高等学校试用教材

概率论与数理统计

北京师范大学

严士健　刘秀芳　徐承彝

高等教育出版社

内 容 提 要

本书主要讲述概率统计的基本理论和基本方法，内容 分为概率、条件概率和独立、分布与期望、 多维随机变量、参数估计与假设检验、回归分析与因子试验的方差分析等六章，可供高等师范院校概率统计专业和数学专业作试用教材或教学参考书使用.

高等学校试用教材

概率论与数理统计

北京师范大学 严士健 刘秀芳 徐承彝

*

高等教育出版社出版

高等教育出版社照排中心照排

新华书店北京发行所发行

北京市制本总厂印刷

*

开本 850×1168 1/32 印张 13.75 字数 350 000

1990 年 4 月第 1 版 1990 年 4 月第 1 次印刷

印数 0001－1 900

ISBN7－04－002836－0/O · 895

定价 3.15 元

编 者 的 话

1986年国家教委理科教材编审委员会概率论与数理统计编审组在拟订"七、五"教材规划时，提出要编写一本高等师范院校适用的概率论与数理统计基础课教材，并且希望能够贯彻教学改革的精神. 1988年12月在宁波师范学院召开了由北京师范大学起草的编写大纲的讨论会.1989年7月在杭州师范学院召开了本书初稿的审查会.

在拟订编写大纲和随后的编写过程中，为了使本书能更好地满足教学的需要，我们认为实现以下一些想法是适宜的.

(1) 加强统计与概率的结合.普及数理统计知识及其应用在国家的现代化进程中具有极其重要的意义，高等师范院校对提高我国人民在这一方面的素养有着独特的作用和方便的条件.概率论在科学和技术领域中有很多重要的应用，其中之一是作为数理统计的基础. 因此我们认为，对初学者来说，以两者结合学习为好. 在教材中尽量介绍统计内容，尽量使统计中所需要的概率理论推导与统计方法的陈述结合在一起，以期使学生能更好地理解概率统计之间的有机联系，理解概率的作用，加深对统计思想的理解，尽早介绍统计的思想和方法. 例如在超几何分布之后即介绍最大似然估计，多重伯努利试验之后即介绍假设检验等. 还想说明即令在高中教学水平上也可以介绍统计的原理及应用，从而更好地推广普及数理统计的知识及应用.

(2) 由于概率统计是大学生首先接触到的关于不确定性现象的数学描述的课程，我们觉得，在叙述上尤其要注意由具体到一般，注意阐述它的基本思想.重要概念的原型及形成.例如概率的概念，我们从最简单明确的古典概型开始，逐步扩充到一般的概

率空间，又说明概率空间的作用更多地表现在理论研究上;又如随机变量是一个十分重要的概念，但初学者不易理解，所以从古典概率开始，我们就用一些变量来表示可以数量化的事件;再如分布的概念是从概率的计算归纳出离散型分布，再进一步引进连续型分布，而后归纳出一般分布;对于独立的概念，我们既注意到它的数学描述，又说明在应用中，独立是根据实际情况判定，而不是数学的验证等.

(3) 在本书中不采用特征函数，测度论等较深的数学工具.这一方面是考虑到课时的限制，另一方面更重要的是，不少学者认为作为随机性数学的第一个课程，主要是让学生理解概率统计的直观背景和基本思想，掌握基本概念及其应用．过早地一般化和运用更深的数学工具会冲淡对概率统计本质的理解，影响学生的长远发展．当然在学生理解了概率统计的特点和实质，要求深造(特别是向理论方面)的时候,进一步的数学工具是完全必要的.

(4) 本书讲述数理统计以正态母体为主.这样做是为了能够以较少的篇幅，较容易处理的方式使学生对数理统计得到较完整而明确的图景，而且以正态母体为基础的统计在应用上还是比较基本和广泛的.

(5) 随机过程及其应用是概率论中十分重要且内容十分丰富的部分．但是由于本书篇幅所限，不能对它作初步介绍，只是在有关章节介绍了它的某些概念(如随机游动、有限状态马氏链)及某些例子(如泊松流—泊松过程的重要例子)，以期读者不致于对随机过程的概念完全陌生，也许对初学者是有益的.

此外，我们在书中每节后面都配有练习性的习题，这是要求读者完成的．章末列有复习题，分 A, B 两种.一般来说，A 种较易，可以选一部分作练习或复习之用; B 种较难，供选用，不一定必作.

本书虽然是为师范院校本科写的教材，但是写作时也照顾到其他类型学校的需要，我们希望也能作为有关专业的教学参考书.

在使用本书时，如果时间不够，可以删掉一些内容，如第六章以及其他一些带星号的节、目和证明．复习题量较大，也有一些较难，不必都做．在此我们向读者提一点参考意见，有些较难的复习题，一时做不出来，不必急于看提示和问老师，留下想一段时间，经过自己努力最终得到解决，这对自己的能力提高、理解加深等都很有好处．

本书的编写大纲最初是由严士健和刘秀芳拟订的．在编写过程中，由严士健、刘秀芳和徐承彝集体讨论内容的取舍与安排．第一至三章由刘秀芳执笔，第四章是严士健在刘秀芳的初稿基础上整理的，第五、六章由徐承彝执笔．全书由严士健统一校阅．习题由陈金文帮助挑选，他和黄荣怀、陈冬青分头核查并做了答案，陈冬青制作了全书的图．在成书过程中得到了多方面的帮助与支持．宁波师范学院的朱作宾、黄世雄副教授，华中师范大学的陆秀丽教授，安徽师范大学的丁万鼎副教授，徐州师范学院的戴朝寿副教授，杭州师范学院的孙志刚教授等认真讨论了编写大纲并提供了他们的宝贵经验．杭州师范学院的孙志刚教授、陈小夏副教授、杭州大学的陆传荣教授对书稿进行　详细审查并提出了修改意见．高等教育出版社及高尚华副编审对本书的出版给于了物质上及业务上各方面的支持、鼓励和帮助．没有这些同志的帮助，本书是不可能出版的．

由于本书写作时间比较仓促，不妥之处肯定不少．我们热切希望各方面给我们提出意见和建议，以能在再版时加以改进，使之更加符合教学改革的实际需要．

编　者

1989年夏于北京师范大学

目　录

编者的话 ……(1)
引言 ……(1)
第一章　概率 ……(5)
§1.1　古典概率 ……(5)
§1.2　超几何分布与产品检验——古典概率的一个应用 ……(22)
§1.3　几何概率 ……(30)
§1.4　频率的稳定性 ……(40)
§1.5　概率的公理化定义 ……(46)
复习题1－A ……(64)
复习题1－B ……(65)
第二章　条件概率和独立 ……(66)
§2.1　条件概率 ……(66)
§2.2　全概率公式和贝叶斯公式 ……(72)
§2.3　事件的独立 ……(79)
§2.4　随机变量及其独立 ……(91)
§2.5　n 重伯努利试验概型 ……(99)
* §2.6　简单随机游动与有限马尔可夫链 ……(116)
复习题2－A ……(122)
复习题2－B ……(124)
第三章　分布与期望 ……(126)
§ 3.1　离散型随机变量及其分布列 ……(126)
§ 3.2　离散型随机变量的期望、方差与母函数 ……(141)
§ 3.3　连续型随机变量 ……(155)
§ 3.4　连续型随机变量的期望 ……(174)
§ 3.5　分布函数 ……(184)
§ 3.6　随机变量的函数及其分布与期望 ……(192)
§ 3.7　期望和方差的性质、切比雪夫不等式 ……(202)

复习题 3－A ……(207)
复习题 3－B ……(209)
第四章 多维随机变量 ……(211)
§ 4.1 多维随机变量的定义及其概率分布 ……(211)
§ 4.2 多维离散型随机变量 ……(225)
§ 4.3 多维连续型随机变量 ……(233)
§ 4.4 多维随机变量函数的分布 ……(248)
§ 4.5 多维随机变量函数的数字特征 ……(271)
复习题 4－A ……(284)
复习题 4－B ……(286)
第五章 参数估计与假设检验 ……(289)
§5.1 概述 ……(289)
§5.2 对未知参数的点估计 ……(294)
§5.3 依概率收敛与大数定律 ……(315)
§5.4 假设检验 ……(324)
§5.5 区间估计 ……(350)
§5.6 中心极限定理与大样本统计推断 ……(355)
第六章 回归分析与因子试验的方差分析 ……(361)
§6.1 方差分析 ……(361)
§6.2 线性回归分析 ……(369)
§6.3 正交试验 ……(392)
习题答案 ……(402)
参考书目 ……(417)
附表 ……(419)
附表 1 泊松分布 ……(420)
附表 2 标准正态分布 ……(422)
附表 3 χ^2 分布 ……(423)
附表 4 t 分布 ……(424)
附表 5 F 分布 ……(426)
附表 6 符号检验表 ……(432)

引　言

概率论是研究随机现象的数量规律的数学分支.

所谓随机现象，是相对于决定性现象而言的. 在一定条件下必然发生(出现)某一结果的现象称为决定性现象. 例如，在没有外力作用下,作匀速直线运动的物体必然继续作匀速直线运动; 又如在标准大气压下，纯水加热到 100 ℃ 时必然会沸腾等等. 这些条件和结果之间存在着必然联系的现象就是决定性现象. 在自然现象和社会现象中还广泛存在着与决定性现象有着本质区别的一类现象，例如当掷一枚硬币时，可能出现正面朝上，也可能出现反面朝上; 每天上午 8:00 — 9:00 记录一个电话交换台收到用户的呼叫次数，可能是 0 次，1 次，2 次 ……;再如，同一门炮向同一目标发射用同一工艺过程生产的炮弹，因为炮弹制造时种种偶然因素对炮弹质量有影响、炮筒位置有差异、空气中气流的变化 …… 都影响着弹着点的位置. 使弹着点在不同次发射中落在不同的位置.这些现象的特点是: 在基本条件不变的情况下，一系列试验或观察会得到不同的结果.每一次试验或观察之前，不能完全肯定会出现哪种结果. 究竟出现哪种结果，呈现出偶然性. 这种现象称为随机现象.

概率论研究随机现象有其独特的方法.它不是企图追索出现每一结果的一切物理因素，从而象研究确定性现象那样确定无疑地预报在哪些条件下出现某一确定的结果，而是通过对随机现象的大量观察，揭示其规律性.例如连续多次掷一枚硬币，随着投掷次数的增加，出现正面的频率(出现正面次数与性掷次数之比)逐渐稳定于 1/2,从而揭示"出现正面"这一结果发生的可能性大小为 1/2;

又如多次测量一物体的长度，其测量结果的平均值随着测量次数的增加逐渐稳定于一个常数等等.

概率论有悠久的历史，它的起源与赌博问题有关.16 世纪，意大利的学者开始研究掷骰子等赌博中的一些简单问题，例如比较掷两个骰子出现点数之和为 9 与 10 的可能性大小.17 世纪中叶，法国数学家帕斯卡(B . Pascal)、费马(P. de Fermat)及荷兰数学家惠更斯(C.Huygens)基于排列组合方法，研究了一些较复杂的赌博问题，他们解决了“分赌注问题”、“赌徒输光问题”等.随着 18、19 世纪科学的发展，人们注意到在某些生物、物理和社会现象与机会游戏之间有一种相似，从而由机会游戏起源的概率论被应用到这些领域中，同时也大大推动了概率论本身的发展.使概率论成为数学的一个分支的奠基人是瑞士数学家伯努利(J. I. Bernoulli)，他建立了概率论中第一个极限定理，即伯努利大数定律，阐明了事件的频率稳定于它的概率.随后棣莫弗(A. de Moivre)和拉普拉斯(P. S. Laplace)又导出了第二个基本极限定理(中心极限定理)的原始形式.拉普拉斯在系统总结前人工作的基础上写出了《分析的概率理论》，明确给出了概率的古典定义，并在概率论中引入了更有力的分析工具，将概率论推向一个新的发展阶段.19 世纪末，俄国数学家切比雪夫(П. Л. Чебышев)、马尔可夫(А. А. Марков)、李亚普诺夫(А. М. Ляпунов)等人用分析方法建立了大数定律及中心极限定理的一般形式，科学地解释了为什么实际中遇到的许多随机变量近似服从正态分布.20 世纪初受物理学的刺激，人们又开始研究随机过程.这方面柯尔莫哥洛夫(А. Н. Колмогоров)、维纳(N. Wiener)、马尔可夫、辛钦(А. Я. Хинчин)、莱维(P. Lévy)及费勒(W. Feller)等人做了杰出的贡献.

如何定义概率，如何把概率论建立在严格的逻辑基础上，是概率理论发展的困难所在，对这一问题的探索一直持续了三个世纪.二十世纪初完成的勒贝格(H. L. Lebesgue)测度与积分理论及

随后发展的抽象测度和积分理论，为概率公理体系的建立奠定了基础．在这种背景下苏联数学家柯尔莫哥洛夫 1933 年在他的《概率论基础》一书中第一次给出了概率的测度论式的定义和一套严密的公理体系．他的公理化方法成为现代概率论的基础，使概率论成为严谨的数学分支，对近几十年概率论的迅速发展起了积极的作用．

数理统计学是概率论的一个姐妹学科．研究怎样有效地收集、整理和分析带有随机性质的数据，以对所观察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议．

统计学自古有之，例如人口统计、社会调查等．但它不是现代意义下的数理统计学，只是数据的记录和整理．

数理统计学是随着概率论的发展而发展起来的．当人们认识到必须把数据看成来自具有一定概率分布的总体，所研究的对象是这个总体而不能局限于数据本身之日，也统是数理统计诞生之时．在 19 世纪中叶以前已出现了若干重要的工作，特别是高斯(C．F．Gauss)和勒让德(A．M．Legendre)关于观测数据的误差分析和最小二乘法．但数理统计学发展成为一门成熟的学科，则是 20 世纪上半叶的事．皮尔森(K．Pearson)、费希尔(R．A．Fisher)作出了重大贡献，1946 年，克拉默(H．Cramér)发表的《统计学的数学方法》是第一部严谨且比较系统的数理统计著作，可以把它作为数理统计学进入成熟阶段的标志．

数理统计学用到很多近代数学知识，但与其关系最密切的是概率论．在很大程度上可以说概率论是数理统计的理论基础，数理统计是概率论的一种应用，并且补充和丰富了概率论．它们是两个并列的数学分支，并无从属关系．

目前，概率论与数理统计的理论与方法已广泛地应用于自然科学、技术科学、社会科学及人文科学的各个领域．近年来随着科学技术的迅速发展，它在经济、管理、工程、技术、物理、气象、海洋、地质等领域中的作用愈益显著．随着计算机的发展

与普及，概率论与数理统计已成为处理信息、制定决策的重要理论和方法.概率论与数理统计向各个领域渗透，产生了许多新的分支和边缘科学，如生物统计、统计物理、数学地质、教育统计等.同时概率论与数理统计又是许多新的重要学科的基础，如信息论、控制论、排队论、预测论、可靠性理论及人工智能等.

概率论与数理统计，作为理论严谨、应用广泛、发展迅速的数学分支正日益受到人们的重视并发挥着重大的作用.

第一章 概 率

§1.1 古典概率

公元1654年7月29日，帕斯卡给著名数学家费马的信中写道，德梅尔(C, DeMere)告诉他，一颗骰子连掷4次"至少出现一个6点"的事件的"概率"(可能性)大于$\frac{1}{2}$；两颗骰子掷一次的结果6倍于一颗骰子掷一次的结果，那么两颗骰子掷24(4×6)次"至少出现一对6点"的事件的"概率"也应大于$\frac{1}{2}$，但为什么在赌场的实践结果不是这样呢？德梅尔愤怒地谴责了数学，粗暴地断言，算术是自相矛盾的.

用什么办法来计算这类事件的"概率"呢？历史上古典概率正是由研究诸如掷骰子一类赌博游戏中的问题引起的，但是经过数学家的长期努力，发现了它既易于理解、饶有兴趣，又有广泛的应用，成为数学的一个论题，因此我们从介绍古典概率开始.

我们先讨论一些简单的问题.

例1.1.1 掷一枚硬币，假如硬币的构造是均匀的，并且投掷的结果只可能是"正面向上"或"反面向上"，问掷得"正面向上"的可能性有多大？

由于硬币的构造是均匀的，因而出现"正面向上"与"反面向上"的机会是均等的，又排除了其它可能，所以我们可以断言，掷一枚硬币，掷得"正面向上"的可能性为$\frac{1}{2}$.

例 1.1.2 掷一颗骰子，设骰子的构造是均匀的，问投掷的可能结果有哪些？掷得 6 点的可能性有多大？

掷一颗骰子，只可能出现以下 6 种结果："掷得 1 点"、"掷得 2 点"、"掷得 3 点"、"掷得 4 点"、"掷得 5 点"和"掷得 6 点"．由于骰子的构造是均匀的，因而出现这 6 种结果的机会是均等的，于是我们可以断言：掷一颗骰子，"掷得 6 点"的可能性是 $\frac{1}{6}$．

例 1.1.3 连续掷两枚硬币，问可能出现的结果有哪些？连续两次均出现正面向上的可能性有多大？

我们分别用"正"表示"正面向上"，"反"表示"反面向上"，那么连续掷两枚硬币可能出现的所有结果是：

$$\Omega=\{(正，正)，(正，反)，(反，正)，(反，反)\}^{*)}$$

其中(正,正)表示两枚都出现正面；(正,反)表示第一枚出现正面，第二枚出现反面；而(反,正)则表示第一枚出现反面，第二枚出现正面；(反,反)则表示两枚都出现反面．不难想见，这四种结果的出现是等可能的，因而我们说：连续掷两枚硬币，均出现"正面向上"的可能性是 $\frac{1}{4}$．

以上三例中所进行的试验有着一些共同的特性：这样一类试验，其结果事先不能准确预言，但是一切可能出现的结果却是已知的．这样的试验称之为**随机试验**．在这类随机试验中，其可能出现的结果只有有限个且它们出现的机会是均等的．我们称这样一类试验为**古典概型**．容易看出上述三例均属古典概型．

称随机试验的一切可能结果的集合为此试验的**样本空间**，常用 Ω 表示；称样本空间的元素为**样本点或基本事件**，常用 ω（或带有脚标）表示．

---

*）集合符号表示法之一，在花括号{……}中将集合的元素一一列出，所列元素彼此不同，不论次序．

由上知，古典概型试验的样本空间是有限集，在例 1.1.1 中的样本空间 $\Omega=\{$正，反$\}$，例 1.1.2 中的样本空间 $\Omega=\{1,2,3,4,5,6\}$，其中"i"表示"掷得 i 点"，$i=1,2,3,4,5,6$. 例 1.1.3 中的样本空间 $\Omega=\{$(正，正)，(正，反)，(反，正)，(反，反)$\}$.

在古典概型中，有时需要考虑进一步的问题. 例如在例 1.1.2 中我们需要求"掷得偶数点"的可能性;在例 1.1.3 中需要求"恰有一次掷得正面"的可能性等等. 我们知道在例 1.1.2 中"掷得偶数点"是由"掷得 2 点"、"掷得 4 点"和"掷得 6 点"这三种结果组成的，出现"掷得偶数点"当且仅当出现"掷得 2 点"、"掷得 4 点"、"掷得 6 点"这三者之一，因而"掷得偶数点"的可能性应该是三者可能性之和，即 $\frac{3}{6}\ (=\frac{1}{2})$. 例 1.1.3 中连续掷两枚硬币"恰有一枚掷得正面"是由(正，反)和(反，正)这两个结果组成的. 在这一试验中，出现"恰有一枚掷得正面"当且仅当出现上述两种结果之一，因为每一个结果出现的可能性为 $\frac{1}{4}$，因而"恰有一枚掷得正面"的可能性为 $\frac{2}{4}\ (=\frac{1}{2})$.

根需上面举出的例子，自然想到称样本空间的子集为**事件**，于是每一事件由若干样本点组成，例如在例 1.1.3 中"恰有一枚掷得正面"就是$\{$(正，反)，(反，正)$\}$,例 1.1.2 中"掷得偶数点就是$\{2,4,6\}$. 由一个样本点组成的单点集"掷得 6 点"$=\{6\}$也是事件. 但不能说样本点本身是事件，因为事件是样本点的集合，而不是样本空间的元素，当我们把基本事件或样本点看作事件时一定是指该样本点组成的单点集.

有时为了方便，称某一试验中不可能发生的事件(即空集 ϕ)为**不可能事件**，称在某一试验中必然发生的事件(即全集 Ω)为**必然事件**. 如在例 1.1.2 中"掷得 7 点"、"掷得 8 点"都是空集，为不可能事件，而"掷得的点数不大于6"$=\{1,2,3,4,5,6\}=\Omega$，为必然

事件.

定义 1.1.4 若某一随机试验的样本空间Ω由有限个样本点组成，设

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

且每个样本点的出现是等可能的(即此试验为古典概型)，如果事件 A 由 m 个样本点组成，设为

$$A=\{\omega_{i_1},\omega_{i_2},\cdots,\omega_{i_m}\},$$

则定义事件 A 的概率

$$P(A)\triangleq\frac{m}{n}=\frac{\text{组成事件 }A\text{ 的样本点数}}{\text{样本空间的点数}}.$$

可以用古典概型计算的概率称之为古典概率.

显然在古典概率中有

$$P(\{\omega_1\})=P(\{\omega_2\})=\cdots=P(\{\omega_n\})=\frac{1}{n},$$

这就表示每个样本点的出现是等可能的，如果一个事件包含着 m 个样本点，我们定义它的概率为 $\frac{m}{n}$，就是认为它发生的可能性应该是 m 个样本点发生的可能性大小的总和. 上述定义是前面各个例子的概括总结.

下面我们进一步举例说明如何选择样本空间，如何计算样本空间的点数和组成某事件的样本点数以及在古典概率计算中应注意的问题.

首先我们仍回到例 1.1.3，假如我们是一次抛掷两枚硬币，而且对硬币也未加区别，这样我们可以认为试验的样本空间为

$$\Omega=\{\text{“两枚全正”},\text{“一正一反”},\text{“两枚全反”}\}$$

那么样本空间是由 3 个样本点组成，能不能说

$$P(\text{“两枚全正”})=\frac{1}{3}$$

呢？ 不能！因为在这一随机试验中 3 个样本点出现的机会是不同的，“一正一反”这一样本点出现的可能性比其它两个样本点出现的可能性要大，因而破坏了古典概型中等可能性的假设. 在这个例子中，为了运用古典概型，必须设想两个硬币是不同的 (例如标上号码)，这样就归结为例 1.1.3 的情形了.

在古典概率计算中，首要的问题是在所构造的样本空间中每个样本点的出现是等可能的.

例 1.1.5 在分别写有2, 6, 8, 11, 13 的五张卡片中任取两张，用卡片上的两个数组成一个分数，试求所得分数是既约分数的概率.

解 在分别写有 2, 6, 8, 11, 13 的五张卡片中任取两张，一切可能的结果组成的样本空间

$$\Omega=\{\{2,6\},\ \{2,8\},\ \{2,11\},\{2,13\},\ \{6,8\},\ \{6,11\},\ \{6,13\},$$
$$\{8,11\},\ \{8,13\},\ \{11,13\}\}^{*)}.$$

Ω 由 10 个点组成，由于任意两张卡片被抽到的机会是均等的，因此这些样本点的出现是等可能的. 用 A 表示两张卡片上的数能组成既约分数这一事件，则

$$A=\{\{2,11\},\ \{2,13\},\ \{6,11\},\ \{6,13\},\ \{8,11\},$$
$$\{8,13\},\ \{11,13\}\}.$$

事件 A 由7 个样本点组成. 因而

$$P(A)=\frac{7}{10}.$$

例 1.1.6 把三个分别标有 a、b、c 的球放入三个盒中，设一切可能的结果是等可能的，试求以下各事件的概率：

---

*) 我们用{2,6}表示其中的元是无序的，用(2,6)表示其中的元为有序的.

A = “第一个盒空”;

B = “三个球落入同一盒内”;

C = “第二个盒内至少有两个球”.

解 表 1－1 列出了三个球放入三个盒中的全部可能结果.其中每一个排列代表一个样本点，样本空间就由这 27 个样本点组成.事件 A 由表中第 2, 3, 16, 17, 18, 19, 20, 21 等 8 个样本点组成;事件 B 由表中第1, 2, 3 等 3 个样本点组成; 而事件 C 由表中第 2, 10, 11, 12, 16, 17, 18 等 7 个样本点组成. 因此所求事件的概率分别为:

$$P(A)=\frac{8}{27}\ ;$$

$$P(B)=\frac{3}{27}=\frac{1}{9}\ ;$$

$$P(C)=\frac{7}{27}\ .$$

表 1－1

1. $\{a\ b\ c\mid —\mid —\}$	10. $\{\ a\mid b\ c\mid —\}$	19. $\{—\mid a\mid bc\}$
2. $\{—\mid a\ b\ c\mid —\}$	11. $\{\ b\mid a\ c\mid —\ \}$	20. $\{—\mid b\mid a\ c\ \}$
3. $\{—\mid —\mid a\ bc\}$	12. $\{\ c\mid a\ b\mid —\ \}$	21. $\{—\mid c\mid a\ b\}$
4. $\{a\ b\mid\ c\mid —\}$	13. $\{\ a\mid —\mid b\ c\ \}$	22. $\{\ a\mid b\mid c\ \}$
5. $\{\ a\ c\mid b\mid —\}$	14. $\{\ b\mid —\mid a\ c\ \}$	23. $\{a\ \mid c\mid b\ \}$
6. $\{\ b\ c\mid a\mid —\}$	15. $\{\ c\mid —\mid a\ b\ \}$	24. $\{\ b\mid a\mid c\ \}$
7. $\{\ a\ b\mid —\mid\ c\ \}$	16. $\{—\mid a\ b\mid c\ \}$	25. $\{\ b\mid c\mid a\ \}$
8. $\{\ a\ c\mid —\mid\ b\ \}$	17. $\{—\mid a\ c\mid b\ \}$	26. $\{\ c\mid a\mid b\ \}$
9. $\{\ b\ c\mid —\mid\ a\ \}$	18. $\{—\mid b\ c\mid a\ \}$	27. $\{\ c\mid b\mid a\ \}$

其中$\{\ a\ b\ c\mid\ —\mid —\}$表示 a、b、c 三球都落入第一盒中，$\{\ a\mid bc\mid —\ \}$表示 a 球落入第一盒中，b、c 两球落入第二盒中等等.

从上面两个例子可以看出，对于比较简单的古典型试验，只要把一切可能的结果(即样本空间的点)一一列出，数清个数，并

能从实际或理论上判断它们出现的可能性是相同的，然后对所述事件 A 的含义分析清楚，把它所包含的样本点一一列出，数清个数，求事件 A 的概率的问题就迎刃而解了. 因此从某种意义上说，求古典概率的问题，是个数数问题.

习题 1－1

1. 试用例 1.1.6 中样本空间的子集表示下列事件(示例："三个球都落入同一盒中"$=\{1,2,3\}$):

(1) a 球不在第三个盒中;

(2) 第二个盒中至少有一个球;

(3) 第一个盒中恰好有一个球.

2. 试用语言叙述例 1.1.6 中的下列事件:

(1) $\{22,23,24,25,26,27\}$;

(2) $\{1,3,7,8,9,13,14,15\}$;

(3) $\{3,13,14,15,19,20,21\}$.

3. 给出随机试验"从包含两件次品 a_1，a_2 和三件正品 b_1，b_2，b_3 的五件产品中依次取出两件(每次取出后不放回)"的样本空间，写出下列事件的集合表示并计算它们的概率:

A_i："恰好取得 i 件次品"，$i=0,1,2$;

B_1："第一次取到的是次品".

