数个两两不相容的事件 $A_1,A_2,\cdots,A_n,\cdots$,有

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

三、概率的公理化定义

任何一个数学概念都是对现实世界的抽象,这种抽象使得其具有广泛的适应性,并成为进一步数学推理的基础.前面指出,概率的频率解释为概率提供了经验基础,但不能作为一个严格的数学定义,它没能抓住"概率"这一概念的抽象本质.如果人们对概率的认识只停留在这一简单的直观上,那么人们对概率论的研究便只能停留在对一些肤浅的问题的零散研究上,概率论的研究和应用就会

受到很大的局限.从人们研究概率论有关问题开始算起,经过近三个世纪的漫长探索历程,伴随着数学的公理化潮流,人们才真正完整地解决概率的严格数学定义.其功劳归于前苏联著名的数学家科尔莫戈罗夫,他在 1933 年发表的《概率论的基本概念》一书中系统地表述了现在已被广泛接受的概率的公理化体系,第一次将概率论建立在严密的逻辑基础上.概率论从此确立了它作为一门严格的数学分支的地位.这一突破性的进展是一个里程碑,它将概率论推向一个全新的发展阶段.

我们知道,一个事件 A 的概率实际上是赋予事件 A 的一个实数值,记作 $P(A)$,那么当 A 在事件域 $\mathscr{F}$ 中变化时 $P(A)$便成为事件域 $\mathscr{F}$ 上的一个函数,记作 $P(\cdot)$.概率的公理化定义并不考虑每一个事件 A 对应的概率 $P(A)$是怎么确定的,值为多大(这依赖于每一个具体的实际问题的结构),而要求作为一个整体,函数 $P(\cdot)$应满足一些必要的条件,这些条件通常总结为一些合理公理,这些公理是对概率的现实直观进行的抽象.

定义 1.2　设 Ω 是一个样本空间,定义在 Ω 的事件域 $\mathscr{F}$ 上的一个实值函数 $P(\cdot)$称为 Ω 上的一个**概率测度**,如果它满足下列三条公理:

公理 1　$P(\Omega)=1$;

公理 2　对任意事件 A,有 $P(A)\geqslant 0$;

公理 3　对任意可数个两两不相容的事件 $A_1, A_2, \cdots, A_n, \cdots$,有

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

其中,对任意给定的具体事件 A,函数值 $P(A)$称为事件 A 的**概率**.此外,一个具有概率测度 $P(\cdot)$的样本空间 Ω 称为一个**概率空间**,记作(Ω, P)①.

概率测度还满足其他许多与直观相符的性质,这些性质均可由上述定义中的三条公理推导出来.

四、概率测度的其他性质

由概率的公理化定义中的三条公理出发,可以推导出概率测度的许多其他性质,这些性质有助于我们对概率概念的进一步了解,同时,它们也是概率计算的重要基础.

性质 1　$P(\varnothing)=0$.

证明　取 $A_i = \varnothing (i=1,2,\cdots)$,显然这是一列两两不相容的事件,且 $\bigcup_{i=1}^{\infty} A_i$

① 严格地讲,概率空间还应指明相应的事件域 $\mathscr{F}$,因而概率空间一般表达为$(\Omega, \mathscr{F}, P)$,在本书中,由于很少对事件域本身进行研究,因而省掉事件域 $\mathscr{F}$

$=\varnothing$,由公理3知

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

由于 $P(\varnothing)$为实数,故必有 $P(\varnothing)=0$.

性质2(有限可加性) $A_1,A_2,\cdots,A_n$是两两不相容的,则有

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

上述性质只需取 $A_i=\varnothing,i=n+1,n+2,\cdots$,即可由公理3及性质1导出.

下面各性质的证明均留作练习.

性质3 $P(\bar{A})=1-P(A)$.

性质4 $P(A-B)=P(A)-P(AB)$.

特别地,若 $A\supset B$,则

(1) $P(A-B)=P(A)-P(B)$;

(2) $P(A)\geqslant P(B)$.

性质5 $0\leqslant P(A)\leqslant 1$.

性质6 $P(A\cup B)=P(A)+P(B)-P(AB)$.

性质6可以推广到任意有限个事件的并的情形,这里我们给出了三个事件的情形,更一般的情形读者可以自行讨论.

推论 $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$.

例1.10 观察某地区未来5天的天气情况,记 A_i为事件:"有 i 天不下雨"($i=0,1,2,3,4,5$).已知 $P(A_i)=iP(A_0)(i=1,2,3,4,5)$.求下列各事件的概率:

(1) 5天均下雨;

(2) 至少一天不下雨;

(3) 至多三天不下雨.

解 显然 $A_0,A_1,\cdots,A_5$是两两不相容事件且 $\bigcup_{i=0}^{5}A_i=\Omega$,从而

$$1=P(\Omega)=P\Big(\bigcup_{i=0}^{5}A_i\Big)=\sum_{i=0}^{5}P(A_i)=P(A_0)+\sum_{i=1}^{5}iP(A_0)=16P(A_0).$$

于是可求得

$$P(A_0)=\frac{1}{16},P(A_i)=\frac{i}{16}(i=1,2,3,4,5).$$

记(1),(2),(3)中三个事件分别为 A,B,C,则

(1) $P(A)=P(A_0)=\frac{1}{16}$;

(2) $P(B)=P\left(\bigcup_{i=1}^{5}A_i\right)=1-P(A_0)=\frac{15}{16}$;

(3) $P(C)=P\left(\bigcup_{i=0}^{3}A_i\right)=P(A_0)+P(A_1)+P(A_2)+P(A_3)=\frac{7}{16}$.

例 1.11 已知 $P(\bar{A})=0.5, P(\bar{A}B)=0.2, P(B)=0.4$. 求:(1) $P(AB)$; (2) $P(A-B)$;(3) $P(A\cup B)$;(4) $P(\bar{A}\bar{B})$.

解 (1) 因为 $AB+\bar{A}B=B$,且 AB 与 $\bar{A}B$ 是不相容的.故有

$$P(AB)+P(\bar{A}B)=P(B).$$

于是

$$P(AB)=P(B)-P(\bar{A}B)=0.4-0.2=0.2;$$

(2) $P(A)=1-P(\bar{A})=1-0.5=0.5$,

$P(A-B)=P(A)-P(AB)=0.5-0.2=0.3$;

(3) $P(A\cup B)=P(A)+P(B)-P(AB)=0.5+0.4-0.2=0.7$;

(4) $P(\bar{A}\bar{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-0.7=0.3$.

§1.3 古典概型与几何概型

一、古典概型

概率的公理化定义给出了概率的严格的数学定义,人们在解决具体问题时,希望明确相应的概率测度是什么,一般而言,这是很困难的.这一节里,介绍两类特殊情形.首先介绍古典概型,它是一类最简单的概率模型.因为其简单,曾经是概率论发展早期的主要研究对象,也正因此,而今被冠以"古典概型".

古典概型是指满足下面两个假设条件的概率模型:

(1) 随机试验只有有限个可能结果;

(2) 每一个可能结果发生的可能性相同.

这两个条件在数学上可表述为:

(1)′ 样本空间有限,记 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$;

(2)′ 每一个基本事件的概率相同,即 $P\{\omega_1\}=P\{\omega_2\}=\cdots=P\{\omega_n\}$.根据概率的公理化定义知

$$1=P(\Omega)=P\left(\bigcup_{i=1}^{n}\{\omega_i\}\right)=\sum_{i=1}^{n}P\{\omega_i\},$$

于是由上述假设(2),知

$$P\{\omega_i\}=\frac{1}{n}, i=1,2,\cdots,n.$$

在古典概型的假设下，可以确定其相应的概率测度，即对任意事件 A，确定其概率 $P(A)$. 设 A 为任意一个事件，并令 A 中含有 m 个样本点，分别为 ω_{i_1}，ω_{i_2}，$\cdots$，ω_{i_m}，即

$$A=\{\omega_{i_1},\omega_{i_2},\cdots,\omega_{i_m}\}=\bigcup_{j=1}^{m}\{\omega_{i_j}\}.$$

由公理化定义中的公理 3 知

$$P(A)=\sum_{j=1}^{m}P\{\omega_{i_j}\}=\frac{m}{n}.$$

因而，古典概型的概率测度可表述为

$$P(A)=\frac{A\text{ 中元素个数}}{\Omega\text{ 中元素个数}}=\frac{\text{使 }A\text{ 发生的基本事件数}}{\text{基本事件总数}},$$

其中 A 为 Ω 中的任意事件.

例 1.12 一个袋子中装有 10 个大小相同的球，其中 3 个黑球，7 个白球. 求：

(1) 从袋子中任取一球，这个球是黑球的概率；

(2) 从袋子中任取两球，刚好一个白球一个黑球的概率以及两个球全是黑球的概率.

解 (1) 10 个球中任取一个，共有 $C_{10}^1=10$ 种取法，10 个球中有 3 个黑球，取到黑球的取法有 $C_3^1=3$ 种. 从而根据古典概率计算，事件 A：“取到的球为黑球”的概率为

$$P(A)=\frac{C_3^1}{C_{10}^1}=\frac{3}{10}.$$

(2) 10 个球中任取两球的取法有 C_{10}^2 种，其中刚好一个白球，一个黑球的取法有 $C_3^1\cdot C_7^1$ 种，两个球均是黑球的取法有 C_3^2 种. 记 B 为事件“刚好取到一个白球一个黑球”，C 为事件“两个球均为黑球”，则

$$P(B)=\frac{C_3^1C_7^1}{C_{10}^2}=\frac{21}{45}=\frac{7}{15},$$

$$P(C)=\frac{C_3^2}{C_{10}^2}=\frac{3}{45}=\frac{1}{15}.$$

例 1.13 将标号为 1,2,3,4 的四个球随意地排成一行，求下列各事件的概率：

(1) 各球自左至右或自右至左恰好排成 1,2,3,4 的顺序；

(2) 第 1 号球排在最右边或最左边；

(3) 第 1 号球与第 2 号球相邻；

(4) 第 1 号球排在第 2 号球的右边(不一定相邻).

解　将 4 个球随意地排成一行有 4！ = 24 种排法，即基本事件总数为 24，记(1)，(2)，(3)，(4)的事件分别为 A，B，C，D.

(1) A 有两种排法，故有

$$P(A)=\frac{2}{24}=\frac{1}{12};$$

(2) B 有 $2\times(3!)=12$ 种排法，故有

$$P(B)=\frac{12}{24}=\frac{1}{2};$$

(3) 先将第 1，2 号球排在任意相邻两个位置，共有 2×3 种排法，其余两个球可在其余两个位置任意排放，共有 2！种排法，因而 C 有 $2\times3\times2=12$ 种排法，故

$$P(C)=\frac{12}{24}=\frac{1}{2};$$

(4) 对于第 1 号球排在第 2 号球的右边的每一种排法，交换第 1 号球和第 2 号球的位置便得到第 1 号球排在第 2 号球的左边的一种排法，反之亦然. 因而第 1 号球排在第 2 号球的右边与第 1 号球排在第 2 号球的左边的排法种数相同，各占总排法数的 $\frac{1}{2}$，故有 $P(D)=\frac{1}{2}$.

例 1.14　将 n 个球随意地放入 N 个箱子中($N\geqslant n$)，其中每个球都等可能地放入任意一个箱子，求下列各事件的概率：

(1) 指定的 n 个箱子各放一球；

(2) 每个箱子最多放入一球；

(3) 某指定的箱子不空；

(4) 某指定的箱子恰好放入 k($k\leqslant n$)个球.

解　将 n 个球随意地放入 N 个箱子，共有 N^n 种放法，记(1)，(2)，(3)，(4)的事件分别为 A，B，C，D.

(1) 将 n 个球放进指定的 n 个箱子，每个箱子一个球，其放法有 $n!$ 种，故有

$$P(A)=\frac{n!}{N^n};$$

(2) 每个箱子最多放入一球等价于将 n 个球放进任意的 n 个箱子中，每箱一个球，其放法有 $C_N^n\cdot(n!)$(或记作 P_N^n)种，于是

$$P(B)=\frac{P_N^n}{N^n};$$

(3) 由于 C 的对立事件 $\overline{C}$ 表示“指定的箱子是空的”，它等价于将 n 个球全部放到其余 $N-1$ 个箱子中，共有 $(N-1)^n$ 种放法，从而

$$P(\bar{C})=\frac{(N-1)^n}{N^n},P(C)=1-P(\bar{C})=\frac{N^n-(N-1)^n}{N^n};$$

(4) 先任取 k 个球(有 C_n^k 种取法)放入指定的箱子中,然后将其余 $n-k$ 个球随意地放入其余 $N-1$ 个箱子(共有 $(N-1)^{n-k}$ 种放法),于是某指定的箱子恰好放入 k 个球的放法有 $C_n^k(N-1)^{n-k}$ 种,故有

$$P(D)=\frac{C_n^k(N-1)^{n-k}}{N^n}.$$

例 1.15 一个袋子中装有 $a+b$ 个球,其中 a 个黑球,b 个白球,随意地每次从中取出一球(不放回),求下列各事件的概率:

(1) 第 i 次取到的是黑球;

(2) 第 i 次才取到黑球;

(3) 前 i 次中能取到黑球.

解 因为所考虑的事件涉及取球的次序,所以基本事件也应考虑顺序,$(a+b)$ 次取球的总取法为 $(a+b)!$,记(1),(2),(3)中的事件分别为 A,B,C.

(1) 第 i 次取到的黑球可以是 a 个黑球中的任意一个,选定其中一个以后,其他各次取球必在 $a+b-1$ 个球中任意选取,共有 $(a+b-1)!$ 种取法,从而 A 中包含的取法有 $a\cdot[(a+b-1)!]$ 种,故

$$P(A)=\frac{a[(a+b-1)!]}{(a+b)!}=\frac{a}{a+b};$$

(2) 第 i 次才取到的黑球可以是 a 个黑球中的任意一个,第 1 到第 $i-1$ 次是在 b 个白球中任选 $i-1$ 个(共有 P_b^{i-1} 种取法),其他各次在剩下的 $a+b-i$ 个球中任意选取(共有 $(a+b-i)!$ 种取法),于是 B 所含的总取法为 $a\cdot P_b^{i-1}\cdot[(a+b-i)!]$,故

$$P(B)=\frac{a\cdot P_b^{i-1}\cdot[(a+b-i)!]}{(a+b)!}=\frac{a\cdot P_b^{i-1}}{P_{a+b}^{i}};$$

(3) 直接考虑事件 C 比较复杂,先考虑其对立事件 $\bar{C}$:"前 i 次未取到黑球",显然 $\bar{C}$ 包含的取法有 $P_b^i[(a+b-i)!]$ 种,于是

$$P(\bar{C})=\frac{P_b^i\cdot[(a+b-i)!]}{(a+b)!}=\frac{P_b^i}{P_{a+b}^i}=\frac{C_b^i}{C_{a+b}^i},$$

故

$$P(C)=1-P(\bar{C})=1-\frac{C_b^i}{C_{a+b}^i}.$$

值得指出的是例 1.15 的问题也可以理解为将 $(a+b)$ 个球随意地排序或理解为将 $(a+b)$ 个球随意地放入 $(a+b)$ 个箱子中,每个箱子放入一个球.此外,从例 1.15 的结果,我们还看出以下几点:

(1) 第 i 次取到的是黑球的概率与 i 无关，均为 $\frac{a}{a+b}$，自然，这也等于第 1 次取到的是黑球的概率，而后者的概率为 $\frac{a}{a+b}$ 则是显然的. 这一结果说明了实际生活中抽签的公平性：一组签中有若干好签，若干坏签，不管你先抽还是后抽，抽到好签的概率总是相同的.

(2) 从例中的结果来看，由于我们考虑的事件只涉及前 i 次，所以也可以只考虑取 i 次球，计算的结果与前面的结果一致.

(3) 对于例中的(3)，我们按只考虑取 i 次球来解答时，问题中的事件并不涉及这 i 次中的次序问题，因而也可以不考虑顺序，例中的答案正说明了这一点.

最后指出，上面所举各例均是古典概型计算中的典型问题，每个例子均是一类问题的一个代表. 我们有意识地编写成“取球”或“放球”问题，是为了使其更有代表性，这里的“球”可以理解为其他任何事物，只要所考虑的问题的结构与相应例子相同，便可套用例子的作法，读者可以通过本章提供的练习来领会这一点.

二、几何概型

古典概型是关于有限等可能结果的随机试验的概率模型. 现在考虑样本空间为一线段、平面区域或空间立体等的等可能随机试验的概率模型.

如果在一个面积为 $S(\Omega)$ 的区域 Ω 中等可能地任意投点(如图 1.2). 这里“等可能”的确切含义是：点落入 Ω 中任意区域 A 的可能性大小与区域 A 的面积 $S(A)$ 成正比，而与其位置和形状无关. 将“点落入区域 A”这一事件仍记为 A，同时有

$$P(A)=tS(A),$$

其中 t 为常数. 于是由

$$P(\Omega)=tS(\Omega)=1$$

知 $t=\frac{1}{S(\Omega)}$. 进而得

$$P(A)=\frac{S(A)}{S(\Omega)}. \qquad (1.1)$$

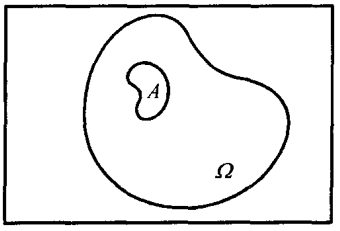


图 1.2　几何概型图示

由(1.1)定义的概率通常称为平面区域 Ω 上的**几何概率**. 容易证明几何概率满足概率的公理化定义. 值得指出的是，如果在一个线段上投点，将(1.1)中的面积改为长度；如果在一个空间立体上投点，将面积改为体积，则可分别定义线段和空间立体上的几何概率.

例 1.16　某人午觉醒来，发觉表停了，他打开收音机，想听电台报时，设电台每正点时报时一次，求他(她)等待时间短于 10 min 的概率.

解 以分钟为单位，记上一次报时时刻为 0，则下一次报时时刻为 60，于是这个人打开收音机的时间必在(0,60)内，记“等待时间短于 10 min”为事件 A，则有 $\Omega=(0,60)$，$A=(50,60)\subset\Omega$，于是

$$P(A)=\frac{10}{60}=\frac{1}{6}.$$

例 1.17(会面问题) 甲、乙两人相约在 7 点到 8 点之间在某地会面，先到者等候另一人 20 min，过时就离开. 如果每个人可在指定的一小时内任意时刻到达，试计算二人能够会面的概率.

解 记 7 点为计算时刻的 0 时，以分钟为单位，x,y 分别记甲、乙到达指定地点的时刻，则样本空间为

$\Omega=\{(x,y)\mid 0\leqslant x\leqslant 60,0\leqslant y\leqslant 60\}$. 以 A 表示事件“两人能会面”，则显然有

$$A=\{(x,y)\mid (x,y)\in\Omega,|x-y|\leqslant 20\}$$

(如图 1.3).

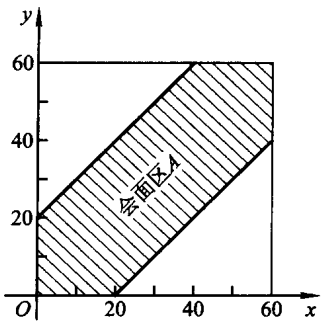


图 1.3 会面问题图示

根据题意，这是一个几何概型问题，于是

$$P(A)=\frac{S(A)}{S(\Omega)}=\frac{60^2-40^2}{60^2}=\frac{5}{9}.$$

§1.4 条件概率

一、引出条件概率的例子

当拥有关于试验结果的额外信息，比如，已知一个事件 A 已经发生时，我们可能需要对另一事件 B 发生的可能性大小重新作出度量，先看一个例子.

例 1.18 一批同型号产品由甲、乙两厂生产，产品结构如下表：

数量 / 厂别 / 等级	甲 厂	乙 厂	合 计
合格品	475	644	1 119
次 品	25	56	81
合 计	500	700	1 200

从这批产品中随意地取一件，则这件产品为次品的概率为

$$\frac{81}{1\,200}=6.83\%.$$

现在假设被告知取出的产品是甲厂生产的，那么这件产品为次品的概率是多大呢？回答这一问题并不困难．当被告知取出的产品是甲厂生产的，我们不能肯定的是该件产品是甲厂生产的 500 件中的哪一件，但由于 500 件中有 25 件次品，自然可得出，在已知取出的产品是甲厂生产的条件下，它是次品的概率为

$$\frac{25}{500}=5\%.$$

记“取出的产品是甲厂生产的”这一事件为 A，“取出的产品为次品”这一事件为 B．在前面实际上已计算了 $P(B)$，同时也算出了在“已知 A 发生”的条件下，B 发生的概率，这个概率称为在 A 发生的条件下，B 发生的条件概率，记作 $P(B|A)$．在本例中，我们注意到：

$$P(B|A)=\frac{25}{500}=\frac{25/1\,200}{500/1\,200}=\frac{P(AB)}{P(A)}.$$

事实上，容易验证，对一般的古典概型，只要 $P(A)>0$，总有

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

在几何概型中（以平面区域情形为例），在平面上的有界区域 Ω 内等可能地投点（如图 1.4），若已知 A 发生，则 B 发生的概率为

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

可见，在古典概型和几何概型这两类“等可能”概率模型中总有

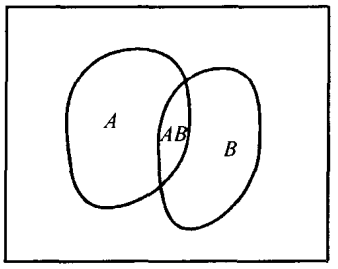


图 1.4　几何概型

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

事实上，还可以验证这一关系对频率也成立．由这些共性中得到启发，我们在一般的概率空间中引入条件概率的数学定义．

二、条件概率的数学定义

定义 1.3　给定概率空间 (Ω,P)，A,B 是其上的两个事件，且 $P(A)>0$，则称

$$P(B|A)=\frac{P(AB)}{P(A)} \tag{1.2}$$

为已知事件 A 发生的条件下，事件 B 发生的**条件概率**．随着 B 在事件域 $\mathscr{F}$ 中变化，$P(B|A)$ 便成为 $\mathscr{F}$ 上的函数，称之为在已知 A 发生的条件下的**条件概率**

测度.

不难验证,对给定的事件 A,$P(A)>0$,条件概率测度 $P(\cdot|A)$满足概率的三条公理:

(1) $P(\Omega|A)=1$;

(2) 对任意事件 B,有 $P(B|A)\geqslant 0$;

(3) 对任意可数个两两不相容的事件 $A_1,A_2,\cdots,A_n,\cdots$,有

$$P\Big(\bigcup_{i=1}^{\infty}A_i|A\Big)=\sum_{i=1}^{\infty}P(A_i|A).$$

由此可见,对给定的概率空间(Ω,P)及事件 A,如果 $P(A)>0$,则条件概率 $P(\cdot|A)$也是一个概率测度.特别地,当 $A=\Omega$ 时,$P(\cdot|A)$就是原来的概率测度 $P(\cdot)$.此外,由于 $P(\cdot|A)$也是概率测度,它也满足概率测度的其他性质.

在计算条件概率时,有时从试验的结构可直接得到条件概率,有时则需要利用(1.2)式来计算条件概率的值.下列例子中的两个问题分别对应于这两种情况.

例 1.19 一袋中装有 10 个球,其中 3 个黑球,7 个白球,先后两次从袋中各取一球(不放回).

(1) 已知第一次取出的是黑球,求第二次取出的仍是黑球的概率;

(2) 已知第二次取出的是黑球,求第一次取出的也是黑球的概率.

解 记 A_i为事件"第 i 次取到的是黑球"$(i=1,2)$.

(1) 在已知 A_1发生,即第一次取到的是黑球的条件下,第二次取球就在剩下的 2 个黑球、7 个白球共 9 个球中任取一个,根据古典概率计算,取到黑球的概率为$\frac{2}{9}$,即有

$$P(A_2|A_1)=\frac{2}{9}.$$

(2) 在已知 A_2发生,即第二次取到的是黑球的条件下,第一次取球发生在第二次取球之前,问题的结构不像(1)那么直观.采用(1.2)式计算 $P(A_1|A_2)$更方便一些.因为

$$P(A_1A_2)=\frac{P_3^2}{P_{10}^2}=\frac{1}{15},P(A_2)=\frac{3}{10},$$

由(1.2)可得

$$P(A_1|A_2)=\frac{P(A_1A_2)}{P(A_2)}=\frac{2}{9}.$$

例 1.19 中两个问题的结果表现出一种有趣的现象:已知第二次取到的是黑球的条件下,第一次取到的是黑球的概率与已知第一次取到的是黑球的条件下,

第二次取到的是黑球的概率相同，这一结论具有一般性. 作为练习，读者可以考虑 a 个黑球、b 个白球的情形. 如何理解这一现象呢？我们可以对例 1.19 的结果作出如下解释：

尽管第一次取球时，可能取到的是 10 个球中的一个，但当我们得知第二次取到的是黑球之后，我们反过来知道第一次取球必是 2 个黑球、7 个白球共 9 个球中的一个，从而结果跟(1)相同.

三、乘法公式

由条件概率定义中的(1.2)式立即可导出下列公式：

$$P(AB)=P(A)P(B|A),\qquad P(A)>0. \tag{1.3}$$

对称地，如果 $P(B)>0$，由

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{1.4}$$

可得

$$P(AB)=P(B)P(A|B),\qquad P(B)>0. \tag{1.5}$$

式(1.3)和(1.5)通常称为两个事件交的概率的**乘法公式**. 在有些问题中，条件概率 $P(B|A)$ 或 $P(A|B)$ 容易得到，于是可用乘法公式计算交的概率 $P(AB)$.

例 1.20　某批产品中，甲厂生产的产品占 60%，已知甲厂的产品的次品率为 10%，从这批产品中随意地抽取一件，求该件产品是甲厂生产的次品的概率.

解　记 A 表示事件“产品是甲厂生产的”，B 表示事件“产品是次品”，由题设知

$$P(A)=60\%,\qquad P(B|A)=10\%.$$

根据乘法公式，有

$$P(AB)=P(A)P(B|A)=60\%\times 10\%=6\%.$$

例 1.21　一袋中装 10 个球，其中 3 个黑球、7 个白球，先后两次从中随意各取一球(不放回)，求两次取到的均为黑球的概率.

这一概率，我们曾用古典概型方法计算过，这里我们使用乘法公式来计算. 在本例中，问题本身提供了分两步完成一个试验的结构，这恰恰与乘法公式的形式相对应，合理地利用问题本身的结构来使用乘法公式往往是使问题得到简化的关键.

解　设 A_i 表示事件“第 i 次取到的是黑球”($i=1,2$)，则 $A_1\cap A_2$ 表示事件“两次取到的均为黑球”. 由题设知：

$$P(A_1)=\frac{3}{10},\qquad P(A_2|A_1)=\frac{2}{9}.$$

于是根据乘法公式，有

$$P(A_1A_2)=P(A_1)P(A_2|A_1)=\frac{3}{10}\times\frac{2}{9}=\frac{1}{15}.$$

最后，我们指出，乘法公式(1.3)和(1.5)可以推广到任意有限个事件的交的概率的乘法公式：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}). \tag{1.6}$$

四、全概率公式

人们在计算某一较复杂的事件的概率时，有时根据事件在不同情况或不同原因或不同途径下发生而将它分解成两个或若干个互不相容的部分的并，分别计算每一部分的概率，然后求和，我们先看两个例子.

例 1.22 一袋中有 10 个球，其中 3 个黑球，7 个白球，从中先后随意各取一球(不放回)，求第二次取到的是黑球的概率.

这一概率，我们前面在古典概型中也计算过，这里我们用一种新的方法来计算.将事件“第二次取到的是黑球”根据第一次取球的情况分解成两个互不相容的部分，分别计算其概率再求和.

解 记 A_i 为事件“第 i 次取到的是黑球”($i=1,2$)，则有

$$\begin{aligned}P(A_2)&=P(A_1A_2)+P(\bar{A}_1A_2)\\&=P(A_1)P(A_2|A_1)+P(\bar{A}_1)P(A_2|\bar{A}_1).\end{aligned}$$

由题设易知

$$P(A_1)=\frac{3}{10},P(\bar{A}_1)=\frac{7}{10},$$

$$P(A_2|A_1)=\frac{2}{9},P(A_2|\bar{A}_1)=\frac{3}{9},$$

于是有

$$P(A_2)=\frac{3}{10}\times\frac{2}{9}+\frac{7}{10}\times\frac{3}{9}=\frac{3}{10}.$$

例 1.23 人们为了解一支股票未来一定时期内价格的变化，往往会去分析影响股票价格的基本因素，比如利率的变化.现在假设人们经分析估计利率下调的概率为 60%，利率不变的概率为 40%.根据经验，人们估计，在利率下调的情况下，该支股票价格上涨的概率为 80%，而在利率不变的情况下，其价格上涨的概率为 40%，求该支股票将上涨的概率.

解 记 A 为事件“利率下调”，那么 $\bar{A}$ 即为“利率不变”，记 B 为事件“股票价格上涨”.据题设知

$$P(A)=60\%,\qquad P(\bar{A})=40\%,$$

$$P(B|A)=80\%,\qquad P(B|\bar{A})=40\%,$$

于是

$$\begin{aligned}P(B)&=P(AB)+P(\bar{A}B)\\&=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})\\&=60\%\times 80\%+40\%\times 40\%\\&=64\%.\end{aligned}$$

上面两个不同的问题所采取的思路是完全一样的，要求一个事件 B 的概率，首先根据另一事件 A 发生与否两种情况（即 A 和 $\bar{A}$）将事件 B 分解成两个不相容的部分：AB 和 $\bar{A}B$，分别计算这两部分的概率．而且人们事先能够得到每一种情况（A 和 $\bar{A}$）发生的概率以及在每一种情况下事件 B 发生的概率，即条件概率 $P(B|A)$ 和 $P(B|\bar{A})$．可由乘法公式分别计算两部分的概率．这一结构可以通过图 1.5 所示的概率树来描述．

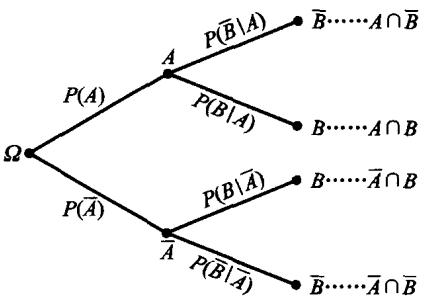


图 1.5　概率树

将上述方法一般化，便得到下面的定理．

定理 1.1（全概率公式）　设 $\{A_i\}$ 是一列有限或可数无穷个两两不相容的非零概率事件，且 $\bigcup_i A_i=\Omega$，则对任意事件 B，有

$$P(B)=\sum_i P(A_i)P(B|A_i).\tag{1.7}$$

证明　$P(B)=P(B\cap\Omega)=P[B\cap(\bigcup_i A_i)]$

$$=P[\bigcup_i (B\cap A_i)]=\sum_i P(B\cap A_i)$$

$$=\sum_i P(A_i)P(B|A_i).$$

例 1.22 和例 1.23 实际上是定理 1.1 的特殊情形．

五、贝叶斯公式

利用全概率公式，人们可以通过综合分析一个事件发生的不同原因、情况或途径及其可能性来求得该事件发生的概率．现在我们来考虑与之完全相反的问题．观察到一个事件已经发生，我们要考察所观察到的事件发生的各种原因、情况或途径的可能性．这里先通过一个简单的例子加以说明．

例 1.24（续例 1.22）　例 1.22 中，如果观察到第二次取到的球是黑球，求第

一次取到的是黑球的概率.

这一问题在例 1.19(2)中曾作出过解答,现在,换一个思路来考虑这一问题.正如例 1.22 那样,可以将"第二次取到的球为黑球"这一事件分解为两种情况下发生,那里利用全概率公式算得"第二次取到的球为黑球"的概率.现在的问题是,假设已经观察到"第二次取到的球为黑球",但不知道是在第一次取到的球为黑球的情况下第二次取到的是黑球,还是在第一次取到的球为白球的情况下第二次取到的是黑球,要求"第一次取到的是黑球"这种"情况"发生的概率.

解 设"第一次取到的是黑球"这一事件为 A,"第二次取到的是黑球"这一事件为 B,则问题归结为求条件概率 $P(A|B)$.

根据条件概率的定义,有

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

由乘法公式知

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

由全概率公式知

$$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A}),$$

于是得

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}. \tag{1.8}$$

据题设易知

$$P(A)=\frac{3}{10},\qquad P(B|A)=\frac{2}{9},$$

$$P(\bar{A})=\frac{7}{10},\qquad P(B|\bar{A})=\frac{3}{9},$$

从而得

$$P(A|B)=\frac{\frac{3}{10}\times\frac{2}{9}}{\frac{3}{10}\times\frac{2}{9}+\frac{7}{10}\times\frac{3}{9}}=\frac{2}{9}.$$

上述例子中的式(1.8)可以推广,我们将其表述为一个一般的公式,称为**贝叶斯公式**,它是以英国数学家贝叶斯(Bayes Thomas)命名的.

定理 1.2 设$\{A_i\}$是有限或可数无穷个两两不相容的非零概率事件,且 $\bigcup_{i=1}^{\infty}A_i=\Omega$,则对任意事件 B,$P(B)>0$,有

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}. \tag{1.9}$$

由全概率公式和乘法公式,该定理的证明是容易的,略去.

例1.25 设某批产品中,甲、乙、丙三厂生产的产品分别占45%、35%、20%,各厂产品的次品率分别为4%、2%、5%,现从中任取一件.

(1) 求取到的是次品的概率;

(2) 经检验发现取到的产品为次品,求该产品是甲厂生产的概率.

解 记"该产品为甲厂生产的"这一事件为 A_1;"该产品为乙厂生产的"这一事件为 A_2;"该产品为丙厂生产的"这一事件为 A_3;"该产品是次品"这一事件为 B.由题设知:

$$P(A_1)=45\%,\quad P(A_2)=35\%,\quad P(A_3)=20\%,$$

$$P(B|A_1)=4\%,\quad P(B|A_2)=2\%,\quad P(B|A_3)=5\%.$$

(1) 由全概率公式得:

$$\begin{aligned}P(B)&=\sum_{i=1}^{3}P(A_i)P(B|A_i)\\&=45\%\times4\%+35\%\times2\%+20\%\times5\%\\&=3.5\%;\end{aligned}$$

(2) 由贝叶斯公式(或条件概率定义)得:

$$P(A_1|B)=\frac{P(A_1B)}{P(B)}=\frac{P(A_1)P(B|A_1)}{P(B)}=\frac{45\%\times4\%}{3.5\%}=51.4\%.$$

§1.5 事件的独立性

一、两个事件的独立性

前一节曾指出,考察同一试验的两个事件,有时一个事件的发生与否会影响另一事件发生的概率.但有时一个事件的发生与否并不影响另一事件发生的概率.比如,投掷一枚硬币和投掷一枚骰子组成的一个试验中,硬币是否出现正面,不会影响骰子出现点数为5的概率.在数学上,一个事件 B 发生与否对另一事件 A 发生的概率没有任何影响,可表述为

$$P(A|B)=P(A),\tag{1.10}$$

其中 $P(B)>0$,则称 A 独立于 B.同样,如果

$$P(B|A)=P(B),\tag{1.11}$$

其中 $P(A)>0$,则称 B 独立于 A.

由于 $P(A)>0$,$P(B)>0$ 时,(1.10)和(1.11)均等价于

$$P(AB)=P(A)P(B).\tag{1.12}$$

因而,此时,A 独立于 B 等价于 B 独立于 A,故通常称 A 与 B 相互独立.注意到 $P(A)=0$,或 $P(B)=0$ 时(1.12)恒成立,为了使独立性概念包括零概率事件的情形,我们以后采用下列定义:

定义 1.4 设(Ω,P)是一个概率空间,A,B 是其上的两个事件,如果

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

则称 A 与 B **相互独立**,简称 A 与 B **独立**.

例 1.26 投掷一枚均匀的骰子.

(1) 设 A 表示事件"点数小于 5",B 表示事件"点数为奇数",则有

$$P(A)=\frac{4}{6},P(B)=\frac{3}{6},P(AB)=\frac{2}{6}=\frac{1}{3}.$$

由于

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

故 A 与 B 独立.

(2) 设 A 表示事件"点数小于 4",B 同(1),则有

$$P(A)=\frac{3}{6},P(B)=\frac{3}{6},P(AB)=\frac{2}{6}=\frac{1}{3}.$$

由于

$$P(AB)\neq P(A)P(B),$$

故 A 与 B 不独立.

二、有限个事件的独立性

定义 1.5 如果 $n(n\geqslant 2)$个事件:$A_1,A_2,\cdots,A_n$ 中任意两个事件均相互独立,即对任意 $1\leqslant i<j\leqslant n$,均有

$$P(A_iA_j)=P(A_i)P(A_j),\tag{1.13}$$

则称 n 个事件 $A_1,A_2,\cdots,A_n$ **两两独立**.

两两独立的概念只涉及 n 个事件中每对事件之间的相互关系,这种关系不涉及第三者.但我们通常还会碰到需要同时考虑多个事件之间的关系,比如,如果三个事件 A_1,A_2,A_3 两两独立,那么粗略地讲,其中任意一个事件发生与否对另一事件发生的概率没有影响,即 $P(A_i\mid A_j)=P(A_i),i\neq j,i,j=1,2,3$.但是,有例子可说明其中两个事件的发生可能会影响另一事件的概率,比如,可能会出现

$$P(A_1\mid A_2A_3)\neq P(A_1).$$

为此,我们引入一个比"两两独立"更强的独立性关系——n 个事件相互独立.

定义 1.6　设 $A_1, A_2, \cdots, A_n$ 为 $n(n \geqslant 2)$ 个事件，如果对其中任何 $k(2 \leqslant k \leqslant n)$ 个事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ $(1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n)$，均有

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}), \tag{1.14}$$

则称 $A_1, A_2, \cdots, A_n$ **相互独立**.

显然，当 $n=2$ 时，两两独立与相互独立是同一概念.但一般地，当 $n>2$ 时，相互独立性是比两两独立性更强的性质.直观地讲，n 个事件相互独立，那么其中任何一个或多个事件的发生与否不会对其他事件发生的概率产生影响.

此外，从定义 1.6 容易看出，如果 n 个事件相互独立，则其中任意 k 个事件 $(2 \leqslant k \leqslant n)$ 也相互独立.

三、相互独立性的性质

事件之间具有独立性的情况下，很多概率计算将变得简单，其中经常用到下面两个性质.

性质 1　如果 n 个事件 $A_1, A_2, \cdots, A_n$ 相互独立，则将其中任何 $m(1 \leqslant m \leqslant n)$ 个事件改为相应的对立事件，形成的新的 n 个事件仍然相互独立.

性质 2　如果 n 个事件 $A_1, A_2, \cdots, A_n$ 相互独立，则有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} P(\overline{A}_i) = 1 - \prod_{i=1}^{n} (1 - P(A_i)). \tag{1.15}$$

性质 1 的证明只需对 $m=1$ 证明，然后使用数学归纳法.性质 2 是性质 1 的直接推论.详细证明过程，留作练习.

例 1.27　甲、乙、丙三人各射一次靶，他们各自中靶与否相互独立，且已知他们各自中靶的概率分别为 0.5、0.6、0.8，求下列事件的概率：

(1) 恰有一人中靶；

(2) 至少有一人中靶.

解　设 $A_i(i=1,2,3)$ 分别表示甲、乙、丙中靶三个事件，则“恰有 1 人中靶”这一事件可表示为 $A_1\overline{A}_2\overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1\overline{A}_2 A_3$，“至少有 1 人中靶”这一事件可表示为 $A_1 + A_2 + A_3$.

(1) $P(A_1\overline{A}_2\overline{A}_3 + \overline{A}_1 A_2\overline{A}_3 + \overline{A}_1\overline{A}_2 A_3)$

$= P(A_1\overline{A}_2\overline{A}_3) + P(\overline{A}_1 A_2\overline{A}_3) + P(\overline{A}_1\overline{A}_2 A_3)$

$= P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3)$

$= 0.5\times(1-0.6)\times(1-0.8) + (1-0.5)\times 0.6\times(1-0.8) +$

$(1-0.5)(1-0.6)\times 0.8$

$= 0.26.$

(2) $P(A_1+A_2+A_3)=1-P(\bar{A}_1)P(\bar{A}_2)P(\bar{A}_3)$

$=1-0.5\times0.4\times0.2$

$=0.96.$

四、伯努利(Bernoulli)概型

有时,我们需要研究的问题涉及多个甚至无穷多个试验,这多个或无穷多个试验通常称为一个**试验序列**.

定义 1.7 一个试验序列称为一个**独立试验序列**,如果它的各试验的结果之间是相互独立的.

在实际中经常碰到一类特殊的试验,它只有两个可能结果,这样的试验称为**伯努利试验**.比如,投掷一枚硬币时观察其出现正面还是反面、抽取一件产品考察其是正品还是次品等.

有些试验的结果虽然不只两个,但我们可能对试验感兴趣的是某事件 A 是否发生,那么可以把 A 作为一个结果,$\bar{A}$ 作为另一结果,从而也可以将试验归结为伯努利试验.比如,一个灯泡的寿命.它可以取不小于 0 的任何数值,但有时根据需要,我们将寿命大于 1 000 小时的灯泡当作合格品,而把寿命不大于 1 000 小时的灯泡当作次品,我们感兴趣的可能是灯泡是合格品还是次品.

定义 1.8 一个试验序列称为**伯努利试验序列**,如果它是由一个伯努利试验独立重复进行形成的试验序列.特别地,由一个伯努利试验独立重复 n 次形成的试验序列称为 n **重伯努利试验**.

设在一次伯努利试验中,$P(A)=p$,$P(\bar{A})=q$,其中 $p>0$,$q>0$,且 $p+q=1$,那么相应的 n 重伯努利试验,根据定义,其特点为:事件 A 在每次试验中发生的概率均为 p,且不受其他各次试验中 A 是否发生的影响.

定理 1.3(伯努利定理) 在一次试验中,事件 A 发生的概率为 $p(0<p<1)$,则在 n 重伯努利试验中,事件 A 恰好发生 k 次的概率(记作 $b(k;n,p)$)为

$$b(k;n,p)=C_n^k p^k q^{n-k},$$

其中 $q=1-p$.

证明 记"第 i 次试验中事件 A 发生"这一事件为 A_i,$i=1,2,\cdots,n$,则"事件 A 恰好发生 k 次"(记作 B_k)是下列 C_n^k 个两两不相容事件的并:

$$A_{i_1}A_{i_2}\cdots A_{i_k}\bar{A}_{j_1}\bar{A}_{j_2}\cdots\bar{A}_{j_{n-k}},$$

其中 $i_1,i_2,\cdots,i_k$ 是取遍 $1,2,\cdots,n$ 中的任意 k 个数(共有 C_n^k 种取法),$j_1,j_2,\cdots,j_{n-k}$ 是取走 $i_1,i_2,\cdots,i_k$ 后剩下的 $n-k$ 个数.

而对任意取出的 $i_1,i_2,\cdots,i_k$,根据独立性及 $P(A_i)=p$,有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}\bar{A}_{j_1}\bar{A}_{j_2}\cdots\bar{A}_{j_{n-k}})$$

$$= P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})P(\bar{A}_{j_1})\cdots P(\bar{A}_{j_{n-k}})$$
$$= p^k q^{n-k}.$$

故有

$$b(k;n,p) = P(B_k) = C_n^k p^k q^{n-k}.$$

定理 1.4 在伯努利试验序列中,设每次试验中事件 A 发生的概率为 p,"事件 A 在第 k 次试验中才首次发生"($k \geqslant 1$)这一事件的概率为

$$g(k,p) = q^{k-1}p.$$

证明 "事件 A 在第 k 次试验中才首次发生"等价于在前 k 次试验组成的 k 重伯努利试验中"事件 A 在前 $k-1$ 次试验中均不发生而第 k 次试验中事件 A 发生",于是根据定理 1.3 的证明可知定理 1.4 成立.

例 1.28 一袋中装有 10 个球,其中 3 个黑球 7 个白球,每次从中随意取出一球,取后放回.

(1) 如果共取 10 次,求 10 次中能取到黑球的概率及 10 次中恰好取到 3 次黑球的概率.

(2) 如果未取到黑球就一直取下去,直到取到黑球为止,求恰好要取 3 次的概率及至少要取 3 次的概率.

解 记 A_i 为事件"第 i 次取到的是黑球",则 $P(A_i) = \dfrac{3}{10}, i = 1,2,\cdots$.

(1) 记 B 为事件"10 次中能取到黑球",B_k 为事件"10 次中恰好取到 k 次黑球",$k = 0,1,\cdots,10$. 于是有

$$P(B) = 1 - P(\bar{B}) = 1 - P(B_0) = 1 - \left(\frac{7}{10}\right)^{10},$$

$$P(B_3) = C_{10}^3 \left(\frac{3}{10}\right)^3 \left(\frac{7}{10}\right)^7.$$

(2) 记 C 为"恰好要取 3 次",D 为"至少要取 3 次". 则由定理 1.4 知

$$P(C) = \left(\frac{7}{10}\right)^2 \cdot \frac{3}{10},$$

$$P(D) = P(\bar{A}_1\bar{A}_2) = P(\bar{A}_1)P(\bar{A}_2) = \left(\frac{7}{10}\right)^2.$$

例 1.29 一辆飞机场的交通车载有 25 名乘客途经 9 个站,每位乘客都等可能在这 9 站中任意一站下车(且不受其他乘客下车与否的影响),交通车只在有乘客下车时才停车,求交通车在第 i 站停车的概率以及在第 i 站不停车的条件下在第 j 站停车的概率. 并判断"第 i 站停车"与"第 j 站停车"两个事件是否独立.

解 记 A_k 为"第 k 位乘客在第 i 站下车"($k = 1,2,\cdots,25$). 考察每一位乘客在第 i 站是否下车,可视为一个 25 重的伯努利试验,记 B 为"第 i 站停车",C

为“第 j 站停车”，则 B,C 分别等价于“第 i 站有人下车”和“第 j 站有人下车”. 于是有

$$P(B)=1-\left(\frac{8}{9}\right)^{25},\qquad P(C)=1-\left(\frac{8}{9}\right)^{25}.$$

在 B 不发生(即 $\overline{B}$ 发生)的条件下，每位乘客均等可能地在第 i 站以外的 8 个站中任意一站下车，于是每位乘客在第 j 站下车的概率为 $\frac{1}{8}$，故有

$$P(C|\overline{B})=1-\left(\frac{7}{8}\right)^{25}.$$

由于 $P(C|\overline{B})\neq P(C)$，故 $\overline{B}$ 与 C 不独立，从而 B 与 C 不独立.

习 题 一

(A)

1. 写出下列各试验的样本空间：

(1) 掷两枚骰子，分别观察其出现的点数；

(2) 观察一支股票某日的价格(收盘价)；

(3) 一人射靶三次，观察其中靶次数；

(4) 一袋中装有 10 个同型号的零件，其中 3 个合格 7 个不合格，每次从中随意取出一个，不合格便放回去，直到取到合格的零件为止，观察所抽取的次数.

2. 试问第 1 题中你所给出的样本空间中，哪些遵循古典概型的假设？

3. 掷一枚骰子，观察其出现的点数，A 表示“出现奇数点”，B 表示“出现的点数小于 5”，C 表示“出现的点数是小于 5 的偶数”，用集合列举法表示下列事件：$\Omega,A,B,C,A+B,A-B,B-A,AB,AC,\overline{A}+B$.

4. 写出例 1.9 中(5)，(8)，(10)，(11)等事件的其他表示方法.

5. 互不相容事件与对立事件的区别何在？说出下列各对事件之间的关系：

(1) $|x-a|<\delta$ 与 $x-a\geqslant\delta$；

(2) $x>20$ 与 $x\leqslant 20$；

(3) $x>20$ 与 $x<18$；

(4) $x>20$ 与 $x\leqslant 22$；

(5) “20 件产品全是合格品”与“20 件产品中恰有一件是废品”；

(6) “20 件产品全是合格品”与“20 件产品中至少有一件是废品”；

(7) “20 件产品全是合格品”与“20 件产品中至多有一件是废品”.

6. 某人用步枪射击目标 5 次，A_i 表示“第 i 次射击击中目标”($i=1,2,3,4,5$). B_i 表示“5 次射击中击中目标 i 次”($i=0,1,2,3,4,5$)，用文字叙述下列各事件，并指出各对事件之间的关系：

(1) $\bigcup_{i=1}^{5}A_i$ 与 $\bigcup_{i=1}^{5}B_i$；

(2) $\bigcup_{i=2}^{5}A_i$ 与 $\bigcup_{i=2}^{5}B_i$；

(3) $\bigcup_{i=1}^{2} A_i$ 与 $\bigcup_{i=3}^{5} A_i$；　　(4) $\bigcup_{i=1}^{2} B_i$ 与 $\bigcup_{i=3}^{5} B_i$；

(5) $\bigcup_{i=0}^{2} B_i$ 与 $\bigcup_{i=3}^{5} B_i$；　　(6) $A_1\bar{A}_2\bar{A}_3A_4A_5$ 与 B_3；

(7) $\bar{A}_1$ 与 B_5；　　(8) $\bigcup_{i=1}^{5} \bar{A}_i$ 与 $\bar{B}_5$.

7. 证明下列关系式：

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

8. 由概率的公理化定义推导概率测度的其他性质；性质 3～性质 6 以及性质 6 的推论.

9. 已知 $P(A)=0.4, P(B)=0.25, P(A-B)=0.25$，求 $P(AB), P(A\cup B), P(B-A), P(\bar{A}\bar{B})$.

10. 已知 $P(A)=0.4, P(B\bar{A})=0.2, P(C\bar{A}\bar{B})=0.1$，求 $P(A+B+C)$.

11. N 件产品中有 N_1 件次品，从中任取 n 件(不放回)，其中 $1\leqslant n\leqslant N$.(1) 求其中恰有 k 件($k\leqslant n$ 且 $k\leqslant N_1$)次品的概率；(2) 求其中有次品的概率；(3) 如果 $N_1\geqslant 2, n\geqslant 2$，求其中至少有两件次品的概率.

12. 一个班共有 30 名同学，其中有 6 名女生，假设他们到校先后次序的所有模式都有同样的可能性.(1) 求男生均比女生先到校的概率；(2) 求班上李明和王菲两位同学中，李明比王菲先到校的概率.

13. 某班级有 n 个同学($n\leqslant 365$)，求至少有两位同学的生日在同一天的概率(设一年按 365 天计).

14. 一辆飞机场的交通车载有 25 名乘客，途经 9 个站，每位乘客都等可能在 9 个站中任意一站下车，交通车只在有乘客下车时才停车，求下列各事件的概率：

(1) 交通车在第 i 站停车；

(2) 交通车在第 i 站和第 j 站至少有一站停车；

(3) 交通车在第 i 站和第 j 站均停车；

(4) 在第 i 站有 3 人下车.

15. 两封信随机地投入 4 个邮筒，求前两个邮筒没有信的概率及第一个邮筒恰有一封信的概率.

16. 求例 1.15 中，前 i 次中恰好取到 k 个黑球的概率.

17. 一串钥匙，共有 10 把，其中有 4 把能打开门，因开门者忘记哪些能打开门，便逐把试开，求下列事件的概率：

(1) 第 3 把钥匙能打开门；

(2) 第 3 把钥匙才打开门；

(3) 最多试 3 把钥匙就能打开门.

18. 连续投掷一枚均匀硬币 10 次，求其中有 3 次是正面的概率.

19. 一个袋子中装有 5 个红球，3 个白球，2 个黑球，从中任取 3 个球，求其中恰有一个红球、一个白球和一个黑球的概率.

20. 将 13 个分别写有 A、A、A、C、E、H、I、I、M、M、N、T、T 的卡片随意地排成一行，

求恰好排单词“*MATHEMATICIAN*”的概率.

21. 某公共汽车站每隔 10 min 有一辆汽车到达,一位乘客到达汽车站的时间是任意的,求他等候时间不超过 3 min 的概率.

22. 两艘轮船都要停靠同一泊位,它们可能在一昼夜的任意时间到达,设两船停靠泊位的时间分别需要 1 小时与两小时,求一艘轮船停靠泊位时,需要等待空出码头的概率.

23. 一袋中装有 a 个黑球,b 个白球.先后两次从袋中各取一球(不放回).

(1) 已知第一次取出的是黑球,求第二次取出的仍是黑球的概率;

(2) 已知第二次取出的是黑球,求第一次取出的也是黑球的概率;

(3) 已知取出的两个球中有一个是黑球,求另一个也是黑球的概率.

24. 已知 A 是概率空间(Ω, P)上的事件,$P(A)>0$.证明:

(1) 如果 $B \supset A$,则 $P(B \mid A)=1$;

(2) 如果 B_1, B_2是(Ω, P)上的两个事件,且 $B_1 B_2 = \varnothing$,则 $P(B_1 + B_2 \mid A) = P(B_1 \mid A) + P(B_2 \mid A)$.

25. 一个家庭中有两个小孩.(1) 已知其中有一个是女孩,求另一个也是女孩的概率;(2) 已知第一胎是女孩,求第二胎也是女孩的概率.

26. 掷两枚均匀的骰子,已知它们出现的点数各不相同,求其中有一个点数为 4 的概率.

27. 设 10 个考题签中有 4 题难答.3 人参加抽签,甲先抽,乙次之,丙最后.求下列事件的概率:

(1) 甲抽到难答签;

(2) 甲未抽到难答签而乙抽到难答签;

(3) 甲、乙、丙均抽到难答签.

28. 有外形相同的球分装 3 个盒子,每盒 10 个球.其中第一个盒子中有 7 个球标有字母 A,3 个球标有字母 B;第二个盒子中有红球和白球各 5 个;第三个盒子则有红球 8 个,白球 2 个.试验按如下规则进行:先在第一个盒子中任取一球,若取得标有字母 A 的球,则在第二个盒子中任取一球;若第一次取到标有字母 B 的球,则在第三个盒子中任取一球.如果第二次取到的是红球,则称试验成功,求试验成功的概率.

29. 某商店收进甲厂生产的产品 30 箱,乙厂生产的同种产品 20 箱,甲厂产品每箱装 100 个,废品率为 0.06,乙厂产品每箱 120 个,废品率为 0.05.

(1) 任取一箱,从中任取一个产品,求其为废品的概率;

(2) 若将所有产品开箱混装,任取一个其为废品的概率.

30. 12 个乒乓球中有 9 个新球,3 个旧球.第一次比赛,取出 3 个球,用完以后放回去,第二次比赛又从中取出 3 个球.(1) 求第二次取出的 3 个球中有 2 个新球的概率;(2) 若第二次取出的 3 个球中有 2 个新球,求第一次取到的 3 个球中恰有一个新球的概率.

31. 血液试验 ELISA(Enzyme Linked Imriunosorbent Assay 酶连接免疫吸附测定)是现今检验艾滋病病毒的一种流行方法.假定 ELISA 试验能正确测出确实带有病毒的人中的 95% 存在艾滋病病毒,又把不带病毒的人中的 1% 不正确地识别为存在病毒.又假定在总人口 1 000人中大约有 1 人确实带有艾滋病病毒,如果对某人的检验结果呈阳性(即认为带有病毒),那么他真的带有艾滋病病毒的概率有多大? 如果被检测者属于“高感染人群”中的一员,

而估计这一高感染人群中大约100人中有1人带有病毒,那么检测为阳性的人,真的带有艾滋病病毒的概率有多大?

32. 一男子到闹市区去,他遇到背后袭击并被抢劫,他断言凶手是个白人,然而当调查这一案件的法院在可比较的光照条件下多次重复展现现场情况时,受害者正确识别袭击者种族的次数约占80%,袭击者确实是白人的概率是0.8吗?试给出说明.

33. 一个均匀的四面体,其第一面染红色;第二面染白色,第三面染黑色,而第四面染红、白、黑三种颜色,以 A、B、C 分别记投掷一次四面体,底面出现红、白、黑的三个事件,判断 A、B、C 是否两两独立?是否相互独立?

34. 设 $P(A)=0$ 或 1,证明 A 与其他任何事件 B 相互独立.

35. 设 $P(A)>0$,$P(B)>0$,且 A、B 互不相容,那么 A 与 B 相互独立吗?为什么?

36. A、B、C 相互独立,证明 $\bar{A}$、$\bar{B}$、C 亦相互独立.

37. 证明相互独立性的性质1和性质2.

38. 加工一个产品要经过三道工序,第一、二、三道工序不出废品的概率分别为0.9、0.95、0.8,若假定各工序是否出废品是独立的,求经过三道工序生产出的是废品的概率.

39. 一个自动报警器由雷达和计算机两部分组成,两部分有任何一个失灵,这个报警器就失灵,若使用100小时后,雷达失灵的概率为0.1,计算机失灵的概率为0.3,若两部分失灵与否独立,求这个报警器使用100小时而不失灵的概率.

40. 高射炮向敌机发射三发炮弹(每弹击中与否相互独立),设每发炮弹击中敌机的概率均为0.3.又知若敌机中一弹,其坠落的概率为0.2;若敌机中两弹,其坠落的概率为0.6;若敌机中三弹则必然坠落.(1) 求敌机被击落的概率;(2) 若敌机被击落,求它中两弹的概率.

(B)

1. 为了减少比赛场次,把20个球队分成两组,每组10队进行比赛,求最强的两队被分在同一组的概率及最强两队分在不同组的概率.

2. 若 n 个人站成一行,其中有 A、B 两人,问夹在 A、B 之间恰有 r 个人的概率是多少?如果 n 个人围成一个圆圈,求从 A 到 B 的顺时针方向,A、B 之间恰有 r 个人的概率.

3. (巴拿赫问题)某数学家有两盒火柴,每一盒装有 N 根.每次使用时,他在任一盒中取一根,问他发现一盒空,而另一盒还有 k 根火柴的概率是多少?

4. 有 k 个坛子,每一个装有 n 个球,分别编号为1至 n,今从每个坛子中任取一球,问 m 是所取的球中的最大编号的概率.

5. 一根长为 l 的棍子在任意两点折断,试计算得到的三段能围成三角形的概率.

6. 今有两名射手轮流对同一目标射击,甲射手命中概率为 p_1,乙射手命中概率为 p_2,甲先射,谁先命中谁得胜,求甲、乙得胜的概率各是多少?

7. 设有来自三个地区的10名、15名、25名考生的报告表,其中女生的报名表分别为3份、7份、5份.随机地取一个地区的报名表,从中先后抽取两份.

(1) 求先抽到的一份是女生表的概率 p;

(2) 已知后抽到的一份是男生表,求先抽到的是女生表的概率 q.

8. 在一通讯渠道中,能传送字符AAAA,BBBB,CCCC三者之一,由于通讯噪声干扰,正

确接收到被传送字母的概率为 0.6,而接收到其他两个字母的概率均为 0.2,假设前后字母是否被歪曲互不影响.

(1) 求收到字符 ABCA 的概率;

(2) 若收到字符为 ABCA,问被传送字符为 AAAA 的概率是多大.

第 2 章

随机变量的分布与数字特征

对于一个随机试验，人们除了对某些特定的事件发生的概率感兴趣以外，往往还会关心某个与随机试验的结果相联系的变量. 由于这一变量的取值依赖于试验结果，而试验结果是不确定的，所以这一变量的取值也是不确定的，这种变量因而被称为随机变量. 对于随机变量，人们无法准确预知其确切取值，但人们可以研究其取值的统计规律性. 对一个随机变量的统计规律性的完整描述被称为随机变量的分布，本章将介绍的两类随机变量——离散型和连续型随机变量及其分布. 但在多数实际应用中，人们很难知道一个随机变量的真实分布，这时人们希望用一些综合指标来反映随机变量的统计规律中的某些重要特征，这些指标被称为随机变量的数字特征. 最常用的数字特征有“数学期望”和“方差”. 随机变量的数字特征不仅容易估计，而且为随机变量的取值提供了综合评价，这恰恰是解决一些概率应用问题的必要前提.

§2.1 随机变量及其分布

一、随机变量的概念

在一些随机试验中，试验的结果本身就是由数量来表示的. 比如，投掷一枚骰子. 观察其出现的点数，可能的结果可分别由 1,2,3,4,5,6 来表示；再比如，观察一个灯泡的使用寿命，实际使用寿命可能是 $[0,+\infty)$ 中的任何一个实数. 在另一些随机试验中，我们可能根据问题的需要对每一个可能结果指定一个数量. 比如，投掷一枚硬币进行打赌时，如果规定投掷者在硬币出现正面时赢 1 元钱，出现反面时输 1 元钱，则可对“出现正面”指定一个数 1，对“出现反面”指定一个数 -1. 无论哪种情形，其共同点是：对每一个可能结果，有惟一一个实数与之对应. 这种对应关系实际定义了样本空间 Ω 上的函数，通常记作 $X=X(\omega),\omega\in\Omega$. 在这里函数 $X=X(\omega),\omega\in\Omega$，其“自变量”$\omega$ 在样本空间 Ω 中“取值”(即定义域为 Ω)，我们主要关心 ω 的不确定性及其统计规律所导致的因变量 X 的不确定性及其统计规律. 我们把取值依赖于一个试验的结果而具有不确定性的变量称

为随机变量,其正式数学定义如下:

定义 2.1 定义在概率空间(Ω,P)上,取值为实数的函数[①]$X=X(\omega)(\omega\in\Omega)$称为$(\Omega,P)$上的一个**随机变量**.

随机变量 X 的取值由样本点 ω 决定.反过来,X 取某一特定值 a 的那些样本点的全体构成 Ω 的一个子集,即$\{\omega\mid X(\omega)=a\}\subset\Omega$.同样,设 I 为实数集 $\mathbf{R}$ 的一个子区间,使得 X 的值落在 I 中的那些样本点全体也是 Ω 的一个子集.为了研究随机变量 X 的统计规律,我们均假设这些子集是随机事件,也假设这些事件的可数并、交及补都是事件[②],并称这些事件为随机变量 X 生成的事件.为表达简便,今后在事件的表示中省掉 ω,比如$\{X=a\}$,$\{a<X\leqslant b\}$等.

在掷骰子的试验中,其出现的点数记为随机变量 X,则 X 作为样本空间 $\Omega=\{1,2,3,4,5,6\}$上的函数定义为

$$X(\omega)=\omega,\omega\in\Omega.$$

在投掷一枚硬币进行打赌时,出现正面时投掷者赢 1 元钱,出现反面时输 1 元钱,记赢钱数为随机变量 X,则 X 作为样本空间$\Omega=\{$正面,反面$\}$上的函数定义为

$$X(\omega)=\begin{cases}1, & \omega=\text{正面},\\ -1, & \omega=\text{反面}.\end{cases}$$

二、离散型随机变量的概率分布

在所有的随机变量中,有一类随机变量最简单,它只有有限个或可数无穷多个可能取值.

定义 2.2 设 X 是定义在概率空间(Ω,P)上的一个随机变量,如果 X 的全部可能取值只有有限个或可数无穷多个,则称 X 是一个**离散型随机变量**.

设 $x_1,x_2,\cdots$是 X 的所有可能取值,对每个 x_i,$\{X=x_i\}$是 Ω 上的一个随机事件,人们往往关心这些事件发生的可能性,也即 X 取每一个可能值 x_i 的概率.

定义 2.3 设 X 是离散型随机变量,其全部可能取值为$\{x_i,i=1,2,\cdots\}$,记

$$p(x_i)=P\{X=x_i\},\qquad i=1,2,\cdots,\tag{2.1}$$

则称$\{p(x_i),i=1,2,\cdots\}$为 X 的**概率分布**.有时也将 $p(x_i)$记为 p_i,用下列表格形式来表示,并称之为 X 的概率分布表:

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

① 严格讲,应为关于事件域 $\mathcal{F}$ 可测的函数 由于本书不对事域作讨论,因而忽略可测性

② 这些假说实际上就是要求随机变量是可测函数

容易看出，任何一个离散型随机变量的概率分布$\{p(x_i)\}$必然满足下列性质：

(1) $p(x_i)\geqslant 0, i=1,2,\cdots$；　　(2.2)

(2) $\sum_i p(x_i)=1$.　　(2.3)

例 2.1　投掷一枚均匀硬币，设 X 为一次投掷中出现正面的次数，即

$$X(\omega)=\begin{cases}1, \omega=\text{正面},\\ 0, \omega=\text{反面}.\end{cases}$$

则有

$$P\{X=1\}=P\{\text{出现正面}\}=\frac{1}{2},$$

$$P\{X=0\}=P\{\text{出现反面}\}=\frac{1}{2}.$$

于是 X 的概率分布为

X	1	0
P	$\frac{1}{2}$	$\frac{1}{2}$

或表示为 $X=\begin{cases}1, & \frac{1}{2},\\ 0, & \frac{1}{2}.\end{cases}$

例 2.2　设离散型随机变量 X 的概率分布为

(1) $P\{X=i\}=a\cdot\left(\frac{2}{3}\right)^i, i=1,2,3$；

(2) $P\{X=i\}=a\cdot\left(\frac{2}{3}\right)^i, i=1,2,\cdots$.

分别求上述各式中的常数 a.

解　(1) 由于

$$1=\sum_{i=1}^{3}P\{X=i\}=\sum_{i=1}^{3}a\cdot\left(\frac{2}{3}\right)^i=a\cdot\frac{38}{27},$$

故有 $a=\frac{27}{38}$.

(2) 由于

$$1=\sum_{i=1}^{\infty}P\{X=i\}=\sum_{i=1}^{\infty}a\cdot\left(\frac{2}{3}\right)^i=a\cdot\frac{\frac{2}{3}}{1-\frac{2}{3}}=2a,$$

故有 $a=\frac{1}{2}$.

一旦知道一个离散型随机变量 X 的概率分布$\{p(x_i)\}$，我们便可求得 X 所生成的任何事件的概率. 特别地，对任意 $a\leqslant b$，有

$$P\{a\leqslant X\leqslant b\}=P\Big(\bigcup_{a\leqslant x_i\leqslant b}\{X=x_i\}\Big)=\sum_{a\leqslant x_i\leqslant b}P\{X=x_i\}$$

$$=\sum_{a\leqslant x_i\leqslant b}p(x_i).\qquad(2.4)$$

一般地,若 I 是一个区间,则

$$P\{X\in I\}=\sum_{x_i\in I}p(x_i). \tag{2.5}$$

例 2.3 设 X 的概率分布由例 2.2(1)给出,则有:

$P\{X<1\}=0.$

$$P\{X\leqslant 1\}=P\{X=1\}=\frac{27}{38}\times\frac{2}{3}=\frac{9}{19},$$

$$P\{X<2\}=P\{X=1\}=\frac{9}{19},$$

$$P\{X<2.5\}=P\{X=1\}+P\{X=2\}=\frac{9}{19}+\frac{27}{38}\times\frac{4}{9}=\frac{15}{19},$$

$$P\{X\leqslant 3\}=P\{X=1\}+P\{X=2\}+P\{X=3\}=1,$$

$$P\{X\leqslant 4\}=P\{X=1\}+P\{X=2\}+P\{X=3\}=1.$$

三、分布函数

离散型随机变量的概率分布为离散型随机变量的统计规律提供了一目了然的描述.然而对那些取值非可数的随机变量,比如,测量的误差、灯泡的寿命等,我们如果同离散型随机变量一样,通过罗列取每一个值及其相应的概率来描述它们会遇到不可克服的困难.其一,这类随机变量的非可数个取值无法一一列举出来;其二,从下面的例子可看到取连续值的随机变量,它取某个特定值的概率往往是 0.不过,对取连续值的随机变量,我们往往关心的是它的取值落在一定范围(比如区间或区间的并)的概率,而不关心它取某个特定值的概率.因此,对这类随机变量,我们希望能够对其取值落于任何一个区间上的概率给出描述.

例 2.4 等可能地在数轴上的有界区间 $[a,b]$ 上投点,记 X 为落点的位置(数轴上的坐标),则 X 是样本空间 $\Omega=[a,b]$ 上的函数

$$X(\omega)=\omega,\omega\in[a,b]. \tag{2.6}$$

根据几何概型,对任意 $c\in[a,b]$,有

$$P\{X=c\}=P\{\omega=c\}=0.$$

而对任意 $B=(c,d]\subset[a,b]$,有

$$P\{c<X\leqslant d\}=P\{\text{落点在 } B \text{ 中}\}=\frac{d-c}{b-a}. \tag{2.7}$$

另一方面,由于

$$P\{c<X\leqslant d\}=P\{X\leqslant d\}-P\{X\leqslant c\} \tag{2.8}$$

为给出 X 取值于任意区间上的概率,实际上只要给出所有 X 取值于形如 $(-\infty,x]$ 的区间上的概率 $P\{X\leqslant x\}$ 即可.记

$$F(x)=P\{X\leqslant x\}.$$

当 x 取遍 $(-\infty,+\infty)$ 上的一切实数时，$F(x)$ 便成为定义在 $(-\infty,+\infty)$ 上的函数. 一旦知道了这个函数 $F(x)$，我们便可得到相应的随机变量取值于任何区间的概率.

上述讨论中的(2.8)式适合于任意随机变量，为此，我们引入下列定义：

定义 2.4　设 X 是一随机变量，则称函数

$$F(x)=P\{X\leqslant x\},x\in(-\infty,+\infty) \tag{2.9}$$

为随机变量 X 的**分布函数**，记作 $X\sim F(x)$.

例 2.5　求例 2.4 中的随机变量 X 的分布函数.

解　当 $x<a$ 时，$\{X\leqslant x\}$ 是不可能事件，于是

$$F(x)=P\{X\leqslant x\}=0.$$

当 $a\leqslant x<b$ 时，由于 $\{X\leqslant x\}=\{a\leqslant X\leqslant x\}$，且 $[a,x]\subset[a,b]$，于是

$$F(x)=P\{X\leqslant x\}=P\{a\leqslant X\leqslant x\}=\frac{x-a}{b-a}.$$

当 $x\geqslant b$ 时，由于 $\{X\leqslant x\}=\{a\leqslant X\leqslant b\}$，于是

$$F(x)=P\{X\leqslant x\}=P\{a\leqslant X\leqslant b\}=\frac{b-a}{b-a}=1.$$

综上，可得 X 的分布函数为

$$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leqslant x<b,\\ 1, & x\geqslant b.\end{cases}$$

根据定义 2.4，可以推导出一个随机变量的分布函数必然满足下列性质：

(1) 单调性　若 $x_1<x_2$，则 $F(x_1)\leqslant F(x_2)$；

(2) $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$，

$F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$；

(3) 右连续性　$F(x+0)=F(x)$.

上述性质中，单调性容易证明，留作练习. 其他性质的证明超出了本书的要求，略去.

另一方面，如果一个函数 $F(x)$ 满足上述三条性质，则可以证明，它一定可以作为某一随机变量 X 的分布函数，因此，通常将满足上述三条性质的函数都称为分布函数.

四、离散型随机变量的分布函数

一个离散型随机变量的分布也可由分布函数来描述，事实上其概率分布与分布函数能够相互确定.

例 2.6 设 X 由例 2.1 给出，求其分布函数.

解 X 只有两个可能取值，其概率分布为

$$P\{X=0\}=P\{X=1\}=\frac{1}{2}.$$

于是，当 $x<0$ 时，

$$F(x)=P\{X\leqslant x\}=0.$$

当 $0\leqslant x<1$ 时，

$$F(x)=P\{X\leqslant x\}=P\{X=0\}=\frac{1}{2}.$$

当 $x\geqslant 1$ 时，

$$F(x)=P\{X\leqslant x\}=P\{X=0\}+P\{X=1\}=\frac{1}{2}+\frac{1}{2}=1.$$

综上，X 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ \frac{1}{2}, & 0\leqslant x<1,\\ 1, & x\geqslant 1.\end{cases}$$

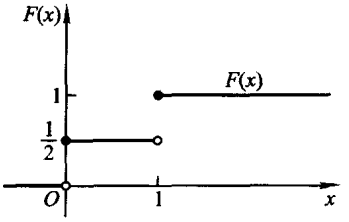


图 2.1 例 2.6 的分布函数

如图 2.1，观察上例中的 X 的分布函数，我们发现 $F(x)$是一个阶梯形的函数，它在 X 的可能取值点 0,1 处发生跳跃，跳跃高度等于相应点处的概率，而在两个相邻跳跃点之间分布函数值保持不变. 这一特征实际上是所有离散型随机变量的共同特征. 反过来，如果一个随机变量 X 的分布函数 $F(x)$ 是阶梯型函数，则 X 一定是一个离散型随机变量，其概率分布可由分布函数 $F(x)$惟一确定：$F(x)$的跳跃点全体构成 X 的所有可能取值，每一跳跃点处的跳跃高度则是 X 在相应点处的概率.

例 2.7 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<1,\\ \frac{9}{19}, & 1\leqslant x<2,\\ \frac{15}{19}, & 2\leqslant x<3,\\ 1, & x\geqslant 3.\end{cases}$$

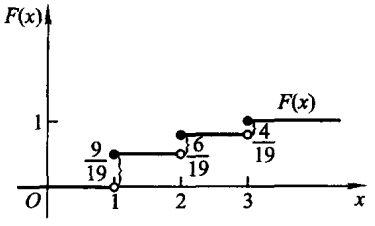


图 2.2 例 2.7 的分布函数

由于 $F(x)$是一个阶梯型函数，故知 X 是一个离散型随机变量，$F(x)$的跳跃点分别为 1,2,3，对应的跳跃高度分别为$\frac{9}{19}$，$\frac{6}{19}$，$\frac{4}{19}$，如图 2.2. 故 X 的概率分布为

$$P\{X=1\}=\frac{9}{19},P\{X=2\}=\frac{6}{19},P\{X=3\}=\frac{4}{19}.$$

五、连续型随机变量及其概率密度

在例2.4中,我们得到了在$[a,b]$上等可能投点时,落点的位置X的分布函数,现在,我们换一个角度来考虑.由于X在$[a,b]$上等可能取值,而X在$[a,b]$上取值的概率为1,我们可将这一概率1视为均匀分布在$[a,b]$的每一点上,为此可求得概率1在区间$[a,b]$上的平均值$\frac{1}{b-a}$,这个平均值称为X在$[a,b]$上的概率密度.这里,由于X在$[a,b]$上等可能取值,因而X在$[a,b]$上的每一点有相同的密度$\frac{1}{b-a}$.而且易见,对任意$[c,d]\subset[a,b]$,有

$$P\{c\leqslant X\leqslant d\}=\frac{d-c}{b-a}=\int_c^d\frac{1}{b-a}\mathrm{d}x.$$

由于$[a,b]$以外的点是X不可能取值点,记其密度为0,令

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b,\\ 0, & \text{其他},\end{cases}\tag{2.10}$$

称$f(x)$为X的概率密度函数.容易验证,X的分布函数可表示为

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t.$$

对一般情形,我们引入下列定义:

定义2.5　一个随机变量X称为**连续型随机变量**,如果存在一个非负可积函数$f(x)$,使得

$$F(x)=P\{X\leqslant x\}=\int_{-\infty}^{x}f(t)\mathrm{d}t,\tag{2.11}$$

并称$f(x)$为X的**概率密度函数**,简称为**密度函数**.

根据定义2.5及分布函数的性质,易知密度函数具有下列性质:

(1) $f(x)\geqslant 0\quad x\in(-\infty,+\infty)$;

(2) $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$.

反过来,可以证明,一个函数满足上述两个性质,一定可以作为某一连续型随机变量的密度函数.

对于一个给定的连续型随机变量X,如果已知其密度函数,根据定义2.5,自然可以求得其分布函数.同时,可以通过密度函数的积分来求X的取值落于任意区间上的概率

$$P\{a<X\leqslant b\}=F(b)-F(a)=\int_a^b f(x)\mathrm{d}x. \tag{2.12}$$

在几何上，$F(x)$和 $P\{a<X\leqslant b\}$可以用相应的曲边梯形的面积来表示(如图 2.3).

由(2.12)式还可得知，对任意实数 x，有

$$P\{X=x\}=0. \tag{2.13}$$

此外，由(2.11)式知，在 $f(x)$的连续点处，有

$$F'(x)=f(x). \tag{2.14}$$

它为我们从分布函数出发确定密度函数提供了途径.

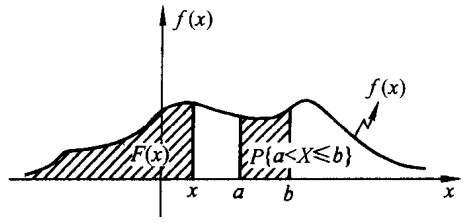


图 2.3 密度函数

例 2.8 设 X 是在$[a,b]$上等可能投点的位置，在例 2.5 中，我们已求得 X 的分布函数，试由分布函数求其密度函数.

解 在例 2.5 中，我们已求出其分布函数为

$$F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leqslant x<b,\\ 1, & x\geqslant b.\end{cases}$$

于是其密度函数为

$$f(x)=F'(x)=\begin{cases}0, & x<a,\\ \dfrac{1}{b-a}, & a<x<b,\\ 0, & x>b.\end{cases} \tag{2.15}$$

在 $x=a$ 和 $x=b$ 处，$F(x)$的导数不存在，可补充定义这两点的密度为

$$f(x)=\frac{1}{b-a},\ x=a \text{ 或 } x=b. \tag{2.16}$$

(原则上，补充定义 $f(a)$和 $f(b)$为其他非负值也是可以的，因为改变密度函数个别点处的值不影响其在区间上的积分值.)

上面求得的结果，与(2.10)一致：随机变量 X 在$[a,b]$上的密度为一个常数$\dfrac{1}{b-a}$，而在$[a,b]$以外的密度为 0，我们称这样的随机变量在$[a,b]$上服从**均匀分布**，即指 X 取$[a,b]$上的每一点是“等可能的”，这正好是对例 2.4 中定义的 X 的“等可能性”的一个更为明确的表述.

§2.2　随机变量的数字特征

一、离散型随机变量的数学期望

我们知道，离散型随机变量的统计规律可以由其概率分布完全描述，但在许多实际问题中，这种完全的描述并不使人感到方便，为了某些目的，我们希望用一些数字指标——**数字特征**来反映随机变量的统计规律的某些层面，数学期望就是其中之一，它主要反映随机变量取值的平均水平.

例如，观察一名射手 20 次射击的成绩如下：

中靶环数(x_i)	0	1	2	3	4	5	6	7	8	9	10
频数(n_i)	1	2	1	2	3	3	2	1	2	2	1
频率(f_i)	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

人们常常使用“平均中靶环数”来对射手的射击水平作出综合评价，记平均中靶环数为 $\bar{x}$，则有

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=0}^{10} x_i \cdot n_i}{n} = \sum_{i=0}^{10} x_i \cdot f_i \\ &= 0\times\frac{1}{20} + 1\times\frac{2}{20} + 2\times\frac{1}{20} + \cdots + 9\times\frac{2}{20} + 10\times\frac{1}{20} \\ &= 5.\end{aligned}$$

我们知道，当试验次数增大时，频率的稳定值就是概率，那么完整描述该射手真实水平的是其射中各环数的概率分布，相应地，观察到的平均中靶环数 $\bar{x}$，随试验次数增大也将趋于一个稳定值. 设中靶环数 X(观察之前为随机变量)的概率分布为

$$P\{X=i\} = p_i, \quad i=0,1,2,\cdots,10,$$

则 $\bar{x}$ 的稳定值为

$$\sum_{i=0}^{10} x_i p_i.$$

它是对该射手的真实水平的综合评价.

定义 2.6　若离散型随机变量 X 的可能值为 x_i ($i=1,2,\cdots$)，其概率分布为

$$P\{X=x_i\}=p_i, i=1,2,\cdots,$$

则当

$$\sum_{i=1}^{\infty}|x_i|p_i<\infty \tag{2.17}$$

时，称 $\sum_{i=1}^{\infty}x_ip_i$ 为随机变量 X 的**数学期望**（简称**期望**），也称为 X 的**均值**. 记作 $E(X)$，为了简明起见，也可记作 EX.

在上述定义中，既然数学期望定义为 $EX=\sum_{i=1}^{\infty}x_ip_i$，那么只要 $\sum_{i=1}^{\infty}x_ip_i$ 收敛就可以了，为什么还要有条件

$$\sum_{i=1}^{\infty}|x_i|p_i<\infty,$$

是不是有点多余呢？我们知道，离散型随机变量可依某种次序一一列举，对同一随机变量，它的取值的列举次序可以不同，但当改变次序时，其数学期望不应该改变，这意味着改变 $\sum_{i=1}^{\infty}x_ip_i$ 的求和次序，其收敛性及和不应改变，为此必须要求 $\sum_{i=1}^{\infty}x_ip_i$ 绝对收敛.

例 2.9 设盒中有 5 个球，其中 2 个白球，3 个黑球，从中随意抽取 3 个球. 记 X 为抽取到的白球数，求 EX.