4. 给出随机试验"从包含两件次品 a_1，a_2 和三件正品 b_1，b_2，b_3 的五件产品中依次取出两件(每次取出后放回)"的样本空间，写出下列事件的集合表示，并计算它们的概率:

A_i："恰好取得 i 件次品"，$i=0,1,2$;

B_1："第一次取到的是次品";

C："两次取到的是同一件产品".

5. 给出随机试验"从包含两件次品和三件正品的五件产品中任取两件"的样本空间，写出下列事件的集合表示并计算它们的概率:

A_i："恰好取到 i 件次品"，$i=0,1,2$.

6. 从 1，2，3，4 四个数字中依次取出三个(每次取出后不放回)，从左到右排成一个三位数，试求

(1) 这个三位数能被 3 整除的概率;

(2) 这个三位数能被 6 整除的概率.

当一个古典概型试验的样本空间所含样本点数比较多时，我们就很难 (或事实上不可能) 把样本点一一列出然后进行数数.这时要用到一些计数的工具，例如排列和组合的知识.即令是样本点不多的情形，有时运用排列和组合的知识求古典概率，也比直接数数来得简单，例 1.1.6 便是这样，读者不妨思考一下.

例 1.1.7 袋中有5 个白球 4 个黑球，今从袋中任取两球，问取得白球黑球各一个的概率是多少?

解 1 袋中共有9 个球，从这 9 个球中任取两个球 (不论次序) 共有 $C_9^2 = \dfrac{9\times 8}{2} = 36$ 种取法，所以样本空间由 36 个样本点组成，这些样本点出现的机会是彼此相同的.

用 A 表示事件"取得白球黑球各一个"，它是在5 个白球中任取其一，4 个黑球中任取其一，因此由 $5\times 4 = 20$ 个样本点组成，故

$$P(A) = \frac{20}{36} = \frac{5}{9}.$$

解 2 从袋中任取两球，我们可以看作是先取一个球(不放回)，再取一个球. 这时一切可能的取法总数为 $9\times 8 = 72$. 它们出现的可能性也是相同的. 这时"取得白球黑球各一个"也可以区分为"先取出一个白球，后取出一个黑球" (有 $5\times 4 = 20$ 种取法) 和"先取出一个黑球，后取出一个白球"(有 $4\times 5 = 20$ 种取法)，这时组成事件 A 的样本点数为40，因此

$$P(A) = \frac{40}{72} = \frac{5}{9}.$$

这个例子说明，在有些问题中样本空间可有不同的选择，但无论怎样选择，样本空间的各个样本点应该具有等可能性，并且在选定样本空间后，我们所关心的事件能够表示成所选的样本空

间的子集，二者必须一致起来.例如在解1中样本空间中的点由两个球组成，不论次序，如果把5个白球标以白$_1$，白$_2$，白$_3$，白$_4$，白$_5$；4个黑球标以黑$_1$，黑$_2$，黑$_3$，黑$_4$，那么{白$_1$，黑$_2$}与{黑$_2$，白$_1$}是同一个样本点(集合{…}中元素的排列是不计次序的).因而在考虑事件A由哪些样本点组成时，也只顾及样本点由哪两个球组成，不计次序.但在解2中，(白$_1$，黑$_2$)与(黑$_2$，白$_1$)视为两个样本点，在考虑事件A由哪些样本点组成时，不仅要考虑由哪两个球组成，还要考虑两个球取出的次序，必须与样本空间的选择一致起来.

现在我们来解决本节开始提出的问题.

例1.1.8 计算一颗骰子掷4次至少出现一个6点的概率和两颗骰子掷24次至少出现一对6点的概率.

解 情形1：一颗骰子连续掷4次算作一次试验，也可以看作标有号码的4颗骰子投掷一次算作一次试验.于是任一次试验的结果可表示为$(\omega_1, \omega_2, \omega_3, \omega_4)$，其中$\omega_i$ $(i=1,2,3,4)$表示第i个骰子出现的点数，而样本空空间Ω可表示为：

$$\Omega=\left\{(\omega_1,\omega_2,\omega_3,\omega_4):\ \begin{matrix}\omega_i=1,2,3,4,5,6;\\ i=1,2,3,4.\end{matrix}\right\}^{*)}$$

由此容易看出：Ω共有6^4个样本点，它们是等可能的.为了表述方便，用ξ表示在4次投掷中6点出现的次数，亦即样本点$(\omega_1, \omega_2, \omega_3, \omega_4)$中$\omega_i=6$的个数.显然$\xi$只能取0,1,2,3,4.我们所关心的事件"至少出现一个6点"就可以表示为"$\xi\geqslant 1$".为计算"$\xi\geqslant 1$"这一事件所含样本点数，我们转而计算"$\xi=0$"这一事件所含样本点数，然后从样本空间点数减去"$\xi=0$"所含样本点数，即得"$\xi\geqslant 1$"所含样本点数.注意到"$\xi=0$"中所含样本点的特点是只出现1,2,3,4,5点而不出现6点，于是有

---

*)集合的另一种表示法.{ω:…}中冒号":"以前表示集合中元素的形式，":"以后表示元素所具有的性质.这个集合是具有":"后所述性质的一切元素的集合.

$$\text{“}\xi=0\text{”}=\left\{(\omega_1,\omega_2,\omega_3,\omega_4):\ \begin{matrix}\omega_i=1,2,3,4,5;\\ i=1,2,3,4.\end{matrix}\right\}.$$

由此容易看出“$\xi=0$”所含样本点数为 5^4. 这样，“$\xi\geqslant 1$”所含样本点数为 6^4-5^4，由定义 1.1.4 知所求概率

$$P(\xi\geqslant 1)=\frac{6^4-5^4}{6^4}\approx 0.5177.$$

情形 2：两颗骰子连续投掷 24 次算作一次试验. 如果把两颗骰子标以号码 1，2，那么第 i 次投掷两颗骰子的结果可以表示为 $\omega_i=(\omega_{i1},\omega_{i2})$，$\omega_{i1},\omega_{i2}=1,2,3,4,5,6$ $(i=1,2,\cdots,24)$. 所以样本空间可表为

$$\Omega=\left\{(\omega_1,\omega_2,\cdots,\omega_{24}):\ \begin{matrix}\omega_i=(\omega_{i1},\omega_{i2}),\omega_{i1},\omega_{i2}=1,\\ 2,3,4,5,6;i=1,2,\cdots,24.\end{matrix}\right\}$$

由于 $\omega_i=(\omega_{i1},\omega_{i2})$，$\omega_{i1},\omega_{i2}=1,2,3,4,5,6$ 有 36 个不同结果，Ω 有样本点 36^{24} 个，它们是等可能的. 用 η 表示在两颗骰子投掷 24 次中两颗骰子都出现 6 点的次数，显然 η 只能取 $0,1,2,\cdots,24$. 我们所关心的事件“至少出现一对 6 点”可表为“$\eta\geqslant 1$”. 类似于情形 1，为计算“$\eta\geqslant 1$”所含样本点数，只需计算“$\eta=0$”所含样本点数. 由前面的表示法，“$\eta=0$”所含的样本点 $(\omega_1,\omega_2,\cdots,\omega_{24})$ 中每一 ω_i 不取 $(6,6)$，即每个 ω_i 有 35 种取法. 故“$\eta=0$”所含样本点数为 35^{24}. 这样“$\eta\geqslant 1$”所含样本点数为 $36^{24}-35^{24}$，故所求概率

$$P(\eta\geqslant 1)=\frac{36^{24}-35^{24}}{36^{24}}\approx 0.4914.$$

上例的计算表明，德梅尔凭经验所得的结论是正确的，但他

的推理是错误的.

下面再讨论几个典型的问题. 许多实际问题可以化作这些模型来计算，或者可以借助于在这些模型计算中用到的方法. 由于以后多处应用，先引进一个有用的符号:

$$\binom{n}{r}=\frac{n(n-1)\cdots(n-r+1)}{r!}, \tag{1.1.9}$$

其中 n 是任意实数，r 为非负整数(约定 $0!=1$). 显然，当 n, r 为任意非负整数时，$\binom{n}{r}=C_n^r$ 为从 n 个元素中取 r 个的组合数，但是它比组合数记号 C_n^r 更方便. 因为当 n 为任意实数时，它仍有意义，而且有

$$\binom{n}{r}=0, \quad \text{当 } n \text{ 为非负整数且} n<r. \tag{1.1.10}$$

例 1.1.11 设有 r 个不同的球，每个球以同样的可能性 $(\frac{1}{n})$ 落入 n 个盒中的任一个. 若 $r \leqslant n$，试求以下事件的概率:

$A=$"指定的 r 个盒中各有一个球";

$B=$"恰有 r 个盒中各有一个球".

例 由于球是可区别的，因面相当于可将 r 个球进行编号，设为 $a_1, a_2, \cdots, a_r$. 将盒子也进行编号，设为 $1, 2, \cdots, n$. 每一个球有 n 种不同的放法，配合起来，r 个球放入 n 个盒中共有 n^r 种不同的方法. 这是整个样本空间的点数. 每个样本点出现的可能性是相同的.

现在来计算 A 的概率，为此要分析 A 的样本点的特征，先指定 r 个盒子，设为 $(i_1, \cdots, i_r)$ (即第 $i_1, i_2, \cdots, i_r$ 个盒子)，对于 $a_1, \cdots, a_r$ 的任一个排列 $a_{j_1}, \cdots, a_{j_r}$，可以从左到右顺序对号入座地放进上述 r 个盒子中，这样就得例一个属于 A 的样本点. 易见任何

两个排列按此法得到的样本点不同，而且 A 的任一个样本点都一定是上述样本点中的一个.因此 A 中的样本点数等于 $a_1,\cdots,a_r$ 的排列数 $r!$，故

$$P(A)=\frac{r!}{n^r}.$$

为计算事件 B 的概率，要分析“恰有 r 个盒中各有一球”是由哪些样本点组成.首先可以从 n 个盒子中选取 r 个盒子，然后再将 r 个球分别放入这 r 个盒中，其它 $n-r$ 个盒子空着.我们知道在 n 个盒子中任取 r 个盒子有 $\binom{n}{r}$ 种取法；对于每取出的 r 个盒子各放一球又有 $r!$ 种不同的放法.根据计数的乘法原理，“恰有 r 个盒中各有一球”的样本点数应为 $\binom{n}{r}\cdot r!$，因此

$$P(B)=\frac{\binom{n}{r}\cdot r!}{n^r}=\frac{n(n-1)\cdots(n-r+1)}{n^r}.$$

例 1.1.12 一个班级有 m 个人，试求他们的生日都不相同的概率.为使这一概率小于 $\frac{1}{2}$，m 需要多大(为简单计，设一年为 365 天，并设生日在哪一天出现是等可能的)?

解 这一问题可以看作把 m 个球投入365 个盒子中，求恰有 m 个盒中各落入一个球的概率.由上例的结果可知所求概率为

$$p_m=\frac{\binom{365}{m}\cdot m!}{365^m}=\prod_{r=1}^{m-1}\left(1-\frac{r}{365}\right).$$

易见 p_m 是 m 的递减函数，且由不等式

$$1-x\leqslant e^{-x},\ x\geqslant 0$$

知

$$p_m \leqslant e^{-\sum_{r=1}^{m-1}\frac{r}{365}} = e^{-m(m-1)/730}.$$

当 $m=23$ 时，上式右端近似于 $0.499998 < \frac{1}{2}$，所以当 $m \geqslant 23$ 时.

$$p_m \leqslant p_{23} < \frac{1}{2}.$$

经过数字计算，还可知当 $m<23$ 时，$p_m > \frac{1}{2}$.

当 $m=50$ 时 $p_m<0.03$，也就是说在 50 个人的班级中每个人的生日都不同的概率小于 0.03，是十分罕见的，这与我们的直观想象有所不同.

例 1.1.11 和例 1.1.12 表面上是两个不同的问题，但数学形式却完全相同，如果善于总结经验，从千变万化的实际问题中看出它们共有的特性，就会做到触类旁通，举一反三.

例 1.1.13 把 r 个不可分辨的质点放入 n 个盒中，假如每个盒子可以容纳任意多个质点，而且所有的放法都是等可能的，试求以下两个事件的概率：

A = "指定的 r 个盒中各落入一个质点"；

B = "恰有 r 个盒中各有一个质点".

解 这个随机试验与例 1.1.11 不同，以三个球(质点)三个盒为例：在例 1.1.11 中 $|ab|c|$，$|ac|b|$ 是两种不同的方法，在本例中，由于质点不可辨，上述两种放法将不加分辨算作一种放法，于是不同的放法如图 1－1 所示.

为了计算这种放法的个数，我们将三个盒连放在一起，用两个公共的墙"|"隔开. 每一种放法都对应着三个质点 0，0，0 和两

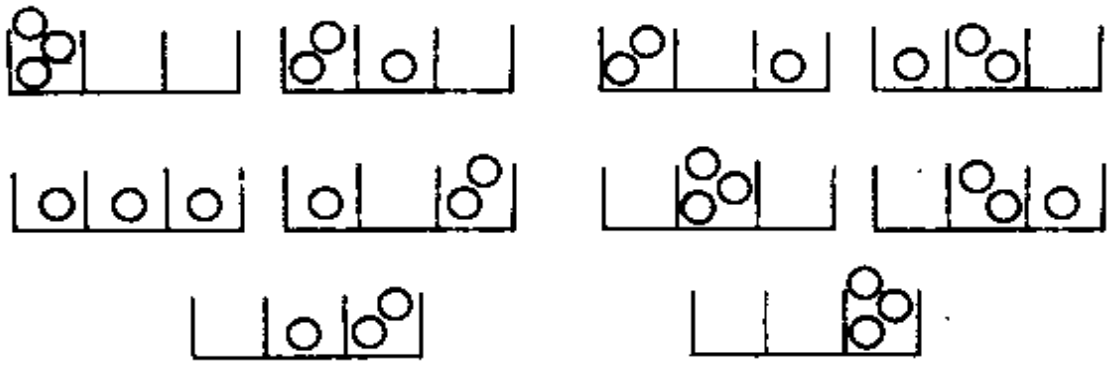


图 1-1

个(公共)墙 |,| 的一个排列. 这个对应是一一对应，所以只需算出三个质点 0,0,0 和两个墙 | ,| 排列的个数. 每一个排列有五个位置，每选出 3 个位置放"0"就得到一个这样的排列. 所以共有

$$\binom{5}{3}=\frac{5\times4\times3}{1\times2\times3}=10$$

种排法. 这就求出了将 3 个不可辨的质点放入 3 个盒中有 $\binom{5}{3}$ 种放法.

一般，将 r 个不可辨的质点放入 n 个盒中，每盒容纳的质点个数不限，相当于 r 个"0"和 $(n-1)$ 个墙"|"的排列，有 $\binom{n+r-1}{r}$ 种放法. 这就是在这一随机试验中样本空间的点数. 根据题设，每种放法具有等可能性.

"指定的 r 个盒中各有一个质点"仅由一个样本点组成，因此

$$P(A)=1\Big/\binom{n+r-1}{r}.$$

"恰有 r 个盒中各有一个质点"，相当于在 n 个盒中选取 r 个

盒子的选法数，因此

$$P(B)=\binom{n}{r}\Big/\binom{n+r-1}{r}.$$

注　有兴趣的读者不妨尝试证明下述事实：把 r 个不可分辨的质点放入 n 个盒子的方法数等于从 n 个元中可重复地取 r 个元的方法数.

例 1.1.14　投掷4颗骰子，试求“点数之和等于9”的概率.

解　在例 1.1.8 中我们已经计算出在投掷 4 颗骰子的试验中，样本空间的点数为 6^4，它们具有等可能性.

为计算“4 颗骰子的点数之和等于 9”这一事件所含样本点数，我们不妨把 1 点用一个“0”来表示，2 点用“0　0”表示等，于是样本点(6, 1, 1, 1)可表示为

$$0\quad 0\quad 0\quad 0\quad 0\quad 0|0|0|0.$$

这样就将计算“4 颗骰子的点数之和等于 9”所含样本点数的问题归结为将 9 个质点(不可分辨)放到 4 个盒子中“没有盒子空着”的放法数. 因为不允许有空盒，相当于在一字排开的 9 个质点(“0”)产生的 8 个缝隙中选择 3 个缝隙装上隔板，共有 $\binom{8}{3}$ 种不同的选择方法，故所求概率为

$$\binom{8}{3}\Big/6^4=\frac{8\times 7\times 6}{1\times 2\times 3}\Big/6^4=\frac{7}{162}.$$

注　请读者考虑，在此例中若要求“4 颗骰子点数之和等于 10”的概率，该怎么办?

以上两例虽然是完全不同的问题，但在方法上却可以互相借鉴.

注　r 个质点放入 n 个盒中，如果将质点加以区别，按照例 1.1.11，共有 n^r 种放法，每种放法的概率都是 n^{-r}，物理学家把这种赋

概方法称为马克斯威尔 － 玻耳兹曼(Maxwell－Boltzman)统计法则. 如果按例 1.1.13 的办法对质点不加区别，r 个质点放入n个盒中共有 $\binom{n+r-1}{r}$ 种放法，每种放法的概率为$\binom{n+r-1}{r}^{-1}$，物理学家把这种赋概方法称为波色－ 爱因斯坦(Bose－Einstein)统计法则. 光子、核子和原子均遵从这一统计法则.还有一种赋概方法称之为费米 － 狄拉克(Fermi－Dirac)统计法则，即质点不可辨且每个盒中至多能放入一个质点. 这时自然要求 $r\leqslant n$, r 个质点落入 n 个盒中共有 $\binom{n}{r}$ 种不同的放法，每种放法的概率为$\binom{n}{r}^{-1}$，电子、质子和中子都遵从这一法则.

例 1.1.15 剧院售票处有$2n$ 个人排队买票，其中 n 个人有 5 角钱一张的钞票，其余 n 个人只有 1 元一张的钞票. 开始售票时售票处无钱可找，而每个人只买一张 5 角钱的票.求售票处不会找不出钱的概率.

解 购票者的一切可能排法都是等可能的. 我们采用下面的几何方法：在坐标系 Oxy 之下，用 $x=1,2,\cdots,2n$ 代表购票者在队列中所占据的位置. 从 $(0,0)$点开始，每遇到一位带 5 角钱的购票者纵坐标减 1，而遇到只带 1 元钱的购票者纵坐标加1，在 Oxy 平面上得到 $2n+1$ 个点，把这 $2n+1$ 个点用折线连接起来. 因为有 n 个带5 角钱的，n 个只带 1 元钱的购票者，因此折线是始于$(0,0)$，终于$(2n,0)$.在组成折线的 $2n$ 节线段中n 节向上，n 节向下，一切可能的不同样式的折线为 $\binom{2n}{n}$ 条，相当于在 $2n$ 个位置中选择n 个位置使在这些位置上纵坐标减 1.其余位置上纵坐标加 1.每一条这样的折线看作一个样本点，它们的出现是等可能的.

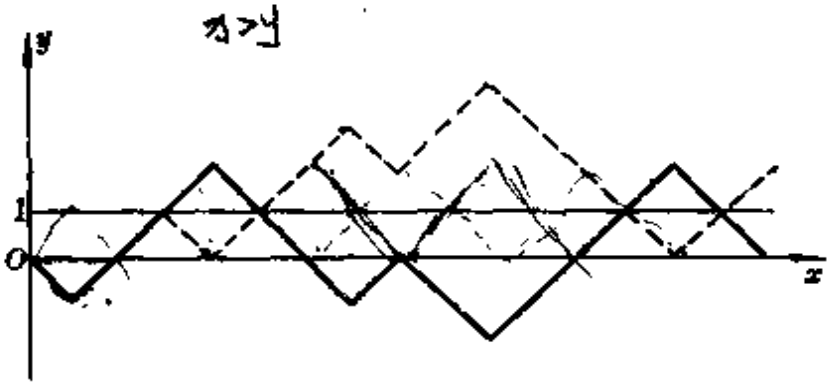


图 1－2

设 A 表示售票处不会找不出钱的事件，它是由所有不越过横坐标轴且在它下面的折线(样本点)组成的．为计算组成事件A的样本点，我们转而计算至少有一次达到或越过直线 $y=1$ 的那些折线的数目．为计算这种折线的数目，我们采用反射的方法构造新的折线:当折线未触及直线 $y=1$ 前，新折线与旧折线一致，一旦接触直线 $y=1$ 后，从接触点开始新折线是旧折线以直线 $y=1$ 的镜面反射.因此新折线是起于(0, 0)而终于(2n, 2)的折线(见图1－2)．这些折线仍由 2n 节组成，但向上的节数比向下的节数多2,共有 $\binom{2n}{n+1}$ 条．因此事件 A 所含样本点数，即不越过直线 $y=1$ 的折线数为 $\binom{2n}{n}-\binom{2n}{n+1}$,事件$A$ 的概率

$$P(A)=\frac{\binom{2n}{n}-\binom{2n}{n+1}}{\binom{2n}{n}}=\frac{1}{n+1}.$$

这一例子说明，计算古典概率有时并非易事，要借助于其它数学方法，要善于分析问题，将问题一步一步转化为我们能够进

行计算的形式.

习题 1－2

1. 把 10 本书任意排在书架的一层上，试求指定的 3 本书放在一起的概率.

2. 袋中装着 5 个白球 6 个黑球，现从袋中依次任意取出 3 个球，每次取出后不放回. 试求"取出的第一个球是黑球、第二个球是白球、第三个球是黑球"的概率.

3. 从 0, 1, 2, …, 9 十个数字中依次任意取出 4 个数字排成一行(每次取后不放回)，试求构成一个 4 位偶数的概率(注意："0"不能放在首位).

4. 在分别写着 1, 2, …, N 的 N 张卡片中任意取出 n 张($n<N$)，然后将这 n 张卡片上的数从小到大排好，试求第 m 个数是 M 的概率($1 \leqslant M \leqslant N$).

5. 将 7 个球任意放入顺序排好的 7 个盒子中，每个球进入每个盒子中的概率都是 $\frac{1}{7}$，试求 7 个盒子中的球数依次是 2, 0, 3, 1, 0, 1, 0 的概率.

6. 甲袋中装有 3 个白球，7 个红球，15 个黑球；乙袋中装有 10 个白球，6 个红球，9 个黑球. 现从两袋中各取一球，试求"两个球颜色相同"的概率.

7. 从 0, 1, 2, …, 9 这十个数字中有放回地任取四个数，试求下列事件的概率：

(1) 四个数字全不相同；

(2) 恰有两个数字相同；

(3) 恰有三个数字相同；

(4) 四个数字全相同；

(5) 只出现两个数字.

§1.2 超几何分布与产品检验

——古典概率的一个应用

古典概型是非常简单的概率模型，但它已经在实际中有许多应用，本节举一个在工业生产的质量检查中经常用到的例子，我们还是从一类抽球模型谈起.

例1.2.1 设袋中有N个球，其中有 M 个是红球，$N-M$个是黑球，从中任取 n 个球，问恰有k个红球的概率是多少?

解 我们用 ξ 表示抽得的 n 个球中所含红球数，它可以取 $\max\{0, n+M-N\}$和 $\min\{n, M\}$之间的任何整数.

从 N 个球中取n 个球共有$\binom{N}{n}$种不同的方法，它们都是等可能的，其中恰有 k 个红球，就是说有 k 个红球 $n-k$ 个黑球，而 k 个红球是从 M 个红球中取的，共有$\binom{M}{k}$ 种取法，$n-k$ 个黑球是从 $N-M$ 个黑球中取的，共有 $\binom{N-M}{n-k}$种取法，由于 k 个红球的任意一种取法都可以和$n-k$ 个黑球的任意一种取法相配合，所以我们所求的概率

(1.2.2)

$$P(\xi=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad \max\{0,n+M-N\}\leqslant k\leqslant \min\{n,M\}$$

由(1.1.10)知当 $k>M$ 或 $n-k>N-M$(即 $k<n+M-N$)时，上述表达式自动为零，因而也可以定义作

(1.2.2)′ $$P(\xi=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,2,\cdots,n.$$

如此定义的这一组概率，称为**超几何分布**.

在实际应用中，有许多问题可以归结为上述模型.

例 1.2.3 由180 只集成电路组成的一批产品中有 8 只是次品，现从中任意抽验 4 只，用 ξ 表示其中的次品数. 试求

(1) 抽验的 4 只中恰好有 k 只次品(即 $\xi=k$)的概率，$k=0,1,2,3,4$;

(2) 抽验的 4 只产品中次品超过一只(即 $\xi>1$)的概率.

解 按照例 1.2.1 中超几何分布的算法，不难求得

$$P(\xi=k)=\frac{\binom{8}{k}\binom{180-8}{4-k}}{\binom{180}{4}},\quad k=0,1,2,3,4.$$

其近似值由下表所示

k	0	1	2	3	4
$P(\xi=k)$	0.8324	0.1576	0.0097	0.0002	0.0000

由于组成事件$\{\xi>1\}$的样本点数是组成$\{\xi=2\}$，$\{\xi=3\}$,$\{\xi=4\}$的样本点数之和，所以

$$\begin{aligned}P(\xi>1)&=P(\xi=2)+P(\xi=3)+P(\xi=4)\\&\approx 0.01\end{aligned}$$

(利用事件的运算及概率的性质更易得出上述等式，这些将在 §1.5 中阐述).

进一步的计算 (直观上也可以看出)表明，如果次品数少于 8 只，那么 $P(\xi>1)$的概率会更小 (详见本节习题 6). 这表明：由 180 只集成电路组成的一批产品中，如果次品数不超过 8 只，则从中抽验 4 只，其中次品数$\xi>1$ 的可能性很小，不超过 1%.

假设有 180 只集成电路组成的一批产品，不知道其中有多少

次品.在抽验的 4 只产品中,发现有3 只是次品.现在问:如果根据这个抽验的结果判断这批产品中次品是否超过8 只,该怎样回答呢?更确切地说,如果规定:180 只为一批,一个合格批中次品不能超过8 只,那么这批产品是合格的,还是不合格的?

因为如果这批产品是合格的,那么根据上面的计算,抽验的 4 只产品中次品数超过1 的可能性很小,如今抽得的4 只产品中竟有3 只次品,可能性很小的事件居然出现了,这使我们有理由怀疑这批产品是合格的,因而可以做出这批产品是不合格的判断.

自然,做出这种判断是会犯错误的,假如这批产品真的只有 3 件次品,抽验的 4 只中碰巧包含了这 3 只次品,这时判断这批产品不合格,犯了把合格批断定为不合格批的错误.不过犯这种错误的概率很小,不超过 1%.称这种错误为第一类错误.允许犯第一类错误的概率用 α 表示.

我们可以根据这一想法为厂家制定一个产品检验方案:用 H 表示需要做出判断的命题(在统计中称为原假设).

H:由 180 只产品组成的一批产品中次品数不超过 8.

抽验方案:在一批产品中抽取 4 只做检验,用 ξ 表示其中次品数.如果 $\xi \leqslant 1$,则接受原假设 H(即认为这批产品是合格的);如果 $\xi > 1$,则拒绝原假设 H (即认为这批产品不合格,次品数大于 8).

犯第一类错误的概率,即在原假设 H 为真的条件下拒绝原假设的概率 $\alpha < 0.01$.

当然也可能犯第二类错误,即在原假设不真时接受原假设.这里我们不去深入探讨.

本例属于数理统计中的假设检验问题,在第五章中将对这类问题做进一步的讨论.

例 1.2.4 (捕鱼问题) 从一个养鱼池里捕出1000尾鱼,涂以红点后放回鱼池,隔了一段时间当鱼充分混合后,又捕出 1000尾鱼,发现其中有 100 尾鱼涂有红点,根据这些数据,应该对池

中鱼的总数做出怎样的估计 (假定两次捕鱼中间池中鱼的总数未变，而且两次捕鱼都是在整个鱼池中随机捕捉的)?

现在我们用数理统计中的最大似然估计法来估计池中鱼的总数，并且通过这个例子介绍最大似然估计的思想.

解 为使问题一般化，我们令

$N=$ 池中鱼的总数(未知);

$M=$ 第一次捕出的鱼数 (涂以红点);

$n=$ 第二次捕出的鱼数;

$k=$ 第二次捕出的 n 条鱼中涂有红点的鱼数;

$q_k(N)=$ 第二次捕出的鱼中恰有 k 尾涂有红点鱼的概率.

由例 1.2.1 知，如果 N 已知，则

$$q_k(N)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

如今 M,k,n 均为已知，N 为未知，我们只知道前后两次共捕到过 $M+n-k$ 尾鱼，N 应该不小于这个数. 在我们的例子中 $M+n-k=1900$, 假定 $N=1900$, 那么由 $M=n=1000$, $k=100$ 易知

$$q_k(N)=\frac{\binom{1000}{100}\binom{900}{900}}{\binom{1900}{1000}}=\frac{(1000!)^2}{100!900!}.$$

由斯特林公式可以证明这个概率的数量级将是 10^{-430}, 因此可以认为，这个假定 ($N=1900$) 是不太合理的，同样如果我们假定 N 很大，例如 $N=1000000$, 经计算 $q_{100}(N)$ 也是非常非常小的，因而

这种假定也不太合理. 那么什么样的 N 才合理呢? 自然我们想到要求使$q_k(N)$达到最大值的那个 N, 也就是要求那样的 N, 使得我们捕捉的 n 条鱼中恰有 k 条带有红点的概率最大. 对于一组特定的观察值M,n,k, 我们把使 $q_k(N)$达到最大值的那个 N 记作$\hat{N}$, 称之为 N 的**最大似然估计**. 为求 $\hat{N}$, 考虑比值

$$\frac{q_k(N)}{q_k(N-1)}=\frac{(N-M)(N-r)}{(N-M-r+k)N}.$$

由简单的计算知, 当 $Nk<Mn$ 时这个比值大于1, 当 $Nk>Mn$ 时这个比值小于1. 因而当 $N<\frac{Mn}{k}$ 时 $q_k(N)>q_k(N-1)$, 当 $N>\frac{Mn}{k}$ 时, $q_k(N)<q_k(N-1)$, 当 $N=\left[\frac{Mn}{k}\right]$ ($[x]$表示不超过x的最大整数)时, $q_k(N)$最大. 即 N 的最大似然估计

$$\hat{N}=\left[\frac{Mn}{k}\right].$$

在我们的例子中 $\hat{N}=\frac{1000\times 1000}{100}=10000$, 即池中鱼的总数估计在10000尾左右.

例 1.2.5 在例1.2.1中从 N 个球中一次取 n 个, 相当于依次取 n 个, 每次取出后不放回. 最如每次取出一个球后放回袋中, 试求 n 次抽取中恰有 k 次抽到红球的概率.

例 从 N 个球中每次取一个, 共取 n 次, 一切可能的取法数为N^n, 每种取法的可能性是相同的, 在 n 次抽取中选定 k 次有$\binom{n}{k}$种不同的选法, 而在指定的 k 次抽到红球另外 $n-k$ 次抽到黑球

共有 $M^k(N-M)^{n-k}$ 种取法. 因此在 n 次抽取中恰有 k 次抽到红球(用 $\xi=k$ 表示)有 $\binom{n}{k}M^k(N-M)^{n-k}$ 种不同的取法. 于是

$$P(\xi=k)=\frac{\binom{n}{k}M^k(N-M)^{n-k}}{N^n}=\binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}.$$

若令 $p=\frac{M}{N}$ 表示红球占的比率, 则

$$P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,2,\cdots,n.$$

如此定义的这一组概率, 称之为二项分布(这里的 p 是有理数, 事实上 p 可以是 0 与 1 之间的任意实数). 关于二项分布, 我们在第二章还要仔细讨论.

在例 1.2.1 中, 如果始终保持 $\frac{M}{N}=p$, 而令 $N\to\infty$, 我们有

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\frac{\frac{M!}{k!\,(M-k)!}\cdot\frac{(N-M)!}{(n-k)!\,(N-M-n+k)!}}{\frac{N!}{n!\,(N-n)!}}$$

$$=\frac{n!}{k!\,(n-k)!}\cdot\frac{M(M-1)\cdots(M-k+1)(N-M)(N-M-1)\cdots(N-M-n+k+1)}{N(N-1)\cdots(N-n+1)}$$

$$=\binom{n}{k}\frac{M}{N}\ \frac{(M-1)}{(N-1)}\cdots\frac{(M-k+1)}{(N-k+1)}\ \frac{(N-M)}{(N-k)}\ \frac{(N-M-1)}{(N-k-1)}$$

$$\cdots\frac{(N-M-n+k+1)}{(N-n+1)}$$

$$=\binom{n}{k}\frac{M}{N}\frac{\left(\frac{M}{N}-\frac{1}{N}\right)}{\left(1-\frac{1}{N}\right)}\cdots\frac{\left(\frac{M}{N}-\frac{k-1}{N}\right)}{\left(1-\frac{k-1}{N}\right)}\frac{\left(1-\frac{M}{N}\right)}{\left(1-\frac{k}{N}\right)}$$

$$\cdot\frac{\left(1-\frac{M}{N}-\frac{1}{N}\right)}{\left(1-\frac{k+1}{N}\right)}\cdots\frac{\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)}{\left(1-\frac{n-1}{N}\right)}$$

$$\to\binom{n}{k}p^k(1-p)^{n-k}.$$

因此，当 N 很大，$\frac{M}{N}$ 一定时，放回抽样与不放回抽样是近似的. 这也说明了超几何分布与二项分布的关系.

习题 1－3

1. 从 1,2,…,n 这 n 个数中任取两个，求其中一个小于 k，另一个大于 k 的概率 $(1<k<n)$.

2. 将 52 张扑克牌均分为两部分，求每一部分中红花色牌的张数与黑花色牌的张数相同的概率.

3. N 只螺钉中有 M 只次品. 从这批螺钉中任取 5 只, 发现其中有 k 只次品.

(1) 若已知 $N=200$, $M=5$, 求"$k=2$"的概率;

(2) 若 $N=200$, M 未知, 设原假设为 H: $M\leqslant 5$, 问 k 不超过多少才能使犯第一类错误的概率不超过0.001?

(3) 若已知 $M=5$, $k=2$, 求 N 的最大似然估计.

4. 四个人打桥牌, 试求你手中有 5 张黑桃, 而另 8 张黑桃全在你的同伴手中的概率.

5. 假设 60 只灯管中有一等品 15 只, 二等品 40 只, 三等品 5 只, 从中任意抽取 10 只, 问其中一、二、三等品分别为 5 只、3 只、2 只的概率是多大?

6. 从 N 个产品中取 n 个进行检验, 用 ξ 表示其中的次品数. 令 $P_M(\xi=k)$表示在 N 个产品中有 M 个次品时检验的 n 个产品中次品数 $\xi=k$ 的概率. 试证: 当 $M>1$ 时恒有

$$P_M(\xi=0)+P_M(\xi=1)<P_{M-1}(\xi=0)+P_{M-1}(\xi=1).$$

§1.3 几何概率

在概率论发展的早期, 人们就已经注意到只考虑那种仅有有限个等可能结果的随机试验是不够的, 还必须考虑试验结果是无穷多个的情形. 这中间最简单的一类是试验结果有无穷多个而又有某种等可能性的场合, 这就是属于几何概型的随机试验. 1777 年法国科学家蒲丰(Buffon)提出的著名的投针试验就是一例: 平面上画有一族平行线, 其中任何相邻的两条线距离是 $a(a>0)$, 向平面任意投掷一枚长为 l $(l<a)$的针, 试求针与平行线相交的概率.

我们还是从简单的问题入手.

例 1.3.1 某人午觉醒来, 发觉表停了, 他打开收音机, 想听电台报时, 求他等待的时间不多于 10 分钟的概率.

一种很自然的想法是所求的概率为 $\frac{1}{6}$. 因为电台每小时报

时一次，用 η 表示某人打开收音机的时刻(不计小时，单位为"分")，它是在两次报时之间，即 $0\leqslant\eta\leqslant 60$，我们假设他在 60 分钟内什么时刻打开收音机都是等可能的，这时等可能性表现在：设某区间包含在 $[0,60]$ 内，只要它长度一定，不论它位于何处，则在此区间内打开收音机的可能性是相同的．用公式表达就是：对任何正数 s, t，$s+t\leqslant 60$ 来说，$P(0\leqslant\eta\leqslant t)=P(s\leqslant\eta\leqslant s+t)$．这就是说，$\eta$ 在区间 $[a,b]\subset[0,60]$ 中取值的概率与区间的长度成比例，即有正数 λ 使

$$P(a\leqslant\eta\leqslant b)=\lambda\ (b-a),\quad 当\ \ 0\leqslant a<b\leqslant 60\ .$$

由于"$0\leqslant\eta\leqslant 60$"为必然率件，所以

$$P(0\leqslant\eta\leqslant 60)=\lambda\cdot\ 60=1\ ,$$

即 $\lambda=\frac{1}{60}$．我们所关心的事件恰好是打开收音机的时刻位于 $[50,60]$ 时间区间内(即 $50\leqslant\eta\leqslant 60$)．因此

$$P(50\leqslant\eta\leqslant 60)=\frac{1}{60}\ (60-50)=\frac{1}{6}\ .$$

即某人打开收音机等待报时"等待时间不超过 10 分钟"的概率为 1/6.

在这个例子中，样本空间 $\Omega=\{x: 0\leqslant x\leqslant 60\}$是一个不可数无穷集．我们关心的率件$\{50\leqslant\eta\leqslant 60\}=\{x: 50\leqslant x\leqslant 60\}$也是一个不可数无穷集.对于 Ω 的任一可求长(也称为可测)的子集 B，我们均可定义

$$P(B)=\frac{\mu(B)}{\mu(\Omega)}\ .$$

其中 $\mu(\Omega)$ 表示Ω 的长度(也称测度)，$\mu(B)$表示集合 B 的长度(当 B 是几个区间的不交并的情形，$\mu(B)$就等于这几个区间长度的

和). 由于不可测集的存在，我们不可能对 Ω 的任意子集合理地定义概率，只能对 Ω 的一部分子集(可测集)定义概率. 关于不可测集存在的论证超出了本书的范围，读者可参看[12，第二章 § 5]*). 下面给出一般定义.

定义 1.3.2 若某随机试验具有以下特点:

(1) 进行一次试验相当于向一个几何体(有限线段，平面区域，曲面，空间区域或它们的并)G 中取一点 g，G 有非零的有限测度(长度、面积或体积). 在建立了坐标系后点 g 的坐标仍用 g 表示. 若用 ξ 表示此随机试验中将要取得的点的坐标，则"取得点 g"可用"$\xi = g$"来表示;

(2) 对 G 内任何可测的子几何体 G_1，事件"点取自 G_1"(即 $\xi \in G_1$)的概率与 G_1 的测度成正比，与 G_1 的形状及 G_1 在 G 中的位置无关.

称此随机试验为**几何概型**.在几何概型中，事件"点取自 G_1"($G_1 \subset G$ 为可测几何体)的概率定义如下:

$$P(\xi \in G_1) = \frac{\mu(G_1)}{\mu(G)},$$

其中 $\mu(A)$ 表示几何体 A 的测度. 如此定义的概率称为**几何概率**.

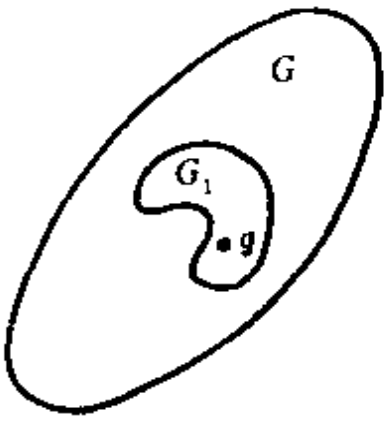


图 1-3

下面我们举一些例子说明在实际问题中是如何把随机试验归结为向一个几何体取点的问题.

例 1.3.3 (会面问题) 甲、乙两人约定在 0 到 T 时间区间内

*) 本书所谓的测度是几何体(线段、平面区域、空间区域或它们的并)的长度、面积或体积. 可以证明：几何体如果有测度则一定有勒贝格测度，而且它们相等. 因此，如果一个几何体没有勒贝格测度，则一定不可测. 在[12，第二章 § 5] 中证明了在直线中存在没有勒贝格测度的子集.

在某地会面，先到者等 $t(t\leqslant T)$ 时后离去，试求两人能会面的概率.

解 以 ξ,η 分别表示甲、乙两人到达的时刻，每次试验 $\zeta=(\xi,\eta)$ 相当于在边长为 T 的正方形区域

$$\Omega=\{(x,y):0\leqslant x,y\leqslant T\}$$

中取一点. 由于甲、乙两人在 $[0,T]$ 时间区间中任何时刻到达是等可能的，而且两人互不影响，所以 (ξ,η) 在 Ω 的任何可求积子区域内取点的可能性只与区域的面积大小成正比而与其形状、位置无关. 于是我们把会面问题化成了向平面区域 Ω 随机取点的问题，而所关心的事件"两人能会面"可表示为

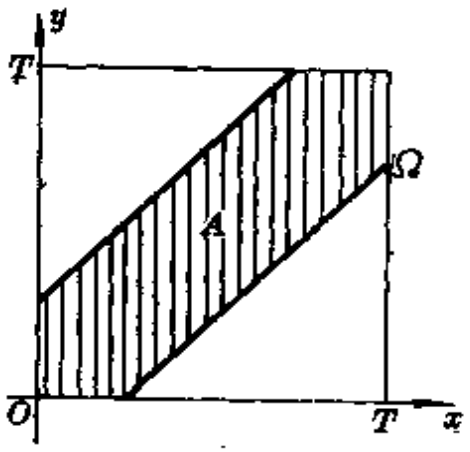


图 1-4

$$\{(\xi,\eta)\in A\},$$

其中 $A=\{(x,y):\ |x-y|\leqslant t\ ,0\leqslant x\ ,y\leqslant T\}$,

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{A\text{的面积}}{\Omega\text{的面积}}=\frac{T^2-(T-t)^2}{T^2}$$

$$=1-\left(1-\frac{t}{T}\right)^2.$$

现在我们来计算本节开始提到的蒲丰投针问题的概率.

例 1.3.4 (投针问题)平面上有一族平行线，其中任何相邻的两线距离都是 $a(a>0)$. 向平面任意投一长为 $l(l<a)$ 的针，试求针与一条平行线相交的概率.

解 设 x 是针的中点 M 到最近的平行线的距离，φ 是针与此平行线的交角. 于是投针问题就相当于向平面区域

$$\Omega=\{(\varphi\ ,\ x)\colon\ 0\leqslant\varphi\leqslant\pi\ ,0\leqslant x\leqslant\frac{a}{2}\ \}$$

取点的几何概型.

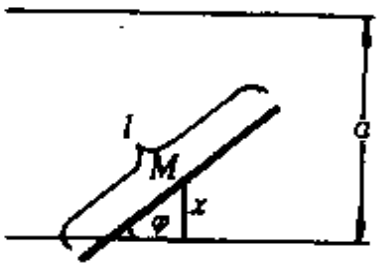


图 1-5

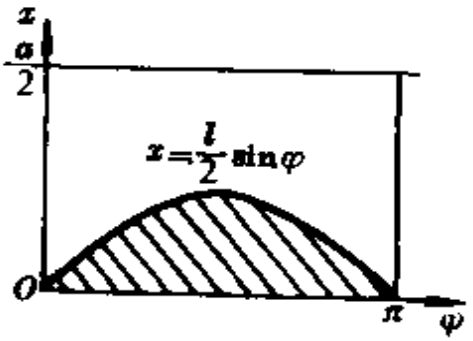


图 1-6

针与平行线之一相交的充分必要条件是$(\varphi\ ,x)$满足条件

$$0\leqslant x\leqslant\frac{l}{2}\sin\varphi\ ,0\leqslant\varphi\leqslant\pi\ \ .$$

即随机点取自平面区域

$$A=\{(\varphi\ ,x)\colon\ 0\leqslant x\leqslant\frac{l}{2}\sin\varphi,0\leqslant\varphi\leqslant\pi\ \ \}$$

中. 于是所求事件的概率

$$p=\frac{\mu(A)}{\mu(\Omega)}=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{\int_0^{\pi}\frac{l}{2}\sin\varphi d\varphi}{\frac{a}{2}\pi}$$

$$=\frac{2l}{\pi a}\ .$$

假如有另外的方法可以求出p(或其近似值),我们就可以利用蒲丰投针试验求得π值(或其近似值).下节还要提到这个例子.

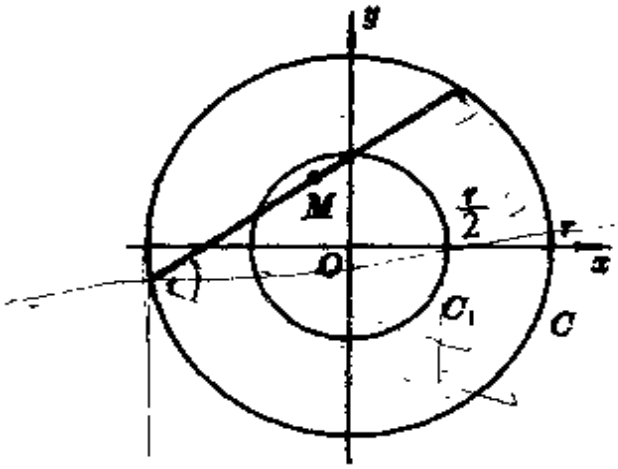


图 1－7

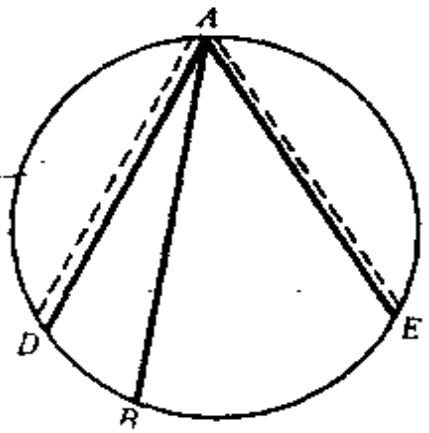


图 1－8

例 1.3.5 （贝特朗(Bertrand)悖论） 在一半径为 r 的圆 C 内"任意"作一弦，试求此弦长度 l 大于圆内接等边三角形的边长 $\sqrt{3}\,r$ 的概率 p.

解 1 作半径 $\frac{r}{2}$ 的同心圆 C_1. 弦的中点 M 是"任意"地取自圆 C 内的一点，则当且仅当 M 取自圆 C_1 内时弦长 $l>\sqrt{3}r$ (见图 1－7)，于是所求概率

$$p=\frac{\pi\left(\frac{r}{2}\right)^2}{\pi r^2}=\frac{1}{4}.$$

解2 如果我们固定弦的一端 A，让另一端 B"任意"地在圆周上变动. 如图设 $\triangle ADE$ 为等边三角形，则当且仅当 B 点取自弧 $\overset{\frown}{DE}$ 上时 $l>\sqrt{3}\,r$ (见图 1－8). 故所求概率

$$p=\frac{\overset{\frown}{DE}\text{的弧长}}{\text{圆周长}}=\frac{1}{3}.$$

解 3 如果固定弦的方向使其与直径 EF 垂直，弦的中点 M

在 EF 上“任意”地运动，当且仅当 M 在距圆心 $\frac{r}{2}$ 的 G、H 两点之间的线段上时，弦 AB 之长 $l>\sqrt{3}r$（见图 1－9）.于是所求概率

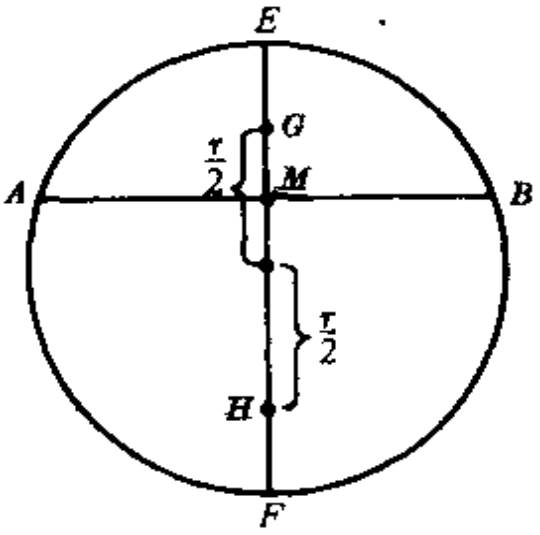


图 1－9

$$p=\frac{GH\text{之长}}{EF\text{之长}}=\frac{2\times\frac{r}{2}}{2r}=\frac{1}{2}.$$

三种解法得到了三种不同的答案.原因在于题中所示：在圆内“任意”作一弦，其中“任意”的意义不明确，三种解法是对“任意”的三种不同理解，从而在三种不同的随机试验中解题.解 1 是理解为“在圆 C 内随机地取点”，求事件“随机点取自圆 C_1 中”的概率，这时对“任意”理解为在圆内随机选择弦的中点 M，而不顾及弦的方向；解 2 是理解为在圆上随机取点，所求的事件是“随机点取自弧 $\overset{\frown}{DE}$ 上”.此时“任意”作一弦，理解为一个端点固定，随机地选择另一个端点；解 3 的随机试验是在线段 EF 上随机取点，我们所求的事件是“随机点取自线段 GH”上.此时“任意”作一弦理解为方向固定任意选中点.因此，三种解法对应于在三种不同的随机试验中求三个不同事件的概率.

这个问题之所以有不同的解答，从根本上说是由于当一个随机试验有无穷多个可能结果时，有时很难客观地规定“等可能”这

一概念，特别是将一个实际问题归结为向一个几何区域取点时本身就等于规定了一种“等可能性”，这种规定是否合于实际，只能用实践来检验.由此可见，和古典概率一样，几何概率的正确运用，有赖于“等可能性”的正确规定.

例 1.3.6 在线段 AD 上任意取两个点 B、C，在 B、C 处折断此线段 AD 而得三折线，试求此三折线能构成三角形的概率.

解 首先弄清楚“任意取两个点”的含义.它是指 B 可以等可能地取自 AD 线段上，C 也等可能地取自 AD 线段上，B 可以在 AC 之间，C 也可以在 AB 之间，如图 1－10 的 (a)，(b).

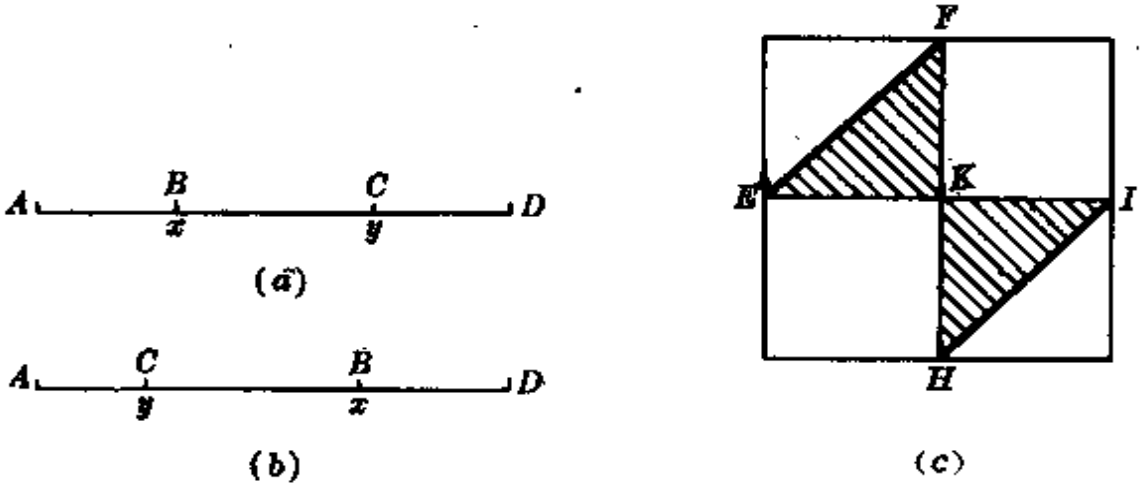


图 1－10

为计算我们感兴趣的事件的概率，不妨假设 AD 的长为 1，于是在线段 AD 上投两个点 B，C，就相当于在正方形区域

$$\Omega=\{(x,y):\quad 0\leqslant x,y\leqslant 1\}$$

内取点，其中 x 表示 AB 之长，y 表示 AC 之长，若 $x<y$，则 AB、BC、CD 能够成三角形的条件为：

$$\overline{AB}+\overline{BC}>\overline{CD}\text{，即}\quad y>1-y\text{，即}\quad y>\frac{1}{2}\text{；}$$

$\overline{BC}+\overline{CD}>\overline{AB}$，即 $1-x>x$，即 $x<\frac{1}{2}$；

$\overline{CD}+\overline{AB}>\overline{BC}$，即 $1-y+x>y-x$，即 $y-x<\frac{1}{2}$.

若 $y>x$，则 AC, CB, BD 能够成三角形的条件为：

$\overline{AC}+\overline{CB}>\overline{BD}$，即 $x>1-x$，即 $x>\frac{1}{2}$；

$\overline{CB}+\overline{BD}>\overline{AC}$，即 $1-y>y$，即 $y<\frac{1}{2}$；

$\overline{BD}+\overline{AC}>\overline{CB}$，即 $1-x+y>x-y$，即 $x-y<\frac{1}{2}$.

因此，在 B、C 处折断线段 AD，三线段能够成三角形的事件相当于随机点取自区域 G 中，其中

$$G=\{(x,y): x<\frac{1}{2}, y>\frac{1}{2}\quad y-x<\frac{1}{2}\}$$

$$\cup\{(x,y): x>\frac{1}{2}, y<\frac{1}{2}\quad x-y<\frac{1}{2}\},$$

它由正方形区域 Ω 中的 $\triangle EKF$ 和 $\triangle HKI$ 组成（图 $1-10(c)$），这两个三角形面积之和为 $\frac{1}{4}$，而区域 Ω 的面积为 1. 因此所求概率为 $\frac{1}{4}$.

本书今后用 R 表示全体实数组成的集，R^n 表示 n 维实空间（n 为正整数），即 $R^1=R$，R^n 由全体 n 维实向量 $(x_1, \cdots, x_n)$，$x_1, \cdots, x_n \in R$ 组成.

例 1.3.7 在 R^3 中随机地选择方向，试求随机方向上单位向量在 x 轴上投影的长度 ξ 不大于 $\frac{1}{2}$ 的概率.