解 X 只可能取 0,1,2 这三个实数值，而且根据古典概型计算，有

$$P\{X=0\}=\frac{C_3^3}{C_5^3}=\frac{1}{10},$$

$$P\{X=1\}=\frac{C_3^2C_2^1}{C_5^3}=\frac{6}{10},$$

$$P\{X=2\}=\frac{C_3^1C_2^2}{C_5^3}=\frac{3}{10}.$$

于是

$$EX=0\times\frac{1}{10}+1\times\frac{6}{10}+2\times\frac{3}{10}=1.2.$$

二、连续型随机变量的数学期望

在微积分中，我们知道对离散取值（可数或有限个）的函数——数列，我们可以考虑其和，而对连续取值（即值域为一个区间）的函数，我们不能简单地考虑函数值的和，这时函数的和的概念被扩充为定积分. 回顾函数在 $[a,b]$ 上的定积分的定义，我们知道定积分实际上是和式的极限，受此启发，我们来考虑连续型随

机变量的数学期望.

设 X 是连续型随机变量,密度函数为 $f(x)$,为简单起见,设 $f(x)$只在有限区间$[a,b]$上取不为零的值,即对一切 $x\notin[a,b]$,有 $f(x)=0$.取分点

$$a=x_0<x_1<\cdots<x_{n+1}=b,$$

则 X 落在区间 $\Delta x_i=(x_i,x_{i+1})$中的概率为

$$P\{X\in\Delta x_i\}=\int_{x_i}^{x_{i+1}}f(x)\mathrm{d}x. \tag{2.18}$$

当 Δx 很小时,

$$P\{X\in\Delta x_i\}\approx f(x_i)\Delta x_i. \tag{2.19}$$

这时,概率分布

x_i	x_0	x_1	$\cdots$	x_n
p_i	$f(x_0)\Delta x_0$	$f(x_1)\Delta x_1$	$\cdots$	$f(x_n)\Delta x_n$

可视为 X 的离散近似,服从上述分布的离散型随机变量的数学期望为

$$\sum_{i=0}^{n}x_if(x_i)\Delta x_i. \tag{2.20}$$

它近似地表达了连续型随机变量 X 的平均值——数学期望,当分点越来越密时,近似会越来越好.根据定积分的定义,上述和式(2.20)以定积分

$$\int_a^b xf(x)\mathrm{d}x=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$$

为极限,于是这一定积分的值便是 X 的精确的数学期望值.如果 X 在无穷区间上取值,上述定义还应扩充到无穷区间上的积分,这里就不再讨论了.一般地,我们给出下列定义:

定义 2.7　若 X 为连续型随机变量,$f(x)$为其密度函数,如果

$$\int_{-\infty}^{+\infty}|x|f(x)\mathrm{d}x<\infty,$$

则称 $\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$ 为随机变量 X 的**数学期望**(简称**期望**),也称为**均值**.记作 $E(X)$,也可记作 EX.

例 2.10　设随机变量 X 的密度函数为

$$f(x)=\begin{cases}\dfrac{2}{\pi}\cos^2 x, & |x|\leqslant\dfrac{\pi}{2},\\ 0, & \text{其他}.\end{cases}$$

求 EX.

解　因为 $f(x)$只在有限区间$\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$上不为 0,且在该区间上为连续函

数,所以 EX 存在,且

$$EX = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x\cos^2 x\mathrm{d}x,$$

根据奇函数的性质知,$EX=0$.

例 2.11 设随机变量 X 的密度函数为

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

求 EX.

解 显然 EX 存在,且

$$EX = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_0^1 x^2\mathrm{d}x + \int_1^2 x(2-x)\mathrm{d}x$$
$$= \frac{x^3}{3}\Big|_0^1 + \left(x^2 - \frac{1}{3}x^3\right)\Big|_1^2 = 1.$$

三、随机变量函数的数学期望

在许多实际问题中,我们除了对某一随机变量进行研究以外,往往还要对该随机变量的函数进行研究.设 X 是一个随机变量,$g(x)$是任意实函数,则 X 与 $g(x)$复合成

$$Y=g(X).$$

显然 Y 也是一个随机变量,但值得指出的是 $g(x)$本身是一个确定性的函数,即 X 与 Y 的关系是确定性的,这意味着当 X 取定某值时,Y 的值将由函数关系 $g(x)$惟一确定.正因为此,Y 的随机性完全由 X 的随机性所决定,进而,Y 的分布原则上由 X 的分布所确定.然而,从 X 的分布出发求 Y 的分布并非易事.关于随机变量函数的分布,我们将在本章 §2.4 中讨论,这里,我们先讨论随机变量函数的数学期望.$Y=g(X)$作为一个随机变量,其数学期望按定义应根据其分布来计算,然而下面的结论表明,我们不必求 Y 的分布,而根据 X 的分布即可求其函数 $Y=g(X)$的数学期望,这显然为计算随机变量函数的数学期望提供了极大的方便.

定理 2.1 设 X 是一个随机变量,$g(x)$是一个实函数.

(1) 若 X 为离散型随机变量,概率分布为

$$P\{X=x_i\}=p_i, i=1,2,\cdots.$$

且 $\sum\limits_{i=1}^{\infty} |g(x_i)|p_i<\infty$,则 $Eg(X)$存在,且

$$Eg(X)=\sum_{i=1}^{\infty} g(x_i)p_i. \tag{2.21}$$

(2) 若 X 为连续型随机变量，$f(x)$是其密度函数，且 $\int_{-\infty}^{+\infty}|g(x)|f(x)\mathrm{d}x<\infty$，则 $Eg(X)$存在，且

$$Eg(X)=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x. \tag{2.22}$$

证明　(这里仅就离散型情形给以证明，连续型的证明已超过本书要求，读者可与离散型对比来理解相应结果.)

令 $Y=g(X)$，则 Y 仍然是一个离散型随机变量，设其可能取值为 y_j，$j=1,2,\cdots$. 于是

$$\begin{aligned}P\{Y=y_j\}&=P\Big(\bigcup_{i:\,g(x_i)=y_j}\{X=x_i\}\Big)\\&=\sum_{i:\,g(x_i)=y_j}P\{X=x_i\}.\end{aligned}$$

由数学期望定义有

$$\begin{aligned}Eg(X)&=\sum_{j=1}^{\infty}y_jP\{Y=y_j\}\\&=\sum_{j=1}^{\infty}y_j\sum_{i:\,g(x_i)=y_j}P\{X=x_i\}\\&=\sum_{j=1}^{\infty}\sum_{i:\,g(x_i)=y_j}g(x_i)P\{X=x_i\}\\&=\sum_{i=1}^{\infty}g(x_i)P\{X=x_i\}.\end{aligned}$$

例 2.12　设 X 如例2.9，即 X 的概率分布如右表，求 $E(X-EX)^2$.

X	0	1	2
P	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$

解　由例2.9已解得 $EX=1.2$，于是根据定理2.1(1)有

$$\begin{aligned}E(X-EX)^2&=(0-1.2)^2\times\frac{1}{10}+(1-1.2)^2\times\frac{6}{10}+(2-1.2)^2\times\frac{3}{10}\\&=1\ 44\times\frac{1}{10}+0.04\times\frac{6}{10}+0.64\times\frac{3}{10}\\&=0.36.\end{aligned}$$

例 2.13　设 X 的密度函数如例2.11，即

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

求 $E|X-EX|$.

解　由例2.11已算得 $EX=1$，于是由定理2.1(2)有

$$
\begin{aligned}
E|X-EX| = E|X-1| &= \int_{-\infty}^{+\infty} |x-1| f(x)\mathrm{d}x \\
&= \int_0^1 |x-1| x\mathrm{d}x + \int_1^2 |x-1|(2-x)\mathrm{d}x \\
&= \int_0^1 (1-x)x\mathrm{d}x + \int_1^2 (x-1)(2-x)\mathrm{d}x \\
&= \int_0^2 (1-x)x\mathrm{d}x + 2\int_1^2 (x-1)\mathrm{d}x \\
&= \frac{1}{3}.
\end{aligned}
$$

四、数学期望的性质

将数学期望的一些基本性质概括出来可能会为许多问题的考虑提供方便，下面列举的性质，对离散型和连续型随机变量而言，是定理 2.1 的直接推论，但事实上，它们对一般的随机变量也都成立.

(1) 对任意常数 a，有 $Ea = a$；

(2) 设 α_1, α_2 为任意实数，$g_1(x), g_2(x)$ 为任意实函数，如果 $Eg_1(X)$，$Eg_2(X)$ 均存在，则

$$E[\alpha_1 g_1(X) + \alpha_2 g_2(X)] = \alpha_1 Eg_1(X) + \alpha_2 Eg_2(X); \tag{2.23}$$

(3) 如果 EX 存在，则对任意实数 a，有

$$E(X+a) = EX + a. \tag{2.24}$$

证明 (这里仅对离散型给出证明，连续型的证明与之类似，读者可以自行证明.)

(1) 取 $g(x) \equiv a$，由定理 2.1 即得.

(2) 首先，由

$$|\alpha_1 g_1(x) + \alpha_2 g_2(x)| \leqslant |\alpha_1||g_1(x)| + |\alpha_2||g_2(x)|$$

可知

$$\sum_i |\alpha_1 g_1(x_i) + \alpha_2 g_2(x_i)| p_i \leqslant |\alpha_1| \sum_i |g_1(x_i)| p_i + |\alpha_2| \sum_i |g_2(x_i)| p_i.$$

因为 $Eg_1(X), Eg_2(X)$ 存在，故由上式得知 $E[\alpha_1 g_1(X) + \alpha_2 g_2(X)]$ 存在. 令 $g(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x)$，代入(2.21)即可证得(2.23)成立.

(3) 在(2)中取 $g_1(x) = x, g_2(x) \equiv a$，即得(3).

例 2.14 设 EX, EX^2 均存在，证明

$$E(X-EX)^2 = EX^2 - (EX)^2. \tag{2.25}$$

证明 因为 $(X-EX)^2 = X^2 - 2X \cdot EX + (EX)^2$，于是由(2.23)得

$$E(X-EX)^2 = E[X^2 - 2X \cdot EX + (EX)^2]$$

$$= EX^2 - 2EX \cdot EX + (EX)^2 = EX^2 - (EX)^2.$$

五、随机变量的方差

随机变量的数学期望是对随机变量取值水平的综合评价，在许多问题中，我们还需要了解随机变量的其他特征. 比如，在投资决策中，我们选择投资某一项目或购买某种资产(如股票、债券等)，不仅关心其未来的收益水平，还关心其未来收益的不确定性程度，前者通常用数学期望来度量，后者通常称为风险程度，有许多种衡量方法，最简单、直观的方法就是用方差来度量. 一个随机变量的方差，粗略地讲，反映随机变量偏离其中心——数学期望的平均偏离程度.

定义 2.8　设 X 为一个随机变量，其数学期望 EX 存在，则称 $X - EX$ 为 X 的**离差**，进一步，如果 $E(X - EX)^2$ 也存在，则称 $E(X - EX)^2$ 为随机变量 X 的**方差**，记作 $D(X)$ 或 $\mathrm{Var}X$，为简明起见，也可记作 DX，并称 $\sqrt{DX}$ 为 X 的**标准差**.

由于 X 是一个随机变量，其离差 $X - EX$ 因而也是一个随机变量，平均来讲，X 的正、负离差相互抵消，因而 $E(X - EX) = 0$. 为考虑 X 对 EX 的偏离程度，我们必须消除符号的影响，为此用 $(X - EX)^2$ 来衡量 X 对 EX 的偏离，从而方差 $DX = E(X - EX)^2$ 即为 X 对 EX 的平均偏离. 当然为消除离差中的符号，也可考虑使用绝对离差 $|X - EX|$，但由于 $E|X - EX|$ 中绝对值不便于处理，所以人们习惯于使用方差来作为随机变量偏离其期望的偏离程度的度量.

根据随机变量函数的期望的计算方法(定理 2.1)，在(2.21)和(2.22)中，取 $g(x) = (x - EX)^2$，即可计算方差：

若 X 为离散型随机变量，其概率分布为 $P\{X = x_i\} = p_i$，$i = 1, 2, \cdots$，则

$$DX = E(X - EX)^2 = \sum_i (x_i - EX)^2 p_i; \tag{2.26}$$

若 X 为连续型随机变量，$f(x)$ 为其密度函数，则

$$DX = E(X - EX)^2 = \int_{-\infty}^{+\infty} (x - EX)^2 f(x)\mathrm{d}x. \tag{2.27}$$

此外，由例 2.14，也可通过(2.25)来计算方差，即

$$DX = EX^2 - (EX)^2. \tag{2.28}$$

由数学期望的性质及方差的定义，容易导出方差的一些基本性质：

设 X 的方差 DX 存在，a 为任意常数，则

(1) $Da = 0$; (2.29)

(2) $D(X + a) = DX$; (2.30)

(3) $D(aX) = a^2 DX$. (2.31)

例 2.15　设 X 的分布如例 2.12，例 2.12 实际上已算得 X 的方差 $DX =$

$E(X-EX)^2=0.36$,其中 $EX=1.2$ 在例 2.9 中算得.现在我们改用(2.28)式来计算方差:

$$EX^2=0^2\times\frac{1}{10}+1^2\times\frac{6}{10}+2^2\times\frac{3}{10}=1.8,$$

$$DX=EX^2-(EX)^2=1.8-(1.2)^2=0.36.$$

例 2.16 设 X 的密度函数如例 2.11,即

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

求 DX.

解 在例 2.11 中,我们已算得 $EX=1$,又

$$EX^2=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\int_0^1x^3\mathrm{d}x+\int_1^2x^2(2-x)\mathrm{d}x$$

$$=\frac{x^4}{4}\Big|_0^1+\left(\frac{2}{3}x^3-\frac{x^4}{4}\right)\Big|_1^2=\frac{7}{6},$$

从而

$$DX=EX^2-(EX)^2=\frac{7}{6}-1=\frac{1}{6}.$$

例 2.17 X 为一随机变量,方差存在,令

$$l(C)=E(X-C)^2. \tag{2.32}$$

证明:当且仅当 $C=EX$ 时,$l(C)$达到最小值,此时最小值为 DX.

证明 $l(C)=E(X-C)^2=E[(X-EX)+(EX-C)]^2$

$$=E[(X-EX)^2+2(X-EX)(EX-C)+(EX-C)^2]$$

$$=E(X-EX)^2+(EX-C)^2$$

$$\geqslant E(X-EX)^2=DX.$$

显然,当且仅当 $C=EX$ 时,最后一个不等式的等号成立,故 $l(C)$在 $C=EX$ 时达到最小值,且最小值为 DX.

例 2.17 实际上给出了数学期望的预测含义:随机变量 X 的取值是不确定的,如果我们企图用一个常数 C 来对 X 作点值预测,未来 X 的实际取值与该预测值 C 存在偏差 $X-C$,平均意义下的偏差程度用 $E(X-C)^2$(通常称之为**均方误差**)来衡量.人们当然希望预测的偏差程度越小越好,于是最好的预测 C 应该使得 $E(X-C)^2$ 达到最小.例 2.17 实际上指出 $C=EX$ 是均方误差最小意义下 X 的最好的点值预测,且最小的均方误差为 DX.

六、随机变量的矩与切比雪夫不等式

数学期望和方差可以纳入一个更一般的概念范畴之中,那就是随机变量的

矩.

定义 2.9　X 为一随机变量，k 为正整数，如果 EX^k 存在(即 $E|X|^k<\infty$)，则称 EX^k 为 X 的 k **阶原点矩**，称 $E|X|^k$ 为 X 的 k **阶绝对矩**.

定理 2.2　随机变量 X 的 t 阶矩存在，则其 s 阶矩($0<s<t$)也存在.

证明　设 X 为连续型(离散型类似)，其密度函数为 $f(x)$，则

$$\begin{aligned}E|X|^s &= \int_{|x|^s\leqslant 1}|x|^s f(x)\mathrm{d}x+\int_{|x|^s>1}|x|^s f(x)\mathrm{d}x\\ &\leqslant \int_{|x|^s\leqslant 1} f(x)\mathrm{d}x+\int_{|x|^s>1}|x|^t f(x)\mathrm{d}x\\ &\leqslant P\{|X|^s\leqslant 1\}+E|X|^t<\infty.\end{aligned}$$

推论　设 k 为正整数，C 为常数，如果 EX^k 存在，则 $E(X+C)^k$ 存在，特别地，$E(X-EX)^k$ 存在.

定义 2.10　X 为一随机变量，k 为正整数，如果 EX^k 存在，则称 $E(X-EX)^k$ 为 X 的 k **阶中心矩**，称 $E|X-EX|^k$ 为 X 的 k **阶绝对中心矩**.

显然，数学期望 EX 是 X 的一阶原点矩，方差 $DX=E(X-EX)^2$ 是 X 的二阶中心矩. 而且，根据定理 2.2 及其推论知：如果 $EX^2<\infty$，则 X 的数学期望和方差均存在.

接下来，我们介绍一类矩的不等式.

定理 2.3　设 $h(x)$ 是 x 的一个非负函数，X 是一个随机变量，且 $Eh(X)$ 存在，则对任意 $\varepsilon>0$，有

$$P\{h(X)\geqslant\varepsilon\}\leqslant\frac{Eh(X)}{\varepsilon}. \tag{2.33}$$

证明　(这里仅证明连续型，离散型类似可证.)

设 X 的密度函数为 $f(x)$，则

$$\begin{aligned}Eh(X)&=\int_{-\infty}^{+\infty}h(x)f(x)\mathrm{d}x\\ &=\int_{h(x)\geqslant\varepsilon}h(x)f(x)\mathrm{d}x+\int_{h(x)<\varepsilon}h(x)f(x)\mathrm{d}x\\ &\geqslant\int_{h(x)\geqslant\varepsilon}h(x)f(x)\mathrm{d}x\\ &\geqslant\varepsilon\int_{h(x)\geqslant\varepsilon}f(x)\mathrm{d}x\\ &=\varepsilon P\{h(x)\geqslant\varepsilon\}.\end{aligned}$$

于是可得(2.33)成立.

推论 1(马尔可夫不等式)　设 X 的 k 阶矩存在，即 $E|X|^k<\infty$，则对任意 $\varepsilon>0$ 有

$$P\{|X|\geqslant\varepsilon\}\leqslant\frac{E|X|^k}{\varepsilon^k}. \tag{2.34}$$

推论 2(切比雪夫不等式) 设 X 的方差存在,则对任意 $\varepsilon>0$ 有

$$P\{|X-EX|\geqslant\varepsilon\}\leqslant\frac{DX}{\varepsilon^2}. \tag{2.35}$$

推论 3 随机变量 X 的方差为 0 当且仅当存在一个常数 a,使得 $P\{X=a\}=1$.

证明 充分性显然,下证必要性.

首先注意到

$$\{|X-EX|>0\}=\bigcup_{n=1}^{\infty}\{|X-EX|>\frac{1}{n}\},$$

从而有

$$P\{|X-EX|>0\}=P\left(\bigcup_{n=1}^{\infty}\{|X-EX|>\frac{1}{n}\}\right)$$

$$\leqslant\sum_{n=1}^{\infty}P\{|X-EX|>\frac{1}{n}\}.$$

由于 $DX=0$,对一切 $n\geqslant1$,据切比雪夫不等式,有

$$P\{|X-EX|>\frac{1}{n}\}\leqslant\frac{DX}{1/n^2}=0.$$

从而得

$$P\{|X-EX|>0\}=0.$$

所以

$$P\{|X-EX|=0\}=1,$$

即推论 3 得证,且常数 a 为 EX.

§2.3 常用的离散型分布

一、退化分布

在所有分布中,最简单的分布是退化分布.一个随机变量 X 以概率 1 取某一常数,即

$$P\{X=a\}=1,$$

则称 X 服从 a 处的**退化分布**.由定理 2.3 的推论 3 知,X 服从退化分布的充要条件是 $DX=0$.且若 X 服从 a 处的退化分布,则 $EX=a$.

退化分布之所以称为退化分布是因为其取值几乎是确定的,即这样的随机

变量退化成了一个确定的常数.

二、两点分布

另一个简单分布是两点分布.一个随机变量只有两个可能取值,设其分布为

$$P\{X=x_1\}=p, P\{X=x_2\}=1-p, 0<p<1, \tag{2.36}$$

则称 X 服从 x_1, x_2 处参数为 p 的**两点分布**.容易求得

$$EX=px_1+(1-p)x_2, \tag{2.37}$$

$$DX=p(1-p)(x_1-x_2)^2. \tag{2.38}$$

特别地,如果 X 服从 $x_1=1, x_2=0$ 处的参数为 p 两点分布,即

$$P\{X=1\}=p, P\{X=0\}=1-p, 0<p<1. \tag{2.39}$$

通常称为 X 服从参数为 p 的两点分布或称 X 服从参数为 p 的**0－1分布**,也称 X 是参数为 p 的**伯努利随机变量**.此时

$$EX=p, DX=p(1-p). \tag{2.40}$$

在实际中,一个两点分布的随机变量通常根据某试验中一特定事件 A 的发生与否构造出来.例如,设 $P(A)=p, P(\bar{A})=1-p$,则随机变量

$$X(\omega)=\begin{cases} x_1, \omega\in A(\text{即 } A \text{ 发生}), \\ x_2, \omega\in \bar{A}(\text{即 } A \text{ 不发生}) \end{cases} \tag{2.41}$$

便服从 x_1, x_2 处参数为 p 的两点分布.特别地在一次试验中,观察 A 是否发生,记 A 发生的次数为 X,则 X 要么取值为1,要么取值为0,于是 X 服从参数为 p 的0－1分布.

三、n 个点上的均匀分布

有一类特殊的随机变量,它共有 n 个不同的可能取值,且取每一个值的可能性相同,即有

$$P\{X=x_i\}=\frac{1}{n}, i=1,2,\cdots,n, \tag{2.42}$$

则称 X 服从 n 个点 $\{x_1, x_2, \cdots, x_n\}$ 上的**均匀分布**.

容易算得:

$$EX=\frac{1}{n}\sum_{i=1}^{n}x_i \xlongequal{\text{def}} \bar{x}, \tag{2.43}$$

$$DX=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2. \tag{2.44}$$

我们可以将古典概型与均匀分布联系起来.古典概型中,试验共有 n 个不同的可能结果,且每个结果出现的可能性相同,设 $\Omega=\{\omega_1, \omega_2, \cdots, \omega_n\}$,则

$P\{\omega_i\}=\frac{1}{n}(i=1,2,\cdots,n)$. 如果随机变量 X 是 Ω 上的一一对应的函数,那么 X 便服从均匀分布.一个简单的例子是:设 X 表示投掷一枚均匀的骰子出现的点数,此时 $\Omega=\{1,2,\cdots,6\}$,令

$$X(\omega)=\omega,\omega\in\Omega,$$

则 X 服从$\{1,2,\cdots,6\}$上的均匀分布.

四、二项分布

在 n 重伯努利试验中,每次试验中事件 A 发生的概率为 $p(0<p<1)$,记 X 为 n 次试验中事件 A 发生的次数,则 X 的可能取值为 $0,1,2,\cdots,n$. 且对每一个 $k,k=0,1,2,\cdots,n$,事件$\{X=k\}$即为"在 n 次试验中事件 A 恰好发生 k 次",根据伯努利概型,有

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},k=0,1,2,\cdots,n. \tag{2.45}$$

一般地,如果一个随机变量 X 的概率分布由(2.45)给出,则称 X 服从参数为 n,p 的**二项分布**,并记作 $X\sim b(n,p)$,且记 $b(k;n,p)=C_n^k p^k(1-p)^{n-k}$. 事实上 $b(k;n,p)(k=0,1,2,\cdots,n)$是二项式$(p+q)^n$ 的展开式中各项(其中 $q=1-p$),这也正是上述分布称为"二项分布"的原因.

显然,当 $n=1$ 时,二项分布 $b(1,p)$实际上就是参数为 p 的 0-1 分布.

例 2.18 一个袋子中装有 N 个球,其中 N_1 个白球,N_2 个黑球($N_1+N_2=N$)每次从中任取一球,查看完其颜色后再放回去,一共取 n 次,求取到的白球数 X 的分布.

解 每次取球视为一次试验,n 次取球视为 n 重伯努利试验,每次取球,取到白球的概率为 $p=\frac{N_1}{N}$,故 $X\sim b\left(n,\frac{N_1}{N}\right)$,其分布为

$$P\{X=k\}=C_n^k\left(\frac{N_1}{N}\right)^k\left(\frac{N_2}{N}\right)^{n-k},0\leqslant k\leqslant n. \tag{2.46}$$

接下来,我们来求二项分布的数学期望和方差:

$$\begin{aligned}EX&=\sum_{k=0}^{n}kb(k;n,p)=\sum_{k=1}^{n}k\cdot C_n^k p^k q^{n-k}\\&=np\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}q^{(n-1)-(k-1)}\\&=np\sum_{k=0}^{n-1}C_{n-1}^{k}p^{k}q^{(n-1)-k}\\&=np\sum_{k=0}^{n-1}b(k;n-1,p)\\&=np.\end{aligned} \tag{2.47}$$

其中最后一个等式，是因为 $b(k;n-1,p)(0\leqslant k\leqslant n-1)$ 是以 $n-1$ 和 p 为参数的二项分布. 由类似的计算可得

$$\begin{aligned} EX^2 &= \sum_{k=0}^{n} k^2 b(k;n,p) \\ &= \sum_{k=0}^{n} [k(k-1)+k]b(k;n,p) \\ &= \sum_{k=2}^{n} k(k-1)b(k;n,p) + \sum_{k=0}^{n} kb(k;n,p) \\ &= n(n-1)p^2 \sum_{k=2}^{n} b(k-2;n-2,p) + np \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} b(k;n-2,p) + np \\ &= n(n-1)p^2 + np = n^2p^2 + npq, \end{aligned} \tag{2.48}$$

于是

$$DX = EX^2 - (EX)^2 = n^2p^2 + npq - n^2p^2 = npq. \tag{2.49}$$

五、几何分布

在独立重复试验中，事件 A 发生的概率为 p，设 X 为直到 A 发生为止所进行的试验的次数，显然 X 的可能取值是全体正整数，且由定理 1.4 知其分布为

$$P\{X=k\} = q^{k-1}p \overset{\text{def}}{=\!=} g(k,p), k\geqslant 1. \tag{2.50}$$

由于 $g(k,p)=q^{k-1}p$ 是一个几何数列（或称等比数列），因而将以(2.50)为概率分布的随机变量称为服从参数为 p 的**几何分布**.

几何分布的数学期望和方差的计算可利用无穷级数求和的常规方法——转化为求幂级数的和函数，下面给出计算结果，计算过程留作练习.

$$EX = \sum_{n=1}^{\infty} npq^{n-1} = \frac{1}{p}, \tag{2.51}$$

$$EX^2 = \sum_{n=1}^{\infty} n^2 pq^{n-1} = \frac{2q}{p^2} + \frac{1}{p}, \tag{2.52}$$

$$DX = EX^2 - (EX)^2 = \frac{q}{p^2}. \tag{2.53}$$

例 2.19　设 X 服从几何分布，则对任何两个正整数 m，n，有

$$P\{X>m+n \mid X>m\} = P\{X>n\}. \tag{2.54}$$

证明　由 $P\{X>m+n \mid X>m\} = \dfrac{P\{X>m+n\}}{P\{X>m\}}$，据(2.50)知

$$P\{X>m\} = \sum_{k=m+1}^{\infty} q^{k-1}p = q^m \sum_{j=1}^{\infty} q^{j-1}p = q^m,$$

同理,有

$$P\{X>m+n\}=q^{m+n},P\{X>n\}=q^{n}.$$

于是得

$$P\{X>m+n\mid X>m\}=\frac{q^{m+n}}{q^{m}}=q^{n}=P\{X>n\}.$$

式(2.54)通常称为几何分布的**无记忆性**,意指几何分布对过去的 m 次失败的信息在后面的计算中被遗忘了.事实上还可以证明:一个取正整数值的随机变量,如果具有无记忆性,即满足(2.54),则一定服从几何分布.可见,无记忆性实际上是几何分布的一个特征性质.

六、超几何分布

一个袋子中装有 N 个球,其中 N_1 个白球,N_2 个黑球($N=N_1+N_2$),从中不放回地抽取 n 个球,X 表示取到白球的数目,那么 X 的分布容易根据古典概型计算(参考例 1.12)得到:

$$P\{X=k\}=\frac{C_{N_1}^{k}C_{N_2}^{n-k}}{C_{N}^{n}},0\leqslant k\leqslant n. \tag{2.55}$$

这里约定:当 $a<b$ 时,$C_a^b=0$.以(2.55)为概率分布的随机变量通常称为服从**超几何分布**.

比较例 2.18 与上述取球问题,惟一的区别是:前者是放回的,后者是不放回的.在实际中,当 N 很大时,且 N_1 和 N_2 均较大,而 n 相对很小时,通常将不放回近似地当作放回来处理,从而用二项分布(见例 2.18)作为超几何分布的近似,即

$$\frac{C_{N_1}^{k}C_{N_2}^{n-k}}{C_{N}^{n}}\approx C_{n}^{k}\left(\frac{N_1}{N}\right)^{k}\left(\frac{N_2}{N}\right)^{n-k}. \tag{2.56}$$

这一近似关系的严格数学表述是:当 $N\to\infty$ 时,$N_1\to\infty$,$N_2\to\infty$,且 $\frac{N_1}{N}\to p$,$\frac{N_2}{N}\to 1-p$,则对任意给定的 n 和 k,有

$$\lim_{N\to\infty}\frac{C_{N_1}^{k}C_{N_2}^{n-k}}{C_{N}^{n}}=C_{n}^{k}p^{k}(1-p)^{n-k}. \tag{2.57}$$

下面给出超几何分布的数学期望和方差,略去其计算过程:

$$EX=n\cdot\frac{N_1}{N}, \tag{2.58}$$

$$D(X)=n\cdot\frac{N_1}{N}\cdot\frac{N_2}{N}\cdot\frac{N-n}{N-1}. \tag{2.59}$$

七、泊松(Poisson)分布

如果一个随机变量 X 的概率分布为

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots \tag{2.60}$$

其中 $\lambda>0$ 为参数,则称 X 服从参数为 λ 的**泊松分布**,记作 $X\sim P(\lambda)$.

注意到概率分布(2.60)中 $\frac{\lambda^k}{k!}$ 实际上是 e^{λ}(关于 λ 的函数)的幂级数展开式的一般项,因而有

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\cdot e^{-\lambda}=e^{\lambda}\cdot e^{-\lambda}=1.$$

这说明(2.60)确实是一个概率分布.此外容易算得

$$EX=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=\lambda. \tag{2.61}$$

由同样的方法可得

$$EX^2=\lambda^2+\lambda.$$

从而有

$$DX=EX^2-(EX)^2=\lambda. \tag{2.62}$$

泊松分布是实际中经常遇到的一类分布,比如,电话交换台在一给定时间内收到用户的呼叫次数,售票口到达的顾客人数,保险公司在一给定时期内被索赔的次数等,均可近似地用泊松分布来描述.

例 2.20　某商店根据过去的销售记录知道某种商品每月的销售量可以用参数为 $\lambda=10$ 的泊松分布来描述,为了以 95% 以上的概率保证不脱销,问商店在月底应存多少件该种商品(设只在月底进货)?

解　设该商店每月销售该商品的件数为 X,月底存货为 a 件,则当 $X\leqslant a$ 时就不会脱销,据题意,要求 a 使得

$$P\{X\leqslant a\}\geqslant 0.95.$$

由于已知 X 服从参数为 $\lambda=10$ 的泊松分布,上式即为

$$\sum_{k=0}^{a}\frac{10^k}{k!}e^{-10}\geqslant 0.95,$$

由附录的泊松分布表知

$$\sum_{k=0}^{14}\frac{10^k}{k!}e^{-10}\approx 0.916\ 6<0.95,$$

$$\sum_{k=0}^{15}\frac{10^k}{k!}e^{-10}\approx 0.951\ 3>0.95.$$

于是,这家商店只要在月底保证存货不低于 15 件就能以 95% 以上的概率保证下个月该种商品不会脱销.

下面的定理给出了二项分布与泊松分布间的近似关系.

定理 2.4(泊松定理) 在 n 重伯努利试验中,事件 A 在每次试验中发生的概率为 p_n(注意这与试验的次数 n 有关),如果 $n\to\infty$ 时,$np_n\to\lambda$($\lambda>0$ 为常数),则对任意给定的 k,有

$$\lim_{n\to\infty} b(k;n,p_n)=\frac{\lambda^k}{k!}e^{-\lambda}. \tag{2.63}$$

该定理的证明略去.由该定理,我们可以将二项分布用泊松分布来近似:当二项分布 $b(n,p)$的参数 n 很大,而 p 很小时,可以将它用参数为 $\lambda=np$ 的泊松分布来近似,即有

$$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np}. \tag{2.64}$$

例 2.21 纺织厂女工照顾 800 个纺锭,每一纺锭在某一短时间内发生断头的概率为 0.005(设短时间内最多只发生一次断头).求在这段时间内总共发生的断头次数超过 2 的概率.

解 设 X 为 800 个纺锭在该段时间内发生的断头次数.则 $X\sim b(800,0.005)$,它可近似于参数为 $\lambda=800\times0.005=4$ 的泊松分布,从而有

$$\begin{aligned}P\{0\leqslant X\leqslant 2\}&=\sum_{k=0}^{2}P\{X=k\}=\sum_{k=0}^{2}b(k;800,500)\\&\approx\sum_{k=0}^{2}\frac{4^k}{k!}e^{-4}=e^{-4}\left(1+4+\frac{16}{2!}\right)\approx 0.238\ 1,\end{aligned}$$

从而

$$P\{X>2\}=1-P\{0\leqslant X\leqslant 2\}\approx 1-0.238\ 1=0.761\ 9.$$

§2.4 常用的连续型分布

一、均匀分布

均匀分布是连续型分布中最简单的一种分布.它用来描述一个随机变量在一个区间上取每一个值的可能性均等的分布规律.它是离散型情形 n 个点上的均匀分布在连续型情形的推广.在离散型情形,n 个点的均匀分布描述为随机变量在 n 个点中取值,取每一点的概率相同.在连续型情形,$[a,b]$上的均匀分布,则描述为该区间上每一点的概率密度相同,或等价地描述为取值落在该区间中每一个子区间上的概率与子区间的长度成正比,在例 2.5 及例 2.8 中,我们实际上已给出了详细的讨论.在第一章中讨论的几何概型,如果用随机变量 X 来描述落点的位置,则 X 服从均匀分布.下面明确给出均匀分布的定义.

一个随机变量 X,如果其密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b,\\ 0, & \text{其他},\end{cases} \tag{2.65}$$

则称 X 服从$[a,b]$上的**均匀分布**,记作 $X\sim U[a,b]$.

$[a,b]$上的均匀分布的惟一特征是密度函数在$[a,b]$以外为 0,而在$[a,b]$上为常数.根据密度函数的性质:

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_a^b f(x)\mathrm{d}x=1.$$

这个常数必然是$\dfrac{1}{b-a}$.

由(2.65)容易导出均匀分布的分布函数为

$$F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\leqslant x\leqslant b,\\ 1, & x\geqslant b,\end{cases} \tag{2.66}$$

且可算得其数学期望和方差分别为

$$EX=\frac{a+b}{2}, \tag{2.67}$$

$$DX=\frac{(b-a)^2}{12}. \tag{2.68}$$

二、指数分布

指数分布通常用来描述对某一事件发生的等待时间,比如,乘客在公共汽车站等车的时间,灯泡的使用寿命(等待用坏的时间),电话交换台收到两次呼叫的时间间隔.在离散型分布中,我们知道,几何分布用来描述伯努利试验中,直到某事件 A 发生为止共进行的试验次数.如果将每次试验视为经历一个单位时间,那么直到事件 A 发生为止进行的试验的次数可视为直到 A 发生为止的等待时间(离散时间).在这个意义上,指数分布可视为离散型情形的几何分布在连续型情形的推广.下面给出指数分布的定义.

一个随机变量 X,如果其密度函数为

$$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0,\end{cases} \tag{2.69}$$

其中 $\lambda>0$ 为参数,则称 X 服从参数为λ 的**指数分布**,记作 $X\sim \mathrm{e}(\lambda)$.

由(2.69)容易导出 X 的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0,\end{cases} \tag{2.70}$$

且可算得其数学期望和方差分别为

$$EX=\frac{1}{\lambda}, \tag{2.71}$$

$$DX=\frac{1}{\lambda^2}. \tag{2.72}$$

与几何分布类似，**无记忆性**是连续型随机变量的指数分布的特征性质，即有下列结果：

定理 2.5 非负连续型随机变量 X 服从指数分布的充要条件是：对任意正实数 r 和 s，有

$$P\{X>r+s\mid X>s\}=P\{X>r\}. \tag{2.73}$$

该定理的必要性的证明留作练习，充分性的证明超出了本书的要求，略去.

例 2.22 某元件的寿命 X 服从指数分布，已知其平均寿命为 1 000 h，求 3 个这样的元件使用 1 000 h，至少已有一个损坏的概率.

解 由题设知，$EX=1\ 000$ h，于是该指数分布的参数为

$$\lambda=\frac{1}{EX}=\frac{1}{1\ 000}.$$

从而 X 的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-\frac{x}{1\ 000}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

由此得

$$P\{X>1\ 000\}=1-P\{X\leqslant 1\ 000\}=1-F(1\ 000)=\mathrm{e}^{-1}.$$

各元件的寿命是否超过 1 000 h 是独立的，于是 3 个元件使用 1 000 h 都未损坏的概率为 e^{-3}，从而至少有一个已损坏的概率为 $1-\mathrm{e}^{-3}$.

三、正态分布

一个连续型随机变量 X，如果其密度函数为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<\infty, \tag{2.74}$$

其中 μ,σ 为常数，且 $\sigma>0$. 则称 X 服从参数为 μ 和 σ^2 的**正态分布**，记作 $X\sim N(\mu,\sigma^2)$.

首先，根据泊松积分

$$\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x=\sqrt{\pi} \tag{2.75}$$

不难验证 $\int_{-\infty}^{+\infty}\varphi(x)\mathrm{d}x=1$，并可算得 $N(\mu,\sigma^2)$ 的数学期望和方差（留作练习）为

$$EX=\mu, DX=\sigma^2. \tag{2.76}$$

可见，正态分布的两个参数实际上分别为其数学期望和方差.

其次，让我们考察一下正态分布的密度函数 $\varphi(x)$ 的特征，如图 2.4 所示，$\varphi(x)$ 具有钟型的图像，且以 x 轴为渐近线；关于 $x=\mu$ 对称，在 $x=\mu$ 处达到函数最大值 $\frac{1}{\sqrt{2\pi}\sigma}$.

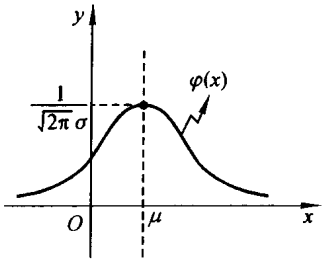


图 2.4　正态分布的密度函数

当 $\mu=0,\sigma^2=1$ 时，即 $X\sim N(0,1)$，称 X 服从**标准正态分布**，其密度函数记作 $\varphi_0(x)$，即

$$\varphi_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \tag{2.77}$$

正态分布的“钟型”特征与实际中很多随机变量的“中间大，两头小”的分布规律相吻合. 比如考察一群人的身高，个体的身高作为一个随机变量，其分布的特点是，在平均身高附近的人较多，特别高和特别矮的人较少. 一个班的一次考试成绩、测量误差等均有类似的特征. 正态分布是概率论中最重要的分布，高斯在研究误差理论时曾用它来刻画误差，所以很多文献中亦称之为**高斯分布**. 进一步的理论研究表明，一个变量如果受到大量的独立因素的影响(无主导因素)，则它一般服从正态分布，这一问题将在下一章进行讨论.

1. 正态分布的分布函数

对许多随机变量，利用密度函数可求出分布函数的表达式. 然而，由于正态分布的密度函数 $\varphi(x)$ 的原函数没有解析表达式，因而其分布函数(记作 $\Phi(x)$)

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)\mathrm{d}t=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t \tag{2.78}$$

不能表示为解析式. 对于任意给定 x，需要利用数值计算方法来求上述反常积分的近似值. 然而对每一个具体的正态分布及每个给定 x 都进行近似计算显然极为不便，解决这一问题的方法是，将标准正态分布的分布函数(记作 $\Phi_0(x)$)在各点的值计算出来制成表，在实际计算时可直接查表. 而一般正态分布，则可经过适当变换转化为标准正态分布来计算.

2. 标准正态分布表

在附录中列出了标准正态分布的密度函数值表和分布函数值表，但表中只列出 $x\geqslant 0$ 时 $\varphi_0(x)$ 和 $\Phi_0(x)$ 的值，这是因为由正态分布的对称性(严格地讲，密度函数 $\varphi_0(x)$ 关于 $x=0$ 对称!)可以导出 $\varphi_0(x)$ 和 $\Phi_0(x)$ 在 $x<0$ 时的值.

对于 $\varphi_0(x)$ 而言，直接由其对称性有

$$\varphi_0(-x)=\varphi_0(x).$$

因而，当 $x<0$ 时，$\varphi_0(x)=\varphi_0(-x)$，在表中查 $\varphi_0(-x)$ 即得 $\varphi_0(x)$.

对于 $\Phi_0(x)$，由于 $\varphi_0(x)$ 关于 $x=0$ 对称，因而有

$$\Phi_0(-x)=\int_{-\infty}^{-x}\varphi_0(t)\mathrm{d}t=\int_{x}^{+\infty}\varphi_0(t)\mathrm{d}t=1-\Phi_0(x) \tag{2.79}$$

或写作

$$\Phi_0(x)+\Phi_0(-x)=1. \tag{2.80}$$

特别地，有 $\Phi_0(0)=0.5$. 当 $x<0$ 时，由 $\Phi_0(x)=1-\Phi_0(-x)$，查表得 $\Phi_0(-x)$，即可得 $\Phi_0(x)$.

例 2.23 设 $X\sim N(0,1)$，(1) 求 $P\{X\leqslant 1.96\}$，$P\{X\leqslant -1.96\}$，$P\{|X|\leqslant 1.96\}$，$P\{-1<X\leqslant 2\}$. (2) 已知 $P\{X\leqslant a\}=0.701\ 9$，$P\{|x|<b\}=0.924\ 2$，$P\{X\leqslant c\}=0.298\ 1$，求 a,b,c.

解 (1) 直接查表可得

$$P\{X\leqslant 1.96\}=\Phi_0(1.96)=0.975,$$

根据 $\varphi_0(x)$的对称性，有

$$P\{X\leqslant -1.96\}=\Phi_0(-1.96)=1-\Phi_0(1.96)=1-0.975=0.025,$$

$$\begin{aligned}P\{|X|\leqslant 1.96\}&=P\{-1.96\leqslant X\leqslant 1.96\}=\Phi_0(1.96)-\Phi_0(-1.96)\\&=2\Phi_0(1.96)-1=2\times 0.975-1=0.95,\end{aligned}$$

$$\begin{aligned}P\{-1<X\leqslant 2\}&=\Phi_0(2)-\Phi_0(-1)=\Phi_0(2)-[1-\Phi_0(1)]\\&=\Phi_0(2)+\Phi_0(1)-1\\&=0.977\ 25+0.841\ 3-1=0.818\ 55,\end{aligned}$$

(2) 直接查表可得 $a=0.53$. 由

$$P\{|X|<b\}=2\Phi_0(b)-1=0.924\ 2$$

得

$$\Phi_0(b)=\frac{1}{2}(1+0.924\ 2)=0.962\ 1,$$

查表即得 $b\approx 1.78$.

由于 $P\{X\leqslant c\}=0.298\ 1<0.5$，所以 $c<0$，根据对称性，有

$$\Phi_0(-c)=1-\Phi_0(c)=0.701\ 9,$$

查表得 $-c=0.53$，$c=-0.53$.

3. 一般正态分布与标准正态分布的关系

我们先看一个有关正态分布的一般性结论.

定理 2.6 设 $X\sim N(\mu,\sigma^2)$，$Y=aX+b$，a,b 为常数，且 $a\neq 0$，则 $Y\sim N(a\mu+b,a^2\sigma^2)$.

证明 记 Y 的分布函数为 $F_Y(y)$，密度函数为 $f_Y(y)$，X 的分布函数为 $\Phi(x)$，密度函数为 $\varphi(x)$，则有

$$F_Y(y)=P\{Y\leqslant y\}=P\{aX+b\leqslant y\}. \tag{2.81}$$

当 $a>0$ 时，

$$F_Y(y)=P\left\{X\leqslant\frac{y-b}{a}\right\}=\Phi\left(\frac{y-b}{a}\right),\tag{2.82}$$

$$f_Y(y)=F_Y'(y)=\frac{1}{a}\Phi'\left(\frac{y-b}{a}\right)=\frac{1}{a}\varphi\left(\frac{y-b}{a}\right).\tag{2.83}$$

当 $a<0$ 时，

$$F_Y(y)=P\left\{X\geqslant\frac{y-b}{a}\right\}=1-\Phi\left(\frac{y-b}{a}\right),\tag{2.84}$$

$$f_Y(y)=F_Y'(y)=-\frac{1}{a}\Phi'\left(\frac{y-b}{a}\right)=-\frac{1}{a}\varphi\left(\frac{y-b}{a}\right).\tag{2.85}$$

综上，有

$$\begin{aligned}f_Y(y)&=\frac{1}{|a|}\varphi\left(\frac{y-b}{a}\right)=\frac{1}{\sqrt{2\pi}|a|\sigma}\mathrm{e}^{-\frac{\left[\frac{y-b}{a}-\mu\right]^2}{2\sigma^2}}\\&=\frac{1}{\sqrt{2\pi}|a|\sigma}\mathrm{e}^{-\frac{[y-(a\mu+b)]^2}{2\sigma^2a^2}},\end{aligned}\tag{2.86}$$

因而 $Y\sim N(a\mu+b,a^2\sigma^2)$.

推论 1　如果 $X\sim N(\mu,\sigma^2)$，则 $\xi=\dfrac{X-\mu}{\sigma}\sim N(0,1)$.

ξ 通常称为 X 的**标准化**.

推论 2　$X\sim N(\mu,\sigma^2)$ 的充要条件是存在一个随机变量 $\xi\sim N(0,1)$，使得 $X=\sigma\xi+\mu$.

推论 3　设 $X\sim N(\mu,\sigma^2)$，$\Phi_0(x)$，$\varphi(x)$ 分别为其分布函数与密度函数，$\Phi_0(x)$，$\varphi_0(x)$ 是标准正态分布的分布函数和密度函数，则有

$$\Phi(x)=\Phi_0\left(\frac{x-\mu}{\sigma}\right),\tag{2.87}$$

$$\varphi(x)=\frac{1}{\sigma}\varphi_0\left(\frac{x-\mu}{\sigma}\right).\tag{2.88}$$

推论 1 和推论 2 留给读者作为练习. 推论 3 的证明，可利用推论 2 及定理 2.6的证明中的(2.82)和(2.83).

4. 一般正态分布的概率计算

前面讨论了一般正态分布与标准正态分布的关系，特别是定理 2.6 的推论 1，为一般正态分布的概率计算提供了有效的途径. 一般地，对于一般正态分布的有关问题，尤其是概率计算，都可以转化为标准正态分布来解决. 下面通过一些例子加以说明.

例 2.24　已知 $X\sim N(8,0.5^2)$，求(1) $\Phi(9)$，$\Phi(7)$；(2) $P\{7.5\leqslant X\leqslant 10\}$；(3) $P\{|X-8|\leqslant 1\}$；(4) $P\{|X-9|<0.5\}$.

解 (1) $\Phi(9)=P\{X\leqslant 9\}=P\left\{\dfrac{X-8}{0.5}\leqslant\dfrac{9-8}{0.5}\right\}=P\left\{\dfrac{X-8}{0.5}\leqslant 2\right\}$

$=\Phi_0(2)=0.977\ 25;$

$\Phi(7)=P\{X\leqslant 7\}=P\left\{\dfrac{X-8}{0.5}\leqslant\dfrac{7-8}{0.5}\right\}=P\left\{\dfrac{X-8}{0.5}\leqslant -2\right\}$

$=\Phi_0(-2)=1-\Phi_0(2)=0.022\ 75;$

(2) $P\{7.5\leqslant X\leqslant 10\}=P\left\{\dfrac{7.5-8}{0.5}\leqslant\dfrac{X-8}{0.5}\leqslant\dfrac{10-8}{0.5}\right\}=P\left\{-1\leqslant\dfrac{X-8}{0.5}\leqslant 4\right\}$

$=\Phi_0(4)-\Phi_0(-1)=\Phi_0(4)+\Phi_0(1)-1$

$=0.9^46833+0.841\ 3-1\approx 0.841\ 3;$

(3) $P\{|X-8|\leqslant 1\}=P\left\{\left|\dfrac{X-8}{0.5}\right|<2\right\}=2\Phi_0(2)-1=2\times 0.977\ 25-1$

$=0.954\ 5;$

(4) $P\{|X-9|<0.5\}=P\{8.5<X<9.5\}=P\left\{1<\dfrac{X-8}{0.5}<3\right\}$

$=\Phi_0(3)-\Phi_0(1)=0.9^286\ 50-0.841\ 3\approx 0.157\ 3.$

例 2.25 某种型号电池的寿命 X 近似服从正态分布 $N(\mu,\sigma^2)$,已知其寿命在 250 小时以上的概率和寿命不超过 350 小时的概率均为 92.36%,为使其寿命在 $\mu-x$ 和 $\mu+x$ 之间的概率不小于 0.9,x 至少为多大?

解 由 $P\{X>250\}=P\{X<350\}$,根据密度函数关于 $x=\mu$ 对称,有 $\mu=\dfrac{350+250}{2}=300$. 又由

$$P\{X<350\}=P\left\{\frac{X-300}{\sigma}<\frac{350-300}{\sigma}\right\}=\Phi_0\left(\frac{50}{\sigma}\right)=0.923\ 6,$$

查表得 $\dfrac{50}{\sigma}\approx 1.43$,于是 $\sigma\approx 35$. 故 $X\sim N(300,35^2)$. 又

$$P\{\mu-x<X<\mu+x\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<\frac{x}{\sigma}\right\}=2\Phi_0\left(\frac{x}{\sigma}\right)-1\geqslant 0.9,$$

即 $\Phi_0\left(\dfrac{x}{\sigma}\right)\geqslant\dfrac{1.9}{2}=0.95$,查表得 $\dfrac{x}{\sigma}\geqslant 1.645$,于是 $x\geqslant 1.645\times 35=57.58$.

§2.5 随机变量函数的分布

一、随机变量的函数

在上一节中,我们在讨论正态分布时,对服从 $N(\mu,\sigma^2)$ 的随机变量 X,引进

了线性变换 $Y=aX+b, a\neq 0$. 事实上，Y 是随机变量 X 的函数，对 X 的每一个取值，Y 有一个惟一取值与之对应. 由于 X 是随机变量，其取值是不确定的，因而 Y 的取值也随之不确定，即 Y 也是一个随机变量.

一般地，如果存在一个函数 $g(x)$，使得随机变量 X, Y 满足

$$Y=g(X), \tag{2.89}$$

则称随机变量 Y 是随机变量 X 的函数. 在微积分中，也讨论变量间的函数关系，但在那里，我们主要研究函数关系中的确定性特征，比如因变量随自变量变化而变化的变化率——导数. 在概率论中，我们主要研究的是随机变量函数的随机性特征，即由自变量的统计规律出发研究因变量的统计规律性. 如何从自变量 X 的统计规律导出其函数 $Y=g(X)$ 的统计规律呢？一般地，对任意区间（或区间的并）B，令 $C=\{x \mid g(x)\in B\}$，则

$$\{Y\in B\}=\{g(X)\in B\}=\{X\in C\}, \tag{2.90}$$

从而

$$P\{Y\in B\}=P\{g(X)\in B\}=P\{X\in C\}. \tag{2.91}$$

(2.91)说明，X 的统计规律确实决定了 Y 的统计规律.

例 2.26　设 X 是一随机变量，且 $Y=X^2$，则对任意 $x\geqslant 0$，有

$$P\{Y\leqslant x\}=P\{X^2\leqslant x\}=P\{-\sqrt{x}\leqslant X\leqslant\sqrt{x}\}, \tag{2.92}$$

$$P\{Y=x\}=P\{X^2=x\}=P(\{x=\sqrt{x}\}\cup\{x=-\sqrt{x}\}). \tag{2.93}$$

为对随机变量 X 的函数 Y 的统计规律进行完整描述，我们希望从 X 的分布出发导出 Y 的分布，(2.91)式实际上为此提供了可行的途径. 本节仍然只对离散型和连续型这两种情形进行讨论.

二、离散型随机变量函数的分布

离散型随机变量 X 的函数 $Y=g(X)$ 显然还是离散型随机变量，因此，我们希望由 X 的概率分布出发导出 Y 的概率分布. 先看一个简单的例子.

例 2.27　设随机变量 X 的分布为

$$P\{X=-1\}=\frac{1}{4}, P\{X=0\}=\frac{1}{2}, P\{X=1\}=\frac{1}{4}.$$

求 $Y=X^2$ 的分布.

解　我们注意到 Y 的可能取值为 0,1，根据(2.90)，有

$$\begin{aligned}
P\{Y=0\}&=P\{X^2=0\}=P\{X=0\}=\frac{1}{2},\\
P\{Y=1\}&=P\{X^2=1\}=P(\{X=1\}\cup\{X=-1\})\\
&=P\{X=1\}+P\{X=-1\}\\
&=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}.
\end{aligned}$$

上面的例子虽然简单,却反映了求离散型随机变量函数的概率分布的一般方法:先根据自变量 X 的可能取值确定因变量 Y 的所有可能取值,然后对 Y 的每一个可能取值 $y_i(i=1,2,\cdots)$ 确定相应的 $C_i=\{x_j \mid g(x_j)=y_i\}$,于是正如(2.90)和(2.91)一样,有

$$\{Y=y_i\}=\{g(X)=y_i\}=\{X\in C_i\}, \tag{2.94}$$

$$P\{Y=y_i\}=P\{X\in C_i\}=\sum_{x_j\in C_i}P\{X=x_j\}, \tag{2.95}$$

从而求得 Y 的概率分布.此外,上述过程还说明 Y 的概率分布完全由 X 的概率分布所确定.

三、连续型随机变量函数的分布

一般说来,连续型随机变量的函数不一定是连续型随机变量.这时讨论随机变量函数的分布的目标是导出其分布函数,这里主要讨论连续型随机变量的函数还是连续型的情形,这时希望不仅求出随机变量函数的分布函数,而且还希望求出其密度函数.不过只要求出分布函数,密度函数也就不成问题了.我们将要说明导出随机变量函数的分布函数的一般方法,并介绍一些典型的随机变量函数的例子.

在上一节中讨论正态分布时,我们引入了服从正态分布 $N(\mu,\sigma^2)$ 的随机变量 X 的线性变换 $Y=aX+b$,定理 2.6 导出了 Y 的分布仍是正态分布,参数分别为 $a\mu+b$ 和 $a^2\sigma^2$.实际上其推导过程体现了求连续型随机变量函数的分布的一般方法.

一般地,已知 X 的分布函数 $F_X(x)$ 或密度函数 $f_X(x)$,为求 $Y=g(X)$ 的分布函数,正如(2.91)所示

$$F_Y(x)=P\{Y\leqslant x\}=P\{g(X)\leqslant x\}=P\{X\in C_x\}, \tag{2.96}$$

其中 $C_x=\{t \mid g(t)\leqslant x\}$.

而 $P\{X\in C_x\}$ 往往可由 X 的分布函数 $F_X(x)$ 来表达或用其密度函数 $f_X(x)$ 的积分来表达:

$$P\{X\in C_x\}=\int_{C_x}f_X(t)\mathrm{d}t. \tag{2.97}$$

进而,Y 的密度函数,可直接从 $F_Y(x)$ 导出.

例 2.28 设 X 是一个连续型随机变量,其分布函数 $F(x)$ 是严格单调递增的,则 $F(X)$ 服从[0,1]上的均匀分布.

证明 由于 $F(x)$ 是严格单调递增函数,因而其反函数存在,记作 $F^{-1}(x)$.设 $Y=F(X)$ 的分布函数为 $F_Y(x)$,则

$$F_Y(x)=P\{Y\leqslant x\}=P\{F(X)\leqslant x\}.$$

由于 $F(x)$的值域为$[0,1]$,故当 $x<0$ 时,$F_Y(x)=P\{F(X)\leqslant x\}=0$,当 $x>1$ 时,$F_Y(x)=P\{F(X)\leqslant x\}=1$,而当 $0\leqslant x\leqslant 1$ 时,由于 $F(x)$严格单调递增,故有

$F_Y(x)=P\{F(X)\leqslant x\}=P\{X\leqslant F^{-1}(x)\}=F(F^{-1}(x))=x$.综上,

$$F_Y(x)=\begin{cases}0, & x<0,\\ x, & 0\leqslant x\leqslant 1,\\ 1, & x>1.\end{cases}$$

即 $Y=F(X)$服从$[0,1]$上的均匀分布.

例 2.28 中的结论对 $F(x)$非严格单调递增的情形同样成立,证明略有不同,这里就不讨论了.

例 2.29($\chi^2(1)$分布)　设 $X\sim N(0,1)$,求 $Y=X^2$ 的密度函数.

解　记 Y 的分布函数为 $F_Y(x)$,则 $F_Y(x)=P\{Y\leqslant x\}=P\{X^2\leqslant x\}$.显然,当 $x<0$ 时,

$$F_Y(x)=P\{X^2\leqslant x\}=0,$$

当 $x\geqslant 0$ 时,

$$F_Y(x)=P\{X^2\leqslant x\}=P\{-\sqrt{x}<X<\sqrt{x}\}=2\Phi_0(\sqrt{x})-1.$$

从而 $Y=X^2$ 的分布函数为

$$F_Y(x)=\begin{cases}2\Phi_0(\sqrt{x})-1, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

由

$$F'_Y(x)=\begin{cases}\dfrac{1}{\sqrt{x}}\varphi_0(\sqrt{x}), & x>0,\\ 0, & x\leqslant 0\end{cases}$$

于是其密度函数为

$$F_Y(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}\tag{2.98}$$

以(2.98)为密度函数的随机变量称为服从 $\chi^2(1)$**分布**,它是一类更广泛的分布——自由度为 n 的 χ^2 分布——$\chi^2(n)$在 $n=1$ 时的特例.关于 $\chi^2(n)$分布的细节将在第四章中给出.

例 2.30(对数正态分布)　随机变量 X 称为服从参数为μ,σ^2 的**对数正态分布**,如果 $Y=\ln X$ 服从正态分布 $N(\mu,\sigma^2)$.试求对数正态分布的密度函数.

解　由于 $Y=\ln X\sim N(\mu,\sigma^2)$,等价地有 $X=e^Y$,$Y\sim N(\mu,\sigma^2)$,于是,当 $x>0$ 时,

$$F_X(x)=P\{X\leqslant x\}=P\{e^Y\leqslant x\}=P\{Y\leqslant \ln x\}=\Phi(\ln x),$$

当 $x\leqslant 0$ 时，显然 $F_X(x)=0$. 继而可得 X 的密度函数为

$$f_X(x)=F_X'(x)=\begin{cases}\dfrac{1}{x}\varphi(\ln x), & x>0,\\ 0, & x\leqslant 0\end{cases}$$

$$=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

在实际中，通常用对数正态分布来描述价格的分布，特别是在金融市场的理论研究中，如著名的期权定价公式——Black－Scholes 公式，以及许多实证研究都用对数正态分布来描述金融资产的价格. 设某种资产当前价格为 P_0，考虑单期投资问题，到期时该资产的价格为一个随机变量，记作 P_1，设投资于该资产的连续复合收益率为 r，则有

$$P_1=P_0e^r,$$

从而

$$r=\ln\frac{P_1}{P_0}=\ln P_1-\ln P_0.$$

注意到 P_0 为当前价格，是已知常数，因而假设价格 P_1 服从对数正态分布实际上等价于假设连续复合收益率 r 服从正态分布.

例 2.31(对数正态分布的矩) 设 X 服从参数为 μ 和 σ^2 的对数正态分布，即 $Y=\ln X\sim N(\mu,\sigma^2)$，则由 $X=e^Y$，有

$$EX^k=Ee^{kY}=\int_{-\infty}^{+\infty}e^{ky}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y-\mu)^2-2k\sigma^2(y-\mu)+k^2\sigma^4}{2\sigma^2}+k\mu+\frac{k^2\sigma^2}{2}\right]dy=e^{k\mu+\frac{k^2\sigma^2}{2}}.$$

特别地，$k=1$ 时，

$$EX=e^{\mu+\frac{\sigma^2}{2}}.$$

进而有

$$DX=EX^2-(EX)^2=e^{2\mu+2\sigma^2}-e^{2\mu+\sigma^2}=e^{2\mu+\sigma^2}(e^{\sigma^2}-1).$$

习 题 二

(A)

1. 甲、乙两人分别拥有赌本 30 元和 20 元，他们利用投掷一枚均匀硬币进行赌博，约定

如果出现正面，甲赢 10 元、乙输 10 元．如果出现反面，则甲输 10 元、乙赢 10 元，分别用随机变量表示投掷一次后甲、乙两人的赌本，并求其概率分布和分布函数，画出分布函数的图形．

2．离散型随机变量 X 的概率分布为：

(1) $P\{X=i\}=a2^i, i=1,2,\cdots,100$；

(2) $P\{X=i\}=2a^i, i=1,2,\cdots,$

分别求(1)、(2)中 a 的值．

3．设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<-5,\\ \frac{1}{5}, & -5\leqslant x<-2,\\ \frac{3}{10}, & -2\leqslant x<0,\\ \frac{1}{2}, & 0\leqslant x<2,\\ 1, & x\geqslant 2.\end{cases}$$

求 X 的概率分布

4．一批产品共 10 件，其中 7 件正品，3 件次品，每次从中任取一件，求下面两种情形下直到取到正品为止所需抽取次数的概率分布．(1) 每次取出后再放回去；(2) 每次取出后不放回．

5．设 X 的分布函数如第 3 题所示，求下列概率：$P\{X>-3\}$，$P\{|X|<3\}$，$P\{|X+1|>2\}$．

6．设 X 的分布函数为

$$F(x)=\begin{cases}0, & x\leqslant 0,\\ Ax^2, & 0<x\leqslant 1,\\ 1, & x>1.\end{cases}$$

求 A 及 $P\{0.5<X\leqslant 0.8\}$．

7．函数 $F(x)=\dfrac{1}{1+x^2}$ 是否可作为某一随机变量的分布函数，如果

(1) $-\infty<x<+\infty$；　(2) $0<x<+\infty$，其他场合适当定义；

(3) $-\infty<x<0$，其他场合适当定义．

8．X 的分布函数分别如下，判断它是否为连续型随机变量，如果是，则求其密度函数．

(1) $F(x)=\begin{cases}0, & x<0,\\ x^2, & 0\leqslant x<1,\\ 1, & x\geqslant 1;\end{cases}$　(2) $F(x)=\begin{cases}0, & x<0,\\ \frac{1}{2}x^2, & 0\leqslant x<1,\\ 1, & x\geqslant 1.\end{cases}$

9．已知连续型随机变量 X 的密度函数为

(1) $f(x)=ae^{-|x|}$；　(2) $f(x)=\begin{cases}x, & 0\leqslant x<1,\\ 2-x, & 1\leqslant x<a,\\ 0, & \text{其他}.\end{cases}$

求 a 及分布函数 $F(x)$，$P\left\{-1<X\leqslant\frac{\sqrt{2}}{2}\right\}$，$P\left\{\frac{\sqrt{2}}{2}\leqslant X\leqslant\sqrt{2}\right\}$，$P\{X>1\}$.

10. 设随机变量 X 的密度函数关于 $x=\mu$ 对称，证明其分布函数满足以下性质：

$$F(\mu+x)+F(\mu-x)=1,\ -\infty<x<+\infty.$$

11. 求第 2 题中(2)及第 3 题的随机变量 X 的数学期望 EX.

12. 已知投资某一项目的收益率 R 是一随机变量，其分布为

R	1%	2%	3%	4%	5%	6%
p_i	0.1	0.1	0.2	0.3	0.2	0.1

一位投资者在该项目上投资 10 万元，求他预期获得多少收入？收入的方差是多大？

13. 一张贴现债券(期中不附息，期末还本附息的债券)承诺到期还本付息共偿还 1 100 元，根据分析，市场上同类债券的收益率为一随机变量，记作 $K\%$，设 K 的密度函数为

$$f(x)=\begin{cases}\frac{1}{5}, & 0\leqslant x\leqslant 5,\\ 0, & \text{其他}.\end{cases}$$

求这张债券现在平均值多少钱.

14. 在连续型情形证明(2.25).

15. 利用数学期望的性质，证明方差的性质(2.31)～(2.33).

16. 利用定理 2.3 的证明方法，分别就离散型和连续型情形直接证明“切比雪夫不等式”即(2.37).

17. 假设一厂家生产的每台仪器以概率 0.7 可直接出厂，以概率 0.3 需进一步调试，经调试后以概率 0.8 可出厂，概率 0.2 不合格而不能出厂，现该厂生产 $n(n\geqslant 2)$ 台仪器(设仪器生产过程相互独立).求(1) 能出厂的仪器数 X 的分布列；(2) n 台仪器能全部出厂的概率；(3) 至少有两台不能出厂的概率；(4) 不能出厂的仪器数的期望和方差.

18. 自动生产线在调整后出现废品的概率为 p，当在生产过程中出现废品时.立即重新进行调整，求在两次调整之间生产的合格品数 X 的分布列及其数学期望.

19. 设 X 服从泊松分布，已知 $P\{X=1\}=2P\{X=2\}$，求 EX，DX，EX^2，$P\{X=3\}$.

20. 在某公共汽车站甲、乙、丙三人分别等 1,2,3 路公共汽车.设每个人等车时间(单位：min)均服从[0,5]上的均匀分布，求 3 人中至少有两人等车时间不超过 2 min 的概率.

21. 某保险公司设置某一险种，规定每一保单有效期为一年，有效理赔一次，每个保单收取保费 500 元.理赔额为 40 000 元.据估计每个保单索赔概率为 0.01，设公司共卖出这种保单 8 000 个，求该公司在该险种上获得的平均利润.

22. 由指数分布的密度函数导出指数分布的分布函数以及数学期望和方差.

23. 给出定理 2.5 中的必要性的证明.

24. 3 个电子元件并联成一个系统，只有当 3 个元件损坏两个或两个以上时，系统便报废，已知电子元件的寿命服从参数为 $\frac{1}{1\,000}$ 的指数分布，求系统的寿命超过 1 000 h 的概率.

25. 导出正态分布 $N(\mu,\sigma^2)$的数学期望和方差.

26. 设测量的随机误差 $X\sim N(0,10^2)$,试求 100 次独立重复测量,至少有三次测量误差的绝对值大于 19.6 的概率 α,并用泊松分布求 α 的近似值.

27. 已知电源电压 X 服从正态分布 $N(220,25^2)$,在电源电压处于 $X\leqslant 200$ V,200 V$<X<240$ V,$X>240$ V 三种情况下,某电子元件损坏的概率分别 0.1,0.01,0.2.(1) 试求该电子元件损坏的概率 X;(2) 该电子元件损坏时,电源电压在 200～240 V 的概率 β.

28. 某班数学考试成绩呈正态分布 $N(70,100)$,老师将最高成绩的 5% 定为优秀,那么成绩为优秀的最少成绩是多少?

29. 测量一圆的半径 R,其概率分布为

R	10	11	12	13
p_i	0.1	0.4	0.3	0.2

求圆的面积 X 和周长 Y 的分布.

30. 设 X 服从$[a,b]$上的均匀分布,证明 $\alpha X+\beta(\alpha>0)$服从$[a\alpha+\beta,b\alpha+\beta]$上的均匀分布.

31. 设 X 服从$[-1,1]$上的均匀分布,求 X^2 的分布函数和密度函数.

32　X 服从参数为 1 的指数分布,求 $Y=\alpha X+\beta(\alpha>0)$的分布函数和密度函数.

33. 对球直径作测量,设其服从$[a,b]$上的均匀分布,求球的体积的均值.

34. 设 X 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{\frac{x}{\theta}}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

其中 $\theta>0$ 为参数,证明 $Y=\dfrac{2}{\theta}X$ 的密度函数为

$$g(y)=\begin{cases}\dfrac{1}{2}e^{-\frac{y}{2}}, & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

服从这一密度函数的分布称为自由度为 2 的 χ^2 分布,即 $Y\sim\chi^2(2)$.

35. X 服从参数为 2 的指数分布,求 $Y=1-e^{-2X}$的分布函数和密度函数.

36. 已知某股票的一年以后价格 X 服从对数正态分布,当前价格为 10 元,且 $EX=15$,$DX=4$.求其连续复合年收益率的分布.

37. 设 X 为非负随机变量,密度函数为 $f(x)$,证明 $Y=\sqrt{X}$的密度函数为

$$f_Y(y)=\begin{cases}2yf(y^2), & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

38. 设随机变量 X 的密度函数为 $f_X(x)$,令 $Y=aX+b(a\neq 0)$.证明:Y 的密度函数为

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

(B)

1. 设 $F(x)$是一个连续型随机变量的分布函数，$a>0$. 证明：$\int_{-\infty}^{+\infty}[F(x+a)-F(x)]\mathrm{d}x=a$.

2. 某型号电子管的寿命 ξ 服从指数分布，如果它的平均寿命为 $E\xi=1\,000$ h. (1) 一个该型号的旧电子管的寿命记为 η，求 η 的密度函数；(2) 一个系统由 n 个该型号的电子管并联组成，求该系统的寿命 X 的密度函数；(3) 一个系统由 n 个该型号电子管串联而成. 求该系统的寿命 Y 的密度函数.

3. 某商场经统计发现顾客对某商品的日需求量 $X\sim N(\mu,\sigma^2)$，且日平均需求量 $\mu=40$ (件)，销售在 30～50(件)之间的概率为 0.5. 若进货不足每件损失利润 70 元. 进货过量每件损失 100 元，求日最优进货量.

4. 设袋中装有 m 只颜色各不相同的球，有放回地摸取 n 次，摸到的球的颜色种数为 X，求证：$EX=m\left[1-\left(1-\frac{1}{m}\right)^n\right]$.

5. 利用概率论的思想证明：$\left[\int_a^b f(x)\mathrm{d}x\right]^2\leqslant(b-a)\int_a^b f^2(x)\mathrm{d}x$，其中 $f(x)$在$[a,b]$上连续.

第 3 章

随机向量

一个随机变量是定义在概率空间上的一个函数，因而，它实际上是一个与某一特定试验相联系的变量.这一章，我们将把讨论扩展到多个随机变量——随机向量，重点讨论二维随机向量.对二维随机向量的讨论很容易扩展到二维以上的随机向量.本章的主要内容包括随机向量的(联合)分布、独立性以及数字特征.在最后，我们还将专门介绍在概率论的理论和应用中都占有重要地位的极限理论.

§3.1 随机向量的分布

一、随机向量及其分布函数

定义 3.1 设 $X_1, X_2, \cdots, X_n$ 是定义在概率空间(Ω, P)上的 n 个随机变量，则称$(X_1, X_2, \cdots, X_n)$是(Ω, P)上的一个 ***n* 维随机向量**.

对于任何一个随机变量，我们都可以用分布函数来描述其统计特性，同样，我们也引入 n 维随机向量的分布函数.

定义 3.2 设$(X_1, X_2, \cdots, X_n)$是(Ω, P)上的一个 n 维随机向量，则称 n 元函数

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\} \tag{3.1}$$

为随机向量$(X_1, X_2, \cdots, X_n)$的**分布函数**或 n 个随机变量$X_1, X_2, \cdots, X_n$ 的**联合分布函数**.其中$\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}$表示$\{X_1 \leqslant x_1\}, \{X_2 \leqslant x_2\}, \cdots, \{X_n \leqslant x_n\}$的交事件.

为简单起见，我们主要讨论二维情形，设(X, Y)为一个二维随机向量，对给定的实向量(x, y)，$F(x, y) = P\{X \leqslant x, Y \leqslant y\}$实际上是$(X, Y)$取值于区域$\{(t, s) \mid t \leqslant x, s \leqslant y\}$(见图 3.1 中的阴影部分)的概率.根据概率的加法法则，(X, Y)取值于图 3.2 中的区域 Ⅰ 的概率 $P\{x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2\}$可用分布

函数来表示：

$$\begin{aligned}&P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}=P\{(X,Y)\in \mathrm{I}\}\\&=P\{(X,Y)\in(\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV})\}-P\{(X,Y)\in(\mathrm{III}\cup\mathrm{IV})\}\\&-P\{(X,Y)\in(\mathrm{II}\cup\mathrm{IV})\}+P\{(X,Y)\in\mathrm{IV}\}\\&=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1).\end{aligned}\tag{3.2}$$

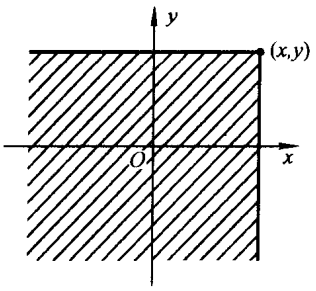


图 3.1 联合分布函数

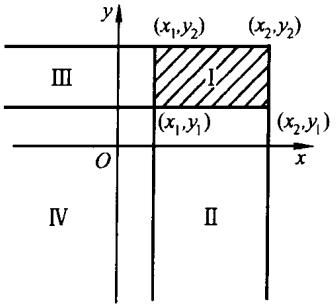


图 3.2 矩形区域上的概率

与一维情形类似，由(联合)分布函数的定义可导出其如下性质：

(1) $0\leqslant F(x,y)\leqslant 1$；

(2) $F(x,y)$关于 x 和 y 均单调非降、右连续；

(3) $F(-\infty,y)=\lim\limits_{x\to-\infty}F(x,y)=0$，

$F(x,-\infty)=\lim\limits_{y\to-\infty}F(x,y)=0$，

$F(-\infty,-\infty)=\lim\limits_{(x,y)\to(-\infty,-\infty)}F(x,y)=0$，

$F(+\infty,+\infty)=\lim\limits_{(x,y)\to(+\infty,+\infty)}F(x,y)=1$.

此外，如果(X,Y)的分布函数 $F(x,y)$已知，则由 $F(x,y)$可导出 X 和 Y 各自的分布函数$F_X(x)$和 $F_Y(y)$：

$$F_X(x)=P\{X\leqslant x\}=P\{X\leqslant x,Y<+\infty\}=F(x,+\infty),\tag{3.3}$$

$$F_Y(y)=P\{Y\leqslant y\}=F(+\infty,y).\tag{3.4}$$

通常称 $F_X(x)$和 $F_Y(y)$为(联合)分布函数 $F(x,y)$的**边缘分布函数**.

一般地，设 n 维随机向量$(X_1,X_2,\cdots,X_n)$的分布函数为 $F(x_1,x_2,\cdots,x_n)$，则称

$$F_i(x_i)=F(+\infty,\cdots,+\infty,x_i,+\infty,\cdots,+\infty),\quad i=1,2,\cdots,n\tag{3.5}$$

为 $F(x_1,x_2,\cdots,x_n)$的边缘分布函数.

二、离散型随机向量的概率分布

定义 3.3 如果二维随机向量(X,Y)只取有限个或可数个值，则称(X,Y)

为**二维离散型随机向量**.

显然,(X,Y)为二维离散型随机向量,当且仅当 X,Y 均为离散型随机变量.

自然,对二维离散型随机向量的最直接的描述是给出其取每一可能值的概率,为此,我们引入下列定义.

定义 3.4 设随机向量(X,Y)的所有可能取值为(x_i,y_j),$i,j=1,2,\cdots$,如果已知

$$P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots, \tag{3.6}$$

则称(3.6)为随机向量(X,Y)的**概率分布**,或 X 和 Y 的**联合概率分布**.

容易看出 p_{ij} 满足下列性质:

(1) $p_{ij}\geqslant 0,i,j=1,2,\cdots$;

(2) $\sum\limits_i\sum\limits_j p_{ij}=1$.

为了直观,有时也将联合概率分布用表格形式表示(见表 3.1),并称之为**联合概率分布表**.

表 3.1 联合概率分布表

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	$P\{X=x_i\}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$\sum\limits_j p_{1j}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$\sum\limits_j p_{2j}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$\sum\limits_j p_{ij}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$P\{Y=y_j\}$	$\sum\limits_i p_{i1}$	$\sum\limits_i p_{i2}$	$\cdots$	$\sum\limits_i p_{ij}$	$\cdots$	

由 X 和 Y 的联合概率的分布,可以求出 X,Y 各自的概率分布 $p_i^X(i=1,2,\cdots)$,$p_j^Y(j=1,2,\cdots)$.

$$\begin{aligned} p_i^X &= P\{X=x_i\}=P\Big(\bigcup_j\{X=x_i,Y=y_j\}\Big) \\ &= \sum_j P\{X=x_i,Y=y_j\}=\sum_j p_{ij},i=1,2,\cdots. \end{aligned} \tag{3.7}$$

同理

$$p_j^Y=P\{Y=y_j\}=\sum_i p_{ij},j=1,2,\cdots. \tag{3.8}$$

通常称(3.7),(3.8)为联合概率分布 $P\{X=x_i,Y=y_j\}=p_{ij}(i,j=1,2,\cdots)$的

边缘概率分布.在联合概率分布表中,边缘分布分别列在表中的最后一列和最后一行,它们分别等于联合概率分布的行和或列和,见表 3.1.

例 3.1 将两封信随意地投入 3 个邮筒,设 X, Y 分别表示投入第 1,2 号邮筒中信的数目,求 X 和 Y 的联合概率分布及边缘概率分布.

解 X, Y 各自的可能取值显然均为 0,1,2,由题设知,(X,Y)取(1,2),(2,1),(2,2)均不可能,因而相应的概率均为 0,我们将其标在联合概率分布表中相应位置(见下表).(X,Y)取其他值的概率可由古典概型计算,由于对称性,我们实际上只需计算下列概率:

$$P\{X=0,Y=0\}=\frac{1}{3^2}=\frac{1}{9},$$

$$P\{X=0,Y=1\}=\frac{2}{3^2}=\frac{2}{9},$$

$$P\{X=1,Y=1\}=\frac{2}{3^2}=\frac{2}{9}.$$

边缘概率分布可直接在联合概率分布表中计算,其中 X 的概率分布由行和产生,Y 的概率分布由列和产生(见下表).

X \ Y	0	1	2	p_i^X
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{4}{9}$
1	$\frac{2}{9}$	$\frac{2}{9}$	0	$\frac{4}{9}$
2	$\frac{1}{9}$	0	0	$\frac{1}{9}$
p_j^Y	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$	

对离散型随机向量而言,联合概率分布不仅比联合分布函数更加直观,而且它能够更加方便地确定(X,Y)取值于任何区域 D 上的概率,事实上,有

$$P\{(X,Y)\in D\}=\sum_{(x_i,y_j)\in D}p_{ij}. \tag{3.9}$$

特别地,由联合概率分布可以确定联合分布函数:

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\}=\sum_{x_i\leqslant x,y_j\leqslant y}p_{ij}. \tag{3.10}$$

例 3.2 设(X,Y)的概率分布由下表给出,求 $P\{X\neq 0,Y=0\}$, $P\{X\leqslant 0,Y\leqslant 0\}$, $P\{XY=0\}$, $P\{X=Y\}$, $P\{|X|=|Y|\}$.

X \ Y	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

解 $P\{X\neq 0, Y=0\} = P\{X=1, Y=0\} + P\{X=2, Y=0\}$
$= 0.05 + 0 = 0.05,$

$P\{X\leqslant 0, Y\leqslant 0\} = P\{X=0, Y=-1\} + P\{X=0, Y=0\}$
$= 0.1 + 0.2 = 0.3,$

$P\{XY=0\} = P(\{X=0\}\cup\{Y=0\})$
$= P\{X=0\} + P\{X\neq 0, Y=0\}$
$= 0.1 + 0.2 + 0 + 0.05 = 0.35,$

$P\{X=Y\} = P\{X=0, Y=0\} + P\{X=2, Y=2\}$
$= 0.2 + 0.1 = 0.3,$

$P\{|X| = |Y|\}$
$= P\{X=0, Y=0\} + P\{X=1, Y=-1\} + P\{X=2, Y=2\}$
$= 0.2 + 0.3 + 0.1 = 0.6.$

三、连续型随机向量的概率密度函数

定义 3.5 设(X, Y)为二维随机向量,分布函数为 $F(x, y)$,如果存在一个非负可积的二元函数 $f(x, y)$,使得对任意实向量(x, y),有

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s, t)\mathrm{d}s\mathrm{d}t, \tag{3.11}$$

则称(X, Y)为**二维连续型随机向量**,并称 $f(x, y)$为(X, Y)的**概率密度函数**(简称**密度函数**)或 X 与 Y 的**联合密度函数**.