解 在 R^3 中随机地选择方向，相当于在单位球面上随机地取点，随机方向上的单位向量即以球心为始点、球面上随机点为终点的向量，它在 x 轴上的射影的长度即随机点的 x 坐标的绝对值，于是我们把问题归结为在球面

$$\Omega=\{(x,y,z):\ x^2+y^2+z^2=1\}$$

上取点的问题，所关心的事件$\{\xi\leqslant\frac{1}{2}\}$相当于随机点取自

$$A=\{(x,y,z):\ x^2+y^2+z^2=1,\ |x|\leqslant\frac{1}{2}\}$$

内，它是单位球面上横坐标绝对值小于或等于 $\frac{1}{2}$ 的球带.

已知单位球面的面积

$$\mu(\Omega)=4\pi,$$

面球带的面积正比于球带的高

$$\frac{\mu(A)}{2\times\frac{1}{2}}=\frac{4\pi}{2}.$$

故所求概率

$$P(\xi\leqslant\frac{1}{2})=\frac{\mu(A)}{\mu(\Omega)}=\frac{1}{2}.$$

习题 1－4

1. 将一根棒任意折成两段，求其中一段的长度大于另一段长度4倍的概率.

2. 从(0,1)中随机地取两个数，求下列事件的概率：

(1) 两个数中较小的小于 $\frac{1}{2}$；

(2) 两个数中较大的小于 $\frac{1}{2}$；

(3) 两个数之和小于 $\frac{3}{2}$；

(4) 两个数之积小于 $\frac{1}{4}$.

3. 在线段 (O,a) 上任意取三个点 x，y，z，求

(1) y 位于 x，z 之间的概率；

(2) 线段 Ox，Oy，Oz 能够成三角形的概率.

4. 平面上画了一族平行线，其中相邻两线的距离都是 $2a$，将一枚半径 $r<a$ 的硬币任意投掷在平面上，求硬币不与任一平行线相交的概率.

5. 甲、乙两船同日欲靠同一码头，设两船各自独立地到达，并且每艘船到达在一昼夜间是等可能的，若甲船需停泊一小时，乙船需停泊 2 小时，而该码头只能停泊一艘船，试求其中一艘船要等待码头空出的概率.

§1.4 频率的稳定性

不是所有的随机试验其结果都具有象 §1.1 和 §1.3 所示的等可能性，因而不能用计算古典概率或几何概率的方法求出事件的概率——事件发生的可能性大小. 在一般情形，如何求出或近似求出事件的概率呢?

在 §1.1 和 §1.3 中，我们介绍了两种类型的随机试验，在这些试验中人们做了各种各样等可能性的假设，这些假设是否合于实际，都有待于在实践中检验.

我们还是从最简单的随机试验——抛掷硬币谈起. 假如硬币是均匀的，由古典概率的计算法则，我们得到“掷得正面”的概率为 $\frac{1}{2}$. 历史上曾经有许多人做过抛掷硬币的试验以验证这一结果的正确性.

若在相同条件下试验 n 次，其中事件 A 出现了 n_A 次，我们

称比值

$$f_n(A)=\frac{n_A}{n} \tag{1.4.1}$$

为 n 次试验中事件 A 出现的频率.

直观上，事件 A 的频率应该能反映事件 A 发生的可能性大小. 表 1－2 是历史文献中记录的硬币投掷试验的结果，这三个频率都在 0.5 的附近.

表 1－2*)

实 验 者	投掷次数	正面向上次数	频 率
蒲丰(Buffon)	4040	2048	0.5069
皮尔逊(K. Pearson)	12000	6019	0.5016
皮尔逊(K. Pearson)	24000	12012	0.5005

频率应该反映事件发生的可能性大小，是因为如果事件发生的可能性大，出现的机会就多，频率就应该大些. 但由于试验结果的随机性，频率不能完全反映概率. 例如甲抛掷 15 次硬币，正面出现 5 次，正面出现的频率为 $\frac{1}{3}$；乙抛掷 20 次，正面出现 12 次，正面出现的频率为 $\frac{3}{5}$. 频率既随着试验次数的不同而不同，而且这 n 次试验和另外 n 次试验的频率也可能不同.

现在设想有一千个人投掷硬币，如果每人投 5次，计算每个人投出正面的频率，在这一千个频率中，一般说，0，$\frac{1}{5}$，$\frac{2}{5}$，$\frac{3}{5}$，$\frac{4}{5}$，1 都会有，而且还会有不少是 0 或 1；如果要求每个人投 20 次，这时频率为 0，0.05，0.95，1 的将会变少，多数频率在 0.35 到 0.65 之间，甚至比较集中在0.4 到 0.6 之间；如果要求每个人投掷1000 次，这时绝大多数的频率会集中在 0.5 的附近. 和 0.5 有较大距离的频率值也会有，但这样的频率值很少. 而且随着投掷次数的增多，频率越来越明显地集中在 0.5 附近. 当然，即使

*) 引自格涅坚科《概率论教程》.

投掷次数再多，也不能绝对排除出现与0.5差距较大的频率值，只不过这种情形极少.

下面，我们再观察“由0，1，2，…，9中任意取出一个数字”的随机试验中，事件“取出的数字是1”的规律.表1－3列出了10个2000次的观察结果，这10个频率虽然不尽相同，但都在0.10附近摆动.

表 1－3*)

观察次数	2000	2000	2000	2000	2000
“1”出现次数	194	203	218	185	212
频　　率	0.0970	0.1015	0.1090	0.0925	0.1060
观察次数	2000	2000	2000	2000	2000
“1”出现次数	205	183	204	204	205
频　　率	0.1025	0.0915	0.1020	0.1020	0.1025

像上面所看到的以及许多其它经验的积累，人们逐渐认识到：在多次重复试验中，同一事件发生的频率在一个数值附近摆动，而且随着试验次数的增加，一般摆动得越小，观察到的大偏差越稀少，呈现出一定的稳定性.**频率的稳定性**揭示出随机事件发生的可能性有一定的大小.事件的频率稳定于某一数值，我们就用这一数值表征事件发生的可能性的客观大小，称之为事件的概率，概率的这种定义即**概率的统计定义**.在实践中很多时候用这种方法求事件概率的估计值.

在人口统计中，有一个著名的例子也说明概率存在的客观性.就是男婴女婴出生的比率问题.一般简单地认为生男生女各半，事实上并非如此.1795年拉普拉斯在《概率的哲学探讨》中叙述到：他对伦敦、彼得堡、柏林、全法兰西的众多统计资料的研究给出了几乎完全一致的男婴出生数与全体出生数的比值.所有这些比值在十年间总摆动于同一数的左右，这数大约等于22/43.同时

*)取自严士健、王隽骧、徐承彝著《概率论与数理统计基础》，表是利用电子计算机进行模拟试验所得的结果.

根据巴黎四十年间(1745 至 1784)类似的统计资料的研究，得出了另一数值 25/49.拉普拉斯对如此显著的差异感觉兴趣，并着手寻求其合理的解释.经对档案资料的详细研究，发现是由弃婴引起的. 除去这个因素后，看出巴黎这四十年间男婴的出生率也稳定地接近于 22/43(参看格涅坚科著《概率论教程》中译本 43 页).

对于由古典概率和几何值率能够计算的事件的概率,历史上也曾有很多人通过试稳来验证频率在概率的值附近摆动，除表 1－2 所列的掷硬币的试验数据外，掷骰子是一个有趣的例子，德梅尔就是根据长期的赌博经稳(即频率的比较)体会到一对骰子掷 24 次至少出现一对 6点的可能性要比一颗骰子掷 4 次至少出现一次 6 点的可能性小.

§1.3 中我们计算出在蒲丰投针试验中针与平行线相交的概率

$$p=\frac{2l}{\pi a}.$$

这个试验的设计和公式是蒲丰 1777 年提出的. 如果利用频率的稳定性，当试验次数 n 很大时，数出针与平行线相交的次数 m，则

$$p\approx\frac{m}{n}.$$

于是人们又可以利用

$$\frac{m}{n}\approx\frac{2l}{\pi a}$$

求出 π 的近似值，即

$$\pi\approx\frac{2ln}{am}. \tag{1.4.2}$$

历史上有不少人曾亲自数过这个试验，其结果列在表 1－4 中(表

中资料把 a 折算为 1).

表 1-4*)

实验者	年份	针长	投掷次数	相交次数	π 的实验值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan.C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

这些实验不仅说明我们的模型假设(即把问题化作向平面区域$\{(\varphi, x): 0 \leqslant \varphi \leqslant \pi, 0 \leqslant x \leqslant \frac{a}{2}\}$随机取点的问题)是正确的,而且利用(1.4.2)式求 π 的近似值的方法也是极有启发性的.

上面的方法概括起来是:建立一个概率模型,使它与我们感兴趣的量(本例中是常数 π)有关,然后设计适当的随机试验,并通过这个试验的结果来确定这些量.随着电子计算机的发展,已按这种思路建立起一类新的计算方法,称为蒙特卡洛(Monte-Carlo)方法.作为例子,我们讨论如何用蒙特卡洛方法计算定积分

$$I=\int_0^1 f(x)dx.$$

不妨假定

$$0 \leqslant f(x) \leqslant 1, x \in [0,1].$$

设想在正方形区域

$$\Omega=\{(x,y): 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$$

*)转引自复旦大学编概率论,第一册,第 39 页,原资料见 Gridgeman, N.T. Geometric Probability and the Number π. Scripta Mathematica, 25(1960), 183—195.

内随机取一点 M，则由几何概率的定义知，点 M 取自曲线 $y=f(x)$ 下方区域的概率

$$p=\int_0^1 f(x)dx.$$

现在由电子计算机产生 n 个在单位正方形中的随机点 (x_i, y_i)，$i=1,2,\cdots,n$，然后逐个检查这些点是否在曲线 $y=f(x)$ 的下方区域内，即检验是否 $y_i \leqslant f(x_i)$. 若共有 m 个点满足 $y_i \leqslant f(x_i)$，

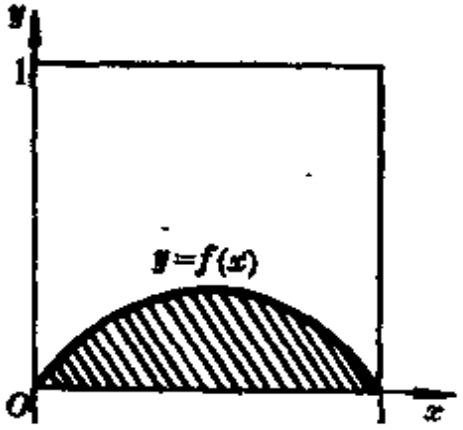


图 1-11

则 n 个随机点取自曲线 $y=f(x)$ 下方的频率为 $\frac{m}{n}$. 当 n 很大时，$\frac{m}{n}$ 一般和 p 很接近，因此我们就用 $\frac{m}{n}$ 作为积分的近似值，即

$$\int_0^1 f(x)dx \approx \frac{m}{n}.$$

蒙特卡洛方法现在已经有了广泛的应用，特别是在高维空间上的积分近似计算方面及模拟一些物理、化学过程的计算中有很重要的作用.

习题 1-5

1. 用事件出现的频率与事件的概率之间的关系说明：

(1) 不可能事件的概率是零；

(2) 必然事件的概率是 1；

(3) 任何事件 A 的概率满足 $0 \leqslant P(A) \leqslant 1$；

(4) 概率具有有限可加性，即若 $A_1, A_2, \cdots, A_n$ 是 n 个两两不会同时发生的事件，A 表示这些事件中至少发生其一的事件，则

$$P(A)=P(A_1)+P(A_2)+\cdots+P(A_n).$$

2. 对靶子射击一次，以 ξ 表示射中的环数，

A 表示事件："$\xi=10$"；

B 表示事件："$\xi=10$ 或 $\xi=9$"；

C 表示事件："$\xi<9$"，

试用事件出现的频率与概率的关系说明

(1) $P(A)\leqslant P(B)$；

(2) $P(B)+P(C)=1$.

3. 利用频率与概率的关系，按你的想法构造一个适当的概率模型，给出一个求 ln2 的近似值的方法.

§1.5 概率的公理化定义

到本世纪初，概率论的各个领域已经得到了大量的成果，而且人们对概率论在其它学科如物理学和统计学的应用也产生了越来越大的兴趣. 但是直到那时，关于概率论的一些基本概念. 例如事件及其概率等都没有严格的、明确的定义，这是一个很大的矛盾，以至于产生人们无法解释的贝特朗悖论(实质上几种不同的理解相应于对等可能性几种不同的规定)，致使许多人对概率的客观含义甚至概率论结论的正确性都产生了怀疑. 这一状况严重阻碍了概率论的进一步发展和应用，又落后于当时数学的其他分支的公理化潮流. 1900 年希尔伯特在世界数学家大会上公开提出了建立概率论公理化体系的问题.

二十世纪初完成的勒贝格测度和积分理论以及后来发展起来的抽象测度和积分理论. 为概率论公理化体系的确立奠定了理论基础. 人们通过对概率论的两个最基本的概念即事件与概率的长期研究，发现事件的运算与集合的运算完全类似，概率与测度有相同的性质. 到了 30 年代，随着大数定律研究的深入，概率论与测度论的联系愈来愈明显. 在这种背景下，柯尔莫哥洛夫于 1933 年在他的《概率论基础》*)一书中，第一次给出了概率的测度论

*) 这是柯尔莫哥洛夫著作 1956 年英文版的名字. 原书是 1933年用德文出版的，1936 年出版了俄文版. 德文、俄文版的书名都是《概率论的基本概念》.

式的定义和一套严密的公理化体系.这一公理体系着眼于规定事件及事件概率的最基本的性质和关系，并用这些规定来表明概率的运算法则．它们是从客观实际中抽象出来的，既概括了概率的古典定义、几何定义及统计定义的基本特性，又避免了各自的局限性和含混之处．因此这一公理体系一经提出，便迅速获得举世公认．柯尔莫哥洛夫公理化的出现，是概率论发展史上的一个重要里程碑．它为近代概率论的蓬勃发展打下了坚实的基础.

本节将简要叙述这一公理化体系，为此还要叙述一些基本的概念.

1. 样本空间和样本点

在§1.1和§1.3中，对古典概型和几何概型定义了样本空间和样本点，事实上对任何试验来讲，都可以有

定义 1.5.1 随机试验的每一可能出现的结果称为样本点，一切可能出现的结果的集合称为样本空间.

这样就将样本空间抽象地看成一个集合 Ω，样本点 ω 看作 Ω 的元素.

2. 事件、事件的关系与运算

在研究古典概率时，将事件看作样本空间的任一子集；在研究几何概率时，将事件看作样本空间中具有有限测度的(即可测的)子集．总而言之，将事件 A 看作样本空间的一个子集，**事件 A 发生**，当且仅当 A 所包含的样本点中有一个发生.

在前几节中似乎是孤立地研究一个个事件的概率，事实上，对于一个随机试验而言，有许多相关联的事件，必须在它们的相互联系中去认识它们，把握它们.在前面的讨论中也并不能完全避免这些相互联系．例如在例1.2.3中，要考虑"次品数 $\xi>1$"这一事件的样本点数，它是三个事件即"$\xi=2$"、"$\xi=3$"和"$\xi=4$"所含样本点发的和，而且"$\xi>1$"的概率也可以表示成这三个事件的概率之和；在例1.1.8中，要计算一颗骰子掷4次至少有一次掷样6

点的样本点个数，首先计算的是"没有一次掷得 6 点"的样本点个数；又如在例 1.3.6 中，为计算在线段 AD 上任意取两个点 B、C，在 B、C 处折断此线段 AD，三折线能够构成三角形的概率，可以把这一事件分割成两个事件即 B 在 A、C 之间及 C 在 A、B 之间来计算，叙述和计算都会更简单一些．以后我们还将看到，如果不利用事件的运算与关系，有些事件所含的样本点数很难计算(在古典概率的情形)．下面给出它们的定义及其与集合的运算和关系之间的联系．

我们用 Ω 表示样本空间，用 $A, B, \cdots$ 表示事件，它们是样本空间的子集，Ω 称为**必然事件**，因为每次试验必有 Ω 中所含的样本点之一发生，因而每次试验 Ω 必然发生．空集 ϕ *)称为**不可能事件**，它不含任何样本点，因而每次试验它一定不发生．

(1) 若事件 A 中的样本点都是事件 B 中的样本点，则称事件 B **包含**事件 A，这时事件 A 发生必导致事件 B 发生．记作 $A \subset B$ 或 $B \supset A$；

(2) 若 $A \supset B$ 且 $B \supset A$，则称 A 与 B 等价，或称 A 与 B **相等**，记作 $A = B$．等价的两个事件同时发生或同时不发生．

(3)对于事件 A，所有不包含在 A 中的样本点所组成的集称为 A 的**余事件**，或称为 A 的**对立事件**，记作 A^c(或 $\overline{A}$)．显然 A 发生则 A^c 必不发生，反之亦然．

(4) 集合 A、B 的交所表示的事件称为事件 A 与 B 的**交**(或**积**)，仍记作 $A \cap B$(或 AB)，它的意义是事件 $A \cap B$ 发生当且仅当事件 A 与事件 B 同时发生，因为 $A \cap B$ 所含的样本点是同属于 A，B 的样本点，$A \cap B$ 发生则它所含的样本点之一发生，而此样本点又同时属于 A 及 B，因此 $A \cap B$ 发生则 A 与 B 同时发生，反之亦然．

(5) 集合 A、B 的并所表示的事件称为事件 A 与 B 的**并**(或

*) 此处及以后，空集永远用 ϕ 表示．

和)，记作 $A \bigcup B$，它的意义是 $A \bigcup B$ 发生当且仅当 A,B 之中至少有一发生.

(6) 集合 A 与 B 的差所表示的事件称为事件 A 与 B 的差，记作 $A \setminus B$，它的意义是 $A \setminus B$ 发生当且仅当 A 发生而 B 不发生. 显然我们有

$$A \setminus B = AB^c,\ A^c = \Omega \setminus A,\ A \bigcup A^c = \Omega,\ A \bigcap A^c = \phi\ .$$

(7) 若事件 A, B 没有公共的样本点，即 $A \bigcap B = \phi$，则称事件 A, B **互不相容**.

我们举例来说明这些关系和运算.

例 1.5.2 对于一个电话系统来说，收到用户的呼叫是具有偶然性的. 人们研究在 $(a\ , a+t\ ]$ 时间间隔内电话局收到的用户呼叫次数问题. 用 ω_k 表示在 $(a\ , a+t]$ 内电话局收到用户 k 次呼叫，则样本空间

$$\Omega = \{\omega_k : k = 0, 1, 2, \cdots\}.$$

若事件 A 表示"收到的呼叫次数不多于 100"；事件 B 表示"收到的呼叫次数不多于 50"；事件 C 表示"收到的呼叫次数多于 50"；事件 D 表示"收到的呼叫次数在 30 到 200 次之间"；事件 E 表示"收到的呼叫次数少于 10"，则

$$A = \{\omega_k : 0 \leqslant k \leqslant 100\};$$

$$B = \{\omega_k : 0 \leqslant k \leqslant 50\};$$

$$C = \{\omega_k : k > 50\};$$

$$D = \{\omega_k : 30 \leqslant k \leqslant 200\ \};$$

$$E = \{\omega_k : 0 \leqslant k < 10\ \}.$$

显然 B 发生时 A 一定发生，即 $B \subset A$，且 $B^c = \{\omega_k : k > 50\}$，即 $B^c = C$.

$A \bigcup D=\{\omega_k : 0 \leqslant k \leqslant 200\}$，表示事件“收到的呼叫次数不多于200”；

$A \bigcap D=\{\omega_k : 30 \leqslant k \leqslant 100\}$，表示“收到的呼叫次数在30次到100次之间”；

$A \setminus D=\{\omega_k : 0 \leqslant k < 30\}$，表示“收到的呼叫次数少于30”；

$A \bigcup C=\Omega=\{\omega_k : k=0, 1, 2, \cdots\}$，表示必然事件；

$C \setminus D=\{\omega_k : k>200\}$，表示收到的呼叫次数大于200；

$D \bigcap E=\varnothing$，因而事件D，E互不相容，同理事件C，E也互不相容.

例1.5.3 如图1－12所示，Ω表示正方形区域，A，B分别表示Ω中的两个小区域，从正方形区域Ω内取点，事件A，B分别表示“点取自A”，“点取自B”，于是事件$A \bigcup B$，$A \bigcap B$，$A \setminus B$，A^c分别表示点取自图1－12中(1)，(2)，(3)，(4)的阴影部分.

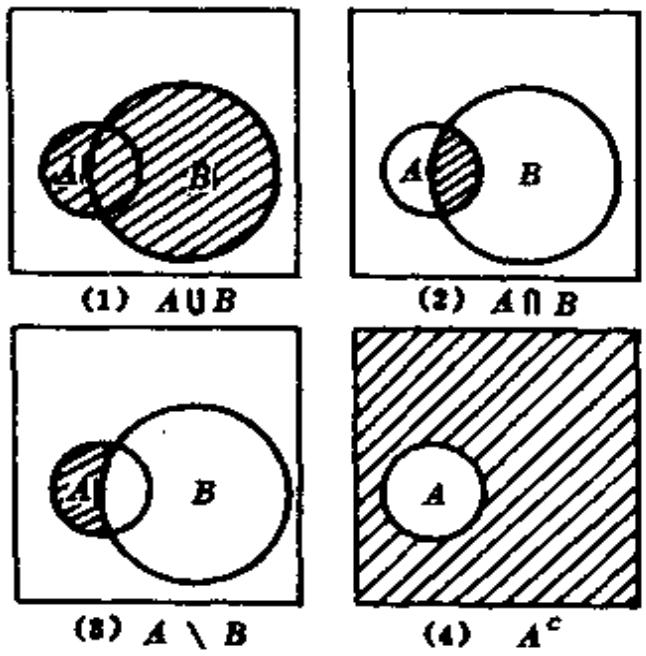


图 1－12

事件的运算完全等同于集合的运算，事件之间的关系也等同于集合之间的关系，除两个事件之间的交和并的运算外，同样可以定义任意有限个或可数个事件的交和并：

$\bigcup_{i=1}^{n} A_i$ 发生当且仅当 $A_1, A_2, \cdots, A_n$ 中至少有一发生;

$\bigcap_{i=1}^{n} A_i$ 发生当且仅当 $A_1, A_2, \cdots, A_n$ 同时发生;

$\bigcup_{i=1}^{\infty} A_i$ 发生当且仅当 $A_1, A_2, \cdots$ 中至少有一发生;

$\bigcap_{i=1}^{\infty} A_i$ 发生当且仅当 $A_1, A_2, \cdots$ 同时发生.

集合论中的许多关系式同样适用于事件的运算，特别是联系并、交、余三种运算的德摩根(De Morgan)定律也同样成立，即

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c \quad , \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c .$$

读者不仅要熟悉集合论中的运算和关系，还应学会熟练地把事件的语言和集合的语言相互转换. 下面再举几例.

例 1.5.4 对某目标进行三次射击，用 A_i 表示事件: "第 i 次射击时射中目标", $i=1,2,3$. 试用 A_1, A_2, A_3 表示事件:

(1) "恰好有一次射中目标"(记为 B);

(2) "最多有一次射中目标"(记为 C).

解 事件 B: "恰好有一次射中目标"就是事件"恰好第一次射中目标或恰好第二次射中目标或恰好第三次射中目标"，而"恰好第一次射中目标"就是"第一次射中目标，而第二、三次没有射中目标"，是 A_1, A_2^c, A_3^c 的积，等等，所以

$$B = (A_1 A_2^c A_3^c) \bigcup (A_1^c A_2 A_3^c) \bigcup (A_1^c A_2^c A_3).$$

事件 C : "最多有一次射中目标"就是"三次都没有射中目标或恰好有一次射中目标"，所以

$$C = (A_1^c A_2^c A_3^c) \bigcup B.$$

事件 C 还可以表示为“至少有两次射中目标”的对立事件，即

$$C=(A_1A_2\cup A_2A_3\cup A_3A_1)^c,$$

或“至少有两次没射中目标”即

$$C=(A_1^cA_2^c)\cup(A_2^cA_3^c)\cup(A_3^cA_1^c).$$

习题 1－6

1. 将红、白两个球任意放入Ⅰ、Ⅱ、Ⅲ三个盒中(每个盒可容纳两个球)；A_1，A_2，A_3 分别表示事件“红球在Ⅰ盒中”，“红球在Ⅱ盒中”，“红球在Ⅲ盒中”；用 B_1，B_2，B_3 分别表示事件“白球在Ⅰ盒中”，“白球在Ⅱ盒中”，“白球在Ⅲ盒”中.

用语言叙述下列事件：

(1) $A_1\cup B_1$；(2) A_2B_2；(3) $A_3\setminus B_3$；(4) B_2^c；

(5) $A_1\cap B_2^c$.

2. 字母含意同第1题，试用 A_1，A_2，A_3，B_1，B_2，B_3 表示下列事件：

(1) “红球白球都在Ⅲ盒中”；

(2) “白球在Ⅱ盒中，但红球不在Ⅱ盒中”；

(3) “红球不在Ⅰ盒或Ⅱ盒中”；

(4) “红球不在Ⅰ盒中，白球也不在Ⅰ盒中”.

3. 字母含意同第1题，试判断下面的命题哪些是正确的，哪些是不正确的：

(1) A_1 与 A_2 互不相容；

(2) A_3 与 $A_1\cup A_2$ 互不相容；

(3) $A_2\setminus B_2\subset B_2$；

(4) $B_2\cap A_3\subset A_3$；

(5) $B_1\subset B_1\cup B_2$；

(6) $(A_1\cap B_1)^c\subset A_1^c\cap B_1^c$.

4. 有两个电路(图1－13，图1－14)，用 A_i 表示事件“接点开关 i 闭合”，$i=1,2,3,4,5,6$. 试用 A_i，$i=1,\cdots,6$ 表示事件

(1) B_1：“L_1R_1 是通路”；

(2) B_2：“L_2R_2 是通路”.

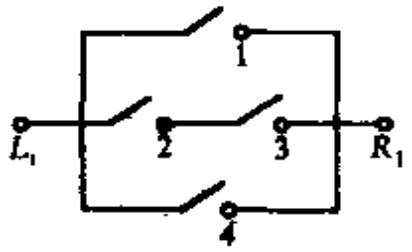


图 1－13

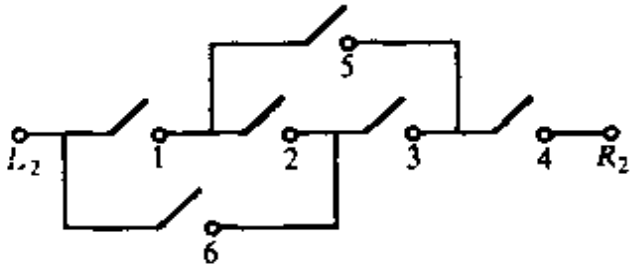


图 1－14

5. A, B, C 是三个事件，试用 A, B, C 表示

(1) A, B, C 中恰有两个出现；(2) A, B, C 中不多于两个出现.

3. 事件域和事件 σ－域

象在例 1.1.6, 1.1.8, .1.1.13, 1.2.3, 1.3.6 中的情况那样，在讨论一个随机现象时，与所要考虑的问题有关的常常不止一个事件，而是若干个事件，为了讨论这些事件，又涉及到它们的并、交、差及对立事件等等. 另一方面，如在讨论几何概率时看到的，又不能把样本空间的所有子集都作为事件包括在我们的讨论范围之内，那样将会给进一步的讨论带来困难. 通常的作法是：根据具体情况确定一组事件，记它们组成的集合为 $\mathscr{F}$，要求 $\mathscr{F}$ 包含与问题有关的事件，而且对事件的可数并、交、差、余运算都封闭. 这就是下面的定义.