从定义 3.5,易知联合密度函数 $f(x, y)$具有下列性质:

(1) $f(x, y)\geqslant 0$;

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\mathrm{d}x\mathrm{d}y = 1.$

反过来,满足上述两条性质的函数 $f(x, y)$,可以称为联合密度函数,因为能证明,它一定是某两个随机变量的联合密度函数.此外,通过密度函数,可以表达随机向量取值于任何平面区域的概率;

(3) 若 D 是平面上的一个区域,则

$$P\{(X,Y)\in D\}=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y. \tag{3.12}$$

特别地,由(3.12),边缘分布函数 $F_X(x)$可表示为

$$\begin{aligned}F_X(x)&=P\{X\leqslant x\}=P\{X\leqslant x,Y<+\infty\}\\&=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(s,t)\mathrm{d}s\mathrm{d}t\\&=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}f(s,t)\mathrm{d}t\right]\mathrm{d}s.\end{aligned} \tag{3.13}$$

由(3.13)知,X 是连续型随机变量,且其密度函数为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y. \tag{3.14}$$

同理,Y 是连续型随机变量,其密度函数为

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x. \tag{3.15}$$

通常称(3.14)、(3.15)中的 $f_X(x)$和 $f_Y(y)$为(X,Y)或联合密度函数 $f(x,y)$的**边缘密度函数**.

例 3.3(均匀分布) 设 G 是平面上的一个有界区域,其面积记作 $S(G)$,二维连续的随机向量(X,Y)只在 G 中取值,且取 G 中的每一点是"等可能的".于是,(X,Y)的密度函数为

$$f(x,y)=\begin{cases}C, & (x,y)\in G,\\0, & \text{其他}.\end{cases}$$

由密度函数的性质

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1,$$

可得 $C=\dfrac{1}{S(G)}$,即有

$$f(x,y)=\begin{cases}\dfrac{1}{S(G)}, & (x,y)\in G,\\0, & \text{其他}.\end{cases} \tag{3.16}$$

通常,如果一个二维随机向量(X,Y)以(3.16)为密度函数,则称(X,Y)服从区域 G 上的**均匀分布**.

由(3.12),若(X,Y)服从 G 上的均匀分布,则对任何平面区域 D,有

$$\begin{aligned}P\{(X,Y)\in D\}&=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y=\iint_{D\cap G}\frac{1}{S(G)}\mathrm{d}x\mathrm{d}y\\&=\frac{S(D\cap G)}{S(G)},\end{aligned} \tag{3.17}$$

其中 $S(D\cap G)$是区域 $D\cap G$ 的面积.

例 3.4 设随机向量(X_1, Y_1)的密度函数$f(x,y)$,(X_2, Y_2)的密度函数$g(x,y)$分别为

$$f(x,y)=\begin{cases}k_1 xy, & 0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1,\\ 0, & \text{其他};\end{cases}$$

$$g(x,y)=\begin{cases}k_2 xy, & 0\leqslant x\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

求(1) 参数k_1的值及(X_1, Y_1)的边缘密度函数;

(2) 参数k_2的值及(X_2, Y_2)的边缘密度函数.

解 (1) 由密度函数的性质,有

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^1\int_0^1 k_1 xy\mathrm{d}x\mathrm{d}y=1,$$

由此易得$k_1=4$,从而

$$f(x,y)=\begin{cases}4xy, & 0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

记(X_1, Y_1)的边缘密度函数分别为$f_{X_1}(x)$,$f_{Y_1}(y)$,则由

$$f_{X_1}(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y,$$

得:当$0\leqslant x\leqslant 1$时,

$$f_{X_1}(x)=\int_0^1 4xy\mathrm{d}y=2x.$$

当$x<0$或$x<1$时,

$$f_{X_1}(x)=0.$$

即

$$f_{X_1}(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

根据对称性,有

$$f_{Y_1}(y)=\begin{cases}2y, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

(2) 由密度函数的性质,有

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x\mathrm{d}y=\iint\limits_{0\leqslant x\leqslant y\leqslant 1} k_2 xy\mathrm{d}x\mathrm{d}y=\int_0^1 \mathrm{d}x\int_x^1 k_2 xy\mathrm{d}y$$

$$=\frac{1}{8}k_2=1,$$

从而得$k_2=8$,故知

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

记(X_2, Y_2)的边缘密度函数为$g_{X_2}(x)$,$g_{Y_2}(y)$,则由

$$g_{X_2}(x)=\int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y$$

得:当 $0\leqslant x\leqslant 1$ 时,

$$g_{X_2}(x)=\int_x^1 8xy\mathrm{d}y=4x(1-x^2),$$

当 $x<0$ 或 $x>1$ 时,

$$g_{X_2}(x)=0,$$

即

$$g_{X_2}(x)=\begin{cases}4x(1-x^2), & 0\leqslant x\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

同理,可得

$$g_{Y_2}(y)=\begin{cases}4y^3, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

四、二元正态分布

在讨论一元随机变量时,我们曾指出,一元正态分布是实际应用中最常见的分布之一.对二维随机向量,在实际中会常用到二元正态分布:设随机向量(X,Y)的密度函数为

$$\varphi(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]},\qquad(3.18)$$

其中 $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$ 均为参数,且 $\sigma_1>0,\sigma_2>0,|\rho|<1$,则称$(X,Y)$服从参数为$(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$的二元正态分布,记作$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$.

如图 3.3,二元正态分布以(μ_1,μ_2)为中心,在中心附近具有较高的密度,离中心越远,密度越小,这与实际中很多现象相吻合.

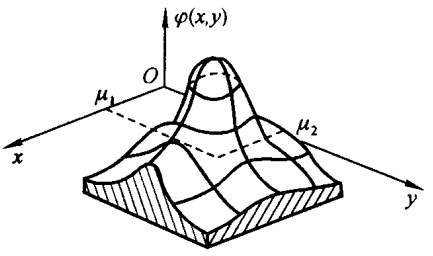


图 3.3 二元正态分布的密度函数

若$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$,记其边缘密度函数分别为 $\varphi_X(x)$和$\varphi_Y(y)$,

令 $u=\dfrac{x-\mu_1}{\sigma_1}$,$t=\dfrac{y-\mu_2}{\sigma_2}$,则

$$\begin{aligned}\varphi_X(x)&=\int_{-\infty}^{+\infty}\varphi(x,y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}[u^2-2\rho ut+t^2]}\mathrm{d}t\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{(t-\rho u)^2}{2(1-\rho^2)}}\mathrm{d}t\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\end{aligned}$$

可见 $X\sim N(\mu_1,\sigma_1^2)$. 对称地,可知

$$\varphi_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

即 $Y\sim N(\mu_2,\sigma_2^2)$.

比较联合密度函数 $\varphi(x,y)$和边缘密度函数 $\varphi_X(x),\varphi_Y(y)$,我们注意到,当且仅当 $\rho=0$ 时,对一切(x,y),有

$$\varphi(x,y)=\varphi_X(x)\varphi_Y(y). \tag{3.19}$$

上述讨论实际上说明:

(1) 二元正态分布的边缘分布是一元正态分布,它们的参数对应于二元正态分布的前 4 个参数;

(2) 不同的二元正态分布,比如不同的 ρ,可以有相同的边缘分布,因而由边缘分布不能惟一确定联合分布. 为了确定一个二元正态分布的密度函数,除了知道边缘分布以外,还须知道参数 ρ 的值. 特别地,如果 $\rho=0$,则联合密度函数可由(3.19)确定.

另外,值得进一步指出的是:

两个边缘分布为正态分布的二维随机向量未必服从二元正态分布(见习题三,第 22 题).

§3.2 条件分布与随机变量的独立性

一、条件分布与独立性的一般概念

设 X 是一个随机变量,我们知道,其统计规律可由其分布函数 $F_X(x)=P\{X\leqslant x\}$, $-\infty<x<+\infty$来完整描述. 现在假设我们观察到一个额外的信息:已知事件 A 已经发生,那么 A 的发生可能会对事件$\{X\leqslant x\}$发生的概率产生影响. 对每个给定的实数 x,我们记条件概率 $P\{X\leqslant x\mid A\}$为 $F(x\mid A)$,并称 $F(x\mid A)$, $-\infty<x<+\infty$为在 A 发生的条件下 X 的**条件分布函数**.

例 3.5 设 X 服从$[0,1]$上的均匀分布,求在已知 $X>\frac{1}{2}$的条件下 X 的条件分布函数.

解 由条件分布函数的定义,有

$$F(x\mid X>\frac{1}{2})=\frac{P\{X\leqslant x,X>\frac{1}{2}\}}{P\{X>\frac{1}{2}\}}.$$

由于 X 服从[0,1]上的均匀分布,故 $P\{X>\frac{1}{2}\}=\frac{1}{2}$.

而当 $x\leqslant\frac{1}{2}$时,

$$P\{X\leqslant x, X>\frac{1}{2}\}=0,$$

当 $x>\frac{1}{2}$时,

$$P\{X\leqslant x, X>\frac{1}{2}\}=F(x)-F\left(\frac{1}{2}\right)=F(x)-\frac{1}{2},$$

其中 $F(x)$为 X 的分布函数.我们知道,

$$F(x)=\begin{cases}0, & x<0,\\ x, & 0\leqslant x\leqslant 1,\\ 1, & x>1.\end{cases}$$

于是,当 $x>\frac{1}{2}$时

$$P\{X\leqslant x, X>\frac{1}{2}\}=\begin{cases}x-\frac{1}{2}, & \frac{1}{2}<x\leqslant 1,\\ \frac{1}{2}, & x>1.\end{cases}$$

从而可得

$$F(x\mid X>\frac{1}{2})=\begin{cases}0, & x\leqslant\frac{1}{2},\\ 2x-1, & \frac{1}{2}<x\leqslant 1,\\ 1, & x>1.\end{cases}$$

设 A 是另一随机变量 Y 所生成的事件:$A=\{Y\leqslant y\}$,且 $P\{Y\leqslant y\}>0$,则有

$$F(x\mid Y\leqslant y)=\frac{P\{X\leqslant x, Y\leqslant y\}}{P\{Y\leqslant y\}}=\frac{F(x,y)}{F_Y(y)}. \tag{3.20}$$

一般地,两个随机变量 X 和 Y 之间存在着相互联系,因而一个随机变量的取值可能会影响另一随机变量取值的统计规律性.(3.20)表明联合分布函数包含了 X 与 Y 相互联系的内容.

对给定的 x 和 y,如果事件$\{X\leqslant x\}$与事件$\{Y\leqslant y\}$独立,则有

$$P\{X\leqslant x, Y\leqslant y\}=P\{X\leqslant x\}P\{Y\leqslant y\},$$

亦即

$$F(x,y)=F_X(x)F_Y(y), \tag{3.21}$$

此时,

$$F(x \mid Y \leqslant y) = F_X(x).$$

如果对任意 x, y，$\{X \leqslant x\}$ 与 $\{Y \leqslant y\}$ 均独立，则称随机变量 X 与 Y 相互独立，由(3.21)我们给出下列定义：

定义 3.6 设随机变量 X, Y 的联合分布函数为 $F(x, y)$，边缘分布函数分别为 $F_X(x), F_Y(y)$，如果对任意实数 x 和 y，恒有

$$F(x, y) = F_X(x) F_Y(y),$$

则称随机变量 X 和 Y **相互独立**.

表面上，上述定义不太自然，随机变量 X 与 Y 独立似乎应该要求 X 所生成的任何事件与 Y 所生成的任何事件都相互独立，事实上，我们有下列定理.

定理 3.1 随机变量 X 与 Y 相互独立的充要条件是 X 所生成的任何事件与 Y 生成的任何事件独立，即，对任意实数集 A 和 B[①]，有

$$P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}. \tag{3.22}$$

定理 3.1 的充分性显然，必要性的完整证明超出了本书的要求，作为练习，读者可以证明 $A = (x_1, x_2], B = (y_1, y_2]$ 这种简单的情形.

定理 3.2 如果随机变量 X 和 Y 相互独立，则对任意函数 $g_1(x), g_2(y)$，均有 $g_1(X)$ 与 $g_2(Y)$ 相互独立.

证明 令 $\zeta = g_1(X), \eta = g_2(Y)$，对任意 x, y，记 $D_x^1 = \{t \mid g_1(t) \leqslant x\}$，$D_y^2 = \{t \mid g_2(t) \leqslant y\}$，则由定理 3.1，有

$$\begin{aligned} P\{\zeta \leqslant x, \eta \leqslant y\} &= P\{g_1(X) \leqslant x, g_2(Y) \leqslant y\} \\ &= P\{X \in D_x^1, Y \in D_y^2\} \\ &= P\{X \in D_x^1\} P\{Y \in D_y^2\} \\ &= P\{\zeta \leqslant x\} P\{\eta \leqslant y\}, \end{aligned}$$

从而由定义 3.6 知 ζ 与 η 独立.

关于两个随机变量的独立性的概念和讨论可以推广到 n 个随机变量的情形.

定义 3.7 设 $X_1, X_2, \cdots, X_n$ 是 n 个随机变量，其联合分布函数为 $F(x_1, x_2, \cdots, x_n)$，边缘分布函数为 $F_i(x_i)(i = 1, 2, \cdots, n)$，如果对任意实数 $x_1, x_2, \cdots, x_n$ 恒有

$$F(x_1, x_2, \cdots, x_n) = F_1(x_1) F_2(x_2) \cdots F_n(x_n),$$

则称 $X_1, X_2, \cdots, X_n$ 相互独立.

条件分布函数为条件分布和随机变量的独立性提供了一般性描述和讨论，但应用起来十分不便，对于本书重点讨论的离散型和连续型的随机变量(向量)，

① 严格地讲，A, B 应为一类特殊的集合——Borel 集，以保证 $\{X \in A\}, \{X \in B\}$ 为事件

我们希望有更加直观而方便的描述.

二、离散型随机变量的条件概率分布与独立性

设(X,Y)是二维离散型随机向量,其概率分布为

$$P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots,$$

则由条件概率公式,当$P\{Y=y_j\}>0$时,有

$$P\{X=x_i\mid Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_j^Y},\tag{3.23}$$

其中$P\{X=x_i\mid Y=y_j\}$是在事件"$Y=y_j$"发生的条件下,事件"$X=x_i$"发生的条件概率,通常记作$p_{i|j}$.

对固定的j,$P\{X=x_i\mid Y=y_j\}=p_{i|j}(i=1,2,\cdots)$完整地给出了在$Y=y_j$的条件下,$X$取每一个可能值$x_i(i=1,2,\cdots)$的概率,且不难验证,数列$p_{i|j}$$(i=1,2,\cdots)$满足概率分布所要求的性质:

(1) $p_{i|j}\geqslant 0$;

(2) $\sum\limits_i p_{i|j}=1$.

因而,我们称

$$P\{X=x_i\mid Y=y_j\}=p_{i|j},i=1,2,\cdots$$

为已知$Y=y_j$的条件下X的**条件概率分布**.

对称地,如果$P\{X=x_i\}>0$,那么在$X=x_i$的条件下Y的条件概率分布为

$$P\{Y=y_j\mid X=x_i\}=p_{j|i}=\frac{p_{ij}}{p_i^X},j=1,2,\cdots.\tag{3.24}$$

此外,根据乘法公式或(3.23)、(3.24),对任意$i,j=1,2,\cdots$,

$$p_{ij}=p_i^X\cdot p_{j|i}=p_j^Y\cdot p_{i|j}.\tag{3.25}$$

例 3.6 设X与Y的联合概率分布由例3.2给出,求$Y=0$时X的条件概率分布以及$X=0$时Y的条件概率分布.

解 $P\{Y=0\}=0.2+0.05+0=0.25$,在$Y=0$时,X的条件概率分布为:

$$P\{X=0\mid Y=0\}=\frac{P\{X=0,Y=0\}}{P\{Y=0\}}=\frac{0.2}{0.25}=0.8;$$

$$P\{X=1\mid Y=0\}=\frac{P\{X=1,Y=0\}}{P\{Y=0\}}=\frac{0.05}{0.25}=0.2;$$

$$P\{X=2\mid Y=0\}=\frac{P\{X=2,Y=0\}}{P\{Y=0\}}=\frac{0}{0.25}=0.$$

又$P\{X=0\}=0.1+0.2+0=0.3$,故在$X=0$时,Y的条件概率分布为:

$$P\{Y=-1\mid X=0\}=\frac{P\{Y=-1,X=0\}}{P\{X=0\}}=\frac{0.1}{0.3}=\frac{1}{3};$$

$$P\{Y=0|X=0\}=\frac{P\{Y=0,X=0\}}{P\{X=0\}}=\frac{0.2}{0.3}=\frac{2}{3};$$

$$P\{Y=2|X=0\}=\frac{P\{Y=2,X=0\}}{P\{X=0\}}=\frac{0}{0.3}=0.$$

接下来考虑离散型随机变量的独立性.由定理3.1,如果 X 与 Y 相互独立,那么对任意 i,j,必有

$$P\{X=x_i,Y=y_j\}=P\{X=x_i\}P\{Y=y_j\}. \tag{3.26}$$

另一方面,如果对任意 i,j,(3.26)成立,则 X 与 Y 相互独立,这是因为:由(3.26),对任意 x 和 y,

$$\begin{aligned}P\{X\leqslant x,Y=y_j\}&=P\Big(\bigcup_{x_i\leqslant x}\{X=x_i,Y=y_j\}\Big)\\&=\sum_{x_i\leqslant x}P\{X=x_i,Y=y_j\}\\&=\sum_{x_i\leqslant x}P\{X=x_i\}P\{Y=y_j\}\\&=P\Big(\bigcup_{x_i\leqslant x}\{X=x_i\}\Big)P\{Y=y_j\}\\&=P\{X\leqslant x\}P\{Y=y_j\},\end{aligned}$$

进而,对任意 x,y,有

$$\begin{aligned}P\{X\leqslant x,Y\leqslant y\}&=P\Big(\bigcup_{y_j\leqslant y}\{X\leqslant x,Y=y_j\}\Big)\\&=\sum_{y_j\leqslant y}P\{X\leqslant x,Y=y_j\}\\&=\sum_{y_j\leqslant y}P\{X\leqslant x\}P\{Y=y_j\}\\&=P\{X\leqslant x\}P\Big(\bigcup_{y_j\leqslant y}\{Y=y_j\}\Big)\\&=P\{X\leqslant x\}P\{Y\leqslant y\}.\end{aligned}$$

于是,我们有下列定理.

定理3.3　设 X,Y 是离散型随机变量,其联合概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}(i,j=1,2,\cdots)$,边缘概率分布分别为 p_i^X 和 $p_j^Y(i,j=1,2,\cdots)$,则 X 与 Y 相互独立的充要条件是

$$p_{ij}=p_i^Xp_j^Y,i,j=1,2,\cdots. \tag{3.27}$$

例3.7　设 X 与 Y 的联合概率分布由例3.2给出,判断 X 与 Y 是否相互独立?

解　因为

$$P\{X=0\}=0.1+0.2=0.3,$$

$$P\{Y=-1\}=0.1+0.3+0.15=0.55,$$

而 $P\{X=0,Y=-1\}=0.1$,可见 $P\{X=0,Y=-1\}\neq P\{X=0\}P\{Y=-1\}$,所以 X 与 Y 不独立.

在前一节讨论中,我们得知,由联合概率分布可以确定边缘概率分布.但是由边缘概率分布一般不能确定联合概率分布,比较表 3.2 中的两个不同联合概率分布,我们注意到它们具有相同的边缘概率分布.

如果已知随机变量 X 与 Y 相互独立,那么由边缘概率分布可以确定联合概率分布,因为此时

$$p_{ij}=p_i^X p_j^Y.$$

比如,设 X 的概率分布为

$$P\{X=0\}=P\{X=1\}=P\{X=2\}=\frac{1}{3},$$

Y 的概率分布为

$$P\{Y=0\}=P\{Y=1\}=P\{Y=2\}=\frac{1}{3},$$

且已知 X 与 Y 相互独立,那么 X 与 Y 的联合概率分布可由(3.27)求得,它实际上就是表 3.2 中右表所列的联合概率分布.

表 3.2 具有相同边缘概率分布的两个不同的联合概率分布

X \ Y	0	1	2	p_i^X
0	$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{1}{3}$
1	0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
2	$\frac{1}{9}$	0	$\frac{2}{9}$	$\frac{1}{3}$
p_j^Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

X \ Y	0	1	2	p_i^X
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
p_j^Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

三、连续型随机变量的条件密度函数与独立性

设(X,Y)是连续型随机向量,分布函数和密度函数分别为 $F(x,y)$和 $f(x,y)$.我们希望考虑在 $Y=y$ 的条件下 X 的条件分布.但由于$\{Y=y\}$是一个零概率事件,

$$P\{X\leqslant x\mid Y=y\}=\frac{P\{X\leqslant x,Y=y\}}{P\{Y=y\}} \tag{3.28}$$

的分子、分母均为 0,因而直接根据条件概率定义来考虑 X 的条件分布行不通.

为此，我们将(3.28)视为$\frac{0}{0}$型的未定式，即有

$$\begin{aligned} P\{X\leqslant x \mid Y=y\} &= \lim_{\Delta y\to 0} P\{X\leqslant x \mid y-\Delta y<Y\leqslant y\} \\ &= \lim_{\Delta y\to 0} \frac{P\{X\leqslant x, y-\Delta y<Y\leqslant y\}}{P\{y-\Delta y<Y\leqslant y\}} \\ &= \lim_{\Delta y\to 0} \frac{\int_{y-\Delta y}^{y}\left[\int_{-\infty}^{x} f(u,t)\mathrm{d}u\right]\mathrm{d}t}{\int_{y-\Delta y}^{y} f_Y(t)\mathrm{d}t} \\ &= \frac{\int_{-\infty}^{x} f(u,y)\mathrm{d}u}{f_Y(y)}. \end{aligned} \tag{3.29}$$

并且，对给定的 y，如果 $f_Y(y)>0$，则称 $P\{X\leqslant x \mid Y=y\}$ 为 $Y=y$ 的条件下 X 的条件分布函数，记作 $F_{X|Y}(x\mid y)$. 由(3.29)知

$$F_{X|Y}(x\mid y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}\mathrm{d}u, \tag{3.30}$$

因而，称$\frac{f(x,y)}{f_Y(y)}$为 $Y=y$ 的条件下 X 的**条件密度函数**，记作 $f_{X|Y}(x\mid y)$.

同样，对给定的 x，如果 $f_X(x)>0$，那么在 $X=x$ 的条件下，Y 的条件分布函数为

$$F_{Y|X}(y\mid x)=\int_{-\infty}^{y}\frac{f(x,t)}{f_X(x)}\mathrm{d}t, \tag{3.31}$$

条件密度函数为

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}. \tag{3.32}$$

由条件密度函数的定义，我们容易知道，密度函数有下列乘法公式：

$$f(x,y)=f_X(x)f_{Y|X}(y\mid x)=f_Y(y)f_{X|Y}(x\mid y). \tag{3.33}$$

例 3.8　设(X,Y)是在 $D=\{(x,y)\mid x^2+y^2\leqslant 1\}$ 上服从均匀分布的随机向量，求 $f_{Y|X}(y\mid x)$和 $f_{X|Y}(x\mid y)$.

解　由于(X,Y)的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

于是其边缘密度函数 $f_X(x)$为

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\begin{cases}\dfrac{2\sqrt{1-x^2}}{\pi}, & |x|\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

对称地，$f_Y(y)$为

$$f_Y(y)=\begin{cases}\dfrac{2\sqrt{1-y^2}}{\pi}, & |y|\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

于是，对一切 $x(|x|<1)$，有

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & |y|\leqslant\sqrt{1-x^2},\\ 0, & \text{其他}.\end{cases}$$

同样，对一切 $y(|y|<1)$，有

$$f_{X|Y}(x|y)=\begin{cases}\dfrac{1}{2\sqrt{1-y^2}}, & |x|\leqslant\sqrt{1-y^2},\\ 0, & \text{其他}.\end{cases}$$

例 3.9 设$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，求 $f_{X|Y}(x|y)$和 $f_{Y|X}(y|x)$.

解 由 §3.1 知，$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，于是

$$\begin{aligned}f_{X|Y}(x|y)&=\frac{f(x,y)}{f_Y(y)}\\&=\frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}}{\dfrac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}}\\&=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)^2}\\&=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-\mu_1-\frac{\sigma_1}{\sigma_2}\rho(y-\mu_2)\right]^2}.\end{aligned}$$

故，在 $Y=y$ 的条件下，X 服从正态分布 $N\left(\mu_1+\dfrac{\sigma_1}{\sigma_2}\rho(y-\mu_2),\sigma_1^2(1-\rho^2)\right)$.

对称地，在 $X=x$ 的条件下，Y 服从正态分布 $N\left(\mu_2+\dfrac{\sigma_2}{\sigma_1}\rho(x-\mu_1),\sigma_2^2(1-\rho^2)\right)$.

接下来考虑连续型随机变量的独立性，我们希望用密度函数来刻画独立性. 为此，给出以下定理.

定理 3.4 设连续型随机向量(X,Y)的密度函数为 $f(x,y)$，边缘密度函数分别为 $f_X(x)$和 $f_Y(y)$，则 X 与 Y 相互独立的充要条件是

$$f(x,y)=f_X(x)f_Y(y). \tag{3.34}$$

证明 必要性 如果 X 与 Y 相互独立,则对任意 x,y,有

$$F(x,y)=F_X(x)F_y(y)=\int_{-\infty}^{x}f_X(u)\mathrm{d}u\cdot\int_{-\infty}^{y}f_Y(t)\mathrm{d}t$$

$$=\int_{-\infty}^{x}\int_{-\infty}^{y}f_X(u)f_Y(t)\mathrm{d}u\mathrm{d}t,$$

于是 $f_X(x)f_Y(y)$是(X,Y)的密度函数,即

$$f(x,y)=f_X(x)f_Y(y).$$

充分性 若 $f(x,y)=f_X(x)f_Y(y)$,则

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_X(u)f_Y(t)\mathrm{d}u\mathrm{d}t$$

$$=\int_{-\infty}^{x}f_X(u)\mathrm{d}u\cdot\int_{-\infty}^{y}f_Y(t)\mathrm{d}t$$

$$=F_X(x)F_Y(y),$$

从而 X 与 Y 相互独立.

例 3.10 判断例 3.4 中 X_1 与 Y_1,X_2 与 Y_2 是否相互独立.

解 由例 3.4 知,X_1 与 Y_1 的联合密度函数及边缘密度函数分别为:

$$f(x,y)=\begin{cases}4xy, & 0\leqslant x\leqslant 1,0\leqslant y\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

$$f_{X_1}(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

$$f_{Y_1}(y)=\begin{cases}2y, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

显然有

$$f(x,y)=f_{X_1}(x)f_{Y_1}(y),$$

故 X_1 与 Y_1 相互独立.

由例 3.4 知,X_2 与 Y_2 的联合密度函数及边缘密度函数分别为:

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

$$g_{X_2}(x)=\begin{cases}4x(1-x^2), & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

$$g_{Y_2}(y)=\begin{cases}4y^3, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

对任意满足 $1>x_0>y_0>0$ 的点(x_0,y_0),有 $g(x_0,y_0)=0$,但 $g_{X_2}(x_0)g_{Y_2}(y_0)$ $=16x_0(1-x_0^2)y_0^3\neq 0$,即

$$g(x_0,y_0)\neq g_{X_2}(x_0)g_{Y_2}(y_0),$$

因而 X_2 与 Y_2 不相互独立.

例 3.11 若$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$,证明 X 与 Y 相互独立的充要条件是$\rho=0$.

解 由§3.1知 $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$,于是比较 $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$与 $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$的密度函数:$\varphi(x,y)$, $\varphi_X(x)$, $\varphi_Y(y)$.易知,当且仅当 $\rho=0$ 时,

$$\varphi(x,y)=\varphi_X(x)\varphi_Y(y).$$

§3.3 随机向量的函数的分布与数学期望

一、离散型随机向量的函数的分布

设(X,Y)是二维离散型随机向量,$g(x,y)$是一个二元函数,则 $g(X,Y)$作为(X,Y)的函数是一个随机变量.如果(X,Y)的概率分布为

$$P\{X=x_i,Y=y_j\}=p_{ij},\ i,j=1,2,\cdots,$$

记 $z_k(k=1,2,\cdots)$为 $Z=g(X,Y)$的所有可能取值,则由(3.9), Z 的概率分布为

$$P\{Z=z_k\}=P\{g(X,Y)=z_k\}=\sum_{g(x_i,y_j)=z_k}P\{X=x_i,Y=y_j\},\quad (3.35)$$

$$k=1,2,\cdots.$$

例 3.12 设(X,Y)的概率分布由例 3.2 给出,求 $\xi=X+Y$ 及 $\eta=XY$ 的概率分布.

解 显然 $\xi=X+Y$ 的可能取值有:$-1,0,1,2,3,4$, $\eta=XY$ 的可能取值有:$-2,-1,0,4$.由(3.35), ξ 的概率分布为:

$$P\{\xi=-1\}=P\{X+Y=-1\}=P\{X=0,Y=-1\}=0.1,$$

$$P\{\xi=0\}=P\{X+Y=0\}=P\{X=0,Y=0\}+P\{X=1,Y=-1\}$$
$$=0.2+0.3=0.5,$$

$$P\{\xi=1\}=P\{X+Y=1\}=P\{X=1,Y=0\}+P\{X=2,Y=-1\}$$
$$=0.05+0.15=0.2,$$

$$P\{\xi=2\}=P\{X+Y=2\}=P\{X=0,Y=2\}+P\{X=2,Y=0\}$$
$$=0+0=0,$$

$$P\{\xi=3\}=P\{X+Y=3\}=P\{X=1,Y=2\}=0.1,$$

$P\{\xi=4\}=P\{X+Y=4\}=P\{X=2,Y=2\}=0.1.$

η 的概率分布为：

$P\{\eta=-2\}=P\{X=2,Y=-1\}=0.15,$

$P\{\eta=-1\}=P\{X=1,Y=-1\}=0.3,$

$$\begin{aligned}P\{\eta=0\}&=P\{X=0,Y=-1\}+P\{X=0,Y=0\}+P\{X=0,Y=2\}\\&\quad+P\{X=1,Y=0\}+P\{X=2,Y=0\}\\&=0.1+0.2+0+0.05+0=0.35,\end{aligned}$$

$P\{\eta=2\}=P\{X=1,Y=2\}=0.1,$

$P\{\eta=4\}=P\{X=2,Y=2\}=0.1.$

我们也可用一种更直接的方式——表上作业法来求上述概率分布.这种表上作业法一般分为三步：

第一步：将(X,Y)的概率分布表达为下列形式.

X \ Y	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

第二步，将每一对(x_i,y_j)对应的函数值 $g(x_i,y_j)$算出标在对应小方框的右下角($X+Y$ 见下列左表，XY 见下列右表).

X \ Y	-1	0	2
0	0.1 -1	0.2 0	0 2
1	0.3 0	0.05 1	0.1 3
2	0.15 1	0 2	0.1 4

X \ Y	-1	0	2
0	0.1 0	0.2 0	0 0
1	0.3 -1	0.05 0	0.1 2
2	0.15 -2	0 0	0.1 4

第三步：将第二步表格中 $g(x_i,y_j)$的值相同的项合并，即得 $g(X,Y)$的概率分布，对本例，有

$P\{\xi=-1\}=0.1,P\{\xi=0\}=0.5,P\{\xi=1\}=0.15+0.05=0.2,$

$P\{\xi=2\}=0,P\{\xi=3\}=0.1,P\{\xi=4\}=0.1,$

$P\{\eta=-2\}=0.15, P\{\eta=-1\}=0.3, P\{\eta=0\}=0.1+0.2+0.05=0.35,$
$P\{\eta=2\}=0.1, P\{\eta=4\}=0.1.$

例 3.13 设 X, Y 是两个相互独立的随机变量，分别服从参数为 λ_1 和 λ_2 的泊松分布，求 $\xi=X+Y$ 的分布.

解
$$\begin{aligned} P\{\xi=k\} &= P\{X+Y=k\} \\ &= \sum_{i=0}^{k} P\{X=i, Y=k-i\} \\ &= \sum_{i=0}^{k} P\{X=i\}P\{Y=k-i\} \\ &= \sum_{i=0}^{k} \frac{\lambda_1^i}{i!}e^{-\lambda_1}\cdot\frac{\lambda_2^{k-i}}{(k-i)!}e^{-\lambda_2} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{k}\frac{\lambda_1^i\lambda_2^{k-i}}{i!\,(k-i)!} \\ &= \frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}, \\ & k=0,1,2,\cdots, \end{aligned}$$

可见 $\xi=X+Y$ 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布.

二、连续型随机向量的函数的分布

设 (X,Y) 是二维连续型随机向量，其概率密度函数为 $f(x,y)$，令 $g(x,y)$ 为一个二元函数，则 $g(X,Y)$ 是 (X,Y) 的函数. 我们可以类似于求一元随机变量函数分布的方法来求 $Z=g(X,Y)$ 的分布.

先求分布函数 $F_Z(z)$.

$$\begin{aligned} F_Z(z) &= P\{Z\leqslant z\} = P\{g(X,Y)\leqslant z\} = P\{(X,Y)\in D_z\} \\ &= \iint_{D_z} f(x,y)\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{3.36}$$

其中 $D_z=\{(x,y)\mid g(x,y)\leqslant z\}$.

继而，其密度函数 $f_Z(z)$，对几乎所有的 z，有

$$f_Z(z)=F_Z'(z) \tag{3.37}$$

例 3.14（随机变量的和） 设 (X,Y) 的联合密度函数为 $f(x,y)$，求 $X+Y$ 的密度函数.

解 对任意 z，令 $D_z=\{(x,y)\mid x+y\leqslant z\}$，如图 3.4，则

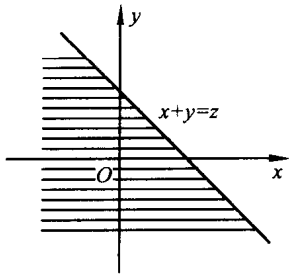


图 3.4 $D_z=\{(x,y)\mid x+y\leqslant z\}$

$$
\begin{aligned}
F_Z(z) &= P\{Z \leqslant z\} = P\{X+Y \leqslant z\} \\
&= \iint_{D_z} f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \mathrm{d}y \int_{-\infty}^{z-y} f(x,y)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \mathrm{d}y \int_{-\infty}^{z} f(u-y,y)\mathrm{d}u \\
&= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} f(u-y,y)\mathrm{d}y\right]\mathrm{d}u. \qquad (3.38)
\end{aligned}
$$

于是,有

$$
f_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y, \qquad (3.39)
$$

易见,交换积分次序,我们亦可得到

$$
f_Z(z) = \int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x. \qquad (3.40)
$$

特别地,如果 X 与 Y 是相互独立的随机变量,则由(3.39)和(3.40)有

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\mathrm{d}y. \qquad (3.41)
\end{aligned}
$$

(3.41)最后一个积分给出的积分运算通常称为函数 $f_X(x)$ 与 $f_Y(y)$ 的卷积,记作 $f_X * f_Y(z)$,由(3.41)知

$$
f_Z(z) = f_X * f_Y(z) = f_Y * f_X(z). \qquad (3.42)
$$

我们把(3.41)或(3.42)称为**卷积公式**.

作为卷积公式的一个应用,下面给出一个重要的例子.

例3.15(独立正态随机变量之和) 设随机变量 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$,且 X 与 Y 独立,证明 $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

证明 $f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\mathrm{d}x$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_1 \sqrt{2\pi}\sigma_2} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot \mathrm{e}^{-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}} \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} \mathrm{e}^{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(z-x-\mu_2)^2}{\sigma_2^2}\right]} \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} \mathrm{e}^{\left[-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(x-\mu_1-\frac{\sigma_1^2(z-\mu_1-\mu_2)}{\sigma_1^2+\sigma_2^2}\right)^2 - \frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right]} \mathrm{d}x
\end{aligned}
$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{\left[-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right]} \int_{-\infty}^{+\infty} \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} e^{\left[-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(x-\mu_1-\frac{\sigma_1^2(z-\mu_1-\mu_2)}{\sigma_1^2+\sigma_2^2}\right)^2\right]} dx$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{\left[-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right]}, \tag{3.43}$$

于是证得 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

例 3.15 实际上得到了正态分布的一个重要性质:正态分布关于独立和运算具有封闭性,同时例 3.15 的结论可以写成更一般的形式:X,Y 相互独立且分别服从正态分布 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$,则其任意非零线性组合仍服从正态分布,且

$$aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2), \tag{3.44}$$

其中 a,b 不全为 0,这一结论还可以推广到 n 个随机变量的情形,请读者自己完成.

例 3.16(随机变量的商) 设二维随机向量(X,Y)的密度函数为 $f(x,y)$,求 $Z=\frac{X}{Y}$的密度函数.

解 对任意 z,令 $D_z=\{(x,y)\mid \frac{x}{y}\leqslant z\}$(如图 3.5),则有

$$F_Z(z)=P\left\{\frac{X}{Y}\leqslant z\right\}=\iint_{D_z} f(x,y)\,dx\,dy$$

$$=\int_0^{+\infty}\left[\int_{-\infty}^{zy} f(x,y)\,dx\right]dy+\int_{-\infty}^{0}\left[\int_{zy}^{+\infty} f(x,y)\,dx\right]dy.$$

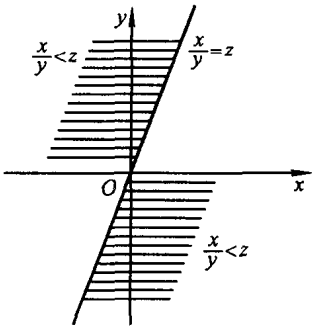


图 3.5 $D_z=\{(x,y)=\frac{x}{y}\leqslant z\}$

于是 Z 的密度函数为

$$f_Z(z)=F_Z'(z)$$

$$=\int_0^{+\infty} yf(zy,y)\,dy-\int_{-\infty}^{0} yf(zy,y)\,dy$$

$$=\int_{-\infty}^{+\infty}|y|f(zy,y)\,dy. \tag{3.45}$$

例 3.17(最大值与最小值) 设 X,Y 的分布函数分别为 $F(x),G(x)$,密度函数分别为 $f(x),g(x)$,且 X 与 Y 相互独立,令 $M=\max\{X,Y\}$,$N=\min\{X,Y\}$,求 M 和 N 的分布函数与密度函数.

解 $F_M(z)=P\{M\leqslant z\}=P\{X\leqslant z,Y\leqslant z\}=P\{X\leqslant z\}P\{Y\leqslant z\}$

$$=F(z)G(z). \tag{3.46}$$

于是 M 的密度函数为

$$f_M(z)=F'_M(z)=F'(z)G(z)+F(z)G'(z)=f(z)G(z)+F(z)g(z). \tag{3.47}$$

而 N 的分布函数为

$$\begin{aligned}F_N(z)&=P\{N\leqslant z\}=P(\{X\leqslant z\}\cup\{Y\leqslant z\})=1-P\{X>z,Y>z\}\\&=1-P\{X>z\}P\{Y>z\}=1-[1-F(z)][1-G(z)].\end{aligned} \tag{3.48}$$

于是 N 的密度函数为

$$f_N(z)=F'_N(z)=f(z)[1-G(z)]+g(z)[1-F(z)]. \tag{3.49}$$

例 3.18　设二维随机向量 (X,Y) 在矩形 $G=\{(x,y)\mid 0\leqslant x\leqslant 2,0\leqslant y\leqslant 1\}$ 上服从均匀分布，试求边长为 X 和 Y 的矩形面积 S 的密度函数 $f(s)$.

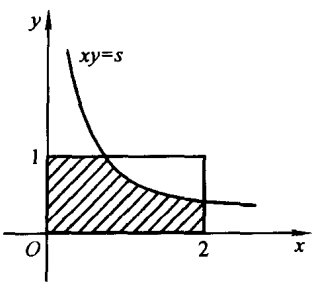


图 3.6

解　二维随机向量 (X,Y) 的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{2}, & (x,y)\in G,\\ 0, & (x,y)\notin G.\end{cases}$$

令 $F(s)$ 为 S 的分布函数，则

$$F(s)=P\{S\leqslant s\}=\iint\limits_{xy\leqslant s} f(x,y)\mathrm{d}x\mathrm{d}y,$$

显然 $s\leqslant 0$ 时 $F(s)=0$，$s\geqslant 2$ 时 $F(s)=1$，而当 $0<s<2$ 时，如图 3.6，有

$$\begin{aligned}\iint\limits_{xy\leqslant s} f(x,y)\mathrm{d}x\mathrm{d}y&=1-\frac{1}{2}\int_s^2 \mathrm{d}x\int_{\frac{s}{x}}^1 \mathrm{d}y\\&=\frac{s}{2}(1+\ln 2-\ln s).\end{aligned}$$

于是

$$F(s)=\begin{cases}0, & s\leqslant 0,\\ \dfrac{s}{2}(1+\ln 2-\ln s), & 0<s<2,\\ 1, & s\geqslant 2.\end{cases}$$

从而

$$f(s)=F'(s)=\begin{cases}\dfrac{1}{2}(\ln 2-\ln s), & 0<s<2,\\ 0, & \text{其他}.\end{cases}$$

三、随机向量的函数的数学期望

设随机向量 (X,Y) 的函数 $Z=g(X,Y)$ 的数学期望存在，为求 EZ，与一元随机变量函数的数学期望一样，并不必先求函数 Z 的分布（事实上，求 Z 的分布

往往十分困难).类似于一元情形,我们可以按以下方式来计算 $Z=g(X,Y)$的数学期望.

如果(X,Y)是二维离散型随机向量,令其概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,\cdots$,则

$$EZ=Eg(X,Y)=\sum_{i,j}g(x_i,y_j)p_{ij}. \tag{3.50}$$

如果(X,Y)是二维连续型随机向量,令其密度函数为 $f(x,y)$,则

$$EZ=Eg(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y. \tag{3.51}$$

例如,设 $g(X,Y)=XY$,则有

$$EXY=\begin{cases}\sum\limits_{i,j}x_iy_jp_{ij}, & \text{若}(X,Y)\text{为离散型},\\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)\mathrm{d}x\mathrm{d}y, & \text{若}(X,Y)\text{为连续型}.\end{cases} \tag{3.52}$$

例 3.19 设(X,Y)的密度函数如例 3.4 的 $g(x,y)$,即

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

求 EXY.

解
$$\begin{aligned}EXY&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyg(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_0^1\mathrm{d}x\int_x^1xy\cdot 8xy\cdot\mathrm{d}y\\&=\frac{4}{9}.\end{aligned}$$

例 3.20 二维离散型随机向量(X,Y)的概率分布由例 3.2 给出,求 EXY.

解
$$\begin{aligned}EXY=&0\times(-1)\times0.1+0\times0\times0.2+0\times2\times0\\&+1\times(-1)\times0.3+1\times0\times0.5+1\times2\times0.1\\&+2\times(-1)\times0.15+2\times0\times0+2\times2\times0.1\\=&0.\end{aligned}$$

例 3.21 一商店经销某种商品,每周进货量 X 与顾客对该商品的需求量 Y 是相互独立的随机变量,且都服从区间$[10,20]$上的均匀分布,商店每售出一单位商品可得利润 1 000 元,若需求量超过进货量,商店可从其他商店调剂供应,这时每单位商品获利润为 500 元,试计算此商品经销商经销该种商品每周所获平均利润.

解 设 Z 表示商店每周所获利润,由题设有

$$Z=g(X,Y)=\begin{cases}1\,000Y, & Y\leqslant X,\\ 1\,000X+500(Y-X)=500(X+Y), & Y>X.\end{cases}$$

由于(X,Y)的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{100}, & 10\leqslant x\leqslant 20,10\leqslant y\leqslant 20,\\ 0, & \text{其他},\end{cases}$$

所以有

$$\begin{aligned}EZ &= \int_{10}^{20}\int_{10}^{20} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y\\ &= \int_{10}^{20}\mathrm{d}y\int_{y}^{20} 1\,000y\cdot\frac{1}{100}\mathrm{d}x+\int_{10}^{20}\mathrm{d}y\int_{10}^{y}500(x+y)\frac{1}{100}\mathrm{d}x\\ &= 10\int_{10}^{20} y(20-y)\mathrm{d}y+5\int_{10}^{20}\left(\frac{3}{2}y^2-10y-50\right)\mathrm{d}y\\ &= \frac{20\,000}{3}+5\times 1\,500\approx 14\,166.67(\text{元}).\end{aligned}$$

四、数学期望的进一步性质

在一元随机变量的讨论中,我们已经给出了数学期望的若干性质,在那里只涉及一个随机变量,下面我们给出的性质则涉及两个或多个随机变量.

(1) 对任意两个随机变量 X,Y,如果其数学期望均存在,则 $E(X+Y)$存在,且

$$E(X+Y)=EX+EY; \tag{3.53}$$

(2) 设 X,Y 为任意两个相互独立的随机变量,数学期望均存在,则 EXY 存在,且

$$EXY=EXEY. \tag{3.54}$$

这两个性质是针对最一般的随机变量给出的,在本书要求范围内,我们只就离散型和连续型进行讨论,这里给出连续型情形的证明,离散型情形留作练习.

证明 设(X,Y)的密度函数为 $f(x,y)$,边缘密度函数为 $f_X(x)$和$f_Y(y)$,据(3.51)有[①]:

$$\begin{aligned}(1)\ E(X+Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\mathrm{d}x\mathrm{d}y\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\mathrm{d}x\mathrm{d}y\\ &= \int_{-\infty}^{+\infty}\left[x\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\right]\mathrm{d}x+\int_{-\infty}^{+\infty}\left[y\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\right]\mathrm{d}y\\ &= \int_{-\infty}^{+\infty}xf_X(x)\mathrm{d}x+\int_{-\infty}^{+\infty}yf_Y(y)\mathrm{d}y\\ &= EX+EY;\end{aligned}$$

① 关于存在性的证明,这里略去,事实上,跟后面的证明过程一样,我们可得 $E(|X|+|Y|)=E|X|+E|Y|$,再由 $E|X+Y|\leqslant E(|X|+|Y|)$,即得 $E(X+Y)$的存在性.由 $E|XY|=E(|X||Y|)=E|X|E|Y|$可得 EXY 的存在性

(2) 由 X, Y 相互独立,知 $f(x,y)=f_X(x)f_Y(y)$,从而

$$\begin{aligned} EXY &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} xf_X(x)\mathrm{d}x \cdot \int_{-\infty}^{+\infty} yf_Y(y)\mathrm{d}y \\ &= EX \cdot EY. \end{aligned}$$

上述两个性质可以推广到 n 个变量情形.

(1)′ $X_1, X_2, \cdots, X_n$ 是任意 n 个随机变量,数学期望均存在,则 $E(X_1+X_2+\cdots+X_n)$存在,且

$$E(X_1+X_2+\cdots+X_n)=EX_1+EX_2+\cdots+EX_n; \tag{3.55}$$

(2)′ 设 $X_1, X_2, \cdots, X_n$ 是 n 个相互独立的随机变量,且数学期望均存在,则 $E(X_1X_2\cdots X_n)$存在,且

$$E(X_1X_2\cdots X_n)=EX_1EX_2\cdots EX_n. \tag{3.56}$$

§3.4 随机向量的数字特征

一、协方差

同随机变量一样,我们希望用一些数字指标来从不同角度综合反映随机向量的分布中的某些重要统计特征,这些数字指标统称为数字特征.由于随机向量的分布中不仅包含单个分量自身的统计规律性,还包含了分量之间相互联系的统计规律性.因而除了各单个分量的数字特征仍是我们需要关注的以外,我们还需要反映分量之间相互联系的一些数字特征."协方差"就是这样的数字特征之一.

定义 3.8 设(X,Y)为二维随机向量,EX, EY 均存在,如果 $E[(X-EX)(Y-EY)]$存在,则称其为随机变量 X 与 Y 的**协方差**,记作 $\mathrm{cov}(X,Y)$.即

$$\mathrm{cov}(X,Y)=E[(X-EX)(Y-EY)]. \tag{3.57}$$

由(3.50)和(3.51)知,如果(X,Y)为离散型随机向量,概率分布为

$$P\{X=x_i, Y=y_j\}=p_{ij}, i,j=1,2,\cdots,$$

则 X 与 Y 的协方差可由下式计算:

$$\mathrm{cov}(X,Y)=\sum_{i,j}(x_i-EX)(y_j-EY)p_{ij}. \tag{3.58}$$

如果(X,Y)是连续型随机向量,密度函数为 $f(x,y)$,则 X 与 Y 的协方差可按下式计算:

$$\operatorname{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-EX)(y-EY)f(x,y)\mathrm{d}x\mathrm{d}y. \tag{3.59}$$

此外,我们常常通过将(3.57)化简为

$$\operatorname{cov}(X,Y)=EXY-EXEY \tag{3.60}$$

来计算协方差.

下一定理总结了协方差的一些基本性质,其证明是容易的,我们将它留给读者完成.

定理 3.5 (1) $\operatorname{cov}(X,X)=DX$;

(2) $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$;

(3) $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$,$a,b$ 为任意常数;

(4) $\operatorname{cov}(C,X)=0$,$C$ 为任意常数;

(5) $\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y)$;

(6) 如果 X 与 Y 相互独立,则 $\operatorname{cov}(X,Y)=0$.

推论 设 X,Y 为任意两个随机变量,如果其方差均存在,则 $X+Y$ 的方差也存在,且

$$D(X+Y)=DX+DY+2\operatorname{cov}(X,Y). \tag{3.61}$$

特别地,如果 X 与 Y 相互独立,则

$$D(X+Y)=DX+DY. \tag{3.62}$$

证明 可以证明(这里略去),如果 X,Y 的方差存在,则协方差 $\operatorname{cov}(X,Y)$ 一定存在且满足下列不等式:

$$|\operatorname{cov}(X,Y)|\leqslant E|(X-EX)(Y-EY)|\leqslant\sqrt{DX}\sqrt{DY}. \tag{3.63}$$

由协方差的性质(1)和(5)有

$$\begin{aligned}D(X+Y)&=\operatorname{cov}(X+Y,X+Y)\\&=\operatorname{cov}(X,X)+2\operatorname{cov}(X,Y)+\operatorname{cov}(Y,Y)\\&=DX+DY+2\operatorname{cov}(X,Y).\end{aligned}$$

当 X 与 Y 相互独立时,由协方差的性质(6)知 $\operatorname{cov}(X,Y)=0$,从而

$$D(X+Y)=DX+DY.$$

例 3.22 已知离散型随机向量(X,Y)的概率分布由例 3.2 给出,求 $\operatorname{cov}(X,Y)$.

解 容易求得 X 的概率分布为

$$P\{X=0\}=0.3,P\{X=1\}=0.45,P\{X=2\}=0.25;$$

Y 的概率分布为

$$P\{Y=-1\}=0.55,P\{Y=0\}=0.25,P\{Y=2\}=0.2.$$

于是有

$$EX=0\times0.3+1\times0.45+2\times0.25=0.95,$$

$$EY=(-1)\times 0.55+0\times 0.25+2\times 0.2=-0.15.$$

在例 3.20 中,我们已算得

$$EXY=0,$$

于是

$$\operatorname{cov}(X,Y)=EXY-EXEY=0.95\times 0.15=0.142\,5.$$

例 3.23 设连续型随机向量(X,Y)的密度函数由例 3.4 中 $g(x,y)$给出,求 $\operatorname{cov}(X,Y)$和 $D(X+Y)$.

解 由于(X,Y)的密度函数为

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

在例 3.4 中,我们已求得其边缘密度函数分别为

$$g_X(x)=\begin{cases}4x(1-x^2), & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}$$

$$g_Y(y)=\begin{cases}4y^3, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

于是

$$EX=\int_{-\infty}^{+\infty}xg_X(x)\mathrm{d}x=\int_0^1 x\cdot 4x(1-x^2)\mathrm{d}x=\frac{8}{15},$$

$$EY=\int_{-\infty}^{+\infty}yg_Y(y)\mathrm{d}y=\int_0^1 y\cdot 4y^3\mathrm{d}y=\frac{4}{5},$$

在例 3.19 中我们已算得

$$EXY=\frac{4}{9},$$

从而得

$$\operatorname{cov}(X,Y)=EXY-EXEY=\frac{4}{9}-\frac{4}{5}\times\frac{8}{15}=\frac{4}{225}.$$

又

$$EX^2=\int_{-\infty}^{+\infty}x^2g_X(x)\mathrm{d}x=\int_0^1 x^2\cdot 4x(1-x^2)\mathrm{d}x=\frac{1}{3},$$

$$EY^2=\int_{-\infty}^{+\infty}y^2g_Y(y)\mathrm{d}y=\int_0^1 y^2\cdot 4y^3\mathrm{d}y=\frac{2}{3},$$

从而得

$$DX=EX^2-(EX)^2=\frac{1}{3}-\left(\frac{8}{15}\right)^2=\frac{11}{225},$$

$$DY=EY^2-(EY)^2=\frac{2}{3}-\left(\frac{4}{5}\right)^2=\frac{2}{75}.$$

故

$$D(X+Y)=DX+DY+2\operatorname{cov}(X,Y)$$

$$= \frac{11}{225} + \frac{2}{75} + \frac{8}{225}$$

$$= \frac{1}{9}.$$

下面的定理将(3.61)和(3.62)推广到 n 维随机向量的情形.

定理 3.6 设$(X_1, X_2, \cdots, X_n)$是 n 维随机向量,如果 $X_i (i = 1, 2, \cdots, n)$的方差均存在,则对任意实向量$(\lambda_1, \lambda_2, \cdots, \lambda_n)$, $\sum_{i=1}^{n} \lambda_i X_i$的方差必存在,且

$$D\left(\sum_{i=1}^{n} \lambda_i X_i\right) = \sum_{i=1}^{n} \lambda_i^2 DX_i + 2 \sum_{1 \leqslant i < j \leqslant n} \lambda_i \lambda_j \operatorname{cov}(X_i, X_j). \tag{3.64}$$

特别地,如果 $X_1, X_2, \cdots, X_n$ 两两独立,则

$$D\left(\sum_{i=1}^{n} \lambda_i X_i\right) = \sum_{i=1}^{n} \lambda_i^2 DX_i. \tag{3.65}$$

该定理的证明与(3.61)和(3.62)的证明完全类似,留作练习.

二、协方差矩阵

定义 3.9 设$(X_1, X_2, \cdots, X_n)$为一个 n 维随机向量,X_i 的方差 $DX_i (i = 1, 2, \cdots, n)$均存在,则以 $\sigma_{ij} = \operatorname{cov}(X_i, Y_j)$为第$(i, j)$元素$(i, j = 1, 2, \cdots, n)$的矩阵 $(\sigma_{ij})_{n \times n}$称为随机向量$(X_1, X_2, \cdots, X_n)$的**协方差矩阵**,简称**协差阵**.

如果记 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)'$,其协差阵通常记作 $D\boldsymbol{X}$.容易验证(3.64)可表示为矩阵形式.对任意实向量 $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)'$,有

$$D(\lambda' \boldsymbol{X}) = \lambda' D\boldsymbol{X} \lambda. \tag{3.66}$$

由(3.66)知,对任意实向量 $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)'$,

$$\lambda' D\boldsymbol{X} \lambda = D(\lambda' \boldsymbol{X}) \geqslant 0.$$

且由于 DX 显然为对称阵,从而知协差阵 $D\boldsymbol{X}$ 一定是非负定阵.

三、相关系数

从协方差的定义可以看出,协方差是对两个随机变量的协同变化的度量,因而反映了随机变量间的相互联系的内容,然而协方差的值还受各随机变量自身取值水平的影响,比如 X 和 Y 同时增大到 K 倍,即 $X_1 = KX$, $Y_1 = KY$,这时 X_1 与 Y_1 之间的统计关系和 X 与 Y 之间的统计关系应该是一样的,然而协方差却增大到 K^2 倍,即

$$\operatorname{cov}(X_1, Y_1) = K^2 \operatorname{cov}(X, Y).$$

为了避免随机变量自身取值水平对度量相互关系的影响,我们先将每个随机变量标准化,即取

$$X^* = \frac{X - EX}{\sqrt{DX}}, Y^* = \frac{Y - EY}{\sqrt{DY}}.$$

然后计算 $\mathrm{cov}(X^*, Y^*)$,将 $\mathrm{cov}(X^*, Y^*)$作为 X 与 Y 之间的相互关系的一种度量,易知,

$$\mathrm{cov}(X^*, Y^*) = \frac{\mathrm{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}}. \tag{3.67}$$

定义 3.10 设(X, Y)是一个二维随机向量,X 和 Y 的方差均存在,且均为正,则称

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} \tag{3.68}$$

为 X 与 Y 之间的**相关系数**.

由(3.63)知,随机变量间的相关系数恒满足

$$|\rho_{X,Y}| \leqslant 1.$$

例 3.24 设 X, Y 是两个随机变量,且 $Y = aX + b$,$(a \neq 0, a$ 与 b 均为常数$)$,DX 存在且不为零,求 $\rho_{X,Y}$.

解 $\mathrm{cov}(X, Y) = \mathrm{cov}(X, aX + b) = a\,\mathrm{cov}(X, X) = aDX$,

$$DY = D(aX + b) = a^2 DX,$$

于是

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} = \frac{aDX}{\sqrt{DX}\sqrt{a^2 DX}} = \frac{a}{|a|}. \tag{3.69}$$

故,当 $a > 0$ 时,$\rho_{X,Y} = 1$;当 $a < 0$ 时,$\rho_{X,Y} = -1$.

例3.24 表明,当 X 与 Y 之间具有线性函数关系时,相关系数的绝对值 $|\rho_{X,Y}|$达到最大值 1.事实上例 3.24 中的条件可以略为放宽,只要

$$P\{Y = aX + b\} = 1. \tag{3.70}$$

即除一个零概率事件以外,X 与 Y 之间具有线性函数关系,则必有 $|\rho_{X,Y}| = 1$.通常将零概率事件忽略不计,如果(3.70)成立,则称 X 与 Y 之间存在**线性函数关系**,或简称为 X 与 Y 具有线性关系.那么 $|\rho_{X,Y}| = 1$ 是否是对 X 与 Y 具有线性关系的完整刻画呢?下述定理对此作出了肯定的回答.

定理 3.7 设(X, Y)是一个二维随机向量,DX, DY 均存在且为正,则 $|\rho_{X,Y}| = 1$的充要条件是 X 与 Y 具有线性关系,即存在常数 $a \neq 0$ 及常数 b,使得

$$P\{Y = aX + b\} = 1.$$

而且,当 $a > 0$ 时,$\rho_{XY} = 1$;当 $a < 0$ 时,$\rho_{XY} = -1$.

该定理的证明略去,有兴趣的读者可以参考有关文献.

与 $|\rho_{X,Y}| = 1$ 相对的另一个极端情形是 $\rho_{X,Y} = 0$.比如,当随机变量 X 与 Y

独立时，设 DX, DY 均存在且为正，则由定理 3.5(6)知 $\operatorname{cov}(X, Y)=0$，从而 $\rho_{X,Y}=0$. 但当 $\rho_{X,Y}=0$ 时，X 与 Y 不一定相互独立(下面的例 3.25 将对此说明). 事实上，两个随机变量相互独立表明随机变量取值之间不存在任何联系，而 $\rho_{X,Y}=0$ 上表明 X 与 Y 之间不存在线性联系，此时，称 X 与 Y **不相关**.

易见，如果 DX, DY 均存在且为正，那么 X 与 Y 不相关等价下列任何一个条件：

(1) $\operatorname{cov}(X, Y)=0$；

(2) $EXY=EXEY$；

(3) $D(X+Y)=DX+DY$.

例 3.25 设 θ 服从$[-\pi,\pi]$上的均匀分布，

$$X=\sin\theta, Y=\cos\theta,$$

判断 X 与 Y 是否不相关，是否独立.

解 由于

$$EX=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin\theta\mathrm{d}\theta=0, EY=\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos\theta\mathrm{d}\theta=0,$$

$$DX=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin^2\theta\mathrm{d}\theta=\frac{1}{2}, DY=EY^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^2\theta\mathrm{d}\theta=\frac{1}{2},$$

$$EXY=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin\theta\cos\theta\mathrm{d}\theta=0.$$

所以，$EXY=EXEY$，从而 X 与 Y 不相关. 但由于 X 与 Y 满足关系

$$X^2+Y^2=1,$$

所以 X 与 Y 不独立.

介于 $|\rho_{X,Y}|=1$ 和 $\rho_{X,Y}=0$ 这两个极端情形之间的一般情形是 $|\rho_{X,Y}|<1$，此时，X 与 Y 之间的关系既不是准确的线性函数关系，但也不是没有任何线性联系，比如，设 X 与 Y 满足下列关系：

$$Y=aX+b+\varepsilon, \tag{3.71}$$

其中 $a\neq0$，a 与 b 均为常数，ε 是与 X 独立的零均值的随机变量，显然 Y 与 X 之间不是线性函数关系，因为 Y 的取值不能由 X 的取值所决定，但 Y 与 X 之间也存在一定的线性联系，当给定 X 的取值时，Y 的取值部分地由 X 的线性函数 $aX+b$ 所确定，但 Y 的取值同时还受与 X 独立的因素(变量)ε 所影响. 作为练习，读者可以验证，当 X 与 Y 满足(3.71)时，其相关系数 $\rho_{X,Y}$ 满足：$0<|\rho_{X,Y}|<1$.

综上，相关系数在某种意义上度量了两个随机变量之间的线性联系的程度，随着 $|\rho_{X,Y}|$ 从 0 增加到 1，这种线性联系的程度越来越高.

四、条件数学期望

由于随机变量之间存在相互联系，一个随机变量的取值可能会对另一随机变量的分布产生影响，这种影响会在数字特征上得到反映.下面我们将要讨论的就是，当已知一个随机变量的取值时，与之相联系的另一个随机变量的平均水平(数学期望).

定义 3.11 (1) 设(X,Y)是离散型随机向量，在 $Y=y_j$ 的条件下，X 的条件概率分布为

$$P\{X=x_i \mid Y=y_j\}=p_{i|j},\ i=1,2,\cdots,$$

如果

$$\sum_i |x_i|\,p_{i|j}<\infty,$$

则称 $\sum_i x_i p_{i|j}$ 为 X 在 $Y=y_j$ 的条件下的**条件数学期望**，记作 $E[X \mid Y=y_j]$.

(2) 设(X,Y)为连续型随机向量，在 $Y=y$ 的条件下，X 的条件密度函数为 $f_{X|Y}(x \mid y)$，如果

$$\int_{-\infty}^{+\infty} |x|\,f_{X|Y}(x \mid y)\mathrm{d}x<\infty,$$

则称

$$\int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)\mathrm{d}x$$

为 X 在 $Y=y$ 的条件下的**条件数学期望**，记作 $E[X \mid Y=y]$.

例 3.26 设(X,Y)的概率分布由例 3.2 给出，求出 $Y=0$ 的条件下，X 的条件数学期望.

解 在例 3.6 中，我们已求得 $Y=0$ 的条件下，X 的条件概率分布为

$$P\{X=0 \mid Y=0\}=0.8,P\{X=1 \mid Y=0\}=0.2,P\{X=2 \mid Y=0\}=0,$$

于是

$$E[X \mid Y=0]=0\times 0.8+1\times 0.2+2\times 0=0.2.$$

例 3.27 设(X,Y)的密度函数如例 3.8 所示，求 $E[X \mid Y=y]\ (|y|<1)$.

解 由例 3.8 知，$|y|<1$ 时，

$$f_{X|Y}(x \mid y)=\begin{cases}\dfrac{1}{2\sqrt{1-y^2}}, & |x|\leqslant\sqrt{1-y^2},\\ 0, & \text{其他}.\end{cases}$$

从而，$|y|<1$ 时，

$$E[X \mid Y=y]=\int_{-\infty}^{+\infty} x\cdot f_{X|Y}(x \mid y)\mathrm{d}x=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\cdot\frac{1}{2\sqrt{1-y^2}}\mathrm{d}x=0.$$

容易验证条件数学期望具有数学期望具有的所有性质：

(1) C 为常数，则 $E[C|Y=y]=C$；　　(3.72)

(2) k_1,k_2 为常数，且 $E[X_i|Y=y](i=1,2)$均存在，则

$$E[k_1X_1+k_2X_2|Y=y]=k_1E[X_1|Y=y]+k_2E[X_2|Y=y];\quad(3.73)$$

(3) 如果 X 与 Y 独立，则 $E[X|Y=y]=EX$.　　(3.74)

记 $h(y)=E(X|Y=y)$，则随 y 的变化，它是 y 的一个函数.进一步定义随机变量 Y 的函数 $E[X|Y]=h(Y)$，并称之为随机变量 X 关于随机变量 Y 的条件期望.由于它是随机变量 Y 的函数，因而它也是一个随机变量，根据前面列举的 $E[X|Y=y]$的性质，可得 $E[X|Y]$的下列性质：

(1) C 为常数，则 $E[C|Y]=C$；　　(3.75)

(2) k_1,k_2 为常数，则 $E[k_1X_1+k_2X_2|Y]=k_1E[X_1|Y]+k_2E[X_2|Y]$；　　(3.76)

(3) X 与 Y 独立，则 $E[X|Y]=EX$.　　(3.77)

除此之外，我们还可得到下列性质：

(4) $g(x)$是一个任意函数，则 $E[g(Y)X|Y]=g(Y)E[X|Y]$，特别地，有 $E[g(Y)|Y]=g(Y)$；　　(3.78)

(5) 全期望公式.$E(E[X|Y])=EX$.　　(3.79)

证明　(4) 由于当 $Y=y$ 时，$g(Y)\equiv g(y)$(为常数)，从而有

$$E[g(Y)X|Y=y]=E[g(y)X|Y=y]=g(y)E[X|Y=y],$$

于是得

$$E[g(Y)X|Y]=g(Y)E[X|Y].$$

(5) 我们只证连续型情形，离散型留作练习.

记 $g(y)=E[X|Y=y]$，从而 $g(Y)=E[X|Y]$，

$$\begin{aligned}E(E[X|Y])=E(g(Y))&=\int_{-\infty}^{+\infty}g(y)f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}E[X|Y=y]f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}xf_{X|Y}(x|y)\mathrm{d}x\right]f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[xf_{X|Y}(x|y)f_Y(y)]\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[xf(x,y)]\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}x\left[\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\right]\mathrm{d}x\\&=\int_{-\infty}^{+\infty}xf_X(x)\mathrm{d}x\\&=EX.\end{aligned}$$

例 3.28 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,求 $E[X|Y]$.

解 由于在 $Y=y$ 的条件下,X 的条件分布为

$$N\left(\mu_1+\frac{\sigma_1}{\sigma_2}\rho(y-\mu_2),\sigma_1^2(1-\rho^2)\right),$$

从而

$$E[X|Y=y]=\mu_1+\frac{\sigma_1}{\sigma_2}\rho(y-\mu_2), \tag{3.80}$$

于是得

$$E[X|Y]=\mu_1+\frac{\sigma_1}{\sigma_2}\rho(Y-\mu_2). \tag{3.81}$$

例 3.29 设 X,Y 满足下列关系

$$Y=aX+b+\varepsilon, \tag{3.82}$$

其中 ε 是一个与 X 独立的、期望为 0 的随机变量,求 $E[Y|X]$.

解
$$\begin{aligned}E[Y|X]&=E[aX+b+\varepsilon|X]\\&=E[aX|X]+E[b|X]+E[\varepsilon|X]\\&=aX+b+E\varepsilon\\&=aX+b.\end{aligned} \tag{3.83}$$

五、条件期望的预测含义

在数学期望的讨论中,我们曾指出,在均方误差最小意义下,数学期望 EX 实际上是对 X 的最优点值预测,即

$$E(X-EX)^2=\min_C E(X-C)^2. \tag{3.84}$$

在许多实际问题中,为对某随机变量 X 进行预测,我们对一个与 X 有关的随机变量 Y 进行观察,通常称 Y 为一个信息变量.一旦得到 Y 的观察值,由于 X 与 Y 存在一定的统计联系,我们便能利用这种联系来对 X 进行预测,其预测值依 Y 的取值而定,也就是说 X 的预测值是 Y 的函数 $g(Y)$(通常称为**预测函数**).为寻求一个合理的预测,实际上就是寻求一个适当的函数 g_0,使得在一定准则下 $g_0(Y)$是 X 的最优预测.比如在均方误差最小意义下,$g_0(Y)$作为 X 的最优预测应满足

$$E[X-g_0(Y)]^2=\min_g E[X-g(Y)]^2. \tag{3.85}$$

另一方面,当我们得到 Y 的观察值 y 时,利用 X 与 Y 的统计联系,我们重新将 X 的分布修正为在 $Y=y$ 的条件下 X 的条件分布.在这一新的分布下,直观上容易想象,在均方误差最小意义下 X 的新的合理预测应该是新的分布下的数学期望,即 $E[X|Y=y]$,可见 $E[X|Y]$应该是一个合理的预测函数.严格地讲,由(3.85),为说明其合理性,我们需要证明

$$E(X-E(X|Y))^2=\min_g E(x-g(Y))^2,$$

其证明过程大致与(3.84)的证明相似.

$$\begin{aligned}E(X-g(Y))^2&=E((X-E[X|Y])+(E[X|Y]-g(Y)))^2\\&=E(X-E[X|Y])^2+2E(X-E[X|Y])(E[X|Y]-g(Y))\\&\quad+E(E[X|Y]-g(Y))^2,\end{aligned}$$

又由(3.79)和(3.78)可知

$$\begin{aligned}&E(X-E[X|Y])(E[X|Y]-g(Y))\\&=E(E[(X-E[X|Y])(E[X|Y]-g(Y))|Y])\\&=E((E[X|Y]-g(Y))E[(X-E[X|Y])|Y])\\&=E((E[X|Y]-g(Y))(E[X|Y]-E[X|Y])\\&=0,\end{aligned}$$

从而有

$$\begin{aligned}E(X-g(Y))^2&=E(X-E[X|Y])^2+E(E[X|Y]-g(Y))^2\\&\geqslant E(X-E[X|Y])^2.\end{aligned}$$

§3.5 大数定律与中心极限定理

一、依概率收敛

在微积分中,收敛性及其极限是一个基本而重要的概念.一个数列 a_n 收敛到 a,记作

$$\lim_{n\to\infty}a_n=a,$$

是指对任意 $\varepsilon>0$,总存在正整数 N,当 $n>N$ 时恒有

$$|a_n-a|<\varepsilon.$$

在概率论中,我们研究的对象是随机变量,因而,我们需要考虑随机变量序列的收敛性.设 $X_1,X_2,\cdots,X_n,\cdots$ 是一列随机变量,如果我们以定义数列的极限完全相同的方式来定义随机变量序列的收敛性,那么 $X_1,X_2,\cdots,X_n,\cdots$ 收敛到一个随机变量 X 是指,对任意 $\varepsilon>0$,存在正整数 N,当 $n>N$ 时,恒有

$$|X_n-X|<\varepsilon.$$

由于 X_n,X 均为随机变量,于是 $|X_n-X|$ 也是随机变量,要求一个随机变量取值小于给定的足够小的 ε 未免太苛刻了,这种苛刻的收敛性关系,在概率论的理论和应用问题中只能是一些罕见的特例,因而对概率论中进一步问题的研究意义并不大.为此,我们需要对上述定义进行修正,以适合随机变量本身的特征.一

种比较普遍的修正方式是:我们并不要求 $n>N$ 时,$|X_n-X|<\varepsilon$ 恒成立,而只要求 n 足够大时,出现

$$|X_n-X|>\varepsilon$$

的可能性可以任意小,于是有下列定义.

定义 3.12 设 $X,X_1,X_2,\cdots,X_n,\cdots$ 是一列随机变量,如果对任意 $\varepsilon>0$,恒有

$$\lim_{n\to\infty}P\{|X_n-X|>\varepsilon\}=0, \tag{3.86}$$

则称 $\{X_n\}$ **依概率收敛到** X,记作 $X_n\xrightarrow{P}X$ 或 $P-\lim_{n\to\infty}X_n=X$.

二、大数定律

在第一章中,我们曾指出,如果一个事件 A 的概率为 p,那么大量重复试验中事件 A 发生的频率将逐渐稳定(靠近)到概率 p,这只是一种直观的说法.我们将对这一说法给出严格的数学表述和论证.记 n 次试验中事件 A 发生的次数为 μ_n,那么事件 A 在 n 次试验中发生的频率为 $\frac{\mu_n}{n}$,μ_n 和 $\frac{\mu_n}{n}$ 均为随机变量,我们将 $\frac{\mu_n}{n}$ 逐渐稳定到常数 p(A 的概率)表述为:$\frac{\mu_n}{n}\xrightarrow{P}p$,下面的定理说明了这一表述的正确性,从而为"频率稳定到概率"这一经验事实提供了理论依据.

定理 3.8(伯努利大数定律) 设 μ_n 是 n 重伯努利试验中事件 A 发生的次数,已知在每次试验中 A 发生的概率为 $p(0<p<1)$,则对任意 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|>\varepsilon\right\}=0, \tag{3.87}$$

即 $\frac{\mu_n}{n}\xrightarrow{P}p$ 或 $P-\lim_{n\to\infty}\frac{\mu_n}{n}=p$.

证明 显然,$E\left(\frac{\mu_n}{n}\right)=p$,$D\left(\frac{\mu_n}{n}\right)=\frac{1}{n^2}D(\mu_n)=\frac{pq}{n}$,其中 $q=1-p$,由切比雪夫不等式,对任意给定 $\varepsilon>0$,

$$P\left\{\left|\frac{\mu_n}{n}-p\right|>\varepsilon\right\}\leqslant\frac{pq}{n\varepsilon^2},$$

从而有

$$0\leqslant\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|>\varepsilon\right\}\leqslant\lim_{n\to\infty}\frac{pq}{n\varepsilon^2}=0,$$

故有

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|>\varepsilon\right\}=0.$$

如果记

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次试验中 } A \text{ 发生},\\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不发生},\end{cases}\quad i=1,2,\cdots,n,$$

则 $\mu_n=\sum_{i=1}^{n}\xi_i$，$E\xi_i=p$，从而(3.87)可改写为

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}E\sum_{i=1}^{n}\xi_i\right|<\varepsilon\right\}=1.$$

其中 $\xi_i(i=1,2,\cdots)$是相互独立服从同一个 0－1 分布的随机变量序列.我们不难将其推广到更一般的情形，得到下列定理.

定理 3.9(切比雪夫大数定律) 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是一列两两不相关的随机变量，它们的数学期望 $E\xi_i$ 和方差 $D\xi_i$ 均存在，且方差有界，即存在常数 C，使得 $D\xi_i\leqslant C(i=1,2,\cdots)$，则对任意 $\varepsilon>0$ 有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}\sum_{i=1}^{n}E\xi_i\right|<\varepsilon\right\}=1. \tag{3.88}$$

证明 由切比雪夫不等式，有

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}E\sum_{i=1}^{n}\xi_i\right|\geqslant\varepsilon\right\}\leqslant\frac{D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)}{\varepsilon^2}$$

$$=\frac{D\left(\sum_{i=1}^{n}\xi_i\right)}{n^2\varepsilon^2}=\frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}D(\xi_i)\leqslant\frac{n\cdot C}{n^2\varepsilon^2}=\frac{C}{n\varepsilon^2}\to 0,$$

从而

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}\sum_{i}^{n}E\xi_i\right|<\varepsilon\right\}=1.$$

推论 1 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是一列独立同分布的随机变量，其数学期望和方差均存在，记 $E\xi_i=\mu$，则对任意 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|<\varepsilon\right\}=1. \tag{3.89}$$

即

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i\xrightarrow{P}\mu.$$

推论 1 中要求方差存在，实际上这一条件可以去掉，相应的结论仍然成立.

定理 3.10(辛钦大数定律) 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是一列相互独立同分布的随机变量，且数学期望存在，记 $E\xi_i=\mu$，则有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|<\varepsilon\right\}=1. \tag{3.90}$$

伯努利大数定律表明：当 n 很大时，事件发生的频率会“靠近”其概率.而辛

钦大数定律则表明，n 次观察的算术平均值$\frac{1}{n}\sum_{i=1}^{n}\xi_i$会“靠近”它的期望值 $\mu=E\xi_i$，这为估计期望值提供了一条实际可行的途径.

三、中心极限定理

我们知道正态分布在概率论中占有极其重要的地位，实际中许多随机现象均可用正态分布来描述，为什么会这样呢？其原因与高斯对误差的研究中使用正态分布来描述误差的分布是相似的，我们不妨通过“误差”的考察来加以说明. 以一门炮的射程为例. 一门炮生产出来以后其制造技术和工艺等内在因素决定了其射程的基准值，但在每次射击中，由于震动会造成误差 ξ_1，每发炮弹外形上的细小差别会造成误差 ξ_2，每发炮弹内炸药数量或质量上的微小差异会造成误差 ξ_3，炮弹在前进中遇到空气流速的微小扰动而造成误差 ξ_4 等等. 这些误差有的为正，有的为负，都是随机的，而炮弹射程的总误差 ξ 是许多这种随机小误差的总和，即

$$\xi=\sum_i \xi_i .$$

而且这许多小误差 ξ_i 可以看成是相互独立的，因此要讨论独立随机变量和的分布问题. 中心极限定理要研究的正是大量的独立随机变量和的近似分布问题，其结论将告诉我们，ξ 实际上近似服从正态分布. 继而射程等于基准值加上总误差 ξ，也近似服从正态分布. 实际中，许多随机现象与上面的例子类似，我们考虑某个量受许多随机因素（主导因素除外）的共同影响而随机取值，那么它的分布便会近似服从正态分布，为了使问题简单并便于掌握，我们在这里讨论的是中心极限定理中比较特殊的情形——独立同分布的随机变量和.

定理 3.11（林德伯格－勒维） 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是一列独立同分布的随机变量，且 $E\xi_i=\mu, D(\xi_i)=\sigma^2>0, i=1,2,\cdots$，则有

$$\lim_{n\to\infty}P\left\{\frac{\sum_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\sigma}\leqslant x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t. \tag{3.91}$$

该定理的证明需要进一步的知识，超出了本书要求，在此不作介绍.

定理 3.11 实际上说明独立同分布随机变量和 $\sum_{i=1}^{n}\xi_i$ 的标准化变量在 $n\to\infty$ 时渐近服从标准正态 $N(0,1)$，通常记作

$$\frac{\sum_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\sigma}\overset{a}{\sim}N(0,1). \tag{3.92}$$

由此得知，当 n 充分大时，$\sum_{i=1}^{n}\xi_i \overset{a}{\sim} N(n\mu, n\sigma^2)$，$\frac{1}{n}\sum_{i=1}^{n}\xi_i \overset{a}{\sim} N\left(\mu, \frac{1}{n}\sigma^2\right)$.

我们来比较一下大数定律与中心极限定理：大数定律实际上告诉我们，当 n 趋向于无穷大时，独立同分布随机变量的算术平均值 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 依概率收敛到其期望值 μ，即对任意指定的 $\varepsilon>0$，有 $P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|>\varepsilon\right\}\to 0$. 那么，对固定的 $\varepsilon>0$，n 充分大时，$\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|>\varepsilon\right\}$ 的概率究竟有多大呢？大数定律没有告诉我们任何内容，但中心极限定理表明

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu \overset{a}{\sim} N\left(0,\frac{1}{n}\sigma^2\right).$$

因而

$$\begin{aligned}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|>\varepsilon\right\}&=1-P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|\leqslant\varepsilon\right\}\\&=1-P\left\{\left|\frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu}{\sigma/\sqrt{n}}\right|\leqslant\frac{\varepsilon}{\sigma/\sqrt{n}}\right\}\\&\approx 1-\left[2\Phi_0\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)-1\right]\\&=2\left[1-\Phi_0\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)\right].\end{aligned}\tag{3.93}$$

其中 $\Phi_0(x)$ 是 $N(0,1)$ 的分布函数，且 $\Phi_0\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)\to 1$. 因而(3.93)不仅导出 $P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|>\varepsilon\right\}\to 0$，而且得到了对较大的 n，$\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu\right|>\varepsilon\right\}$ 的概率近似值，可见，中心极限定理比大数定律的结论更加“精细”.

例 3.30　一盒同型号螺丝钉共有 100 个，已知该型号的螺丝钉的重量是一个随机变量，期望值是 100 g，标准差是 10 g，求一盒螺丝钉的重量超过 10.2 kg 的概率.

解　设 ξ_i 为第 i 个螺丝钉的重量($i=1,2,\cdots,100$)，且它们之间独立同分布，于是一盒螺丝钉的重量为 $\xi=\sum_{i=1}^{100}\xi_i$，且由 $E\xi_i=100$，$\sqrt{D(\xi_i)}=10$，知 $E\xi=100\times E\xi_i=10\ 000$，$\sqrt{D\xi}=100$，由中心极限定理有

$$P\{\xi>10\ 200\}=P\left\{\frac{\xi-10\ 000}{100}>\frac{10\ 200-10\ 000}{100}\right\}$$

$$=P\left\{\frac{\xi-10\,000}{100}>2\right\}$$

$$=1-P\left\{\frac{\xi-10\,000}{100}\leqslant 2\right\}$$

$$\approx 1-\Phi_0(2)=1-0.977\,25=0.022\,75.$$

作为林德伯格-勒维中心极限定理的推论,我们给出历史上著名的棣莫弗-拉普拉斯中心极限定理.