定义 1.5.5 如果样本空间 Ω 的子集类 $\mathscr{F}$ 满足如下条件，则称 $\mathscr{F}$ 为 Ω 上的**事件 σ－域或事件域**：

(1) $\Omega \in \mathscr{F}$；

(2) 若 $A \in \mathscr{F}$，则 $A^c \in \mathscr{F}$；

(3) 若 $A_i \in \mathscr{F}$，$i=1, 2, \cdots$，则 $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

事件域的定义中只要求对求余及可数并运算封闭，事实上这时事件域对求交、差等运算也封闭.

命题 1.5.6 若 $\mathscr{F}$ 为 Ω 上的 σ－域，则

(1) $\phi \in \mathscr{F}$；

(2) 若 $A_1, A_2, \cdots, A_n \in \mathscr{F}$，则 $\bigcup_{i=1}^{n} A_i \in \mathscr{F}$；

(3) 若 $A_1, A_2, \cdots, \in \mathscr{F}$，则 $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$；且定义1.5.5中的(3)可由此性质代替；

(4) 若 $A_1, A_2, \cdots A_n \in \mathscr{F}$，则 $\bigcap_{i=1}^{n} A_i \in \mathscr{F}$；

(5) 若 $A, B \in \mathscr{F}$，则 $A \setminus B \in \mathscr{F}$.

证 (1) 由 $\Omega \in \mathscr{F}$，知 $\phi = \Omega^c \in \mathscr{F}$；

(2) 由 $\phi \in \mathscr{F}$，令 $A_{n+1} = A_{n+2} = \cdots = \phi$，则 $A_i \in \mathscr{F}$，$i = 1, 2, \cdots$. 故由定义中的(3)知 $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$；

(3) 由 $A_i \in \mathscr{F}$，知 $A_i^c \in \mathscr{F}$，$i = 1, 2, \cdots$. 故 $\bigcup_{i=1}^{\infty} A_i^c \in \mathscr{F}$，再由德摩根公式 $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathscr{F}$，故(3)成立.

(4) 令 $A_{n+1} = A_{n+2} = \cdots = \Omega \in \mathscr{F}$，故 $A_i \in \mathscr{F}$，$i = 1, 2, \cdots$.

由(3)知 $\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$.

(5) $A \setminus B = A \cap B^c$，故由 $A, B \in \mathscr{F}$ 知 $A, B^c \in \mathscr{F}$，再由(4)知 $A \cap B^c \in \mathscr{F}$.

定义 1.5.7 称事件域 $\mathscr{F}$ 中的元为事件，若 $A, B \in \mathscr{F}$，称 $A \cup B$ 为事件 A 与 B 的**并(或和)**；AB 为事件 A 与 B 的**交(或积)**；$A \setminus B$ 为事件 A 与 B 的**差**，A^c 为事件 A 的**余事件或对立事件**.

例 1.5.8 若 A 为 Ω 的非空真子集，则 $\mathscr{F} = \{\phi, A, A^c, \Omega\}$ 为事件域.

例 1.5.9 若 Ω 为可数集，即

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

则 Ω 的一切子集做成的类 $\mathscr{F}$ 为事件域.

例 1.5.10 若 Ω 为一实数区间(有限的或无限的)，则通常取 $\mathscr{F}$

为包含 Ω 的一切子区间的一个最小的 $\sigma-$ 域.这个 $\sigma-$ 域叫做 Ω 上的**波莱尔域**，其中的**莱合叫波莱尔莱**.

4. 概率的公理化定义

在前四节中对于各种不同类型的随机试验给出了随机事件概率的定义，用以表示事件发生的可能性大小.那里人们着眼于某一事件的概率.但是在一些例子中已经看到，在一个随机试验中会出现各种各样的随机事件，它们发生的可能性大小不是彼此孤立的，而是彼此联系、相互制约的.例如在例 1.2.3 中废品个数 $\xi>1$ 这一事件就可以分解为三个事件的并，即

$$\{\xi>1\}=\{\xi=2\}\bigcup\{\xi=3\}\bigcup\{\xi=4\},$$

而且等式右端的三个事件是彼此互不相容的.因此$\{\xi>1\}$发生的可能性大小应是右端三个事件发生的可能性的大小的和，也就是说，$\{\xi>1\}$的概率应是右端三个事件各自概率的和.

又如在几何概率中，若 Ω 是具有非零的有限测度的几何体，B_1，B_2 是 Ω 的两个不交的可测 的子几何体，那么 $B=B_1\bigcup B_2$也是 Ω 的可测的子几何体，由 B_1，B_2 不交易知 $\mu(B_1\bigcup B_2)=\mu(B_1)+\mu(B_2)$，因而我们定义的几何概率

$$P(B_1\bigcup B_2)=\frac{\mu(B_1\bigcup B_2)}{\mu(\Omega)}=\frac{\mu(B_1)+\mu(B_2)}{\mu(\Omega)}$$

$$=P(B_1)+P(B_2).$$

对于概率的统计定义同样具有上述性质.

为了摆脱古典概率、几何概率及概率的统计定义的局限性和含混之处，人们从这些定义中抽象出它们具有的最本质的共同特性，作为概率的最基本的公理，使它不仅适用于人们已经熟悉了的古典概率、几何概率等特殊情形而且适用于更广泛的其它情形.

定义 1.5.11 设$\mathscr{F}$ 为样本空间Ω 上的 $\sigma-$ 域，P 为定义在$\mathscr{F}$

上的集函数，如果它满足下述条件:

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

(2) $P(A)\geqslant 0$, 对一切 $A\in\mathscr{F}$;

(3) 若 $A_i\in\mathscr{F}$, $i=1,2,\cdots$ 且两两互不相容，则

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

则称 $P(\cdot)$为定义在$(\Omega,\mathscr{F})$上的**概率**，$P(A)$称为事件 A 的**概率**，(3)称为概率的**可数可加性**，$(\Omega,\mathscr{F},P)$称为**概率空间**(或**概率场**).

* 下面举例说明如何构造概率场.

例 1.5.12 某人有n 把钥匙，其中只有一把能打开他的门.他随机地用这些钥匙去试开，打不开仍放回去.试问某人恰巧在第r次打开他的门的概率为多大? 构造一个概率空间，描述这一随机现象.

解 我们把钥匙标上号码 $1,2,\cdots,n$, 若只有 1 号钥匙可以打开门.某人用钥匙试开,直到打开一次门算作一次试验.若用j_l表示第l次试开时用的钥匙号码. 令

$$\Omega=\{(j_1,j_2,\cdots,j_{k-1},1):\ j_l=2,3,\cdots,n;\ l=1,2,\cdots,k-1;$$

$$k=1,2,\cdots\}.$$

它具有可列个样本点，因而可令

$$\mathscr{F}=\{A:A\subset\Omega\}.$$

$\mathscr{F}$ 为Ω 上的 $\sigma-$ 域.

为定义$\mathscr{F}$ 上的集函数P.我们首先对样本点组成的单点集定义

$$P(\{j_1,j_2,\cdots,j_{k-1},1\})=n^{-k},$$

$$j_l=2,3,\cdots,n\ ;\ l=1,2,\cdots,k-1;k=1,2,\cdots.$$

对任意集$A\subset\Omega$，令$P(A)$为A中所包含的样本点所成的单点集的概率的和．即

$$P(A)=\sum_{\omega\in A}P(\{\omega\})$$

令A_r表示事件“恰巧在第r次打开门”，则

$$A_r=\{(j_1,j_2,\cdots,j_{r-1},1):\ j_l=2,3,\cdots,n,\ l=1,2,\cdots,r-1\},$$

$A_1,A_2,\cdots,A_r,\cdots$两两不交且$\Omega=\bigcup_{r=1}^{\infty}A_r$．按照定义

$$P(A_r)=\sum_{\omega\in A_r}P(\{\omega\})=\sum_{j_1=2}^{n}\sum_{j_2=2}^{n}\cdots\sum_{j_{r-1}=2}^{n}P(\{j_1,j_2,\cdots,j_{r-1},1\})$$

$$=(n-1)^{r-1}\cdot n^{-r}=\left(1-\frac{1}{n}\right)^{r-1}\cdot\frac{1}{n},$$

$$P(\Omega)=\sum_{\omega\in\Omega}P(\{\omega\})=\sum_{r=1}^{\infty}\sum_{\omega\in A_r}P(\{\omega\})$$

$$=\sum_{r=1}^{\infty}\left(1-\frac{1}{n}\right)^{r-1}\cdot\frac{1}{n}=1.$$

因此$(\Omega,\mathcal{F})$上的集函数$P(\cdot)$具有性质

(1) $P(\Omega)=1$；

(2) $P(A)\geqslant 0$；

(3) 若$B_1,B_2,\cdots,B_m,\cdots$两两不交，$B_m\in\mathcal{F}$，$m=1,2,\cdots$，则$P\left(\bigcup_{m=1}^{\infty}B_m\right)=\sum_{m=1}^{\infty}P(B_m)$．

因而P是$(\Omega,\mathcal{F})$上的概率，$(\Omega,\mathcal{F},P)$构成概率空间(注意P的可列各加性可直接由$P(A)$的定义得到).

在本书中，计算一些事件的概率时，通常只要解决问题需要，就对与问题有关的事件的概率直接应用定义1.5.11中的(1)、(2)、(3)及第5目中各命题所列性质，即并不先去构造出概率空间或证明它的存在，而是假定有适当的概率空间存在，在这种情况下，例1.5.12的解可以如下进行:

设B_i表示第i次打开门(即用第1号钥匙试开)，由题设B_i、$i=1,2,\cdots$独立(关于事件的独立参考§2.3)，

$$P(B_i)=\frac{1}{n},\ P(B_i^c)=1-\frac{1}{n},\ i=1,2,\cdots.$$

于是某人"恰巧在第r次首次打开门"

$$A_r=B_1^cB_2^c\cdots B_{r-1}^cB_r,\ r=1,2,\cdots.$$

由$B_1,B_2,\cdots,B_r$独立知

$$\begin{aligned}P(A_r)&=P(B_1^c)P(B_2^c)\cdots P(B_{r-1}^c)P(B_r)\\&=\left(1-\frac{1}{n}\right)^{r-1}\cdot\frac{1}{n},\ r=1,2,\cdots.\end{aligned}$$

这就得到了例1.5.12中所提问题的解答.

不证明概率空间存在的理由在于:比较初等的概率问题的概率空间总是存在的，而这种存在性的证明对于初学者是难于理解和非本质的(当然对于一些进一步的概率问题是重要的).因此，本书和一般初等概率统计的书籍一样，今后不再强调概率空间的存在.例1.5.12只是向读者展示一下，从数学的严格性讲，还需要有构造概率空间或证明它的存在这一步骤.

5.概率的性质

正如任何公理化方法一样，我们可以从概率满足定义1.5.11中的(1)、(2)、(3)推导演绎出概率所具有的一系列性质来.它们在

计算复杂事件的概率和讨论概率理论问题时都是非常有用的.

命题 1.5.13 不可能事件的概率为零，即

$$P(\phi)=0.$$

证 因为

$$\Omega=\Omega\cup\phi\cup\phi\cup\cdots,$$

由定义 1.5.11(3)知

$$P(\Omega)=P(\Omega)+P(\phi)+P(\phi)+\cdots.$$

由(1)知$P(\Omega)=1$，又由(2)，$P(\phi)\geqslant 0$，从而得知

$$P(\phi)=0.$$

命题 1.5.14 概率具有有限可加性，即若$A_1, A_2, \cdots, A_n$两两互不相容，则

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

证 因为

$$A_1\cup A_2\cup\cdots\cup A_n=A_1\cup A_2\cup\cdots\cup A_n\cup\phi\cup\phi\cup\cdots,$$

由可列可加性及命题 1.5.13 立得所欲证.

命题 1.5.15 对任何事件A有

(1.5.16) $$P(A^c)=1-P(A).$$

证 由于$A\cup A^c=\Omega$，$A\cap A^c=\phi$，故由(1)及命题 1.5.14 得

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

事实上我们已经多次用到了(1.5.16)，例如，在例 1.1.8 中，

我们为计算一颗骰子掷4次"至少出现一次6点"的样本点数，转而去计算"不出现6点"的样本点数，实质上是通过计算对立事件的概率得到所要求的概率.

命题 1.5.17 如果$A \supset B$，则

$$P(A \setminus B) = P(A) - P(B),$$

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

证 当$A \supset B$时，$A = B \cup (A \setminus B)$，且$B \cap (A \setminus B) = \phi$，由概率的有限可加性得

$P(A) = P(B) + P(A \setminus B)$，即 $P(A \setminus B) = P(A) - P(B)$.

再由概率的非负性知

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

命题 1.5.18 (概率加法公式) 对任何事件A, B有

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

证 易见$A \cup B = A \cup (B \setminus AB)$，$A \cap (B \setminus AB) = \phi$及$AB \subset B$，再应用命题1.5.14，1.5.17得

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

命题 1.5.19 (有限加法公式) 对任何事件$A_1, A_2, \cdots, A_n$有

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_iA_j)$$

$$+ \sum_{1 \leqslant i < j < k \leqslant n} P(A_iA_jA_k) - \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n).$$

此命题由命题1.5.18及数学归纳法可证，证明留给读者. 请读者注意，概率的有限加法公式实际上是组合论中容斥原理的

变形. 我们举一个例子说明它的应用.

例 1.5.20 (配对问题) 某人写了 n 封信，又写了 n 个地址(彼此不同)，然后在每个信封中随意地装入一封信，试求没有一封信碰对地址的概率.

解 令

$$A=\text{“没有一封信碰对地址”}.$$

为求事件 A 的概率，转而考虑它的对立事件

$$A^c=\text{“至少有一封信碰对地址”}.$$

用 A_i 表示第 i 封信碰对地址(装对信封)，则

$$A^c=\bigcup_{i=1}^{n}A_i.$$

利用概率的有限加法公式，

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\cdots$$

$$+(-1)^{k-1}\sum_{i_1<i_2<\cdots<i_k}P(A_{i_1}A_{i_2}\cdots A_{i_k})$$

$$+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n).$$

如果将试验的结果看作由 $1,2,\cdots,n$ 组成的向量，即 $(j_1,j_2,\cdots,j_n)$，它表示第 k 封信装入第 j_k 个信封里，$k=1,2,\cdots,n$. 若 $j_k=k$ 就表示第 k 封信装对信封. 由全排列公式知样本空间由 $n!$ 个点组成，它们具有等可能性，第 $i_1,i_2,\cdots,i_k$ 封信同时装对了信封的事件 $A_{i_1}A_{i_2}\cdots A_{i_k}$ 表示 $j_{i_1}=i_1,j_{i_2}=i_2,\cdots,j_{i_k}=i_k$，而其它 $n-k$ 封信可以随意装在剩下来的 $n-k$ 个信封里，因此 $A_{i_1}A_{i_2}\cdots A_{i_k}$ 由 $(n-k)!$ 个样本点组成. 于是

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=\frac{(n-k)!}{n!}.$$

在和式 $\sum_{i_1<i_2<\cdots<i_k}P(A_{i_1}A_{i_2}\cdots A_{i_k})$ 中共有 $\binom{n}{k}$ 项，故

$$\sum_{i_1<i_2<\cdots<i_k} P(A_{i_1}A_{i_2}\cdots A_{i_k}) = \binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!} .$$

于是有

$$P(\bigcup_{k=1}^{n} A_k) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!} .$$

因而没有一封信碰对地址的概率为

$$\frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n}\frac{1}{n!} .$$

请读者应用这些结果计算出恰有 k 封信装对信封的概率.

最后，介绍一下关于概率的上、下连续性的结果.

命题 1.5.20 若事件 $A_1 \subset A_2 \subset \cdots$，则

$$\lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$$

这一性质称为概率的**下连续性**.

证 令 $B_1 = A_1$，由于 $\bigcup_{i=1}^{k} A_i = A_k$（利用了$\{A_n\}$是上升集序列），令

$$B_n = A_n \left(\bigcup_{i=1}^{n-1} A_i\right)^c = A_n \setminus A_{n-1} , n > 1$$

即 B_n 由属于 A_n，不属于一切 $A_i, i < n$ 的样本点组成，因此 B_n 两两互不相容，并满足

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \text{ 及 } \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i , \text{对一切 } n \geqslant 1 .$$

因此利用可数可加性及有限可加性得：

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i)$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} B_i)$$

$$=\lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} A_i\right)=\lim_{n\to\infty} P(A_n).$$

这就证明了本命题.

利用本命题及用对立事件求概率的方法可证

命题 1.5.21 若事件$A_1\supset A_2\supset\cdots$，则

$$\lim_{n\to\infty} P(A_n)=P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

该性质称为概率的**上连续性**.

习题 1－7

1. 事件 A、B 互不相容，已知 $P(A)=\frac{1}{2}$，$P(B)=\frac{1}{3}$，求

(1) $P(A\bigcup B)$; (2) $P(A\bigcap B)$; (3) $P(A\bigcup B^c)$;

(4) $P(A\bigcap B^c)$; (5) $P(A^c\bigcap B^c)$.

2. 证明

$$P(A_1\bigcup A_2\bigcup A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1\bigcap A_2)$$
$$-P(A_2\bigcap A_3)-P(A_1\bigcap A_3)+P(A_1\bigcap A_2\bigcap A_3).$$

3. 甲、乙两人对同一目标各射击一次，已知甲射中目标的概率是 0.9，乙射中目标的概率是 0.8，两人都射中目标的概率为 0.72，试求恰有一个人射中目标的概率.

4. 证明：若 A_1，A_2，$A_3\subset A$，则

$$P(A)\geqslant P(A_1)+P(A_2)+P(A_3)-2.$$

5. 对任意 n 个事件 A_1，A_2，…，A_n，证明

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

6. 证明命题 1.5.21.

复习题 1－A

1. 袋中装有 m 个红球，n 个黑球，现将球任意依次取出，试求第 k 次 $(k \leqslant m+n)$ 取得红球的概率，并同任取一球得红球的概率进行比较.

2. 在 N 个相同的球上分别写上号码 $1, 2, \cdots, N$，放在一个袋中. 现从袋中取 n 次球，每次任取一个，记录下号码然后放回，设 n 次取球中最大号码为 ξ，最小号码为 η，试求：(1) $P(\xi = k)$；(2) $P(\eta = k)$.

3. 袋中装着 n 个球，分别记有号记 $1, 2, \cdots, n$. 求

(1) 任意取出的两个球号码是 1, 2 的概率 p_1；

(2) 任意取出的三个球中没有号码 1 的概率 p_2；

(3) 任意取出的五个球中号码 1, 2, 3 至少出现一个的概率 p_3.

4. 袋中装有 7 个白球，4 个红球，现将球一个一个地摸出. 试求"不连续摸出两个红球"的概率.

5. 从 6 双不同的手套中任取 4 只，求"恰有一双配对"的概率.

6. 袋中装着三个白球，两个红球. 将球随机地一个一个往外取，直到取出两个红球为止. 若用 ξ 表示取出的球数，问 ξ 能取哪些值？ξ 取每一个值的概率是多少？

7. 某工作人员规定了一周内哪些天接见来访者，每个来访者在规定的这些天等可能地来访. 一个观察者观察了 12 次访问，恰巧都发生在星期二和星期四. 试问假如该工作人员规定每周三天(例如星期二、四、六)接受来访，那么上述事件发生的概率有多大？根据小概率事件实际不可能发生的原理，他是否能做出该工作人员只在星期二和星期四接受访问的判断？又假如这 12 次来访没有一次发生在星期日，是否可以很大的概率断言他星期日不接见来访者？

8. 一信号随机地在区间 $[0, T]$ 中某时刻发出，它的持续时间为 Δ，收报机也在 $[0, T]$ 中随机地打开，它的工作时间为 t，求收报机发现信号的概率.

9. 在 $(0,1)$ 中随机地取两个数 ξ，η.

(1) 证明：对任意 $a, b \in (0,1)$，成立等式

$$P(\xi < a, \eta < b) = P(\xi < a)P(\eta < b) = ab;$$

(2) 给定 $a \in (0,1)$，求 $P(\min\{\xi, \eta\} \leqslant a)$，$P(\max\{\xi, \eta\} < a)$ 及 $P(\xi + \eta < a)$.

10. 平面上画有横、纵两组互相垂直的等距离平行线，它们把平面划分

成一些边长为a、b的长方形. 现在向平面任投一长为l ($l<a, l<b$)的针. 试求:

(1) 针与横、纵线各一条同时相交的概率;

(2) 针与任一平行线相交的概率.

11. 设$\{A_n\}$为一列随机事件, 若$\sum_{n=1}^{\infty} P(A_n)<\infty$, 证明

$$P\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c\right)=1.$$

复习题 1-B

1. 袋中装着标号分别为$1, 2, \cdots, N$的N个球; 现有放回地摸出n个球, 依次记下号码, 试求:

(1) 这些号码按严格上升次序排列的概率;

(2) 这些号码按不降的次序排列的概率.

2. 将一颗骰子掷3次, 用ξ表示3次掷得的点数之和, 试比较$P(\xi=12)$与$P(\xi=11)$的大小.

3. N个球随机地落入n个盒中($N\geqslant n$), 求没有空盒的概率.

4. 设A, a为正整数, $A>a$, 利用概率论想法证明

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots+\frac{(A-a)\cdots 2\cdot 1}{(A-1)\cdots(a+1)a}=\frac{A}{a}.$$

5. 在一水平面上画有间隔为d的等距离平行线, 现向此平面任意投掷一个三角形, 这个三角形的边长分别为a, b, c(都小于d), 试求三角形与平行线相交的概率.

6. 利用概率论的想法求n阶行列式的展开式中含主对角线元素的项数.

7. 盒中有n个编号从1到n的相同的球, 将球不放回地依次取出. 试求恰有k($1\leqslant k\leqslant n$)次取出的球其号码与试验号码相同的概率.

8. 从一副扑克牌(52张)中有放回地每次取出一张, 试求:

(1) "抽取的n张牌($n\geqslant 4$)包含了全部四种花色"的概率;

(2) "直到第n次抽牌, 才使抽出的牌包含了全部四种花色"的概率.

第二章　条件概率和独立

条件概率是概率计算的重要工具，独立是概率论与数理统计中的一个重要概念．它们在理论上和实际的概率计算中起着非常重要的作用．

§2.1 条 件 概 率

应该说，每一个随机试验都是在一定条件下进行的．而这里要讨论的条件概率则是当试验结果的一部分信息已知(即在原随机试验的条件上，再加上一定的条件)，例如某一事件 B 已经发生时，求另一事件 A 在此条件下发生的概率．

例 2.1.1 设在 N 件产品中有 M 件次品，今依次取出两件(每次取出后不放回)．如果已经知道取出的第一件产品为次品，求在此条件下第二件产品也是次品的概率．

解 随机试验是从 N 件产品中依次抽取两件，因而样本空间 Ω 由 $N(N-1)$ 个样本点组成，且 Ω 中每个样本点的出现都是等可能的，令

$$B_i = \text{“第}i\text{ 次抽到次品”}, \quad i=1,2,$$

则 B_1 是由 $M(N-1)$ 个样本点组成的集．已知 B_1 发生，在此条件下 B_2 发生，相当于 B_1B_2 发生．它由 $M(M-1)$ 个样本点组成．要求在 B_1 发生的条件下 B_2 发生的概率(记作 $P(B_2|B_1)$)相当于把 B_1 看作新的样本空间来计算 B_1B_2 发生的概率，即

$$P(B_2|B_1)=\frac{B_1B_2\text{所含的样本点数}}{B_1\text{所含的样本点数}}=\frac{M(M-1)}{M(N-1)}=\frac{M-1}{N-1}.$$

若用原来的样本空间 Ω 中的概率来表示，就有

$$P(B_2|B_1)=\frac{B_1B_2\text{所含样本点数}/\Omega\text{所含样本点数}}{B_1\text{所含样本点数}/\Omega\text{所含样本点数}}$$

$$=\frac{P(B_1B_2)}{P(B_1)}.$$

例 2.1.2 由正方形区域Ω 内随机取点，若已知点取自区域B内，求在此条件下点取自区域 A 内的概率.(如图 2－1).

解 用 $\mu(E)$表示区域 E 的面积，E 表示"点取自集E"的事件. 已知B发生，在此条件下 A 发生的概率 $P(A|B)$相当于在 B 中随机取点，点落在集合 AB 中的概率，因此

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

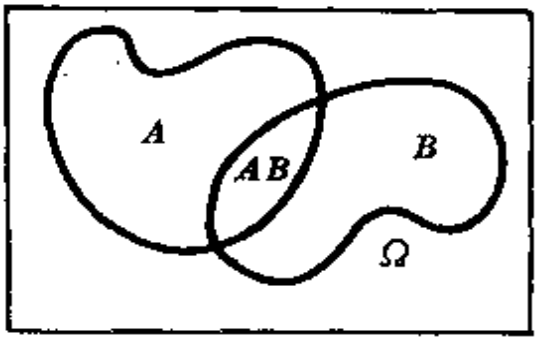


图 2－1

用原来 Ω 空间中的概率来表示这个结果，就有

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

例 2.1.3 某随机试验进行了n次，其中事件 B、AB 分别发生了 n_B、n_{AB}次，试求在事件 B 发生的条件下事件A 发生的频率.

解 在 n 次试验中事件 B 发生了 n_B 次,其中事件 A 发生了n_{AB}次，因此在事件 B 发生的条件下事件 A 发生的频率

$$\mu_{A|B}=\frac{n_{AB}}{n_B}=\frac{n_{AB}/n}{n_B/n}=\frac{\mu_{AB}}{\mu_B},$$

其中 μ_B、μ_{AB}分别为 n 次试验中事件 B、AB 发生的频率.

当试验次数 n 很大且 $P(B)>0$ 时，μ_B 接近于 $P(B)$(因而 $\mu_B>0$)，μ_{AB} 接近于 $P(AB)$，而 $\mu_{A|B}$ 应接近于 $P(A|B)$(注意：这里说的接近是指当 n 很大时频率稳定于概率意义下的接近)，即

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

从以上三个例子，可以给出如下的

定义 2.1.4 设 $(\Omega,\mathcal{F},P)$ 为概率空间，$B\in\mathcal{F}$ 且 $P(B)>0$，则对任意的 $A\in\mathcal{F}$，称

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

为在事件 B 发生的条件下事件 A 发生的**条件概率**.

例 2.1.5 掷两颗均匀的骰子，问在已知它们点数不同的条件下，至少有一颗是6点的条件概率是多少？又问至少有一颗是6点的(无条件)概率又是多少？

解 对两颗骰子加以区别，则样本空间由 36 个样本点组成，它们是等可能的. 令

$$A=\text{“至少有一颗是 6 点”},$$

$$B=\text{“两颗骰子点数不同”}.$$

事件 AB 由 10 个样本点组成，事件 B 由 $6\times 5=30$ 个样本点组成，因而所求的条件概率

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{10/36}{30/36}=\frac{1}{3}.$$

而事件 A 由 11 个样本点组成，故

$$P(A)=11/36.$$

这里条件概率 $P(A|B)$和(无条件)概率 $P(A)$是不同的.事件 B 已发生的条件改变了事件 A 发生的可能性大小.

在概率计算中常常需要计算事件乘积的概率 $P(AB)$,例如在求 $P(A\bigcup B)$时，如果 $P(A)$，$P(B)$均易于求得，那么就归结为求 $P(AB)$的问题.如果直接计算 $P(AB)$较难，但计算 $P(B)$及 $P(A|B)$比较容易时，我们可以用下述公式计算 $P(AB)$.