定理 3.12 设 $X_n\sim b(n,p)$,$0<p<1$,则

$$\lim_{n\to\infty}P\left\{\frac{X_n-np}{\sqrt{npq}}\leqslant x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt. \tag{3.94}$$

定理 3.12 表明,当 n 充分大时,二项分布可用正态分布来近似.比如,设 $X_n\sim b(n,p)$,要计算 $P\{a\leqslant X_n<b\}=\sum\limits_{a\leqslant k<b}C_n^k p^k q^{n-k}$,当 n 很大时,这个计算量是相当大的.现在,根据定理 3.12,由于 $\dfrac{X_n-np}{\sqrt{npq}}\overset{a}{\sim}N(0,1)$ 或等价地 $X_n\overset{a}{\sim}N(np,npq)$,于是可以近似地用正态分布来计算上述概率.即

$$P\{a\leqslant X_n<b\}=P\left\{\frac{a-np}{\sqrt{npq}}\leqslant\frac{X_n-np}{\sqrt{npq}}<\frac{b-np}{\sqrt{npq}}\right\}$$

$$\approx\Phi_0\left(\frac{b-np}{\sqrt{npq}}\right)-\Phi_0\left(\frac{a-np}{\sqrt{npq}}\right).$$

只要查一查标准正态分布函数表就很容易得到 $P\{a\leqslant X_n<b\}$ 的相当精确的值.

例 3.31 设某电站供电网有 10 000 盏电灯,夜晚每盏灯开灯的概率为0.7,而假定开关时间彼此独立,估计夜晚同时开着的灯的盏数在 6 800 与 7 200 之间的概率.

解 ξ 表示在夜晚同时开着的灯的盏数,则 $\xi\sim b(10\,000,0.7)$,于是 $E\xi=10\,000\times0.7=7\,000$,$D\xi=10\,000\times0.7\times0.3=2\,100$.由中心极限定理,有

$$P\{6\,800<\xi<7\,200\}=P\left\{\frac{-200}{\sqrt{2\,100}}<\frac{\xi-7\,000}{\sqrt{2\,100}}<\frac{200}{\sqrt{2\,100}}\right\}$$

$$=P\left\{\left|\frac{\xi-7\,000}{45.83}\right|<4.36\right\}$$

$$\approx 2\Phi_0(4.36)-1=0.999\,99.$$

例 3.32 某仪器由 n 个电子元件组成,每个电子元件的寿命服从[0,1 000]上的均匀分布(单位:h),当有 20% 的元件烧坏时,仪器便报废,求为使该仪器的寿命超过 100 h 的概率不低于 0.95,n 至少为多大?

解 设 X 表示仪器的寿命,X_i 表示第 i 个电子元件的寿命,记 $A_i=\{X_i\geqslant$

100|，ξ 表示 n 个事件 A_i $(i=1,2,\cdots,n)$ 中发生的个数，由 $P\{A_i\}=0.9$，知 $\xi\sim b(n,0.9)$，显然 $\{X\geqslant 100\}=\left\{\frac{\xi}{n}\geqslant 0.8\right\}$，于是有

$$\begin{aligned}P\{X\geqslant 100\}&=P\left\{\frac{\xi}{n}\geqslant 0.8\right\}=P\{\xi>0.8n\}\\&=P\left\{\frac{\xi-0.9n}{\sqrt{n\times 0.9\times 0.1}}>\frac{0.8n-0.9n}{\sqrt{n\times 0.9\times 0.1}}\right\}\\&=1-\Phi_0\left(-\frac{\sqrt{n}}{3}\right)=\Phi_0\left(\frac{\sqrt{n}}{3}\right)\geqslant 0.95,\end{aligned}$$

从而 $\frac{\sqrt{n}}{3}\geqslant 1.64$，$n\geqslant 25$.

习 题 三

(A)

1. 设 $F(x,y)$ 是一个二维随机向量 (X,Y) 的分布函数，$x_1<x_2$，$y_1<y_2$，证明：

$$F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\geqslant 0.$$

2. 一个袋内装有5个白球，3个红球.第一次从袋内任意取一个球，不放回，第二次又从袋内任意取两个球，X_i 表示第 i 次取到的白球数 $(i=1,2)$.求(1) (X_1,X_2) 的分布及边缘分布；(2) $P\{X_1=0,X_2\neq 0\}$，$P\{X_1=X_2\}$，$P\{X_1X_2=0\}$.

3. 设二维随机向量 (X_1,Y_1) 及 (X_2,Y_2) 的密度函数 $f(x,y)$ 及 $g(x,y)$ 分别为：

$$f(x,y)=\begin{cases}k_1e^{-3x-4y}, & x>0,y>0,\\0, & \text{其他；}\end{cases}$$

$$g(x,y)=\begin{cases}k_2e^{-3x-4y}, & x>y>0,\\0, & \text{其他.}\end{cases}$$

求(1)常数 k_1,k_2；(2) 边缘密度函数.

4. 设 (X,Y) 服从 $G=\{(x,y)|0\leqslant x\leqslant 2,0\leqslant y\leqslant 1\}$ 上的均匀分布，求(1) (X,Y) 的密度函数及分布函数；(2) X 和 Y 的边缘密度函数和边缘分布函数；(3) $P\{Y<X^2\}$.

5. 设 (X,Y) 服从 $G=\{(x,y)|1>y>x>0\}$ 上的均匀分布，求(1) (X,Y) 的密度函数；(2) X 和 Y 的边缘密度函数.

6. 已知 X 和 Y 的分布函数 $F_X(x)$，$F_Y(y)$ 分别为：

$$F_X(x)=\begin{cases}0, & x<0,\\\frac{x}{2}, & 0\leqslant x\leqslant 2,\\1, & x>2;\end{cases}\qquad F_Y(y)=\begin{cases}0, & y<1,\\y-1, & 1\leqslant y\leqslant 2,\\1, & y>2,\end{cases}$$

且 X 与 Y 相互独立.(1) 求 (X,Y) 的分布函数；(2) 令 $\xi=X^2$，$\eta=Y^2$，求 (ξ,η) 的分布函数 $G(x,y)$；(3) 求 $P\left\{X<1,Y>\frac{3}{2}\right\}$.

7. 设 X 与 Y 独立，证明：对任意实数 $x_1, x_2, y_1, y_2 (x_1 < x_2; y_1 < y_2)$，事件 $\{x_1 < X \leqslant x_2\}$ 与事件 $\{y_1 < Y \leqslant y_2\}$ 独立.

8. 求第 2 题中 $X_2 = 1$ 的条件下 X_1 的条件分布.

9. 已知 (X, Y) 为离散型分布，X 的分布为：$P\{X = x_i\} = p_i^X, i = 1,2,\cdots$，对每个 $i (i = 1,2,\cdots)$ 在 $X = x_i$ 的条件下，Y 的条件分布为 $P\{Y = y_j \mid X = x_i\} = p_{j|i}, j = 1,2,\cdots$. (1) 求 (X, Y) 的分布 $P\{X = x_i, Y = y_j\} = p_{ij}, i, j = 1,2,\cdots$；(2) 求 Y 的分布 $P\{Y = y_j\} = p_j^Y$；(3) 求 $Y = y_j$ 的条件下 X 的条件分布 $P\{X = x_i \mid Y = y_j\} = p_{i|j}, i = 1,2,\cdots$.

10. 已知 (X, Y) 的分布及边缘分布如下表：

X \ Y	-1	0	1	p_i^X
0	p_{11}	p_{12}	p_{13}	$\frac{1}{2}$
1	0	p_{22}	0	$\frac{1}{2}$
p_j^Y	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

(1) 求 (X, Y) 的联合分布表中 $p_{11}, p_{12}, p_{13}, p_{22}$ 的值；

(2) 判断 X 与 Y 是否独立.

11. 设 X 与 Y 独立，下表列出了二维随机向量 (X, Y) 的分布、边缘分布中的部分概率值，试将其余概率值填入表中空白处.

X \ Y	y_1	y_2	y_3	p_i^X
x_1		$\frac{1}{8}$		
x_2	$\frac{1}{8}$			
p_j^Y	$\frac{1}{6}$			

12. 设 (X, Y) 是二维离散型随机向量，其分布为 $P\{X = x_i, Y = y_j\} = p_{ij}, i = 1,2,\cdots, m, j = 1,2,\cdots, n$，称 $(p_{ij})_{m\times n}$ 为**联合概率矩阵**. 证明：X 与 Y 相互独立的充要条件是 $(p_{ij})_{m\times n}$ 的秩为 1.

13. 独立投掷一枚均匀骰子两次，记 B, C 为两次中各出现的点数，求一元二次方程 $x^2 + Bx + C = 0$ 有实根的概率 p 和有重根的概率 q.

14. 求第 3 题中的条件密度函数：$f_{X_1|Y_1}(x \mid y)$，$f_{Y_1|X_1}(y \mid x)$，$g_{X_2|Y_2}(x \mid y)$，$g_{Y_2|X_2}(y \mid x)$.

15. 设区域 D 是由直线 $y=x-1, y=x+1, x=2$ 及坐标轴围成的区域.(X,Y)服从区域 D 上的均匀分布.求条件密度函数 $f_{Y|X}(y|x)$和 $f_{X|Y}(x|y)$.

16. 设 $D=\{(x,y)\mid a\leqslant x\leqslant b, \varphi(x)<y<\psi(x)\}$,其中 $\varphi(x),\psi(x)$是两个任意函数,且 $\varphi(x)<\psi(x)$在$[a,b]$上恒成立,(X,Y)服从 D 上的均匀分布.证明条件分布为均匀分布.

17. 设(X,Y)为连续型随机向量,已知 X 的密度函数 $f_X(x)$及对一切 x,在 $X=x$ 的条件下 Y 的条件密度 $f_{Y|X}(y|x)$.(1) 求密度函数 $f(x,y)$;(2) 求 Y 的密度函数 $f_Y(y)$;(3) 求条件密度函数 $f_{X|Y}(x|y)$.

18. 判断第 3 题中 X_1 与 Y_1,X_2 与 Y_2 分别是否独立.

19. 判断第 5 题中 X 与 Y 是否独立.

20. 已知 X 服从$[0,1]$上的均匀分布,$Y\sim N(0,1)$,且 X 与 Y 相互独立,求(X,Y)的密度函数.

21. 举例说明边缘密度不能确定联合密度(以二元正态为例).

22. 如果 $X\sim N(\mu_1,\sigma_1^2)$,$Y\sim N(\mu_2,\sigma_2^2)$那么$(X,Y)$一定服从二元正态分布吗?分析下面的例子:

$$(X,Y)\sim f(x,y)=\frac{1}{2\pi}\mathrm{e}^{\frac{x^2+y^2}{2}}(1+\sin x\sin y).$$

23. (X,Y)如例 3.1 所示.(1) 求 $\xi=X+Y$,$\eta=X-Y$ 各自的分布;(2) 求(ξ,η)的分布.

24. B,C 如第 13 题所示,求 $\xi=B+C$ 的分布.

25. 证明:如果 $X_i\sim P(\lambda_i)$(参数为 λ_i 的泊松分布)$(i=1,2,\cdots,n)$且相互独立,则 $\sum\limits_{i=1}^{n}X_i\sim P\left(\sum\limits_{i=1}^{n}\lambda_i\right)$.

26. 设 X_1,X_2 均服从参数为 λ 的指数分布,且相互独立,求 X_1+X_2 的密度函数.

27. 已知(X,Y)为一个二维随机向量,$X_1=X+Y$,$X_2=X-Y$,(X_1,X_2)的密度函数为

$$\varphi(x_1,x_2)=\frac{1}{2\sqrt{3}\pi}\mathrm{e}^{-\frac{1}{2}\left[\frac{(x_1-4)^2}{3}+(x_2-2)^2\right]}.$$

分别求 X 和 Y 的密度函数.

28. 将(3.44)推广到 n 个相互独立的随机变量的情形.

29. 已知(X,Y)服从 $G=\{(x,y)\mid 0<x\leqslant 2, 0<y\leqslant 1\}$上的均匀分布,求 $\xi=\dfrac{X}{Y}$的分布函数和密度函数.

30. 设(X,Y)的分布为

X \ Y	-1	0	1
0	0.3	0	0.3
1	0.1	0.2	0.1

求 EXY,$E(X+Y)$.

31. 在一次拍卖中,两人竞买一幅名画,拍卖以暗标形式进行,并以最高价成交.设两人的出价相互独立且均服从[1,2]上的均匀分布,求这幅画的期望成交价.

32. 对离散型情形证明(3.53)及(3.54).

33. 一袋中装有 60 个黑球,40 个红球,从中任取 20 个,求取到的红球数的期望值.

34. 求例 1.29 中交通车的停车次数的数学期望.

35. 证明定理 3.3.

36. 求第 2 题中 X_1 与 X_2 的协方差 $\mathrm{cov}(X_1, X_2)$.

37. 求第 3 题中(X_1, Y_1)和(X_2, Y_2)的协方差 $\mathrm{cov}(X_1, Y_1)$,$\mathrm{cov}(X_2, Y_2)$.

38. 求第 27 题中的 $\mathrm{cov}(X, Y)$,$\mathrm{cov}(X_1, X_2)$.

39. 设(X, Y)服从 $G=\{(x,y)\mid x^2+y^2\leqslant 1\}$上的均匀分布,讨论 X 与 Y 的独立性与相关性.

40. 求例 3.23 中 X 与 Y 的相关系数 $\rho_{X,Y}$.

41. 两支股票 A 和 B,在一个给定时期内的收益率 r_A,r_B 均为随机变量,已知 r_A 和 r_B 的协差阵为 $\boldsymbol{V}$:

$$\boldsymbol{V}=\begin{pmatrix}16 & 6\\ 6 & 9\end{pmatrix}.$$

现将一笔资金按比例 x,$1-x$ 分别投资到股票 A 和 B 上形成一个投资组合 P,记其收益率为 r_P.

(1) 求 r_A 和 r_B 的相关系数;

(2) 求 Dr_P;

(3) 在不允许卖空的情况下(即 $0\leqslant x\leqslant 1$),x 为何值时 Dr_P 最小,何时 $Dr_P\leqslant\min(Dr_A, Dr_B)$.

42. 两种证券 A,B 的收益率为 r_A 和 r_B,人们常用收益率的方差来衡量证券的风险,收益率的方差为正的证券称为**风险证券**.如果 A,B 均为风险证券,且$|\rho_{AB}|\neq 1$,证明 A 与 B 的任意投资组合 P(允许卖空)必然也是风险证券;若$|\rho_{AB}|=1$,何时能得到无风险组合?并构造相应的无风险组合;当 ρ_{AB} 满足什么条件时,我们能在不允许卖空的情况下,得到比 A,B 的风险都小的投资组合?

43. 已知某支股票价格变化率 r 与银行利率 r_f 存在一定的联系,设 r 和 r_f 的联合分布如下:

r / r_f	−3%	1%	2%	3%	4%	5%	6%	7%
1%	0.015	0.015	0.045	0.09	0.03	0.06	0.03	0.015
1.5%	0.025	0.05	0.1	0.15	0.075	0.05	0.025	0.025
2%	0.06	0.04	0.03	0.02	0.02	0.02	0.01	0

(1) 求该股票价格的平均变化率;

(2) 如果已知利率 $r_f=1.5\%$,求股票价格的平均变化率.

44. 求第 40 题中 $E[Y|X=x]$和 $E[Y|X]$.

45. 设(X,Y)服从 $D=\{(x,y)|x^2+y^2\leqslant 2x\}$上的均匀分布,求 $E[X|Y]$和 $E[Y|X]$.

46. 计算机有 120 个终端,每个终端在一小时内平均有 3 分钟使用打印机,假定各终端使用打印机与否相互独立,求至少有 10 个终端同时使用打印机的概率.

47. 某车间有同型号机床 200 部,每部开动的概率为 0.7,假定各机床开关是相互独立的,开动时每部要消耗电能 15 个单位,问电厂最少要供应该车间多少单位电能,才能以 95% 的概率保证不致因供电不足而影响生产?

48. 计算机在进行加法时,每个加数取整数(按四舍五入取最为接近的整数),设所有加数的取整误差是相互独立的,且它们服从$[-0.5,0.5]$上的均匀分布.

(1) 若将 300 个数相加,求误差总和绝对值超过 15 的概率;

(2) 至多 n 个数加在一起,其误差总和的绝对值小于 10 的概率为 0.9,求 n 值.

(B)

1. 一个盒子内放有 12 个大小相同的球,其中有 5 个红球,4 个白球,3 个黑球.第一次随机地摸出 2 个球,观察后不放回,第二次随机地摸出 3 个球,记 X_i 表示第 i 次摸到的红球的数目$(i=1,2)$;Y_j 表示第 j 次摸到的白球数,求

(1) (X_1,X_2)的分布;

(2) X_2 的分布;

(3) 在分别已知 $X_2=j(j=0,1,2,3)$时,X_1 的条件分布;

(4) 两次摸到的红球总数 Y 的分布;

(5) (X_1,Y_1)及(X_1,Y_2)的分布.

2. 一辆机场交通车载有 25 名乘客途经 9 个站,每位乘客都等可能在这 9 个站中任意一站下车(且不受其他乘客下车与否的影响),交通车只在有乘客下车时才停车,令随机变量 Y_i 表示在第 i 站下车的乘客数,$i=1,2,\cdots,X_i$ 在有乘客下车时取值为 1,否则取值为 0.求

(1) $(Y_i,Y_j)(i\neq j)$的分布及边缘分布,并判断 Y_i 与 Y_j 是否独立;

(2) $(X_i,X_j)(i\neq j)$的分布及边缘分布,并判断 X_i 与 X_j 是否独立;

(3) $\operatorname{cov}(X_i,X_j)$,$\rho_{X_i,X_j}$;

(4) 交通车停车次数 X 的方差.

3. 设 $X\sim b(25,p_1)$,$Y\sim b(25-X,p_2)$,求

(1) 已知 $X=k(k=0,1,2,\cdots,25)$时,Y 的条件分布;

(2) (X,Y)的分布;

(3) $E[Y|X]$,$E[Y^2|X]$;

(4) EY,DY.

4. 设(X,Y)的联合分布为:

$$P\{X=K,Y=S\}=\frac{N!}{K!\ S!\ (N-K-S)!}p_1^K p_2^S(1-p_1-p_2)^{N-K-S},$$

其中 N 是正整数，K,S 为非负整数且 $K+S\leqslant N$，p_1,p_2 是(0,1)上的实数，且 $p_1+p_2<1$.

(1) 证明：$X\sim b(N,p_1)$，$Y\sim b(N,p_2)$；

(2) 证明：在 $Y=S$ 的条件下，X 的条件分布为 $b\left(N-S,\dfrac{p_1}{1-p_2}\right)$，在 $X=K$ 的条件下 Y 的条件分布为 $b\left(N-K,\dfrac{p_2}{1-p_1}\right)$；

(3) 求第 3 题中 Y 的边缘分布及在 $Y=S$ 的条件下 X 的条件分布.

5. Y 服从参数 X 的指数分布，而 X 是服从[1,2]上的均匀分布的随机变量.

(1) 求 (X,Y) 的密度函数；

(2) 求 Y 的边缘密度函数；

(3) 求已知 $Y=y$ 时 X 的条件密度函数；

(4) $Y=1$ 时 X 的条件期望；

(5) 求 $P\{Y\leqslant X\}$.

6. 设 (X,Y) 服从 $D=\{(x,y)\mid 0\leqslant y\leqslant 1,y\leqslant x\leqslant 3-y\}$ 上的均匀分布.

(1) 求 X,Y 的边缘密度函数，并判断 X,Y 是否独立；

(2) 求 (X,Y) 的协差阵，判断 X 与 Y 是否相关.

(3) 求密度函数 $f_{Y|X}(y|x)$ 和 $f_{X|Y}(x|y)$；

(4) 求 $E[Y|X]$ 和 $E[X|Y]$.

7. 设 $X_1\sim P(\lambda_1)$，$X_2\sim P(\lambda_2)$ 证明在 $X_1+X_2=n(n=1,2,\cdots)$ 的条件下，X_j 的条件分布为 $b\left(n,\dfrac{\lambda_j}{\lambda_1+\lambda_2}\right)$.

8. 假设随机变量 Y 服从参数为 $\lambda=1$ 的指数分布，随机变量

$$X_k=\begin{cases}0, & \text{若 } Y\leqslant k,\\ 1, & \text{若 } Y>k.\end{cases}\qquad (k=1,2).$$

求(1) (X_1,X_2) 的联合分布；

(2) $\operatorname{cov}(X_1,X_2)$，$\rho_{X_1,X_2}$.

9. 假设二维随机变量 (X,Y) 在矩形 $G=\{(x,y)\mid 0\leqslant x\leqslant 2,0\leqslant y\leqslant 1\}$ 上服从均匀分布，记

$$U=\begin{cases}0, & \text{若 } X\leqslant Y,\\ 1, & \text{若 } X>Y;\end{cases}\qquad V=\begin{cases}0, & \text{若 } X\leqslant 2Y,\\ 1, & \text{若 } X>2Y.\end{cases}$$

(1) 求 (U,V) 的分布；

(2) 求 (U,V) 的相关系数.

10. 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$，利用条件期望 $E[X|Y]=\mu_1+\dfrac{\sigma_1}{\sigma_2}\rho(Y-\mu_2)$，证明 $\rho_{X,Y}=\rho$.

11. 设 $D[Y|X]\overset{\text{def}}{=\!=}E[(Y-E[Y|X])^2|X]$，通常称之为 Y 关于 X 的**条件方差**.证明：

(1) $D[Y|X]=E[Y^2|X]-(E[Y|X])^2$；

(2) $DY=D[E(Y|X)]+E[D(Y|X)]$.

12. 设 X,Y 的方差存在,相关系数为 $\rho_{X,Y}$,现用 X 对 Y 作线性预测.证明在均方误差最小意义下的最优线性预测为

$$L(X)=EY+\rho_{X,Y}\cdot\frac{\sigma_Y}{\sigma_X}(X-EX).$$

13. 设二维随机向量(X,Y)服从 $D=\{(x,y)|0\leqslant x\leqslant 1,0\leqslant y\leqslant 2\}$上的均匀分布.求(1) $P\{3X\geqslant Y\}$;(2) $Z=\min\{X,Y\}$的密度函数.

第 4 章

数理统计的基础知识

从本章开始，我们将讨论另一主题：数理统计. 数理统计是研究统计工作的一般原理和方法的科学，它主要阐述搜集、整理、分析统计数据，并据以对研究对象进行统计推断的理论和方法，是统计学的核心和基础. 本章将介绍数理统计的基本概念：总体、样本、统计量与抽样分布. 抽样分布泛指统计量服从的分布. 为研究抽样分布，必须补充一些在前三章中未提及的概率分布，因而本章还较为详尽地介绍三种常用的统计分布：χ^2 分布、F 分布与 t 分布.

§4.1 总体与样本

一、总体与总体分布

总体是具有一定的共同属性的研究对象全体. 总体的大小与范围由具体研究与考察的目的确定. 譬如，为了了解某校一年级学生“高等数学”的学习情况，该校学习“高等数学”的全体一年级学生便构成了待研究的总体. 又如要了解北京市男大学生的身高和体重的分布情况，北京市的全体男大学生便组成了总体. 一旦总体确定了，便称总体的每一个别成员为**个体**. 个体与总体的关系，即集合论中元素与集合之间的关系. 统计学中关心的不是每个个体的所有具体特性，而仅仅是它的某一项或某几项数量指标. 在上述第一个例子中，数量指标可取为该校一年级学生“高等数学”的期末考试成绩. 在第二个例子中，北京市男大学生的身高与体重是数量指标. 对于选定的数量指标 X（可以是向量）而言，每个个体所取的值是不同的. 在试验中，抽取了若干个个体就观察到了 X 的这样或那样的数值，因而这一数量指标 X 是一个随机变量（或向量），而 X 的分布就完全描述了总体中我们所关心的这一数量指标的分布情况. 由于我们关心的只是此数量指标，因此我们以后就把总体与数量指标 X 等同起来，并把数量指标 X 的分布称为总体的分布，由此导出下述定义：

定义 4.1 统计学中称随机变量（或向量）X 为**总体**，并把随机变量（或向

量)X 的分布称为**总体分布**.

对于上述定义,作三点说明.

首先,表示总体的 X 既可以是随机变量,也可以是随机向量.如果当事者关心的不是个体的一项数量指标,而是两项或两项以上的数量指标时,X 便是随机向量.但为简化讨论,本书只限于考察一项数量指标的情形.这样,今后总体指的皆是随机变量.

其次,有时个体的特性本身不是直接由数量指标来描述的.例如,服装厂生产的各式时装,玩具厂生产的儿童玩具等.这时产品的检验部门通常将产品分成若干等级.譬如说,划分成一、二、三等品和等外品四个等级.当出现这样的情形时,我们仍可用一个随机变量 X 来表示产品的质量.它可取 1、2、3 和 4 四个值,分别对应于产品为一、二、三等和等外品.一般说来,个体的定性指标皆可类似地通过上述方式转化成一个数量指标,从而也就可设定一个随机变量来表示所研究的总体.

第三,总体分布就是设定的表示总体的随机变量 X 的分布.总体的分布,一般说来是未知的.有时虽已知总体分布的类型(如正态分布、伯努利分布等),但不知这些分布中所含的参数(如 μ,σ^2,p 等),有时甚至连分布所属的类型也不能肯定.统计学的主要任务正是要对总体的未知分布进行推断.

二、样本与样本分布

前面指出,总体的分布一般来说是未知的.为了获取对于总体的分布的知识,一般的方法是对总体进行抽样观察.通过观察可得到总体 X 的一组数值 $(x_1,x_2,\cdots,x_n)$,其中每一 x_i 是从总体中抽取的某一个体的数量指标 X 的观察值.统计学的任务就是提供科学的方法,借助这组观察值对未知的总体分布进行合理的推断.有时,我们还需进行多次抽样观察.显然,再作一次抽样观察所得的一组值 $(x'_1,x'_2,\cdots,x'_n)$ 往往和前一次得到的观察值有别.这样,考虑问题时,就不能把每次抽样观察得到的值看成是一组确定的数值.一种合理的解释是把它看成随机向量 $(X_1,X_2,\cdots,X_n)$ 的一次实现值,既然抽样观察的目的是为了对总体的分布进行各种分析推断,因而要求抽取的样本能很好地反映总体的特性.由此导出对样本的下述定义:

定义 4.2　称 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的**简单随机样本**,若 $X_1,X_2,\cdots,X_n$ 是独立同分布的随机变量,且与总体 X 同分布.样本中所含分量的个数 n 称为该样本的**容量**.

要求样本中的每一分量 X_i 与总体 X 同分布,表明抽样观察时,每一个体都是从同一总体中抽取的.要求样本中诸分量是独立的,则表明每一观察结果既不

影响其他观察结果，也不受其他观察结果的影响．获得上述简单随机样本的方法称为**简单随机抽样**．

显然，简单随机样本是一种非常理想化的样本．在实际中要获得真正意义下的简单随机样本并非易事．严格地说，现实的总体总是有限的．在有限总体中，若是有放回地抽样，且能保证抽样的方式是充分随机的，那么可得到一个简单随机样本．但实际中通常是不放回的抽样，样本中诸分量既不是同分布的，也不是相互独立的．不过，如果所考察的总体规模很大，不放回抽样与有放回抽样的区别就很小，此时可近似地把所得的样本看成是一个简单随机样本．本书不准备对抽样方法作详细介绍，故以下假定所考虑的样本均为简单随机样本，并简称为**样本**．

对于样本，我们应持有双重的理解．在未观察具体的抽样结果时，应把样本$(X_1, X_2, \cdots, X_n)$视为一个随机向量．在理论上探讨统计方法时，通常都是如此．但在观察具体的抽样结果后，样本便理解为所得的一组具体的观察值$(x_1, x_2, \cdots, x_n)$．在实际统计应用中，通常使用样本观察值．今后约定，以大写的英文字母 X_i 表示随机变量，而以相应的小写英文字母 x_i 表示它的观察值，并称样本$(X_1, X_2, \cdots, X_n)$的一组具体的观察值$(x_1, x_2, \cdots, x_n)$为**样本值**，全体样本值组成的集合称为**样本空间**．显然，容量为 n 的样本的样本空间是 n 维实空间 $\mathbf{R}^n$ 中的一个子集．

设总体 X 的分布函数为 $F(x)$，则由定义 4.2 知，样本$(X_1, X_2, \cdots, X_n)$的分布函数为

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i), \tag{4.1}$$

称之为**样本分布**．

特别地，若总体 X 为连续型随机变量，其密度函数为 $f(x)$，则样本的密度函数为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i). \tag{4.2}$$

如总体 X 为离散型随机变量，概率分布为 $p(x) = P\{X = x\}$，x 取遍 X 所有可能取值，则样本的概率分布为

$$p(x_1, x_2, \cdots, x_n) = P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i).$$

例 4.1 称总体 X 为**正态总体**，如它服从正态分布．正态总体是统计应用中最常见的总体．现设总体 X 服从正态分布 $N(\mu, \sigma^2)$，则其样本密度由下式给出：

$$f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}$$

$$=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}. \quad (4.3)$$

例 4.2　称总体 X 为**伯努利总体**，如果它服从以 $p(0<p<1)$ 为参数的伯努利分布，即

$$P\{X=1\}=p, P\{X=0\}=1-p.$$

不难算出其样本 $(X_1,X_2,\cdots,X_n)$ 的概率分布为

$$P\{X_1=i_1,X_2=i_2,\cdots,X_n=i_n\}=p^{s_n}(1-p)^{n-s_n}, \quad (4.4)$$

其中 $i_k(1\leqslant k\leqslant n)$ 取 1 或 0，而 $s_n=i_1+i_2+\cdots+i_n$，它恰等于样本中取值为 1 的分量之总数.

服从伯努利分布的总体也具有较广泛的应用背景. 概率 p 通常可视为某一实际总体(如工厂的某一批产品)中具有某一特征(如废品)的个体所占的比例，亦称为**比率**. 从总体中随机抽取一个个体，可视为一个随机试验，试验结果可用一随机变量 X 来刻画：若恰好抽到具有该特征的个体，记 $X=1$；否则，记 $X=0$. 这样，X 便服从以 p 为参数的伯努利分布. 通常参数 p 是未知的，故需通过抽样对其作统计推断.

例 4.3　设总体 X 服从参数为 λ 的泊松分布，$(X_1,X_2,\cdots,X_n)$ 为其样本，则样本的概率分布为

$$P\{X_1=i_1,X_2=i_2,\cdots,X_n=i_n\}=\prod_{k=1}^{n}P\{X=i_k\}$$

$$=\prod_{k=1}^{n}\frac{\lambda^{i_k}}{i_k!}e^{-\lambda}=\frac{\lambda^{s_n}}{i_1!\ i_2!\ \cdots i_n!}e^{-n\lambda}, \quad (4.5)$$

其中 $i_k(1\leqslant k\leqslant n)$ 取非负整数，而 $s_n=i_1+i_2+\cdots+i_n$.

实际应用中，人们对总体的分布通常不是毫无所知的. 在很多情形，根据知识与经验能肯定分布所属的类型. 例如，通常认为钢筋强度服从正态分布，描述产品是否合格的随机变量自然是服从伯努利分布的；再如记录电话呼唤次数的随机变量通常认为是服从泊松分布的. 这样，对于总体的分布，仅有参数是未知的. 一旦能确定总体分布的参数，总体分布就完全已知了.

三、统计推断问题简述

前面的阐述已使我们对统计学要解决的问题有了一个概要的认识，即借助总体 X 的一个样本 $(X_1,X_2,\cdots,X_n)$，对总体 X 的未知分布进行推断. 我们把这

类问题统称为**统计推断问题**.为利用样本对未知的总体分布进行推断,我们需借助样本构造样本的适当的函数,正是利用这些函数所反映的总体分布的信息来对总体分布所属的类型,或总体分布中所含的未知参数作出统计推断.

§4.2 统计量

一、统计量的定义

如前所述,在进行统计推断时,构造样本的适当的函数是关键,为此,我们给出下列定义.

定义 4.3 设$(X_1, X_2, \cdots, X_n)$为总体 X 的一个样本,称此样本的任一不含总体分布未知参数的函数为该样本的**统计量**.

例 4.4 设总体 X 服从正态分布,$EX = 5$,$DX = \sigma^2$,σ^2 未知.$(X_1, X_2, \cdots, X_n)$为总体 X 的一个样本,令

$$S_n = X_1 + X_2 + \cdots + X_n, \overline{X} = \frac{S_n}{n},$$

则 S_n 与 $\overline{X}$ 均为样本$(X_1, X_2, \cdots, X_n)$的统计量.但若令

$$U = \frac{n(\overline{X} - 5)}{\sigma},$$

则 U 不是该样本的统计量,因 U 的表示式中含有总体分布中的未知参数σ.

对于一个给定的样本,根据统计量的定义,尽管可以构造出很多统计量来,但常用的统计量并不多.以下将介绍统计学中常用的一些统计量.

二、常用的统计量

设$(X_1, X_2, \cdots, X_n)$为总体 X 的一个样本.

1. 样本均值

称样本的算术平均值为**样本均值**,记为 $\overline{X}$,即

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

2. 样本方差

样本方差是用来描述样本中诸分量与样本均值的均方差异的.它有两种定义方式.较直观的定义是

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

并称 S_0^2 为样本的**未修正样本方差**.

统计学中更常用的是另一种定义,即

$$S^2=\frac{n}{n-1}S_0^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

并称 S^2 为样本的**修正样本方差**.

初看起来,S^2 没有 S_0^2 那么自然,但用于统计推断,S^2 比 S_0^2 具有更好的统计性质.比如,当用 S^2 或 S_0^2 来作为总体 X 的方差 σ^2 的估计时,S^2 满足 $ES^2=\sigma^2$,而 $ES_0^2\neq\sigma^2$,因而用 S^2 估计 σ^2 不会产生系统的偏差,第 5 章将对此作出讨论.由于今后使用的主要是修正样本方差,故以下简称修正样本方差为**样本方差**.

3. 样本标准差

正如总体的方差与标准差的关系一样,**样本标准差** S 定义为样本方差的算术平方根,即

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}.$$

4. 样本原点矩

记

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,k\geqslant 1,$$

并称 A_k 为样本的 k **阶原点矩**.显然,一阶原点矩即为样本均值.因此可把样本原点矩视为样本均值概念的推广.

5. 样本中心矩

记

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,k\geqslant 1,$$

并称 B_k 为样本的 k **阶中心矩**.显然,二阶中心矩即为未修正样本方差.因此可把样本中心矩视为未修正样本方差概念的推广.

上述五种统计量可统称为样本的**矩统计量**,简称为**样本矩**.它们皆可表为样本的显式函数.注意到函数关系强调的仅是变量之间的依赖关系,除样本矩外,尚可定义如下一些统计量,它们不能表为样本的显式函数.

6. 顺序统计量

设 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的一个样本,将样本中的诸分量按由小到大的次序排列成

$$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)},$$

则称 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 为样本的一组**顺序统计量**,称 $X_{(i)}$ 为样本的第 i 个顺

序统计量.特别地,称 $X_{(1)}=\min(X_1,X_2,\cdots,X_n)$ 与 $X_{(n)}=\max(X_1,X_2,\cdots,X_n)$ 分别为**样本极小值**与**样本极大值**,并称 $X_{(n)}-X_{(1)}$ 为样本的**极差**.

三、枢轴量

前面已提及,样本的统计量中是不应包含总体分布的未知参数的.不过,有时在统计推断问题中,会遇到下述情形.设 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的样本,现需对总体分布中某一未知参数 θ 进行推断(总体分布也许还含其他的未知参数).这时常需构造样本的一个仅含未知参数 θ,但不含其他未知参数的函数 $U(X_1,X_2,\cdots,X_n;\theta)$.而如此构造的样本函数,尽管含未知参数 θ,却服从一个已知的分布.利用这一已知分布的知识,即可对未知参数 θ 进行统计推断.我们将这种仅含一个未知参数,但其分布却已知的样本函数称为**枢轴量**.

例 4.5 设总体 $X\sim N(\mu,\sigma_0^2)$,其中 σ_0^2 已知,μ 未知,$(X_1,X_2,\cdots,X_n)$ 为总体 X 的一个样本,令

$$U=\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma_0}.$$

样本的上述函数 U 中尽管含未知参数 μ,但不论其真值为何值,总有 $U\sim N(0,1)$.故 U 是一枢轴量,可用来对未知参数 μ 作统计推断.

§4.3 常用的统计分布

前面已指出,当取得总体 X 的样本 $(X_1,X_2,\cdots,X_n)$ 后,通常是借助样本的统计量(或枢轴量)对未知的总体分布进行推断的.为了实现推断的目的必须进一步确定相应的统计量(或枢轴量)所服从的分布.这样就有必要补充一些在本书概率论部分中未曾提及.但在统计学中经常用到的分布.本节将依次介绍 χ^2 分布、F 分布与 t 分布.鉴于这些分布在统计学中的重要性,通常统称其为常用的统计分布.为突出本节的统计内容,我们将略去一些命题的证明.不过,为了兼顾理论的完整性,相关的证明将在本章的附录中给出,以供对此感兴趣的读者参考.

一、分位数

在统计推断中,经常用到统计分布的一类数字特征——分位数.在即将讨论一些常用的统计分布之前,我们首先给出分位数的一般概念与性质.熟悉这些概念与性质,对于稍后查阅常用统计分布表是非常有用的.

定义 4.4　设随机变量 X 的分布函数为 $F(x)$，对给定的实数 $\alpha(0<\alpha<1)$，如果实数 F_α 满足

$$P\{X>F_\alpha\}=\alpha, \tag{4.6}$$

即

$$1-F(F_\alpha)=\alpha \text{ 或 } F(F_\alpha)=1-\alpha, \tag{4.7}$$

则称 F_α 为随机变量 X 的分布的水平 α 的**上侧分位数**.或直接称为分布函数 $F(x)$的水平 α 的上侧分位数.

显然，如果 $F(x)$是严格单调增的，那么其水平 α 的上侧分位数 F_α 为

$$F_\alpha=F^{-1}(1-\alpha). \tag{4.8}$$

当 X 是连续型随机变量时，设其密度函数为 $f(x)$，则其水平 α 的上侧分位数 F_α 满足

$$\int_{F_\alpha}^{+\infty} f(x)\mathrm{d}x=\alpha.$$

在图形上(如图 4.1)，介于密度函数曲线下方，x 轴上方与垂直直线 $x=F_\alpha$ 右方之间的阴影区域的面积恰等于 α.

例如，标准正态分布 $N(0,1)$的水平 α 的上侧分位数 通常记作 u_α，则 u_α 满足

$$1-\Phi_0(u_\alpha)=\alpha,$$

即

$$\Phi_0(u_\alpha)=1-\alpha. \tag{4.9}$$

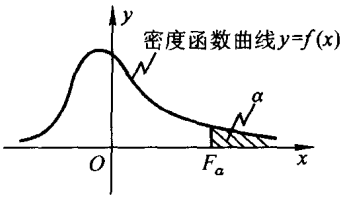


图 4.1　上侧分位数

图 4.2 给出了标准正态分布的水平 α 的上侧分位数的图示.

一般来讲，直接求解分位数是很困难的，对常见的统计分布，在本书附录中给出了分布函数值表或分位数表，通过查表，可以很方便地得到分位数的值.比如，对给定的 α，由(4.9)查标准正态分布的分布函数值表，可得到 u_α 的值.另外根据第 2 章对正态分布的讨论，一般正态分布可借助标准化化为标准正态分布来考虑.

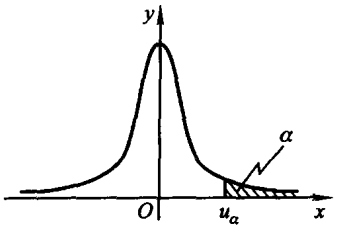


图 4.2　标准正态分布的上侧分位数

为简便以后的讨论，我们列出与分布函数的上侧分位数有关的两个等式：

$$P\{X\leqslant F_{1-\alpha}\}=\alpha. \tag{4.10}$$

$$P\{F_{1-\alpha/2}<X\leqslant F_{\alpha/2}\}=1-\alpha. \tag{4.11}$$

上述两式的证明留待读者自己完成.

对于像标准正态分布那样的**对称分布**(即其密度函数为偶函数，关于 y 轴

对称!),统计学中还用到其另一种分位数——双侧分位数.

定义 4.5 设 X 是对称分布的连续型随机变量,其分布函数为 $F(x)$,对给定的实数 $\alpha(0<\alpha<1)$,如果正实数 T_α 满足

$$P\{|X|>T_\alpha\}=\alpha, \tag{4.12}$$

即

$$F(T_\alpha)-F(-T_\alpha)=1-\alpha. \tag{4.13}$$

则称 T_α 为随机变量 X 的分布的水平 α 的**双侧分位数**,也简称为**分位数**,或直接称为分布函数 $F(x)$ 的水平 α 的分位数.

由于对称性(见习题四(B)的第 6 题),(4.13)可改写为

$$F(T_\alpha)=1-\frac{\alpha}{2}$$

或

$$P\{X>T_\alpha\}=1-F(T_\alpha)=\frac{\alpha}{2}. \tag{4.14}$$

对比(4.6)与(4.14)两式知,水平 α 的分位数实际等于水平 $\frac{\alpha}{2}$ 的上侧分位数.即有

$$T_\alpha=F_{\alpha/2}. \tag{4.15}$$

图 4.3 以标准正态分布为例给出了双侧分位数的图示.

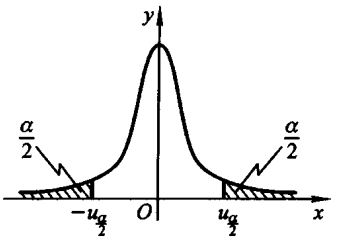


图 4.3 标准正态分布的水平 α 的双侧分位数

此外,对于具有对称密度函数的分布函数的上侧分位数,恒有

$$F_\alpha=-F_{1-\alpha}. \tag{4.16}$$

请读者自己验证上式.

例 4.6 设 $\alpha=0.05$,求标准正态分布的水平 0.05 的上侧分位数和双侧分位数.

解 由于

$$\Phi_0(u_{0\,05})=1-0.05=0.95,$$

查标准正态分布函数值表可得

$$u_{0\,05}=1.645.$$

而水平 0.05 的双侧分位数为 $u_{0\,025}$,它满足

$$\Phi_0(u_{0\,025})=1-0.025=0.975.$$

查表得

$$u_{0.025}=1.96.$$

二、χ^2 分布

在第二章中，我们曾证明(见例 2.29)：若 $X\sim N(0,1)$，则 X^2 的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}x},x>0. \tag{4.17}$$

这一结果实际上是下述命题的特例.

命题 4.1　设 $X_1,X_2,\cdots,X_n$ 是 n 个相互独立的随机变量，且 $X_i\sim N(0,1)$，$i=1,2,\cdots,n$，则

$$X=X_1^2+X_2^2+\cdots+X_n^2$$

的密度函数为

$$\chi^2(x;n)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{1}{2}x},x>0. \tag{4.18}$$

其中 $\Gamma(a)=\int_0^{+\infty}x^{a-1}\mathrm{e}^{-x}\mathrm{d}x(a>0)$ 是 Γ(伽马)函数.

证明见附录 4.1.

现给出 χ^2 分布的定义.

定义 4.6　一个随机变量 X 称为服从以 n 为自由度的 **χ^2 分布**，如果其密度函数由(4.18)给出，记作 $X\sim\chi^2(n)$.

当自由度 $n\geqslant 3$ 时，χ^2 分布密度函数的曲线皆为单峰曲线. 密度曲线从原点开始递增，在 $x=n-2$ 处达最大值，然后递减，渐近于 x 轴. 显然，函数图形关于垂直线 $x=n-2$ 不对称. 随着自由度 n 的增大，密度曲线的峰值向右移动，图形变得比较平缓并趋于对称，而且可用正态分布来近似(为什么？请读者思考). 当自由度 $n=2$ 时，此时分布实际上就是参数为 $\frac{1}{2}$ 的指数分布，密度曲线在 $x=0$ 处达最大值，然后递减；当自由度 $n=1$ 时，密度曲线在 $x=0$ 处取无穷大值，这时 y 轴为其垂直渐近线. 在后两种情形，密度曲线也以 x 轴为水平渐近线. 图 4.4 给出了 $n=1,4,10,20$ 时 χ^2 分布的密度函数曲线.

下述命题给出了关于 χ^2 分布的若干重要性质.

命题 4.2

(1) 若 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立，则 $X+Y\sim\chi^2(m+n)$.

(2) 若 $X\sim\chi^2(n)$，则 $EX=n$，$DX=2n$.

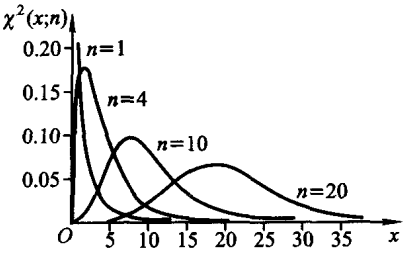


图 4.4　χ^2 分布的密度函数曲线

证明　设 $X_1,X_2,\cdots,X_{m+n}$ 独立同分

布,且均服从标准正态分布.

(1) 由于 $X\sim\chi^2(m)$,据定义 4.6 与命题 4.1,X 与 $X_1^2+X_2^2+\cdots+X_m^2$ 同分布,Y 与 $X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2$ 同分布,再由 X 与 Y 独立知,$X+Y$ 与 $X_1^2+X^2{}_2+\cdots+X_{m+n}^2$ 同分布,从而得 $X+Y\sim\chi^2(m+n)$.

(2) 设 $X_1,X_2,\cdots,X_n$ 相互独立且均服从标准正态分布,则由 $X\sim\chi^2(n)$ 知 X 与 $X_1^2+X_2^2+\cdots+X_n^2$ 同分布,于是

$$EX=E\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}EX_i^2=\sum_{i=1}^{n}DX_i=n.$$

此外,由于 $E[X_i^4]=3$(见习题四(B)的第 4 题),便知

$$D[X_i^2]=E[X_i^4]-(E[X_i^2])^2=3-1=2.$$

再因 $X_1,X_2,\cdots,X_n$ 相互独立,即得

$$D[X]=D\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}D[X_i^2]=2n.$$

上述命题中第一个结论实际上说明 χ^2 分布同正态分布一样具有**可加性**.

由于 χ^2 分布是常用的统计分布,但又难于利用其密度函数 $\chi^2(x;n)$ 进行直接计算,通常也为其制定了统计用表.附表 3 中对自由度 $n\leqslant45$ 的 χ^2 分布给出了水平 α 的上侧分位数 $\chi_\alpha^2(n)$ 之值.由前述关于一般分布的上侧分位数的定义与性质,当 $X\sim\chi^2(n)$ 时,由(4.6)与(4.10)两式立得

$$P\{X>\chi_\alpha^2(n)\}=P\{X<\chi_{1-\alpha}^2(n)\}=\alpha.$$

因 $\chi^2(x;n)$ 不是对称函数,故对 χ^2 分布而言,不存在双侧分位数,但在以后统计推断中,将用到等式

$$P\left(\{X<\chi_{1-\frac{\alpha}{2}}^2(n)\}\cup\{X>\chi_{\frac{\alpha}{2}}^2(n)\}\right)=\alpha$$

或

$$P\{\chi_{1-\frac{\alpha}{2}}^2(n)<X<\chi_{\frac{\alpha}{2}}^2(n)\}=1-\alpha.$$

上式是(4.11)式的直接推论.

例如,设 $X\sim\chi^2(10)$,取水平 $\alpha=0.05$,查表可知

$$P\{X<3.940\}=P\{X>18.307\}=0.05,$$

$$P\{3.247\leqslant X\leqslant20.483\}=0.95.$$

当自由度 n 充分大(如 $n>45$)时,χ^2 分布可近似地看作正态分布,于是由正态分布的分位数可近似地求得 χ^2 分布的分位数.

三、F 分布

接下来,我们考虑另一个常用的统计分布.

设 $X\sim\chi^2(m)$,$Y\sim\chi^2(n)$,且 X 与 Y 相互独立,记

$$Z=\frac{X/m}{Y/n}=\frac{n}{m}\frac{X}{Y}. \tag{4.19}$$

下述命题给出了 Z 的密度函数的表示式.

命题 4.3　设 Z 由(4.19)所定义,则 Z 的密度函数为

$$f(x;m,n)=\frac{1}{\mathrm{B}\left(\frac{m}{2},\frac{n}{2}\right)}\left(\frac{m}{n}\right)\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{1}{2}(m+n)},x>0. \tag{4.20}$$

其中 $\mathrm{B}(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}\mathrm{d}x(p>0,q>0)$ 是 B(贝塔)函数.

证明见附录 4.1.

定义 4.7　如果一个随机变量 X 的密度函数由(4.20)给出,则称其服从第一自由度为 m,第二自由度为 n 的 F **分布**,记作 $X\sim F(m,n)$.

由命题 4.3 不难推知,若 $X\sim F(m,n)$,则 $X^{-1}\sim F(n,m)$.

F 分布的密度函数曲线也为单峰曲线,当第一自由度 $m\geqslant 3$ 时,曲线在

$$x^*=\frac{m-2}{m}\cdot\frac{n}{n+2}$$

处达最大值.显见 $x^*<1$,即图形的峰值恒在小于 1 处取到.此外,不难看出,当两个自由度 m 与 n 都变得越来越大时,x^* 接近 1,从而函数图形就在非常接近 1 的地方达到最高点.图 4.5 给出了若干 F 分布的密度函数曲线.

由于 F 分布是常用的统计分布,且也难于利用其密度函数直接进行计算,因此对它也制出了统计用表.附表 4 中对一些充分小的 α 值列出了 F 分布的水平 α 的上侧分位数 $F_\alpha(m,n)$ 之值.根据上侧分位数的定义与性质,当 $X\sim F(m,n)$时,也可推得

$$P\{X>F_\alpha(m,n)\}=P\{X<F_{1-\alpha}(m,n)\}=\alpha.$$

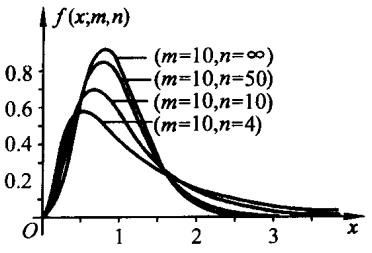


图 4.5　F 分布的密度函数曲线

由于 X 为非负随机变量,其密度函数 $f(x;m,n)$自然不是对称函数.这样,对 F 分布而言,也不存在双侧分位数.但在统计推断中将使用关系式

$$P\left(\left\{X<F_{1-\frac{\alpha}{2}}(m,n)\right\}\cup\left\{X>F_{\frac{\alpha}{2}}(m,n)\right\}\right)=\alpha$$

或

$$P\left\{F_{1-\frac{\alpha}{2}}(m,n)<X<F_{\frac{\alpha}{2}}(m,n)\right\}=1-\alpha.$$

上式也是(4.11)式的直接推论.

作为例子,设 $X\sim F(5,10)$,查表 4 知

$$P\{X>3.33\}=0.05,P\{X>4.24\}=0.025.$$

又设 $Y\sim F(10,5)$,查表可得

$$P\{Y>4.74\}=0.05, P\{Y>6.62\}=0.025.$$

由于附表 4 仅对一些充分小的 α 值给出了 F 分布的上侧分位数,当 α 接近于 1 时,可利用下式求出所需的上侧分位数

$$F_{\alpha}(m,n)=\frac{1}{F_{1-\alpha}(n,m)}, \tag{4.21}$$

这是因为由 $X\sim F(m,n)$,可推知 $X^{-1}\sim F(n,m)$.

例如,由(4.21)式知,

$$F_{0.95}(m,n)=\frac{1}{F_{0.05}(n,m)}, \quad F_{0.975}(m,n)=\frac{1}{F_{0.025}(n,m)}.$$

这样,当 $X\sim F(5,10)$时,查表可知

$$P\left\{X<\frac{1}{4.74}\right\}=0.05,$$

$$P\left\{\frac{1}{6.62}\leqslant X\leqslant 4.24\right\}=0.95.$$

四、t 分布

最后,我们介绍 t 分布,它与 F 分布有紧密联系.

设 $X\sim N(0,1)$, $Y\sim\chi^2(n)$,且 X 与 Y 相互独立.记

$$T=\frac{X}{\sqrt{Y/n}}, \tag{4.22}$$

由(4.22)式可推知

$$T^2=\frac{X^2}{Y/n}\sim F(1,n).$$

利用这一关系可以简便地导出由(4.22)定义的随机变量 T 的密度函数的表示式.

命题 4.4 由(4.22)所定义的随机变量 T 的密度函数为

$$t(x;n)=\frac{1}{\mathrm{B}\left(\frac{1}{2},\frac{n}{2}\right)}\frac{1}{\sqrt{n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, -\infty<x<\infty. \tag{4.23}$$

证明见附录 4.1.

定义 4.8 如果一个随机变量 X 的密度函数由(4.23)给出,则称其为服从自由度为 n 的 t 分布,记作 $X\sim t(n)$,

t 分布的密度曲线也为单峰曲线,但关于 y 轴对称,且在 $x=0$ 处取到最大值. x 轴为其水平渐近线.图 4.6 给出了自由度 $n=1,5,10,\infty$时 t 分布的密度函数曲线.

当自由度 n 很大时，t 分布接近于标准正态分布，这是因为

$$\lim_{n\to\infty}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}=\mathrm{e}^{-\frac{1}{2}x^2}.$$

图 4.6 中，$n=\infty$ 时的 t 分布的密度函数曲线，即为标准正态分布的密度函数曲线. 从图 4.6 中看出，与标准正态密度曲线相比，t 分布的曲线以较慢速率趋于 x 轴. 换言之，t 分布的尾部比标准正态分布的尾部有更大的概率.

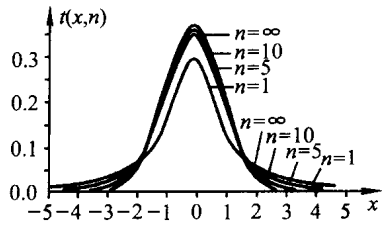


图 4.6　t 分布的密度函数曲线

附表 5 对于一些充分小的 α 值给出了 t 分布的水平 α 的上侧分位数 $t_\alpha(n)$ 之值. 由于 t 分布具有对称的密度函数，当 α 接近 1 时，可按下式求出相应的上侧分位数：

$$t_\alpha(n)=-t_{1-\alpha}(n). \tag{4.24}$$

上式是(4.16)式的直接推论.

根据上侧分位数的定义与性质，如 $X\sim t(n)$，由(4.6)，(4.10)与上式立得：

$$P\{X>t_\alpha(n)\}=P\{X<-t_\alpha(n)\}=\alpha.$$

再由于 t 分布具有对称的密度函数，从而具有双侧分位数，故由(4.12)与(4.15)两式知

$$P\{|X|>t_{\alpha/2}(n)\}=\alpha.$$

例如，设 $X\sim t(8)$，取水平 $\alpha=0.05$，查表可知 $t_\alpha(8)=1.860$，$t_{\frac{\alpha}{2}}(8)=2.306$，故有

$$P\{X>1.860\}=P\{X<-1.860\}=P\{|X|>2.306\}=0.05.$$

此外，由于自由度 n 充分大时，t 分布近似于标准正态分布，故有 $t_\alpha(n)\approx u_\alpha$，其中 u_α 为标准正态分布的上侧分位数.

§4.4　抽样分布

总体的分布往往是未知的，或是部分地未知的. 根据实际问题的需要，有时需对总体未知的重要数字特征(如总体数学期望、总体方差)或总体分布中所含的未知参数进行统计推断. 这类问题称为**参数统计推断**. 在参数统计推断问题中，经常需要利用总体的样本构造出合适的统计量(或枢轴量)，并使其服从或渐近地服从已知的确定分布. 统计学中泛称统计量(或枢轴量)的分布为**抽样分布**. 讨论抽样分布的途径有两个：一是精确地求出抽样分布，并称相应的统计推断为

小样本统计推断;另一种方式是让样本容量趋于无穷,并求出抽样分布的极限分布.然后,在样本容量充分大时,可利用该极限分布作为抽样分布的近似分布,再对未知参数进行统计推断,因此称相应的统计推断为**大样本统计推断**.本节重点讨论正态总体的抽样分布,属小样本统计范畴;此外,也简要地讨论一般总体的某些抽样分布的极限分布,属大样本统计范畴.

一、正态总体的抽样分布

三种常用的统计分布(χ^2分布、F 分布与 t 分布)为讨论正态总体的抽样分布作了必要的准备.不过,在一般地讨论正态分布的抽样分布以前,我们还需首先介绍一个涉及正态总体样本均值与样本方差的抽样分布的定理.它是讨论正态总体抽样分布的一个基础性定理.

定理 4.1 设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,X_2,\cdots,X_n)$是其容量为 n 的一个样本,$\overline{X}$与S^2分别为此样本的样本均值与样本方差,则有

(1) $\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$;

(2) $\dfrac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$;

(3) $\overline{X}$与S^2相互独立.

证明 结论(1)可由独立正态非零线性组合的一个重要性质(见(3.44)式)导出.

结论(2)与(3)的严格证明需要用到关于多重积分的变量替换公式,此外还要利用正交矩阵的一些性质,数学推导的技巧性很强,故此处略去,有兴趣的读者可参考附录 4.2.

有了上述关于正态总体的样本均值与样本方差的抽样分布的基础性定理,再结合上一节中关于常用统计分布的有关论述,便可容易地构造出单正态总体与双正态总体中样本的一些统计量(或枢轴量),并使之服从确定的已知分布,以下分别细述之.

1. 单正态总体的抽样分布

定理 4.2 设$(X_1,X_2,\cdots,X_n)$为正态总体$X\sim N(\mu,\sigma^2)$的样本,$\overline{X}$与S^2分别为该样本的样本均值与样本方差,则有

(1) $U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$;

(2) $\dfrac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$;

(3) $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

证明　结论(1)是定理 4.1(1)的直接推论. 结论(2)已由定理 4.1 给出,再因

$$T=\frac{U}{\sqrt{\frac{n-1}{\sigma^2}S^2/(n-1)}},$$

且由定理 4.1 知 U 与 $\frac{n-1}{\sigma^2}S^2$ 相互独立,故由本定理的结论(1)、(2)与命题 4.4,知 $T\sim t(n-1)$.

作为概率论的推导结果,定理 4.2 的结论无需涉及正态分布中的两个参数 μ 与 σ^2 是否已知. 换言之,只要 μ 与 σ^2 分别表示该正态总体的期望与方差,不论它们是否已知,相应的结论都是成立的. 不过,在统计学的应用中,则需关心这两个参数是否已知. 只有当 μ 与 σ^2 都已知时,定理 4.2 中提及的三个样本函数方可视为样本的统计量. 退一步,若 μ 未知,T 可视为一枢轴量;若 σ^2 未知,$\frac{n-1}{\sigma^2}S^2$ 也可视为一枢轴量,因此它们也可用于对未知参数 μ 与 σ^2 作统计推断. 但在 μ 与 σ^2 都未知时,U 甚至不能作为枢轴量. 也正是基于这一原因,在关于单正态总体参数的统计推断中,当 σ^2 未知时,需设计出枢轴量 T 来替换样本函数 U.

2. 双正态总体的抽样分布

在统计学的应用中,有时要比较两个正态总体的参数. 下述定理为比较两个正态总体的参数提供了合适的统计量(或枢轴量).

定理 4.3　设 $X\sim N(\mu_1,\sigma_1^2)$ 与 $Y\sim N(\mu_2,\sigma_2^2)$ 是两个相互独立的正态总体. 又设 $(X_1,X_2,\cdots,X_{n_1})$ 是总体 X 的容量为 n_1 的样本,$\overline{X}$ 与 S_1^2 分别为该样本的样本均值与样本方差. 再设 $(Y_1,Y_2,\cdots,Y_{n_2})$ 是总体 Y 的容量为 n_2 的样本,$\overline{Y}$ 与 S_2^2 分别为此样本的样本均值与样本方差. 记 S^2 是 S_1^2 与 S_2^2 的加权平均:

$$S^2=\frac{n_1-1}{n_1+n_2-2}S_1^2+\frac{n_2-1}{n_1+n_2-2}S_2^2,$$

则有

(1) $U=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$;

(2) $F=\left(\dfrac{\sigma_2}{\sigma_1}\right)^2\dfrac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$;

(3) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时,

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2).$$

证明 (1) 由定理 4.1 的结论(1)知

$$\overline{X}\sim N\left(\mu_1,\frac{\sigma_1^2}{n_1}\right),\overline{Y}\sim N\left(\mu_2,\frac{\sigma_2^2}{n_2}\right).$$

再因两个总体 X 与 Y 相互独立,从而它们的样本均值$\overline{X}$与$\overline{Y}$也相互独立,故

$$\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right). \tag{4.25}$$

将上述分布标准化即证得(1).

(2) 由定理 4.1 的结论(2)知

$$\frac{n_1-1}{\sigma_1^2}S_1^2\sim\chi^2(n_1-1),\frac{n_2-1}{\sigma_2^2}S_2^2\sim\chi^2(n_2-1). \tag{4.26}$$

再因两个总体 X 与 Y 相互独立,从而它们的样本方差也相互独立,故由命题 4.3 知

$$F=\frac{\dfrac{n_1-1}{\sigma_1^2}S_1^2/(n_1-1)}{\dfrac{n_2-1}{\sigma_2^2}S_2^2/(n_2-1)}\sim F(n_1-1,n_2-1).$$

(3) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$时,由(4.25)即知

$$U_1=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1).$$

另由(4.26),再结合命题 4.2 知

$$V=\frac{1}{\sigma^2}\{(n_1-1)S_1^2+(n_2-1)S_2^2\}\sim\chi^2(n_1+n_2-2).$$

此外,可以证明 U_1 与 V 相互独立(见习题四(A)第 18 题).

综上,由命题 4.4 即知

$$\frac{U_1}{\sqrt{\dfrac{V}{n_1+n_2-2}}}\sim t(n_1+n_2-2).$$

再因

$$\frac{V}{n_1+n_2-2}=\frac{1}{\sigma^2}S^2,$$

即知

$$T=\frac{(\overline{X}-\overline{Y})-\mu}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}=\frac{U_1}{\sqrt{\dfrac{V}{n_1+n_2-2}}}\sim t(n_1+n_2-2).$$

如单正态总体情形中所述一样,作为概率论推导的结果,只要 μ_1 与 μ_2 分别

表示总体 X 与 Y 的数学期望，σ_1^2 与 σ_2^2 分别表示总体 X 与 Y 的方差，则不论它们是否已知，定理 4.3 的结论总是成立的. 又在统计应用中，为比较两个总体的参数，实际上仅对 $\mu=\mu_1-\mu_2$ 与 $r=\sigma_1^2/\sigma_2^2$ 发生兴趣，其中 μ 表示两个总体数学期望之间的差异，r 则表示两个总体方差的比值. 这样，当两正态总体的方差相等时，μ 未知，T 可作为枢轴量；而当 r 未知时，F 可作枢轴量. 从而，T 和 F 可分别用于推断未知的期望差异值 μ 或方差的比值 r. 此外，当两个正态总体的方差相等但未知时，如期望差异值 μ 也未知，样本函数 U_1 甚至连枢轴量也不是，这也就是要设计枢轴量 T 来替换样本函数 U_1 的原由.

二、一般总体抽样分布的极限分布

本分节将取消定理 4.2 中总体服从正态分布的条件，这时将不能精确地导出样本函数 U 与 T 的分布函数的显式. 于是，考虑当样本容量 n 趋于无穷时，相应的分布函数是否存在极限分布？为此，首先需引入概率论中有别于依概率收敛的另一种收敛性概念，即依分布收敛的概念. 现设 $F_n(x)$ 为随机变量 X_n 的分布函数，$F(x)$ 为随机变量 X 的分布函数，并记 $C(F)$ 为由 $F(x)$ 的全体连续点组成的集合，若

$$\lim_{n\to\infty}F_n(x)=F(x),\ \forall\, x\in C(F),$$

则称随机变量 X_n **依分布收敛至** X，或称分布函数 $F_n(x)$ 依分布收敛至 $F(x)$，简记为

$$X_n\xrightarrow{d}X \quad 或 \quad F_n(x)\xrightarrow{d}F(x).$$

有了上述收敛性概念，我们便可回答本分节一开始提出的问题.

定理 4.4　设 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的样本，并设总体 X 的数学期望与方差均存在，分别记为 $EX=\mu$，$DX=\sigma^2$. 再记

$$U_n=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}},\ T_n=\frac{\overline{X}-\mu}{S/\sqrt{n}},$$

其中 $\overline{X}$ 与 S 分别表示上述样本的样本均值与样本方差，则有

(1) $F_{U_n}(x)\xrightarrow{d}\Phi_0(x)$，

(2) $F_{T_n}(x)\xrightarrow{d}\Phi_0(x)$，

以上 $F_{U_n}(x)$，$F_{T_n}(x)$ 与 $\Phi_0(x)$ 分别表示 U_n，T_n 及标准正态分布的分布函数.

证明　(1) 由于

$$U_n=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\frac{(X_1+X_2+\cdots+X_n)-n\mu}{\sigma\sqrt{n}}$$

在总体 X 的方差存在的前提下，由 §3.5 节中所述的(林德伯格－勒维)中心极限定理，立知

$$\lim_{n\to\infty} F_{U_n}(x) = \lim_{n\to\infty} P(U_n \leqslant x) = \Phi_0(x), \forall\, x \in \mathbf{R}^1.$$

(2) 证明略，有兴趣的读者可参看附录 4.3 中的相关内容.

定理 4.4 的适用范围很宽泛，惟一的条件是总体的方差存在. 这样，当样本容量 n 充分大时，U_n 与 T_n 都近似地服从标准正态分布. 因此，如总体的方差 σ^2 已知，便可利用枢轴量 U_n 近似地对总体未知的数学期望 μ 进行统计推断. 如 σ^2 未知，则可选用枢轴量 T_n 近似地对 μ 作统计推断.

最后，我们指出，前面介绍的关于正态总体抽样分布的精确分布与一般总体的若干抽样分布的极限分布，都用作关于单参数的统计推断. 在第 5 章中，我们还要讨论关于有限离散型总体的多参数假设检验.

附录 4.1　三个常用统计分布密度函数的推导

以下依次给出 §4.3 中命题 4.1，4.3 与 4.4 的证明.

1. 命题 4.1 的证明

证明　由(4.17)知，当 $n=1$ 时，(4.18)成立. 使用数学归纳法. 设 $n=k$ 时，(4.18)成立，令

$$\zeta = X_1^2 + X_2^2 + \cdots + X_k^2,\ \eta = X_{k+1}^2,\ \gamma = X + Y.$$

由归纳假设及(4.17)知 ζ, η 的密度函数分别为

$$f_\zeta(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}x},\ x>0,$$

$$f_\eta(y) = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)} y^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}y},\ y>0.$$

由于 ζ 与 η 皆为非负的随机变量且相互独立，由第 3 章的卷积公式(见例 3.14)可推知，当 $z>0$ 时，γ 的密度函数可按下式计算：

$$\begin{aligned}
f_\gamma(z) &= \int_0^z f_\zeta(z-y) f_\eta(y)\mathrm{d}y \\
&= \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^z (z-y)^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}(z-y)} y^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}y}\mathrm{d}y \\
&= \frac{\mathrm{e}^{-\frac{1}{2}z}}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^z (z-y)^{\frac{k}{2}-1} y^{-\frac{1}{2}}\mathrm{d}y \\
&= \frac{\mathrm{e}^{-\frac{1}{2}z} z^{\frac{k+1}{2}-1}}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^z \left(1-\frac{y}{z}\right)^{\frac{k}{2}-1} \left(\frac{y}{z}\right)^{-\frac{1}{2}} \mathrm{d}\left(\frac{y}{z}\right)
\end{aligned}$$

$$= \frac{e^{-\frac{1}{2}z}z^{\frac{k+1}{2}-1}}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)}\int_0^1 (1-t)^{\frac{k}{2}-1}t^{\frac{1}{2}-1}dt \quad \left(令\ t=\frac{y}{z}\right)$$

$$= \frac{B\left(\frac{k}{2},\frac{1}{2}\right)}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)}z^{\frac{k+1}{2}-1}e^{-\frac{1}{2}z}$$

$$= \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k+1}{2}\right)}z^{\frac{k+1}{2}-1}e^{-\frac{1}{2}z},$$

其中倒数第二个等式中使用了贝塔函数的定义：

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

在最后一个等式中，使用了贝塔函数 B 与伽马函数 Γ 间的关系：

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

这表明，当 $n=k+1$ 时，(4.18)式也成立.

2. 命题 4.3 的证明

证明　首先，因 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，由定义 4.6 知，X 与 Y 的密度函数分别为

$$f_X(x)=\frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}x^{\frac{m}{2}-1}e^{-\frac{1}{2}x},x>0,$$

$$f_Y(y)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y},y>0.$$

设 $Z_0=\frac{X}{Y}$，从而 $Z=\left(\frac{n}{m}\right)Z_0$.

由于 X 与 Y 皆为非负的随机变量且相互独立，由第 3 章的例 3.16 知，当 $z>0$ 时，随机变量 Z_0 的密度函数可按下式计算：

$$f_{Z_0}(z)=\int_0^\infty f_X(yz)f_Y(y)ydy$$

$$= \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\int_0^\infty (yz)^{\frac{m}{2}-1}e^{-\frac{1}{2}yz}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}ydy$$

$$= \frac{z^{\frac{m}{2}-1}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\int_0^\infty y^{\frac{m+n}{2}-1}e^{-\frac{y}{2}(z+1)}dy$$

$$= \frac{z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\int_0^\infty t^{\frac{m+n}{2}-1}e^{-t}dt \quad \left(令\ t=\frac{y(z+1)}{2}\right)$$

$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}$$

$$= \frac{1}{B\left(\frac{m}{2},\frac{n}{2}\right)} z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}.$$

再由于 $Z=\left(\frac{n}{m}\right)Z_0$，当 $z>0$ 时，即知随机变量 Z 的密度函数可表示为

$$f_Z(z)=\frac{m}{n}f_{Z_0}\left(\frac{m}{n}z\right)=\frac{1}{B\left(\frac{m}{2},\frac{n}{2}\right)}\left(\frac{m}{n}\right)\left(\frac{m}{n}z\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}z\right)^{-\frac{m+n}{2}}.$$

3. 命题 4.4 的证明

证明　首先注意到，T 的表示式(4.22)中的分母为非负连续型随机变量且与 X 独立；而 X 服从标准正态分布，从而具有对称的密度函数. 因此，由两独立随机变量商的密度函数公式可推出随机变量 T 的密度函数也是对称函数(见习题四(B)的第 5 题).

其次，以 $f_T(t)$ 与 $f_{|T|}(t)$ 分别表示 T 与 $|T|$ 的密度函数，由于 T 具有对称的密度函数，不难证明，当 $t>0$ 时，$f_T(t)=\frac{1}{2}f_{|T|}(t)$(见习题四(B)的第 6 题).

现设 $F=\frac{X^2}{Y/n}$，则 $F\sim F(1,n)$，且由命题 4.3 知，随机变量 F 的密度函数为

$$f_F(x)=\frac{1}{B\left(\frac{1}{2},\frac{n}{2}\right)}\frac{1}{n}\left(\frac{x}{n}\right)^{-\frac{1}{2}}\left(1+\frac{x}{n}\right)^{-\frac{n+1}{2}},\quad x>0.$$

再注意到

$$|T|=\frac{|X|}{\sqrt{Y/n}}=\sqrt{F},$$

由第 2 章习题二(A)的第 37 题便知，当 $t>0$ 时，应有

$$f_{|T|}(t)=2tf_F(t^2)=\frac{2t}{B\left(\frac{1}{2},\frac{n}{2}\right)}\frac{1}{n}\left(\frac{t^2}{n}\right)^{-\frac{1}{2}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

$$=\frac{2}{B\left(\frac{1}{2},\frac{n}{2}\right)}\frac{1}{\sqrt{n}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

于是，当 $t>0$ 时，

$$f_T(t)=\frac{1}{2}f_{|T|}(t)=\frac{1}{B\left(\frac{1}{2},\frac{n}{2}\right)}\frac{1}{\sqrt{n}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

这表明，当 $x>0$ 时，(4.23)式是成立的. 再由于 $f_T(t)$ 是对称函数，即知当 $x<0$ 时，(4.23)式也成立.

附录 4.2　正态总体抽样分布的基础性定理

在证明定理 4.1 之前，先给出两个预备性命题.

命题 4.5　设 $X_1, X_2, \cdots, X_n$ 为相互独立同分布的随机变量，而且 $X_i\sim N(0,1)$，$i=1,2,\cdots,n$. 再设 $\boldsymbol{A}$ 为 n 阶正交矩阵，即 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{E}$，其中 $\boldsymbol{A}^{\mathrm{T}}$ 表示矩阵 $\boldsymbol{A}$ 之转置，$\boldsymbol{E}$ 为单位矩阵，记

$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$，$\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}=(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$，则 $Y_1,Y_2,\cdots,Y_n$ 仍为相互独立同分布的随机变量，且 $Y_i\sim N(0,1)$，$i=1,2,\cdots,n$.

证明　由例 4.1 知，n 维随机向量 $\boldsymbol{X}$ 的联合密度函数为

$$f(\boldsymbol{x})=\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right),$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，则 n 维随机向量 $\boldsymbol{Y}$ 的分布函数为

$$F(y_1,y_2,\cdots,y_n)=\int_{\boldsymbol{A}\boldsymbol{x}\leqslant\boldsymbol{y}}\cdots\int\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n.$$

在上述 n 重积分中作变量替换 $\boldsymbol{t}=\boldsymbol{A}\boldsymbol{x}$. 由于 $\boldsymbol{A}$ 为正交矩阵，便知 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{t}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{t}$，而变换的雅可比行列式为

$$J=\left|\frac{\partial x_i}{\partial t_j}\right|=|\boldsymbol{A}^{\mathrm{T}}|.$$

再由于 $\boldsymbol{A}$ 为正交矩阵，即知行列式 $|\boldsymbol{A}^{\mathrm{T}}|$ 的值为 1 或 -1，从而雅可比行列式的绝对值 $|J|=1$. 于是

$$\begin{aligned}F(y_1,y_2,\cdots,y_n)&=\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_n}\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left\{-\frac{1}{2}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{t})^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{t})\right\}\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n\\&=\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_n}\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left\{-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{t}\right\}\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n=\prod_{i=1}^{n}\Phi_0(y_i).\end{aligned}$$

这表明 $Y_1,Y_2,\cdots,Y_n$ 相互独立同分布，且 $Y_i\sim N(0,1)$，$i=1,2,\cdots,n$.

命题 4.6　设 $X_1,X_2,\cdots,X_n$ 为相互独立同分布的随机变量，且 $X_i\sim N(0,1)$，$i=1,2,\cdots,n$. 记

$$\overline{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n),S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

则 $\overline{X}\sim N\left(0,\frac{1}{n}\right)$，$(n-1)S^2\sim\chi^2(n-1)$，且 $\overline{X}$ 与 S^2 相互独立.

证明　首先，由第 3 章中的(3.44)式可推知 $\overline{X}\sim N\left(0,\frac{1}{n}\right)$. 其次，不难验证

$$(n-1)S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n(\overline{X})^2.$$

现令　$Y_k=\dfrac{1}{\sqrt{k(k+1)}}(X_1+X_2+\cdots+X_k-X_{k+1}),k=1,2,\cdots,n-1;$

$$Y_n=\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)=\sqrt{n}\,\overline{X}.$$

可以验证这是由 $(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 至 $(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$ 的正交变换，即若记 $\boldsymbol{Y}=(Y_1,T_2,\cdots,Y_n)^{\mathrm{T}}$，$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$，当把上述诸式统写成矩阵等式 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$ 后，可以验证 $\boldsymbol{A}$ 为一正交矩阵. 这样，由上述命题知 $Y_1,Y_2,\cdots,Y_n$ 相互独立同分布，且 $Y_i\sim N(0,1)$，$i=1,2,\cdots,n$.

再因 A 为正交矩阵，即知

$$Y_1^2+Y_2^2+\cdots+Y_n^2=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}=(\boldsymbol{A}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=X_1^2+X_2^2+\cdots+X_n^2,$$

从而

$$(n-1)S^2=\sum_{i=1}^{n}X_i^2-n(\overline{X})^2=\sum_{i=1}^{n}Y_i^2-Y_n^2=Y_1^2+Y_2^2+\cdots+Y_{n-1}^2.$$

这样,由命题 4.1 与定义 4.6 即知$(n-1)S^2\sim\chi^2(n-1)$.

此外,$(n-1)S^2$与 Y_n独立,从而也和$\overline{X}$独立.

定理 4.1 的证明.

证明　记

$$X_i^*=\frac{X_i-\mu}{\sigma},i=1,2,\cdots,n,$$

则　$X_1^*,X_2^*,\cdots,X_n^*$相互独立,且有 $X_1^*\sim N(0,1),i=1,2,\cdots,n$.

现记

$$\overline{X}^*=\frac{1}{n}(X_1^*+X_2^*+\cdots+X_n^*),\quad (S^*)^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i^*-\overline{X}^*)^2.$$

首先,由命题 4.6 知,$\overline{X}^*\sim N\left(0,\frac{1}{n}\right)$,$(n-1)(S^*)^2\sim\chi^2(n-1)$且$\overline{X}^*$与$(S^*)^2$相互独立.其次,不难验证

$$\overline{X}^*=\frac{\overline{X}-\mu}{\sigma},\text{即}\overline{X}=\sigma\overline{X}^*+\mu,\quad (S^*)^2=\frac{S^2}{\sigma^2},\text{即 }S^2=\sigma^2(S^*)^2.$$

故知$\overline{X}$与S^2也相互独立,且$\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$,$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

附录 4.3　一般总体枢轴量 T_n 的极限分布

为了给出定理 4.4 的证明,先给出一个预备性命题.

命题 4.7　设

$$X_n\xrightarrow{d}X\quad,\quad Y_n\xrightarrow{P}1,$$

则

$$X_nY_n\xrightarrow{d}X.$$

证明　因 $Y_n\xrightarrow{P}1$,任取 $0<\varepsilon<1$,恒有

$$\lim_{n\to\infty}P\{1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\}=1.$$

记 $A_n=\{1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\}$,便有$\lim\limits_{n\to\infty}P(\overline{A}_n)=0$.由于

$$P\{X_nY_n\leqslant\varepsilon\}=P\{X_nY_n\leqslant x,1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\}+P(\{X_nY_n\leqslant x\}\Lambda\overline{A}_n)$$

且

$$P\{X_nY_n\leqslant x,1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\}\leqslant P\left\{X_n\leqslant\frac{x}{1-\varepsilon},1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\right\}\leqslant P\left\{X_n\leqslant\frac{x}{1-\varepsilon}\right\}$$

$$P\{X_nY_n\leqslant x,1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\}\geqslant P\left\{X_n\leqslant\frac{x}{1+\varepsilon},1-\varepsilon\leqslant Y_n\leqslant 1+\varepsilon\right\}$$

$$=P\left\{X_n\leqslant\frac{x}{1+\varepsilon}\right\}-P\left(\left\{X_n\leqslant\frac{x}{1+\varepsilon}\right\}\Lambda\overline{A}_n\right)$$

$$\geqslant P\left\{X_n \leqslant \frac{x}{1+\varepsilon}\right\} - P(\bar{A}_n),$$

于是

$$P\left\{X_n \leqslant \frac{x}{1+\varepsilon}\right\} - P(\bar{A}_n) \leqslant P\{X_n Y_n \leqslant x\} \leqslant P\left\{X_n \leqslant \frac{x}{1-\varepsilon}\right\} + P(\bar{A}_n).$$

由于对 X 的分布函数 $F(x)$ 的任意连续点有

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \left[P\left\{X_n \leqslant \frac{x}{1+\varepsilon}\right\} - P(\bar{A}_n)\right] = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \left[P\left\{X_n \leqslant \frac{x}{1-\varepsilon}\right\} + P(\bar{A}_n)\right] = F(x),$$

从而由夹逼定理知：

$$\lim_{n \to \infty} P\{X_n Y_n \leqslant x\} = F(x)，\text{对一切 } F(x) \text{的连续点 } x.$$

定理 4.4(2)的证明

由习题四(B)的第 10 题知 $\frac{S}{\sigma} \xrightarrow{P} 1$. 再因 $T_n = \frac{\overline{X} - \mu}{S/\sqrt{n}} = U_n \frac{\sigma}{S}$，

故由定理 4.4(1)与命题 4.7 即知　$F_{T_n}(x) \to \Phi_0(x)$.

习　题　四

(A)

1. 设总体 X 服从以 $\alpha(\alpha>0)$ 为参数的指数分布，$(X_1, X_2, \cdots, X_n)$ 为其一个样本，求该样本的样本密度.

2. 设总体 X 服从闭区间 $[0,1]$ 上的均匀分布，$(X_1, X_2, \cdots, X_n)$ 为其一个样本，求该样本的样本密度.

3. 设总体 X 服从以 $p(0<p<1)$ 为参数的几何分布，$(X_1, X_2, \cdots, X_n)$ 为其一个样本，求该样本的离散样本密度.

4. 设 $(x_1, x_2, \cdots, x_n)$ 与 $(u_1, u_2, \cdots, u_n)$ 为两组样本的样本值，它们有下列关系：

$$u_i = \frac{x_i - a}{b} \qquad (b \neq 0, a \text{ 为常数}).$$

求样本均值 $\overline{u}, \overline{x}$ 及样本方差 s_u^2 与 s_x^2 之间的关系.

5. 设 $(X_1, X_2, \cdots, X_n)$ 是总体 X 的一个容量为 n 的样本，S^2 为该样本的样本方差. 试证 $\frac{n-1}{n} S^2 = A_2 - (A_1)^2$.

6. 说明统计量与枢轴量之间的差别. 试给出统计量与枢轴量的一些例子.

7. 验证(4.10)与(4.11)两式.

8. 验证(4.21)式.

9. 设总体 X 的分布函数为 $F(x)$，$(X_1, \cdots, X_n)$ 为其一个样本，试以 $F(x)$ 表示该样本的极小值 $X_{(1)}$ 与极大值 $X_{(n)}$ 的分布函数. 特别地，当总体 X 服从以 $\alpha(\alpha>0)$ 为参数的指数分布时，试求 $X_{(1)}$ 与 $X_{(n)}$ 的分布函数.

10. 设总体 X 服从自由度为 m 的 χ^2 分布，$(X_1, X_2, \cdots, X_n)$ 为其一个样本，试求该样本

的样本均值$\overline{X}$的密度函数.

11. 设总体 X 服从标准正态分布,从此总体中取出一个容量为 6 的样本 $(X_1,X_2,\cdots,X_6)$,令

$$Y=(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2.$$

试决定常数 c,使得随机变量 cY 服从 χ^2 分布,并求该 χ^2 分布的自由度.

12. 设总体 X 服从以 $\lambda(\lambda>0)$ 为参数的泊松分布,$(X_1,X_2,\cdots,X_n)$ 为其一个样本,试求样本和 $S_n=X_1+X_2+\cdots+X_n$ 的确切分布与渐近分布.

13. 查表求标准正态分布的下列上侧分位数:

$$u_{0.4},u_{0.2},u_{0.1}\text{与}\ u_{0.05}.$$

14. 设总体 $X\sim N(\mu,\sigma^2)$,假如我们要以 99.7% 的概率保证偏差 $|\overline{X}-\mu|<0.1$,试问在 $\sigma^2=0.5$ 时,样本容量应取多大?

15. 查表求 χ^2 分布的下列上侧分位数:

$$\chi^2_{0.95}(5),\chi^2_{0.05}(5),\chi^2_{0.99}(10)\text{与}\ \chi^2_{0.01}(10).$$

16. 证明 F 分布上侧分位数的关系式(4.21)式,并查表求 F 分布的下列上侧分位数:

$$F_{0.95}(4,6),F_{0.975}(3,7)\text{与}\ F_{0.99}(5,5).$$

17. 查表求 t 分布的下列上侧分位数:

$$t_{0.05}(3),t_{0.01}(5),t_{0.10}(7)\text{与}\ t_{0.005}(10).$$

18. 设 $\overline{X},\overline{Y},S_1^2,S_2^2,S^2,U_1$ 由定理 4.3 及其证明中定义,证明

(1) $\overline{X},\overline{Y},S_1^2,S_2^2$ 相互独立;

(2) U_1 与 S^2 相互独立.

(B)

1. 利用切比雪夫不等式求一枚均匀钱币需抛多少次才能使样本均值$\overline{X}$落在 0.4 到 0.6 之间的概率至少为 0.9? 这里$\overline{X}$是样本 $(X_1,X_2,\cdots,X_n)$ 的均值,而 X_i 表示第 i 次抛均匀钱币时正面出现的次数.

2. 设总体 X 的方差为 $\sigma^2=4$,而$\overline{X}$是容量为 100 的样本均值.利用切比雪夫不等式求出一个下限和一个上限,使得$\overline{X}-\mu$(μ 为总体 X 的数学期望)落在这两个界限之间的概率至少为 0.90.

3. 设总体 X 的二阶原点矩存在,$(X_1,X_2,\cdots,X_n)$ 为其一个样本,$\overline{X}$为该样本的样本均值,试证 $X_i-\overline{X}$与$X_j-\overline{X}(i\neq j)$的相关系数为 $\rho=-\dfrac{1}{n-1}$.

4. 设 $X\sim N(0,1)$.利用伽马函数的定义与性质,证明 $E[X^{2n}]=1\cdot3\cdot5\cdot\cdots\cdot(2n-1)$.

5. 记 $Z=\dfrac{X}{Y}$,其中 X 是具有对称密度函数的连续型随机变量,Y 是非负连续型随机变量.试证 Z 的密度函数也是对称函数.

6. 设随机变量 X 具有对称的密度函数 $f(x)$,$F(x)$为其分布函数.再设 $g(x)$与 $G(x)$ 分别为$|X|$的密度函数与分布函数.

(1) 证明 $F(x)+F(-x)=1;G(x)=2F(x)-1,\forall x\geqslant0$.

(2) 证明　$f(x)=\frac{1}{2}g(|x|)$.

7. 验证(4.16)式.

8. (1) 已知 $X_n \xrightarrow{P} a$，　$Y_n \xrightarrow{P} b$. 试证　$X_n + Y_n \xrightarrow{P} a + b$.

(2) 设 $a_n(n\geqslant 1)$是实数，且有 $\lim\limits_{n\to\infty} a_n = a$. 又已知 $Y_n \xrightarrow{P} b$，试证 $a_n Y_n \xrightarrow{P} ab$. 特别地有 $aY_n \xrightarrow{P} ab$.

(3) 已知 $X_n \xrightarrow{P} a$，试证 $X_n^2 \xrightarrow{P} a^2$.

(4) 设 $X_n(n\geqslant 1)$为非负随机变量. 已知 $X_n \xrightarrow{P} a(a\geqslant 0)$，试证 $\sqrt{X_n} \xrightarrow{P} \sqrt{a}$.

9. 设 $(X_1, X_2, \cdots, X_n)$为总体 X 的一个样本，A_k 为此样本的 k 阶原点矩. 若总体 X 的 k 阶原点矩 α_k 存在，利用大数定律证明

$$A_k \xrightarrow{P} \alpha_k (n\to\infty).$$

10. 设 $(X_1, X_2, \cdots, X_n)$是总体 X 的一个容量为 n 的样本，S_n^2 为该样本的样本方差. 另设总体 X 的方差 $DX=\sigma^2$ 存在，试证 $S_n^2 \xrightarrow{P} \sigma^2$，　$\frac{S_n}{\sigma} \xrightarrow{P} 1$.

第 5 章

参数估计与假设检验

统计推断是统计学的重要内容，它大致可分为两类：估计问题与假设检验问题，其中估计问题又分为参数估计和非参数估计，假设检验也分为参数假设检验和非参数假设检验. 本章将介绍参数估计和假设检验的基本思想和方法，并着重讨论求参数的点估计的经典方法以及正态总体参数的假设检验与区间估计，同时也对其他总体（包括小样本和大样本情形）的区间估计问题以及多参数和非参数假设检验问题作了介绍.

§5.1 点估计概述

我们常常会面临这样一类问题：已知总体的分布类型，但不知道其中某些参数的真值. 例如已知总体服从泊松分布，但不知其参数 λ 到底等于多少. 这时我们希望通过所拥有的样本来对未知参数作出估计，这就是参数估计问题. 其实，上述泊松分布总体（$\lambda>0$）代表着一族总体，估计 λ 无非是要推断样本究竟来自这族总体中的哪一个. 另外，总体分布中未知参数的实值函数通常也叫参数. 因此，利用样本估计未知参数的实值函数也属于参数估计问题. 参数估计有点估计与区间估计之分. 我们首先讨论参数的点估计.

一、什么叫点估计

设 $(X_1,\cdots,X_n)$ 为来自总体 X 的样本，$(x_1,\cdots,x_n)$ 为相应的样本值. θ 是总体分布中的未知参数，$\theta\in\Theta$. 这里 Θ 表示 θ 的取值范围，称为**参数空间**. 尽管 θ 是未知的，但它的参数空间 Θ 是事先知道的. 为了估计未知参数 θ，我们构造一个统计量 $h(X_1,\cdots,X_n)$，然后用 $h(X_1,\cdots,X_n)$ 的值 $h(x_1,\cdots,x_n)$ 来估计 θ 的真值. 称 $h(X_1,\cdots,X_n)$ 为 θ 的**估计量**，记作 $\hat{\theta}(X_1,\cdots,X_n)$；称 $h(x_1,\cdots,x_n)$ 为 θ 的**估计值**，记作 $\hat{\theta}(x_1,\cdots,x_n)$. 在不会引起误会的场合，估计量与估计值统称为**点估计**，简称为**估计**，并简记为 $\hat{\theta}$. 事实上，θ 的估计值是数轴上的一个点，用 θ 的估计值 $\hat{\theta}$ 作为 θ 的真值的近似值就相当于用一个点来估计 θ，因此得名为点估计.

如果总体分布中含有多个未知参数 $\theta_1,\cdots,\theta_r,(\theta_1,\cdots,\theta_r)\in\Theta$，那么称统计量 $\hat{\theta}_i(X_1,\cdots,X_n)$ 为 θ_i 的估计量，称相应的值为 θ_i 的估计值，$i=1,2,\cdots,r$.

至于待估参数为未知参数 θ 的实值函数 $g(\theta)$ 时，则称用来估计 $g(\theta)$ 的统计量 $\hat{g}(\theta)$ 为 $g(\theta)$ 的估计量，称相应的值为 $g(\theta)$ 的估计值.

例 5.1　设某种型号的电子元件的寿命 X（以小时计）$\sim f(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$，$(x>0)$. θ 为未知参数，$\theta>0$. 现得样本值为

$$168,130,169,143,174,198,108,212,252,$$

试估计未知参数 θ.

由题意知总体 X 的均值为 θ，即 $\theta=EX$，因此用样本均值 $\overline{X}$ 作为 θ 的估计量看起来是最自然的. 对给定的样本值计算得

$$\overline{x}=\frac{1}{9}(168+130+\cdots+252)=172.7,$$

故 $\hat{\theta}=\overline{X}$ 与 $\hat{\theta}=\overline{x}=172.7$ 分别为 θ 的估计量与估计值.

我们注意到上例中除 $\hat{\theta}=\overline{X}$ 以外 $\hat{\theta}_1=X_1$ 也可作为 θ 的估计量，相应的估计值是 $\hat{\theta}_1=168$；若记 $X_{(1)}=\min\{X_1,\cdots,X_n\}$，$X_{(n)}=\max\{x_1,\cdots,x_n\}$，则 $\hat{\theta}_2=\frac{1}{2}(X_{(1)}+X_{(n)})$ 同样有资格作 θ 的估计量，相应的估计值 $\hat{\theta}_2=\frac{1}{2}(108+252)=180$. 由此可见，点估计的概念相当宽松. 对一个未知参数，原则上可以随意构造其估计. 因此有必要建立一些评价估计量好坏的标准. 下面简单介绍三条最基本的标准：无偏性、有效性及相合性.

二、评价估计量的标准

1. 无偏性

我们希望估计量 $\hat{\theta}$ 的取值不要偏高也不要偏低，即 $\hat{\theta}$ 的平均取值与 θ 的真值一致. 于是导出了无偏性标准.

定义 5.1　设 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 为参数 θ 的估计量，若 $E\hat{\theta}=\theta$，则称 $\hat{\theta}$ 是 θ 的**无偏估计量**，否则称 $\hat{\theta}$ 为 θ 的**有偏估计量**.

若 $\lim\limits_{n\to\infty}E\hat{\theta}=\theta$，则称 $\hat{\theta}$ 是 θ 的**渐近无偏估计量**.

例 5.2　设 $(X_1,\cdots,X_n)$ 为取自总体 X 的样本，总体 X 的均值为 μ，方差为 σ^2. 则

(1) 样本均值 $\overline{X}$ 是 μ 的无偏估计量；

(2) 样本方差 S^2 是 σ^2 的无偏估计量；

(3) 未修正的样本方差,即样本二阶中心矩$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$是$\sigma^2$的有偏估计量,但它是$\sigma^2$的渐近无偏估计量.

解 (1) 因为$EX_i=EX=\mu, i=1,2,\cdots,n$.

$E\overline{X}=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}EX_i=EX=\mu$,故$\hat{\mu}=\overline{X}$是$\mu$的一个无偏估计量.

(2) $DX_i=DX=\sigma^2, i=1,2,\cdots,n$. $D\overline{X}=\frac{1}{n}DX=\frac{\sigma^2}{n}$,

于是

$$\begin{aligned}ES^2&=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]\\&=E\left\{\frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2-n(\overline{X})^2\right]\right\}\\&=\frac{1}{n-1}\left[\sum_{i=1}^{n}EX_i^2-nE(\overline{X})^2\right]\\&=\frac{1}{n-1}\left\{\sum_{i=1}^{n}(\mu^2+\sigma^2)-n[D\overline{X}+(E\overline{X})^2]\right\}\\&=\frac{1}{n-1}(n\sigma^2-\sigma^2)\\&=\sigma^2.\end{aligned}$$

故$\hat{\sigma}^2=S^2$是σ^2的一个无偏估计量.

(3) $E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=E\left(\frac{n-1}{n}S^2\right)$

$$\begin{aligned}&=\frac{n-1}{n}ES^2\\&=\frac{n-1}{n}\sigma^2\\&\neq\sigma^2.\end{aligned}$$

故样本二阶中心矩是σ^2的有偏估计量.但

$$\lim_{n\to\infty}E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=\sigma^2,$$

因此它是σ^2的一个渐近无偏估计量.

注意,如果$\hat{\theta}$是θ的无偏估计量,$g(\theta)$是θ的函数,未必能推出$g(\hat{\theta})$是$g(\theta)$的无偏估计量.例如总体$X\sim N(\mu,\sigma^2)$,$\overline{X}$是μ的无偏估计量,但$(\overline{X})^2$却不是μ^2的无偏估计量.因为

$$E(\overline{X})^2=D\overline{X}+(E\overline{X})^2=\frac{\sigma^2}{n}+\mu^2,$$

而 $\sigma^2>0$，所以 $E(\overline{X})^2\neq\mu^2$.

2. 有效性

有时一个参数存在许多无偏估计量，选用哪一个好呢？显然应该看它们中间哪一个取值更集中，即方差更小. 也就是说，一个好的估计量应具有尽量小的方差. 由此引出了第二个标准——有效性.

定义 5.2　设 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 为参数 θ 的两个无偏估计量，若 $D\hat{\theta}_1<D\hat{\theta}_2$，则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**.

例 5.3　设总体 X 的方差存在且大于零，$EX=\mu$，(X_1,X_2) 为 X 的一个样本，则 $\hat{\mu}_1=\overline{X}$ 与 $\hat{\mu}_2=X_1$ 都是 μ 的无偏估计量，但 $D\hat{\mu}_1=\frac{1}{2}DX<D\hat{\mu}_2=DX$，故 $\hat{\mu}_1$ 比 $\hat{\mu}_2$ 有效.

例 5.4　设总体 $X\sim N(1,\sigma^2)$，其中参数 σ^2 未知，$\sigma^2>0$，$(X_1,\cdots,X_n)$ 为来自总体 X 的样本 $(n>1)$. 考虑 σ^2 的两个估计量：

$$(1)\ \hat{\sigma}_1^2=S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2;(2)\ \hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-1)^2.$$

因为

$$E\hat{\sigma}_1^2=ES^2=\sigma^2,$$

$$E\hat{\sigma}_2^2=E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-1)^2\right]=\frac{1}{n}\sum_{i=1}^{n}E(X_i-EX_i)^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}DX_i=\sigma^2.$$

所以它们都是 σ^2 的无偏估计量. 下面来比较它们的方差. 由于

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n),$$

$$D\left[\frac{(n-1)S^2}{\sigma^2}\right]=2(n-1),$$

$$D\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right]=2n,$$

因此

$$D\hat{\sigma}_1^2=DS^2=\left(\frac{\sigma^2}{n-1}\right)^2D\left[\frac{(n-1)S^2}{\sigma^2}\right]=\frac{2\sigma^4}{n-1},$$

$$D\hat{\sigma}_2^2=D\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-1)^2\right]=\left(\frac{\sigma^2}{n}\right)^2D\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-1)^2\right]$$

$$=\frac{2\sigma^4}{n}.$$

后者小于前者,故 $\hat{\sigma}_2^2$ 较 $\hat{\sigma}_1^2$ 有效.

3. 相合性

所谓相合性就是当样本容量无限增大时估计量 $\hat{\theta}$ 与 θ 的真值任意接近的概率趋于 1.相合性也称一致性,它反映了估计量的一种大样本性质.

定义 5.3 设 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 为未知参数 θ 的估计量,若 $\hat{\theta}$ 依概率收敛于 θ,即对任意 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|<\varepsilon\}=1$$

或

$$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|\geqslant\varepsilon\}=0,$$

则称 $\hat{\theta}$ 为 θ 的(弱)**相合估计量**.

例 5.5 设 $(X_1,\cdots,X_n)$ 是取自总体 X 的样本,且 DX^k 存在,$k=1,2,\cdots,n$. 则 $\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 为 EX^k 的相合估计量,$k=1,2,\cdots,n$.

事实上,对指定的 k,令 $Y=X^k$,$Y_i=X_i^k$,$\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i=\frac{1}{n}\sum_{i=1}^{n}X_i^k$,

由大数定律知

$$P-\lim_{n\to\infty}\overline{Y}=EY=EX^k.$$

从而 $\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是 EX^k 的相合估计量.

作为特例,样本均值 $\overline{X}$ 是总体均值 EX 的相合估计量.

例 5.6 设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,\cdots,X_n)$ 为其样本.试证样本方差 S^2 是 σ^2 的相合估计量.

证 因 $ES^2=\sigma^2$,$DS^2=\frac{2\sigma^4}{n-1}$(可参见例 5.4),故由切比雪夫不等式推得,对任意 $\varepsilon>0$,

$$0\leqslant P\{|S^2-ES^2|\geqslant\varepsilon\}=P\{|S^2-\sigma^2|\geqslant\varepsilon\}$$
$$\leqslant\frac{1}{\varepsilon^2}DS^2=\frac{2\sigma^4}{\varepsilon^2(n-1)}.$$

当 $n\to\infty$ 时上式左、右端均趋于 0,根据相合性定义可知 S^2 是 σ^2 的相合估计量.

§5.2 参数的最大似然估计与矩估计

无偏性、有效性以及相合性从各个不同角度描述了估计量的合理性.那么如

何求得比较理想的估计量呢？本节将介绍两种既经典又很流行的求参数点估计的方法，即最大似然法和矩法.

一、最大似然估计

1. 最大似然法的基本思想

最大似然法的思想很简单：在已经得到试验结果的情况下，我们应该寻找使这个结果出现的可能性最大的那个 θ 作为真 θ 的估计. 下面我们仅就离散型总体和连续型总体这两种情况做进一步分析.

设 $(X_1,\cdots,X_n)$ 为来自总体 X 的样本，X 的分布类型已知，但参数 θ 未知，$\theta\in\Theta$.

(1) 若 X 为离散型随机变量，其概率分布的形式为 $P\{X=x\}=p(x;\theta)$，则样本 $(X_1,\cdots,X_n)$ 的概率分布为 $P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i;\theta)$，在 θ 固定时，上式表示 $(X_1,\cdots,X_n)$ 取值 $(x_1,\cdots,x_n)$ 的概率；反之，当样本值 $(x_1,\cdots,x_n)$ 给定时，它可看作 θ 的函数，我们把它记作 $L(\theta)$，并称

$$L(\theta)=\prod_{i=1}^{n}p(x_i;\theta),\theta\in\Theta \tag{5.1}$$

为**似然函数**. 似然函数 $L(\theta)$ 的值的大小意味着该样本值出现的可能性的大小. 既然已经得到了样本值 $(x_1,\cdots,x_n)$，那它出现的可能性应该是大的，即似然函数的值应该是大的. 因而我们选择使 $L(\theta)$ 达到最大值的那个 θ 作为真 θ 的估计.

(2) 若 X 为连续型随机变量，其密度函数为 $f(x;\theta)$，则样本 $(X_1,\cdots,X_n)$ 的密度函数为 $\prod_{i=1}^{n}f(x_i;\theta)$. 在 θ 固定时，它是 $(X_1,X_2,\cdots,X_n)$ 在 $(x_1,x_2,\cdots,x_n)$ 处的密度，它的大小与 $(X_1,\cdots,X_n)$ 落在 $(x_1,\cdots,x_n)$ 附近的概率的大小成正比. 而当样本值 $(x_1,\cdots,x_n)$ 给定时，它是 θ 的函数. 我们仍把它记为 $L(\theta)$ 并称

$$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta),\theta\in\Theta \tag{5.2}$$

为**似然函数**. 类似于刚才的讨论，我们选择使 $L(\theta)$ 最大的那个 θ 作为真 θ 的估计.

总之，在有了试验结果即样本值 $(x_1,\cdots,x_n)$ 时，似然函数 $L(\theta)$ 反映了 θ 的各个不同值导出这个结果的可能性的大小. 我们选择使 $L(\theta)$ 达到最大值的那个 θ 作为真 θ 的估计. 这种求点估计的方法就叫做**最大似然法**.

定义 5.4　若对任意给定的样本值 $(x_1,\cdots,x_n)$，存在 $\theta^*=$

$\theta^*(x_1,\cdots,x_n)$,使

$$L(\theta^*)=\max_{\theta\in\Theta}L(\theta), \tag{5.3}$$

则称 $\theta^*(x_1,\cdots,x_n)$为θ的**最大似然估计值**,称相应的统计量$\theta^*(X_1,\cdots,X_n)$为θ的**最大似然估计量**.它们统称为 θ 的最大似然估计,可简记为 MLE.

如果未知参数为 $\theta_1,\theta_2,\cdots,\theta_r$,那么似然函数是多元函数 $L(\theta_1,\cdots,\theta_r)$.若对任意给定的样本值$(x_1,\cdots,x_n)$存在$\theta_i^*=\theta_i^*(x_1,\cdots,x_n)$,$i=1,2,\cdots,r$,使

$$L(\theta_1^*,\cdots,\theta_r^*)=\max_{(\theta_1,\cdots,\theta_r)\in\Theta}L(\theta_1,\cdots,\theta_r), \tag{5.4}$$

则称 θ_i^* 为 θ_i 的 MLE,$i=1,2,\cdots,r$.

2. 最大似然估计的一般求法

当似然函数关于未知参数可微时,一般可通过求导数得到 MLE,其主要步骤是:

① 写出似然函数 $L(\theta_1,\cdots,\theta_r)$;

② 令$\frac{\partial}{\partial\theta_i}L=0$ 或$\frac{\partial}{\partial\theta_i}\ln L=0$,$i=1,\cdots,r$,从中解得驻点.注意,函数 L 与 $\ln L$有相同的最值点,而使用后者往往更方便;

③ 判断驻点为最大值点;

④ 求得各参数的 MLE.

例 5.7 设总体 $X\sim N(\mu,\sigma^2)$,μ 与 σ^2 均未知,$-\infty<\mu<+\infty$,$\sigma^2>0$.$(X_1,\cdots,X_n)$为来自 X 的样本,$(x_1,\cdots,x_n)$为样本值.试求 μ 与 σ^2 的最大似然估计.

解 X 的密度为$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$,似然函数

$$\begin{aligned}L(\mu,\sigma^2)&=\prod_{i=1}^{n}f(x_i;\mu,\sigma^2)\\&=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},\end{aligned}$$

于是 $\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$.

$$\text{令}\begin{cases}\frac{\partial}{\partial\mu}\ln L(\mu,\sigma^2)=0,\\ \frac{\partial}{\partial(\sigma^2)}\ln L(\mu,\sigma^2)=0\end{cases}\quad\text{即}\begin{cases}\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0,\\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0,\end{cases}$$

从中解得 $\mu_0=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$,$\sigma_0^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

分别计算

$$A=\frac{\partial^2}{\partial\mu^2}\ln L(\mu,\sigma^2)\Big|_{(\mu_0,\sigma_0^2)}=-\frac{n}{\sigma_0^2},$$

$$B=\frac{\partial^2}{\partial\mu\partial(\sigma^2)}\ln L(\mu,\sigma^2)\Big|_{(\mu_0,\sigma_0^2)}=0,$$

$$C=\frac{\partial^2}{\partial(\sigma^2)^2}\ln L(\mu,\sigma^2)\Big|_{(\mu_0,\sigma_0^2)}=-\frac{n}{2\sigma_0^4}.$$

由于 $A<0$ 且 $AC-B^2=\left(-\frac{n}{\sigma_0^2}\right)\left(-\frac{n}{2\sigma_0^4}\right)>0$，因此 μ_0,σ_0^2 是极大值点同时也是最大值点.从而 $\hat{\mu}=\overline{X}$ 与 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别为 μ 与 σ^2 的最大似然估计量，$\hat{\mu}=\overline{x}$ 与 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$ 分别为 μ 与 σ^2 的最大似然估计值.

正态总体均值与方差的 *MLE* 都是相合估计，并且均值的 *MLE* 是无偏估计，方差的 *MLE* 是渐近无偏估计.

这里有一点需要说明：按照本课程的要求，当似然函数的驻点惟一时，不必验证该驻点是否为最大值点.可直接把驻点作为所求参数的最大似然估计.

最大似然估计有一个简单但很有用的性质：如果 $\hat{\theta}$ 是 θ 的最大似然估计，$u=g(\theta)$ 是 θ 的函数且存在单值反函数 $\theta=h(u)$.那么 $g(\hat{\theta})$ 是 $g(\theta)$ 的最大似然估计.这种性质叫做**最大似然估计的不变性**.该性质还可以推广到多个参数的情形.

例 5.8　试求§5.1例5.1中元件的平均寿命以及概率 $P\{X>180\}$ 的最大似然估计值.

解　先求平均寿命 EX 即 θ 的最大似然估计量.由于似然函数为

$$L(\theta)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i}{\theta}},x_i>0,i=1,\cdots,n,$$

$$\ln L(\theta)=-n\ln\theta-\frac{1}{\theta}\sum_{i=1}^{n}x_i.$$

令　$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L=0$　即　$-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}x_i=0,$

解得　$\theta_0=\overline{x}$，故 $\hat{\theta}=\overline{X}$ 是 θ 的最大似然估计量.从而平均寿命 θ 的最大似然估计值为 $\overline{x}=172.7$.

$P\{X>180\}=e^{-\frac{180}{\theta}}$ 是 θ 的函数.由不变性可得 $P(X>180)$ 的最大似然估计量为 $e^{-\frac{180}{\overline{X}}}$，从而 $P(X>180)$ 的最大似然估计值为 $e^{-\frac{180}{172.7}}=0.353$.

综上所述,求最大似然估计的要点在于掌握最大似然法的基本思想.最大似然法充分利用了已知总体分布类型所蕴涵的信息,因此最大似然估计一般具有较多的优良性质.当然它的计算也往往比较麻烦,在不少场合只能通过数值方法并借助于计算机求得近似解.

二、矩估计

1. 矩法的基本思想

除最大似然法外,矩法也是求点估计的常用方法.它的基本思想是:用相应的样本矩去估计总体矩;用相应的样本矩的函数去估计总体矩的函数.例如

$$\hat{EX}=\overline{X},\hat{DX}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

一般地,若记

$$\alpha_k=EX^k,\beta_k=E(X-EX)^k,A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,$$

则总体的 k 阶原点矩用相应的样本 k 阶原点矩来估计;而总体的 k 阶中心矩用相应的样本 k 阶中心矩来估计,即

$$\hat{\alpha}_k=A_k,k=1,2,\cdots,$$
$$\hat{\beta}_k=B_k,k=2,3,\cdots.$$

这种求点估计的方法叫做**矩法**.用矩法确定的估计量称为**矩估计量**,相应的估计值称为**矩估计值**,矩估计量与矩估计值统称为矩估计,可简记为 ME.

矩法是一种古老的方法,它的特点是并不要求已知总体分布的类型.只要未知参数可以表示成总体矩的函数,就能求出其矩估计.当总体分布类型已知时,由于没有充分利用总体分布所提供的信息,矩法估计不一定是理想的估计.但因矩法估计简便易行,而且具有一定的优良性,所以应用仍然十分广泛.

2. 矩估计的求法

按照矩法的基本思想求矩估计的一般步骤为:

① 从总体矩入手将待估参数 θ 表示为总体矩的函数,即

$$\theta=g(\alpha_1,\cdots,\alpha_l;\beta_2,\cdots,\beta_s);$$

② 用 A_k,B_k 分别替换 g 中的 α_k,β_k;

③ $\hat{\theta}=g(\hat{\alpha}_1,\cdots,\hat{\alpha}_l;\hat{\beta}_2,\cdots,\hat{\beta}_s)=g(A_1,\cdots,A_l;B_2,\cdots,B_s)$ 即为 θ 的 ME.

例 5.9 设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,\cdots,X_n)$ 为取自总体 X 的样本.试求 μ,σ^2 的矩估计量.

解 $\mu=EX,\sigma^2=DX$,故 $\hat{\mu}=\hat{EX}=\overline{X}$ 与 $\hat{\sigma}^2=\hat{DX}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别为 μ 与 σ^2 的矩估计量.

由此可见，正态总体 $N(\mu,\sigma^2)$ 中 μ 与 σ^2 的最大似然估计和矩估计是完全一样的.

例 5.10　设总体 X 服从参数为 m，p 的二项分布，m 已知，p 未知. $(X,\cdots,X_n)$ 为其样本. 试求

(1) p 的矩估计量；

(2) p 与 q 之比的矩估计量，其中 $q=1-p$.

解(1) 因 $EX=mp$，$p=\dfrac{EX}{m}$，又 $\hat{EX}=\overline{X}$，故 $\hat{p}=\dfrac{\overline{X}}{m}$ 为 p 的矩估计量.

(2) 令 $g(p)=\dfrac{p}{q}=\dfrac{p}{1-p}$，$\hat{p}=\dfrac{\overline{X}}{m}$，故

$$\hat{g(p)}=\frac{\hat{p}}{1-\hat{p}}=\frac{\dfrac{\overline{X}}{m}}{1-\dfrac{\overline{X}}{m}}=\frac{\overline{X}}{m-\overline{X}}$$

即为 $g(p)$ 的矩估计量.

但要注意，因 $DX=mpq$，$g(p)=\dfrac{p}{q}=\dfrac{mp^2}{DX}$，故

$$\hat{g(p)}=\frac{m\left(\dfrac{\overline{X}}{m}\right)^2}{\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}=\frac{(\overline{X})^2}{\dfrac{m}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

也是 $g(p)$ 的矩估计量.

由上例看出，利用不同的矩得到的矩估计一般是不同的，这是因为不同的矩包含了总体分布的不同信息，如何充分利用不同的矩所包含的总体分布的信息从而得到更有效的估计，导致了近年在计量经济学中十分流行的**广义矩方法(GMM)**.

例 5.11　设总体 X 的密度函数为

$$f(x;\theta,\mu)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x-\mu}{\theta}}, & x\geqslant\mu,\\ 0, & x<\mu,\end{cases}$$

其中参数 θ,μ 均未知，$\theta>0$. $(X_1,\cdots,X_n)$ 为取自 X 的样本. 试求 θ,μ 的矩估计量.

解　令 $\dfrac{x-\mu}{\theta}=t$，则

$$EX=\int_{-\infty}^{+\infty}xf(x;\theta,\mu)dx=\int_{\mu}^{+\infty}\frac{x}{\theta}e^{-\frac{x-\mu}{\theta}}dx$$

$$=\int_0^{+\infty}(\theta t+\mu)e^{-t}dt=\theta\Gamma(2)+\mu\Gamma(1)=\theta+\mu.$$

$$DX = \int_{-\infty}^{+\infty}(x - EX)^2 f(x;\theta,\mu)\mathrm{d}x = \int_{\mu}^{+\infty}\frac{(x-\theta-\mu)^2}{\theta}\mathrm{e}^{-\frac{x-\mu}{\theta}}\mathrm{d}x$$
$$= \theta^2\Gamma(3) - 2\theta^2\Gamma(2) + \theta^2\Gamma(1) = \theta^2.$$

于是从方程组$\begin{cases} EX = \theta + \mu, \\ DX = \theta^2, \end{cases}$得$\begin{cases} \theta = \sqrt{DX}, \\ \mu = EX - \sqrt{DX}. \end{cases}$

又 $\hat{EX} = \overline{X}, \hat{DX} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$

从而 θ 与 μ 的矩估计量分别为

$$\hat{\theta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2},$$

$$\hat{\mu} = \overline{X} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_0 - \overline{X})^2}.$$

§5.3 置信区间

前面讨论了参数的点估计，即用一个点去估计未知参数. 比如正态总体 $X \sim N(\mu,\sigma^2)$，我们可用$\overline{X}$作为μ 的估计量，用相应的点$\overline{x}$作为 μ 的近似值，而且 $\hat{\mu} = \overline{X}$是 μ 的优良估计. 但是任何一种估计，如果不注明估计的误差，这种估计恐怕是没有多大意义的. 点估计不能直接提供其估计的误差，因此需要引入另一类估计——区间估计. 在区间估计理论中，被广泛接受的一种观点是置信区间. 它是由奈曼(Neyman)于 1934 年提出的.

一、置信区间的概念

粗略地说，置信区间就是一个随机区间，它能以足够大的概率套住我们感兴趣的参数. 例如 $\hat{\theta}$是未知参数 θ 的一个估计量，$(\hat{\theta}-\delta,\hat{\theta}+\delta)$为一个随机区间，其中 $\delta > 0$. 若能使该区间套住 θ 的概率等于事先指定的数 $1-\alpha$，即 $P\{\hat{\theta}-\delta < \theta < \hat{\theta}+\delta\} = 1-\alpha$，区间$(\hat{\theta}-\delta,\hat{\theta}+\delta)$便是 θ 的一个置信区间. 下面给出置信区间的一般定义.

定义 5.5 设 θ 为总体分布的未知参数，$\theta \in \Theta$. $(X_1,\cdots,X_n)$为来自总体 X 的样本. 对给定的数 $1-\alpha(0<\alpha<1)$，若存在两个统计量$\underline{\theta} = \underline{\theta}(X_1,\cdots,X_n)$与$\overline{\theta} = \overline{\theta}(X_1,\cdots,X_n)$，使得

$$P\{\underline{\theta} < \theta < \overline{\theta}\} = 1-\alpha, \tag{5.5}$$

则称随机区间$(\underline{\theta},\overline{\theta})$为 θ 的 $1-\alpha$ **置信区间**，称 $1-\alpha$ 为**置信水平**，又分别称$\underline{\theta}$ 与

$\overline{\theta}$ 为 θ 的**置信下限**与**置信上限**.

一旦有了样本值 $(x_1,\cdots,x_n)$，区间 $(\underline{\theta},\overline{\theta})$ 的端点也随之确定. 称区间 $(\underline{\theta}(x_1,\cdots,x_n),\overline{\theta}(x_1,\cdots,x_n))$ 为置信区间 $(\underline{\theta},\overline{\theta})$ 的一个**实现**，它是一个普通的区间，也简称为置信区间.

定义 5.5 中置信水平 $1-\alpha$ 的含义是：随机区间 $(\underline{\theta},\overline{\theta})$ 套住 θ 的概率为 $1-\alpha$，而不套住 θ 的概率为 α. 用频率来解释就是：如果重复试验 100 次获得了样本 $(X_1,\cdots,X_n)$ 的 100 个样本值，相应地得到了 $(\underline{\theta},\overline{\theta})$ 的 100 个实现，那么在这 100 个普通区间中，大约有 $100(1-\alpha)$ 个区间套住了 θ，而不套住 θ 的区间大约有 100α 个. 若令 $1-\alpha=0.95$，则其中大约有 95 个区间套住 θ，大约有 5 个区间不套住 θ.

事实上，置信区间 $(\underline{\theta},\overline{\theta})$ 也是对未知参数 θ 的一种估计. 区间的长度意味着误差，因此可以说区间估计与点估计是互补的两种参数估计.

二、寻求置信区间的方法

通常采用基于点估计构造置信区间的方法来获得置信区间. 我们通过一个简单的例子来说明求置信区间的一般步骤.

例 5.12　设总体 $X\sim N(\mu,\sigma^2)$，σ^2 已知，μ 未知，$(X_1,\cdots,X_n)$ 为来自 X 的样本. 试求 μ 的 $1-\alpha$ 置信区间.

样本均值 $\overline{X}$ 是 μ 的最大似然估计，从 $\overline{X}$ 出发考虑一个样本与未知参数 μ 的函数 $U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$，它不包含其他未知参数，并且它的分布已知，$U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，因而它是一个枢轴量. 对事先指定的 $1-\alpha$，确定 λ_1，λ_2，使 $P\left\{\lambda_1<\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}<\lambda_2\right\}=1-\alpha$. 显然满足条件的 λ_1,λ_2 不止一对，我们选取 $-u_{\frac{\alpha}{2}}$ 为 λ_1，$u_{\frac{\alpha}{2}}$ 为 λ_2，其中 $u_{\frac{\alpha}{2}}$ 是水平 α 的双侧分位数（如图 5.1），由

$$P\left\{-u_{\frac{\alpha}{2}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<u_{\frac{\alpha}{2}}\right\}=1-\alpha,$$

经不等式变形得

$$P\left\{\overline{X}-u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right\}=1-\alpha.$$

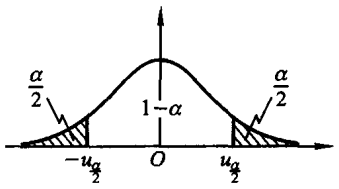


图 5.1　$N(0,1)$ 分布的分位数

于是 $\left(\overline{X}-u_{\frac{\alpha}{2}}\cdot\dfrac{\sigma}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\cdot\dfrac{\sigma}{\sqrt{n}}\right)$ 为 μ 的 $1-\alpha$ 置信区间. 若令 $1-\alpha=95\%$，则 μ 的 95% 置信区间为 $\left(\overline{X}-1.96\cdot\dfrac{\sigma}{\sqrt{n}},\overline{X}+1.96\cdot\dfrac{\sigma}{\sqrt{n}}\right)$. 用该区间估计 μ，我们不仅直接可知估计成功的概率是 95%，而且能够以 95%

的把握断言以$\overline{X}$代替μ的绝对误差小于$1.96\cdot\frac{\sigma}{\sqrt{n}}$.

由此看来,基于点估计构造置信区间的关键在于寻找一个合适的样本的函数u,在它的基础上推导出所要的区间.现在把求未知参数θ的置信区间的一般步骤归纳如下:

① 选取θ的一个较优的点估计$\hat{\theta}$;

② 围绕$\hat{\theta}$寻找一个依赖于样本与θ的函数$u=u(X_1,\cdots,X_n;\theta)$.$u$的分布为已知分布.像$u$这样的函数,在前一章中被称为枢轴量;

③ 对给定的置信水平$1-\alpha$,确定λ_1与λ_2,使

$$P\{\lambda_1<u<\lambda_2\}=1-\alpha,$$

一般可选取满足$P\{u\leqslant\lambda_1\}=P\{u\geqslant\lambda_2\}=\frac{\alpha}{2}$的$\lambda_1$与$\lambda_2$;

④ 利用不等式变形导出套住θ的置信区间$(\underline{\theta},\overline{\theta})$.

例 5.13 设总体X的密度为

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

未知参数$\theta>0$,$(X_1,\cdots,X_n)$为取自X的样本.

(1) 试证$\frac{2n\overline{X}}{\theta}\sim\chi^2(2n)$;

(2) 试求θ的$1-\alpha$置信区间.

解 (1) 由第2章习题(A)34知$\frac{2}{\theta}X_i\sim\chi^2(2)$,$i=1,\cdots,n$.

又由χ^2分布的可加性得

$$\sum_{i=1}^{n}\frac{2}{\theta}X_i\sim\chi^2(2n).$$

而

$$\sum_{i=1}^{n}\frac{2}{\theta}X_i=\frac{2}{\theta}\sum_{i=1}^{n}X_i=\frac{2n}{\theta}\overline{X}.$$

故

$$\frac{2n}{\theta}\overline{X}\sim\chi^2(2n).$$

(2) $\overline{X}$是θ的最大似然估计,从$\overline{X}$出发考虑枢轴量$u=\frac{2n}{\theta}\overline{X}$,由(1)知$u$的分布只依赖于样本容量$n$,即$u=\frac{2n}{\theta}\overline{X}\sim\chi^2(2n)$.对给定的$1-\alpha$,由

$$P\left\{\chi^2_{1-\frac{\alpha}{2}}(2n)<\frac{2n}{\theta}\overline{X}<\chi^2_{\frac{\alpha}{2}}(2n)\right\}=1-\alpha,$$

经不等式变形得

$$P\left\{\frac{2n\overline{X}}{\chi^2_{\frac{\alpha}{2}}(2n)}<\theta<\frac{2n\overline{X}}{\chi^2_{1-\frac{\alpha}{2}}(2n)}\right\}=1-\alpha.$$

于是 $\left(\frac{2n\overline{X}}{\chi^2_{\frac{\alpha}{2}}(2n)},\frac{2n\overline{X}}{\chi^2_{1-\frac{\alpha}{2}}(2n)}\right)$ 为所求置信区间.

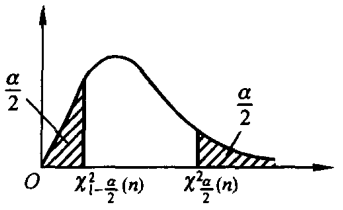


图 5.2　χ^2 分布的分位数

综上,为求参数的置信区间,关键是构造一个样本和待估参数的一个函数—枢轴量,并确定该枢轴量的分布.构造枢轴量的方法一般基于点估计,但对一般总体而言,确定枢轴量的分布往往十分困难.因此,接下来我们只讨论:①对一类特殊的总体——正态总体求精确的置信区间;②在大样本情形利用枢轴量的极限(渐近)分布求近似置信区间.

三、正态总体参数的置信区间

与其他总体相比,正态总体参数的置信区间是最完善的,应用也最广泛.在构造正态总体参数的置信区间的过程中,t 分布、χ^2 分布、F 分布以及 $N(0,1)$ 分布扮演了重要角色.

设总体 $X\sim N(\mu,\sigma^2)$,$-\infty<\mu<+\infty$,$\sigma^2>0$.$(X_1,\cdots,X_n)$ 为来自 X 的样本.

1. 均值 μ 的置信区间

(1) 方差 σ^2 已知的情形

我们在上一节例 5.12 已讨论了在 σ^2 已知的条件下 μ 的 $1-\alpha$ 置信区间为 $\left(\overline{X}-u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$,可简记为 $\left(\overline{X}\pm u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$.

事实上,在给定的置信水平为 $1-\alpha$ 时,对任意的 $\alpha_1>0$,$\alpha_2>0$,$\alpha_1+\alpha_2=\alpha$,凡是满足

$$P\left(u_{1-\alpha_2}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<u_{\alpha_1}\right)=1-\alpha$$

的区间 $\left(\overline{X}-u_{\alpha_1}\cdot\frac{\sigma}{\sqrt{n}},\overline{X}-u_{1-\alpha_2}\cdot\frac{\sigma}{\sqrt{n}}\right)$ 都是 μ 的置信区间,只不过在所有这类区间中就数 $\alpha_1=\alpha_2=\frac{\alpha}{2}$ 时的区间长度最短(见图 5.3).这个区间正是前面所推导出的 $\left(\overline{X}\pm u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$.

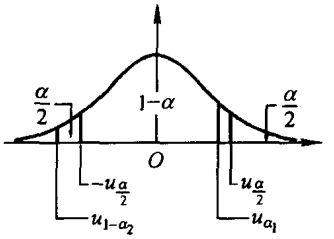


图 5.3　比较 $N(0,1)$ 分布的分位数

例 5.14　设总体 $X\sim N(\mu,\sigma^2)$,其中 μ 未知,$\sigma=4.50$.现取得样本容量为 36 的一个样本,

测得样本均值为$\bar{x}=54.40$. 试求 μ 的 95% 置信区间.

解 由题意知 $n=36,\sigma=4.50,\alpha=0.05$,查表得:$u_{\frac{0.05}{2}}=1.96$,于是

$$u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}=1.96\times\frac{4.50}{\sqrt{36}}=1.47.$$

从而 μ 的 95% 置信区间为

$$(\bar{x}-1.47,\bar{x}+1.47)=(52.93,55.87).$$

这里要注意,(52.93,55.87)是一个普通区间,"μ 属于该区间"这件事情的可信程度为 95%,或该区间属于那些套住 μ 的区间的概率为 95%.

例 5.15 设总体 $X\sim N(\mu,\sigma^2)$,其中 μ 未知,$\sigma^2=4$. 设$(X_1,\cdots,X_n)$为其一个样本.

(1) 当 $n=16$ 时,试求置信水平分别为 0.9 及 0.95 的 μ 的置信区间的长度.

(2) n 多大方能使μ 的 90% 置信区间的长度不超过 1?

(3) n 多大方能使μ 的 95% 置信区间的长度不超过 1?

解 (1) 记 μ 的置信区间长度为Δ,则

$$\Delta=\left(\overline{X}+u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)-\left(\overline{X}-u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)=2u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}.$$

于是当 $1-\alpha=90\%$ 时,$\Delta=2\times1.65\times\frac{2}{\sqrt{16}}=1.65$,当 $1-\alpha=95\%$ 时 $\Delta=2\times1.96\times\frac{2}{\sqrt{16}}=1.96$.

(2) 欲使 $\Delta\leqslant1$ 即 $2u_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\leqslant1$,必须 $n\geqslant\left(2\sigma u_{\frac{\alpha}{2}}\right)^2$,于是当 $1-\alpha=90\%$ 时,$n\geqslant(2\times2\times1.65)^2$即 $n\geqslant44$. 也就是说,样本容量 n 至少为 44 时,μ 的 90% 置信区间的长度才不超过 1.

(3) 当 $1-\alpha=95\%$ 时,类似可得 $n\geqslant62$.

在用置信区间对未知参数进行区间估计时,我们总是希望估计的可靠度要高,即置信区间包含未知参数的概率 $1-\alpha$ 越大越好;另一方面希望估计的精确度要高,即置信区间的长度 Δ 越小越好. 但上面的例子给我们的启示是:当样本容量 n 固定时可靠度与精确度是互相制约的. 如果提高可靠度即增大 $1-\alpha$ 的值,势必拉长区间,从而精确度就减小;反之,缩短区间提高了精确度,则可靠度降低. 总之,"鱼和熊掌不可兼得",若要可靠度与精确度两者都高除非增加样本容量.

(2) 方差 σ^2 未知的情形

用 σ^2 的无偏估计 S^2 代替 σ^2，选取枢轴量为 $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$，由定理 4.2 知 $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$. 对给定的置信水平 $1-\alpha$，由

$$P\left\{-t_{\frac{\alpha}{2}}(n-1)<\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\frac{\alpha}{2}}(n-1)\right\}=1-\alpha,$$

经不等式变形得

$$P\left\{\overline{X}-t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}<\mu<\overline{X}+t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}\right\}=1-\alpha,$$

因此，μ 的 $1-\alpha$ 置信区间为

$$\left(\overline{X}-t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}},\overline{X}+t_{\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}\right).$$

可简记为 $\left(\overline{X}\pm t_{\frac{\alpha}{2}}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right)$.

2. 方差 σ^2 的置信区间

在实际问题中 μ 与 σ^2 往往都是未知的，下面主要讨论在 μ 未知时方差 σ^2 的置信区间.

从 σ^2 的无偏估计 S^2 出发考虑枢轴量 $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$，由定理 4.2 知 $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$. 对给定的 $1-\alpha$，由

$$P\left\{\chi^2_{1-\frac{\alpha}{2}}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\frac{\alpha}{2}}(n-1)\right\}=1-\alpha,$$

得

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right\}=1-\alpha,$$

因此 σ^2 的 $1-\alpha$ 置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

也即

$$\left(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right).$$

另外，标准差 σ 的 $1-\alpha$ 置信区间为

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}},\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right).$$

如果 μ 已知，只需考虑枢轴量为 $\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$. 由 χ^2 分布的定义知

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n).$$

类似地可导出 σ^2 的 $1-\alpha$ 置信区间为

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right).$$

例 5.16 为考察某大学成年男性的胆固醇水平,现抽取了样本容量为 25 的一个样本,并测得样本均值为 $\overline{x}=186$,样本标准差为 $s=12$. 假定胆固醇水平 $X\sim N(\mu,\sigma^2)$,μ 与 σ^2 均未知. 分别求 μ 以及 σ 的 90% 置信区间.

解 μ 的 $1-\alpha$ 置信区间为 $\left(\overline{X}\pm t_{\frac{\alpha}{2}}(n-1)\cdot\frac{s}{\sqrt{n}}\right)$. $\alpha=0.1$, $s=12$, $n=25$, 查表得 $t_{\frac{0.1}{2}}(25-1)=1.710\ 9$, 于是 $t_{\frac{\alpha}{2}}(n-1)\cdot\frac{s}{\sqrt{n}}=1.710\ 9\times\frac{12}{\sqrt{25}}=4.106$, 从而 μ 的 90% 置信区间为 (186 ± 4.106), 即 $(181.89,190.11)$.

σ 的 $1-\alpha$ 置信区间为 $\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}},\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right)$. 查表得

$$\chi^2_{\frac{0.1}{2}}(25-1)=36.42,\ \chi^2_{1-\frac{0.1}{2}}(25-1)=13.85,$$

于是置信下限为 $\sqrt{\frac{24\times12^2}{36.42}}=9.74$, 置信上限为 $\sqrt{\frac{24\times12^2}{13.85}}=15.80$, 从而 σ 的 90% 置信区间为 $(9.74,15.80)$.

四、大样本情形的渐近置信区间

如果枢轴量的分布不易确定,有时可用极限分布来构造近似的置信区间,当然此时要求样本容量足够大. 近似置信区间的求法与精确的置信区间求法类似,不同的只是将枢轴量的精确分布改为极限分布. 以下仅举例说明.

例 5.17 设总体 X 服从参数为 p 的两点分布,p 未知,$0<p<1$. $(X_1,\cdots,X_n)$ 为其样本. 试求 p 的置信区间.

解 我们知道 p 的最大似然估计量为 $\overline{X}$. 基于 $\overline{X}$ 考虑 $u=\frac{\overline{X}-EX}{\sqrt{DX/n}}=\frac{\overline{X}-p}{\sqrt{p(1-p)/n}}$. 根据定理 3.10 知当 n 足够大时 u 近似服从 $N(0,1)$ 分布. 因此,对给定的置信水平 $1-\alpha$, 由

$$P\left\{\left|\frac{\overline{X}-p}{\sqrt{p(1-p)/n}}\right|<u_{\frac{\alpha}{2}}\right\}\approx1-\alpha,$$

经不等式变形得

$$P\{ap^2+bp+c<0\}\approx 1-\alpha,$$

其中

$$a=n+\left(u_{\frac{\alpha}{2}}\right)^2, b=-2n\overline{X}-\left(u_{\frac{\alpha}{2}}\right)^2, c=n(\overline{X})^2.$$

又由 $a>0$ 知 $ap^2+bp+c<0$ 等价于 $p_1<p<p_2$，其中

$$p_1=\frac{1}{2a}\left(-b-\sqrt{b^2-4ac}\right), p_2=\frac{1}{2a}\left(-b+\sqrt{b^2-4ac}\right).$$

总之，对给定的 $1-\alpha$，存在 p_1 与 p_2 使

$$P(p_1<p<p_2)\approx 1-\alpha,$$

于是 (p_1,p_2) 是 p 的一个置信水平近似为 $1-\alpha$ 的置信区间.

然而在实际问题中往往采用下面简化了的置信区间：

$$\left(\overline{X}-u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}, \overline{X}+u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}\right).$$

这是因为当 n 足够大时不仅 $\dfrac{\overline{X}-p}{\sqrt{p(1-p)/n}}$ 近似服从 $N(0,1)$ 分布，而且可以证明 $\overline{X}$ 代替分母中的 p 后，$\dfrac{\overline{X}-p}{\sqrt{\overline{X}(1-\overline{X})/n}}$ 仍然近似服从 $N(0,1)$ 分布. 所以当 n 足够大时，由 $P\left(\left|\dfrac{\overline{X}-p}{\sqrt{\overline{X}(1-\overline{X})/n}}\right|<u_{\frac{\alpha}{2}}\right)\approx 1-\alpha$，得

$$P\left\{\overline{X}-u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}<p<\overline{X}+u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}\right\}\approx 1-\alpha.$$

从而得到上述近似置信区间.

例 5.18　为了研究在一指定时间段内某地区的国际互联网用户所占的比例，随机地调查了该地区的 400 名居民，发现其中有 108 名居民为上网者. 试求该地区居民的上网率 p 的 95% 置信区间.

解　由题意知总体服从 0－1 分布，参数 p 即上网率，上网率是一种比率. 现在要求 p 的置信区间.

p 的置信水平近似为 $1-\alpha$ 的置信区间是

$$\left(\left(\overline{X}-u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}, \overline{X}+u_{\frac{\alpha}{2}}\sqrt{\overline{X}(1-\overline{X})/n}\right)\right).$$

现已知 $n=400$，$\overline{X}=\dfrac{108}{400}=0.27$，$\alpha=0.05$，$u_{\frac{0.05}{2}}=1.96$，故所求置信区间为 $(0.27-1.96\times\sqrt{0.27\times 0.73/400}, 0.27+1.96\times\sqrt{0.27\times 0.73/400})$，即 $(0.23, 0.31)$.

此外，如果一个总体 X，其均值 μ，方差 σ^2 是两个独立的参数，或者我们根本就不知道 X 的分布类型，那么在大样本情形. 对参数 μ 或 σ^2 的区间估计完全

类似于正态总体情形，只不过那里枢轴量的精确分布在这里均变为渐近分布，相应的置信区间变为近似的置信区间.

例如，σ^2未知，求 μ 的区间估计.

由定理 4.4，当样本容量 n 充分大时，枢轴量

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

渐近服从 $N(0,1)$，于是 μ 的近似置信区间为

$$\left(\overline{X}-u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right).$$

例 5.19 某厂新研究并开发了某类设备所需的关键部件，现无法确定此部件的连续使用寿命 X(单位:kh)所服从的分布类型.通过加速失效试验法，测试 100 个此类部件的连续使用寿命.测得样本平均值为$\overline{x}=17.84$，样本标准差为 $s=1.25$，试由试验结果求 $\mu=EX$ 的置信水平为 99% 的近似置信区间.

解 由题设$\overline{x}=17.84$，$s=1.25$，$n=100$，给定 $\alpha=0.01$，查附表 $u_{\frac{\alpha}{2}}=2.56$，计算可得

$$\overline{x}\pm u_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}}=17.84\pm0.32.$$

故 μ 的置信水平为 99% 的近似置信区间为(17.52,18.16).

§5.4 假设检验概述

本章前三节讨论了统计推断中的估计问题，在接下来的几节中，我们将讨论另一类统计推断问题——假设检验.

一、假设检验问题的提法

为说明什么是假设检验问题，我们先来看几个实际例子.

例 5.20 某厂生产的一批产品，其出厂标准为:次品率不超过 4%，现抽测 60 件产品，发现有 3 件次品，问这批产品能否出厂?

若以 X 记一次抽测的结果，$X=1$ 与 $X=0$ 分别表示抽到正品与次品，则 X 为伯努利总体且 $P\{X=1\}=p$，这里 p 为次品率.进而例中提出的问题实际上是要判断是否有:$p\leqslant4\%$，我们把它称为一个假设.p 是未知的，自然可以用样本的次品率去估计总体的次品率 p.但由于抽样的随机性，点估计值不一定是参数的真值，我们不能简单地根据估计值$\frac{3}{60}=5\%$认定次品率超过 4%，而必须提

出合理的方法去判断假设是否成立.

例 5.21　某厂宣称已采取大力措施治理废水污染,根据经验,废水中所含某种有毒物质的浓度 X(单位:mg/kg)服从正态分布.现环保部门抽测了 9 个水样,测得样本平均值为 $\overline{x}=17.4$,样本标准差为 $s=2.4$,以往该厂废水中有毒物质的平均浓度为 18.2,试问有毒物质的浓度有无显著变化?

本例中,$X\sim N(\mu,\sigma^2)$,其中 μ,σ^2 均是未知的,直观上看,有毒物质的浓度有所降低,但这种差异也有可能是抽样的随机性造成的.我们可以先提出假设:有毒物质的浓度无显著变化,即 $\mu=18.2$,然后根据样本去判断假设是否成立.

例 5.22　随机抽测了 50 名 2000 年 1 月出生的男婴的体重,希望确定男婴的体重 X 是否服从正态分布.

设 $F(x)$为 X 的分布函数,同样先提出假设:$F(x)$是 $N(\mu,\sigma^2)$分布,然后利用样本去判断假设是否成立.

上述各例所述问题的共同点是:对总体分布函数的类型或分布函数中的参数提出假设,希望通过抽样并根据样本提供的信息对假设是否成立进行推断.这类问题即是统计推断的另一类基本问题——假设检验.

在假设检验问题中,我们把任何一个有关总体未知分布的假设称为**统计假设**,简称**假设**.通常把待检验的假设称为**原假设**或**零假设**,记为 H_0,与之对立的假设则称为**备择假设**或**对立假设**,记为 H_1.一个假设检验问题通常简记为 $H_0\longleftrightarrow H_1$,比如上述三例假设问题可分别简述为:

$$H_0:p\leqslant 4\%\longleftrightarrow H_1:p>4\%.$$

$$H_0:\mu=18.2\longleftrightarrow H_1:\mu\neq 18.2.$$

$$H_0:F(x)\text{是 }N(\mu,\sigma^2)\text{分布}\longleftrightarrow H_1:F(x)\text{不是 }N(\mu,\sigma^2)\text{分布}.$$

通常为了叙述更加简洁,可以不列出备择假设.

上述三例中给出的假设检验问题又有一定区别.在例 5.20 和例 5.21 中,总体的分布类型是已知的,未知的只是其中的一个或几个参数,统计假设只与这些未知参数有关,我们称为**参数假设**,相应的检验称为**参数假设检验**.而在例 5.22 中,总体的分布类型也是未知的,在这种情形往往需要直接针对总体分布的具体形式或总体分布的某些特征提出假设,我们称此类假设为**非参数假设**,相应的检验称为**非参数假设检验**.

进一步地,在参数假设检验问题中,总体中可能有多个未知的参数,实际问题有时只需对其中某一个参数提出假设并进行检验,有时可能需要对多个参数一起提出一个假设并进行检验,根据这一区别,我们可将参数假设检验细分为单参数假设检验与多参数假设检验.

我们将重点介绍参数假设检验,对于非参数假设检验只予以简单介绍.而对

参数假设检验，我们会涉及多参数假设检验，但重点讨论的是单参数假设检验.

二、假设检验的思想和原理

我们以例 5.21 为例来说明如何进行假设检验.

首先建立统计假设

$$H_0:\mu=\mu_0 \longleftrightarrow H_1:\mu\neq\mu_0(\text{这里 } \mu_0=18.2).$$

由于 μ 未知，我们不能直接对假设 $H_0:\mu=\mu_0$ 成立与否作出判断，自然地，我们借助于样本对 μ 作出估计. 设 $(X_1,X_2,\cdots,X_n)$ 是总体 X 一个样本容量为 n 的样本，$(x_1,x_2,\cdots,x_n)$ 是相应的样本观察值，那么 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 是 μ 的一个估计量，$\overline{x}$ 为相应的估计值. 我们也许想到用 μ 的估计值 $\overline{x}$ 代替 μ 来检验 H_0，但由于样本的随机性造成的估计误差使得 $\overline{x}$ 几乎不会真正等于 μ，所以即使 $H_0:\mu=\mu_0$ 为真，由于估计误差的存在，$\overline{x}$ 也不会真正等于 μ_0，而是处于 μ_0 的附近. 因而我们不能简单地根据是否有 $\overline{x}=\mu_0$ 来判断 $H_0:\mu=\mu_0$ 是否成立. 但是，如果 $H_0:\mu=\mu_0$ 为真，$\overline{x}$ 偏离 μ_0 仅仅是由于随机误差的原因，那么 $\overline{x}$ 会以很大概率落在 μ_0 附近一定的范围内，而远离 μ_0 的概率会很小. 一般认为“一个小概率事件在一次试验中不太可能发生”(通常称为**小概率原理**)，如果观察的样本值 $\overline{x}$ 实际上是远离 μ_0 而使得小概率事件发生了，也就是说一个不太可能发生的事件发生了，我们自然有充足的理由否定假设 $H_0:\mu=\mu_0$ 的正确性. 相反，如果 $\overline{x}$ 事实上真的落在 μ_0 附近的一定的范围内，没有使小概率事件发生，也就是说 $\overline{x}$ 偏离 μ_0 可以认为是由于样本随机性造成的合理偏离，我们便没有从样本信息中找到足够的证据来否定假设 $H_0:\mu=\mu_0$ 的正确性，于是不能拒绝假设 H_0.

上面的论述事实上提供了假设检验的一般思想，我们可以对照确定性论证中的反证法：如果条件 A 成立导出一个矛盾的或不可能成立的结论 B，那么 A 就是错误的. 统计上的假设检验不可能像反证法一样导出确定的结论，但论证的逻辑却惊人地相似，它依据的是前面提到的小概率原理：如果 H_0 成立导致一个小概率事件发生了，那么我们就拒绝假设 H_0.

根据上述说明，要实施检验，首先要确定小概率的大小，这一小概率在假设检验中称为检验的**显著性水平**，通常记作 α. 它是根据具体问题需要事先确定的一个很小的正数，比如 0.01，0.05，0.10 等. 在后面我们还将对 α 的选取作出进一步的说明.

其次，对给定的显著性水平 α，我们要确定一个由样本所描述的概率不超过显著性水平 α 的小概率事件，这一小概率事件对应的样本取值区域通常称为假设检验的**拒绝域**. 一旦样本观察值落入拒绝域，便拒绝零假设，否则，不能拒绝零

假设.为确定拒绝域,需要构造一个含待检验参数、分布已知的枢轴量,通过枢轴量的已知分布并结合零假设确定拒绝域.例如,在上面的例子中,选取枢轴量

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1),$$

在 H_0 下,有统计量

$$T_0=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1).$$

于是可构造小概率事件

$$A=\left\{\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|>t_{\frac{\alpha}{2}}(n-1)\right\},$$

由 $T_0\sim t(n-1)$,知 $P(A)=\alpha$,记

$$C=\left\{(x_1,x_2,\cdots,x_n):\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|>t_{\frac{\alpha}{2}}(n-1)\right\},$$

则 $A=\{(X_1,X_2,\cdots,X_n)\in C\}$,$C$ 即为检验水平 α 的拒绝域.

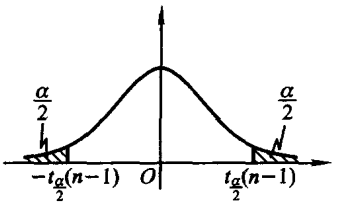


图 5.4　小概率事件 A

三、假设检验的一般步骤

根据前面的讨论和例子,假设检验的一般步骤可归纳如下:

1. 建立零假设 H_0;

2. 构造一个含待检参数 θ(不含其他未知参数!)且分布已知的枢轴量 $u(X_1,X_2,\cdots,X_n;\theta)$,并确定其分布;

3. 对给定的显著性水平 α,由上述枢轴量及其分布,结合零假设 H_0,确定拒绝域 C,使得

$$P\{(X_1,X_2,\cdots,X_n)\in C\mid H_0\}\leqslant\alpha;\tag{5.6}$$

4. 根据样本值 $(x_1,x_2,\cdots,x_n)$ 是否落在 C 中作出是否拒绝 H_0 的统计决断:如果 $(x_1,x_2,\cdots,x_n)\in C$,则拒绝 H_0,如果 $(x_1,x_2,\cdots,x_n)\notin C$,则不能拒绝 H_0.

例 5.23　在 $\alpha=0.05$ 的显著性水平下检验例 5.21 中有毒物质的浓度有无显著的变化.

解　① 建立零假设

$$H_0:\mu=18.2;$$

② 构造枢轴量

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(8);$$

③ 在 H_0 假设下

$$T_0=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(8)$$

从而 $$P\{|T_0|>t_{0.025}(8)\}=P\left\{\left|\frac{\bar{X}-\mu_0}{S/\sqrt{9}}\right|>t_{0.025}(8)\right\}=0.05,$$

其中 $t_{0.025}(8)=2.306$.

于是拒绝域为： $C=\left\{(x_1,x_2,\cdots,x_n):\left|\frac{\bar{x}-18.2}{s/\sqrt{9}}\right|>2.306\right\}$；

④ 由观察到的样本值 $(x_1,x_2,\cdots,x_n)$ 满足

$$|t_0|=\left|\frac{\bar{x}-18.2}{s/\sqrt{9}}\right|=\left|\frac{17.4-18.2}{2.4/\sqrt{9}}\right|=1.00<2.306.$$

知样本值不在拒绝域中，不能拒绝 $H_0:\mu=18.2$，即不能认为有毒物质浓度有明显变化.

四、检验的显著性水平与两类错误

一般说来，统计推断的特点是由样本提供的信息来推断总体. 这样，推断时所下的结论未必总是正确的. 就假设检验问题而言，如零假设 H_0 事实上是成立的，但却因样本观察值落入拒绝域而拒绝了零假设 H_0，这便犯了弃真错误，通常称为**第一类错误**；相反，如零假设 H_0 不成立，却因样本的观测值落入接受域，而接受了零假设 H_0 时，便犯了纳伪错误，通常称为**第二类错误**. 根据检验法则和(5.6)式，我们可以断言，当零假设 H_0 成立时，拒绝零假设的概率至多为显著性水平 α，这表明犯第一类错误的概率至多为 α，从而说明检验的显著性水平是用以控制犯第一类错误的概率的. 由此可能会产生一种错觉，以为只要把显著性水平 α 取得越小，假设检验的准确程度就越高. 事实上不然，因为显著性水平只是用来控制犯第一类错误的概率，而在假设检验中还存在着犯第二类错误的可能性. 一般来说，当样本容量给定时，在降低显著性水平的同时，往往会增大犯第二类错误的可能性. 若记犯第二类错误的概率为 β，那么 $1-\beta$ 表示不犯第二类错误的概率，通常称之为检验的**功效**，因而增大犯第二类错误的概率也意味着降低检验的功效.

通常的做法是事先给定显著性水平 α 来限制犯第一类错误的概率，再通过选取好的检验方法尽可能减少犯第二类错误的概率.

为避免事先人为地确定显著性水平 α，许多统计软件同时使用 p 值的概念.

一个假设检验的 p 值是指一个给定的样本观察拒绝该零假设的最小的显著性水平. 根据这一定义,对任意给定显著性水平 α,只要 α 大于 p 值,就在该显著性水平下拒绝零假设,否则不能在显著性水平 α 下拒绝零假设. 下面以例 5.23 为例说明 p 值使用的原理和方法.

例 5.23′　分别在 $\alpha=0.01,0.05,0.10$ 的显著性水平下检验例 5.21 中有毒物质的浓度有无显著的变化.

解　① 建立零假设

$$H_0:\mu=18.2;$$

② 构造枢轴量

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(8);$$

③ 在 H_0 假设下

$$T_0=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(8);$$

④ 计算统计量 T_0 的样本值

$$t_0=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=-1.00.$$

$$|t_0|=1.00;$$

⑤ 计算 p 值

$$p=P\{|T_0|>t_0|\}=P\{|T_0|>1\}=2P\{T_0>1\}\approx 0.36;$$

⑥ 比较 α 值与 p 值并作结论

由于 $\alpha=0.01,0.05,0.10$ 均比 p 值小,因而在三个显著性水平下均不能拒绝零假设,即不能认为有毒物质的浓度有明显的变化.

上面例子中,实际上只要给定的显著性水平比 0.36 小,那么在相应的显著性水平就不能拒绝零假设. 事实上,容易看出利用 p 值检验上述零假设的依据是

$$p>\alpha\Leftrightarrow|t_0|>t_{\frac{\alpha}{2}}(8).$$

五、多参数与非参数假设检验问题

以上我们对单参数假设检验问题作了较详尽的介绍. 原则上说来,以上介绍的所有内容也适用于多参数假设检验或非参数假设检验问题,只需在若干细节上作适当调整即可. 为此,仅说明两点:

1. 对于多参数假设检验问题,可寻求一个包含所有待检验参数的枢轴量,并使之服从或渐近地服从一个已知的确定分布;

2. 非参数假设检验问题可近似地化为一个多参数假设检问题来解决.

鉴于正态总体是统计应用中最常见的总体,我们将首先在接下来的两节分别讨论单正态总体与双正态总体的参数假设检验. 然后讨论伯努利总体及一般的有限离散总体分布的假设检验,这些总体虽然是在未知总体分布情形下对分布进行统计推断,但它们都可归结为参数检验问题,其中有限离散总体分布的假设检验属于多参数检验问题. 在本章最后一节,我们将讨论两类非参数假设检验问题:拟合优度的 χ^2 检验和独立性的检验.

§5.5 单正态总体的参数假设检验

设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,\cdots,X_n)$是总体 X 的容量为 n 的样本,记 $\bar{X}$ 与 S^2 分别为样本均值与样本方差,我们考虑对于均值 μ 与方差 σ^2 的参数假设检验.

一、均值 μ 的检验

通常对于参数 μ 可提出下面的假设检验问题

$$H_0:\mu=\mu_0\longleftrightarrow H_1:\mu\neq\mu_0,\qquad (A)$$

$$H_0:\mu\leqslant\mu_0\longleftrightarrow H_1:\mu>\mu_0,\qquad (B)$$

$$H_0:\mu\geqslant\mu_0\longleftrightarrow H_1:\mu<\mu_0,\qquad (C)$$

其中 μ_0 为指定的常数.

在(A)的备择假设 H_1 中,参数 μ 可取在 μ_0 的两侧,即 $\mu>\mu_0$ 或 $\mu<\mu_0$,这样的假设检验问题称为**双侧假设检验**问题;而(B)与(C)的备择假设 H_1 中,参数 μ 只可取在 μ_0 的一侧,这样的假设检验问题称为**单侧假设检验**问题.

针对参数 σ^2 是否已知的两种情形,可构造相应的检验法.

1. 方差 σ^2 已知的情形

设 $\sigma^2=\sigma_0^2$ 已知,由定理 4.2,可知枢轴量

$$U=\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1),\qquad (5.7)$$

构造相应的统计量

$$U_0=\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}.$$

对于(A)的检验,当 H_0 成立时,$U_0\sim N(0,1)$,统计量 U_0 不应偏大或偏小,拒绝域可取为

$$C=\{(x_1,\cdots,x_n):|u_0|>u_{\frac{\alpha}{2}}\},$$

其中 $u_{\frac{\alpha}{2}}$ 为水平 $\frac{\alpha}{2}$ 的 $N(0,1)$ 分布上侧分位数.

此时　　$P(|U_0|>u_{\frac{\alpha}{2}})=\alpha$.

对于(B)的检验,当 H_0 成立时,由于

$$U_0=\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\leqslant\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}=U,$$

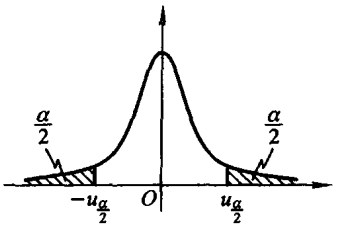


图 5.5　(A)的拒绝域

统计量 U_0 有偏小的倾向,而当 H_1 成立时,U_0 有偏大的趋势,故当 U_0 偏大时拒绝 H_0,拒绝域可取为

$$C=\{(x_1,\cdots,x_n):u_0>u_\alpha\},$$

其中 u_α 为水平为 α 的 $N(0,1)$ 分布的上侧分位数.此时

$$P(U_0>u_\alpha)\leqslant P(U>u_\alpha)=\alpha$$

类似于(B)的方法,容易得到(C)的拒绝域为

$$C=\{(x_1,\cdots,x_n):u_0<-u_\alpha\}.$$

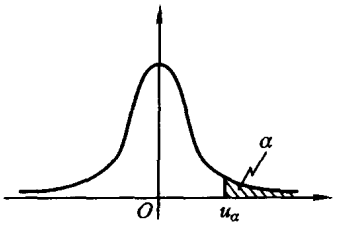


图 5.6　(B)的拒绝域

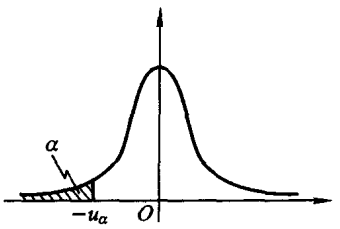


图 5.7　(C)的拒绝域

上述检验问题都运用了 ***U* 检验法**,所谓 ***U* 检验法**是指借助的枢轴量服从或渐进服从标准正态分布的检验法.进一步地,根据拒绝域单侧与双侧的不同,还可细分检验法为单侧检验法与双侧检验法. 这里(A)的检验法是双侧的,(B)与(C)是单侧的.

例 5.24　已知某厂生产灯管的寿命 X(单位:h)服从正态分布 $N(\mu,40\ 000)$,根据经验,灯管的平均寿命不超过 1 500 h,现测试了 25 只采用新工艺生产的灯管的寿命,测得其平均值为 1 575 h,试问新工艺是否提高灯管的寿命?(显著性水平 $\alpha=0.05$)

解　根据问题的特点,建立统计假设

$$H_0:\mu\leqslant 1\ 500\longleftrightarrow H_1:\mu>1\ 500.$$

由题意可知,$\mu_0=1\ 500$,$\sigma_0=200$,$n=25$,$\bar{x}=1\ 575$,给定 $\alpha=0.05$,查附表 $u_\alpha=1.645$,计算可得

$$u_0=\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}=\frac{1\,575-1\,500}{200/\sqrt{25}}=1.875.$$

由于 $u_0>u_\alpha$,故拒绝 H_0接受 H_1,即认为新工艺提高了灯管的寿命.

2. 方差 σ^2未知的情形

由定理 4.2,可知枢轴量

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1), \tag{5.8}$$

构造相应的统计量

$$T_0=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}.$$

类似于方差 σ^2 已知的情形,容易得到相应于(A),(B),(C)的拒绝域分别为

$$C=\{(x_1,\cdots,x_n):|t_0|>t_{\frac{\alpha}{2}}(n-1)\},$$

$$C=\{(x_1,\cdots,x_n):t_0>t_\alpha(n-1)\},$$

$$C=\{(x_1,\cdots,x_n):t_0<-t_\alpha(n-1)\},$$

其中 $t_{\frac{\alpha}{2}}(n-1),t_\alpha(n-1)$分别为水平$\frac{\alpha}{2}$与 α 的$t(n-1)$分布的上侧分位数.

上述检验问题都运用了 **t 检验法**——借助的枢轴量服从或渐近服从 t 分布的检验法. 同 σ^2已知的情形一样,这里(A)运用了双侧检验法. (B)与(C)运用了单侧检验法.

二、方差 σ^2的检验

关于参数 σ^2可类似提出双侧假设检验问题:

$$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\neq\sigma_0^2,$$

与单侧假设检验问题:

$$H_0:\sigma^2\leqslant\sigma_0^2\longleftrightarrow H_1:\sigma^2>\sigma_0^2,$$

$$H_0:\sigma^2\geqslant\sigma_0^2\longleftrightarrow H_1:\sigma^2<\sigma_0^2,$$

其中 $\sigma_0>0$ 为指定的正数.

这里介绍 μ 未知情形下的检验,$\mu=\mu_0$已知的情形留给读者完成.

由定理 4.2,可知枢轴量

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1), \tag{5.9}$$

构造相应的统计量

$$\chi_0^2=\frac{(n-1)S^2}{\sigma_0^2}.$$

对于双侧检验,当 H_0成立时,$\chi_0^2\sim\chi^2(n-1)$,统计量 χ_0^2不应偏大或偏小,

拒绝域可取为

$$C=\{(x_1,\cdots,x_n):\chi_0^2<\chi_{1-\frac{\alpha}{2}}^2(n-1)\text{或 }\chi_0^2>\chi_{\frac{\alpha}{2}}^2(n-1)\},$$

其中 $\chi_{1-\frac{\alpha}{2}}^2(n-1)$，$\chi_{\frac{\alpha}{2}}^2(n-1)$分别为水平 $1-\frac{\alpha}{2}$与$\frac{\alpha}{2}$的 $\chi^2(n-1)$分布的上侧分位数. 此时

$$P(\chi_0^2<\chi_{1-\frac{\alpha}{2}}^2(n-1))=P(\chi_0^2>\chi_{\frac{\alpha}{2}}^2(n-1))=\frac{\alpha}{2}.$$

类似于均值的检验，容易得到单侧检验的拒绝域分别为

$$C=\{(x_1,\cdots,x_n):\chi_0^2>\chi_\alpha^2(n-1)\},$$

$$C=\{(x_1,\cdots,x_n):\chi_0^2<\chi_{1-\alpha}^2(n-1)\},$$

其中 $\chi_\alpha^2(n-1)$，$\chi_{1-\alpha}^2(n-1)$分别为水平 α 与 $1-\alpha$ 的 $\chi^2(n-1)$分布的上侧分位数.

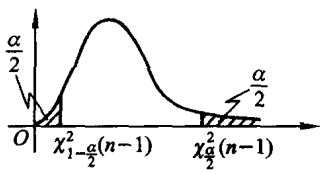


图 5.8　双侧检验的拒绝域

这里关于方差 σ^2的检验，都运用了 **χ^2检验法**，即所借助的枢轴量服从 χ^2分布.

例 5.25　假设某炼铁厂铁水中含碳量 X（单位：kg）服从正态分布 $N(\mu,0.112^2)$，现对工艺进行了改进，从中抽取了 7 炉铁水，测得含碳量数据如下：

$$4.421,\ 4.052,\ 4.357,\ 4.394,\ 4.326,\ 4.287,\ 4.683.$$

试问：新工艺炼出的铁水含碳量的方差是否有明显改变？（显著性水平 $\alpha=0.05$）

解　建立统计假设

$$H_0:\sigma^2=0.112^2\longleftrightarrow H_1:\sigma^2\neq 0.112^2.$$

由题意可知，$\sigma_0^2=0.112^2$，$n=7$，给定 $\alpha=0.05$，查附表 $\chi_{0.975}^2(6)=1.237$，$\chi_{0.025}^2(6)=14.449$，由抽样数据计算可得

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=4.36,$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=0.035\,1,$$

$$\chi_0^2=\frac{(n-1)s^2}{\sigma_0^2}=\frac{6\times 0.035\,1}{0.112^2}=16.789.$$

由于 $\chi_0^2>\chi_{0.025}^2(6)$，故拒绝 H_0接受 H_1，即认为铁水含碳量的方差明显变化.

有时根据实际问题，对于均值 μ 与方差 σ^2都要进行检验.

例 5.26　某工厂用自动生产线生产金属丝，假定金属丝的折断力 X（单位：N）服从正态分布，其合格标准为：平均值为 580 N，方差不超过 64. 某日开工后，

抽取 9 根作折断检测，测得结果如下：

578, 572, 570, 568, 572, 570, 596, 586, 568.

试问：此日自动生产线是否工作正常？（显著性水平 $\alpha=0.05$）

解 建立统计假设

$$H_0:\mu=580 \longleftrightarrow H_1:\mu\neq 580.$$

这里 $\mu_0=580$，$n=9$，给定 $\alpha=0.05$，查附表 $t_{0.025}(8)=2.306$，由抽样数据计算可得

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\approx 575.56,$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\approx 86.02,$$

$$t_0=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=\frac{575.56-580}{\sqrt{86.02}/\sqrt{9}}\approx -1.436.$$

由于 $|t_0|<t_{0.025}(8)$，故不能拒绝 H_0，即认为自动生产线没有系统误差.

进一步建立统计假设

$$H_0':\sigma^2\leqslant 64 \longleftrightarrow H_1':\sigma^2>64.$$

这里 $\sigma_0^2=64$，给定 $\alpha=0.05$，查附表 $\chi^2_{0.05}(8)=15.507$，计算可得

$$\chi_0^2=\frac{(n-1)s^2}{\sigma_0^2}=\frac{8\times 86.02}{64}\approx 10.753.$$

由于 $\chi_0^2<\chi^2_{0.05}(8)$，故不能拒绝 H_0'，即认为自动生产线工作稳定.