定理 2.1.6 (概率乘法公式) 若$(\Omega,\mathscr{F},P)$为概率空间，A，$B\in\mathscr{F}$且 $P(B)>0$，则

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

概率乘法公式在概率计算中非常有用.下面利用它来研究著名的波伊亚(Polya)罐子模型中一些事件的概率以及摸彩不论先后中彩的机会均等等问题.下一节还将多次用到概率乘法公式.

例 2.1.7 (波伊亚罐子模型) 设一个罐中包含b 个黑球 r 个红球，随机地取一个球而后放回，并且加进去 c 个与所抽取的球同样颜色的球.(这个模型可粗糙地解释疾病的传染.每抽出一个球，代表疾病的一次传染，每一次传染都增加这种疾病再传染的概率.)试求

(1) 第一次抽到黑球，第二次抽到红球的概率;

(2) 第一次抽到红球，第二次抽到黑球的概率;

(3) 第一、二次抽到红球，第三次抽到黑球的概率;

(4) 第一、三次抽到红球，第二次抽到黑球的概率.

解 令 B_i,R_i分别表示第 i 次取到黑球和红球的事件，$i=1$，2,3.

(1)中所求概率为 $P(B_1R_2)$，根据概率乘法公式

$$P(B_1R_2)=P(B_1)P(R_2|B_1),$$

在第一次抽取时罐中有 b 个黑球r 个红球，因而

$$P(B_1)=\frac{b}{b+r},$$

当第一次取出的是黑球的条件下第二次抽球时罐中有$b+c$个黑球r个红球，因此第一次抽得黑球的条件下第二次抽到红球的条件概率

$$P(R_2|B_1)=\frac{r}{b+r+c}.$$

于是

$$P(B_1R_2)=P(B_1)P(R_2|B_1)=\frac{b}{b+r}\cdot\frac{r}{b+r+c}.$$

同样地讨论(或根据b和r的对称性)可得(2)中所求概率

$$P(R_1B_2)=P(R_1)P(B_2|R_1)=\frac{r}{b+r}\ \frac{b}{b+r+c}=P(B_1R_2).$$

(3) 连续使用概率乘法公式

$$\begin{aligned}P(R_1R_2B_3)&=P(B_3|R_1R_2)P(R_1R_2)\\&=P(B_3|R_1R_2)P(R_2|R_1)P(R_1).\end{aligned}$$

第一次取得红球后罐中有$r+c$个红球b个黑球，故

$$P(R_2|R_1)=\frac{r+c}{b+r+c},$$

而在前两次取得红球的条件下第三次抽取时罐中有$b+r+2c$个球，其中黑球仍为b个，于是

$$P(B_3|R_1R_2)=\frac{b}{b+r+2c},$$

于是(3)中所求概率

$$P(R_1R_2B_3)=\frac{r(r+c)b}{(b+r)(b+r+c)(b+r+2c)}.$$

(4) 中所求概率可用计算(3)中所求概率同样方法得到，

$$P(R_1B_2R_3)=P(R_3|R_1B_2)P(B_2|R_1)P(R_1)$$

$$=\frac{r+c}{b+r+2c}\cdot\frac{b}{b+r+c}\cdot\frac{r}{b+r}$$

$$=P(R_1R_2B_3).$$

例 2.1.8 制作 n 张彩票，其中只有一张写有“电影票”字样，n 个人依次摸彩，摸到标有“电影票”字样的称为中彩. 试求

(1) 已知前 $k-1$ $(k\leqslant n)$ 个人都没有中彩的条件下第 k 个人中彩的概率；

(2) 第 k $(k\leqslant n)$ 个人中彩的概率.

解 令

$$A_j=\text{“第 }j\text{ 个人中彩”},\ j=1,2,\cdots,n\ .$$

则 A_j 两两互不相容(因为只能有一个人中彩).

(1) 前 $k-1$ 个人均未中彩，第 k 个人摸彩时还有 $n-k+1$ 张彩票，其中有一张标有“电影票”字样，于是(1)中所求条件概率为

$$P\left(A_k|\bigcap_{j=1}^{k-1}A_j^c\right)=\frac{1}{n-k+1}\ .$$

(2) 由于 $A_1,A_2,\cdots,A_k$ 两两互不相容，因此 $A_k\subset\bigcap_{j=1}^{k-1}A_j^c$，$A_k=A_kA_1^cA_2^c\cdots A_{k-1}^c$，重复使用乘法公式：

$$P(A_k)=P(A_kA_1^cA_2^c\cdots A_{k-1}^c)$$

$$=P(A_k|A_1^cA_2^c\cdots A_{k-1}^c)P(A_{k-1}^c|A_1^c\cdots A_{k-2}^c)$$

$$\cdots P(A_2^c|A_1^c)P(A_1^c)$$

$$=\frac{1}{n-k+1}\ \frac{n-k+1}{n-k+2}\ \frac{n-k+2}{n-k+3}\cdots$$

$$\cdot\frac{n-2}{n-1}\ \frac{n-1}{n}=\frac{1}{n}\ ,$$

$$k=1,2,\cdots,n\ .$$

这说明摸彩不论先后，中彩的机会是均等的. 类似地，在体育比赛中抽签不论先后，机会也是均等的.

习题 2-1

1. 四个球连续地放进四个盒中，全部4^4个排列法是等可能的，若已知前两个球放在不同的盒中，求某一盒中恰有三个球的概率.

2. 若 M 件产品中包含m 件废品. 今在其中任取两件，求:

(1) 已知取出的两件中有一件废品的条件下另一件也是废品的概率;

(2) 已知两件中有一件不是废品的条件下另一件是废品的概率;

(3) 取出的两件至少有一件是废品的概率.

3. 设 $P(B)>0$, 证明

$$P(A|B)\geqslant 1-\frac{P(A^c)}{P(B)}\ .$$

4. 设有 N 个袋子，每个袋子中装有 a 个白球，b 个黑球.从第一袋中随机地取出一球放入第二袋中，再从第二袋中取出两球放入第三袋中，如此继续下去，最后从第 N 袋中取出 N 个球. 求每次取出的全为黑球的概率.

5. 某学生可以选修甲、乙两门课程之一，他估计学习课程甲获得成绩"优"的概率为 1/2, 而学习课程乙获得"优"的概率为 1/3. 该生通过掷一枚硬币出现正反而来决定选哪一门课. 求该生选了课程乙且获得 "优"的概率.

§2.2 全概率公式和贝叶斯公式

我们从讨论两个概率计算的例子谈起.

例 2.2.1 继续例2.1.7 的讨论. 如果我们要计算"第二次取得黑球"即事件 B_2 的概率，该怎样进行呢? 很自然的想法是要区别两种情况: 第一次取得是红球还是黑球，即把事件 B_2 分解为两个事件的并

$$B_2 = R_1 B_2 \cup B_1 B_2 .$$

由于 R_1 和 B_1 互不相容，由概率的有限可加性可得

$$P(B_2) = P(R_1 B_2) + P(B_1 B_2) .$$

再利用概率乘法公式及例 2.1.7 的结果及算法，我们有

$$\begin{aligned} P(B_2) &= P(B_2|R_1)P(R_1) + P(B_2|B_1)P(B_1) \\ &= \frac{rb}{(b+r)(b+r+c)} + \frac{b(b+c)}{(b+r)(b+r+c)} \\ &= \frac{b}{b+r} . \end{aligned}$$

例 2.2.2 某工厂有甲、乙、丙三台机器生产同样的零件，它们的产量各占25%，35%，40%，而在各自的产品中废品率分别为5%，4%，2%. 试求该厂生产的这种零件的废品率.

例 本例相当于在该厂生产的这些零件中任取一件，求取得废品的概率. 同例2.2.1一样，必须分情形讨论，区分成它是来自甲、乙、丙三台机器中的哪一台. 令

A = "取得的零件是废品"；

B_1 = "取得的零件为甲台机器所生产"；

B_2 = "取得的零件为乙台机器所生产"；

B_3 = "取得的零件为丙台机器所生产".

由于 B_1，B_2，B_3 两两互不相容且 $\bigcup_{i=1}^{3} B_i = \Omega$，故有

$$P(A) = P(AB_1) + P(AB_2) + P(AB_3) .$$

再用乘法公式及各台机器产量百分比及废品率，我们得到

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

$$=0.05\times0.25+0.04\times0.35+0.02\times0.4$$
$$=0.0345=3.45\%\ .$$

将这两个例中所用的方法推广到一般的情形，就得到在概率计算中非常有用的**全概率公式**，即

定理 2.2.3 若$(\Omega,\mathcal{F},P)$为概率空间，$B_1,B_2,\cdots,B_n$为两两互不相容的正概率事件，$A\subset\bigcup_{k=1}^{n}B_k$，则

$$P(A)=\sum_{k=1}^{n}P(A|B_k)P(B_k)\ .$$

证明请读者按上面说明去完成.

下面的例子一方面使我们进一步了解全概率公式的应用，同时还给我们以新的启示.

例 2.2.4 在标准化考试中，一道考题同时列出四种选择答案，要求学生把其中唯一的正确答案选择出来.设某考生知道正确答案的概率为p，不知道正确答案的概率为$1-p$，不知道正确答案的条件下答对的概率应为$\frac{1}{4}$.考试后已知他答对了，求他知道正确答案的概率.

解 令

$A=$“考生答对了”，

$B=$“考生知道正确答案”.

则B^c表示“考生不知道正确答案”，由题设已知

$$P(B)=p\ ,P(B^c)=1-p\ ;$$
$$P(A|B^c)=\frac{1}{4}\ ,P(A|B)=1\ ,$$

利用乘法公式及全概率公式可得所求的条件概率为

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

$$= \frac{p}{p+(1-p)\cdot\frac{1}{4}}.$$

如果 $p=\frac{1}{2}$，则 $P(B|A)=\frac{4}{5}$；如果 $p=\frac{2}{3}$，则 $P(B|A)=\frac{8}{9}$.

上例中公式的推导，事实上已经得到一个重要的公式——**贝叶斯(Bayes)公式**.

定理 2.2.5(贝叶斯定理) 在定理 2.2.3 的假设下，若还有 $P(A)>0$，则

$$P(B_k|A)=\frac{P(B_k)P(A|B_k)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)},\quad k=1,2,\cdots,n.$$

贝叶斯定理非常著名，发表于 1763 年. 它被想象为在观察到事件 A 已发生的基础上给出每个原因 B_k 的概率. $P(B_k)$和 $P(B_k|A)$分别称为原因 B_k 的**验前概率和验后概率**.在自然现象和人类活动的各个领域有很多应用.它可以帮助人们确定引起事件A发生的最可能的原因是什么.例如在上例中，考生答对了(事件 A 发生)，由 $P(B|A)=\frac{4}{5}$，$P(B^c|A)=\frac{1}{5}$ (当 $p=\frac{1}{2}$ 时)，我们可以得出：最可能的原因是考生知道正确答案而不是乱猜. 又如 A 是"目标被击中"，B_k 是"第 k 号射击手射击"，则$P(B_k|A)$的大小可以用来判断目标被哪位射击手击中的可能性大.

但在使用贝叶斯公式时往往由于不知道验前概率而发生困难，于是发生了关于贝叶斯统计学派的争论.一派人认为既然不知道验前概率,就无法应用贝叶斯公式；而贝叶斯学派则认为概率是对某一陈述相信程度的度量.在完全无知的情况下,可以认为各种原因是等可能的，即 $P(B_k)=\frac{1}{n}$，或者根据已有经验给出一个

关于 $P(B_k)$，$k=1,\cdots,n$ 的主观估计，然后通过试验，不断地用验后概率来修正.例如在例2.2.4中，如果不能肯定考生知道正确答案的概率 p 究竟是 $\frac{1}{2}$ 还是 $\frac{2}{3}$，那么经过几次考试，如果考生接连几次都答对了，这就使人们有理由确信 p 更可能是2/3.这就是贝叶斯统计学派的基本思想.1960年以前，大多数统计学家反对贝叶斯学派，甚至避免使用贝叶斯公式.然而从60年代起，贝叶斯统计学派吸引了众多的追随者；进一步的发展还会为它开辟更为广阔的应用场所.

例 2.2.6 有朋自远方来，他乘火车、汽车、飞机来的概率分别为0.3，0.45，0.25，他乘火车、汽车、飞机来迟到的概率分别为 $\frac{1}{4}$，$\frac{1}{3}$，$\frac{1}{12}$.结果他迟到了，试问他乘火车、汽车、飞机来的概率分别是多少?

解 令 A 表示"迟到"，$B_{火}$，$B_{汽}$，$B_{飞}$分别表示乘火车、汽车、飞机来.则

$$P(B_{汽}|A)=\frac{P(B_{汽})P(A|B_{汽})}{P(B_{火})P(A|B_{火})+P(B_{汽})P(A|B_{汽})+P(B_{飞})P(A|B_{飞})}$$

$$=\frac{0.45\times\frac{1}{3}}{0.3\times\frac{1}{4}+0.45\times\frac{1}{3}+0.25\times\frac{1}{12}}=0.61,$$

$$P(B_{火}|A)=\frac{0.3\times\frac{1}{4}}{0.3\times\frac{1}{4}+0.45\times\frac{1}{3}+0.25\times\frac{1}{12}}=0.31,$$

$$P(B_{飞}|A)=1-0.61-0.31=0.08.$$

因此迟到的原因最可能是因为乘汽车来的.

例 2.2.7 一项血液化验有95%的把握将患肝炎病的人鉴别出来，但用于健康人也会有 1% 的错诊(即一个健康人化验结果为阳性的可能性为 1%). 假设肝炎病人占人口总数的 0.5%. 若某人的化验结果为阳性，那么他确实患肝炎病的概率有多大?

解 令

$A=$"被化验者患肝炎病",

$B=$"被化验者的化验结果为阳性".

已知 $P(A)=0.005$, $P(A^c)=0.995$, $P(B|A)=0.95$, $P(B|A^c)=0.01$. 由贝叶斯公式，所求的条件概率

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

$$=\frac{0.95\times 0.005}{0.95\times 0.005+0.01\times 0.995}$$

$$=0.323.$$

这说明，当某种疾病在人群中只占很小的比例，即使一种诊断方法把健康人误诊为患者的可能性很小，但是仅靠一种方法诊断某人患此疾病的准确性还是比较低的. 在此情况下仍须对可疑者用另外的方法诊断，才能减少误诊.

例 2.2.8 设有三张形状完全相同的卡片，第一张两面全是红色，第二张两面全是黑色；第三张一面红色一面黑色. 将这三张卡片随机地选取一张随机地抛在桌面上. 发现这张卡片朝上的一面是红色，求在此条件下它的另一面是黑色的概率是多少?

解 乍一看，人们很容易认为：既然取出的卡片朝上的一面为红色，那么这张卡片只能是题目所示的第一张或第三张. 取得这两张卡片的任一张是等可能的，面只有取到第三张才可能使得一面红一面黑，因此所求概率为 $\frac{1}{2}$.

事实上，上述结论是错的，在取出的卡片红面朝上的条件下所取的卡片是第一张还是第三张并不是等可能的．因为每张卡片由两个面组成．那么此试验就有 6 个等可能结果．用 R_1，R_2 分别表示第一张卡片第 1，2 个红面朝上；B_1，B_2 分别表示第二张卡片的第 1，2 个黑面朝上；R_3，B_3 分别表示第 3 张卡片的红面，黑面朝上．于是

$$\Omega=\{R_1,R_2,B_1,B_2,R_3,B_3\}.$$

而事件

$R=$“取出的卡片朝上的一面为红色”$=\{R_1,R_2,R_3\}$

由 3 个样本点组成，事件

$B=$“取出的卡片红面朝上，黑面朝下”$=\{R_3\}$

由一个样本点组成，故所求的条件概率

$$P(B\mid R)=\frac{1}{3}.$$

我们也可以用贝叶斯公式来计算所要求的条件概率．令 A_1，A_2，A_3 分别表示抽出的卡片是第一、二、三张，R 表示抛到桌上红面向上．由于 A_1，A_2，A_3 两两互不相容且 $P(A_i)=\frac{1}{3}$，$i=1,2,3$；$P(R\mid A_i)=1$，$P(R\mid A_2)=0$，$P(R\mid A_3)=\frac{1}{2}$，从而

$$P(A_3\mid R)=\frac{P(A_3R)}{P(R)}$$

$$=\frac{P(A_3)P(R\mid A_3)}{P(R\mid A_1)P(A_1)+P(R\mid A_2)P(A_2)+P(R\mid A_3)P(A_3)}$$

$$=\frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}}=\frac{1}{3}.$$

习题 2－2

1. 甲、乙两人各从 1, 2, …, 15 中任取一数(不放回)，求已知甲取到的数是 5 的倍数的条件下，甲数大于乙数的概率.

2. 在第一盒中有 6 个白球，4 个黑球；第二盒中有 5 个白球，10 个黑球. 从第一盒中取出一球放入第二盒中，再从第二盒中取出一球，问这球为黑球的概率是多少?

3. 某保险公司的统计表明，新保险的汽车司机中可按年龄划分为易出事故的和较为谨慎的两类. 易出事故的人在一年中出事故的可能性为 0.4，较为谨慎的人则为 0.2. 假若易出事故的人占新保险的汽车司机的 30%，那么，一个新保险的汽车司机在一年内出事故的概率是多少?

4. 两台钻机按相同的规格钻孔，第一台钻机钻的孔的不合格率为 0.05，第二台为 0.1，现在发现一钻孔不合格，求它是第一台钻机钻出来的概率.

5. 已知产品中 96% 是合格品. 现有一种简化的检查方法. 它把真正合格的确认为合格的概率为 0.98，而把废品误认为合格品的概率为 0.05. 求在这种方法下，按检查为合格品的一个产品确实为合格品的概率.

§2.3 事件的独立

在前两节所述的许多例子中，条件概率 $P(A|B)$ 一般不等于 $P(A)$，也就是说事件 B 已发生的信息会影响事件 A 出现的机会. 但在有些情况下 $P(A|B)=P(A)$，即事件 B 的发生不影响事件 A 发生的概率，这时我们称 A 独立于 B. 由条件概率的定义知: 若 A 独立于 B，即 $P(A|B)=P(A)$，则

$$P(AB)=P(A)P(B). \tag{2.3.1}$$

且除了 $P(B)=0$ 的情形外，两者是等价的. 在(2.3.1)中 A 和 B 的关系是对称的，并且在 $P(B)=0$ 时也有意义且必然成立. 于是我们给出下述

定义 2.3.2 若概率空间 $(\Omega, \mathcal{F}, P)$ 中的事件 A, B 的概率满足(2.3.1)，则称事件 A 与 B 独立.

例 2.3.3 从一副52张的扑克牌中随机地抽取一张，以 A 表示抽出的一张牌是"A"，B 表示抽出的牌是黑桃，则 $P(AB)=\frac{1}{52}$，$P(A)=\frac{4}{52}$，$P(B)=\frac{13}{52}$，于是

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

因此事件 A，B 独立.

例 2.3.4 考虑有三个孩子的家庭，按大小顺序，不同的性别分布有8种等可能的情形. 即

$\Omega=\{$(男，男，男)，(男，男，女)，(男，女，男)，(女，男，男)

(女，男，女)，(女，女，男)，(男，女，女)，(女，女，女)$\}$.

令 $A=\{$家中既有男孩又有女孩$\}$，

$B=\{$家中至多有一个女孩$\}$.

则 A 由6个基本事件组成，B 由4个基本事件组成，AB 由3个基本事件组成. 于是

$$P(AB)=\frac{3}{8},\ P(A)P(B)=\frac{6}{8}\cdot\frac{4}{8}=\frac{3}{8},$$

故 A，B 独立.

若我们考虑的是两个孩子的家庭，则

$$\Omega=\{\text{(男，男)，(男，女)，(女，男)，(女，女)}\},$$

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

此时 $P(AB)\neq P(A)P(B)$，事件 A，B 不独立.

这个例子说明两个事件是否独立不能完全靠直觉，要对随机

现象仔细研究考察.

从直观意义容易得出这样的结论：如果已知 B 发生不影响事件 A 发生的概率，那么事件 B 不发生也应不影响事件 A 发生的概率，这就是

命题 2.3.5 若事件 A, B 独立，则事件 A, B^c 也独立.

证 已知 $P(AB)=P(A)P(B)$，利用概率的性质

$$P(A)=P(AB^c)+P(AB),$$

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

因而 A, B^c 满足(2.3.1)，故 A, B^c 独立.

由上述命题立得：若 A 与 B 独立，则 A 与 B^c，A^c 与 B，A^c 与 B^c 都独立.

对于三个事件 A, B, C 的独立，自然应要求任何一个事件发生的概率不受其它事件发生与否的影响，因此除要求 A 与 B，B 与 C，C 与 A 独立而外，还应要求 A 与 BC，B 与 AC，C 与 AB 独立，即 $P(ABC)=P(A)P(BC)=P(A)P(B)P(C)$ 等. 因此我们给出如下的

定义 2.3.6 称事件 A, B, C 独立，如果下述四个等式同时成立：

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

$$P(BC)=P(B)P(C),$$

$$P(AC)=P(A)P(C),$$

$$P(ABC)=P(A)P(B)P(C).$$

显然由上述定义容易推出若 A, B, C 独立则 A 与 BC，B 与 AC，C 与 AB 都独立.

根据独立的意义，A, B, C 独立，则 A 发生的概率不受 B, C 发生与否的影响，因而容易想到，若 A, B, C 独立，则 A 与 $B \cup C$, A 与 $B \setminus C$ …… 都独立. 下面的命题的证明证实了这些直观想法的正确性.

命题 2.3.7 若 A, B, C 独立，则 A 与 BC; A 与 $B \cup C$; A 与 $B \setminus C$; A, B^c, C^c; 及 $A, B^c; C$…… 都独立.

证 我们只证 A 与 $B \cup C$ 独立，其它结论的证明留给读者. 由概率加法公式及 A, B, C 独立的定义，有

$$\begin{aligned} P(A(B \cup C)) &= P(AB \cup AC) = P(AB) + P(AC) - P(ABC) \\ &= P(A)P(B) + P(A)P(C) - P(A)P(B)P(C) \\ &= P(A)[P(B) + P(C) - P(B)P(C)] \\ &= P(A)[P(B) + P(C) - P(BC)] \\ &= P(A)P(B \cup C), \end{aligned}$$

这就得出 A 与 $B \cup C$ 独立.

仿照定义 2.3.6，给出下述任意多个事件独立的定义.

定义 2.3.8 称事件 $A_1, A_2, \cdots, A_n$ 独立，如果对于任意的 $k(2 \leqslant k \leqslant n)$ 和任意的一组 $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$，都有

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

成立.

称事件序列 $A_1, A_2, \cdots, A_n, \cdots$ 独立，如果它们之中的任意有限个事件独立.

一些读者往往认为：若事件 $A_1, A_2, \cdots, A_n (n \geqslant 3)$ 两两独立，则事件 $A_1, A_2, \cdots, A_n$ 独立. 下面的例子说明这是不对的.

例 2.3.9 一个正四面体，其中三个面分别涂上红、绿、黄三种颜色，第四个面同时涂上红、黄、绿三种颜色. 随意地抛掷该

四面体，设各面着地是等可能的．令 A、B、C 分别表示着地的一面涂有红、绿、黄色．则易知

$$P(A)=P(B)=P(C)=\frac{2}{4}=\frac{1}{2},$$

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

因此 A，B，C 两两独立．但是

$$P(ABC)=\frac{1}{4}\neq P(A)P(B)P(C).$$

这说明三个以上事件两两独立不能保证它们独立．

读者不难举出 $P(ABC)=P(A)P(B)P(C)$ 而 $P(AB)\neq P(A)\cdot P(B)$ 的例子来．

类似于命题 2.3.5 及 2.3.7，对于多个随机事件的情形，有

定理 2.3.10 若 $A_1, A_2, \cdots, A_n$ 独立，则

(1) $A_1', A_2', \cdots, A_n'$ 独立，其中 $A_k'=A_k$ 或 A_k^c；

(2) 将事件 $A_1, A_2, \cdots, A_n$ 分成 k 组(不重不漏)；设 $B_1, B_2, \cdots, B_k$ 分别由第 $1,2,\cdots,k$ 组内的 A_i 经过并、积、差、求余等运算所得，则 $B_1, B_2, \cdots, B_k$ 独立．

我们不去证明这个定理，读者可以作为练习证明定理中所包含的一些特例，这一定理的结论今后常用到．

前面我们介绍过，事件是否独立(特别是多个事件是否独立)不能完全靠直觉，要对随机事件仔细研究，经过计算得出结论．这是讲问题的一个方面，即根据数学定义来检查事件是否独立的问题．但在实际应用时，往往根据问题的实际意义，可以认为某些事件是相互独立的．然后运用独立的概念，进行推导和计算，而且这正是独立性在概率及数理统计的理论上及应用中发挥重要作用的所在．例如在产品检验中，如果是有放回的随意抽样，那么可以认为“第一次抽到废品”，“第二次抽到废品”，……，“第 n 次抽到废品”这些事件是独

立的. 这是因为每次抽取后又将产品放回，在这一批产品中废品的个数与正品的个数都不发生变化，每次抽取是否抽到废品的概率与其它次抽到什么没有关系. 又如 n 个人进行射击，如果互不干扰，则也可以认为"第一个人射中"，"第二个人射中"，……，"第 n 个人射中"诸事件是独立的，等等(当然，在有的情况下难以作出判断，还需要用一些统计方法进行检验). 正是基于在实际问题中对事件独立性的判断，才使许多概率计算和应用较易进行. 下面举几个例子.

例 2.3.11 设有两门高射炮各自独立进行射击，每一门击中飞机的概率都是0.6，问同时各发射一发炮弹击中飞机的概率是多少？欲以99% 以上的概率击中它，问至少需要多少门高射炮同时射击?

解 用 A_i 表示"第 i 门高射炮击中飞机"，$i=1,2$，则两门炮同时各发射一枚炮弹击中飞机的概率为

$$P(A_1 \bigcup A_2)=P(A_1)+P(A_2)-P(A_1A_2).$$

由于两门炮独立射击，因而可以认为 A_1, A_2 独立. 故

$$P(A_1)=P(A_2)=0.6,$$

$$P(A_1A_2)=P(A_1)P(A_2)=0.6^2,$$

即

$$P(A_1 \bigcup A_2)=0.6+0.6-0.6^2=0.84.$$

设有 n 门炮同时射击，击中的概率应为 $P(A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n)$，当 n 较大时用一般概率加法公式计算是很麻烦的事. 我们可以转而求 $(A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n)^c$ 的概率. 利用定理 2.3.10 由 $A_1, A_2, \cdots, A_n$ 独立可以导出 $A_1^c, A_2^c, \cdots, A_n^c$ 也独立，从而

$$P(\bigcup_{k=1}^{n} A_k)=1-P((\bigcup_{k=1}^{n} A_k)^c)$$

$$=1-P(A_1^cA_2^c\cdots A_n^c)=1-0.4^n.$$

欲使上述概率大于0.99，只要 n 满足

$$0.4^n < 0.01\ ,$$

即 $n > \dfrac{\ln 0.01}{\ln 0.4} \approx 5.026$ 即可. 所以只要6门以上高射炮同时射击，击中飞机的概率即可达0.99以上.

例 2.3.12 用 $2n$ 个相同的元件(例如整流二极管)组成一个系统，有两种不同的联结方式，第一种是先串联后并联(如图2－2)；第二种联结方式是先并联后串联(如图2－3所示).