综上所述，可以认为自动生产线工作正常.

为了使用方便，我们将单正态总体的参数假设检验总结于附录 5.2，供读者参考.

§5.6 双正态总体的参数假设检验

设 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且 X 与 Y 相互独立，$(X_1,\cdots,X_{n_1})$，$(Y_1,\cdots,Y_{n_2})$分别是总体 X 与 Y 容量为 n_1，n_2 的样本，记 $\bar{X}$，$\bar{Y}$ 分别为总体 X 与 Y 的样本均值，S_1^2，S_2^2 分别为样本方差，我们将两个总体的比较转化为均值和方差的差异性检验.

一、均值的差异性检验

通常对于均值的比较，可提出下面的假设检验问题：

$$H_0:\mu_1=\mu_2\longleftrightarrow H_1:\mu_1\neq\mu_2;$$
$$H_0:\mu_1\leqslant\mu_2\longleftrightarrow H_1:\mu_1>\mu_2;$$
$$H_0:\mu_1\geqslant\mu_2\longleftrightarrow H_1:\mu_1<\mu_2.$$

若记 $\mu=\mu_1-\mu_2$，显然这些问题可相应地转化为 μ 的假设检验问题：

$$H_0:\mu=0\longleftrightarrow H_1:\mu\neq0;$$
$$H_0:\mu\leqslant0\longleftrightarrow H_1:\mu>0;$$
$$H_0:\mu\geqslant0\longleftrightarrow H_1:\mu<0.$$

类似于上节的方法，针对方差 σ_1^2,σ_2^2 的不同情形，我们简单给出相应的检验法.

1. σ_1^2 与 σ_2^2 均已知的情形

由定理 4.3，可知枢轴量

$$U=\frac{(\overline{X}-\overline{Y})-\mu}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1),\tag{5.10}$$

构造相应的统计量(此时 $\mu_0=0$)

$$U_0=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}.$$

上述检验问题的拒绝域分别为：

$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):|u_0|>u_{\frac{\alpha}{2}}\right\};$$
$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):u_0>u_\alpha\right\};$$
$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):u_0<-u_\alpha\right\}.$$

2. $\sigma_1^2=\sigma_2^2=\sigma^2$ 但未知的情形

由定理 4.3，可知枢轴量

$$T=\frac{(\overline{X}-\overline{Y})-\mu}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2),\tag{5.11}$$

其中 $S^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ 为样本方差 S_1^2 与 S_2^2 的加权平均，它恰是 σ^2 的无偏估计，构造相应的统计量

$$T_0=\frac{\overline{X}-\overline{Y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}.$$

上述检验问题的拒绝域分别为：

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):|t_0|>t_{\frac{\alpha}{2}}(n_1+n_2-2)\};$$

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):t_0>t_\alpha(n_1+n_2-2)\};$$

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):t_0<-t_\alpha(n_1+n_2-2)\}.$$

*3. $\sigma_1^2\neq\sigma_2^2$且未知的情形

可以证明(证略),枢轴量

$$T=\frac{(\bar{X}-\bar{Y})-\mu}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim t(f),\tag{5.12}$$

其中修正的自由度 $f=\dfrac{\left(\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}\right)^2}{\dfrac{1}{n_1}\left(\dfrac{S_1^2}{n_1}\right)^2+\dfrac{1}{n_2}\left(\dfrac{S_2^2}{n_2}\right)^2}$,这里需要指出的是,$t$ 分布的自由度实际上可以是大于 0 的任意实数.

构造相应的统计量 $T_0=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$,拒绝域分别为:

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):|t_0|>t_{\frac{\alpha}{2}}(f)\};$$

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):t_0>t_\alpha(f)\};$$

$$C=\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):t_0<-t_\alpha(f)\}.$$

例 5.27 研究学生的身高与性别的关系,现从某学院学生中抽查了 30 名男生与 24 名女生,测得身高的平均值分别为 172.5 cm,164.2 cm,假定男生、女生的身高服从标准差为 5.4 cm、4.8 cm 的正态分布,试问抽样结果是否可以支持"男生比女生高"这一观点?(显著性水平 $\alpha=0.05$)

解 记 X, Y 分别为男生与女生的身高,由题设,$X\sim N(\mu_1,5.4^2)$, $Y\sim N(\mu_2,4.8^2)$. 且记 $\mu=\mu_1-\mu_2$,建立统计假设

$$H_0:\mu\leqslant 0\longleftrightarrow H_1:\mu>0.$$

显然,此假设等价于均值比较假设 $H_0:\mu_1\leqslant\mu_2\longleftrightarrow H_1:\mu_1>\mu_2$.

这里 $\bar{x}=172.5$, $\bar{y}=164.2$, $n_1=30$, $n_2=24$, $\sigma_1^2=5.4^2$, $\sigma_2^2=4.8^2$,给定 $\alpha=0.05$,查附表 $u_\alpha=1.645$,计算可得

$$u_0=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{172.5-164.2}{\sqrt{\frac{5.4^2}{30}+\frac{4.8^2}{24}}}\approx 5.971.$$

由于 $u_0>u_\alpha$,故拒绝 H_0接受 H_1,即认为抽样结果可以支持题中观点.

例 5.28 某纺织厂生产纱线,为提高纱线强力,试用了新原料. 现在分别测

试了采用新、旧原料生产的纱线的强力(单位:kg),结果如下:

旧原料：　1.42, 1.49, 1.46, 1.34, 1.38, 1.54, 1.38, 1.51

新原料：　1.55, 1.47, 1.52, 1.60, 1.43, 1.53, 1.54.

假定工艺的生产精度相同,试问:抽样结果是否可以支持工厂采用新原料?(显著性水平 $\alpha=0.05$).

解　记 X, Y 分别为旧、新原料生产的纱线的强力,由题设可知, $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$,且 $\sigma_1^2=\sigma_2^2$.

建立统计假设(记 $\mu=\mu_1-\mu_2$)

$$H_0:\mu\geqslant 0\longleftrightarrow H_1:\mu<0.$$

这里 $n_1=8$, $n_2=7$,给定 $\alpha=0.05$,查附表 $t_{0.05}(13)=1.771$,由抽样数据计算可得

$$\bar{x}=1.44,\ s_1^2=0.005\ 1,\ \bar{y}=1.52,\ s_2^2=0.003\ 1,$$

$$s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}=0.004\ 173,\ s=0.064\ 6,$$

$$t_0=\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{1.44-1.52}{0.006\ 46\cdot\sqrt{\frac{1}{7}+\frac{1}{8}}}=-2.395.$$

由于 $t_0<-t_\alpha$,故拒绝 H_0 接受 H_1,即认为采用新原料可以提高纱线的强力.

与上例类似,许多实际问题中,两总体的方差均是未知的,进行均值的比较检验时,通常需要根据经验作出方差等与不等的假定. 实际上,我们可以利用 F 检验法对总体方差进行比较检验.

二、方差的差异性检验

记 $r=\dfrac{\sigma_1^2}{\sigma_2^2}$,类似地我们将方差 σ_1^2, σ_2^2 的比较检验问题

$$H_0:\sigma_1^2=\sigma_2^2\longleftrightarrow H_1:\sigma_1^2\neq\sigma_2^2;$$

$$H_0:\sigma_1^2\leqslant\sigma_2^2\longleftrightarrow H_1:\sigma_1^2>\sigma_2^2;$$

$$H_0:\sigma_1^2\geqslant\sigma_2^2\longleftrightarrow H_1:\sigma_1^2<\sigma_2^2$$

相应转化为 r 的下述检验问题

$$H_0:r=1\longleftrightarrow H_1:r\neq 1;$$

$$H_0:r\leqslant 1\longleftrightarrow H_1:r>1;$$

$$H_0:r\geqslant 1\longleftrightarrow H_1:r<1.$$

下面介绍 μ_1 与 μ_2 未知情形下的检验,μ_1 与 μ_2 为已知的情形留给读者完

成.

由定理 4.3,可知枢轴量

$$F=\frac{1}{r}\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1). \tag{5.13}$$

相应地构造统计量 $F_0=\dfrac{S_1^2}{S_2^2}$,运用上一节的方法,上述检验问题的拒绝域分别为

$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):F_0<F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)\right.$$
$$\left.\text{或 } F_0>F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\right\};$$
$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):F_0>F_{\alpha}(n_1-1,n_2-1)\right\};$$
$$C=\left\{(x_1,\cdots,x_{n_1};y_1,\cdots,y_{n_2}):F_0<F_{1-\alpha}(n_1-1,n_2-1)\right\}.$$

这里的假设检验运用了 **F 检验法**,即所借助的枢轴量服从 F 分布.

例 5.28 中曾经假定工艺的生产精度相同,即 $\sigma_1^2=\sigma_2^2$,这一假定是否合理呢? 下面利用 F 检验法进行检验,经计算

$$F_0=\frac{s_1^2}{s_2^2}=\frac{0.005\,1}{0.003\,1}\approx 1.645.$$

假定显著性水平 $\alpha=0.05$,查附表 $F_{0.025}(7,6)=5.70$,$F_{0.975}(7,6)=\dfrac{1}{F_{0.025}(6,7)}$ $=0.195$,由于 $F_{0.975}(7,6)<F_0<F_{0.025}(7,6)$,故不能拒绝方差相等的假定,即可以认为工艺的生产精度是相同的.

通常实际问题会遇到大量的抽样数据,我们可以借助各种统计软件如 SPSS,SAS 等来处理繁琐的计算,从而使统计方法更为有效实用.

我们将双正态总体的参数假设检验总结于附录 5.3,供读者参考.

§5.7 一般总体的参数假设检验

前面介绍的正态总体均值和方差的假设检验是最常用且相对简单的假设检验,本节我们将介绍伯努利总体与有限离散总体的假设检验,这两类检验本质上属于参数假设检验问题,而且后者属于多参数假设检验问题。对一般总体的参数假设检验问题,精确的检验往往较为复杂,我们在这里只介绍大样本情形下总体均值的假设检验问题.

一、伯努利总体的参数假设检验

设总体 X 服从以 $p(0<p<1)$为参数的伯努利分布,即有

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

伯努利总体在实际中应用广泛,参数 p 可视为实际总体中具有某一统计特征的个体所占的比率. 实际问题通常需要考虑参数 p 的下述假设检验问题:

$$H_0: p=p_0, \qquad (A)$$

$$H_0: p\leqslant p_0, \qquad (B)$$

$$H_0: p\geqslant p_0, \qquad (C)$$

其中 p_0 为指定的正数,$0<p_0<1$.

设 $(X_1,\cdots,X_n)$ 是总体 X 的容量为 n 的样本,$(x_1,\cdots,x_n)$ 为样本值. 记 $N=\sum_{i=1}^{n}X_i$,显然 N 是样本中诸分量取值为 1 的频数,容易证明:$N\sim b(n,p)$,即有

$$P(N=k)=b(k;n,p)=C_n^k p^k(1-p)^{n-k},\ k=0,1,\cdots,n. \qquad (5.14)$$

对于(A)的检验,当 H_0 成立时,$N\sim b(n,p_0)$,统计量 N 不应偏大或偏小,拒绝域可取为

$$C=\{(x_1,\cdots,x_n): N\leqslant k_1 \text{ 或 } N\geqslant k_2\},$$

其中正整数 k_1,k_2 满足

$$\sum_{k=0}^{k_1} b(k;n,p_0)\leqslant\frac{\alpha}{2}<\sum_{k=0}^{k_1+1} b(k;n,p_0),$$

$$\sum_{k=k_2}^{n} b(k;n,p_0)\leqslant\frac{\alpha}{2}<\sum_{k=k_2-1}^{n} b(k;n,p_0),$$

此时

$$P(N\leqslant k_1)\leqslant\frac{\alpha}{2},\ P(N\geqslant k_2)\leqslant\frac{\alpha}{2}.$$

对于(B)的检验,当 H_0 成立时,统计量 N 有偏小的趋势,而当 H_0 不成立时,统计量 N 有偏大的趋势,故拒绝域可取为

$$C=\{(x_1,\cdots,x_n): N\geqslant k_0\},$$

其中正整数 k_0 满足

$$\sum_{k=k_0}^{n} b(k;n,p_0)\leqslant\alpha<\sum_{k=k_0-1}^{n} b(k;n,p_0).$$

类似可得(C)的拒绝域为

$$C=\{(x_1,\cdots,x_n): N\leqslant k_0\},$$

其中正整数 k_0 满足

$$\sum_{k=0}^{k_0} b(k;n,p_0)\leqslant\alpha<\sum_{k=0}^{k_0+1} b(k;n,p_0).$$

上述检验法称为伯努利分布的**二项式检验法**.

例 5.29 某药厂宣称其生产的某种补钙剂的有效率超过 90%，现对 20 名缺钙患者进行试验，发现有 4 人无明显疗效，试问试验结果能否认定药厂的宣称不实？（显著性水平 $\alpha=0.10$）

解 建立统计假设

$$H_0: p \geqslant 90\%.$$

由题意，$n=20, N=16$，给定 $\alpha=0.10$，由于

$$\sum_{k=0}^{15} b(k;20,0.9)=0.046\ 8<\alpha<\sum_{k=0}^{16} b(k;20,0.9)=0.136\ 1,$$

故 $k_0=15$. 由于 $N>k_0$，故接受 H_0，即认为试验结果不能认定药厂的宣称不实.

在大样本情况下，由于二项式检验法的查表计算太过于复杂麻烦，我们可以利用 U 检验法来进行检验.

由定理 4.4，当 n 充分大时，枢轴量

$$U=\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}=\frac{N-np}{\sqrt{np(1-p)}} \tag{5.15}$$

渐近服从 $N(0,1)$，这里 $EX=p, DX=p(1-p)$.

相应地构造统计量 $U_0=\dfrac{N-np_0}{\sqrt{np_0(1-p_0)}}$，容易得到上述问题的拒绝域分别为：

$$C=\{(x_1,\cdots,x_n): |u_0|>u_{\frac{\alpha}{2}}\},$$

$$C=\{(x_1,\cdots,x_n): u_0>u_\alpha\},$$

$$C=\{(x_1,\cdots,x_n): u_0<-u_\alpha\},$$

其中 $u_{\frac{\alpha}{2}}, u_\alpha$ 分别为水平 $\dfrac{\alpha}{2}$ 与 α 的 $N(0,1)$ 的上侧分位数.

我们来考虑例 5.20 中提出的假设检验问题

$$H_0: p \leqslant 4\%$$

这里 $n=60, N=3$，若给定 $\alpha=0.05$，查附表 $u_\alpha=1.64$，计算可得

$$u_0=\frac{N-np_0}{\sqrt{np_0(1-p_0)}}=\frac{3-60\times 4\%}{\sqrt{60\times 4\%\times(1-4\%)}}=0.395\ 3.$$

由于 $u_0<u_\alpha$，故接受 H_0，即认为产品能够出厂.

二、有限离散总体分布的检验

设总体 X 是仅取 $r(r>2)$ 个可能值的离散型随机变量，不失一般性，设其 r 个可能取值为 $1,2,\cdots,r$，并设

$$P(X=i)=p_i,\ i=1,2,\cdots,r.$$

考虑多参数假设检验问题

$$H_0: p_i=p_{i0},\ i=1,2,\cdots,r,$$

这里 $p_{i0}(i=1,2,\cdots,r)$是 r 个指定的正数，$\sum_{i=1}^{r} p_{i0}=1$.

设$(X_1,\cdots,X_n)$是总体 X 的容量为 n 的样本，记 $N_i(i=1,2,\cdots,r)$为样本中诸分量取值为 i 的个数，即事件$(X=i)$出现的频数. 显然 N_i 均是基于样本的统计量且 $\sum_{i=1}^{r} N_i=n$，可以证明，$(N_1,\cdots,N_r)$服从多项分布

$$P(N_1=n_1,\cdots,N_r=n_r)=\frac{n!}{n_1!\cdots n_r!}p_1^{n_1}\cdots p_r^{n_r},\quad \sum_{i=1}^{r} n_i=n.$$

皮尔逊(K. Pearson)证明了(证略)下述结论：

定理 5.1　(皮尔逊定理)　当 $n\to\infty$时，枢轴量

$$\chi^2=\sum_{i=1}^{r}\frac{(N_i-np_i)^2}{np_i} \tag{5.16}$$

渐近服从 $\chi^2(r-1)$分布.

上述枢轴量并不是严格意义下的统计量，但我们约定俗成称为皮尔逊统计量.

皮尔逊定理解决了上述问题的检验. 当 n 充分大时，构造相应的统计量

$$\chi_0^2=\sum_{i=1}^{r}\frac{(N_i-np_{i0})^2}{np_{i0}}.$$

直观上很明显，若 H_0 成立，此时$(p_{10},\cdots,p_{r0})$即为总体的真分布，频率$\frac{N_i}{n}$与概率 p_{i0} 应充分地接近，相应地观察频数 N_i 与理论频数 np_{i0} 的差异愈小，统计量 χ_0^2 有偏小的趋势，否则它有偏大的趋势，因此拒绝域可取为

$$C=\{(x_1,\cdots,x_n):\chi_0^2>\chi_\alpha^2(r-1)\},$$

此时，$P_{H_0}(\chi_0^2>\chi_\alpha^2(r-1))\approx\alpha$.

上述检验法也称为多项分布的 χ^2 检验法.

例 5.30　一家工厂分早、中、晚三班，每班 8 小时，近期发生了一些事故. 在近期记录的 30 次事故中，有 11 次发生在早班，6 次发生在中班，13 次发生在晚班，从而怀疑班次不同与事故出现率有关. 试利用上述记录数据来判断这一猜测是否成立. (显著性水平 $\alpha=0.05$)

解　以 X 记事故发生的班次，$X=1,2,3$ 分别表示事故发生在早班、中班、晚班，并记 $p_i=P(X=i),i=1,2,3$.

若事故的出现与班次无关,则 $p_i=\frac{1}{3}$,$i=1,2,3$,建立统计假设

$$H_0: p_i=\frac{1}{3},\ i=1,2,3.$$

由题设,$n=30$,$r=3$,$n_1=11$,$n_2=6$,$n_3=13$,给定 $\alpha=0.05$,查附表 $\chi_\alpha^2(r-1)=5.991$,计算可得

$$\chi_0^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}=\frac{1}{10}((11-10)^2+(6-10)^2+(13-10)^2)=2.6.$$

由于 $\chi_0^2<\chi_\alpha^2(r-1)$,故接受 H_0,即认为事故的出现与班次无关.

从已记录的事故数据看,11:6:13 的比例似乎应当说明中班事故率要低于平均值$\frac{1}{3}$,但上述分析表明,有时直观的判断未必可靠.

三、一般总体均值的大样本假设检验

设 X 为任一总体,$EX=\mu$,一般地,总体方差 DX 通常是未知的,我们考虑 μ 的下述假设检验问题:

$$H_0: \mu=\mu_0;$$
$$H_0: \mu\leqslant\mu_0;$$
$$H_0: \mu\geqslant\mu_0.$$

在大样本情形下的检验,这里 μ_0 为指定的常数.

设 $(X_1,\cdots,X_n)$ 是总体 X 的容量为 n 的样本,$(x_1,\cdots,x_n)$ 为样本值,由定理 4.4,当 n 充分大时,枢轴量

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}} \tag{5.17}$$

渐近服从 $N(0,1)$ 分布.

相应地构造统计量 $T_0=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$,上述假设检验问题的拒绝域分别为:

$$C=\{(x_1,\cdots,x_n): |t_0|>u_{\frac{\alpha}{2}}\};$$
$$C=\{(x_1,\cdots,x_n): t_0>u_\alpha\};$$
$$C=\{(x_1,\cdots,x_n): t_0<-u_\alpha\}.$$

这里的检验都运用了 U 检验法.

例 5.31 例 5.19 中求出了 $\mu=EX$ 的置信水平为 99% 的近似置信区间.现在考虑假设检验问题:由试验结果能否判定此部件的平均连续使用寿命至少为 2 年?(显著性水平 $\alpha=0.01$)

解 以每年 365 天计算,一部件可连续使用 2 年,则使用的小时数至少应为

$2\times365\times24=17\ 520$,折合为 17.52 kh. 建立统计假设

$$H_0:\mu\leqslant17.52.$$

由题意,$n=100$,$\bar{x}=17.84$,$s=1.25$,$\mu_0=17.52$,给定 $\alpha=0.01$,查附表 $u_\alpha=2.33$,计算可得

$$t_0=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=\frac{17.84-17.52}{1.25/\sqrt{100}}=2.56.$$

由于 $t_0>u_\alpha$,故拒绝 H_0,即认为此部件至少可以连续使用两年.

§5.8 拟合优度 χ^2 检验与独立性检验

本节我们将介绍两类重要的非参数假设检验问题的检验.

一、拟合优度 χ^2 检验

考虑下述非参数假设检验问题

$$H_0:F(x)=F_0(x)\quad(F_0(x)\text{是某个已知的分布}).$$

这里限定 X 为连续型随机变量或取无限个可能值的离散型随机变量,其分布函数为 $F(x)$.

利用皮尔逊定理,容易构造上述问题的检验法.

事实上,我们可以将总体 X 有限离散化,选取 $r-1$ 个实数 $a_1<a_2<\cdots<a_{r-1}$,并依序将实轴划分为 r 个互不相交的子区间 $I_1=(-\infty,a_1]$,$I_2=(a_1,a_2]$,$\cdots$,$I_r=(a_{r-1},+\infty)$. 记

$$p_i=F(a_i)-F(a_{i-1}),\ p_{i0}=F_0(a_i)-F_0(a_{i-1}),\ i=1,2,\cdots,r.$$

这里约定 $a_0=-\infty$;$a_r=+\infty$.

以 Y 记 X 有限离散化的结果,$Y=i\Leftrightarrow X\in I_i$,$i=1,2,\cdots,r$,则 Y 是仅取 r 个可能值的离散型随机变量,且 $P(Y=i)=p_i$,$i=1,2,\cdots,r$.

上述问题可转化为关于 Y 的多参数假设检验问题

$$H_0:p_i=p_{i0},i=1,2,\cdots,r.$$

设 $(X_1,\cdots,X_n)$ 是总体 X 的容量为 n 的样本,$(x_1,\cdots,x_n)$ 为样本值,记 $N_i(i=1,2,\cdots,r)$ 为样本中诸分量落入区间 I_i 频数,即事件 $(Y=i)$ 出现的频数,由皮尔逊定理,统计量

$$\chi_0^2=\sum_{i=1}^{r}\frac{(N_i-np_{i0})^2}{np_{i0}}\tag{5.18}$$

当 H_0 成立时渐近服从 $\chi^2(r-1)$ 分布.

拒绝域类似可取为

$$C=\{(x_1,\cdots,x_n):\chi_0^2>\chi_\alpha^2(r-1)\}.$$

上述检验法称为总体分布的**拟合优度 χ^2 检验法**，它实际上是以一个已知的有限离散总体的分布去拟合一个未知的连续型或无限离散型总体的分布.

必须指出的是，拟合优度 χ^2 检验法有其理论上的缺陷. 它实际上只是检验了假设

$$H_0:F(a_i)=F_0(a_i),\ i=1,2,\cdots,r-1.$$

而并未真正检验总体的分布是否为 $F_0(x)$.

有时虽然 $F(x)\neq F_0(x)$，但在某一种划分下，可有

$$p_i=F(a_i)-F(a_{i-1})=F_0(a_i)-F_0(a_{i-1})=p_{i0},\ i=1,2,\cdots,r.$$

此时拒绝 H_0 是合理的，而接受 H_0 会增大犯第二类错误(纳伪错误)的可能性.

上述检验问题实际上有更为严谨有效的检验法，如 K－S 检验法等，有兴趣的读者可参阅相关文献.

例 5.32 某黑盒中装有偶数个球，分为黑、白两色. 现有放回地摸球，记录下直到摸到白球为止所需的摸球次数. 重复执行上述试验 100 次，结果如下：

抽取次数	1	2	3	4	≥5
频数	43	31	15	6	5

试由此判断黑盒中白球与黑球的个数是否相等？(显著性水平 $\alpha=0.05$)

解 记 X 为首次摸到白球时所需的摸球次数，则 X 服从几何分布

$$P(X=k)=(1-p)^{k-1}p,\ k=1,2,\cdots,$$

其中 p 为黑盒中白球所占的比例.

黑盒中白球与黑球个数相等，当且仅当 $p=\frac{1}{2}$，相应地可计算得

$$P(X=k)=\left(\frac{1}{2}\right)^k,\ k=1,2,3,4,\ P(X\geqslant 5)=1-\sum_{k=1}^{4}\left(\frac{1}{2}\right)^k=\left(\frac{1}{2}\right)^4.$$

可建立统计假设

$$H_0:p_1=\frac{1}{2},\ p_2=\frac{1}{4},\ p_3=\frac{1}{8},\ p_4=\frac{1}{16},\ p_5=\frac{1}{16}.$$

由题意，$n=100$，$r=5$，$n_1=43$，$n_2=31$，$n_3=15$，$n_4=6$，$n_5=5$，给定 $\alpha=0.05$，查附表 $\chi_\alpha^2(r-1)=9.488$，计算可得

$$\chi_0^2=\sum_{i=1}^{r}\frac{(n_i-np_{i0})^2}{np_{i0}}=3.2$$

由于 $\chi_0^2<\chi_\alpha^2(r-1)$，故不能拒绝 H_0，即可认为黑盒中白球与黑球个数相等.

一般地，待检验的分布类型中可能含有未知的参数，进一步考虑下述非参数假设检验问题

$$H_0: F(x) = F_0(x; \theta_1, \cdots, \theta_s),$$

其中 F_0 的具体形式已知，$\theta = (\theta_1, \cdots, \theta_s)$ 为未知参数.

我们可以对上面的检验法进行简单的修改. 设 $\hat{\theta}_1, \cdots, \hat{\theta}_s$ 为未知参数 $\theta_1, \cdots, \theta_s$ 的最大似然估计，相应地求出 p_{i0} 的估计量

$$\hat{p}_{i0} = F_0(a_i; \hat{\theta}) - F_0(a_{i-1}; \hat{\theta}) \quad i = 1, 2, \cdots, r.$$

费希尔(R. A. Fisher)证明了(证略)，当 H_0 成立且分布函数 $F_0(x; \theta)$ 满足一定的正则条件时，相应的统计量

$$\chi_0^2 = \sum_{i=1}^{r} \frac{(N_i - n\hat{p}_{i0})^2}{n\hat{p}_{i0}} \tag{5.19}$$

渐近服从 $\chi^2(r-s-1)$ 分布.

拒绝域可相应修改为

$$C = \{(x_1, \cdots, x_n): \chi_0^2 > \chi_\alpha^2(r-s-1)\}.$$

这里需要指出，实际计算时，应选取恰当的划分，一般地应保证区间的个数 $r > s+2$ 但不应太大且 $n_i \geqslant 5, i = 1, 2, \cdots, r$.

下面是例 5.22 中的具体数据(单位:g).

2 720, 3 660, 2 800, 3 520, 3 320, 3 600, 3 100, 2 620, 3 480, 3 300,
3 180, 3 360, 3 300, 3 660, 2 940, 3 260, 3 900, 3 660, 3 700, 2 100,
3 300, 3 900, 3 480, 3 000, 3 320, 4 000, 3 940, 3 140, 3 780, 3 180,
3 900, 3 660, 3 140, 3 500, 3 180, 3 680, 3 420, 3 260, 3 600, 2 880,
3 540, 2 700, 3 160, 3 100, 4 500, 2 980, 3 540, 2 700, 3 500, 3 840.

假定显著性水平 $\alpha = 0.05$，我们来检验假设

$$H_0: X \sim N(\mu, \sigma^2)$$

由例 5.7，μ, σ^2 的最大似然估计分别为 $\bar{X}, S_0^2$，计算可得

$$\hat{\mu} = \bar{x} = 3\,360, \quad \hat{\sigma}^2 = s_0^2 \approx 424.2^2$$

根据数据的特点，选取 5 个分点 $a_i = 2\,750 + (i-1) \times 250, i = 1, \cdots, 5$，将实轴划分为 6 个区间，相应的统计结果为

区间 I_i	$\leqslant 2\,750$	(2 750,3 000]	(3 000,3 250]	(3 250,3 500]	(3 500,3 750]	$>3\,750$
频数 n_i	5	5	8	13	11	8

利用下述两式

$$\Phi(a_i;\hat{\mu},\hat{\sigma}^2)=\Phi_0\left(\frac{a_i-\hat{\mu}}{\hat{\sigma}}\right),\ i=1,2,\cdots,r-1;$$

$$\hat{p}_{i0}=\Phi(a_i;\hat{\mu},\hat{\sigma}^2)-\Phi(a_{i-1};\hat{\mu},\hat{\sigma}^2),\ i=1,2,\cdots,r.$$

计算可得

$\hat{p}_{10}\approx0.075$, $\hat{p}_{20}\approx0.123$, $\hat{p}_{30}\approx0.199$, $\hat{p}_{40}\approx0.232$, $\hat{p}_{50}\approx0.192$, $\hat{p}_{60}\approx0.179$.

给定 $\alpha=0.05$,由于 $r=6,s=2$,查附表 $\chi^2_\alpha(r-s-1)=7.815$,计算可得

$$\chi_0^2=\sum_{i=1}^{r}\frac{(n_i-n\hat{p}_{i0})^2}{n\hat{p}_{i0}}\approx1.488.$$

由于 $\chi_0^2<\chi_\alpha^2(r-s-1)$,故不能拒绝 H_0,即认为男婴的体重服从正态分布.

二、独立性检验

考虑二元总体(X,Y),由于总体可以有限离散化,不妨假定 X 与 Y 的取值范围可分别分成 r 个和 q 个互不相交的子区间 $A_1,\cdots,A_r$ 和 $B_1,\cdots,B_q$. 下记

$$p_{ij}=P(X\in A_i,\ Y\in B_j),\ i=1,\cdots,r,\ j=1,\cdots,q.$$

$$p_{i\cdot}=P(X\in A_i),\ i=1,\cdots,r.$$

$$p_{\cdot j}=P(Y\in B_j),\ j=1,\cdots,q.$$

显然 $$p_{i\cdot}=\sum_{j=1}^{q}p_{ij},\ p_{\cdot j}=\sum_{i=1}^{r}p_{ij},\ i=1,\cdots,r,\ j=1,\cdots,q.$$

考虑下述非参数假设检验问题

H_0:X 与 Y 独立,

显然它可转化为多参数假设检验问题

$$H_0:p_{ij}=p_{i\cdot}\cdot p_{\cdot j},\ i=1,\cdots,r,\ j=1,\cdots,q.$$

设$((X_1,Y_1),\cdots,(X_n,Y_n))$是总体$(X,Y)$的容量为 n 的样本,记 N_{ij} $(i=1,\cdots,r,\ j=1,\cdots,q)$为样本中诸分量落入矩形区域 $A_i\times B_j$ 的频数. 且记

$$N_{i\cdot}=\sum_{j=1}^{q}N_{ij},\ N_{\cdot j}=\sum_{i=1}^{r}N_{ij},\ i=1,\cdots,r,\ j=1,\cdots,q,$$

显然 $$\sum_{i=1}^{r}N_{i\cdot}=\sum_{j=1}^{q}N_{\cdot j}=\sum_{i=1}^{r}\sum_{j=1}^{q}N_{ij}=n.$$

通常我们把上述统计量列成下面的联立表

		B_j				$\sum_{j=1}^{q}(N_{i\cdot})$
		1	2	$\cdots$	q	
A_i	1	N_{11}	N_{12}	$\cdots$	N_{1q}	N_1
	2	N_{21}	N_{22}	$\cdots$	N_{2q}	N_2
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	r	N_{r1}	N_{r2}	$\cdots$	N_{rq}	N_r.
$\sum_{i=1}^{r}(N_{\cdot j})$		$N_{\cdot1}$	$N_{\cdot2}$	$\cdots$	$N_{\cdot q}$	n

利用前面的方法，我们可以进行联立表的独立性检验.

首先，可以证明，参数 $p_{i\cdot}$ 与 $p_{\cdot j}$ 的最大似然估计为

$$\hat{p}_{i\cdot}=\frac{N_{i\cdot}}{n},\hat{p}_{\cdot j}=\frac{N_{\cdot j}}{n},\ i=1,\cdots,r,\ j=1,\cdots,q.$$

其次，由于 $\sum_{i=1}^{r}p_{i\cdot}=\sum_{j=1}^{q}p_{\cdot j}=1$，故 $r+q$ 个参数 $p_{i\cdot}$ 与 $p_{\cdot j}$ 中仅有 $r+q-2$ 个独立参数. 于是相应的统计量

$$\chi_0^2=n\sum_{i=1}^{r}\sum_{j=1}^{q}\frac{\left(N_{ij}-\frac{N_{i\cdot}\cdot N_{\cdot j}}{n}\right)^2}{N_{i\cdot}\cdot N_{\cdot j}} \tag{5.20}$$

渐近服从 $\chi^2(rq-(r+q-2)-1)$（即 $\chi^2((r-1)(q-1))$）分布.

拒绝域相应为

$$C=\{((x_1,y_1),\cdots,(x_n,y_n)):\chi_0^2>\chi_\alpha^2((r-1)(q-1))\}.$$

上述检验通常称为**联立表的独立性检验**，它在实际中应用广泛.

例 5.33　研究青少年行为与家庭状况的关系，调查结果如下：

青少年行为	家庭状况		合　计
	离异家庭	和睦家庭	
犯　罪	178	272	450
未犯罪	38	502	540
合　计	216	774	990

试问：调查结果能否证实青少年的行为受到家庭状况的影响？（显著性水平 $\alpha=0.05$）

解　以 X 记青少年的行为，Y 记家庭的状况，建立统计假设

$$H_0:X\text{ 与 }Y\text{ 独立}.$$

由题设，$n=990$，$n_{11}=178$，$n_{12}=272$，$n_{21}=38$，$n_{22}=502$，给定 $\alpha=0.05$，由于 $r=q=2$，查附表 $\chi_\alpha^2((r-1)(q-1))=3.841$，计算可得

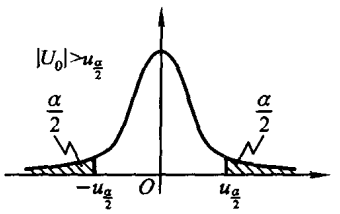
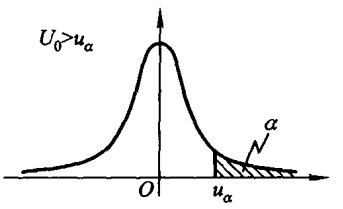
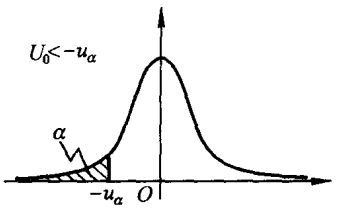
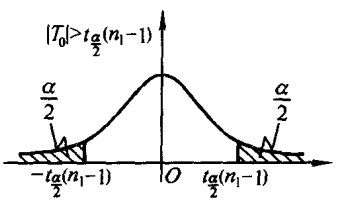
$$\chi_0^2=n\sum_{i=1}^{r}\sum_{j=1}^{q}\frac{\left(n_{ij}-\frac{n_{i\cdot}\cdot n_{\cdot j}}{n}\right)^2}{n_{i\cdot}\cdot n_{\cdot j}}\approx 152.153.$$

由于 $\chi_0^2>\chi_\alpha^2((r-1)(q-1))$，故拒绝 H_0，即认为青少年的行为受到家庭状况的影响，家庭的破裂容易引起青少年的犯罪.

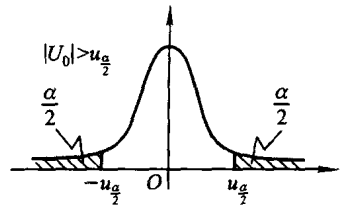
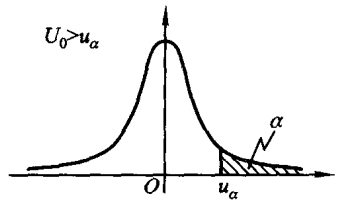
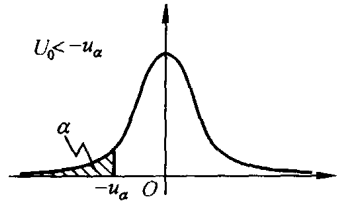
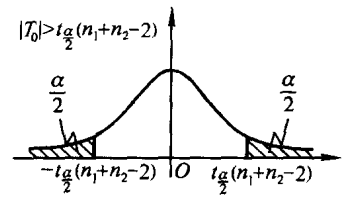
附录 5.1　正态总体参数的置信区间一览表

待估参数	条件	枢轴量	置信下、上限
均值 μ	σ^2 已知	$U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\bar{X}-u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ $\bar{X}+u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
	σ^2 未知	$T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\bar{X}-t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}$ $\bar{X}+t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}$
方差 σ^2	μ 未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}$ $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}$
均值差 $\mu_1-\mu_2$	σ_1^2,σ_2^2 已知	$U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$(\bar{X}-\bar{Y})-u_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$ $(\bar{X}-\bar{Y})+u_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$
	$\sigma_1^2=\sigma_2^2$ 但未知	$T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(f)$ $f=n_1+n_2-2$	$(\bar{X}-\bar{Y})-t_{\alpha/2}(f)S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ $(\bar{X}-\bar{Y})+t_{\alpha/2}(f)S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$
方差比 σ_1^2/σ_2^2	μ_1,μ_2 均未知	$F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\dfrac{1}{F_{\alpha/2}(n_1-1,n_2-1)}\dfrac{S_1^2}{S_2^2}$ $\dfrac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\dfrac{S_1^2}{S_2^2}$

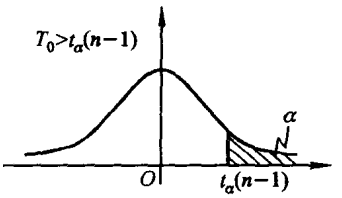
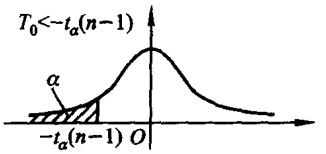
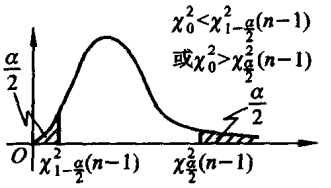
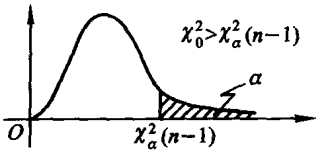
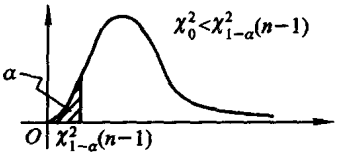
附录 5.2　单正态总体参数假设检验一览表

零假设	条件	枢轴量与统计量	拒绝域
$H_0:\mu=\mu_0$	$\sigma^2=\sigma_0^2$ 已知	$U=\dfrac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1)$ $U_0=\dfrac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}$	
$H_0:\mu\leqslant\mu_0$			
$H_0:\mu\geqslant\mu_0$			
$H_0:\mu=\mu_0$	σ^2 未知	$T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ $T_0=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$	

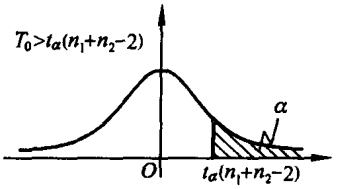
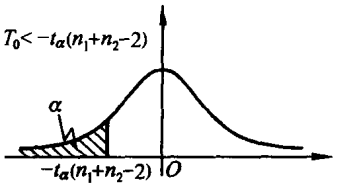
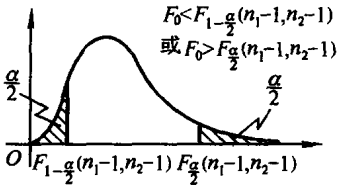
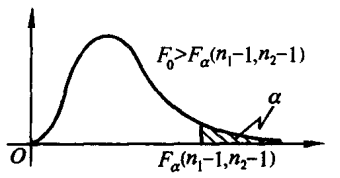
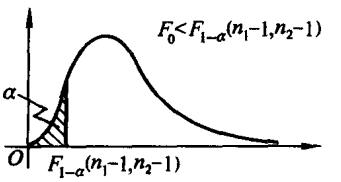
附录 5.3　双正态总体参数假设检验一览表

零假设	条件	枢轴量与统计量	拒绝域
$H_0:\mu_1=\mu_2$	σ_1^2,σ_2^2 已知	$U=\dfrac{(\bar{X}-\bar{Y})-\mu}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$ $U_0=\dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	
$H_0:\mu_1\leqslant\mu_2$			
$H_0:\mu_1\leqslant\mu_2$			
$H_0:\mu_1=\mu_2$	$\sigma_1^2=\sigma_2^2$ 但未知	$T=\dfrac{(\bar{X}-\bar{Y})-\mu}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ $\sim t(n_1+n_2-2)$	

续表

零假设	条件	枢轴量与统计量	拒绝域
$H_0:\mu\leqslant\mu_0$	μ 未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ $\chi_0^2=\dfrac{(n-1)S^2}{\sigma_0^2}$	
$H_0:\mu\geqslant\mu_0$			
$H_0:\sigma^2=\sigma_0^2$			
$H_0:\sigma^2\leqslant\sigma_0^2$			
$H_0:\sigma^2\geqslant\sigma_0^2$			

续表

零假设	条件	枢轴量与统计量	拒绝域
$H_0:\mu_1\leqslant\mu_2$	μ_1,μ_2 未知	$T_0=\dfrac{\bar{X}-\bar{Y}}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	
$H_0:\mu_1\geqslant\mu_2$			
$H_0:\sigma_1^2=\sigma_2^2$		$F=\left(\dfrac{\sigma_2}{\sigma_1}\right)^2\dfrac{S_1^2}{S_2^2}$ $\sim F(n_1-1,n_2-1)$ $F_0=\dfrac{S_1^2}{S_2^2}$	
$H_0:\sigma_1^2\leqslant\sigma_2^2$			
$H_0:\sigma_1^2\geqslant\sigma_2^2$			

习　题　五

(A)

1. 设$(X_1,\cdots,X_n)$是取自总体 X 的样本,试求下列总体分布中未知参数的矩估计与最大似然估计.

(1) X 的密度函数为

$$f(x;\alpha)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1,\\ 0, & \text{其他},\end{cases}$$

其中 $\alpha>-1$ 为未知参数;

(2) X 服从参数为 p 的两点分布,$0<p<1$ 为未知参数;

(3) X 的密度函数为

$$f(x;\lambda,\alpha)=\begin{cases}\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & \text{其他},\end{cases}$$

其中 α 已知,$\lambda>0$ 为未知参数;

(4) X 服从参数为 λ 的泊松分布,$\lambda>0$ 为未知参数.

2. 设$(X_1,\cdots,X_n)$是取自于总体 X 的样本,记 $\mu=EX,\sigma^2=DX$,试证对任意常数 $a_i>0,\sum\limits_{i=1}^{n}a_i=1,i=1,2,\cdots,n,\hat{\mu}=\sum\limits_{i=1}^{n}a_iX_i$ 均是 μ 的无偏估计,其中 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 最为有效.

3. 设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,\cdots,X_n)$是取自于总体 X 的样本,试选择适当的常数 c,使 $c\sum\limits_{i=1}^{n-1}(X_{i+1}-X_i)^2$ 为 σ^2 的无偏估计.

4. 假设某市每月死于交通事故的人数 X 服从参数为 λ 的泊松分布,$\lambda>0$ 为未知参数,现有以下样本值

$$3,\ 2,\ 0,\ 5,\ 4,\ 3,\ 1,\ 0,\ 7,\ 2,\ 0,\ 2.$$

试求月无死亡的概率的最大似然估计值.

5. 设总体 $X\sim N(\mu,9)$,$(X_1,\cdots,X_n)$是取自于总体 X 的样本,欲使 μ 的 $1-\alpha$ 置信区间的长度 L 不超过 2,问在以下两种情况下样本容量 n 至少应取多少?

(1) $\alpha=0.1$;　　(2) $\alpha=0.01$.

6. 假定某毛纺厂生产的羊毛锭的含脂率(%)服从正态分布 $N(\mu,0.8^2)$,现进行了 10 次抽样检测,结果如下

$$7.7,\ 8.2,\ 7.5,\ 9.4,\ 7.1,\ 6.8,\ 8.8,\ 7.3,\ 6.4,\ 7.8.$$

试由样本结果求 μ 的置信水平为 95% 的置信区间.

7. 设按某种工艺生产的金属纤维的长度 $X\sim N(\mu,\sigma^2)$,现抽取了 15 根纤维测得平均长度为 $\bar{x}=5.4$,样本方差为 $s^2=0.16$. 试求 μ 与 σ 的 95% 置信区间.

8. 某市四家大型电器商场的手机销售情况抽样表明,在一周内总计2 154部的手机销量

中,某品牌手机占 568 部,试求此品牌手机的销售比例的 95% 置信区间.

9. 某超市为了增加销售,对营销方式、管理人员等进行了一系列调整,调整后随机抽查了 9 天的日销售额(单位:万元),结果如下

56.4, 54.2, 50.6, 53.7, 48.3, 55.9, 57.4, 58.7, 55.3.

根据统计,调整前的日平均销售额为 51.2 万元,假定日销售额服从正态分布,试问调整措施的效果是否显著?(显著性水平 $\alpha=0.05$)

10. 市质检局接到投诉后,对某金商进行质量调查. 现从其出售的标志 18K 的项链中抽取 9 件进行检测,检测标准为:标准值 18K 且标准差不得超过 0.3K,检测结果如下:

17.3, 16.6, 17.9, 18.2, 17.4, 16.3, 18.5, 17.2, 18.1.

假定项链的含金量服从正态分布,试问检测结果能否认定金商出售的产品存在质量问题?(显著性水平 $\alpha=0.01$)

11. 在相同的条件下对甲、乙两种品牌的洗涤剂分别进行去污试验,测得去污率(%)结果如下:

品牌甲:79.4, 80.5, 76.2, 82.7, 77.8, 75.6;

品牌乙:73.4, 77.5, 79.3, 75.1, 74.7.

假定两品牌的去污率均服从正态分布且方差相同,试问两品牌的去污率有无明显差异?(显著性水平 $\alpha=0.05$)

12. 设某零件厂生产的零件的直径服从正态分布,为了减小方差,提高生产精度,工厂试用了新工艺. 现分别抽样了采用新、旧工艺生产零件的直径(单位:mm),结果如下:

旧工艺:20.0, 19.3, 19.8, 20.0, 20.5, 19.5, 20.4, 20.6, 19.9;

新工艺:20.0, 19.7, 20.3, 20.2, 20.0, 19.7, 19.9, 19.8.

试问抽样结果是否可以支持工厂采用新工艺?(显著性水平 $\alpha=0.05$)

13. 中药厂从某种药材中提取有效成分,为了提高提取率,改革了提炼方法,现对同一质量的药材,分别抽测了采用新、旧方法的提取率(%),结果如下:

旧方法:75.4, 77.3, 76.2, 78.1, 74.3, 73.2, 77.4, 78.8, 76.9;

新方法:77.3, 78.7, 79.1, 81.0, 80.5, 79.8, 82.1, 78.3, 79.6.

假定新、旧方法的提取率均服从正态分布,试问新方法的提取率是否有明显提高?(显著性水平 $\alpha=0.05$)

14. 某市对成年人查体,在随机抽查的 100 人中发现了 59 人患有不同程度的牙疾,试问抽查结果能否说明该市 50% 以上的成年人患有牙疾?(显著性水平 $\alpha=0.05$)

15. 在公路上,每隔 15 秒记录下路过的汽车数,共进行了 200 次观察,得到频数分布如下:

路过汽车数	0	1	2	3	4	≥5
频数	92	68	28	11	1	0

试检验路过的汽车数是否服从泊松分布. (显著性水平 $\alpha=0.05$)

16. 随机抽取 200 只某种电子元件进行寿命试验,测得元件的寿命(单位:h)的频数分布如下:

元件寿命	≤200	(200,300]	(300,400]	(400,500]	>500
频数	94	25	22	17	42

根据计算,平均寿命为325 h,试检验寿命是否服从指数分布?(显著性水平 $\alpha=0.10$)

17. 为了解吸烟习惯与患慢性气管炎的关系,对339名50岁以上的人作了调查. 调查结果如下:

	患慢性气管炎者	未患慢性气管炎者	合　计	患病率
吸　烟	43	162	205	21%
不吸烟	13	121	134	9.7%
合　计	56	283	339	16.5%

试判断吸烟者与不吸烟者慢性气管炎患病率是否不同?(显著性水平 $\alpha=0.01$)

18. 近年来人们对人的智力和出生季节之间的关系感兴趣,下面是随机抽查148人得到的统计结果:

IQ范围 \ 出生季节	春	夏	秋	冬	合　计
55～69IQ	18	29	19	12	78
40～54IQ	20	13	17	20	70
合　计	38	42	36	32	148

其中IQ为智力单位,通常认为人的智力小于70IQ,试判断人的智力高低是否和出生季节有关?(显著性水平 $\alpha=0.05$)

(B)

1. 设总体 X 服从对数正态分布,密度函数为

$$f(x;\mu,\sigma^2)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right\}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

其中 $-\infty<\mu<+\infty$,$\sigma^2>0$ 为未知参数,$(X_1,\cdots,X_n)$是取自于总体 X 的样本,试求 μ 与 σ^2 的矩估计与最大似然估计.

2. 设总体 $X\sim U[a,b]$,其中 $b>a$ 为未知参数,$(X_1,\cdots,X_n)$是取自于总体 X 的样本,试求 a 和 b 的矩估计与最大似然估计.

3. 设$(X_1,\cdots,X_n)$是取自于总体 X 的样本,$EX=\mu$,$DX=\sigma^2$,$\mu\neq 0$,$\sigma^2>0$,试证 $\hat{\mu}=\dfrac{2}{n(n+1)}\sum\limits_{k=1}^{n}kX_k$ 是 μ 的无偏估计与相合估计.

4. 试证在等方差双正态总体的均值差异性检验(见§5.6节)中,对于任意常数 $a>0$, $b>0, a+b=1, \hat{\sigma}^2 = aS_1^2 + bS_2^2$ 均是 σ^2 的无偏估计,其中 $S^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ 的方差最小.

5. 设总体 $X \sim N(\mu, \sigma^2)$,其中 $\mu = \mu_0$ 已知,σ^2 为未知参数,$(X_1, \cdots, X_n)$ 是取自于总体 X 的样本,试对方差 σ^2 的双侧与单侧假设检验问题,构造一个有别于(5.7)式的枢轴量并进行假设检验.

6. 设总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ 且相互独立,μ_1 与 μ_2 均已知,σ_1^2 与 σ_2^2 为未知参数,$(X_1, \cdots, X_{n_1})$, $(Y_1, \cdots, Y_{n_2})$ 分别是 X 与 Y 的样本,试对 $r = \dfrac{\sigma_1^2}{\sigma_2^2}$ 的检验问题,构造一个有别于(5.11)式的枢轴量并进行假设检验.

7. 保险学中一般假定非寿险保单的持有者在保单固定的有效期内,因事故要求索赔的次数服从参数为 λ 的泊松分布,保险公司为合理地厘定保险费,需要对未知参数 λ 作统计推断. 下面是随机抽取的同一非寿险险种的 4 000 张保单的索赔情况,假定每张保单具有相同的有效期.

索赔次数	0	1	2	3
保单数	3 288	642	66	4

(1) 试利用上述抽样结果求 λ 的最大似然估计.

(2) 运用大样本的统计推断方法检验下述假设. ($\alpha = 0.10$)

$$H_0: \lambda = 0.2 \longleftrightarrow H_1: \lambda \neq 0.2$$

8. 下面是某种新型传染病的一个抽样统计结果:

病情 \ 年龄	老年	中青年	青少年	合 计
重 症	64	76	8	148
一 般	21	376	19	416
合 计	85	452	27	564

其中相应死亡人数为 22,14,3,给定显著性水平 $\alpha = 0.05$.

(1) 试利用 χ^2 检验法研究病情是否与年龄有关.

(2) 试利用 U 检验法判断"老年人病死率偏高"这一结论的正确性.

9. 试证在联立表的独立性检验(见§5.7节)中,各参数 $p_{i\cdot}$ 与 $p_{\cdot j}$ 的最大似然估计为

$$\hat{p}_{i\cdot} = \frac{N_{i\cdot}}{n},\ \hat{p}_{\cdot j} = \frac{N_{\cdot j}}{n},\ i = 1, \cdots, r,\ j = 1, \cdots, q.$$

第 6 章

方差分析

方差分析是统计推断中鉴别因素效应的一种有效实用的统计方法，它的主要思想和理论是由费希尔(R. A. Fisher)在 1926 年研究农业试验时首先提出来的，随后迅速发展完善. 方差分析不仅在实际工作中应用广泛，而且在理论上也非常重要，数理统计的许多分支，如回归分析、相关分析等都借鉴了它的思想和方法.

本章将主要介绍单因素方差分析和双因素方差分析的模型结构、思想和方法.

§6.1 方差分析概述

实际工作中，通常我们关心的变量或指标会受到许多因素的影响，我们需要找出那些对变量或指标有显著影响的因素，有时还需要确定它们的合理搭配，使这些因素的影响效果最好. 例如农业生产中，农作物的产量受到作物的品种、施用肥料的种类及数量、耕种方式等因素的影响，我们需要考虑它们之中哪些因素的影响是显著的，进一步地，如何选择它们的合理搭配，以提高农作物的产量. 又如市场营销中，商品的销量受到商品的价格、包装样式、广告宣传形式等因素的影响. 同样地，我们需要找出有显著影响的因素，并确定它们的合理搭配，制定出较优的市场营销策略等等. 实际中所遇到的这类问题即为方差分析问题，解决这类问题的统计方法即为方差分析.

在方差分析中，若仅仅考虑一个因素的影响，这时的方差分析称为**单因素方差分析**；若需要同时考虑多个因素的影响，这时的方差分析称为**多因素方差分析**. 特别地，若同时仅考虑两个因素的影响，则称之为**双因素方差分析**. 鉴于两个以上的多因素方差分析较为复杂，其处理的思想与方法与双因素方差分析类似，本书将仅介绍单因素方差分析与双因素方差分析.

我们先来看一个单因素方差分析问题，这是一个早期的农业试验的例子.

例 6.1 比较研究某种农作物的四个不同品种对产量的影响. 选取 16 块大小相同、肥沃程度相近的土地，每个品种选种四块，采用相同的耕种方式，测得产

量结果如下：

品　种	产　量				平均值
A_1	202	215	225	218	215.00
A_2	237	215	205	226	220.75
A_3	340	325	315	334	328.50
A_4	250	267	242	254	253.25

我们希望由此来推断作物的品种对产量的影响是否显著，如影响显著，那么哪一个品种的作物产量最高？

这里可以看到，不仅不同品种的作物产量是不同的，而且即使是同一品种其产量也不尽相同. 在这一试验中，造成作物产量差异的主要原因有两个方面：一是作物品种的不同，这是一个可控制的试验条件；二是自然条件、管理差异等随机因素的变化. 在方差分析中，通常把可控制的试验条件称为**因素**，因素的不同状态称为**水平**. 本例中作物的品种即是一个因素，它有四个水平. 我们要确定品种对产量的影响是否显著，即是要判断作物产量的差异主要是由随机因素造成的，还是由品种因素的不同水平造成的.

上例中我们只需考虑品种这一个因素对产量的影响，这是一个单因素方差分析问题. 进一步地，如果同时考虑农作物的四个不同品种和肥料的五个不同种类对产量的影响，此时的问题即是一个双因素方差分析问题. 这里影响农作物产量有两个因素：一个是作物的品种；另一个是施用肥料的种类. 同样地，我们需要确定这两个因素对农作物的产量影响是否显著，也就是说，作物的不同品种与肥料的不同种类是否分别造成了作物产量的显著差异. 由于存在两个因素的影响，可能还需要考虑它们之间的交互作用. 实际中经常会出现这样的情况：分别使产量达到最高的作物品种及肥料种类搭配在一起时，并不一定会使作物的产量进一步提高，可能反而会降低. 而并非最好的作物品种与肥料种类其合理搭配却可能获得最高的作物产量. 也就是说，作物的不同品种与肥料的不同种类对产量的联合影响并不一定正好等于它们各自影响的简单相加. 在方差分析中，我们把这种因素间不同水平的搭配所产生的新的影响称为**交互作用**. 在单因素方差分析中没有交互作用问题，但在双因素方差分析中有时必须考虑交互作用问题. 从根本上说，交互作用的存在正是双因素方差分析与单因素方差分析的本质区别所在.

一般地，在双因素方差分析中，如果不考虑两个因素间的交互作用，或者说假定无交互作用存在，通常只需对两个因素的各种水平搭配进行一次独立试验. 如果需要考虑两个因素间的交互作用，就必须对它们的相互搭配重复进行独立

试验,只有这样才能判断出因素间的交互作用是否显著.

下面给出两个双因素方差分析的例子. 例 6.2 不考虑交互作用,例 6.3 则需要考虑交互作用.

例 6.2 考虑三种不同形式的广告与五种不同的价格对某种商品销量的影响. 我们选取某市 15 家大超市,每家超市选用其中的一种组合,统计出一个月的销量如下:

广告 \ 价格	B_1	B_2	B_3	B_4	B_5
A_1	276	352	178	295	273
A_2	114	176	102	155	128
A_3	364	547	288	392	378

希望由上述统计结果判断:

(1) 不同的广告形式下商品的销量是否有显著差异;

(2) 不同的价格下商品的销量的差异是否显著.

例 6.3 比较研究三种温度和四种催化剂对某种工业产品质量的影响. 现对温度和催化剂的不同水平搭配各重复三次独立试验,测得某主要质量指标的结果如下:

温度 \ 催化剂	B_1	B_2	B_3	B_4
A_1	85,84,84	92,90,91	83,84,82	74,76,72
A_2	78,75,76	82,84,85	79,74,77	69,72,71
A_3	92,94,95	94,95,94	90,90,91	84,80,83

希望由上述试验数据推断温度、催化剂,以及它们的交互作用对产品的质量指标是否有显著影响.

§6.2 单因素方差分析

一、模型的结构

我们进一步来分析例 6.1 中的问题. 若假定四个不同品种作物的产量为四

个不同的总体，那么每一品种作物的产量可以认为是取自于同一总体的样本，它们之间的差异可以认为是随机因素造成的，而不同品种作物产量的差异是由于作物品种以及随机因素造成的，于是判断品种对作物产量的影响是否显著，可归结为判断四个不同的总体是否具有相同分布.

一般地，设单个因素 A 具有 r 个水平，分别记为 $A_1, A_2, \cdots, A_r$，对应于 r 个总体 $X_1, X_2, \cdots, X_r$. 在方差分析中，通常假定 $X_i \sim N(\mu_i, \sigma^2)(i=1,2,\cdots,r)$ 且相互独立. 这一假定是十分合理的，一方面实际中经常遇到的是具有正态分布的总体，而且对于因素的各个水平我们总是分别进行独立试验；另一方面为了突出因素的影响，我们总是尽量把其他试验条件做到一致，也就是说，每个总体的方差是相等的. 于是因素 A 的影响是否显著，进一步可归结为几个等方差的正态总体均值是否相等，也即下述假设检验问题：

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_r. \tag{6.1}$$

很明显，若 H_0 成立，则 r 个总体服从相同的正态分布，变量取值的差异性只是由于随机因素造成的，因素 A 的影响就是不显著的；否则，因素 A 的各个水平间就存在系统差异，因素 A 的影响就是显著的.

由上述分析，我们容易看出，单因素方差分析问题实际上是双正态总体均值比较在多正态总体上的推广.

设 $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ 是取自于 X_i 的容量为 n_i 的样本，$i=1,2,\cdots,r$，记 $n = \sum_{i=1}^{r} n_i$，我们称

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ij}$$

分别为**组平均值**与**总平均值**. 通常将样本结果列成下表：

单因素方差分析数据结构

A 的水平	抽样结果				平均
A_1	X_{11}	X_{12}	$\cdots$	X_{1n_1}	$\bar{X}_1$
A_2	X_{21}	X_{22}	$\cdots$	X_{2n_2}	$\bar{X}_2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	X_{r1}	X_{r2}	$\cdots$	X_{rn_r}	$\bar{X}_r$

由前面的假定，$X_{ij} \sim N(\mu_i, \sigma^2)$，若以 ε_{ij} 表示因素 A 的第 i 个水平下第 j 个样本分量 X_{ij} 的随机误差，则有

$$X_{ij} = \mu_i + \varepsilon_{ij}, \ i=1,2,\cdots,r, \ j=1,2,\cdots,n_i,$$

其中 $\varepsilon_{ij} \sim N(0, \sigma^2)$ 且相互独立.

记 $\mu=\frac{1}{n}\sum_{i=1}^{r}n_i\mu_i$，$\alpha_i=\mu_i-\mu$，$i=1,2,\cdots,r$，则有

$$\mu_i=\mu+\alpha_i,\quad \sum_{i=1}^{r}n_i\alpha_i=0.$$

显然若 H_0 成立，则 $\mu_i=\mu$，于是 $\alpha_i=0$，$i=1,2,\cdots,r$；若 H_0 不成立，则必存在 $1\leqslant i_0\leqslant r$，使得 $\alpha_{i_0}\neq 0$，这里 α_i 实际上体现了水平 A_i 对变量取值的特殊影响，我们称之为**水平 A_i 的效应**. 由此，我们可以将单因素方差分析模型归结为

$$\begin{cases}X_{ij}=\mu+\alpha_i+\varepsilon_{ij},\\ \varepsilon_{ij}\sim N(0,\sigma^2)\text{且相互独立},\\ \sum_{i=1}^{r}n_i\alpha_i=0.\end{cases}\quad i=1,2,\cdots,r,\ j=1,2,\cdots,n_i, \tag{6.2}$$

在这一模型下，(6.1)式可进一步明确为

$$H_0:\alpha_i=0,i=1,2,\cdots,r\longleftrightarrow H_1:\text{存在 }1\leqslant i_0\leqslant r,\alpha_{i_0}\neq 0. \tag{6.3}$$

这里我们实际上将不同总体的均值比较问题转化为各水平的效应鉴别问题，也就是说，我们需要判断因素 A 的所有水平的效应是否均不存在. 接下来的问题是：如何对(6.3)式的假设进行检验？

二、离差平方和分解与方差分析表

在方差分析中，通常利用离差平方和分解方法来进行假设检验. 考虑所有样本 X_{ij} 与总平均值 $\bar{X}$ 的离差平方和

$$Q=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2.$$

显然 Q 反映了所有样本 $X_{ij}(i=1,2,\cdots,r,\ j=1,2,\cdots,n_i)$ 取值的离散程度，我们称之为**总离差平方和**，它可以进行下面的分解

$$\begin{aligned}Q&=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i+\bar{X}_i-\bar{X})^2\\&=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2+\sum_{i=1}^{r}n_i(\bar{X}_i-\bar{X})^2+2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)(\bar{X}_i-\bar{X}).\end{aligned}$$

容易验证 $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)(\bar{X}_i-\bar{X})=0$，若记

$$Q_A=\sum_{i=1}^{r}n_i(\bar{X}_i-\bar{X})^2,\quad Q_e=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2,$$

则有

$$Q=Q_A+Q_e. \tag{6.4}$$

(6.4)式通常称为离差平方和分解式,其中 Q_A 是各组平均值与总平均值的离差平方和,它度量了组与组之间的系统性差异,我们称之为**组间平方和**或**系统平方和**;Q_e 是每一组内的样本与其组平均值的离差平方和,它是由于随机因素产生的离差平方和,我们称之为**组内平方和**或**误差平方和**.

现在我们来阐述方差分析的思想. 我们知道,变量取值的差异性可以归结为因素 A 与随机因素的影响. 我们以总离差平方和 Q 来度量所有样本取值的差异程度,Q_A,Q_e 则分别反映了因素 A 的不同水平与随机因素造成的变量取值差异性的程度. 根据离差平方和分解式,Q 恰好可以分解为 Q_A 与 Q_e 之和,显然 Q_A 与 Q_e 的相对大小可以反映出因素 A 的显著程度. 若 Q_A 显著地大于 Q_e,说明因素 A 的各水平差异是造成变量取值差异的主要原因,因素 A 的影响就是显著的,否则,因素 A 的影响就是不显著的.

根据上述思想,我们来寻求检验用的统计量. 可以证明,当 H_0 成立时,有

$$\frac{Q_A}{\sigma^2}\sim\chi^2(r-1),\frac{Q_e}{\sigma^2}\sim\chi^2(n-r),$$

且它们相互独立. 于是

$$F=\frac{Q_A/(r-1)}{Q_e/(n-r)}\sim F(r-1,n-r). \tag{6.5}$$

根据上面的分析,拒绝域可取为

$$C=\{(x_{11},\cdots,x_{1n_1};\cdots;x_{r1},\cdots,x_{rn_r}):F\geqslant F_\alpha(r-1,n-r)\},$$

其中 $F_\alpha(r-1,n-r)$ 是水平为 α 的 $F(r-1,n-r)$ 分布的上侧分位数.

在方差分析中,为了表达的直观清晰,通常使用下面的**方差分析表**:

单因素方差分析表

方差来源	平方和	自由度	均 方	F 值
因 素	$Q_A=\sum_{i=1}^{r}n_i(\bar{X}_i-\bar{X})^2$	$r-1$	$S_A=\frac{Q_A}{r-1}$	$\frac{S_A}{S_e}$
误 差	$Q_e=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2$	$n-r$	$S_e=\frac{Q_e}{n-r}$	
总 和	$Q=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2$	$n-1$		

下面利用方差分析来解决例 6.1 中的问题. 经过计算,可以得到下面的方差分析表:

方差来源	平方和	自由度	均方	F 值
品 种	32 707.25	3	10 902.42	85.255
误 差	1 534.50	12	127.88	
总 和	34 241.75	15		

假如给定 $\alpha=0.05$,查附表 $F_{0.05}(3,12)=3.49$,由于 $F=85.255>F_{0.05}(3,12)$,故拒绝 H_0,即认为作物的品种对产量的影响是显著的.

三、参数估计

上例中,因素 A 的各水平间存在系统差异,那么选择哪一个品种比较适合呢?这里我们需要考虑各水平效应的估计问题.

为此可求单因素方差分析模型(6.2)中,各参数 $\mu,\alpha_1,\cdots,\alpha_r,\sigma^2$ 的最大似然估计. 由 X_{ij} 的独立性,可知似然函数为

$$L(\mu,\alpha_1,\cdots,\alpha_r,\sigma^2)=\prod_{i=1}^{r}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(X_{ij}-\mu-\alpha_i)^2\right\}$$

$$=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\mu-\alpha_i)^2\right\}.$$

由此可解得(作为练习)$\hat{\mu}=\bar{X}$,$\hat{\alpha}_i=\bar{X}_i-\bar{X}$,$\hat{\sigma}^2=\dfrac{Q_e}{n}$ 为各自参数的最大似然估计,并且容易证明 $\hat{\mu}$,$\hat{\alpha}_i$ 分别是 μ,α_i 的无偏估计.

实际工作中,当因素的影响显著时,我们可以简单认为最大的 $\hat{\alpha}_i$(或 $\bar{x}_i$)相应的水平 A_i 的影响效果最大. 例 6.1 中,$\bar{x}_3=328.50$(或 $\hat{\alpha}_3=74.125$)为最大,我们认为品种 A_3 作物的产量最高,选择这一品种效果最好.

对于方差分析问题,我们也可以利用一些统计软件来处理其中大量复杂的计算,下面再来看一个单因素方差分析的例子.

例 6.4 将抗生素注入人体会产生抗生素与血浆蛋白质结合的现象,这种结合会降低药效. 下表给出了四种常用的抗生素注入到牛的体内时,抗生素与血浆蛋白质结合的百分比.

抗生素	结合百分比	平均值
A_1	29.6, 24.3, 28.5, 32.0, 28.6, 31.5, 25.7	28.60
A_2	27.3, 32.6, 30.8, 34.8, 31.4	31.38
A_3	21.6, 17.5, 18.3, 19.0, 23.4, 14.8	19.10
A_4	29.2, 32.8, 25.0, 24.2, 28.2, 27.4	27.80

试由上述试验数据推断不同抗生素的结合百分比是否有显著差异.(显著性水平 $\alpha=0.05$).

利用统计软件 SPSS 处理的结果如下:

Test of Homogeneity of Variances

Levene Statistic	df_1	df_2	*Sig*.
.063	3	20	0.979

ANOVA

Source	Sum of Squares	*df*	Mean Squares	F	*Sig*.
Between Groups	488.382	3	162.794	18.867	.000
Within Groups	172.568	20	8.628		
Total	660.950	23			

这里先对各总体等方差的假设进行检验,具体的检验法我们不予介绍. 由于 p 值 *Sig*. $=0.979>\alpha=0.05$,故可以认为各总体方差相等.

我们再来看方差分析的结果. 这里 ANOVA 是 Analysis of Variance 的缩写,即方差分析. 表中给出了平方和(Sum of Squares),自由度(df),均方(Mean Squares),F 值(F)及 p 值(*Sig*.). 由于 p 值 *Sig*. $=.000<\alpha=0.05$,故可以认为不同抗生素的结合百分比有显著差异.

由于 $\bar{x}_2=31.38$ 为最大,$\bar{x}_3=19.10$ 为最小,可以简单认为抗生素 A_2 的结合百分比最高,抗生素 A_3 的结合百分比最低.

§6.3 双因素方差分析

一、相关效应与统计假设

我们来考虑双因素方差分析. 设因素 A 有 r 个水平 $A_1,\cdots,A_r$,因素 B 有 s 个水平 $B_1,\cdots,B_s$,各水平的每一组合(A_i,B_j)相应于一个总体 X_{ij},我们假定 $X_{ij}\sim N(\mu_{ij},\sigma^2)$且相互独立($i=1,2,\cdots,r$, $j=1,2,\cdots,s$),显然这是对单因素方差分析的一个自然的推广.

我们记

$$\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij},$$

$$\mu_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\mu_{ij},i=1,2,\cdots,r,\ \mu_{\cdot j}=\frac{1}{r}\sum_{i=1}^{r}\mu_{ij},j=1,2,\cdots,s,$$

则有

$$\mu=\frac{1}{r}\sum_{i=1}^{r}\mu_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\mu_{\cdot j}.$$

进一步地,记

$$\alpha_i=\mu_{i\cdot}-\mu,\quad i=1,2,\cdots,r,\quad \beta_j=\mu_{\cdot j}-\mu,\quad j=1,2,\cdots,s,$$

则有

$$\sum_{i=1}^{r}\alpha_i=0,\ \sum_{j=1}^{s}\beta_j=0.$$

这里 μ 为各个总体均值 μ_{ij} 的总平均,而 $\mu_{i\cdot},\mu_{\cdot j}$ 分别是水平 A_i,B_j 下所有总体均值的平均,很明显,α_i 与 β_j 分别度量了水平 A_i,B_j 对变量取值的特殊影响,我们分别称为水平 A_i,B_j 的**效应**. 于是判断因素 A 的影响是否显著,即是要检验假设

$$H_{01}:\alpha_i=0,\ i=1,2,\cdots,r.$$

判断因素 B 的影响是否显著,即是要检验假设

$$H_{02}:\beta_j=0,j=1,2,\cdots,s.$$

那么如何度量因素 A,B 的交互作用呢? 我们记

$$\begin{aligned}\gamma_{ij}&=\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu\\&=(\mu_{ij}-\mu)-(\alpha_i+\beta_j)\quad i=1,2,\cdots,r,\quad j=1,2,\cdots,s,\end{aligned}$$

则有

$$\sum_{i=1}^{r}\gamma_{ij}=0,\quad j=1,2,\cdots,s,\quad \sum_{j=1}^{s}\gamma_{ij}=0,\quad i=1,2,\cdots,r.$$

这里 γ_{ij} 恰好等于组合(A_i,B_j)下水平 A_i 与 B_j 的联合效应$(\mu_{ij}-\mu)$与水平 A_i,B_j 各自效应之和$(\alpha_i+\beta_j)$的差,我们称为水平 A_i 与 B_j 的**交互效应**,显然它度量了水平 A_i 与 B_j 搭配而产生的交互作用的影响. 于是判断因素 A 与因素 B 交互作用的影响是否显著,即是要检验假设

$$H_{03}:\gamma_{ij}=0,\quad i=1,2,\cdots,r,\quad j=1,2,\cdots,s.$$

一般地,在双因素方差分析中,根据上述各效应的定义,可有

$$\mu_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij},\ i=1,2,\cdots,r,j=1,2,\cdots,s.$$

这里若事先假定所有 $\gamma_{ij}=0$,我们考虑的即为无交互作用的双因素方差分析,若事前不作此假定,则考虑的即为有交互作用的双因素方差分析.

为了说明得更为清晰,我们先考虑无交互作用的情况,再考虑有交互作用的情况.

二、无交互作用的情况

由于不考虑交互作用的影响，或者说假定无交互作用存在，我们对每一组合(A_i, B_j)只进行一次独立试验，通常称这种试验为**无重复试验**. 这里不妨仍以X_{ij}来标记样本，我们将样本结果列成下表：

无交互作用的双因素方差分析数据结构

B 的水平 / A 的水平	B_1	B_2	$\cdots$	B_s	平均
A_1	X_{11}	X_{12}	$\cdots$	X_{1s}	$\bar{X}_{1\cdot}$
A_2	X_{21}	X_{22}	$\cdots$	X_{2s}	$\bar{X}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	X_{r1}	X_{r2}	$\cdots$	X_{rs}	$\bar{X}_{r\cdot}$
平均	$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$	$\cdots$	$\bar{X}_{\cdot s}$	$\bar{X}$

其中

$$\bar{X} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s} X_{ij},$$

$$X_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s} X_{ij}, i = 1,2,\cdots,r, \quad X_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} X_{ij}, j = 1,2,\cdots,s.$$

若以ε_{ij}记X_{ij}的随机误差，我们可将无交互作用的双因素方差分析模型归结为

$$\begin{cases} X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \\ \varepsilon_{ij} \sim N(0,\sigma^2) \text{ 且相互独立}, \\ \sum_{i=1}^{r}\alpha_i = \sum_{j=1}^{s}\beta_j = 0. \end{cases} \quad i = 1,2,\cdots,r, \quad j = 1,2,\cdots,s, \tag{6.6}$$

在上述模型下，我们需要分别对下述假设进行检验

$$H_{01}: \alpha_i = 0, \quad i = 1,2,\cdots,r, \tag{6.7}$$

$$H_{02}: \beta_j = 0, \quad j = 1,2,\cdots,s. \tag{6.8}$$

我们同样利用离差平方和分解方法. 可以证明，总离差平方和Q可进行下面的分解

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X})^2$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}[(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}) + (\bar{X}_{i\cdot} - \bar{X}) + (\bar{X}_{\cdot j} - \bar{X})]^2$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2 + s\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{X})^2 + r\sum_{j=1}^{s}(\bar{X}_{\cdot j} - \bar{X})^2.$$

若记

$$Q_A = s\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{X})^2, \quad Q_B = r\sum_{j=1}^{s}(\bar{X}_{\cdot j} - \bar{X})^2,$$

$$Q_e = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2,$$

则有

$$Q = Q_A + Q_B + Q_e. \tag{6.9}$$

上式即为离差平方和分解式，其中 Q_A，Q_B 分别反映了因素 A 与因素 B 的不同水平造成的变量取值差异性的程度，我们分别称之为**因素 A 的离差平方和**、**因素 B 的离差平方和**；Q_e 则反映了随机因素造成的变量取值差异性的程度，我们称之为**误差平方和**.

类似于单因素方差分析，可以证明，当 H_{01} 成立时，有

$$\frac{Q_A}{\sigma^2} \sim \chi^2(r-1), \quad \frac{Q_e}{\sigma^2} \sim \chi^2((r-1)(s-1)),$$

且它们相互独立. 于是

$$F_A = \frac{Q_A/(r-1)}{Q_e/(r-1)(s-1)} \sim F(r-1,(r-1)(s-1)). \tag{6.10}$$

(6.7)的拒绝域可取为

$$C = \{(x_{11},\cdots,x_{rs}): F_A \geqslant F_\alpha(r-1,(r-1)(s-1))\}.$$

同样地，当 H_{02} 成立时，有

$$F_B = \frac{Q_B/(s-1)}{Q_e/(r-1)(s-1)} \sim F(s-1,(r-1)(s-1)). \tag{6.11}$$

(6.8)的拒绝域可取为

$$C = \{(x_{11},\cdots,x_{rs}): F_B \geqslant F_\alpha(s-1,(r-1)(s-1))\}.$$

与单因素方差分析相似，我们通常也使用下面的方差分析表：

无交互作用双因素方差分析表

方差来源	平方和	自由度	均方	F 值
因素 A	$Q_A = s\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{X})^2$	$r-1$	$S_A = \dfrac{Q_A}{r-1}$	$F_A = \dfrac{S_A}{S_e}$
因素 B	$Q_B = r\sum_{j=1}^{s}(\bar{X}_{\cdot j} - \bar{X})^2$	$s-1$	$S_B = \dfrac{Q_B}{s-1}$	$F_B = \dfrac{S_B}{S_e}$
误　差	$Q_e = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2$	$(r-1)(s-1)$	$S_e = \dfrac{Q_e}{(r-1)(s-1)}$	
总　和	$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X})^2$	$rs-1$		

另外，类似于单因素方差分析情形，可以证明水平 A_i, B_j 的效应 α_i, β_j 的最大似然估计分别为

$$\hat{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}, i = 1,2,\cdots,r, \quad \hat{\beta}_j = \bar{X}_{\cdot j} - \bar{X}, j = 1,2,\cdots,s.$$

下面我们来解决例 6.2 中的问题，假定显著性水平 $\alpha = 0.05$. 经过计算，可以得到下面的方差分析表：

方差来源	平方和	自由度	均方	F 值
广　告	167 804.13	2	83 902.07	63.09
价　格	44 568.40	4	11 142.10	8.38
误　差	10 639.20	8	1 329.90	
总　和	223 011.73	14		

给定 $\alpha = 0.05$，查附表 $F_{0.05}(2,8) = 4.46$，$F_{0.05}(4,8) = 3.84$，这里由于 $F_A = 63.09 > F_{0.05}(2,8)$，$F_B = 8.38 > F_{0.05}(4,8)$，故拒绝 H_{01}, H_{02}，即认为不同的广告形式、不同的价格均造成商品销量的显著差异.

进一步地，通过计算，$\hat{\alpha}_3 = \bar{x}_{3\cdot} - \bar{x} = 125.93$，$\hat{\beta}_2 = \bar{x}_{\cdot 2} - \bar{x} = 90.46$ 分别为各自因素中最大水平效应，故水平组合 (A_3, B_2) 下商品的销量最高，实际销售中应该采用这一组合方式.

三、有交互作用的情况

由于因素可能存在交互作用，我们需要对每一组合 (A_i, B_j) 分别进行 t 次 $(t \geqslant 2)$ 重复试验. 在后面的分析中，我们将会看到，只有这样才能将交互作用平方和从误差平方和中分离出来，我们称这种试验为**等重复试验**. 通常将样本结果列成下表：

有交互作用的双因素方差分析数据结构

A 的水平 \ B 的水平	B_1	B_2	…	B_s
A_1	$X_{111},\cdots,X_{11t}$	$X_{121},\cdots,X_{12t}$	…	$X_{1s1},\cdots,X_{1st}$
A_2	$X_{211},\cdots,X_{21t}$	$X_{221},\cdots,X_{22t}$	…	$X_{2s1},\cdots,X_{2st}$
⋮	⋮	⋮		⋮
A_r	$X_{r11},\cdots,X_{r1t}$	$X_{r21},\cdots,X_{r2t}$	…	$X_{rs1},\cdots,X_{rst}$

显然有交互作用的双因素方差分析模型可归结为

$$\begin{cases} X_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}, \\ \varepsilon_{ijk}\sim N(0,\sigma^2)\text{ 且相互独立}, \end{cases} i=1,\cdots,r,\quad j=1,\cdots,s,k=1,\cdots,t,$$

$$\begin{cases} \sum_{i=1}^{r}\alpha_i=\sum_{j=1}^{s}\beta_j=0, \\ \sum_{i=1}^{r}\gamma_{ij}=\sum_{j=1}^{s}\gamma_{ij}=0, \end{cases} \tag{6.12}$$

其中 ε_{ijk} 为 X_{ijk} 的随机误差.