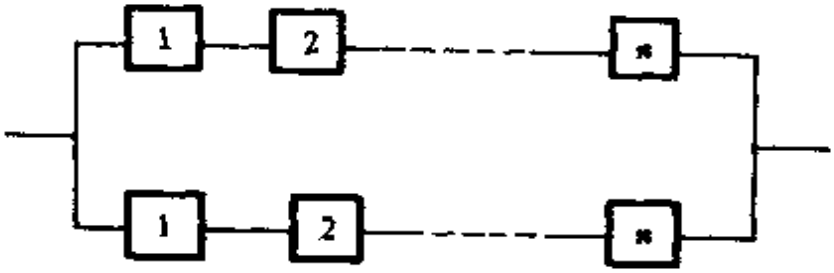


图 2－2

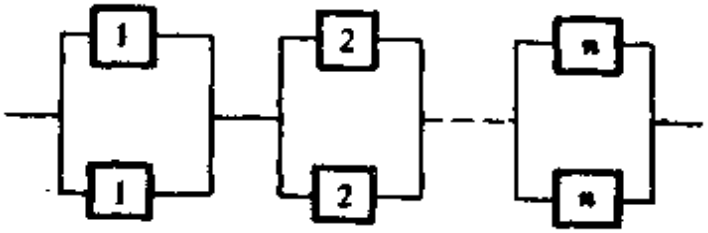


图 2－3

如果各个元件能否正常工作是相互独立的，每个元件能正常工作的概率为 γ. 元件或系统能正常工作的概率称为可靠度，试比较两个系统哪一个的可靠度更大.

解 对于系统Ⅰ，它有两条通路工作，分别记两条通路的可靠度为 $R_{\mathrm{I},1}, R_{\mathrm{I},2}$. 因为各个元件能否正常工作是相互独立的，因而

$$R_{\mathrm{I},1} = R_{\mathrm{I},2} = \gamma^n .$$

系统Ⅰ正常工作的充分必要条件是两条通路至少有一条能正常

工作，因而系统 I 的可靠度

$$R_{\mathrm{I}}=1-(1-R_{\mathrm{I},1})(1-R_{\mathrm{I},2})=1-(1-\gamma^{n})^{2}$$
$$=\gamma^{n}(2-\gamma^{n}).$$

对于系统 II，先求出每一个并联的小节(图 2－4)的可靠度 $R_{\mathrm{II},i}$，$i=1,2,\cdots,n$，由独立性知

$$R_{\mathrm{II},i}=1-(1-\gamma)^{2}=\gamma(2-\gamma).$$

系统 II 能正常工作当且仅当每一小节都正常工作，再次应用独立性 (用到了定理 2.3.10)得

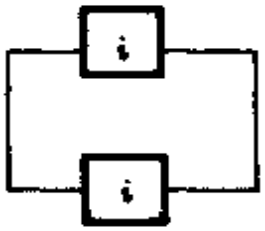


图 2－4

$$R_{\mathrm{II}}=R_{\mathrm{II},1}R_{\mathrm{II},2}\cdots R_{\mathrm{II},n}$$
$$=[\gamma(2-\gamma)]^{n}$$
$$=\gamma^{n}(2-\gamma)^{n}.$$

易证当 $n\geqslant 2$ 时

$$(2-\gamma)^{n}>2-\gamma^{n}.$$

因此系统 II 可靠度大.

例 2.3.13 某人进行射击，每次中靶的概率为 p，设各次射击中靶与否彼此独立，用 ξ 表示首次中靶时射击的次数，试求 $P(\xi=n)$ 及某人永远射不中的概率.

解 用 A_i 表示事件“第 i 次中靶”，则

$$\{\xi=n\}=A_1^{c}A_2^{c}\cdots A_{n-1}^{c}A_n,$$

由 $A_1,A_2,\cdots,A_n$ 独立及定理 2.3.10 知 $A_1^{c},A_2^{c},\cdots,A_{n-1}^{c},A_n$ 也独立，故

$$P(\xi=n)=(1-p)^{n-1}p,\ n=1,2,\cdots.$$

下面我们再来计算某人永远射不中的概率.

用 B 表示事件“永远射不中”，B_n 表示“前 n 次没射中”，则

$$B_n=\bigcap_{k=1}^{n}A_k^c,\quad n=1,2,\cdots.$$

且 $B_1\supset B_2\supset\cdots\supset B_n\supset\cdots$，而

$$B=\bigcap_{n=1}^{\infty}B_n.$$

由概率的上连续性及 $A_1, A_2, \cdots, A_n, \cdots$ 的独立性及定理 2.3.10 易知

$$P(B)=P(\bigcap_{n=1}^{\infty}B_n)=\lim_{n\to\infty}P(B_n)$$

$$=\lim_{n\to\infty}P(\bigcap_{k=1}^{n}A_k^c)$$

$$=\lim_{n\to\infty}(1-p)^n.$$

当 $p=0$ 时 $P(B)=1$，当 $p>0$ 时 $P(B)=0$.

最后让我们用独立和条件概率来解决在概率论历史上非常著名的被帕斯卡和费马讨论过的赌徒输光问题.

例 2.3.14 (赌徒输光问题) 甲乙二人进行赌博，甲胜一局，乙给甲1元，乙胜一局，甲给乙1元. 若每局甲胜的概率为 $p(0<p<1)$，乙胜的概率为 $1-p$，且每局的胜负彼此独立. 开始时甲有 i 元，乙有 $N-i$ 元，他们一直赌到某人输光为止. 试求最终甲赢得全部赌本的斯率. 赌博永远续续下去的概率有多大?

解 令 A_i 表示从甲有 i 元，乙有 $N-i$ 元开始甲赢得全部赌本这一事件. 记 $q=1-p$，

$$P_i=P(A_i),\quad i=0,1,2,\cdots,N-1,N.$$

显然

$$P_0=0\ ;\ P_N=1.$$

用 B 表示"第一局甲胜"，则 $P(B)=p$，$P(B^c)=q$. 利用全概率公

式知，当 $1 \leqslant i \leqslant N-1$ 时有

$$P_i = P(A_i) = P(A_i|B)P(B) + P(A_i|B^c)P(B^c)$$

$$= pP(A_i|B) + q\ P(A_i|B^c)\ .$$

如果已知 B 发生，则第一局赌博结束后，甲有 $i+1$ 元，乙有 $N-(i+1)$元，由于假定了每局胜负是相互独立的，且每局甲胜的概率皆为 p，所以在 B 发生的条件下甲赢得全部赌本的概率正好等于从甲有 $i+1$ 元乙有 $N-(i+1)$元开始赌博最终甲赢得全部赌本的概率，即

$$P(A_i|B) = P(A_{i+1}) = P_{i+1}\ ,$$

同理

$$P(A_i|B^c) = P(A_{i-1}) = P_{i-1}\ .$$

于是得到一组方程

$$P_i = p\ P_{i+1} + q\ P_{i-1}\ ,\quad i = 1, 2, \cdots, N-1\ .$$

利用 $p+q=1$ 得到

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}\ ,\ i = 1, 2, \cdots, N-1$$

或

$$P_{i+1} - P_i = \frac{q}{p}\ (P_i - P_{i-1})\ \ i = 1, 2, \cdots, N-1.$$

利用 $P_0 = 0$，得

$$P_2 - P_1 = \frac{q}{p}\ \ P_1\ ,$$

$$P_3 - P_2 = \left(\frac{q}{p}\right)^2 P_1\ ,$$

$$\cdots\cdots$$

$$P_i - P_{i-1} = \left(\frac{q}{p}\right)^{i-1} P_1,$$

$$\cdots\cdots$$

$$P_N - P_{N-1} = \left(\frac{q}{p}\right)^{N-1} P_1.$$

将前 $i-1$ 个等式相加，即得

$$P_i - P_1 = P_1\left(\frac{q}{p} + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right).$$

$$P_i = \begin{cases} \dfrac{1-\left(\dfrac{q}{p}\right)^i}{1-\dfrac{q}{p}} P_1, & \text{若 } \dfrac{q}{p} \neq 1 \ (\text{即} p \neq \dfrac{1}{2}), \\ iP_1, & \text{若 } \dfrac{q}{p} = 1 \ (\text{即} p = \dfrac{1}{2}). \end{cases}$$

再利用 $P_N = 1$，得

$$P_1 = \begin{cases} \dfrac{1-\dfrac{q}{p}}{1-\left(\dfrac{q}{p}\right)^N}, & \text{若} p \neq \dfrac{1}{2}, \\ \dfrac{1}{N}, & \text{若 } p = \dfrac{1}{2}. \end{cases}$$

从而得到问题的解

$$P_i=\begin{cases}\dfrac{1-\left(\dfrac{q}{p}\right)^i}{1-\left(\dfrac{q}{p}\right)^N}, & \text{若}p\neq\dfrac{1}{2},\\ & i=0,1,2,\cdots,N.\\ \dfrac{i}{N}, & \text{若}p=\dfrac{1}{2},\end{cases}$$

我们再进一步研究，赌博会不会永不休止的继续下去，即两个人都不能赢得全部赌本的概率有多大?

为此我们用 Q_i 表示从甲有 i 元赌本，乙有 $N-i$ 元赌本开始最终乙赢得全部赌本的概率.由对称性易知

$$Q_i=\begin{cases}\dfrac{1-\left(\dfrac{p}{q}\right)^{N-i}}{1-\left(\dfrac{p}{q}\right)^N}, & q\neq\dfrac{1}{2},\\ & i=0,1,2,\cdots,N.\\ \dfrac{N-i}{N}, & q=\dfrac{1}{2},\end{cases}$$

由于 $p=\frac{1}{2}$ 与 $q=\frac{1}{2}$ 等价，故当 $p\neq\frac{1}{2}$ 时

$$P_i+Q_i=\frac{1-\left(\dfrac{q}{p}\right)^i}{1-\left(\dfrac{q}{p}\right)^N}+\frac{1-\left(\dfrac{p}{q}\right)^{N-i}}{1-\left(\dfrac{p}{q}\right)^N}$$

$$=\frac{p^N-qp^{N-i}}{p^N-q^N}-\frac{q^N-p^{N-i}q^i}{p^N-q^N}=1,\quad i=0,1,\cdots,N.$$

而当 $p=q=\frac{1}{2}$ 时仍有 $P_i+Q_i=1, i=0,1,2,\cdots,N$. 这说明甲、乙两人中某人赢得全部赌本的概率为 1，赌博永无休止地进行下去的概率为零.

赌徒输光问题在数学上经常有人提到和介绍的原因可能在于它是一种具有吸收壁的随机游动，这一点将在 §2.6 中看到.

习题 2-3

1. 设三次独立射击中至少命中一次的概率为 0.875，求在一次射击中命中的概率.

2. 设 $A_1, A_2, \cdots, A_n$ 独立且 $P(A_i)=p_i$，$i=1,2,\cdots,n$. 求下列事件的概率:

(1) n 个事件全不发生;

(2) n 个事件至少发生一个;

(3) n 个事件恰好发生一个.

3. 生产零件需要经过三道工序，在第一、二、三道工序中生产出废品的概率分别为 0.02, 0.03, 0.02，假设每道工序生产废品是独立事件. 试求经过三道工序后得到的零件不是废品的概率.

4. 一个人的血型为 O、A、B、AB 型的概率分别为 0.46, 0.40, 0.11, 0.03，任意挑选 5 人，求下列事件的概率:

(1) 两人为 O 型，其他三人分别为其他三种血型;

(2) 三人为 O 型，两人为 A 型;

(3) 没有一人为 AB 型.

5. 证明：若 A, B, C 独立，则 $A \cup B, AB^c$ 都与 C 独立.

§2.4 随机变量及其独立

在讨论德梅尔提出的掷骰子问题时，我们曾用 ξ 表示一颗骰子掷 4 次中 6 点出现的次数. 这里 ξ 的取值是随机的，ξ 取什么值在试验之前不能确定，而是随着试验的结果而定. 例如试验结果为 (6, 3, 1, 2)(即 4 次投掷依次为 6 点，3 点，1 点，2 点)，则 $\xi=1$；试验的结果为 (4, 6, 2, 6)，则 $\xi=2$，……. 也就是说，给定

一个样本点，ξ 就有确定的值.这时“$\xi \geqslant 1$”，“$\xi=2$”，…… 都表示事件，它们的概率都可以计算.

又如讨论会面问题时，用 ξ, η 分别表示甲、乙两人到这的时刻，由于两人到达的时刻是随机的，它们的取值也是随机的，在试验之前不能确定它们取什么值，但试验之后，如果甲、乙两人分别到达的时刻是 x , y, 则 $\xi=x, \eta=y$, 事实上 ξ, η 都是试验结果的函数，即样本空间上的函数，我们可以明确地表示为:

$$\begin{aligned} &\xi(x, y)=x, \\ &\eta(x, y)=y, \end{aligned} \qquad (x, y) \in \Omega .$$

这时“$|\xi-\eta| \leqslant t$”为随机事件，它可用样本点的集合

$$\{(x, y) \in \Omega:|x-y| \leqslant t\}$$

来表示，有着确定的概率.

再如在例 1.3.7 中，将在 R^3 中随机选择方向化作在单位球面上随机选点的问题，并且用 ξ 表示随机方向单位向量在 x 轴上投影的长度，这里的 ξ 也是随机取值的，方向没有确定前它不能确定，当方向一经取定后，这个方向上的单位向量可以用单位球面上的一点 (x, y, z) 来表示，这时 $\xi=|x|$, 也就是说 ξ 是样本空间 $\Omega=\{(x, y, z) \in R^3: x^2+y^2+z^2=1\}$ 上的函数，

$$\xi(x, y, z)=|x|, \quad (x, y, z) \in \Omega .$$

读者不难体会，在例 1.2.1, 例 1.2.3, 例 1.3.1, …… 中都有类似的情况.为了描述这些随机现象，使用这种随机取值的变量是很方便的，它是定义在样本空间 Ω 上的函数，样本点 $\omega \in \Omega$ 一经取定，$\xi(\omega)$ 的值也就确定. 设 B 为一实数集合，用 $\{\xi \in B\}$ 表示使得 $\xi(\omega)$ 的值在 B 中的一切样本点 ω 的集合，即

$$\{\xi \in B\}=\{\omega \in \Omega: \xi(\omega) \in B\}. \tag{2.4.1}$$

这时 $\{\xi \in (a, b]\}$, $\{\xi \leqslant b\}$, $\{\xi=a\}$, …… 都应该是随机事件，可

以赋予概率. 由于事件$\{\xi\in(a,b]\}=\{\xi\leqslant b\}\setminus\{\xi\leqslant a\}$,$\{\xi=a\}=\bigcap_{n=1}^{\infty}\{\xi\in(a-\frac{1}{n},a]\}$等等，在实践中关心的事件$\{\xi\in B\}$，均可利用$\{\xi\leqslant x\}$，$x\in R$经过集合的有限或可数运算生成. 因为$\mathscr{F}$对有限或可数运算封闭，所以我们对于随机变量只要求对一切$x\in R$，$\{\xi\leqslant x\}\in\mathscr{F}$. 下面给出

定义2.4.2 设$(\Omega,\mathscr{F},P)$为概率空间，$\xi(\omega)$为定义在Ω上的实值函数，如果对一切$x\in R$，

$$\{\omega:\xi(\omega)\leqslant x\}\in\mathscr{F},$$

则称ξ为定义在概率空间$(\Omega,\mathscr{F},P)$上的**随机变量**.

我们也可以同时在一个概率空间上定义几个随机变量，如ξ,η,ζ等等.

例2.4.3 从一副52张的扑克牌中任取一张，用ξ表示其点数，η表示其花色($\eta=1,2,3,4$分别表示取得黑桃、红心、梅花、方块)，于是

$$P(\xi=k)=\frac{4}{52}=\frac{1}{13},k=1,2,\cdots,13;$$

$$P(\eta=j)=\frac{13}{52}=\frac{1}{4},j=1,2,3,4;$$

$$P(\xi=k,\eta=j)=\frac{1}{52}=P(\xi=k)P(\eta=j),$$

$$k=1,2,\cdots,13;j=1,2,3,4.$$

从这个例子我们还看到，本来黑桃、红心……并非数量指标，仍然可以使它数量化，使问题的表述简洁明了.

不仅如此，对于任何一个随机事件A的发生与否都可以用随机变量来表示. 例如在掷一枚硬币的试验中，可以用η表示正面出现的次数，即

$$\eta=\begin{cases}1, \text{如“正面出现”},\\ 0, \text{如“正面不出现”}.\end{cases}$$

一般地，有

定义 2.4.4 设$(\Omega,\mathcal{F},P)$为概率空间，$A\in\mathcal{F}$，令

$$I_A(\omega)=\begin{cases}1, \omega\in A,\\ 0, \omega\notin A\end{cases}.$$

则 I_A 称为事件 A 的**示性函数**，表示在一次试验中 A 出现的次数.

显然有

$$\{I_A\leqslant x\}=\begin{cases}\phi, x<0,\\ A^c, 0\leqslant x<1,\\ \Omega, x\geqslant 1.\end{cases}$$

于是对一切 $x\in R$，$\{I_A\leqslant x\}\in\mathcal{F}$，因而随机事件$A$ 的示性函数为$(\Omega,\mathcal{F},P)$上的随机变量.

对于只取有限个值的随机变量来说，只要知道它能取哪些值以及取相应的每一个值的概率，就完全掌握了它的概率规律. 如例 2.4.3 中的ξ,η及定义 2.4.4 中的 I_A(知道 $P(I_A=1)=P(A)$).

在 §2.3 中讨论了事件$A_1,A_2,\cdots,A_n$的独立性. 独立性的概念在概率论中是基本的，它来自实践，又在概率论的理论和应用中有重要的意义. 那么，反映在随机变量上，应该如何刻划独立性呢? 很自然地应该要求，与这些随机变量有关的事件

$$\{\xi_1\in B_1\},\ \{\xi_2\in B_2\},\cdots,\{\xi_n\in B_n\}$$

独立，其中 B_k 为任何开、闭、半开半闭，有限或无限区间. 若某一$B_k=(-\infty,\infty)$，则$\{\xi_k\in B_k\}=\Omega$，这时在集合的求积运算中$\{\xi_k\in B_k\}$可以自动消失，且 $P(\xi_k\in B_k)=1$，于是我们可以给出

定义 2.4.5 若随机变量$\xi_1,\xi_2,\cdots,\xi_n$满足

$$P(\xi_1\in B_1,\xi_2\in B_2,\cdots,\xi_n\in B_n)=P(\xi_1\in B_1)P(\xi_2\in B_2)\cdots P(\xi_n\in B_n),\tag{2.4.6}$$

其中B_k为任意有限、无限的开、闭或半开半闭区间，$k=1,2,\cdots,n$，则称$\xi_1,\xi_2,\cdots,\xi_n$**独立**.

命题 2.4.7 若随机变量$\xi_1,\xi_2,\cdots,\xi_n$独立，则对任何$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$，$\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_k}$也独立.

证 对任何有限、无限的开、闭或半开半闭区间$B_{i_1},B_{i_2},\cdots,B_{i_k}$，取$B_i=(-\infty,\infty)$，$i\neq i_1,i_2,\cdots,i_k$，$1\leqslant i\leqslant n$，此时由于$\{\xi_i\in B_i\}=\Omega$以及$\xi_1,\xi_2,\cdots,\xi_n$独立，我们有

$$\begin{aligned}
&P(\xi_{i_1}\in B_{i_1},\xi_{i_2}\in B_{i_2},\cdots,\xi_{i_k}\in B_{i_k})\\
&=P(\xi_1\in B_1,\xi_2\in B_2,\cdots\xi_n\in B_n)\\
&=P(\xi_1\in B_1)P(\xi_2\in B_2)\cdots P(\xi_n\in B_n)\\
&=\prod_{l=1}^{k}P(\xi_{i_l}\in B_{i_l})\prod_{\substack{i\neq i_l\\ l=1,2,\cdots k}}P(\xi_i\in B_i)\\
&=P(\xi_{i_1}\in B_{i_1})P(\xi_{i_2}\in B_{i_2})\cdots P(\xi_{i_k}\in B_{i_k}).
\end{aligned}$$

这就证明了$\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_k}$独立.

值得再次强调的是：在实践中往往不是先用定义验证$\xi_1,\xi_2,\cdots,\xi_n$独立，而是从问题的实际意义出发认为$\xi_1,\xi_2,\cdots,\xi_n$独立，然后使用定义中的(2.4.6)进行推导和计算.

这里我们以数理统计中的基本概念**样本**为例来说明这一问题.

工农业生产和社会生活都离不开统计，在统计问题中人们经常关心的是某一产品、某一社会现象的数量指标.例如对一个灯泡厂，人们关心它生产的灯泡的寿命，由于种种随机因素的影响，该厂生产的一个灯泡的寿命是一个随机变量ξ，我们称它为**总体**(或母体)，它有一定的取值的概率规律.为了对灯泡的寿命进行推断，需要对ξ进行若干次观察，即取出产品对寿命进行检测，每次观察得到ξ的一个值，若作了n次观察就得到了n个数据，依次记为$x_1,x_2,\cdots,x_n$.如果我们从理论上来进行分析，每次取哪个产

品是随机的，因而每次观察都可以看作是一个随机变量．n 次观察的结果就应 是 n 个随机变量，依次记作 $\xi_1, \xi_2, \cdots, \xi_n$，称为 ξ 的一个大小为 n 的**样本**．而n 次具体的观测值$x_1, x_2, \cdots, x_n$ 只是 $\xi_1, \xi_2, \cdots, \xi_n$ 各自的一次取值．为了对ξ 的取值的概率规律进行推断．应当对样本有一定的要求．首先，每次观察要有代表性，应该能看作是对总体 ξ 的一次观察，这就要 求每一 ξ_k 与ξ 具有同样的取值的概率规律．例如在生产过程中对灯泡寿命进行观察，就要求是在基本生产条件未发生变化的条件下来观察．这样，每一个 ξ_k 都有同ξ 一样的取值的概率规律，都能代表这一工厂灯泡寿命的状况．又如在产品数量较小的情况下，要求抽检又不改变它的性能指标，这时抽检应该是有放回的，否则就破坏了 ξ_k 与总体ξ 有同样的取值规律这一条件．另外，为了数学处理上简单一些，尽量使每次抽检不受其它次抽检结果的影响.这样，就使得在研究数理统计中的许多问题时，常常假定样本 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立的随机变量．

为了考察只取有限个值的随机变量之间的抽立，我们证明下述判定定理．

定理 2.4.8 若$\xi_1, \xi_2, \cdots, \xi_n$是只取有限个值的随机变量，设 ξ_i 可能取的值是$x_{i_1}, x_{i_2}, \cdots, x_{ik_i}$，$i=1, 2, \cdots, n$.则 $\xi_1, \xi_2, \cdots, \xi_n$ 独立的充分必要条件是

$$P(\xi_1 = x_{1j_1}, \xi_2 = x_{2j_2}, \cdots, \xi_n = x_{nj_n})$$

$$= P(\xi_1 = x_{1j_1})P(\xi_2 = x_{2j_2})\cdots P(\xi_n = x_{nj_n}), \tag{2.4.9}$$

$$j_i = 1, 2, \cdots, k_i; \ i = 1, 2, \cdots, n.$$

证 若(2.4.9)成立，则对一切实数集合 $B_1, B_2, \cdots, B_n$ 由命题 1.5.14 有

$$P(\xi_1 \in B_1, \xi_2 \in B_2, \cdots, \xi_n \in B_n)$$

$$= P\Big(\bigcup_{x_{1j_1} \in B_1} \{\xi_1 = x_{1j_1}\}, \bigcup_{x_{2j_2} \in B_2} \{\xi_2 = x_{2j_2}\}, \cdots, \bigcup_{x_{nj_n} \in B_n} \{\xi_n = x_{nj_n}\}\Big)$$

$$=\sum_{x_{1j_1}\in B_1}\sum_{x_{2j_2}\in B_2}\cdots\sum_{x_{nj_n}\in B_n}P(\xi_1=x_{1j_1},\xi_2=x_{2j_2},\cdots,\xi_n=x_{nj_n})$$

$$=\sum_{x_{1j_1}\in B_1}\sum_{x_{2j_2}\in B_2}\cdots\sum_{x_{nj_n}\in B_n}P(\xi_1=x_{1j_1})P(\xi_2=x_{2j_2})\cdots P(\xi_n=x_{nj_n})$$

$$=\sum_{x_{1j_1}\in B_1}P(\xi_1=x_{1j_1})\sum_{x_{2j_2}\in B_2}P(\xi_2=x_{2j_2})\cdots\sum_{x_{nj_n}\in B_n}P(\xi_n=x_{nj_n})$$

$$=P(\xi_1\in B_1)P(\xi_2=B_2)\cdots P(\xi_n\in B_n).$$

因而 $\xi_1,\xi_2,\cdots,\xi_n$ 独立. 反之，若$\xi_1,\xi_2,\cdots,\xi_n$独立，取 $B_1=\{x_{1j_1}\}$, $B_2=\{x_{2j_2}\},\cdots,B_n=\{x_{nj_n}\}$(注意：任意单点集$\{a\}$可以看作闭区间$[a,a]$)，由(2.4.6)知 (2.4.9)成立.

例 2.4.10 在例2.4.3 中的 ξ 和 η 独立.

例 2.4.11 设N 件产品中有 M 件次品，每次独检 n 件而后放回，设抽检不损坏产品. 用 ξ_1,ξ_2,ξ_3 表示第一、二、三次抽得的次品数，求 $P(\xi_1=i,\xi_2=j,\xi_3=k)$.

例 不妨设 $P(\xi_1=i)>0$, $P(\xi_2=j)>0$, 否则所求概率显然为零. 用概率乘法公式，我们有

$$P(\xi_1=i,\xi_2=j,\xi_3=k)=P(\xi_3=k|\xi_1=i,\xi_2=j)$$

$$\cdot\ P(\xi_2=j|\xi_1=i)P(\xi_1=i).$$

由于第一次抽检后把产品放回，无论第一次抽到几件次品都不影响第二次抽检时$\{\xi_2=j\}$发生的概率. 因此

$$P(\xi_2=j|\xi_1=i)=P(\xi_2=j)$$

$$=\binom{M}{j}\binom{N-M}{n-j}\Big/\binom{N}{n}.$$

同理

$$P(\xi_3=k|\xi_1=i,\xi_2=j)=P(\xi_3=k)=\binom{M}{k}\binom{N-M}{n-k}\Big/\binom{N}{n}.$$

于是

$$P(\xi_1=i,\xi_2=j,\xi_3=k)=P(\xi_1=i)P(\xi_2=j)P(\xi_3=k)$$

$$=\frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}\cdot\frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}$$

$$\cdot\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

$$i,j,k=0,1,2,\cdots,n.$$

进一步由定理 2.4.8 还知道 ξ_1，ξ_2，ξ_3 独立.

随机变量及其独立性是概率论的重要课题，以后我们还要进一步讨论. 这里只是建立一点初步概念.

习题 2－4

1. 设随机变量 ξ 表示掷两颗骰子所得点数之和. 指出它能取哪些值以及取各值的概率.

2. 袋中有编号从 1 到 5 的 5 只同样大小的球，从中取出 3 只球，以 ξ 表示取出的球的最小号码. 指出 ξ 的取值及相应的概率.

3. 将一枚均匀硬币掷三次，定义

$$\xi_i=\begin{cases}1, \text{若第 } i \text{ 次出现正面},\\ 0, \text{若第 } i \text{ 次出现反面},\end{cases}\qquad i=1,2,3.$$

证明：ξ_1, ξ_2, ξ_3 独立.