在上述模型下,我们需要分别对下述假设进行检验

$$H_{01}:\alpha_i=0,\quad i=1,2,\cdots,r, \tag{6.13}$$

$$H_{02}:\beta_j=0,\quad j=1,2,\cdots,s, \tag{6.14}$$

$$H_{03}:\gamma_{ij}=0,\quad i=1,2,\cdots,r,\quad j=1,2,\cdots,s. \tag{6.15}$$

我们记

$$\bar{X}=\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk},$$

$$\bar{X}_{ij\cdot}=\frac{1}{t}\sum_{k=1}^{t}X_{ijk},\quad i=1,2,\cdots,r,\quad j=1,2,\cdots,s,$$

$$\bar{X}_{i\cdot\cdot}=\frac{1}{s}\sum_{j-1}^{s}\bar{X}_{ij\cdot},i=1,2,\cdots,r,\quad \bar{X}_{\cdot j\cdot}=\frac{1}{r}\sum_{i=1}^{r}\bar{X}_{ij\cdot},j=1,2,\cdots,s.$$

可以证明,总离差平方和可进行下面的分解

$$\begin{aligned} Q&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\bar{X})^2 \\ &=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}[(X_{ijk}-\bar{X}_{ij\cdot})+(\bar{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X})+(\bar{X}_{i\cdot\cdot}-\bar{X})+ \\ &\quad(\bar{X}_{\cdot j\cdot}-\bar{X})]^2 \end{aligned}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\bar{X}_{ij\cdot})^2 + t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X})^2$$
$$+ st\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})^2 + rt\sum_{j=1}^{s}(\bar{X}_{\cdot j\cdot}-\bar{X})^2.$$

若记

$$Q_A = st\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})^2,\quad Q_B = rt\sum_{j=1}^{s}(\bar{X}_{\cdot j\cdot}-\bar{X})^2,$$
$$Q_{A\times B} = t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X})^2,$$
$$Q_\varepsilon = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\bar{X}_{ij\cdot})^2,$$

则有

$$Q = Q_A + Q_B + Q_{A\times B} + Q_\varepsilon. \tag{6.16}$$

这里 $Q_A, Q_B, Q_{A\times B}$ 分别反映了因素 A，因素 B 以及交互因素 $A\times B$ 对变量取值的影响，我们分别称之为因素 A 的离差平方和、因素 B 的离差平方和以及交互因素 $A\times B$ 的离差平方和；Q_ε 则反映了随机因素对变量取值的影响，我们称之为误差平方和.

可以证明，当 H_{01} 成立时，有

$$\frac{Q_A}{\sigma^2}\sim\chi^2(r-1),\quad \frac{Q_\varepsilon}{\sigma^2}\sim\chi^2(rs(t-1)),$$

且它们相互独立. 于是

$$F_A = \frac{Q_A/(r-1)}{Q_\varepsilon/[rs(t-1)]}\sim F(r-1, rs(t-1)). \tag{6.17}$$

(6.13)的拒绝域可取为

$$C = \{(x_{111},\cdots,x_{rst}): F_A \geqslant F_\alpha(r-1, rs(t-1))\}.$$

同样，当 H_{02} 成立时，有

$$F_B = \frac{Q_B/(s-1)}{Q_\varepsilon/[rs(t-1)]}\sim F(s-1, rs(t-1)). \tag{6.18}$$

(6.14)的拒绝域可取为

$$C = \{(x_{111},\cdots,x_{rst}): F_B \geqslant F_\alpha(s-1, rs(t-1))\}.$$

当 H_{03} 成立时，有

$$F_{A\times B} = \frac{Q_{A\times B}/[(r-1)(s-1)]}{Q_\varepsilon/[rs(t-1)]}\sim F((r-1)(s-1), rs(t-1)). \tag{6.19}$$

(6.15)的拒绝域可取为

$$C = \{(x_{111},\cdots,x_{rst}): F_{A\times B} \geqslant F_\alpha((r-1)(s-1), rs(t-1))\}.$$

同样地,我们使用下面的方差分析表:

有交互作用的双因素方差分析表

方差来源	平方和	自由度	均方	F 值
因素 A	$Q_A = st\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot} - \bar{X})^2$	$r-1$	$S_A = \dfrac{Q_A}{r-1}$	$F_A = \dfrac{S_A}{S_e}$
因素 B	$Q_B = rt\sum_{j=1}^{s}(\bar{X}_{\cdot j\cdot} - \bar{X})^2$	$s-1$	$S_B = \dfrac{Q_B}{s-1}$	$F_B = \dfrac{S_B}{S_e}$
交互因素 $A\times B$	$Q_{A\times B} = t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X})^2$	$(r-1)(s-1)$	$S_{A\times B} = \dfrac{Q_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B} = \dfrac{S_{A\times B}}{S_e}$
误 差	$Q_e = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk} - \bar{X}_{ij\cdot})^2$	$rs(t-1)$	$S_e = \dfrac{Q_e}{rs(t-1)}$	
总 和	$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk} - \bar{X})^2$	$rst-1$		

另外,可以证明水平 A_i,B_j 及 A_i 与 B_j 的交互效应 α_i,β_j,γ_{ij} 的最大似然估计分别为

$$\hat{\alpha}_i = \bar{X}_{i\cdot\cdot} - \bar{X}, i = 1,2,\cdots,r, \quad \hat{\beta}_j = \bar{X}_{\cdot j\cdot} - \bar{X}, j = 1,2,\cdots,s,$$

$$\hat{\gamma}_{ij} = \bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} - \bar{X}, i = 1,2,\cdots,r, j = 1,2,\cdots,s.$$

相应地各水平组合(A_i, B_j)下总体均值 μ_{ij} 的最大似然估计为

$$\hat{\mu}_{ij} = \bar{X}_{ij\cdot}, \quad i = 1,2,\cdots,r, j = 1,2,\cdots,s.$$

下面我们来解决例 6.3 中的问题,假定显著性水平 $\alpha = 0.05$. 经过计算,可以得到下面的方差分析表:

方差来源	平方和	自由度	均方	F 值
温 度	1 068.06	2	534.03	237.35
催化剂	908.75	3	302.92	134.63
温度 * 催化剂	59.50	6	9.92	4.41
误 差	54.00	24	2.25	
总 和	2 090.31	35		

给定 $\alpha = 0.05$,查附表 $F_{0.05}(2,24) = 3.40$,$F_{0.05}(3,24) = 3.01$,$F_{0.05}(6,24) = 2.51$,由于 $F_A = 237.35 > F_{0.05}(2,24)$,$F_B = 134.63 > F_{0.05}(3,24)$,$F_{A\times B} = 4.41 > F_{0.05}(6,24)$,故拒绝 H_{01},H_{02},H_{03},即认为温度、催化剂以及它们的交互作用对产品的质量指标均有显著影响.

进一步地,经过计算,$\bar{x}_{31\cdot} = 93.67$,$\bar{x}_{32\cdot} = 94.33$ 为所有 $\bar{x}_{ij\cdot}$ 中最大的两个,

它们比较相近，实际生产时，可根据情况，选择(A_3,B_1)或(A_3,B_2)其中之一的生产方案.

例 6.5 火箭的射程(单位:海里)与燃料的种类与推进器的型号有关，现对四种不同的燃料与三种不同型号的推进器进行试验，每种组合各发射火箭两次，测得火箭的射程结果如下：

燃料 \ 推进器	B_1	B_2	B_3
A_1	58.2,52.6	56.2,41.2	65.3,60.8
A_2	49.1,42.8	54.1,50.5	51.6,48.4
A_3	60.1,58.3	70.9,73.2	39.2,40.7
A_4	75.8,71.5	58.2,51.0	48.7,41.4

试由上述试验结果判断燃料、推进器以及它们的交互作用对火箭的射程是否有显著影响.(显著性水平 $\alpha=0.01$)

利用 SPSS 可得到下面的方差分析表：

ANOVA

Source	Sum of Squares	*df*	Mean Square	F	*Sig.*
燃 料	261.675	3	87.225	4.417	.026
推进器	370.981	2	185.490	9.394	.004
燃料 * 推进器	1 768.693	6	294.782	14.929	.000
Error	236.950	12	19.746		
Total	2 638.299	23			

由表中结果可以看出，对于 F_A，F_B及$F_{A\times B}$对应的 p 值，分别有

$Sig.(A)=.026>\alpha$， $Sig.(B)=.004<\alpha$， $Sig.(A\times B)=.000<\alpha$.

这里 $\alpha=0.01$，故可以认为燃料对火箭的射程没有显著影响，而推进器及燃料与推进器的交互作用影响显著.

由于 $\bar{x}_{41.}=73.65$ 为最大，故可以简单认为采用(A_4,B_1)的搭配火箭的射程最远，这一组合的效果最好.

习 题 六

(A)

1. 比较研究三种不同品牌小汽车的耗油量，现从每种品牌选取 5 辆在同一试验场行驶

相同规定的里程,测得每辆汽车的耗油量结果如下:

汽车品牌	耗油量					平均值
A	25.8,	23.4,	21.8,	25.2,	24.8	24.20
B	29.5,	27.8,	28.1,	28.3,	25.4	27.82
C	25.9,	25.4,	24.2,	27.3,	23.4	25.24

试判断不同品牌小汽车的耗油量是否有显著差异.(显著性水平 $\alpha=0.05$)

2. 一批由同一种原料织成的布,用不同的印染工艺处理,然后进行缩水率试验.现采用五种不同的工艺,每种工艺处理4块布样,测得缩水率的百分数如下:

工艺	缩水率				平均值
A_1	4.3,	6.8,	3.2,	5.5	4.95
A_2	6.1,	7.4,	4.2,	5.1	5.70
A_3	4.8,	8.7,	7.2,	8.1	7.20
A_4	6.5,	8.3,	9.2,	7.2	7.80
A_5	9.6,	8.8,	10.4,	7.8	9.15

试分析不同工艺下布的缩水率有无显著差异.(显著性水平 $\alpha=0.05$)

3. 下面给出小白鼠在接种三种不同菌型伤寒杆菌后的存活日数:

菌型	存活日数
Ⅰ	2, 4, 3, 2, 4, 7, 7, 2, 5, 4
Ⅱ	5, 6, 8, 5, 10, 7, 12, 6, 6
Ⅲ	7, 11, 6, 6, 7, 9, 5, 10, 6, 3, 10

试由上述结果判断三种不同菌型下白鼠的存活日数是否有显著差异.(显著性水平 $\alpha=0.05$)

4. 下面给出了用于计算器的四种类型的电路的响应时间(单位:ms):

电路类型	响应时间
Ⅰ	19, 22, 20, 18, 15
Ⅱ	24, 31, 27, 28
Ⅲ	17, 15, 17, 23, 18
Ⅳ	18, 23, 19

试判断不同类型的电路响应时间是否有显著差异.(显著性水平 $\alpha=0.05$)

5. 研究原料的三种不同产地与四种不同的生产工艺对某种化工产品纯度的影响,现对各种组合各进行一次试验,测得产品的纯度结果如下:

工艺 / 产地	B_1	B_2	B_3	B_4
A_1	94.5	97.8	96.1	95.4
A_2	95.8	98.6	97.2	96.4
A_3	92.7	97.1	97.7	93.9

试问:不同的原料产地、不同的生产工艺下产品的纯度是否有显著差异?(显著性水平 $\alpha=0.05$)

6. 下面给出了三位操作工人分别在四台不同机器上操作三天的日产量:

工人 / 机 器	甲	乙	丙
A_1	15, 15, 17	19, 19, 16	16, 18, 21
A_2	17, 17, 17	15, 15, 15	19, 22, 22
A_3	18, 20, 22	15, 16, 17	17, 17, 17
A_4	15, 17, 16	18, 17, 16	18, 18, 18

给定显著性水平 $\alpha=0.05$,试分别检验操作工人、机器以及它们的交互作用是否显著.

7. 比较研究某种农作物的三个不同品种及施用的四种不同肥料对作物产量的影响,现对品种和肥料的不同组合各重复进行二次试验,测得产量结果如下:

肥料 / 品种	B_1	B_2	B_3	B_4
A_1	195, 204	224, 217	238, 251	265, 256
A_2	254, 241	258, 272	294, 276	267, 278
A_3	279, 292	302, 289	325, 308	297, 283

试由上述试验数据推断品种、肥料以及它们的交互作用对作物产量的影响是否显著.(显著性水平 $\alpha=0.05$)

(B)

1. 在单因素方差分析模型(6.2)中,试证若 $r=2$,则此时的 F 检验法与两等方差正态总体的 t 检验法是等价的.

2. 试证在单因素方差分析模型(6.2)中,各参数 $\mu,\alpha_1,\cdots,\alpha_r,\sigma^2$ 的最大似然估计分别为

$\hat{\mu}=\bar{X}$，$\hat{\alpha}_i=\bar{X}_{i\cdot}-\bar{X}$，$\hat{\sigma}^2=\dfrac{Q_e}{n}$且$\hat{\mu}$，$\hat{\alpha}_i$分别是$\mu$，$\alpha_i$的无偏估计.

3. 试证在无交互作用双因素方差分析模型(6.6)中，参数α_i，β_j的最大似然估计分别为：

$$\hat{\alpha}_i=\bar{X}_{i\cdot}-\bar{X},i=1,2,\cdots,r,\quad \hat{\beta}_j=\bar{X}_{\cdot j}-\bar{X},j=1,2,\cdots,s,$$

且它们都是无偏的.

4. 试证在有交互作用双因素方差分析模型(6.12)中，无论假设H_{01}，H_{02}，H_{03}是否成立，均有

$$\frac{Q_e}{\sigma^2}\sim\chi^2(rs(t-1)),$$

且$\dfrac{Q_e}{rs(t-1)}$是σ^2的无偏估计.

第 7 章

回归分析

回归分析是研究两个或两个以上变量之间的相互关系的一种重要的统计方法.与变量之间的函数关系不同,这种关系描述的是变量之间相互依存相伴发生的性质,它不是一种确定性的关系.比如不能根据对某种商品的给定的一种广告投入来完全确定该种商品的销售额,但是增加广告投入往往可以提高销售额,因此这两者之间确实存在一定的关联性.回归分析通过建立统计模型来研究这种关系,并由此对相应的变量进行预测和控制.回归分析具有广泛的应用,在经济学中,它是进行经济计量分析的主要工具.

本章介绍线性回归模型的估计、检验以及相应的预测和控制等问题,我们将重点讨论一元线性回归模型.

§7.1 一元线性回归模型及其参数估计

一、回归模型

无论是在物理、生物等自然科学还是在经济、管理等社会学科中,都要研究变量与变量的关系问题,通常这种关系有两种形式:

一种是我们熟知的函数关系,比如匀速运动物体运动的距离 s,速度 v 和运动的时间 t 之间具有明确的关系

$$s = v \cdot t,$$

任意给定 v 和 t 的值,s 也被完全确定.这种确定性是函数关系的重要特征.

另一种关系则无法通过明确的函数关系来表达,我们可以先看下面的例子.

例 7.1 某广告公司为了研究某一类产品的广告费用与其销售额之间的关系,对多个厂家进行了调查,获得数据资料如表 7.1 所示.

表 7.1 广告投入与销售额资料 (单位:万元)

厂 家	1	2	3	4	5	6	7	8	9	10
广告费	35	60	25	30	35	40	25	20	50	45
销售额	440	520	380	475	385	525	450	365	540	500

为了分析的方便我们用 X 表示广告费用，用 Y 表示销售额. 从表 7.1 中不难看出，X 与 Y 之间不可能存在一个明确的函数关系式. 事实上，即便不同厂家投入了相同的广告费用，其各自的销售额也不会是完全相同的，影响销售额的因素是多种多样的，除了广告投放的影响，还受厂家产品的特色、定价、销售渠道、售后服务以及其他一些偶然性因素的影响. 因此尽管厂家 1 和 5 以及厂家 3 和 7 广告投入分别相同，其销售额却是不同的.

但是，也不能就此认为这两者之间没有什么关系. 为了更加清楚地看清其中的规律，通常采用“散点图”的方式将各个厂家的广告费用和销售额成对数据 $(X_i, Y_i)(i=1,2,\cdots,10)$ 在平面直角坐标系中用点表示出来，如图 7.1 所示.

在图 7.1 中可以看出，随着广告投入费用的增加，销售额基本上也呈上升趋势，图中的点大致分布在一条向右上方倾斜的直线附近，高投入的广告费用对应的是一组比较高的销售额，伴随着比较低的广告投入结果则是一组比较低的销售额.

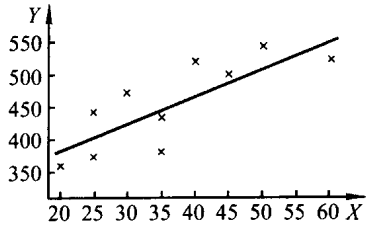


图 7.1 广告投入与销售额的散点图

统计学上将类似于上述广告费用与销售额之间这种不具有确定函数关系的两个变量之间的统计关系称为**相关关系**.

研究两个变量之间的相关关系主要从两个方向进行：一个方向是进行相关分析，即通过引入一定的统计指标来量化变量之间相关的程度，此时两个变量的地位是对等的，没有方向上的差异. 比如相关系数就是这样一个指标.

人们往往通过对 (X, Y) 的一个观察样本 $(X_i, Y_i)(i=1,2,\cdots,n)$ 来对 X 与 Y 的相关系数作出估计，常用的估计是所谓的**样本相关系数**

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]^{1/2}}. \tag{7.1}$$

可以证明，r 是相关系数 ρ 的渐近无偏估计.

另一个研究变量间相关关系的方向就是回归分析.

回归分析是如何对变量间的相关关系进行研究的呢？我们回到前面关于广

告费用与销售额的例子.从图 7.1 和前面的分析,可以将销售额的观察结果 y 看作由两部分叠加而成,一部分由广告投入的线性函数引起,记作 $\beta_0+\beta_1 x$;另一部分由随机因素引起,记作 ε.即

$$y=\beta_0+\beta_1 x+\varepsilon.$$

相应地,变量 Y 与变量 X 之间的关系可表示为

$$Y=\beta_0+\beta_1 X+\varepsilon, \tag{7.2}$$

其中 Y 是随机变量,X 是一个普通变量,对它可加以控制.ε 是一随机变量,它可看成随机误差,因而可以假定 $E\varepsilon=0$.于是,在给定 X 的取值时,有

$$E[Y|X]=\beta_0+\beta_1 X.$$

对本例而言回归分析的主要任务是确定(7.2)中的线性函数 $\beta_0+\beta_1 X$ 的系数 β_0 和 β_1.一旦通过估计和检验得到一个合适的线性函数 $\beta_0+\beta_1 X$,便可利用(7.2)对 Y 进行预测和实施控制.例 7.1 所述的是变量 X 和 Y 的一种特别简单的关系,但它在实际中具有广泛的应用,在对它进行深入讨论之前,我们先对回归分析作出一般的阐述.

正如前面所描述的,**回归分析**是寻求一个随机变量 Y 对另一个或一组(随机或非随机)变量 $X_1,\cdots,X_m$ 的相关性的一种统计方法,它主要通过对变量的观察所获得的统计数据来确定反映变量间关系的经验公式,并通过所得公式进行统计描述、分析和推断,进而解决预测、控制和优化问题.

用来进行回归分析的数学模型通常有如下一般形式:

$$Y=f(X_1,X_2,\cdots,X_m)+\varepsilon, \tag{7.3}$$

其中 $f(X_1,\cdots,X_m)$ 是 $X_1,\cdots,X_m$ 的一个确定的函数,这一函数,通常称为 Y 对 $X_1,\cdots,X_m$ 的**回归函数**.ε 是数学期望为 0 的随机变量,称为**随机误差**.方程 $y=f(x_1,x_2,\cdots,x_m)$ 称作**回归方程**.自变量 $X_1,X_2,\cdots,X_m$ 称为"回归变量"或"解释变量",在预测问题中也称之为"预报因子";因变量 Y 称作"响应变量","被解释变量",在预测问题中也称之为"预测量".

用来进行回归分析的数学模型(含有关假设)称为**回归模型**.只含有一个回归变量的回归模型称为**一元回归模型**,否则称为**多元回归模型**.在应用回归分析时,一般假设回归函数 f 的形式是已知的,但含有未知参数.最简单、最重要的且最便于处理,因而也最常用的情形是函数 f 关于未知参数 $\beta_0,\beta_1,\cdots,\beta_k$ 是线性的情形,这样的回归模型称为**线性回归模型**,否则称为**非线性回归模型**.许多非线性回归模型经过变换可化为线性回归模型.因此,我们重点介绍线性回归模型.

二、一元线性回归模型

在所有回归模型中,最简单的情形是两个变量间的线性回归模型.这一模型

中回归函数是单个解释变量的、系数未知的线性函数，即形如：

$$f(X)=\beta_0+\beta_1 X, \tag{7.4}$$

其中 β_0、β_1 为待定的参数.此时的回归方程表现为一条直线.

除此之外，为了处理问题的方便，通常我们只考虑 X 为可控制变量，即 X 不是一个随机变量.这一假定在大多数实际情况下是合理的，比如例 7.1 中的广告费用的确是一个人们可以控制的确定的量.在今后我们用 x 代替 X 以示它为一确定的量.n 组样本则以 $(x_i, Y_i)(i=1,2,\cdots,n)$ 来表示.

根据样本抽取原则，可以设 $Y_1, Y_2, \cdots, Y_n$ 是相互独立的，这意味着对任意 i，Y_i 的取值不依赖于其他 Y_j 的值$(j\neq i, i=1,2,\cdots,n)$.由(7.4)式易知

$$Y_i=\beta_0+\beta_1 x_i+\varepsilon_i, i=1,2,\cdots,n, \tag{7.5}$$

其中 $\varepsilon_i(i=1,2,\cdots,n)$ 是 n 个独立的随机变量，期望 $E(\varepsilon_i)=0(i=1,2,\cdots,n)$.

这里的 $\varepsilon_i(i=1,2,\cdots,n)$ 又可以理解为 Y_i 偏离回归直线的大小，通常还假定这种偏离的程度不会随 x_i 的不同而发生改变，即假定

$$E(\varepsilon_i^2)=D(\varepsilon_i)=\sigma^2, i=1,2,\cdots,n, \tag{7.6}$$

其中 σ^2 是待定的正常数.上式又被称为**同方差**(Homoscedasticity)**假设**.

至此，我们给出了在经典统计中对一元线性回归模型做出的几个基本假设.现在总结如下：

(H_1) 回归函数是自变量 x 的一次线性函数；

(H_2) 自变量 x 被看作确定变量；

(H_3) n 个样本 $Y_1, Y_2, \cdots, Y_n$ 是独立的；

(H_4) (7.5)式中的 ε_i 满足同方差条件，即 $D(\varepsilon_i)=\sigma^2, i=1,2,\cdots,n$，从而 $Y_1, Y_2, \cdots, Y_n$ 的方差也相等，即

$$D(Y_i)=\sigma^2, i=1,2,\cdots,n. \tag{7.7}$$

对一元线性回归模型，我们要解决下面问题：

(1) 参数估计：给出 $\beta_0, \beta_1, \sigma^2$ 的估计.

(2) 显著性检验：检验线性函数 $\beta_0+\beta_1 x$ 用来描述 Y 和 X 间的关系是否合适.

(3) 模型检查：检查对模型做的假设是否成立.

(4) 预测或控制.

以下分别加以介绍.

三、最小二乘估计

对 β_0, β_1 的估计实际上就是在平面直角坐标系中“估计”一条直线

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1 x, \tag{7.8}$$

使得它尽可能地接近回归直线 $Y=\beta_0+\beta_1 x$. 如图 7.2 所示.

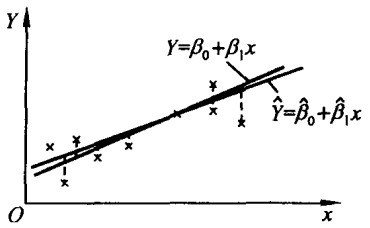


图 7.2 估计的回归直线

一种比较直观的思路是希望观测得到的样本点(x_i, Y_i)与直线(7.8)上相应的点$(x_i, \hat{Y}_i)(i=1,2,\cdots,n)$尽可能地接近. 或者说使它们每一对点之间的距离都尽可能地短,从而要求直线(7.8)使得以下平方和达到最小值:

$$Q(\beta_0,\beta_1)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 x_i)^2 \tag{7.9}$$

即

$$Q(\hat{\beta}_0,\hat{\beta}_1)=\min_{\beta_0,\beta_1} Q(\beta_0,\beta_1), \tag{7.10}$$

这里的 $Q(\beta_0,\beta_1)$是一个非负的二元函数. 根据多元函数微分学的知识,$\hat{\beta}_0,\hat{\beta}_1$应该满足

$$\left.\frac{\partial Q}{\partial \beta_0}\right|_{\beta_0=\hat{\beta}_0,\beta_1=\hat{\beta}_1}=\left.\frac{\partial Q}{\partial \beta_1}\right|_{\beta_0=\hat{\beta}_0,\beta_1=\hat{\beta}_1}=0, \tag{7.11}$$

将上式展开,即得

$$\begin{cases}\sum\limits_{i=1}^{n}(Y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)=0,\\ \sum\limits_{i=1}^{n}(Y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)x_i=0.\end{cases} \tag{7.12}$$

经过整理,得到关于 $\hat{\beta}_0$和 $\hat{\beta}_1$的一个方程组

$$\begin{cases}\overline{Y}=\hat{\beta}_0+\hat{\beta}_1\bar{x},\\ \sum\limits_{i=1}^{n}x_i Y_i=n\bar{x}\hat{\beta}_0+\left(\sum\limits_{i=1}^{n}x_i^2\right)\hat{\beta}_1,\end{cases} \tag{7.13}$$

其中 $\bar{x}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$, $\overline{Y}=\frac{1}{n}\sum\limits_{i=1}^{n}Y_i$.

求解上述方程组,即可得到

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\bar{x}, \tag{7.14}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}. \tag{7.15}$$

容易验证,这样得到的 $\hat{\beta}_0, \hat{\beta}_1$ 的确是(7.9)式定义的二元函数 Q 的最小值点.我们称 $\hat{\beta}_0, \hat{\beta}_1$ 是 β_0, β_1 的**最小二乘估计**(Least Squares Estimators,简记作LSE).将 $\hat{\beta}_0, \hat{\beta}_1$ 代入(7.8)式即可得到估计出来的回归直线方程为

$$\hat{Y} = \bar{Y} + \hat{\beta}_1 (x - \bar{x}). \tag{7.16}$$

这是一条过点 $(\bar{x}, \bar{Y})$ 和 $(0, \hat{\beta}_0)$ 的直线.

为了书写简洁,我们引入记号

$$l_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2,$$

$$l_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

$$l_{xY} = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}_i) = \sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y},$$

则 $\hat{\beta}_1$ 可以表示成

$$\hat{\beta}_1 = \frac{l_{xY}}{l_{xx}}. \tag{7.17}$$

对于每一个 $x_i (i = 1, 2, \cdots, n)$,称

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{7.18}$$

为相应的真实值 Y_i 的**回归值**或者**拟合值**,记

$$e_i = Y_i - \hat{Y}_i, i = 1, 2, \cdots, n, \tag{7.19}$$

称 $e_i (i = 1, 2, \cdots, n)$ 为**拟合残差**或简称**残差**.

任给常数 a 和 b,由(7.12)式容易验证下列简单的性质:

$$\sum_{i=1}^{n} e_i \cdot (ax_i + b) = 0. \tag{7.20}$$

特别地,有 $\sum_{i=1}^{n} e_i = 0$ 成立.

对于例7.1中给出的样本数据,根据(7.14)和(7.15)式可以求得 β_0, β_1 的最小二乘估计值分别为 $\hat{\beta}_0 = 309.527\,6, \hat{\beta}_1 = 4.067\,736$,由此得到回归直线方程的估计是

$$Y = 309.527\,6 + 4.067\,736x.$$

四、$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的性质

最小二乘法得到的 β_0 和 β_1 的估计量是否是无偏的,其有效性如何?这些问

题的解答将有助于我们对最小二乘估计进行评价.

从(7.15)式不难将 $\hat{\beta}_1$ 写成如下形式

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{l_{xx}} \cdot Y_i, \tag{7.21}$$

即 $\hat{\beta}_1$ 是 $Y_1, Y_2, \cdots, Y_n$ 的线性组合.于是,

$$\begin{aligned} E(\hat{\beta}_1) &= \sum_{i=1}^{n} \frac{x_i - \bar{x}}{l_{xx}} \cdot E(Y_i) \\ &= \sum_{i=1}^{n} \frac{x_i - \bar{x}}{l_{xx}} \cdot (\beta_0 + \beta_1 \cdot x_i) \\ &= \beta_0 \cdot \sum_{i=1}^{n} \frac{x_i - \bar{x}}{l_{xx}} + \beta_1 \cdot \sum_{i=1}^{n} \frac{(x_i - \bar{x}) \cdot x_i}{l_{xx}} \\ &= \beta_1. \end{aligned} \tag{7.22}$$

因此,$\hat{\beta}_1$ 是 β_1 的无偏估计.

考虑到 $Y_1, Y_2, \cdots, Y_n$ 之间的独立性以及同方差性,易知

$$D(\hat{\beta}_1) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{l_{xx}^2} \cdot D(Y_i) = \frac{\sigma^2}{l_{xx}}. \tag{7.23}$$

类似地,利用(7.14)式以及(7.21)式也可以将 $\hat{\beta}_0$ 写成 $Y_1, Y_2, \cdots, Y_n$ 的线性组合的形式:

$$\hat{\beta}_0 = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{l_{xx}} \right) \cdot Y_i. \tag{7.24}$$

同样地可以验证

$$E(\hat{\beta}_0) = \beta_0 \tag{7.25}$$

以及

$$D(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}} \right) \cdot \sigma^2. \tag{7.26}$$

因此,$\hat{\beta}_0$ 也是 β_0 的无偏估计.

由于 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 都可以表示成样本 $Y_1, Y_2, \cdots, Y_n$ 的线性组合的形式,通常称这种形式的估计为**线性估计**.

可以证明,对于最小二乘估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 来说有下面的定理成立.

定理 7.1(高斯 - 马尔可夫定理) 在假设(H_1)~(H_4)满足的条件下,模型(7.5)中参数 β_0 和 β_1 的最小二乘估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$(分别由(7.14)和(7.15)给出)是最优线性无偏估计量(Best Linear Unbiased Estimators,简记作 BLUE).

上述定理中的"最优"一词的含义为在所有关于 β_0 和 β_1 的线性无偏估计量中 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 分别是其中方差最小的一个.

另外，由(7.8)式定义的 $\hat{Y}$ 可以作为 $E(Y|x)$ 的一个估计，而且由于

$$E(\hat{Y}) = E(\hat{\beta}_0) + E(\hat{\beta}_1)\cdot x = \beta_0 + \beta_1 x = E(Y|x),$$

可知 $\hat{Y}$ 是 $E(Y|x)$ 的无偏估计. 由于

$$\begin{aligned}
\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) &= \operatorname{cov}(\overline{Y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) \\
&= \operatorname{cov}(\overline{Y}, \hat{\beta}_1) - \operatorname{cov}(\hat{\beta}_1 \bar{x}, \hat{\beta}_1) \\
&= \operatorname{cov}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} \frac{x_i - \bar{x}}{l_{xx}} \cdot Y_i\right] - \bar{x} \cdot \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_1) \\
&= \sum_{i=1}^{n} \frac{x_i - \bar{x}}{n l_{xx}} \cdot D(Y_i) - \bar{x} \cdot D(\hat{\beta}_1) \\
&= \sum_{i=1}^{n} \frac{x_i - \bar{x}}{n l_{xx}} \cdot \sigma^2 - \bar{x} \cdot \frac{\sigma^2}{l_{xx}} \\
&= -\frac{\bar{x}}{l_{xx}} \cdot \sigma^2, \qquad (7.27)
\end{aligned}$$

于是，$\hat{Y}$ 的方差为

$$D(\hat{Y}) = D(\hat{\beta}_0) + D(\hat{\beta}_1)\cdot x^2 + 2\cdot \operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1)\cdot x = \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{l_{xx}}\right]\cdot \sigma^2. \qquad (7.28)$$

§7.2 一元线性回归模型的检验

不难看出，对于一元线性回归模型而言，无论变量 x 对变量 Y 有无影响，总可以利用上一节中给出的最小二乘估计公式(7.14)和(7.15)，由给定的一组样本值求出 β_0 和 β_1 的估计值 $\hat{\beta}_0, \hat{\beta}_1$ 并得到相应的回归方程，正如我们对例 7.1 所做的一样. 但是，是否广告费用真正对销售额有影响和解释作用呢？如果答案是否定的，那么上面求解出来的回归方程就是没有意义的. 因此对回归模型的显著性进行检验，也就是对 x 的变化是否影响到 Y 进行检验，是非常有必要的.

在一元线性回归模型中，回归函数 $E(Y|x)$ 是 x 的线性函数，如果 x 的变化与 Y 无关，则说明 $E(Y|x)$ 与 x 无关，即有 $\beta_1 = 0$；反之，若 $\beta_1 = 0$，则回归函数是一个常数，从而 x 变化对 Y 不产生影响. 因此检验变量 x 是否对 Y 有解释作用等价于检验下面假设

$$H_0: \beta_1 = 0 \qquad (7.29)$$

能否被拒绝.

为了寻求上述假设的检验统计量，我们常要对一元线性回归模型(7.5)作出一个附加的假设条件：

(H_5) ε_i服从正态分布.即 $\varepsilon_i \sim N(0,\sigma^2), i=1,2,\cdots,n$.

易知,在(H_5)成立时,$Y_1, Y_2, \cdots, Y_n$也服从正态分布,且 $Y_i \sim N(\beta_0+\beta_i x_i, \sigma^2), i=1,2,\cdots,n$.另外,$\hat{\beta}_0, \hat{\beta}_1, \hat{Y}$显然也都服从正态分布,且

$$\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n}+\frac{\bar{x}^2}{l_{xx}}\right)\cdot\sigma^2\right), \tag{7.30}$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{l_{xx}}\right), \tag{7.31}$$

$$\hat{Y} \sim N\left(\beta_0+\beta_1\cdot x, \left[\frac{1}{n}+\frac{(x-\bar{x})^2}{l_{xx}}\right]\cdot\sigma^2\right). \tag{7.32}$$

本章后面的分析都是基于假设条件$(H_1)\sim(H_5)$成立而作出的.

一、方差分析与 F 检验

正如我们在前面一章中介绍的,分析变量 x 是否对 Y 有影响,可以通过将 Y 的总离差平方和进行分解来确定诸因素对 Y 的各个样本之间的差异做出的贡献.事实上,导致这种差异性的原因无非是变量 x 的变化或者是其他一些随机性因素的作用.对比这两种因素对 Y 的样本之间差异的贡献是寻找假设(7.29)的检验统计量的一个直观思路.

首先,容易证明总的离差平方和为

$$SST = l_{YY}\sum_{i=1}^{n}(Y_i-\bar{Y})^2. \tag{7.33}$$

可以分解为:

$$\begin{aligned} SST &= \sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2 \\ &\xlongequal{\text{def}} SSE+SSR, \end{aligned} \tag{7.34}$$

其中 $SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}e_i^2$称为**残差平方和**,$SSR=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2=\hat{\beta}_i^2\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2=\hat{\beta}_1^2\cdot l_{xx}$称为**回归平方和**.

为了能够更加清楚地了解这两个平方和各自的含义,并用于对回归模型的检验我们首先给出下列命题.

命题 7.1 (1) $E(SSR)=\sigma^2+\beta_1^2 l_{xx}$ (7.35)

(2) $E(SSE)=(n-2)\sigma^2$ (7.36)

我们将证明放在附录 7.1 中.

由命题 7.1(2)知

$$E\left(\frac{SSE}{n-2}\right)=\sigma^2, \tag{7.37}$$

从而$\frac{SSE}{n-2}$是σ^2的无偏估计量.

从(7.35)和(7.37)知

$$E(SSR)=\left(1+\frac{\beta_1^2 l_{xx}}{\sigma^2}\right)E\left(\frac{SSE}{n-2}\right)\geqslant E\left(\frac{SSE}{n-2}\right).$$

当$\beta_1=0$时,$E(SSR)=E\left(\frac{SSE}{n-2}\right)$. 可见,当$\frac{SSR}{SSE/(n-2)}$足够大时应拒绝$H_0:\beta_1=0$,即$H_0$的拒绝域可取为所有满足条件

$$\frac{SSR}{SSE/(n-2)}>c \tag{7.38}$$

的样本集合,这里c是待定的正常数.为在给定显著性水平下,确定相应的c的值,我们要明确相应统计量的分布.

首先,由于$Y_i\sim N(\beta_0+\beta_i x_i,\sigma^2)(i=1,2,\cdots,n)$,所以在$H_0$成立时,$Y_i\sim N(\beta_0,\sigma^2),i=1,2,\cdots,n$.易知此时

$$\frac{SST}{\sigma^2}\sim\chi^2(n-1). \tag{7.39}$$

又因为$\hat{\beta}_1\sim N\left(\beta_1,\frac{\sigma^2}{l_{xx}}\right)$,故在$H_0$成立时,$\hat{\beta}_1\sim N\left(0,\frac{\sigma^2}{l_{xx}}\right)$,于是

$$\frac{SSR}{\sigma^2}=\frac{\hat{\beta}_1^2\cdot l_{xx}}{\sigma^2}\sim\chi^2(1). \tag{7.40}$$

另外,$e_i=Y_i-\hat{Y}_i(i=1,2,\cdots,n)$显然可以写作$Y_1,Y_2,\cdots,Y_n$的线性组合形式.因此$e_i(i=1,2,\cdots,n)$服从正态分布,$SSE$是$n$个正态随机变量的平方和,考虑到(7.12)中的两个约束等式,可知$\frac{SSE}{\sigma^2}$服从自由度为$n-2$的χ^2分布.①

进一步还可以证明,在H_0成立的条件下,SSR与SSE独立.因此,由(7.39)和(7.40)知,在H_0成立时,

$$F=\frac{SSR}{SSE/(n-2)}\sim F(1,n-2) \tag{7.41}$$

于是,给定显著性水平α,查表获得临界值$c=F_\alpha(1,n-2)$,即$F(1,n-2)$分布的水平α的上侧分位数,H_0的拒绝可取为

$$C=\{(x_1,y_1;x_2,y_2;\cdots;x_n,y_n):F>F_\alpha(1,n-2)\} \tag{7.42}$$

这样就得到了H_0的一个显著性水平为α的检验.

类似于前面一章,上述分析也可以用"方差分析表"表达出来.如表7.2所示.

① 关于这一点的严格证明可参阅书后所附有关文献.

表 7.2 回归方程显著性检验的方差分析表

方差来源	平 方 和	自由度	均方	F 值
回归	$SSR=\sum_{i=1}^{n}(\hat{Y}_i-\overline{Y})^2$	1	$MSR=\frac{SSR}{1}$	$F=\frac{MSR}{MSE}$
残差	$SSE=\sum_{i=1}^{n}(Y-\hat{Y}_i)^2$	$n-2$	$MSE=\frac{SSE}{n-2}$	
总计	$SST=\sum_{i=1}^{n}(Y_i-\overline{Y}_i)^2$	$n-1$		

作为例子,对例 7.1 中给出的广告费用及销售额的数据(表 7.1)可以按照上面给出的有关公式分别求得各个平方和,从而求出 F 值,与临界值对比即可获得有关结论.

使用统计软件可以方便地获得类似表 7.2 的方差分析表.表 7.3 是我们针对例 7.1 中的数据使用 SPSS 软件得到的方差分析表.

表 7.3 例 7.1 数据的方差分析

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23206.435	1	23206.435	12.934	.007
	Residual	14353.565	8	1794.196		
	Total	37560.000	9			

a. Predictors:(Constant),广告费

b. Dependent Variable:销售额

表 7.3 中的"Sig."一项即是 F 的 P 值,由于这一值 0.007 远远小于 0.05,故可以在给定显著性水平为 0.05 的前提下拒绝原假设 $H_0:\beta_1=0$,从而可以认为广告费用的确对销售额有影响,回归方程是显著的.另外,通过查表也可以得到自由度为 1 和 8 的 F 分布的 0.05 上侧分位数为 $F_{0.05}(1,8)=5.32$,将这一临界值同 F 值相对比,由 $F=12.934>F_{0.05}(1,8)$,因此可以获得相同的结论.

二、可决系数与相关系数检验

根据前面离差平方和的分解公式(7.34)以及 SSR 和 SSE 的各自含义,不难看出,回归平方和 SSR 在总的离差平方和 SST 中所占比重越大,从而残差平方和 SSE 所占比重越小,则越发说明了变量 x 对变量 Y 有一定的影响作用.因此,我们引入如下的指标 R^2 来度量变量 x 对变量 Y 的解释能力:

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST}. \tag{7.43}$$

R^2 称为**可决系数**(或**判定系数**).由定义知

$$R^2 \leqslant 1. \tag{7.44}$$

R^2作为一个相对的指标,测度了拟合的回归直线所导致离差平方和占样本的总离差平方和的百分比,因此它也是对回归方程拟合优度的一种测度.R^2越接近于1,则回归方程对样本点拟合得越好,x对Y的解释能力越强.特别地,如果$R^2=1$,则$SSE=0$,说明每个样本点都落在了回归直线上,即被完全拟合了,此时由x的变化完全可以解释Y的变化.

利用前面得到的$SSR=\hat{\beta}_1^2 \cdot l_{xx}$以及$\hat{\beta}_1$的计算公式(7.17),可得

$$\begin{aligned} R^2 &= \frac{\hat{\beta}_1^2 \cdot l_{xx}}{l_{YY}} = \frac{l_{xY}^2}{l_{xx} \cdot l_{YY}} \\ &= \frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(Y_i-\bar{Y})^2}. \end{aligned} \tag{7.45}$$

容易看出可决系数R^2就是变量x和Y之间的样本相关系数r(此时将x看成随机变量)的平方,即

$$R^2 = r^2. \tag{7.46}$$

这样一来,样本相关系数r和R^2尽管定义的角度不同,但是对于一元线性回归模型而言,它们反映的性质是等价的,即x与Y的相关程度越强,x的回归方程对Y的解释能力也越强.

基于上述认识,我们可以对零假设$H_0:\beta_1=0$设置拒绝域为满足条件

$$|r| > d \tag{7.47}$$

的样本集合,其中d为某一待定正常数.

事实上,由于

$$\begin{aligned} r^2 = R^2 &= \frac{SSR}{SST} = \frac{1}{1+\dfrac{SSE}{SSR}} \\ &= \frac{1}{1+\dfrac{n-2}{F}} = \frac{F}{F+(n-2)} \end{aligned} \tag{7.48}$$

在$n>2$时是F的增函数,所以F大于上侧分位数$F_\alpha(1,n-2)$时,

$$r^2 > \frac{F_\alpha(1,n-2)}{F_\alpha(1,n-2)+(n-2)}. \tag{7.49}$$

因此令(7.48)式中的d为

$$d = \sqrt{\frac{F_\alpha(1,n-2)}{F_\alpha(1,n-2)+(n-2)}}. \tag{7.50}$$

这样设置拒绝域为满足条件$|r|>d$的样本集合即可得到H_0的一个显著性水平

为 α 的检验.为此,有时也称 d 为 r 的显著性水平为 α 的**临界值**,记为 $r_{\frac{\alpha}{2}}(n-2)=d$.

对例 7.1 而言,可以算得广告费用与销售额之间的样本相关系数 $r=0.786\,034$,给定显著性水平 $\alpha=0.05$,查表得临界值 $r_{0.025}(8)=0.632$,由 $|r|>r_{0.025}(8)$,所以拒绝原假设 $H_0:\beta_1=0$,得到的结论跟前面 F 检验是一致的.

三、t 检验

如果单纯地把 β_0 和 β_1 理解为模型中的参数,由于前面已经得到了 β_0 和 β_1 的点估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 及其分布,所以利用点估计可以方便地构造关于相应参数的假设检验问题的检验统计量.

在前面已经得到

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{l_{xx}}\right),$$

故

$$\frac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{l_{xx}}} \sim N(0,1). \tag{7.51}$$

又由于

$$\frac{SSE}{\sigma^2} \sim \chi^2(n-2).$$

而且可以证明 SSE 与 $\hat{\beta}_1$ 独立,根据 t 分布的定义,则

$$\frac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{l_{xx}}}\Bigg/\sqrt{\frac{SSE}{\sigma^2(n-2)}}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-2), \tag{7.52}$$

其中

$$\hat{\sigma}_{\hat{\beta}_1}=\sqrt{\frac{SSE}{n-2}}\cdot\frac{1}{\sqrt{l_{xx}}} \tag{7.53}$$

称为 $\hat{\beta}_1$ 的**标准误差**,它是 $\hat{\beta}_1$ 的标准差 $\frac{\sigma}{\sqrt{l_{xx}}}$ 的估计值.

在零假设 $H_0:\beta_1=0$ 成立的条件下,

$$t_1=\frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-2). \tag{7.54}$$

给定显著性水平 α,令 $t_{\frac{\alpha}{2}}(n-2)$ 表示自由度为 $n-2$ 的 t 分布的水平 $\frac{\alpha}{2}$ 上侧分位数,则可设定拒绝域

$$C=\{(x_1,y_1;x_2,y_2;\cdots;x_n,y_n):|t_1|>t_{\frac{\alpha}{2}}(n-2)\}. \tag{7.55}$$

类似地,可以证明

$$\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\hat{\beta}_0}}\sim t(n-2), \tag{7.56}$$

其中

$$\hat{\sigma}_{\hat{\beta}_0}=\sqrt{\frac{SSE}{n-2}}\cdot\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{l_{xx}}} \tag{7.57}$$

称为 $\hat{\beta}_0$ 的标准误差,它是 $\hat{\beta}_0$ 的标准差 $\sigma\cdot\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{l_{xx}}}$ 的估计值.

在原假设 $H_0:\beta_0=0$ 成立时,

$$t_0=\frac{\hat{\beta}_0}{\hat{\sigma}_{\hat{\beta}_0}}\sim t(n-2). \tag{7.58}$$

从而设定拒绝域为满足条件

$$|t_0|>t_{\frac{\alpha}{2}}(n-2) \tag{7.59}$$

的样本集合.

对于例 7.1 中给出的数据,我们已经求得回归方程的两个系数的估计值分别为 $\hat{\beta}_0=309.528$,$\hat{\beta}_1=4.068$,由公式(7.57)和(7.53)可以分别求得两个估计值的标准误差分别为 $\hat{\sigma}_{\hat{\beta}_0}=43.402$,$\hat{\sigma}_{\hat{\beta}_1}=1.131$.从而相应的两个 t 值分别为 $t_0=7.132$和 $t_1=3.596$,查表可得自由度为 8 的 t 分布的 0.025 上侧分位数为 2.306,t_0,t_1 都大于该临界值,从而在给定显著性水平为 0.05 的前提下 $H_0:\beta_0=0$和 $H_0:\beta_1=0$ 都可以被拒绝.

如果使用诸如 SPSS 等统计软件,除了给出上述各值的计算结果之外,通常还会给出 t_0和 t_1的 p 值.表 7.4 是我们针对例 7.1 中的数据使用 SPSS 得到的回归系数估计表.

表 7.4 例 7.1 数据的回归系数估计表

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	309.528	43.402		7.132	.000
	广告费	4.068	1.131	.786	3.596	.007

a. Dependent Variable:销售额

表 7.4 中第一列是系数对应的项,分别为常数项和自变量广告费,第二列和

第三列分别为系数的估计值 $\hat{\beta}_i$ 和 $\hat{\beta}_i$ 的标准误差 $\hat{\sigma}_{\hat{\beta}_i}$ $(i=0,1)$. 表中的"t"所在的列是前面所说的两个 t 值 t_0 和 t_1,"Sig."的列则给出了 t_0 和 t_1 的 p 值分别为 0.000 和 0.007(保留三位小数). 两个 p 值都远小于显著水平 0.05,因此利用这均拒绝原假设的结论.

也许有人已经注意到,对 $H_0:\beta_1=0$ 的 F 检验和 t 检验中 F 值和 t_1 值的 p 值都是 0.007,实际上,对于一元线性回归模型来说,上述两个检验是等价的(即有相同的拒绝域).

§7.3 一元线性回归的残差分析

上一节讨论的检验都有一个前提,就是一元线性回归模型符合假设(H_1)、(H_2)、(H_3)、(H_5). 如何从数据出发判断这些假设是否成立呢? 这是回归分析处理实际数据时的重要问题.

一元线性回归模型的假设简单概括起来就是"独立"、"线性"、"正态"、"同方差". 我们可以通过分析残差来考察这些假设是否成立.

根据前面的定义,残差为观察值与拟合值的差,即

$$e_i = Y_i - \hat{Y}_i, \quad i=1,2,\cdots,n.$$

残差是回归方程未能解释的量. 若模型合适,残差就是随机误差 ε_i $(i=1,2,\cdots,n)$ 的估计. 我们通过对残差的分析来寻找数据违反模型假设的证据,如果结论是违背了模型的假设,则需要调整数据分析方案.

残差分析所包含的内容很多,下面我们介绍几种较为常用的分析方法.

一、残差图

许多残差分析都是在对残差图形的仔细考察基础上完成的,常用的有关于自变量 x 的值的残差图和关于因变量的预测值 $\hat{Y}$ 的残差图.

关于自变量 x 的值的残差图是用横轴表示自变量的值,用纵轴表示残差值. 每个残差用图上的一个点来表示,每个点的第一个坐标由 x_i 的值给出,第二个坐标由对应的第 i 个残差值 e_i 给出.

如果假定对所有的 x 值,ε 的方差都是相同的,并且假定描述变量 x 和 Y 之间关系的回归模型是合理的,那么所有的点应随机地分布在与横轴平行且对称的两直线围成的带域内,如图 7.3(a)所示. 如果残差图不是这种形状,呈现出明显的其他规律,则表明回归模型的某些假设不成立. 图 7.3(b)和(c)是两种较常见的情形. 其中图(b)表明随着 x 的值的增大,ε 的方差也增大,此时应采取加

权回归或对数据进行对数变换.图(c)表明假定的回归模型不能合理地描述变量间的关系,这时我们应考虑曲线回归模型或多元回归模型.

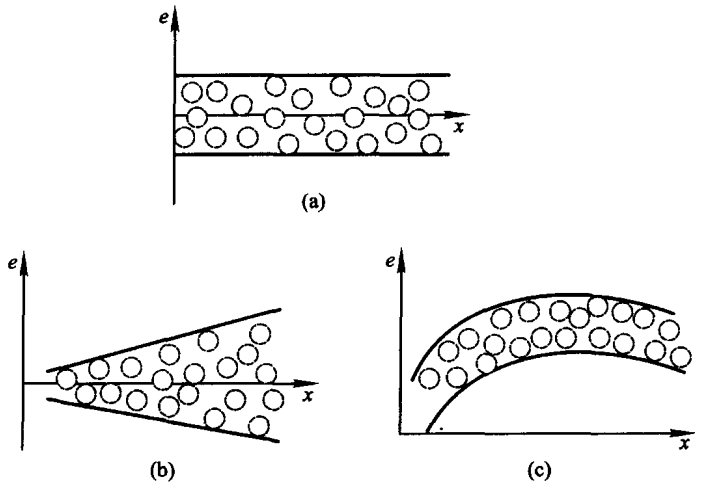


图 7.3 残差图的几种模式

另一种残差图是用横轴表示因变量的预测值 $\hat{Y}$ 和用纵轴表示残差值,即点的坐标为$(\hat{Y}_i, e_i)(i=1,2,\cdots,n)$.对于一元线性回归,关于 x 的残差图和关于 $\hat{Y}$ 的残差图提供了同样的信息.对于多元回归分析,因为有一个以上的自变量,所以一般采用关于 $\hat{Y}$ 的残差图.

二、标准化残差

残差减去它的平均值(零),再除以它的标准差就得到标准化残差,即

$$\tilde{e}_i = \frac{e_i}{\sqrt{\frac{SSE}{n-2}}\sqrt{1-h_i}}, \quad i=1,2,\cdots,n,$$

其中

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{l_{xx}}.$$

利用标准化残差可以对模型的正态假设进行检验.数学上可以证明,如果假设$(H_1)\sim(H_5)$成立,则标准化残差近似相互独立,且近似服从标准正态分布.换句话说,当样本量 n 比较大时,大约有 95% 的标准化残差介于 -2 和 $+2$ 之间.

另外一个考察正态性的方法是做标准化残差的正态 $Q-Q$ 图.具体做法是:

(1) 求出标准正态$\left(i-\frac{3}{8}\right)\Big/\left(n+\frac{1}{4}\right)$分位点 Z_i，即

$$\Phi(Z_i)=\left(i-\frac{3}{8}\right)\Big/\left(n+\frac{1}{4}\right),$$

Φ 为标准正态分布函数.

(2) 将标准化残差从小到大排序得 $\tilde{e}_{(1)}\leqslant\tilde{e}_{(2)}\leqslant\cdots\leqslant\tilde{e}_{(n)}$；

(3) 将$(Z_i,\tilde{e}_{(i)})(i=1,\cdots,n)$各点标在坐标平面上.

若图中各点基本上在一条直线附近，则认为没有充分的证据说明正态假设不成立.

三、Durbin－Watson 检验

在许多回归研究中，特别是在经济学的回归研究中，所利用的数据通常是按时间顺序采集的，即时间序列数据. 我们用 Y_t 表示 Y 在时刻 t 的值，而 Y_t 的值又常常依赖于 Y 在以前时刻的值. 这时我们说数据存在**自相关**(也叫**序列相关**). 当数据存在自相关时，违背了回归模型的假设，误差项 $\varepsilon_1,\cdots,\varepsilon_n$ 已不再是独立的.

我们可以用下面的 Durbin－Watson 统计量来检测自相关：

$$D=\frac{\sum_{t=2}^{n}(e_t-e_{t-1})^2}{\sum_{t=1}^{n}e_t^2}.$$

如果残差的相邻值彼此之间比较接近，那么 Durbin－Watson 统计量的值是比较小的. 此时存在正自相关性，表现在数据上就是：若 e_t 是正的，我们期待着下一个时刻的残差 e_{t+1} 也是正的. 若 e_t 是负的，则我们期待着下一个时刻的残差也是负的. 如果残差的相邻值彼此之间相距比较远，那么 Durbin－Watson 统计量的值将是比较大的. 此时存在负自相关性，表现在数据上就是残差是正负相间的，即若 e_t 是正的，则 e_{t+1} 是负的，e_{t+1} 又是正的，等等.

Durbin－Watson 检验统计量的取值介于 0 到 4 之间，一个粗略的判断是：若 D 的数值在 2 的附近(介于 1.6 和 2.4 之间)，则没有充分的证据说明存在自相关. 若 $D<1.6$，则存在正的自相关. 若 $D>2.4$，则存在负的自相关. Durbin－Watson 检验精确的临界值可查表获得，各种统计软件也一般会给出 D 的值.

§7.4 一元线性回归的预测和控制

如果根据样本估计出来的回归方程通过了显著性的检验和模型检查，那么

就可以利用这一回归方程进行预测和控制问题的研究.

一、预测问题

所谓的预测,就是给定自变量 x 的一个值 x_0,求 x_0 对应的因变量 Y_0 的过程.

譬如对例 7.1 来说,广告公司希望根据对 10 个厂家的广告费用与销售额的资料,进一步了解如果另外一个厂家对同一类产品投入了广告费用为 55 万元,该厂家的销售额是多少.这就是一个典型的预测问题.

根据在前面两节中对模型的设定,易知 Y_0 为一随机变量,而且

$$Y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0, \tag{7.60}$$

其中 $\varepsilon_0 \sim N(0,\sigma^2)$.

如果 β_0,β_1,σ^2 都是已知常数,那么此时 Y_0 的分布也是已知的,即 $Y_0 \sim N(\beta_0+\beta_1 x_0,\sigma^2)$.

进一步,如果此时需要找出一个数值作为 Y_0 的代表值,显然在均方误差最小的原则下,$E(Y_0 \mid x_0)=\beta_0+\beta_1 x_0$ 是惟一的选择,即

$$E([Y_0 - E(Y_0 \mid x_0)]^2) = \min_c E([Y_0 - c]^2). \tag{7.61}$$

但是,在实际问题当中,β_0,β_1,σ^2 都是未知的,正如 §7.1,§7.2 两节中所做的,只能利用样本对其进行估计.因此实际上的预测问题就是利用拟合得到的回归方程 $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1 x$,进而做以下两个方面的工作:

(1) 对 $E(Y_0 \mid x_0)=\beta_0+\beta_1 x_0$ 进行估计;

(2) 对 Y_0 进行直接的估计.

下面我们分别来看.

1. 对 $E(Y_0 \mid x_0)$ 的估计

设 $\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1 x_0$,则类似于(7.32)式可知

$$\hat{Y}_0 \sim N\left(\beta_0+\beta_1 x_0, \left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}\right)\cdot\sigma^2\right). \tag{7.62}$$

所以,$\hat{Y}_0$ 是 $E(Y_0 \mid x_0)$ 的无偏估计量.

为了求 $E(Y_0 \mid x_0)$ 的区间估计,由(7.62)式,易知

$$\frac{\hat{Y}_0 - (\beta_0+\beta_1 x_0)}{\sigma_{\hat{Y}_0}} \sim N(0,1), \tag{7.63}$$

其中 $\sigma_{\hat{Y}_0}$ 表示 $\hat{Y}_0$ 的标准差,即

$$\sigma_{\hat{Y}_0} = \sigma\cdot\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}. \tag{7.64}$$

又由于 SSE 与 $\hat{\beta}_0,\hat{\beta}_1$ 都独立以及 $\frac{SSE}{n-2}\sim\chi^2(n-2)$ 可知 SSE 与 $\hat{Y}_0$ 独立，而且

$$\frac{\hat{Y}_0-(\beta_0+\beta_1 x_0)}{\hat{\sigma}_{\hat{Y}_0}}\sim t(n-2), \tag{7.65}$$

其中

$$\hat{\sigma}_{\hat{Y}_0}=\sqrt{\frac{SSE}{n-2}}\cdot\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}} \tag{7.66}$$

是 $\sigma_{\hat{Y}_0}$ 的估计.

由(7.65)式，设自由度为 $n-2$ 的 t 分布的水平 $\frac{\alpha}{2}$ 上侧分位数为 $t_{\frac{\alpha}{2}}(n-2)$，那么 $E(Y_0\mid x_0)=\beta_0+\beta_1 x_0$ 的置信水平为 $1-\alpha$ 的置信区间为

$$\hat{Y}_0-\hat{\sigma}_{\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2)\leqslant E(Y_0\mid x_0)\leqslant\hat{Y}_0+\hat{\sigma}_{\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2). \tag{7.67}$$

容易看出，这一置信区间的长度是 $2\cdot\hat{\sigma}_{\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2)$，它是 x_0 的函数，特别地，当 $x_0=\bar{x}$ 时，即 x_0 取所有样本点的算术平均值时，该置信区间的长度最短.

2. 对 Y_0 的预测

正如前面提到的，Y_0 是一个随机变量，服从分布 $N(\beta_0+\beta_1 x_0,\sigma^2)$；$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1 x_0$ 是由样本值构成的统计量，其分布满足(7.62)式，显然 $\hat{Y}_0$ 也是 Y_0 的无偏预测.

那么用 $\hat{Y}_0$ 作为对 Y_0 的预测其预测精度是多少？这里的预测精度通常按照以下的方式衡量.给定 α 为一个小正数，求 δ 使之满足

$$P(|Y_0-\hat{Y}_0|<\delta)=1-\alpha,$$

则 δ 越小就说明 $\hat{Y}_0$ 的预测精度越高，并且称区间

$$(\hat{Y}_0-\delta,\hat{Y}_0+\delta)$$

为 Y_0 的概率为 $1-\alpha$ 的**预测区间**.考虑到 Y_0 与已有样本 $Y_1,Y_2,\cdots,Y_n$ 之间的独立性，易知 Y_0 与 $\hat{Y}_0$ 相互独立，所以 $Y_0-\hat{Y}_0$ 服从正态分布，而且

$$\begin{aligned}&E(Y_0-\hat{Y}_0)=0;\\&D(Y_0-\hat{Y}_0)=D(Y_0)+D(\hat{Y}_0)\\&\qquad=\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}\right)\cdot\sigma^2\triangleq\sigma^2_{Y_0-\hat{Y}_0},\end{aligned}$$

故有

$$\frac{Y_0-\hat{Y}_0}{\sigma_{Y_0-\hat{Y}_0}}\sim N(0,1). \tag{7.68}$$

同样地，利用 SSE 与 Y_0 及 $\hat{Y}_0$ 之间的独立性可知

$$\frac{Y_0-\hat{Y}_0}{\hat{\sigma}_{Y_0-\hat{Y}_0}}\sim t(n-2), \tag{7.69}$$

其中

$$\hat{\sigma}_{Y_0-\hat{Y}_0}=\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}\cdot\sqrt{\frac{SSE}{n-2}} \tag{7.70}$$

是 $\sigma_{Y_0-\hat{Y}_0}$ 的估计.

由(7.69)式可知

$$P\{|Y_0-\hat{Y}_0|\leqslant\hat{\sigma}_{Y-\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2)\}=1-\alpha. \tag{7.71}$$

因此可以得到 Y_0 的概率为 $1-\alpha$ 的预测区间是

$$(\hat{Y}_0-\hat{\sigma}_{Y_0-\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2),\hat{Y}_0+\hat{\sigma}_{Y_0-\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2)), \tag{7.72}$$

其预测精度可以用

$$\delta=\hat{\sigma}_{Y_0-\hat{Y}_0}\cdot t_{\frac{\alpha}{2}}(n-2)$$

来反映.

由(7.70)式可以看出，在样本给定的前提下，δ 是 x_0 的函数，不妨记之为 $\delta(x_0)$，x_0 越靠近样本的算术平均值 $\bar{x}$，δ 越小，表示预测精度越高，在 $x_0=\bar{x}$ 时，预测精度达到最高.见图 7.4.

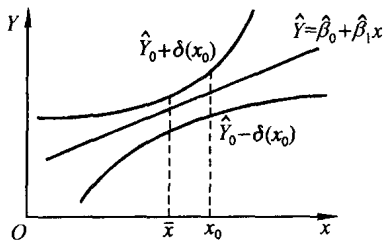


图 7.4　Y_0 的预测区间图示

另外，增加样本的容量 n 或者扩大 $x_1,x_2,\cdots,x_n$ 的分散程度以致使 l_{xx} 增加，都会导致 δ 减小，从而提高预测的精度.

特别地，如果 n 非常大，而且 x_0 在 $\bar{x}$ 附近取值时，近似地有 $\hat{\sigma}_{Y_0-\hat{Y}_0}\approx\sqrt{\frac{SSE}{n-2}}$，此时自由度为 $n-2$ 的 t 分布与 $N(0,1)$ 也会十分接近，因此可以用标准正态分布的 $\frac{\alpha}{2}$ 上侧分位数 $u_{\frac{\alpha}{2}}$ 来取代 $t_{\frac{\alpha}{2}}(n-2)$，这样一来，$Y_0$ 的概率为 $1-\alpha$

的预测区间就可以近似地记成

$$(\hat{Y}_0 - \hat{\sigma}\cdot u_{\frac{\alpha}{2}}, \hat{Y}_0 + \hat{\sigma}\cdot u_{\frac{\alpha}{2}}), \tag{7.73}$$

其中 $\hat{\sigma}=\sqrt{\frac{SSE}{n-2}}$ 是 σ 的估计.

在例 7.1 中如果一个厂家投入了 55 万元的广告费用,我们来对销售额进行预测.令 $x_0=55$,设 Y_0 为要预测的销售额.由拟合的回归方程可知

$$\hat{Y}_0 = 309.527\ 6 + 4.067\ 736\times 55 = 533.253,$$

因此 Y_0 的均值 $E(Y_0)$ 的估计值是 533.253 万元,它是对 Y_0 的平均水平的预测.$E(Y_0)$ 的 95% 的置信区间是(533.253 − 57.291 66,533.253 + 59.291 66) = (475.961 3,590.544 7).而且,由

$$\delta(55) = 113.239\ 6$$

可知 Y_0 的概率为 95% 的预测区间是(420.013 4,646.492 6).

在此例中,如果取 $x_0 = 36.5 = \bar{x}$,则此时对应的销售额的概率为 95% 的预测区间是(458 − 102.444 9,458 + 102.444 9) = (355.555 1,560.444 9),它是所有 95% 的预测区间中最短的一个.如果使用近似公式(7.73),则可以得到这一预测区间的近似表示是

$$(458 - 2\times 42.358, 458 + 2\times 42.358) = (373.284, 542.716),$$

其中采用了 $u_{0\ 025}\approx 2$.

二、控制问题

对于例 7.1 中考虑的问题,如果某个厂家感兴趣的是为了有 95% 的把握使得它的销售额高于 380 万元,那么该厂家至少应该投入多少广告费用.这就是一个控制问题.

控制问题可以看作是预测的反问题.它的较为一般的提法是,给定某一小正数 $\alpha(0<\alpha<1)$,如何控制自变量 x 的取值,才能以 $1-\alpha$ 的概率保证因变量 Y 落在某个预先指定的区间(y_L, y_U)之内.

利用图形可以直观地求解上述问题.在图 7.5 上,通过 Y 轴上的$(0, y_L)$,$(0, y_U)$两点各画一条平行于 x 轴的直线,并设它们分别交 $\hat{Y}-\delta$ 及 $\hat{Y}+\delta$ 于 L,U 两点,设 L,U 两点对应的横坐标分别为 x_L,x_U.容易看出,如果自变量 x 取 x_L 和 x_U 之间的任一值,相应的因变量 Y 的概率为 $1-\alpha$ 的预测区间都将包含在指定的区间(y_L, y_U)之内,因此,x_L 与 x_U 之间的任一 x 都符合控制问题的要求.

当然从这里也可以看出,预先指定的区间(y_L, y_U)的长度不能太短,至少应该不少于 $2\delta(x_L)$,即 $x = x_L$ 时对应的 Y 的概率为 $1-\alpha$ 的预测区间的长度,否则,比如取(y_L, y_U)小于该长度,则正如图 7.5 所示,x_L 与 x_U 之间的任一 x 值

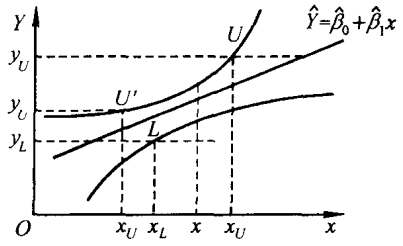


图 7.5 控制问题的直观求解

对应的 Y 的 $1-\alpha$ 预测区间都比(y_L, y_U)更大,都不是控制问题的解.此时按照上述回归模型无法求解该控制问题.

在实际问题当中,通常利用(7.73)式表示的预测区间的近似表达式来求解 x,即通过方程

$$\begin{cases} \hat{Y}(x_L) - \hat{\sigma} \cdot u_{\frac{\alpha}{2}} = y_L, \\ \hat{Y}(x_U) + \hat{\sigma} \cdot u_{\frac{\alpha}{2}} = y_U \end{cases}$$

可以解出

$$x_L = \frac{y_L + \hat{\sigma} \cdot u_{\frac{\alpha}{2}} - \hat{\beta}_0}{\hat{\beta}_1}, \tag{7.74}$$

$$x_U = \frac{y_U - \hat{\sigma} \cdot u_{\frac{\alpha}{2}} - \hat{\beta}_0}{\hat{\beta}_1}, \tag{7.75}$$

那么 x_L 与 x_U 之间的 x 即所求.

利用以上公式可以十分方便地求解本小节开始提出的广告费用的控制问题,即至少投入多少广告费用才能以 95% 的概率保证销售额不低于 380 万元.由公式(7.74)易得

$$x_L = \frac{380 + 2 \times 42.358 - 309.527\,6}{4.067\,736} \approx 38.15,$$

因此至少要投入 38.15 万元的广告费用.①

§7.5 一元非线性问题的线性化

在一元线性回归模型中,假设因变量 Y 对自变量 x 的回归函数是一个线性

① 使用 Y 的概率为 $1-\alpha$ 的单侧预测区间求解该问题更为直接和合理.参见有关文献.

函数(见(7.4)式),这意味着在样本数据的散点图上这些点将在一条直线的附近变动.但是在许多实际问题当中,Y 与 x 的关系不一定都是线性的,它们之间可能存在着某种复杂的非线性的联系,表现为散点图上的点围绕着某条曲线波动.请看下面的例子.

例 7.2 经过调查得到 8 个厂家对同种类型的产品年新增加投资额和年利润额的数据资料,如表 7.5 所示.

表 7.5 8 个厂家年投资额与利润数据资料

厂家	1	2	3	4	5	6	7	8
年新增投资额 x(万元)	4	6	10	11	15	17	18	20
年利润额 Y(万元)	6	7	9	10	17	24	23	26
ln Y	1.79	1.95	2.20	2.30	2.83	3.18	3.14	3.26

图 7.6 给出了年利润额 Y 与年新增加投资额 x 的散点图,从图中可以清楚地看出来,随着 x 的增大 Y 也有明显的增加的趋势,因此两者之间存在着相关关系,但是这种相关关系与其用一条直线来描述倒不如用曲线描述更加合适,因此 Y 与 x 之间更加倾向于被认为是一种非线性关系.回归方程也需要用一些非线性函数来刻画,比如

$$\hat{Y} = \hat{\beta}_0 \cdot e^{\beta_1 x} \tag{7.76}$$

或者

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x^2 \tag{7.77}$$

等等.图 7.7 给出的是变量 ln Y 与变量 x 的散点图,从中可以看出这些点基本上是围绕一条直线波动,说明变量 ln Y 与 x 之间近似是一种线性关系,从而也印证了回归方程取(7.76)形式的合理性.

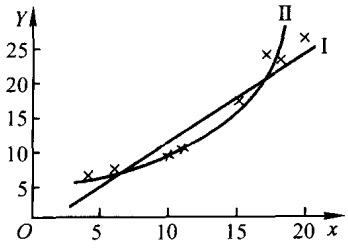


图 7.6 年投资额与利润数据的散点图

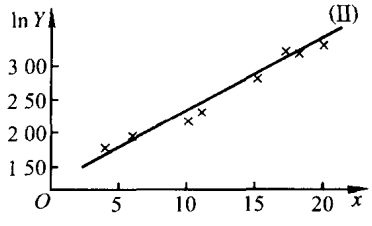


图 7.7 经过对数变换后的散点图

同时,图 7.7 也提示给我们一种求解回归方程(7.76)的思路,即通过求解变量 ln Y 对 x 的线性回归方程即可得到相应的(7.76)式所表示的 Y 对 x 的回归

方程.具体来说,首先对样本数据$(x_i, Y_i)(i=1,2,\cdots,n)$作对数变换

$$Z_i = \ln Y_i, i=1,2,\cdots,n, \tag{7.78}$$

然后利用最小二乘法求出变量 Z 对 x 的回归方程

$$\hat{Z} = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot x, \tag{7.79}$$

即图 7.7 中的直线方程,则相应的形如(7.76)式的 Y 对 x 的回归方程是

$$\hat{Y} = e^{\hat{Z}} = e^{\hat{\alpha}_0} \cdot e^{\hat{\alpha}_1 x}, \tag{7.80}$$

即 $\hat{\beta}_0 = e^{\hat{\alpha}_0}, \hat{\beta}_1 = \hat{\alpha}_1$.

利用表 7.5 中给出的数据,可以得到 $\ln Y$ 对 x 的线性回归方程是

$$\hat{Z} = 1.313\,9 \quad + \quad 0.100\,3 \quad x,$$
$$(14.234) \qquad (14.934)$$

其中下面括号中的数是对应系数的 t 值.由此可得 Y 对 x 的回归方程是

$$\hat{Y} = 3.720\,8 \cdot e^{0.100\,3x}. \tag{7.81}$$

如果采用形如(7.77)式的抛物线型回归方程,容易看出,令 $w = x^2$,则(7.77)式可表示为变量 Y 对 w 的线性回归方程

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot w. \tag{7.82}$$

对样本数据做变换 $w_i = x_i^2 (i=1,2,\cdots,n)$,利用$(w_i, Y_i)(i=1,2,\cdots,n)$解出(7.82)中的系数估计值 $\hat{\beta}_0, \hat{\beta}_1$,代入(7.77)式即得到 Y 对 x 的回归方程.

对表 7-4 中的数据计算结果为

$$\hat{Y} = 4.413 \quad + \quad 0.057 \quad x^2 \tag{7.83}$$
$$(4.547) \qquad (13.630)$$

通过上面的例子可以看出,对于一些常见的非线性回归方程,可以通过对样本进行适当的变换将其化成新的变量之间的线性回归方程,这种求解思路除了适用于(7.76)、(7.77)两种类型的非线性函数形式外,还适用于以下常见的函数.

1. 双曲线函数

函数形式为

$$\frac{1}{Y} = a + b \cdot \frac{1}{x}$$

或

$$Y = \frac{x}{ax + b},$$

令 $Z = \frac{1}{Y}, w = \frac{1}{x}$,则

$$Z = a + bw.$$

因此,需对原始样本数据(x_i, Y_i)进行变换$\left(\frac{1}{x_i}, \frac{1}{Y_i}\right)(i=1,2,\cdots,n)$,利用后者

求解$\frac{1}{Y}$对$\frac{1}{x}$之间的线性回归方程,即可得到 Y 对 x 的非线性回归方程的形式.

以下其他函数形式的回归方程与之类似,因此只给出相应的变换形式.

2. 对数曲线函数

函数形式为

$$Y = a + b \cdot \ln x,$$

令 $w = \ln x$,则

$$Y = a + b \cdot w.$$

3. 幂函数形式

函数形式为

$$Y = a \cdot x^b,$$

令 $Z = \ln Y, \nu = \ln a, w = \ln x$,则

$$Z = \nu + b \cdot w.$$

4. S 形曲线函数

函数形式为

$$Y = \frac{1}{a + b\mathrm{e}^{-x}},$$

令 $Z = \frac{1}{Y}, w = \mathrm{e}^{-x}$ 则

$$Z = a + b \cdot w.$$

对回归方程选择一种合适的函数形式,必须事先对散点图进行认真的分析,此时对上述各种函数类型所描述的曲线形状有充分的了解是十分必要的.我们在图 7.8 中给出了上述几种函数的图示,可供参考.

正如对例 7.2 的分析中所看到的,对同一种散点图所呈现的 Y 与 x 之间的关系,可以选择不同的函数形式来描述回归方程.那么如何判断并比较不同回归方程的拟合优度呢?

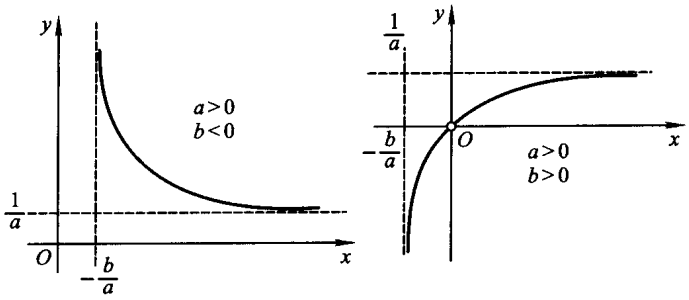
通常使用的比较准则有下面两个.

1. 相关指数 R

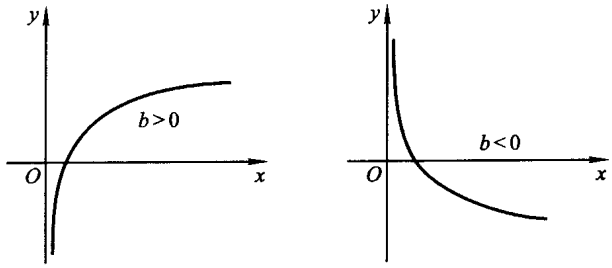
定义 R 是

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{7.84}$$

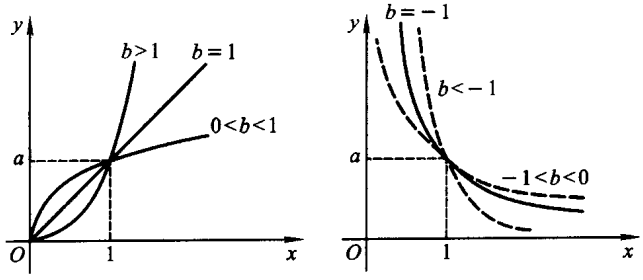
的平方根,上式与线性回归模型中定义的可决系数是一致的,等号右边的第二项表示了样本点对回归曲线的偏离占样本点总的离散程度的比重,如果回归曲线对样本点拟合得很好,这一项自然就会比较小,从而 R^2 较大,因此对于不同的曲



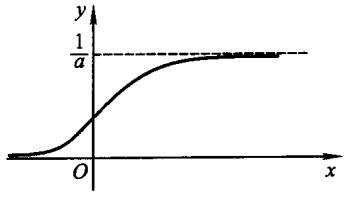
(a) 双曲函数 $\frac{1}{y}=a+b\cdot\frac{1}{x}$



(b) 对数函数 $y=a+b\cdot\ln x$



(c) 幂函数 $y=a\cdot x^b$



(d) 双曲函数 $y=\frac{1}{a+b\cdot e^{-x}}$

图 7.8 常用的曲线形式

线回归方程通常选择 R^2 较大的一个.

2. 剩余标准差 S

定义 S 为

$$S=\sqrt{\frac{\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2}{n-2}}, \tag{7.85}$$

它反映了样本点偏离回归曲线的平均大小,当然 S 小一点比较好.

从(7.84)、(7.85)两个式子中不难发现上述两个准则是一致的,即 R 越大则 S 越小,反之亦然.

对于前面利用例 7.2 中的数据得到的两个曲线回归方程即指数形式(7.81)和抛物线形式(7.83),分别计算它们的相关指数 R 和剩余标准差 S,计算结果列在了表 7.5 中.为了便于比较,同时还列出回归方程为线性函数时所对应的 R 与 S 的值.

从表 7.6 中可以清楚地看出,三种不同形式的回归方程当中抛物线型方程对应的相关指数最大,而剩余标准差最小,因此相比之下抛物线型方程是最可取的.直线回归方程的拟合效果在三者之中是最差的,这也进一步说明了选择适当的非线性回归模型的必要性.

表 7.6 三种不同回归方程的比较

曲线形式	回归方程	相关指数 R	剩余标准差 S
直线	$\hat{Y}=-2.02+1.37x$	0.963 7	2.376 2
指数曲线	$\hat{Y}=3.720\ 8\cdot e^{0.100\ 3x}$	0.980 4	1.752 2
抛物线	$\hat{Y}=4.413\ 2+0.057\ 4x^2$	0.984 2	1.574 6

§7.6 多元线性回归分析

在许多实际问题当中,往往需要研究多个变量之间的相关关系,比如某种产品的销售额不仅受到投入的广告费用的影响,通常还与产品的价格、消费者的收入状况、社会保有量以及其他可替代产品的价格等诸多因素有关系,研究一个变量同其他多个变量之间的关系的主要方法是运用多元回归分析.多元线性回归分析是一元线性回归分析的自然推广形式,两者在参数估计、显著性检验等技术方面是非常相似的.

一、多元线性回归模型

设影响因变量 Y 的自变量有 p 个，并分别记为 $x_1, x_2, \cdots, x_p$，所谓**多元线性模型**是指这些自变量对 Y 的影响是线性的，即

$$Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_p \cdot x_p + \varepsilon, \tag{7.86}$$

其中

$$f(x_1, x_2, \cdots, x_p) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_p \cdot x_p. \tag{7.87}$$

并称 Y 为 p 个自变量 $x_1, x_2, \cdots, x_p$ 的**线性回归函数**. 类似于一元线性回归，在这里我们也假定 p 个自变量都是确定变量.

记 n 组样本分别是 $(x_{i1}, x_{i2}, \cdots, x_{ip}; Y_i)(i = 1, 2, \cdots, n)$. 那么由(7.86)式得到

$$Y_i = \beta_0 + \beta_i x_{i1} + \beta_2 \cdot x_{i2} + \cdots + \beta_p \cdot x_{ip} + \varepsilon_i, i = 1, 2, \cdots, n. \tag{7.88}$$

类似于一元情形，通常假设

$$E(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2, i = 1, 2, \cdots, n, \tag{7.89}$$

而且 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 相互独立. (7.90)

为了检验回归模型及其系数的显著性，通常还假设 $\varepsilon_i (i = 1, 2, \cdots, n)$ 服从正态分布，此时(7.89)、(7.90)等价于

$$\varepsilon_i \sim N(0, \sigma^2), \text{且 } \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, (i \neq j, i, j = 1, 2, \cdots, n). \tag{7.91}$$

这样一来，多元线性回归分析的主要问题就是基于模型(7.88)、(7.91)对于未知参数 $\beta_0, \beta_1, \cdots, \beta_p$ 和 σ^2 进行估计、检验并利用得到的回归模型进行预测等.

为了表达的简洁起见，通常使用矩阵将上述模型(7.88)、(7.91)统一记成

$$\begin{cases} \boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\varepsilon} \sim N(\boldsymbol{O}, \sigma^2 \boldsymbol{I}_n). \end{cases} \tag{7.92}$$

其中

$$\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)^{\mathrm{T}},$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^{\mathrm{T}},$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}},$$

$\boldsymbol{I}_n$ 是 n 阶单位阵，$\boldsymbol{O}$ 表示 n 维零向量.

二、回归系数的最小二乘估计

利用最小二乘法估计回归系数即是求解一组 $\beta_0,\beta_1,\cdots,\beta_p$ 使如下定义的平方和 Q 达到最小：

$$Q(\beta_0,\beta_1,\cdots,\beta_p)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1x_{i1}-\cdots-\beta_p\cdot x_{ip})^2 \tag{7.93}$$

即

$$Q(\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_p)=\min_{\beta_0,\beta_1,\cdots,\beta_p}Q(\beta_0,\beta_1,\cdots,\beta_p).$$

由多元函数微分学的知识可知 Q 在$(\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_p)$点的各个一阶偏导数均为零，于是 $\hat\beta_0,\hat\beta_1,\hat\beta_2,\cdots,\hat\beta_p$ 满足以下 $p+1$ 个方程：

$$\begin{cases}-2\cdot\sum\limits_{i=1}^{n}(Y_i-\hat\beta_0-\hat\beta_1x_{i1}-\cdots-\hat\beta_p\cdot x_{ip})=0,\\ -2\cdot\sum\limits_{i=1}^{n}(Y_i-\hat\beta_0-\hat\beta_1x_{i1}-\cdots-\hat\beta_p\cdot x_{ip})\cdot x_{i1}=0,\\ \cdots\cdots\cdots\cdots\\ -2\cdot\sum\limits_{i=1}^{n}(Y_i-\hat\beta_0-\hat\beta_1x_{i1}-\cdots-\hat\beta_p\cdot x_{ip})\cdot x_{ip}=0.\end{cases}$$

整理该方程组可得

$$\begin{cases}n\hat\beta_0+\sum\limits_{i=1}^{n}x_{i1}\hat\beta_1+\cdots+\sum\limits_{i=1}^{n}x_{ip}\hat\beta_p=\sum\limits_{i=1}^{n}Y_i,\\ \sum\limits_{i=1}^{n}x_{i1}\hat\beta_0+\sum\limits_{i=1}^{n}x_{i1}^2\hat\beta_1+\cdots+\sum\limits_{i=1}^{n}x_{i1}x_{ip}\hat\beta_p=\sum\limits_{i=1}^{n}x_{i1}Y_i,\\ \cdots\cdots\cdots\cdots\\ \sum\limits_{i=1}^{n}x_{ip}\hat\beta_0+\sum\limits_{i=1}^{n}x_{ip}x_{i1}\hat\beta_1+\cdots+\sum\limits_{i=1}^{n}x_{ip}^2\hat\beta_p=\sum\limits_{i=1}^{n}x_{ip}Y_i.\end{cases} \tag{7.94}$$

上述方程组又称为**正规方程组**，用矩阵形式可以表示为

$$(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\cdot\hat{\boldsymbol{\beta}}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}, \tag{7.95}$$

其中 $\hat{\boldsymbol{\beta}}=(\hat\beta_0,\hat\beta_1,\cdots,\hat\beta_p)^{\mathrm{T}}$，如果矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 可逆①，则正规方程组的解是

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}. \tag{7.96}$$

经常还用到正规方程的另一种形式，考虑到(7.94)中的第一式可以写成

$$\hat\beta_0=\overline{Y}-\bar x_1\hat\beta_1-\bar x_2\hat\beta_2-\cdots-\bar x_p\hat\beta_p, \tag{7.97}$$

其中 $\overline{Y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}Y_i,\bar x_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_{ik}(k=1,2,\cdots,p)$，将其代入其余 p 个式子可

① 由于一般情况下 $n>p$，所以只要 $\boldsymbol{X}$ 为列满秩矩阵即可保证 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 可逆

以得到

$$\begin{cases} l_{11}\hat{\beta}_1 + l_{12}\cdot\hat{\beta}_2 + \cdots + l_{1p}\hat{\beta}_p = l_{1Y}, \\ l_{21}\hat{\beta}_1 + l_{22}\cdot\hat{\beta}_2 + \cdots + l_{2p}\hat{\beta}_p = l_{2Y}, \\ \cdots\cdots\cdots\cdots \\ l_{p1}\hat{\beta}_1 + l_{p2}\cdot\hat{\beta}_2 + \cdots + l_{pp}\hat{\beta}_p = l_{pY}, \end{cases} \tag{7.98}$$

其中

$$l_{kj} = \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)\cdot(x_{ij} - \bar{x}_j), k, j = 1,2,\cdots,p;$$

$$l_{kY} = \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)\cdot(Y_i - \bar{Y}), k = 1,2,\cdots,p.$$

如果以 $\boldsymbol{L}$ 表示(7.98)的系数矩阵,即 $\boldsymbol{L} = (l_{kj})_{p\times p}$,以 $\boldsymbol{l}_Y$ 表示(7.98)式等号右边的常数项,$\boldsymbol{l}_Y = (l_{1Y}, l_{2Y}, \cdots, l_{pY})^{\mathrm{T}}$,则(7.98)可以简记为

$$\boldsymbol{L}\cdot\hat{\boldsymbol{\beta}}^* = \boldsymbol{l}_Y, \tag{7.99}$$

在这里以 $\hat{\boldsymbol{\beta}}^*$ 表示向量$(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p)^{\mathrm{T}}$.如果 $\boldsymbol{L}$ 可逆,则

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{L}^{-1}\cdot\boldsymbol{l}_Y, \tag{7.100}$$

将其代入(7.97)式即可得到 $\hat{\beta}_0$ 的值.

易知这样得到的 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ 与(7.96)是一致的,而且的确是 Q 的最小值点,我们称这样得到的 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ 是回归系数 $\beta_0, \beta_1, \cdots, \beta_p$ 的最小二乘估计.而

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\cdot x_1 + \cdots + \hat{\beta}_p\cdot x_p \tag{7.101}$$

就是拟合的回归方程.

类似地也可以证明所谓的高斯—马尔可夫定理,即在模型(7.92)的假设条件下,由最小二乘法得到的 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ 分别是 $\beta_0, \beta_1, \cdots, \beta_p$ 的最优线性无偏估计(BLUE).

三、回归方程的显著性检验

回归方程的显著性检验用来考察所选用的自变量 $x_1, x_2, \cdots, x_p$ 的线性函数是否的确对因变量 Y 起到了解释作用,即要检验假设

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0. \tag{7.102}$$

如果 H_0 为真,说明 Y 没有受到这些自变量中任何一个的线性影响,此时回归方程是没有意义的.

以 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1\cdot x_{i1} + \cdots + \hat{\beta}_p\cdot x_{ip}(i = 1,2,\cdots,n)$ 表示第 i 组样本对应的拟合值,类似于§7.2中对一元线性回归问题分析的思路,进行如下的离差平方和分解:

$$\begin{aligned} SST &= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \\ &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \\ &\xlongequal{\text{def}} SSE + SSR, \end{aligned}$$

其中利用到等式

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i) \cdot (\hat{Y}_i - \bar{Y}) = 0,$$

该等式易由正规方程组(7.94)推知.

这里的 SSR, SSE 分别称为回归平方和、残差平方和,各自的含义也与§7.2中的分析相似.而且,可以证明,在 H_0 成立时,如下定义的统计量

$$F = \frac{SSR/p}{SSE/(n-p-1)} \tag{7.103}$$

服从自由度为 p 和 $n-p-1$ 的 F 分布.

给定显著性水平 α,令 $F_\alpha(p, n-p-1)$ 为自由度为 p 和 $n-p-1$ 的 F 分布的水平 α 的上侧分位数,则可设定拒绝域为满足条件

$$F > F_\alpha(p, n-p-1)$$

的样本集合.即可得到 H_0 的一个显著性水平为 α 的检验.

表7.7给出的方差分析表是对上述分析过程的一个总结.

表 7.7 多元线性回归方程显著性检验的方差分析表

方差来源	平 方 和	自由度	均方和	F 值
回归	$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$	p	$MSR = \frac{SSR}{p}$	$F = \frac{MSR}{MSE}$
残差	$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$	$n-p-1$	$MSE = \frac{SSE}{n-p-1}$	
总计	$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$	$n-1$		

同样地,称

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

为**可决系数**,易知 $0 \leqslant R^2 \leqslant 1$, R^2 越接近于1则说明拟合值的离差平方和占样本总的离差平方和的比重越大,因此 R^2 是拟合优度的度量, R^2 的平方根 R 通常称为**复相关系数**.类似于度量两个变量之间的线性相关程度的直线相关系数, R 可以看作是 Y 与 $x_1, x_2, \cdots, x_p$ 之间相关关系及其密切程度的一种度量.

四、回归系数的检验

对于多元线性回归模型来说，回归方程的显著性检验被通过（即 H_0 (7.102)被拒绝）只能说明 p 个自变量在整体上对 Y 是有影响的，但是这并不意味着模型中每一个自变量都是必需的，因此，还需要进一步对每一个自变量前面的系数进行检验.

譬如为了考察变量 x_j 对 Y 的影响是否显著，就需要对如下的原假设进行检验.

$$H_0: \beta_j = 0. \tag{7.104}$$

如果在给定的显著性水平下不能拒绝 H_0，说明 x_j 对 Y 影响并不显著，可以考虑在回归方程中去掉 x_j.

由于 β_j 的最小二乘估计量为 $\hat{\beta}_j$，故可以选择 H_0 的拒绝域为满足条件 $|\hat{\beta}_j| > c$ 的样本集合，其中 c 为某一待定的正常数. 可以证明，在 H_0 式成立的条件下，构造统计量

$$F_j = \frac{\hat{\beta}_j^2 / l^{jj}}{SSE/n-p-1} \tag{7.105}$$

和

$$t_j = \frac{\hat{\beta}_j / \sqrt{l^{jj}}}{\sqrt{SSE/n-p-1}}, \tag{7.106}$$

其中 l^{jj} 是矩阵 $\boldsymbol{L}^{-1}$ 对角线上的第 j 个元素，则

$$F_j \sim F(1, n-p-1),$$

$$t_j \sim t(n-p-1).$$

因此，给定显著性水平 α，令 $F_\alpha(1, n-p-1)$ 表示 $F(1, n-p-1)$ 分布的 α 上侧分位数，令 $t_{\frac{\alpha}{2}}(n-p-1)$ 表示 $t(n-p-1)$ 分布的 $\frac{\alpha}{2}$ 上侧分位数，所以可以将 H_0 的拒绝域设定为满足条件

$$F_j > F_\alpha(1, n-p-1) \tag{7.107}$$

或者满足条件

$$|t_j| > t_{\frac{\alpha}{2}}(n-p-1) \tag{7.108}$$

的样本集合，都可以得到 H_0 的一个显著性水平为 α 的检验. 通常称满足(7.107)式的拒绝域为 F **检验**，称满足(7.108)式的拒绝域为 t **检验**. 对于(7.104)所设定的 H_0 而言，F 检验和 t 检验是等价的.

在实际问题中，如果有多个回归系数不显著，则不能同时将相应的自变量在回归方程中去除掉，而只能先将 F_j 值（或者 $|t_j|$ 值）最小的一个自变量删除，再

对剩余的 $p-1$ 个自变量重新建立回归方程，并进行系数的显著性检验，如此直至所有变量的系数都显著为止.

五、多元线性回归模型的预测

假设经过回归方程和回归系数的显著性检验之后最终得到的拟合方程是 $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1\cdot x_1+\cdots+\hat{\beta}_p\cdot x_p$，那么任意给定一组自变量的值$(x_{01},x_{02},\cdots,x_{0p})$，并令 Y_0 是该组值对应的因变量的值，即

$$Y_0=\beta_0+\beta_1\cdot x_{01}+\cdots+\beta_p\cdot x_{0p}+\varepsilon_0,$$

则

$$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1\cdot x_{01}+\cdots+\hat{\beta}_p\cdot x_{0p} \tag{7.109}$$

是 $E(Y_0)$的无偏估计.

给定 $0<\alpha<1$，$E(Y_0)$的置信水平为 $1-\alpha$ 的置信区间为

$$(\hat{Y}_0-\delta_0\cdot\hat{\sigma}\cdot t_{\frac{\alpha}{2}}(n-p-1),\hat{Y}_0+\delta_0\cdot\hat{\sigma}\cdot t_{\frac{\alpha}{2}}(n-p-1)), \tag{7.110}$$

其中

$$\delta_0=\sqrt{\frac{1}{n}+\sum_{k=1}^{n}\sum_{j=1}^{n}l^{kj}\cdot(x_{0k}-\bar{x}_k)\cdot(x_{0j}-\bar{x}_j)},$$

$$\hat{\sigma}=\sqrt{\frac{SSE}{n-p-1}}\text{是 }\sigma\text{ 的估计},$$

而 l^{kj} 是矩阵 $\boldsymbol{L}^{-1}$ 中第 k 行第 j 列元素，$k,j=1,2,\cdots,p$.

类似地，对于给定的 $0<\alpha<1$，还可以考虑 Y_0 的概率为 $1-\alpha$ 的预测区间，利用

$$\frac{Y_0-\hat{Y}_0}{\delta\cdot\hat{\sigma}}\sim t(n-p-1), \tag{7.111}$$

其中

$$\delta=\sqrt{1+\frac{1}{n}+\sum_{k=1}^{n}\sum_{j=1}^{n}l^{kj}\cdot(x_{0k}-\bar{x}_k)(x_{0j}-\bar{x}_j)},$$

可得 Y_0 的概率为 $1-\alpha$ 的预测区间是

$$(\hat{Y}_0-\delta\cdot\hat{\sigma}\cdot t_{\frac{\alpha}{2}}(n-p-1),\hat{Y}_0+\delta\cdot\hat{\sigma}\cdot t_{\frac{\alpha}{2}}(n-p-1)), \tag{7.112}$$

它要比 $E(Y_0)$的 $1-\alpha$ 的置信区间大一些.

如果 n 很大，而且给定的自变量的值$(x_{01},x_{02},\cdots,x_{0p})$非常接近于样本的中间位置，即$(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_p)$，那么 Y_0 的 $1-\alpha$ 的预测区间可以近似地表示成

$$(\hat{Y}_0-\hat{\sigma}\cdot u_{\frac{\alpha}{2}},\hat{Y}_0+\hat{\sigma}\cdot u_{\frac{\alpha}{2}}), \tag{7.113}$$

其中 $u_{\frac{\alpha}{2}}$ 是标准正态分布的 $\frac{\alpha}{2}$ 上侧分位数.

附录 7.1 命题 7.1 的证明

(1) $E(SSR)=E(\hat{\beta}_1^2\cdot l_{xx})=l_{xx}\cdot E(\hat{\beta}_1^2)$

$$=l_{xx}\cdot[D(\hat{\beta}_1)+(E(\hat{\beta}_1))^2]$$

$$=l_{xx}\cdot\left[\frac{\sigma^2}{l_{xx}}+\beta_1^2\right]$$

$$=\sigma^2+\beta_1^2\cdot l_{xx}.$$

(2) $SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$

$$=\sum_{i=1}^{n}(\beta_0+\beta_1x_i+\varepsilon_i-\hat{\beta}_0-\hat{\beta}_i\cdot x_i)^2$$

$$=\sum_{i=1}^{n}[(\beta_0-\hat{\beta}_0)+(\beta_1-\hat{\beta}_1)x_i+\varepsilon_i]^2$$

$$=\sum_{i=1}^{n}[(\beta_0-\hat{\beta}_0)^2+x_i^2\cdot(\beta_1-\hat{\beta}_1)^2+\varepsilon_i^2+2\cdot(\beta_0-\hat{\beta}_0)\cdot\varepsilon_i+$$

$$2x_i(\beta_1-\hat{\beta}_1)\cdot\varepsilon_i+2\cdot x_i\cdot(\beta_0-\hat{\beta}_0)\cdot(\beta_1-\hat{\beta}_1)]$$

$$=n(\beta_0-\hat{\beta}_0)^2+(\beta_1-\hat{\beta}_1)^2\cdot\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}\varepsilon_i^2+2\sum_{i=1}^{n}(\beta_0-\hat{\beta}_0)\cdot\varepsilon_i$$

$$+2\sum_{i=1}^{n}x_i\cdot(\beta_1-\hat{\beta}_1)\cdot\varepsilon_i+2\sum_{i=1}^{n}x_i\cdot(\beta_0-\hat{\beta}_0)\cdot(\beta_1-\hat{\beta}_1),$$

而且,由(7.24)和(7.21)式,得

$$E[(\beta_0-\hat{\beta}_0)\cdot\varepsilon_i]=-E(\hat{\beta}_0\cdot\varepsilon_i)$$

$$=-\sum_{j=1}^{n}\left(\frac{1}{n}-\frac{(x_j-\bar{x})\cdot\bar{x}}{l_{xx}}\right)E(Y_j\cdot\varepsilon_i)$$

$$=-\left[\frac{1}{n}-\frac{(x_i-\bar{x})\cdot\bar{x}}{l_{xx}}\right]\sigma^2;$$

$$E[x_i\cdot(\beta_1-\hat{\beta}_1)\cdot\varepsilon_i]=-x_i\cdot E(\hat{\beta}_1\cdot\varepsilon_i)$$

$$=-x_i\cdot\sum_{j=1}^{n}\frac{(x_j-\bar{x})}{l_{xx}}\cdot E(Y_j\varepsilon_i)$$

$$=-\frac{(x_i-\bar{x})x_i}{l_{xx}}\cdot\sigma^2.$$

故有

$E(SSE)$

$$
\begin{aligned}
&= nD(\hat{\beta}_0) + D(\hat{\beta}_1)\cdot\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} D(\varepsilon_i) \\
&\quad - 2\sum_{i=1}^{n}\left[\frac{1}{n} - \frac{(x_i-\bar{x})\cdot\bar{x}}{l_{xx}}\right]\sigma^2 - 2\sum_{i=1}^{n}\frac{(x_i-\bar{x})\cdot x_i}{l_{xx}}\cdot\sigma^2 \\
&\quad + 2\sum_{i=1}^{n} x_i\cdot\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1) \\
&= \left[n\left(\frac{1}{n}+\frac{\bar{x}^2}{l_{xx}}\right) + \sum_{i=1}^{n}\frac{x_i^2}{l_{xx}} + n - 2 + 2\sum_{i=1}^{n}\frac{(x_i-\bar{x})\cdot\bar{x}}{l_{xx}}\right. \\
&\quad \left. - 2\sum_{i=1}^{n}\frac{(x_i-\bar{x})\cdot x_i}{l_{xx}} - 2\sum_{i=1}^{n}\frac{x_i\cdot\bar{x}}{l_{xx}}\right]\cdot\sigma^2. \\
&= \left[n - 1 + \left(n\bar{x}^2 + \sum_{i=1}^{n}x_i^2 + 2\sum_{i=1}^{n}(x_i-\bar{x})\cdot\bar{x}\right.\right. \\
&\quad \left.\left. - 2\sum_{i=1}^{n}(x_i-\bar{x})\cdot x_i - 2\sum_{i=1}^{n}x_i\cdot\bar{x}\right)\cdot\frac{1}{l_{xx}}\right]\sigma^2 \\
&= (n-2)\sigma^2.
\end{aligned}
$$

习 题 七

(A)

1. 验证(7.25)及(7.26).

2. 证明离差平方和分解(7.34).

3. 在某个地区随机抽取 9 个家庭,调查得到每个家庭每月人均收入与食品消费支出的数据如下表所示(单位:元):

编 号	1	2	3	4	5	6	7	8	9
人均收入	102	124	156	182	194	229	272	342	495
食品支出	87	92	90	124	150	160	182	216	220

根据上面数据,试求:

(1) 人均收入与食品消费支出的相关系数;

(2) 建立食品支出 Y 对人均收入 x 的一元线性回归模型,并对模型的显著性进行检验;

(3) 根据以上建立的模型,试求该地区一个人均收入为每月 300 元的家庭人均食品支出的概率为 95% 的预测区间.

4. 为了考察某班学生的学习成绩,随机抽取了七名同学,他们两个学期各门课程的平均成绩记录在下表中:

利用下述数据建立第二学期各科平均成绩对第一学期各科平均成绩的一元线性回归模型,并对模型的显著性进行检验(取显著性水平 $\alpha=0.05$),据此你可以得出什么结论?

学生编号	1	2	3	4	5	6	7
第一学期	76	69	82	92	88	70	84
第二学期	80	63	84	90	91	76	71

5. 假设某种钢材的硬度与含铜量百分比以及温度之间服从线性关系,下面给出六次试验的数据资料.

钢材硬度 Y	含铜量(%)x_1	温度 x_2
78.9	0.02	1 000
55.2	0.02	1 200
80.9	0.10	1 000
57.4	0.10	1 200
85.3	0.18	1 000
60.7	0.18	1 200

根据上面数据,

(1) 建立硬度对含铜量及温度的二元线性回归方程;

(2) 对回归方程的显著性进行检验($\alpha=0.05$);

(3) 对回归方程的系数进行显著性检验($\alpha=0.05$),并对系数的含义进行解释.

6. 随机选择了10个地区对某种产品一周之内的销售额 Y,广告费用 x_1 进行调查,数据列在下面的表中,同时列出的还有每个地区的人口密度 x_2,试建立 Y 对 x_1, x_2 的线性回归模型.

地区编号	销售额 Y (千元)	广告费 x_1 (千元)	人口密度 x_2 (人/平方公里)
1	20	0.2	50
2	25	0.2	50
3	24	0.2	50
4	30	0.3	60
5	32	0.3	60
6	40	0.4	70
7	28	0.3	50
8	50	0.5	75
9	40	0.4	70
10	50	0.5	74

7. 下面的表中列出了分别对变量 x 与变量 y 独立进行 8 次观测的结果：

x	3	5	9	13	16	21	27	30
y	102	97	110	128	172	213	326	450

根据以上数据，

(1) 画出散点图；

(2) 选择两种曲线形式建立 y 对 x 的曲线回归模型，并对两种模型的拟合优度进行比较.

(B)

1. 对于一元线性回归模型

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \cdots, n$$

在满足条件(H_1)~(H_5)的情况下，试求 β_0, β_1 以及 σ^2 的最大似然估计.

2. 试求回归系数 $\hat{\beta}_1$ 与相关系数 γ 之间的关系并由此给出 γ 的一种解释.

3. 如果存在 δ 使得

$$P(Y_0 \geqslant \hat{Y}_0 - \delta) = 1 - \alpha$$

成立，则称$(\hat{Y}_0 - \delta, +\infty)$为 Y_0 的概率为 $1-\alpha$ 的**单(上)侧预测区间**；类似地可以定义其下侧预测区间. 给定 α，试求 Y_0 的上下两个单侧预测区间的表达式.

4. 对于(7.95)和(7.99)式，试证明矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 可逆当且仅当 $\boldsymbol{L}$ 可逆.

习题参考答案

习 题 一

(A)

9. 0.15,0.5,0.10,0.5.

10. 0.7

11. (1) $\dfrac{C_{N_1}^{k} C_{N-N_1}^{n-k}}{C_N^n}$; (2) $1-\dfrac{C_{N-N_1}^{n}}{C_N^n}$;

(3) $1-\dfrac{C_{N-N_1}^{n}}{C_N^n}-\dfrac{C_{N_1}^{1} C_{N-N_1}^{n-1}}{C_N^n}$.

12. (1) $\dfrac{24!}{P_{30}^{24}}$ $\left(\text{或}\dfrac{1}{C_{30}^{24}}\right)$; (2) $\dfrac{1}{2}$.

13. $1-\dfrac{P_n^{365}}{365^n}$

14. (1) $1-\left(\dfrac{8}{9}\right)^{25}$; (2) $1-\left(\dfrac{7}{9}\right)^{25}$; (3) $1-2\cdot\left(\dfrac{8}{9}\right)^{25}+\left(\dfrac{7}{9}\right)^{25}$;

(4) $C_{25}^{3}\left(\dfrac{1}{9}\right)^{3}\left(\dfrac{8}{9}\right)^{22}$.

15. $\dfrac{1}{4},\dfrac{3}{8}$.

16. $\dfrac{C_a^k C_b^{i-k}}{C_{a+b}^i}$.

17. (1) $\dfrac{4}{10}$;(2) $\dfrac{1}{6}$;(3) $\dfrac{5}{6}$.

18. $C_{10}^{3}\cdot\dfrac{1}{2^{10}}$.

19. 0.25

20. $\dfrac{48}{13!}$.

21. $\dfrac{3}{10}$.

22. 0.121.

23. (1) $\dfrac{a-1}{a+b-1}$;(2) $\dfrac{a-1}{a+b-1}$;(3) $\dfrac{a(a-1)}{a(a+b)+a(b-1)}$.

25. (1) $\dfrac{1}{3}$;(2) $\dfrac{1}{2}$.

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

27. (1) $\frac{4}{10}$;(2) $\frac{24}{90}$;(3) $\frac{24}{720}$.

28. 0.59.

29. (1) 0.056;(2) $\frac{1}{18}$.

30. (1) 0.455;(2) 0.14.

31. 0.087,0.49.

38. 0.316.

39. 0.63.

40. (1) 0.228 6;(2) 0.049 7.

(B)

1. $\frac{9}{19},\frac{10}{19}$.

2. $\frac{2(n-r-1)}{n(n-1)},\frac{1}{n-1}$.

3. $C_{2N}^{N}\left(\frac{1}{2}\right)^{2N-k}$.

4. $\frac{m^k-(m-1)^k}{n^k}$.

5. 0.25.

6. $\frac{p_1}{1-q_1q_2},\frac{q_1p_2}{1-q_1q_2}$.

7. (1) $p=\frac{29}{90}$;(2) $q=\frac{20}{61}$.

8. (1) 0.08;(2) 0.6.

习 题 二

(A)

1. $X=\begin{cases}40, & \text{出现正面},\\ 20, & \text{出现反面}.\end{cases}$ $Y=\begin{cases}10, & \text{出现正面},\\ 30, & \text{出现反面}.\end{cases}$

$$F_X(x)=\begin{cases}0, & x<20,\\ \frac{1}{2}, & 20\leqslant x<40,\\ 1, & x\geqslant 40.\end{cases}\qquad F_Y(x)=\begin{cases}0, & x<10,\\ \frac{1}{2}, & 10\leqslant x<30,\\ 1, & x\geqslant 30.\end{cases}$$

2. (1) $\frac{1}{2(2^{100}-1)}$;(2) $\frac{1}{3}$.

4. (1) $P\{X=k\}=0.3^{k-1}\cdot 0.7,k=1,2,\cdots$;

(2)

X	1	2	3	4
p_i	$\frac{7}{10}$	$\frac{7}{30}$	$\frac{7}{120}$	$\frac{1}{120}$

.

5. $\frac{4}{5},\frac{4}{5},\frac{7}{10}$.

6. 1,0.39.

8. (1) 是,$f(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他};\end{cases}$ (2) 不是.

9. (1) $\frac{1}{2}$,$F(x)=\begin{cases}\frac{1}{2}e^{x}, & x<0,\\ 1-\frac{1}{2}e^{-x}, & x\geqslant 0,\end{cases}$ $1-\frac{1}{2}(e^{-\frac{\sqrt{2}}{2}}+e^{-1})$,

$\frac{1}{2}(e^{\frac{\sqrt{2}}{2}}-e^{-\sqrt{2}}),\frac{1}{2}e^{-1}$;

(2) 2, $F(x)=\begin{cases}0, & x<0,\\ \frac{1}{2}x^{2}, & 0\leqslant x<1,\\ 2x-\frac{1}{2}x^{2}-1, & 1\leqslant x<2,\\ 1, & x\geqslant 2,\end{cases}$

$\frac{1}{4},2\sqrt{2}-\frac{9}{4},\frac{1}{2}$.

11. $\frac{3}{2},-\frac{1}{5}$.

12. 10.37,0.020 1.

13. 975.6.

17. (1) $P\{X=k\}=C_n^k(0.94)^k(0.06)^{n-k}$; (2) $(0.94)^n$;

(3) $1-(0.94)^n-n\cdot(0.94)^{n-1}\cdot 0.06$; (4) $0.06n$,$0.056\,4n$.

18. pq^k,$\frac{q}{p}$.

19. 1,1,2,$\frac{1}{6e}$.

20. 0.352.

21. 800 000.

24. $3e^{-2}-2e^{-3}$.

26. $1-18.5e^{-5}$.

27. (1) 0.069 31; (2) 0.083.

28. 86.45.

29.

X	20π	22π	24π	26π
p_i	0.1	0.4	0.3	0.2

Y	100π	121π	144π	169π
p_i	0.1	0.4	0.3	0.2

31. $F_{X^2}(x)=\begin{cases}0, & x<0,\\ \sqrt{x}, & 0\leqslant x<1,\\ 1, & y>1;\end{cases}$ $f_{X^2}(x)=\begin{cases}\dfrac{1}{2\sqrt{x}}, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$

32. $F_Y(x)=\begin{cases}0, & x<\beta,\\ 1-\mathrm{e}^{-\frac{\lambda}{\alpha}(x-\beta)}, & x\geqslant\beta;\end{cases}$

$f_Y(x)=\begin{cases}0, & x<\beta,\\ \dfrac{\lambda}{\alpha}\mathrm{e}^{\frac{\lambda}{\alpha}(x-\beta)}, & x\geqslant\beta.\end{cases}$

33. $\dfrac{\pi}{24}(a+b)(a^2+b^2)$.

36. $N\left(\ln 3-\ln 2-\dfrac{1}{2}\ln\left(\dfrac{4}{15^2}+1\right),\ln\left(1+\dfrac{4}{15^2}\right)\right)$.

(B)

2. (1) $f_\eta(x)=\begin{cases}\dfrac{1}{1\,000}\mathrm{e}^{-\frac{x}{1\,000}}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$

(2) $f_X(x)=\begin{cases}\dfrac{n}{1\,000}\left[1-\mathrm{e}^{-\frac{x}{1\,000}}\right]^{n-1}\mathrm{e}^{-\frac{x}{1\,000}}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$

(3) $f_Y(x)=\begin{cases}\dfrac{n}{1\,000}\mathrm{e}^{-\frac{nx}{1\,000}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$

3. 37.

习 题 三

(A)

2. (1)

X_1 \ X_2	0	1	2	p_i^X
0	1/56	5/28	5/28	21/56
1	5/56	5/14	5/28	35/56
p_j^Y	3/28	15/28	5/14	

(2) $\dfrac{5}{14},\dfrac{21}{56},\dfrac{13}{28}$.

3. (1) $k_1=12, k_2=21$;

(2) $f_{X_1}(x)=\begin{cases}3e^{-3x}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$ $f_{Y_1}(y)=\begin{cases}4e^{-4y}, & y\geqslant 0,\\ 0, & y<0;\end{cases}$

$g_{X_2}(x)=\begin{cases}\frac{21}{4}e^{-3x}(1-e^{-4x}), & x\geqslant 0,\\ 0, & x<0;\end{cases}$ $g_{Y_2}(y)=\begin{cases}7e^{-7y}, & y\geqslant 0,\\ 0, & y<0.\end{cases}$

4. (1) $f(x,y)=\begin{cases}\frac{1}{2}, & 0\leqslant x\leqslant 2,\ 0\leqslant y\leqslant 1,\\ 0, & \text{其他};\end{cases}$

$F(x,y)=\begin{cases}0, & x<0 \text{ 或 } y<0,\\ \frac{xy}{2}, & 0\leqslant x\leqslant 2, 0\leqslant y\leqslant 1,\\ \frac{x}{2}, & 0\leqslant x\leqslant 2, y>1,\\ y, & x>2, 0\leqslant y\leqslant 1,\\ 1, & x>2, y>1;\end{cases}$

(2) $f_X(x)=\begin{cases}\frac{1}{2}, & 0\leqslant x\leqslant 2,\\ 0, & \text{其他};\end{cases}$ $f_Y(y)=\begin{cases}1, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他};\end{cases}$

$F_X(x)=\begin{cases}0, & x<0,\\ \frac{x}{2}, & 0\leqslant x\leqslant 2,\\ 1, & x>2;\end{cases}$ $F_Y(y)=\begin{cases}0, & y<0,\\ y, & 0\leqslant y\leqslant 1,\\ 1 & y>1.\end{cases}$

(3) $\frac{2}{3}$.

5. (1) $f(x,y)=\begin{cases}2, & 1>y>x>0,\\ 0, & \text{其他};\end{cases}$

(2) $f_X(x)=\begin{cases}2(1-x), & 0\leqslant x\leqslant 1,\\ 0, & \text{其他};\end{cases}$ $f_Y(y)=\begin{cases}2y, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$

6. (1) $F_{(x,y)}=\begin{cases}0, & x<0 \text{ 或 } y<1,\\ \frac{x(y-1)}{2}, & 0\leqslant x\leqslant 2, 1\leqslant y\leqslant 2,\\ \frac{x}{2}, & 0\leqslant x\leqslant 2, y>2,\\ y-1, & x>2, 1\leqslant y\leqslant 2,\\ 1, & x>2, y>2;\end{cases}$

(2) $G_{(x,y)}=\begin{cases}0, & x<0 \text{ 或 } y<1,\\ \dfrac{\sqrt{x}(\sqrt{y}-1)}{2}, & 0\leqslant x\leqslant 4,1\leqslant y\leqslant 4,\\ \dfrac{\sqrt{x}}{2}, & 0\leqslant x\leqslant 4,y>4,\\ \sqrt{y}-1, & x>4,1\leqslant y\leqslant 4,\\ 1, & x>4,y>4;\end{cases}$

(3) $\frac{1}{4}$.

8. $P\{X_1=0\mid X_2=1\}=\frac{1}{3}$; $P\{X_1=1\mid X_2=1\}=\frac{2}{3}$.

9. (1) $p_{ij}=p_i^X p_{j|i}$; (2) $p_j^Y=\sum\limits_i p_{ij}=\sum\limits_i p_i^X p_{j|i}$;

(3) $p_{i|j}=\dfrac{p_{ij}}{p_j^Y}=\dfrac{p_i^X p_{j|i}}{\sum\limits_i p_i^X p_{j|i}}$.

10. (1) $p_{11}=\frac{1}{4}$, $p_{12}=0$, $p_{13}=\frac{1}{4}$, $p_{22}=\frac{1}{2}$; (2) 不独立.

11.

X \ Y	y_1	y_2	y_3	p_i^X
x_1	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{4}$
x_2	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{4}$
p_j^Y	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	1

13. $\frac{19}{36}$, $\frac{2}{36}$.

14. $f_{X_1|Y_1}(x\mid y)=\begin{cases}3e^{-3x}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$

$f_{Y_1|X_1}(y\mid x)=\begin{cases}4e^{-4y}, & y\geqslant 0,\\ 0, & y<0;\end{cases}$

$f_{X_2|Y_2}(x\mid y)=\begin{cases}3e^{-3x+3y}, & x\geqslant y,\\ 0, & x<y;\end{cases}$ $(y>0)$

$f_{Y_2|X_2}(y\mid x)=\begin{cases}4\cdot\dfrac{e^{-4y}}{1-e^{-4x}}, & 0<y\leqslant x,\\ 0, & \text{其他}.\end{cases}$ $(x>0)$

15. $f_{Y|X}(y\mid x)=\begin{cases}\dfrac{1}{x+1}, & 0<y<x+1,\\ 0, & \text{其他};\end{cases}$ $(0<x<1)$

$$f_{Y|X}(y|x)=\begin{cases}\frac{1}{2}, & x-1<y<x+1,\\ 0, & \text{其他};\end{cases}\qquad (1\leqslant x\leqslant 2)$$

$$f_{X|Y}(x|y)=\begin{cases}\frac{1}{y+1}, & 0<x<y+1,\\ 0, & \text{其他};\end{cases}\qquad (0\leqslant y\leqslant 1)$$

$$f_{X|Y}(x|y)=\begin{cases}\frac{1}{3-y}, & y-1<x<2,\\ 0, & \text{其他}.\end{cases}\qquad (1\leqslant y<3)$$

17. (1) $f(x,y)=f_X(x)f_{Y|X}(y|x)$;

(2) $f_Y(y)=\int_{-\infty}^{+\infty}f_X(x)f_{Y|X}(y\mid x)\mathrm{d}x$;

(3) $f_{X|Y}(x/y)=\dfrac{f_X(x)f_{Y|X}(y\mid x)}{\int_{-\infty}^{+\infty}f_X(x)f_{Y|X}(y\mid x)\mathrm{d}x}$.

18. X_1 与 Y_1 独立，X_2 与 Y_2 不独立.

19. 不独立.

20. $f(x,y)=\begin{cases}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{y^2}{2}}, & 0\leqslant x\leqslant 1,-\infty<y<+\infty,\\ 0, & \text{其他}.\end{cases}$

23.

ζ	0	1	2
p_i	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

，

η	−2	−1	0	1	2
p_i	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

。

ζ \ η	−2	−1	0	1	2
0	0	0	$\frac{1}{9}$	0	0
1	0	$\frac{2}{9}$	0	$\frac{2}{9}$	0
2	$\frac{1}{9}$	0	$\frac{2}{9}$	0	$\frac{1}{9}$

24.

ζ	2	3	4	5	6	7	8	9	10	11	12
p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

26. $f_{X_1+X_2}(z)=\begin{cases}0, & z\leqslant 0,\\ \lambda^2 z\mathrm{e}^{-\lambda z}, & z>0.\end{cases}$

27. $f_X(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-3)^2}{2}}$，$f_Y(y)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-1)^2}{2}}$.

29. $F_\xi(s)=\begin{cases}0, & s\leqslant 0,\\ \frac{1}{4}s, & 0<s\leqslant 2,\\ 1-\frac{1}{s}, & s>2;\end{cases}$ $f_\xi(s)=\begin{cases}0, & s\leqslant 0,\\ \frac{1}{4}, & 0<s\leqslant 2,\\ \frac{1}{s^2}, & s>2.\end{cases}$

30. 0,0.4.

31. $\frac{5}{3}$.

33. 8

34. $9\left[1-\left(\frac{8}{9}\right)^{25}\right]$.

36. $-\frac{15}{224}$.

37. $0,\frac{1}{49}$.

38. $\frac{1}{2},0$.

40. $\frac{4}{\sqrt{66}}$.

41. (1) $\frac{1}{2}$;(2) $13x^2-6x+9$;

(3) $x=\frac{3}{13}$,$0\leqslant x\leqslant\frac{6}{13}$.

42. 若 $\rho_{AB}=1$,且 $\sigma_A\neq\sigma_B$,取 $x_A=\frac{\sigma_B}{\sigma_B-\sigma_A}$时,$P$ 无风险;若 $\rho_{AB}=-1$,取 $x_A=\frac{\sigma_B}{\sigma_A+\sigma_B}$时,$P$ 无风险. $\rho_{AB}<\frac{\min(\sigma_A,\sigma_B)}{\max(\sigma_A,\sigma_B)}$.

43. (1) 2.755%;(2) 3%.

44. $E[Y|X=x]=\begin{cases}\frac{x+1}{2}, & 0<x<1,\\ x, & 1\leqslant x\leqslant 2;\end{cases}$

$E[Y|X]=\begin{cases}\frac{X+1}{2}, & 0<X<1,\\ X, & 1\leqslant X\leqslant 2.\end{cases}$

45. $E[X|Y]=1$,$E[Y|X]=0$.

46. 0.046 5.

47. 2 265.

48. (1) 0.002 7,(2) 440.

(B)

1. (1)

X_1 \ X_2	0	1	2	3
0	21/792	105/792	105/792	21/792
1	70/792	210/792	126/792	14/792
2	35/792	63/792	21/792	1/792

(2)

X_2	0	1	2	3
$p_j^{X_2}$	126/792	378/792	252/792	36/792

(3)

X_1	0	1	2
$p_{i\|0}$	6/36	20/36	10/36
$p_{i\|1}$	10/36	20/36	6/36
$p_{i\|2}$	15/36	18/36	3/36
$p_{i\|3}$	21/36	14/36	1/36

(4)

Y	0	1	2	3	4	5
p_i	21/792	175/792	350/792	210/792	35/792	1/792

(5)

X_1 \ Y_1	0	1	2
0	3/66	12/66	6/66
1	15/66	20/66	0
2	10/66	0	0

X_1 \ Y_2	0	1	2	3
0	102/990	159/990	51/990	3/990
1	125/990	270/990	120/990	10/990
2	25/990	75/990	45/990	5/990

2. (1) $P\{Y_i=k,Y_j=s\}=\dfrac{C_{25}^{k}C_{25-k}^{s}7^{25-k-s}}{9^{25}}$, $k+s\leqslant 25, k,s=0,1,2,\cdots,25$,

$P\{Y_i=k\}=\dfrac{C_{25}^{k}8^{25-k}}{9^{25}}$, $k=0,1,\cdots,25$,

Y_i 与 Y_j 不独立.

(2)

X_i \ X_j	0	1	$p_k^{X_i}$
0	$\left(\frac{7}{9}\right)^{25}$	$\left(\frac{8}{9}\right)^{25}-\left(\frac{7}{9}\right)^{25}$	$\left(\frac{8}{9}\right)^{25}$
1	$\left(\frac{8}{9}\right)^{25}-\left(\frac{7}{9}\right)^{25}$	$1-2\left(\frac{8}{9}\right)^{25}+\left(\frac{7}{9}\right)^{25}$	$1-\left(\frac{8}{9}\right)^{25}$
$p_l^{X_j}$	$\left(\frac{8}{9}\right)^{25}$	$1-\left(\frac{8}{9}\right)^{25}$	

X_i 与 X_j 不独立.

(3) $\operatorname{cov}(X_i,X_j)=\left(\frac{7}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50}$, $(i\neq j)$

$$DX_i=\left(\frac{8}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50},$$

$$\rho_{X_i,X_j}=\left[\left(\frac{7}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50}\right]\Big/\left[\left(\frac{8}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50}\right];$$

(4) $DX=9\left[\left(\frac{8}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50}\right]+72\left[\left(\frac{7}{9}\right)^{25}-\left(\frac{8}{9}\right)^{50}\right]$.

3. (1) $P\{Y=s\mid X=k\}=C_{25-k}^{s}p_2^{s}(1-p_2)^{25-k-s}$, $0\leqslant s\leqslant 25-k$;

(2) $P\{X=k,Y=s\}=C_{25}^{k}p_1^{k}(1-p_1)^{25-k}C_{25-k}^{s}p_2^{s}(1-p_2)^{25-k-s}$

$$=\frac{25!}{k!\,s!\,(25-k-s)!}p_1^{k}[p_2(1-p_1)]^{s}[1-p_1-p_2(1-p_1)]^{25-k-s};$$

(3) $E[Y\mid X]=(25-X)p_2$,

$E[Y^2\mid X]=(25-X)p_2(1-p_2)+(25-X)^2p_2^2$;

(4) $EY=E[E(Y\mid X)]=(25-25p_1)p_2=25(1-p_1)p_2$,

$DY=EY^2-(EY)^2=E(E[Y^2\mid X])-(EY)^2$

$=25(1-p_1)p_2[1-p_2(1-p_1)]$.

4. (3) $Y\sim b(25,(1-p_1)p_2)$, 在 $Y=s$ 的条件下 X 的条件分布 $X\mid Y=s\sim b\left(25-s,\frac{p_1}{1-(1-p_1)p_2}\right)$

5. (1) $f(x,y)=\begin{cases}xe^{-xy}, & 1\leqslant x\leqslant 2,y\geqslant 0,\\ 0, & \text{其他};\end{cases}$

(2) $f_Y(y)=\begin{cases}y(e^{-y}-2e^{-2y})+e^{-y}-e^{-2y}, & y\geqslant 0,\\ 0, & \text{其他};\end{cases}$

(3) $f_{X|Y}(x\mid y)=\begin{cases}\frac{xe^{-xy}}{y(e^{-y}-2e^{-2y})+e^{-y}-e^{-2y}}, & 1\leqslant x\leqslant 2,\ (y>0)\\ 0, & \text{其他};\end{cases}$

(4) 1.474;

(5) 0.865 5.

6. (1) $f_X(x)=\begin{cases}\frac{1}{2}x, & 0<x<1,\\ \frac{1}{2}, & 1\leqslant x\leqslant 2,\\ \frac{3-x}{2}, & 2\leqslant x\leqslant 3,\\ 0, & \text{其他;}\end{cases}$

$f_Y(y)=\begin{cases}\frac{3-2y}{2}, & 0\leqslant y\leqslant 1,\\ 0, & \text{其他.}\end{cases}$

X 与 Y 不独立;

(2) $\begin{pmatrix}\frac{5}{12} & 0\\ 0 & \frac{11}{144}\end{pmatrix}$, X 与 Y 不相关;

(3) $f_{Y|X}(y|x)=\begin{cases}\frac{1}{x}, & 0\leqslant y\leqslant x, \quad (0<x<1)\\ 0, & \text{其他;}\end{cases}$

$f_{Y|X}(y|x)=\begin{cases}1, & 0\leqslant y\leqslant 1, \quad (1<x<2)\\ 0, & \text{其他;}\end{cases}$

$f_{Y|X}(y|x)=\begin{cases}\frac{1}{3-x}, & 0\leqslant y\leqslant 3-x, \quad (2\leqslant x<3)\\ 0, & \text{其他;}\end{cases}$

$f_{X|Y}(x|y)=\begin{cases}\frac{1}{3-2y}, & y\leqslant x\leqslant 3-y, \quad (0\leqslant y\leqslant 1)\\ 0, & \text{其他;}\end{cases}$

(4) $E[Y|X]=\begin{cases}\frac{x}{2}, & 0<x<1,\\ \frac{1}{2}, & 1<x<2,\\ \frac{3-x}{2}, & 2\leqslant x<3;\end{cases}$

$E[X|Y]=\frac{3}{2}$. $(0\leqslant Y\leqslant 1)$

8. (1)

X_1 \ X_2	0	1
0	$1-e^{-1}$	0
1	$e^{-1}-e^{-2}$	e^{-2}

(2) $\operatorname{cov}(X_1,X_2)=e^{-2}-e^{-3}$

$\rho_{X_1,X_2}=\sqrt{\frac{e-1}{e^2-1}}$

9. (1)

U \ V	0	1
0	$\frac{1}{4}$	0
1	$\frac{1}{4}$	$\frac{1}{2}$

(2) $\rho_{UV}=\frac{1}{\sqrt{3}}$.

13. (1) $\frac{2}{3}$;

(2) $f_Z(z)=\begin{cases}\frac{3}{2}-z & 0<z<1,\\ 0, & \text{其他}.\end{cases}$

习 题 四

(A)

1. $f(x_1,x_2,\cdots,x_n)=\alpha^n e^{\alpha(x_1+\cdots+x_n)}$, $x_i\geqslant 0, i=1,2,\cdots,n$.

2. $f(x_1,x_2,\cdots,x_n)=\begin{cases}1, & 0\leqslant x_i\leqslant 1, i=1,2,\cdots,n,\\ 0, & \text{其他}.\end{cases}$

3. $P(X_1=k_1,X_2=k_2,\cdots,X_n=k_n)=\left(\frac{p}{1-p}\right)^n(1-p)^{k_1+k_2+\cdots+k_n}$,

其中 k_i 为正整数, $i=1,2,\cdots,n$.

4. $\bar{y}=\frac{\bar{x}-a}{b}$, $s_y^2=\frac{s_x^2}{b^2}$.

9. $P(X_{(1)}\leqslant x)=1-(1-F(x))^n$;

$P(X_{(n)}\leqslant x)=(F(x))^n$.

若 X 服从以 $\alpha(\alpha>0)$ 为参数的指数分布, 则 $P\{X_{(1)}\leqslant x\}=1-e^{-n\alpha x}$, $x\geqslant 0$; $P\{X_{(n)}\leqslant x\}=(1-e^{-\alpha x})^n$, $x\geqslant 0$.

10. $f(x)=\frac{n^{\frac{mn}{2}}}{2^{\frac{mn}{2}}\Gamma\left(\frac{mn}{2}\right)}x^{\frac{mn}{2}-1}e^{-\frac{nx}{2}}$, $x>0$.

11. $c=\frac{1}{3}$, 自由度为 2.

12. $P(S_n=k)=\frac{(n\lambda)^k}{k!}e^{-n\lambda}$, $k=0,1,2,\cdots$;

S_n 渐近地服从正态分布 $N(n\lambda,n\lambda)$, 即有

$$P(S_n\leqslant x)\approx\Phi_0\left(\frac{x-n\lambda}{\sqrt{n\lambda}}\right).$$

13. $u_{0.4}=0.253$, $u_{0.2}=0.841\,6$, $u_{0.1}=1.28$, $u_{0.05}=1.65$.

14. 441.

15. 1.145, 11.071, 2.558, 23.209.

16. 0.162 3,0.068 4,0.091 2.

17. 2.353,3.365,1.415,3.169.

(B)

1. 250.

2. (−0.63,0.63).

8. (4) $a=0$ 时的证明是简单的.现设 $a>0$,对任意正数 $\varepsilon<\sqrt{a}$,

令 $\eta=\varepsilon(2\sqrt{a}-\varepsilon)$. 因 $X_n \xrightarrow{P} a$, 即知

$\lim\limits_{n\to\infty} P(a-\eta<X_n<a+\eta)=1$,从而 $\lim\limits_{n\to\infty} P(\sqrt{a-\eta}<\sqrt{X_n}<\sqrt{a+\eta})=1$.

再因 $a-\eta=(\sqrt{a}-\varepsilon)^2, a+\eta=a+2\varepsilon\sqrt{a}-\varepsilon^2<(\sqrt{a}+\varepsilon)^2$,

可得 $P(\sqrt{a-\eta}<\sqrt{X_n}<\sqrt{a+\eta})\leqslant P(\sqrt{a}-\varepsilon<\sqrt{X_n}<\sqrt{a}+\varepsilon)\leqslant 1$.

由夹逼定理可得

$$\lim_{n\to\infty} P(\sqrt{a}-\varepsilon<\sqrt{X_n}<\sqrt{a}+\varepsilon)=1, \text{故} \sqrt{X_n}\xrightarrow{P}\sqrt{a}.$$

结论(1),(2)与(3)也可根据依概率收敛的定义类似地证明.

10. 提示:可利用习题四(A)第 5 题以及前两题的结论证之.

习 题 五

(A)

1. (1) $\hat{\alpha}_{MF}=\dfrac{1-2\bar{X}}{\bar{X}-1}, \hat{\alpha}_{MLE}=-\left(1+\dfrac{n}{\sum\limits_{i=1}^{n}\ln X_i}\right)$;

(2) $\hat{p}_{ME}=\hat{p}_{MLE}=\bar{X}$;

(3) $\hat{\lambda}_{ME}=\hat{\lambda}_{MLE}=\dfrac{\alpha}{\bar{X}}$;

(4) $\hat{\lambda}_{MF}=\hat{\lambda}_{MLE}=\bar{X}$.

3. $c=\dfrac{1}{2(n-1)}$.

4. 0.089.

5. (1) $n\geqslant 25$; (2) $n\geqslant 60$.

6. (7.08,8.32).

7. (5.297,5.503),(0.293,0.631).

8. (0.245 1,0.282 3).

9. 认为调整措施效果显著.

10. 先后检验假设 $H_{01}:\mu=18, H_{02}:\sigma^2\leqslant 0.3^2$,经过检验,$H_{01}$ 成立而 H_{02} 不成立,认为金商出售的产品存在质量问题.

11. 认为两品牌的去污率无明显差异.

12. 可以支持工厂采用新工艺.

13. 先后检验假设 $H_{01}:\sigma_1^2=\sigma_2^2$, $H_{02}:\mu_1\geqslant\mu_2$, 经过检验, H_{01} 成立而 H_{02} 不成立, 认为新方法的提取率有明显提高.

14. 可以.

15. 可以认为服从泊松分布.

16. 可以认为服从指数分布.

17. 患病率与吸烟有关.

18. 智力高低与出生季节有关.

(B)

1. $\mu_{MF}=\ln\bar{X}-\frac{1}{2}\ln\left(1+\frac{\bar{X}^2}{S_0^2}\right)$, $\hat{\sigma}_{ME}^2=\ln\left(1+\frac{\bar{X}^2}{S_0^2}\right)$;

   $\mu_{MLE}=\frac{1}{n}\sum_{i=1}^{n}\ln X_i$, $\hat{\sigma}_{MLE}^2=\frac{1}{n}\sum_{i=1}^{n}\left(\ln X_i-\frac{1}{n}\sum_{i=1}^{n}\ln X_i\right)^2$.

2. $\hat{a}_{ME}=\bar{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$, $\hat{b}_{ME}=\bar{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$,

   $\hat{a}_{ME}=\min_{1\leqslant i\leqslant n}X_i$, $\hat{b}_{MLE}=\max_{1\leqslant i\leqslant n}X_i$.

5. $\chi^2=\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sigma^2}\sim\chi^2(n)$.

6. $F=\frac{1}{r}\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}\sim F(n_1,n_2)$.

7. (1) $\hat{\lambda}=\bar{x}=0.196\ 5$;

   (2) $U=\frac{\bar{X}-EX}{\sqrt{DX/n}}$ 渐近服从 $N(0,1)$ 分布, $EX=DX=\lambda$, 经检验, H_0 成立.

8. (1) 认为病情与年龄有关.

   (2) 病死率为 $\hat{p}_0=\frac{39}{564}\approx0.069$, 可检验假设 $H_0:p\leqslant\hat{p}_0$. 经检验, 可以认为老年人病死率偏高

习 题 六

(A)

1. 不同品牌小汽车的耗油量有显著差异.

2. 不同工艺下布的缩水率有显著差异.

3. 不同菌型下小白鼠存活日数有显著差异.

4. 不同类型的电路响应时间有显著差异.

5. 原料产地的不同没有显著影响,生产工艺的不同影响显著.

6. 操作工人的影响不显著,机器以及机器与操作工人的交互作用影响显著.

7. 品种、肥料以及它们的交互作用影响均是显著的.

常用统计分布表

附表 1　泊松分布概率值表

$$P\{X=m\}=\frac{\lambda^m}{m!}e^{-\lambda}$$

m \ λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	0.904 837	0.818 731	0.740 818	0.676 320	0.606 531	0.548 812	0.496 585	0.449 329
1	0.090 484	0.163 746	0.222 245	0.268 128	0.303 265	0.329 287	0.347 610	0.359 463
2	0.004 524	0.016 375	0.033 337	0.053 626	0.075 816	0.098 786	0.121 663	0.143 785
3	0.000 151	0.001 092	0.003 334	0.007 150	0.012 636	0.019 757	0.028 388	0.038 343
4	0.000 004	0.000 055	0.000 250	0.000 715	0.001 580	0.002 964	0.004 968	0.007 669
5		0.000 002	0.000 015	0.000 057	0.000 158	0.000 356	0.000 696	0.001 227
6			0.000 001	0.000 004	0.000 013	0.000 036	0.000 081	0.000 164
7					0.000 001	0.000 003	0.000 008	0.000 019
8							0.000 001	0.000 002
9								
10								
11								
12								
13								
14								
15								
16								
17								

m \ λ	0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	0.406 570	0.367 879	0.223 130	0.135 335	0.082 085	0.049 787	0.030 197	0.018 316
1	0.359 13	0.367 879	0.334 695	0.270 671	0.205 212	0.149 361	0.105 691	0.073 263
2	0.164 661	0.183 940	0.251 021	0.270 671	0.256 516	0.224 042	0.184 959	0.146 525
3	0.049 398	0.061 313	0.125 510	0.180 447	0.213 763	0.224 042	0.215 785	0.195 367
4	0.011 115	0.015 328	0.047 067	0.090 224	0.133 602	0.168 031	0.188 812	0.195 367
5	0.002 001	0.003 066	0.014 120	0.036 089	0.066 801	0.100 819	0.132 169	0.156 293
6	0.000 300	0.000 511	0.003 530	0.012 030	0.027 834	0.050 409	0.077 098	0.104 196
7	0.000 039	0.000 073	0.000 756	0.003 437	0.009 941	0.021 604	0.038 549	0.059 540
8	0.000 004	0.000 009	0.000 142	0.000 859	0.003 106	0.008 102	0.016 865	0.029 770
9		0.000 001	0.000 024	0.000 191	0.000 863	0.002 701	0.006 559	0.013 231

续表

m \ λ	0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
10			0.000 04	0.000 038	0.000 216	0.000 810	0.002 296	0.005 292
11				0.000 007	0.000 049	0.000 221	0.000 730	0.001 925
12				0.000 001	0.000 010	0.000 055	0.000 213	0.000 642
13					0.000 002	0.000 013	0.000 057	0.000 197
14						0.000 002	0.000 014	0.000 056
15						0.000 001	0.000 003	0.000 015
16							0.000 001	0.000 004
17								0.000 001

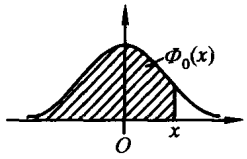
m \ λ	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
0	0.011 109	0.006 738	0.004 087	0.002 479	0.001 503	0.000 912	0.000 553	0.000 335
1	0.049 990	0.033 690	0.022 477	0.014 873	0.009 773	0.006 383	0.004 148	0.002 684
2	0.112 479	0.084 224	0.061 812	0.044 618	0.031 760	0.022 341	0.015 556	0.010 735
3	0.168 718	0.140 374	0.113 323	0.089 235	0.068 814	0.052 129	0.038 888	0.028 626
4	0.189 808	0.175 467	0.155 819	0.133 853	0.111 822	0.091 226	0.072 917	0.057 252
5	0.170 827	0.175 467	0.171 001	0.160 623	0.145 369	0.127 717	0.109 374	0.091 604
6	0.128 120	0.146 223	0.157 117	0.160 623	0.157 483	0.149 003	0.136 719	0.122 138
7	0.082 363	0.104 445	0.123 449	0.137 677	0.146 234	0.149 003	0.146 484	0.139 587
8	0.046 329	0.065 278	0.084 872	0.103 258	0.118 815	0.130 377	0.137 328	0.139 587
9	0.023 165	0.036 266	0.051 866	0.068 838	0.085 811	0.101 405	0.114 441	0.124 077
10	0.010 424	0.018 133	0.028 526	0.041 303	0.055 777	0.070 983	0.085 830	0.099 262
11	0.004 264	0.008 242	0.014 263	0.022 529	0.032 959	0.045 171	0.058 521	0.072 190
12	0.001 599	0.003 434	0.006 537	0.011 264	0.017 853	0.026 350	0.036 575	0.048 127
13	0.000 554	0.001 321	0.002 766	0.005 199	0.008 927	0.014 188	0.021 01	0.029 616
14	0.000 178	0.000 472	0.001 086	0.002 228	0.004 144	0.007 094	0.011 305	0.016 924
15	0.000 053	0.000 157	0.000 399	0.000 891	0.001 796	0.003 311	0.005 652	0.009 026
16	0.000 015	0.000 049	0.000 137	0.000 334	0.000 730	0.001 448	0.002 649	0.004 513
17	0.000 004	0.000 014	0.000 044	0.000 118	0.000 279	0.000 596	0.001 169	0.002 124
18	0.000 001	0.000 004	0.000 014	0.000 039	0.000 100	0.000 232	0.000 487	0.000 944
19		0.000 001	0.000 004	0.000 012	0.000 035	0.000 085	0.000 192	0.000 397
20			0.000 01	0.000 004	0.000 011	0.000 030	0.000 072	0.000 159
21				0.000 001	0.000 004	0.000 010	0.000 026	0.000 061
22					0.000 001	0.000 003	0.000 009	0.000 022
23						0.000 001	0.000 003	0.000 008
24							0.000 001	0.000 003
25								0.000 001

续表

m \ λ	8.5	9.0	9.5	10.0	m \ λ	20	m \ λ	30
0	0.000 203	0.000 123	0.000 075	0.000 045	5	0.000 1	12	0.000 1
1	0.001 730	0.001 111	0.000 711	0.000 454	6	0.000 2	13	0.000 2
2	0.007 350	0.004 998	0.003 378	0.002 270	7	0.000 5	14	0.000 5
3	0.020 826	0.014 994	0.010 696	0.007 567	8	0.001 3	15	0.001 0
4	0.442 55	0.033 737	0.025 403	0.018 917	9	0.002 9	16	0.001 9
5	0.075 233	0.060 727	0.048 265	0.037 833	10	0.005 8	17	0.003 4
6	0.106 581	0.091 090	0.076 421	0.063 055	11	0.010 6	18	0.005 7
7	0.129 419	0.117 116	0.103 714	0.090 079	12	0.017 6	19	0.008 9
8	0.137 508	0.131 756	0.123 160	0.112 599	13	0.027 1	20	0.013 4
9	0.129 869	0.131 756	0.130 003	0.125 110	14	0.038 2	21	0.019 2
10	0.110 303	0.118 580	0.122 502	0.125 110	15	0.051 7	22	0.026 1
11	0.085 300	0.097 020	0.106 662	0.113 736	16	0.064 6	23	0.034 1
12	0.060 421	0.072 765	0.084 440	0.094 780	17	0.076 0	24	0.042 6
13	0.039 506	0.050 376	0.061 706	0.072 908	18	0.081 4	25	0.057 1
14	0.023 986	0.032 384	0.041 872	0.052 077	19	0.088 8	26	0.059 0
15	0.013 592	0.019 431	0.026 519	0.034 718	20	0.088 8	27	0.065 5
16	0.007 220	0.010 930	0.015 746	0.021 699	21	0.084 6	28	0.070 2
17	0.003 611	0.005 786	0.008 799	0.012 764	22	0.076 7	29	0.072 6
18	0.001 705	0.002 893	0.004 644	0.007 091	23	0.066 9	30	0.072 6
19	0.000 762	0.001 370	0.002 322	0.003 732	24	0.055 7	31	0.703
20	0.000 324	0.000 617	0.001 103	0.001 866	25	0.044 6	32	0.065 9
21	0.000 132	0.000 264	0.000 433	0.008 989	26	0.034 3	33	0.059 9
22	0.000 050	0.000 108	0.000 216	0.000 404	27	0.025 4	34	0.052 9
23	0.000 019	0.000 042	0.000 89	0.000 176	28	0.018 2	35	0.045 3
24	0.000 007	0.000 016	0.000 025	0.000 073	29	0.012 5	36	0.037 8
25	0.000 002	0.000 006	0.000 014	0.000 029	30	0.008 3	37	0.030 6
26	0.000 001	0.000 002	0.000 004	0.000 011	31	0.005 4	38	0.024 2
27		0.000 001	0.000 002	0.000 004	32	0.003 4	39	0.018 6
28			0.000 001	0.000 001	33	0.002 0	40	0.013 9
29				0.000 001	34	0.001 2	41	0.010 2
							42	0.007 3
							43	0.050 1
					35	0.000 7	44	0.003 5
					36	0.000 4	45	0.002 3
					37	0.000 2	46	0.001 5
					38	0.000 1	47	0.001 0
					39	0.000 1	48	0.000 6

附表 2 标准正态分布函数值表

$$\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt\ (x\geqslant 0)$$



x	0.00	0.01	0.02	0.03	0.04
0.0	0.500 0	0.504 0	0.508 0	0.512 0	0.516 0
0.1	0.539 8	0.543 8	0.547 8	0.551 7	0.555 7
0.2	0.579 3	0.583 2	0.587 1	0.591 0	0.594 8
0.3	0.617 9	0.621 7	0.625 5	0.629 3	0.633 1
0.4	0.655 4	0.659 1	0.662 8	0.666 4	0.670 0
0.5	0.691 5	0.695 0	0.698 5	0.701 9	0.705 4
0.6	0.725 7	0.729 1	0.732 4	0.735 7	0.738 9
0.7	0.758 0	0.761 1	0.764 2	0.767 3	0.770 3
0.8	0.788 1	0.791 0	0.793 9	0.796 7	0.799 5
0.9	0.815 9	0.818 6	0.821 2	0.823 8	0.826 4
1.0	0.841 3	0.843 8	0.846 1	0.848 5	0.850 8
1.1	0.864 3	0.866 5	0.868 6	0.870 8	0.872 9
1.2	0.884 9	0.886 9	0.888 8	0.890 7	0.892 5
1.3	0.903 20	0.904 90	0.906 78	0.908 24	0.909 88
1.4	0.919 24	0.920 73	0.922 20	0.923 64	0.925 07
1.5	0.933 19	0.934 48	0.935 74	0.936 99	0.938 22
1.6	0.945 20	0.946 30	0.947 38	0.948 45	0.949 50
1.7	0.955 43	0.956 37	0.957 28	0.958 18	0.959 07
1.8	0.964 07	0.964 85	0.965 62	0.966 38	0.967 12
1.9	0.971 28	0.971 93	0.972 57	0.973 20	0.973 81
2.0	0.977 25	0.977 78	0.978 31	0.978 82	0.979 32
2.1	0.982 14	0.982 57	0.983 00	0.983 41	0.983 82
2.2	0.986 10	0.986 45	0.986 79	0.987 13	0.987 45
2.3	0.989 28	0.989 56	0.989 83	0.990 10	0.990 36
2.4	0.991 80	0.992 02	0.992 24	0.992 45	0.992 66
2.5	0.993 79	0.993 96	0.994 13	0.994 30	0.994 46
2.6	0.995 34	0.995 47	0.995 60	0.995 73	0.995 86
2.7	0.996 53	0.996 64	0.996 74	0.996 83	0.996 93
2.8	0.997 45	0.997 52	0.997 60	0.997 67	0.997 74

续表

x	0.00	0.01	0.02	0.03	0.04
2.9	0.998 13	0.998 19	0.998 25	0.998 31	0.998 36
3.0	0.998 65	0.998 69	0.998 74	0.998 78	0.998 82
3.1	0.999 03	0.999 06	0.999 10	0.999 13	0.999 16
3.2	0.999 31	0.999 34	0.999 36	0.999 38	0.999 40
3.3	0.999 52	0.999 53	0.999 55	0.999 57	0.999 58
3.4	0.999 66	0.999 68	0.999 69	0.999 70	0.999 71
3.5	0.999 77	0.999 78	0.999 78	0.999 79	0.999 80
3.6	0.999 84	0.999 85	0.999 85	0.999 86	0.999 86
3.7	0.999 89	0.999 90	0.999 90	0.999 90	0.999 91
3.8	0.999 93	0.999 93	0.999 93	0.999 94	0.999 94
3.9	0.999 95	0.999 95	0.999 96	0.999 96	0.999 96
4.0	0.999 97	0.999 97	0.999 97	0.999 97	0.999 97
4.1	0.999 98	0.999 98	0.999 98	0.999 98	0.999 98
4.2	0.999 99	0.999 99	0.999 99	0.999 99	0.999 99
4.3	0.999 99	0.999 99	0.999 99	0.999 99	0.999 99
4.4	0.999 99	0.999 99	1.000 00	1.000 00	1.000 00

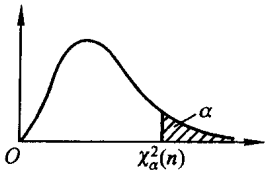
x	0.05	0.06	0.07	0.08	0.09
0.0	0.519 9	0.523 9	0.527 9	0.531 9	0.535 9
0.1	0.559 6	0.563 6	0.567 5	0.571 4	0.575 3
0.2	0.598 7	0.602 6	0.606 4	0.610 3	0.614 1
0.3	0.636 8	0.640 6	0.644 3	0.648 0	0.651 7
0.4	0.673 6	0.677 2	0.680 8	0.684 4	0.687 9
0.5	0.708 8	0.712 3	0.715 7	0.719 0	0.722 4
0.6	0.742 2	0.745 4	0.748 6	0.751 7	0.754 9
0.7	0.773 4	0.776 4	0.779 4	0.782 3	0.785 2
0.8	0.802 3	0.805 1	0.807 8	0.810 6	0.813 3
0.9	0.828 9	0.831 5	0.834 0	0.836 5	0.838 9
1.0	0.853 1	0.855 4	0.857 7	0.859 9	0.862 1
1.1	0.874 9	0.877 0	0.879 0	0.881 0	0.883 0
1.2	0.894 4	0.896 2	0.898 0	0.899 7	0.901 47
1.3	0.911 40	0.913 09	0.914 66	0.916 21	0.917 74
1.4	0.926 47	0.927 85	0.929 22	0.930 56	0.931 89
1.5	0.939 43	0.940 62	0.941 79	0.942 95	0.944 08
1.6	0.950 53	0.951 54	0.952 54	0.953 52	0.954 49
1.7	0.959 94	0.960 80	0.961 64	0.962 46	0.963 27

续表

x	0.05	0.06	0.07	0.08	0.09
1.8	0.967 84	0.968 56	0.969 26	0.969 95	0.970 62
1.9	0.974 41	0.975 00	0.975 58	0.976 15	0.976 70
2.0	0.979 82	0.980 30	0.980 77	0.981 24	0.981 69
2.1	0.984 22	0.984 61	0.985 00	0.985 37	0.985 74
2.2	0.987 78	0.988 09	0.988 40	0.988 70	0.988 99
2.3	0.990 61	0.990 86	0.991 11	0.991 34	0.991 58
2.4	0.992 86	0.993 05	0.993 24	0.993 43	0.993 61
2.5	0.994 61	0.994 77	0.994 92	0.995 06	0.995 20
2.6	0.995 98	0.996 09	0.996 21	0.996 32	0.996 43
2.7	0.997 02	0.997 11	0.997 20	0.997 28	0.997 37
2.8	0.997 81	0.997 88	0.997 95	0.998 01	0.998 07
2.9	0.998 41	0.998 46	0.998 51	0.998 56	0.998 61
3.0	0.998 86	0.998 89	0.998 93	0.998 97	0.999 00
3.1	0.999 18	0.999 21	0.999 24	0.999 26	0.999 29
3.2	0.999 42	0.999 44	0.999 46	0.999 48	0.999 50
3.3	0.999 60	0.999 61	0.999 62	0.999 64	0.999 65
3.4	0.999 72	0.999 73	0.999 74	0.999 75	0.999 76
3.5	0.999 81	0.999 81	0.999 82	0.999 83	0.999 83
3.6	0.999 87	0.999 87	0.999 88	0.999 88	0.999 89
3.7	0.999 91	0.999 92	0.999 92	0.999 92	0.999 92
3.8	0.999 94	0.999 94	0.999 95	0.999 95	0.999 95
3.9	0.999 96	0.999 96	0.999 96	0.999 97	0.999 97
4.0	0.999 97	0.999 98	0.999 98	0.999 98	0.999 98
4.1	0.999 98	0.999 98	0.999 98	0.999 99	0.999 99
4.2	0.999 99	0.999 99	0.999 99	0.999 99	0.999 99
4.3	0.999 99	0.999 99	0.999 99	0.999 99	0.999 99
4.4	1.000 00	1.000 00	1.000 00	1.000 00	1.000 00

附表 3　χ^2 分布上侧分位数表

$P\{\chi^2(n)>\chi_\alpha^2(n)\}=\alpha$



n \ α	$\alpha=0.995$	0.99	0.975	0.95	0.90	0.75
1	—	—	0.001	0.004	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438
13	3.565	4.107	5.009	5.892	7.042	9.299
14	4.075	4.660	5.629	6.571	7.790	10.165
15	4.601	5.229	6.262	7.261	8.547	11.037
16	5.142	5.812	6.908	7.962	9.312	11.912
17	5.697	6.408	7.564	8.672	10.085	12.792
18	6.265	7.015	8.231	9.390	10.865	13.675
19	6.844	7.633	8.907	10.117	11.651	14.562
20	7.434	8.260	9.591	10.851	12.443	15.452
21	8.034	8.897	10.283	11.591	13.240	16.344
22	8.643	9.542	10.982	12.338	14.042	17.240
23	9.260	10.196	11.689	13.091	14.848	18.137
24	9.886	10.856	12.401	13.848	15.659	19.037
25	10.520	11.524	13.120	14.611	16.473	19.939
26	11.160	12.198	13.844	15.379	17.292	20.843
27	11.808	12.879	14.573	16.151	18.114	21.749
28	12.461	13.565	15.308	16.928	18.939	22.657
29	13.121	14.257	16.047	17.708	19.768	23.567
30	13.787	14.954	16.791	18.493	20.599	24.478

续表

n \ α	$\alpha=0.995$	0.99	0.975	0.95	0.90	0.75
31	14.458	15.655	17.539	19.281	21.434	25.390
32	15.134	16.362	18.291	20.072	22.271	26.304
33	15.815	17.074	19.047	20.867	23.110	27.219
34	16.501	17.789	19.806	21.664	23.952	28.136
35	17.192	18.509	20.569	22.465	24.797	29.054
36	17.887	19.233	21.336	23.269	25.643	29.973
37	18.586	19.960	22.106	24.075	26.492	30.893
38	19.289	20.691	22.878	24.884	27.343	31.815
39	19.996	21.426	23.654	25.695	28.196	32.737
40	20.707	22.164	24.433	26.509	29.051	33.660
41	21.421	22.906	25.215	27.326	29.907	34.585
42	22.138	23.650	25.999	28.144	30.765	35.510
43	22.859	24.398	26.785	28.965	31.625	36.436
44	23.584	25.148	27.575	29.787	32.487	37.363
45	24.311	25.901	28.366	30.612	33.350	38.291

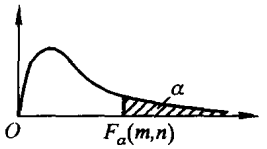
n \ α	$\alpha=0.25$	0.10	0.05	0.025	0.01	0.005
1	1.323	2.706	3.841	5.024	6.635	7.879
2	2.773	4.605	5.991	7.378	9.210	10.597
3	4.108	6.251	7.815	9.348	11.345	12.838
4	5.385	7.779	9.488	11.143	13.277	14.860
5	6.626	9.236	11.071	12.833	15.086	16.750
6	7.841	10.45	12.592	14.449	16.812	18.548
7	9.037	12.017	14.067	16.013	18.475	20.278
8	10.219	13.362	15.507	17.535	20.090	21.955
9	11.389	14.684	16.919	19.023	21.666	23.589
10	12.549	15.987	18.307	20.483	23.209	25.188
11	13.701	17.275	19.675	21.920	24.725	26.756
12	14.845	18.549	21.026	23.337	26.217	28.299
13	15.984	19.812	22.362	24.736	27.688	29.819
14	17.117	21.064	23.685	26.119	29.141	31.319
15	18.245	22.307	24.996	27.488	30.578	32.801
16	19.369	23.542	26.296	28.845	32.000	34.267
17	20.489	24.769	27.587	30.191	33.409	35.718
18	21.605	25.989	28.869	31.526	34.805	37.156
19	22.718	27.204	30.144	32.852	36.191	38.582
20	23.828	28.412	31.410	34.170	37.566	39.997

续表

n \ α	$\alpha=0.25$	0.10	0.05	0.025	0.01	0.005
21	24.935	29.615	32.671	35.479	38.932	41.401
22	26.039	30.813	33.924	36.781	40.289	42.796
23	27.141	32.007	35.172	38.076	41.638	44.181
24	28.241	33.196	36.415	39.364	42.980	45.559
25	29.339	34.382	37.652	40.646	44.314	46.928
26	30.435	35.563	38.885	41.923	45.642	48.290
27	31.528	36.741	40.113	43.194	46.963	49.645
28	32.620	37.916	41.337	44.461	48.278	50.993
29	33.711	39.087	42.557	45.722	49.588	52.336
30	34.800	40.256	43.773	46.979	50.892	53.672
31	35.887	41.422	44.985	48.232	52.191	55.003
32	36.973	42.585	46.194	49.480	53.486	56.328
33	38.058	43.745	47.400	50.725	54.776	57.648
34	39.141	44.903	48.602	51.966	56.061	58.964
35	40.223	46.059	49.802	53.203	57.342	60.275
36	41.304	47.212	50.998	54.437	58.619	61.581
37	42.383	48.363	52.192	55.668	59.892	62.883
38	43.462	59.513	58.384	56.896	61.162	64.181
39	44.539	50.660	54.572	58.120	62.428	65.476
40	45.616	51.805	55.758	59.342	63.691	66.766
41	56.692	52.949	56.942	60.561	64.950	68.053
42	47.766	54.090	58.124	61.777	66.206	69.336
43	48.840	55.230	59.304	62.990	67.459	70.616
44	49.913	56.369	60.481	64.201	68.710	71.893
45	50.985	57.505	61.656	65.410	69.957	73.166

附表 4 F 分布上侧分位数表

$P\{F(m,n)>F_\alpha(m,n)\}=\alpha$



$\alpha=0.10$

n \ m	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.92	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

续表

$\alpha = 0.10$

n \ m	10	12	15	20	24	30	40	60	120	∞
1	60.19	60.17	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.70	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.56
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.10

续表

$\alpha = 0.05$

n \ m	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.06	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

续表

$\alpha = 0.05$

n \ m	10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

续表

$\alpha = 0.025$

n \ m	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.52	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.03	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.42	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.01	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.76	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

续表

$\alpha = 0.025$

n \ m	10	12	15	20	24	30	40	60	120	∞
1	968.6	976.7	984.9	993.1	997.2	1 001	1 006	1 010	1 014	1 018
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

续表

$\alpha = 0.01$

n \ m	1	2	3	4	5	6	7	8	9
1	4 652	4 999.5	5 403	5 626	5 764	5 859	5 928	5 982	6 022
2	98.50	90.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.53	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.45	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	6.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.83	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.30	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.52	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.61	2.41

续表

$\alpha = 0.01$

n \ m	10	12	15	20	24	30	40	60	120	∞
1	6 056	6 106	6 157	6 200	6 235	6 261	6 287	6 313	6 339	6 336
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.53	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

续表

$\alpha = 0.005$

n \ m	1	2	3	4	5	6	7	8	9
1	16 211	20 000	21 615	22 500	23 056	23 437	23 715	23 925	24 091
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.03	4.94
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.68	3.56
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	9.18	6.35	5.24	4.62	4.32	3.95	3.74	3.58	3.45
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	8.18	5.54	4.50	3.92	3.55	3.28	3.00	2.93	2.81
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

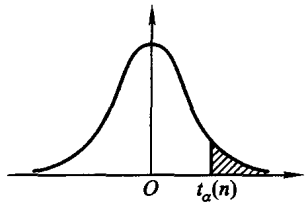
续表

$\alpha = 0.005$

n \ m	10	12	15	20	24	30	40	60	120	∞
1	24 224	24 426	24 630	24 836	24 940	25 044	25 148	25 253	25 359	25 465
2	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	13.62	13.38	13.15	12.90	12.78	12.60	12.53	12.40	12.27	12.14
6	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	5.85	5.66	5.47	5.27	5.17	5.67	4.97	4.86	4.75	4.64
11	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29
28	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25
29	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21
30	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	2.90	2.74	2.57	2.39	2.29	2.19	2.09	1.96	1.83	1.69
120	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

附表 5 t 分布上侧分位数表

$P(t_n > t_\alpha(n)) = \alpha$



n \ α	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

主要参考文献

[1] 洛哈吉 V K. 概率论及数理统计导论. 高尚华译. 北京:高等教育出版社, 1983

[2] 华东师范大学. 概率论与数理统计教程. 北京:高等教育出版社,1983

[3] 陈希孺,倪国熙. 数理统计学教程. 上海:上海科学技术出版社,1988

[4] 复旦大学. 概率论第 2 册:数理统计. 北京:高等教育出版社,1979.

[5] 成世学,严颖,张诒兰等. 经济数学基础(3):概率统计. 北京:中国人民大学出版社,1994

[6] 范培华,袁荫棠等. 经济数学基础,第 3 分册:概率统计. 成都:四川人民出版社,1998

[7] 严士健,刘秀芳,徐承彝. 概率论与数理统计. 北京:高等教育出版社,1990

[8] 陈希孺. 数理统计引论. 北京:科学出版社,1981

[9] 王成名,余鑫晖等. 应用概率统计. 桂林:广西师范大学出版社,1994

[10] Alder H L, Roessler E B. 概率与统计导论. 胡崇能,李隆章校译. 北京:北京大学出版社,1984

[11] 华东师范大学. 概率论与数理统计习题集. 北京:高等教育出版社,1982

[12] 盛骤等. 概率论与数理统计. 北京:高等教育出版社,1989

[13] 陆璇. 数理统计基础. 北京:清华大学出版社,1998

[14] 茆诗松等. 高等数理统计. 北京:高等教育出版社,1998

[15] 王玲玲,周纪芗. 常用统计方法. 上海:华东师范大学出版社,1998

[16] 王学仁等. 实用多元统计分析. 上海:上海科学技术出版社,1990

[17] 王其文. 经济管理计算机基础教程. 北京:高等教育出版社,1999

[18] 周概容. 应用统计方法辞典. 北京:中国统计出版社,1993

Probability Theory & Mathematica Statistics

高等学校经济管理学科数学基础

主编　范培华　胡显佑

	微积分（第二版）	朱来义　主编
	线性代数（第二版）	卢　刚　主编
●	概率论与数理统计（第二版）	龙永红　主编
	微积分中的典型例题分析与习题	朱来义　主编
	线性代数中的典型例题分析与习题	卢　刚　主编
	概率论与数理统计中的典型例题分析与习题	龙永红　主编



ISBN 7-04-013787-9

9 787040 137873 >

定价 19.60 元