4. 甲、乙两盒中分别装有编号从1到5的5张卡片.现从两盒中各取一张卡片，分别以 ξ 、η 表示所得的编号.令 $\zeta=\xi\eta$.分别求ξ,η,ζ 的取值及相应概率.

5. 从{1,2,3,4}中随机地取出一数 ξ，再从数集{1, 2 ,…,ξ}中随机地取出一数 η .求$P(\xi=i,\eta=j)$及 $P(\xi=i)P(\eta=j)$，$i,j=1,2,3,4$. 问ξ 与 η 是否独立?

§2.5 n 重伯努利试验概型

在例 1.1.8 中我们用古典概率计算了“一颗骰子掷 4 次至少出现 1 次 6 点”的概率. 对于这个问题，也可以这样来考虑：一次试验由 4 个子试验组成，每掷一次骰子，看作一次子试验，在每一次子试验中“出现 6 点”的概率都相同(等于 $\frac{1}{6}$)，而且各次试验是否出现 6 点彼此独立. 如果用 A_i 表示在第i次子试验掷得 6 点，$i=1,2,3,4$,用 ξ 表示 4 次子试验中 6 点出现的次数,则 $P(A_i)=\frac{1}{6}$, $P(A_i^c)=1-\frac{1}{6}$,$i=1,2,3,4$;A_1, A_2, A_3, A_4 独立.

$\{\xi=0\}=A_1^cA_2^cA_3^cA_4^c$, $P(\xi=0)=(1-\frac{1}{6})^4$,

$\{\xi=1\}=A_1A_2^cA_3^cA_4^c\cup A_1^cA_2A_3^cA_4^c\cup A_1^cA_2^cA_3A_4^c\cup A_1^cA_2^cA_3^cA_4$ 是四个两两互不相容的事件的并，每一项的概率都是 $\frac{1}{6}(1-\frac{1}{6})^3$,因此

$$P(\xi=1)=4\times\frac{1}{6}\times(1-\frac{1}{6})^3,$$

$$\{\xi=2\}=A_1A_2A_3^cA_4^c\cup A_1A_2^cA_3A_4^c\cup A_1A_2^cA_3^cA_4\cup A_1^cA_2A_3A_4^c\cup A_1^cA_2A_3^cA_4\cup A_1^cA_2^cA_3A_4$$

是 $\binom{4}{2}=6$ 个两两互不相容的事件的并，每一项的概率为 $\left(\frac{1}{6}\right)^2$

$\left(1-\frac{1}{6}\right)^2$，所以

$$P(\xi=2)=\binom{4}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^2.$$

同理

$$P(\xi=3)=\binom{4}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^1,$$

$$P(\xi=4)=\binom{4}{4}\left(\frac{1}{6}\right)^4.$$

于是所求的概率

$$P(\xi\geqslant 1)=1-P(\xi=0)=1-\left(\frac{5}{6}\right)^4=\frac{6^4-5^4}{6^4}.$$

又如在例 1.2.5 中，用古典概率计算了从含有 M 个红球 $N-M$ 个黑球的袋子中有放回地抽取 n 次，恰有 k 次抽得红球的概率. 现在我们可以使用刚才讨论例 1.1.8 的方法重新讨论这个问题. 将连续抽 n 次球看作一次试验，每次抽一个球看作一次子试验，于是在每次子试验中"抽得红球"的概率皆为 $\frac{M}{N}$，各次子试验中是否抽得红球彼此独立. 令 A_i 表示第 i 次试验抽得红球的事件，则 $P(A_i)=\frac{M}{N}$，$P(A_i^c)=1-\frac{M}{N}$，$i=1,2,\cdots,n$，令 ξ 表示 n 次子试验中红球出现的次数，则

$$P(\xi=k)=P\Big(\bigcup_{\substack{i_1<i_2<\cdots<i_k\\ j_1<j_2<\cdots<j_{n-k}\\ \{i_1,\cdots,i_k,j_1,\cdots,j_{n-k}\}\\ =\{1,2,\cdots,n\}}} A_{i_1}A_{i_2}\cdots A_{i_k}A_{j_1}^cA_{j_2}^c\cdots A_{j_{n-k}}^c\Big)$$

$$=\binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k},k=0,1,2,\cdots,n\ .$$

一般地，如果一个随机试验是由 n 个子试验组成.第 i 次子试验有两个结果：成功(A_i)与失败(A_i^c)，并且 $P(A_i)=p$，$i=1,2,\cdots,n$ (因而 $P(A_i^c)=1-p$)，如果这 n 次子试验还是独立的，即 $A_1,A_2,\cdots,A_n$ 独立，则称这个随机试验属于 n **重伯努利试验概型**(或称n 次独立试验概型)．用ξ 表示 n 次子试验中恰好成功 k 次，则用上面相同的讨论方法得

(2.5.1)　　$$P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,2,\cdots,n\ .$$

一个服从(2.5.1)取值的概率规律的随机变量 ξ，称为服从参数为(n,p)的二项分布的随机变量，记作 $\xi\sim B(n,p)$.

注意：在例 1.1.8 和 1.2.5 中$\{\xi=k\}$的概率均可用古典概率求得，那是因为 p 为有理数(在例 1.1.8 中 $p=\frac{1}{6}$，例 1.2.5 中 $p=\frac{M}{N}$)，一般地当 p 为任何实数时不能由古典概率来计算出 $P(\xi=k)$来．因此不能说所有 n 重伯努利试验概型均可用古典概率来计算.例1.1.8 和例 1.2.5 既能用古典概型又能用 n 重伯努利试验概型来计算,是有其特殊性的.

n 重伯努利试验的概率模型(或二项分布的随机变量)在实践中经常遇到.

现在我们再举几例,说明二项分布的应用.

例 2.5.2　某车间有10 台电动机功率为 10 千瓦的机床,每台机床是否工作是独立的且工作的概率为 $\frac{1}{5}$，若供电部门只提供 50 千瓦的电力给这个车间，试求该车间不能正常工作的概率.

解　用 ξ 表示需要开动的机床数．如果需要开动的机床有 6 台

或6台以上，则用电超过50千瓦，该车间就不能正常工作. 所以{ξ≥6}表示事件"该车间不能正常工作". 由于每台机床开动的概率为 $\frac{1}{5}$，各台机床开动与否彼次独立. 因此可看作$p=\frac{1}{5}$ 的10重伯努利试验概型，即 $\xi \sim B(10, \frac{1}{5})$，所以

$$P(\xi=k)=\binom{10}{k}\left(\frac{1}{5}\right)^k\left(1-\frac{1}{5}\right)^{10-k}, k=0,1,2,\cdots,10.$$

由于当 $k\neq j$ 时{ξ= k}与{ξ= j}互不相容，因此所求概率

$$P(\xi\geqslant 6)=\sum_{k=6}^{10}P(\xi=k)=\sum_{k=6}^{10}\binom{10}{k}\left(\frac{1}{5}\right)^k\left(1-\frac{1}{5}\right)^{10-k}$$

$$=1-\sum_{k=0}^{5}\binom{10}{k}\left(\frac{1}{5}\right)^k\left(1-\frac{1}{5}\right)^{10-k}.$$

经计算求得

$$\sum_{k=0}^{5}\binom{10}{k}\left(\frac{1}{5}\right)^k\left(1-\frac{1}{5}\right)^{10-k}=0.9936.$$

故所求概率

$$P(\xi\geqslant 6)=1-0.9936=0.0064.$$

即10台机床有6台或6台以上需要开动，因而由于电力供应不足车间不能正常生产的概率为0.0064，这是一个小概率事件，如果不具有破坏性而且不是重要产品的生产车间，则对生产影响不大，在整个电力供应不足的情况下可以这样安排.

例 2.5.3 某电话交换台有300 个用户，在任何时刻各用户是否需要通话是相互独立的，且每个用户需要通话的概率为 $\frac{1}{60}$. 设该交换台只有 8 条线路供用户同时使用，试求在任一给定时刻用户打不通电话的概率.

解 将每个用户需要通话与否看成一次子试验，因此用户需要使用电话可以看成是$p=\frac{1}{60}$ 的 300 重伯努利试验概型. 用 ξ 表示某给定时刻需要通话的用户数，如果 $\xi>8$，则用户打不通电话，因此由 (2.5.1) 知所求概率

$$P(\xi>8)=1-\sum_{k=0}^{8}\binom{300}{k}\left(\frac{1}{60}\right)^{k}\left(\frac{59}{60}\right)^{300-k}.$$

上式右端的数值直接计算很困难，计算量很大. 即便用查表的方法，一般来说当 n 较大时 (例如 $n>30$) 在二项分布表中就查不到了. 根据下面的定理，我们可以得到近似算法.

定理 2.5.4 设 λ 是一正常数，$p_n=\frac{\lambda}{n}$，则有

$$\lim_{n\to\infty}\binom{n}{k}p_n^{k}(1-p_n)^{n-k}=e^{-\lambda}\frac{\lambda^{k}}{k!},\quad k=0,1,2,\cdots.$$

证

$$\binom{n}{k}p_n^{k}(1-p_n)^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$=\frac{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!}\lambda^{k}\left[\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}\right]^{\lambda}\left(1-\frac{\lambda}{n}\right)^{-k}.$$

由于对任何指定的k，

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1;$$

$$\lim_{n\to\infty}\left(1-\frac{j}{n}\right)=1,\quad 0\leqslant j\leqslant k-1;$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}=e^{-1},$$

所以

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=e^{-\lambda}\frac{\lambda^k}{k!},\ k=0,1,2,\cdots.$$

这个定理告诉我们：当 n 比较大、p 比较小时可用 $\frac{\lambda^k}{k!}e^{-\lambda}$ 作为 $\binom{n}{k}p^k(1-p)^{n-k}$ 的近似值，其中 $\lambda=np$. 由于当 k 固定时，$\frac{\lambda^k}{k!}e^{-\lambda}$ 只有一个参数 λ，表的制作比较容易；而 $\binom{n}{k}p^k(1-p)^{n-k}$ 有两个参数 n,p，表的制作就很困难(即使可以，也会占很大的篇幅).

定义 2.5.5 若ξ是取非负整值的随机变量且

$$P(\xi=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ k=0,1,2,\cdots,$$

则称 ξ 为服从参数为 λ 的泊松(Poisson)分布的随机变量，记作 $\xi\sim\mathscr{P}(\lambda)$.

书末附有泊松分布表(附表 1).

例 2.5.3 中 $n=300$，$p=\frac{1}{60}$，因而 $\lambda=5$，查泊松分布表得

$$\frac{5^0}{0!}e^{-5}=0.006738,\quad \frac{5^5}{5!}e^{-5}=0.175467,$$

$$\frac{5^1}{1!}e^{-5}=0.033690,\quad \frac{5^6}{6!}e^{-5}=0.146223,$$

$$\frac{5^2}{2!}e^{-5}=0.084224,\quad \frac{5^7}{7!}e^{-5}=0.104445,$$

$$\frac{5^3}{3!}e^{-5}=0.140374,\quad \frac{5^8}{8!}e^{-5}=0.065278.$$

$$\frac{5^4}{4!}e^{-5}=0.175467,$$

于是

$$P(\xi>8)=1-\sum_{k=0}^{8}\binom{300}{k}\left(\frac{1}{60}\right)^k\left(1-\frac{1}{60}\right)^{300-k}$$

$$\approx 1-\sum_{k=0}^{8}\frac{5^k}{k!}e^{-5}\approx 1-0.931906$$

$$\approx 0.068.$$

这就是说对于 300 个用户，只有 8 条线路可以同时通话，在通常情况下够用了，用户需要通话而全部占线的概率很小.

我们注意到在泊松分布表(附表 1)中对于很大的参数 λ 相应的 $\frac{\lambda^k}{k!}e^{-\lambda}$ 的值也未列出.因而当 n 很大、p 不很小、np 较大时查泊松分布表也无用.但实践中有时又需要求这样的值，如

例 2.5.6 抛掷100 个硬币，求恰好出现 50 个正面朝上的概率.

解 利用二项分布，立刻可以得到所求概率为

$$\binom{100}{50}\left(\frac{1}{2}\right)^{100}=\frac{100!}{50!\,50!}\cdot\frac{1}{2^{100}}\,. \tag{2.5.7}$$

但这种表达式不能令人满意，我们不知道这一概率的数值大小，它究竟是接近 $\frac{1}{2}$，$\frac{1}{10}$ 还是 $\frac{1}{50}$？问题在于计算 $n!$ 这件事表面上很简单，但当 n 很大时算出它的值并非易事. 为计算(2.5.7)，我们用数学分析中下述著名的公式:

定理 2.5.8 (斯特林(Stirling)公式):

$$n!=n^ne^{-n}\sqrt{2\pi n}\;e^{\theta_n}\,,$$

其中

$$\frac{1}{12(n+1)}<\theta_n<\frac{1}{12n}\,.$$

这一公式的证明可参看[2, II .9] .

利用斯特林公式可以得到

$$\binom{2n}{n}\frac{1}{2^{2n}}=\frac{(2n)!}{n!\,n!}\,\frac{1}{2^{2n}}\sim\frac{(2n)^{2n}e^{-2n}\sqrt{4\pi n}}{(n^ne^{-n})^2 2\pi n}\,\frac{1}{2^{2n}}$$

$$=\frac{1}{\sqrt{\pi n}}\,,$$

当 $n=50$ 时，$\frac{1}{\sqrt{\pi n}}\approx 0.08$. 因此例 2.5.6 的答案是 0.08. 即抛掷 100 枚硬币恰有 50 枚正面朝上的概率近似于 0.08.

一般地，利用斯特林公式还可以证明下述逼近二项分布的定

理.

定理 2.5.9 若$0<p<1, q=1-p$,则对任意给定的实数$a<b$,当k满足$\frac{k-np}{\sqrt{npq}} \in (a, b)$时,下述极限一致地成立:

$$\lim_{n\to\infty} \frac{\binom{n}{k} p^k (1-p)^{n-k}}{(2\pi npq)^{-\frac{1}{2}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}} = 1 . \tag{2.5.10}$$

这称之为局部极限定理.

再有,若$\xi_n \sim B(n, p)$,则对任何有限实数$a, b, a \leqslant b$,或$a=-\infty$,或$b=+\infty$有

$$\lim_{n\to\infty} P\left(a < \frac{\xi_n - np}{\sqrt{npq}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx . \tag{2.5.11}$$

这称之为积分极限定理.

在此,我们不给出这一定理的证明,有兴趣的读者可参看[5,第二章§12, §13].

当n很大$|k-np|$相对于$\sqrt{npq}$不很大时,我们可以用$\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}$来近似$\binom{n}{k} p^k q^{n-k}$,虽然前一表达式看起来复杂,但人们并不难用计算器迅速地算出它的值来.

例 2.5.12 设色盲者在全部人口中所占的比例为$p=0.005$,今抽检10000个人,求其中恰有40人为色盲者的概率.

解 本题可以看成是$p=0.005, n=10000$次的独立试验概型,由(2.5.1)及(2.5.10)($k=40$)知

$$\binom{10000}{40}(0.005)^{40}(1-0.005)^{9960} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2}.$$

现在

$$\sqrt{npq} = \sqrt{10000 \times 0.005 \times 0.995} = \sqrt{49.75} \approx 7.053,$$

$$\frac{k-np}{\sqrt{npq}} \approx -1.418.$$

因此所求概率近似于

$$\frac{1}{7.053\sqrt{2\pi}} e^{-\frac{1.418^2}{2}} \approx 0.0207.$$

例 2.5.13 若事件 A 的概率 $P(A)=p$，$0<p<1$，设 ξ_n 表示在 n 次独立试验中事件 A 出现的频数，则 $\frac{\xi_n}{n}$ 表示 n 次试验中事件 A 出现的频率，于是对任意 $\varepsilon>0$，

$$P\left(\left|\frac{\xi_n}{n}-p\right|<\varepsilon\right)=P\left(-\varepsilon\sqrt{\frac{n}{pq}}<\frac{\xi_n-np}{\sqrt{npq}}<\varepsilon\sqrt{\frac{n}{pq}}\right)$$

当 n 充分大时，对任意给定的正数 A，上述概率大于

$$P\left(-A<\frac{\xi_n-np}{\sqrt{npq}}<A\right).$$

因此

$$\lim_{n\to\infty} P\left(\left|\frac{\xi_n}{n}-p\right|<\varepsilon\right) \geqslant \lim_{A\to\infty}\lim_{n\to\infty} P\left(-A<\frac{\xi_n-np}{\sqrt{npq}}<A\right)$$

$$=\lim_{A\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-A}^{A} e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx.$$

可以证明上式右端等于1(参看§3.3)，再由

$$\overline{\lim_{n\to\infty}} P\left(\left|\frac{\xi_n}{n}-p\right|<\varepsilon\right)\leqslant 1,$$

即可得：对任意 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{\xi_n}{n}-p\right|<\varepsilon\right)=1. \tag{2.5.14}$$

这就是著名的伯努利大数定律．它说明了频率的稳定性，关于大数定律，以后还有更一般的结论，这里只是作为积分极限定理应用的例子．

下面我们讨论事件概率的最大似然估计．

例2.5.15 (废品率的最大似然估计) 把某车间在确定生产条件下生产的产品看作一个总体，为估计这种产品的废品率p，依次抽取n件产品进行检检，抽取产品时尽量注意使各次抽取无关．设

$$\xi_i=\begin{cases}1, \text{若第 } i \text{ 次抽得的是废品},\\ 0, \text{若第 } i \text{ 次抽得的是合格品}.\end{cases}$$

则在未具体检例之前，可以认为$\xi_1,\xi_2,\cdots,\xi_n$是n个独立的随机变量，设具体抽取了n件产品检测的结果得到$\xi_1,\xi_2,\cdots,\xi_n$的一组观测值$x_1,x_2,\cdots,x_n$ ($x_i=0$或1)．我们用这些数据对该产品的废品率进行估计．

与例1.2.4讨论捕鱼问题时估计池中鱼数的想法类似，我们希望求得p的估计量$\hat{p}$(只与$x_1,\cdots,x_n$有关)，使$P(\xi_1=x_1,\xi_2=x_2,\cdots,\xi_n=x_n)$取得最大值(称为$p$的最大似然估计)．由于

$$P(\xi_i=x_i)=p^{x_i}(1-p)^{1-x_i},\ x_i=0,1;\ i=1,2,\cdots,n,$$

再由$\xi_1,\xi_2,\cdots,\xi_n$的独立性知

$$P(\xi_1=x_1,\xi_2=x_2,\cdots,\xi_n=x_n)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}$$

若 $\sum_{i=1}^{n} x_i = 0$，则 $P(\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n) = (1-p)^n$，因而

$$\hat{p} = 0;$$

若 $\sum_{i=1}^{n} x_i = n$，则 $P(\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n) = p^n$，因而 $\hat{p} = 1$.

现在考虑 $0 < \sum_{i=1}^{n} x_i < n$ 的情形，此时当 $p = 0$ 或 1 时，$P(\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n) = 0$. 所以 $\hat{p}$ 如果存在，一定有 $0 < \hat{p} < 1$. 因此假设 $0 < p < 1$，令

$$L(x_1, x_2, \cdots, x_n; p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i},$$

为求得 p 的最大似然估计，令

$$\frac{\partial L(x_1, x_2, \cdots, x_n; p)}{\partial p} = 0$$

即

$$\sum_{i=1}^{n} x_i \frac{1}{p} + (n - \sum_{i=1}^{n} x_i) \frac{-1}{1-p} = 0.$$

解得

$$p = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

且当 $p < \frac{1}{n} \sum_{i=1}^{n} x_i$ 时 $\frac{\partial L(x_1, x_2, \cdots, x_n; p)}{\partial p} > 0$，当 $p > \frac{1}{n} \sum_{i=1}^{n} x_i$ 时 $\frac{\partial L(x_1, x_2, \cdots, x_n; p)}{\partial p} < 0$，因而当 $p = \frac{1}{n} \sum_{i=1}^{n} x_i$ 时 $L(x_1, x_2, \cdots x_n; p)$ 取得极大值，再由前面说明，此时 $0 < \hat{p} < 1$. 故 $p = \frac{1}{n} \sum_{i=1}^{n} x_i$

时，$L(x_1, x_2, \cdots, x_n; p)$取最大值. 总结上面讨论，在一切情形下，都有

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

当检验的产品没有取定时，ξ_i也没有取定，它是随机变量，因此p的估计量也应为随机变量，即

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \xi_i .$$

$\sum_{i=1}^{n} \xi_i$是抽取 n 件产品中所含废品的个数，$\hat{p}$恰是抽取的n 件产品的废品率(频率). 这就是用随机抽取的 n 件产品的废品率去估计整个车间产品的废品率，是合乎情理的.

与讨论超几何分布时废品数的检验(例 1.2.3)问题类似，对于二项分布也可以讨论检验问题.

例 2.5.16 (关于二项分布参数p 的检验问题) 我们先从一个极其明显的例子谈起：设某厂家声称该厂生产的某种电子元件废品率p 不超过 0.01，今从中抽取了 5 件，废品数为 2，自然会从直觉上就怀疑厂家讲的话是否属实. 因为若废品率为 0.01，任取 5 件产品，废品数超过 1 的概率为

$$\sum_{k=2}^{5} \binom{5}{k} 0.01^k 0.99^{5-k} = 10 \times 0.01^2 \times 0.99^3 + 10 \times 0.01^3 \times 0.99^2 + 5 \times 0.01^4 \times 0.99 + 0.01^5 \leqslant 0.001 ,$$

即当$p = 0.01$ 时，5 件随机抽取(有放回)的产品中含有多于 1 件废品的概率小于千分之一. 由于对任何 $n(\geqslant 2)$及p，

$$\frac{d}{dp}\left(\sum_{k=0}^{1} \binom{n}{k} p^k (1-p)^{n-k}\right) = -n(n-1)(1-p)^{n-2} \leqslant 0 ,$$

因而当$p < 0.01$ 时，5 件随机抽取的产品中含有多于 1 件废品的

概率会更小. 这样的小概率事件是不易发生的，所以我们有理由不相信厂家的说法.那么我们一般应该怎样制定一个检验方案呢?

如果我们要通过抽检 n 个产品来检验原假设

$$H_0: p \leqslant p_0$$

是否成立.令相对立的假设(称为备择假设)为

$$H_1: p > p_0 .$$

给定 α (通常取作0.05 或 0.01),称为显著性水平，可以从下面看出它就是当原假设 $H_0: p \leqslant p_0$ 真时被否定的概率(犯第一类错误的概率)的上确界. 求出使

$$\sum_{k=K_\alpha}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \leqslant \alpha ,$$

$$\sum_{k=K_\alpha-1}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} > \alpha ,$$

$$K_\alpha \geqslant np$$

成立的 K_α，则由于当 $k \geqslant np$ 时 $p^k(1-p)^{n-k}$ 是 p 的增函数，因而对一切 $0 \leqslant p < p_0$，

$$\sum_{k=K_\alpha}^{n} \binom{n}{k} p^k (1-p)^{n-k} \leqslant \alpha .$$

令 ξ_n 表示 n 件产品中所含的次品数，若 $p \leqslant p_0$(即原假设 $H_0: p \leqslant p_0$ 为真)，则

$$P(\xi_n \geqslant K_\alpha) = \sum_{k=K_\alpha}^{n} \binom{n}{k} p^k (1-p)^{n-k} \leqslant \alpha .$$

随机地选取 n 件产品，若次品数 $\geqslant K_\alpha$ (此情形发生的概率 $\leqslant \alpha$)，

则否定假设 H_0，接受H_1；如果次品数$<K_\alpha$，则接受原假设，即没有理由认为原假设不成立. 称

$$\{K_\alpha, K_\alpha+1, \cdots, n\}$$

为此检验的判别区域或拒绝区域.

例 2.5.17 从一批电子元件中，有放回地抽取20 件产品进行检验，制定一个显著性水平 $\alpha=0.05$ 的检验方案，检验次品率 p 是否不超过 0.1.

解 设原假设

$$H_0: p \leqslant 0.1,$$

备择假设

$$H_1: p > 0.1.$$

取 $n=20, p_0=0.1$，经计算求得使

$$\sum_{k=K_\alpha}^{20} \binom{20}{k} 0.1^k 0.9^{20-k} \leqslant 0.05$$

的最小的 $K_\alpha=4$，再验证 $4>20\times0.1$，因而得到拒绝区域为

$$\{4, 5, \cdots, 20\}.$$

于是当抽出 20 件产品后，发现次品数 $\geqslant 4$ 时就拒绝假设 $p \leqslant 0.1$，认为次品率超过 0.1，此时犯错误的概率不超过 0.05. 若次品数少于 4 件，则没有理由否定原假设，认为次品率不超过 0.1.

最后，我们再讨论二项分布在统计上的另外一个应用——符号检验法的原理.

例 2.5.18 (符号检验法) 为检验某车间甲、乙两台机床生产的某种产品的一种数量指标有无差异，每小时从甲、乙两台机床各抽取一件产品进行观测，用 x_i、y_i 分别表示第 i 次观测甲、乙两台机床生产的产品的该数量指标. 如果甲、乙两台机床生产的产品的该指标无差异，那么 $x_i>y_i$ 和 $x_i<y_i$ 应是机会均等的，即在

$x_i \neq y_i$ 的条件下两者出现的机会各为 1/2. 假定在若干次抽取中除去相等的情形外，有 n 次 $x_i \neq y_i$，那么 $x_i < y_i$ 发生的次数应服从二项分布 $B(n, \frac{1}{2})$. 我们的问题变成了检验 $p = \frac{1}{2}$ 是否成立的问题. 给定显著性水平 α，由于当 $x_i < y_i$ 的次数 ξ_n 服从 $B(n, \frac{1}{2})$ 时，若求得使

$$\sum_{i=0}^{K} \binom{n}{i} \left(\frac{1}{2}\right)^n \leqslant \frac{\alpha}{2}$$

的最大的 K，令其为 K_α，则由于

$$\sum_{i=n-K_\alpha}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \sum_{i=0}^{K_\alpha} \binom{n}{i} \left(\frac{1}{2}\right)^n \leqslant \frac{\alpha}{2},$$

我们有

$$P\left(\left|\xi_n - \frac{n}{2}\right| \geqslant \frac{n}{2} - K_\alpha\right)$$

$$= P(\xi_n \leqslant K_\alpha) + P(\xi_n \geqslant n - K_\alpha) \leqslant \alpha.$$

因此，若在 n 次 $x_i \neq y_i$ 的观测中，$x_i < y_i$ 的次数

$$K \in \{0, 1, \cdots, K_\alpha\} \bigcup \{n - K_\alpha, n - K_\alpha + 1, \cdots n\},$$

则认为这两台机床生产的产品的这项指标有差异，若

$$K \in \{K_\alpha + 1, K_\alpha + 2, \cdots, n - K_\alpha - 1\},$$

则不认为有差异.

在这个例子中，拒绝区域是双边的，原因是在 n 次观测中 "$x_i < y_i$" 发生的次数过少或过多都说明两台机床生产的产品的该项指标有差异，我们要检验 $p = \frac{1}{2}$ 是否成立. 这与例2.5.16 不同，

称之为双边检验，例 2.5.16 称之为单边检验.

这里所讨论的问题本来是二项分布的检验问题，但由于在统计上应用较多，特别根据上面的讨论，制定出一套称之为符号检验的方案:

检验目的:A、B 两种处理对产品的一个指标的效应是否相同.

检验手续:

(1) 对 A、B 两种处理下产品的这个数量指标进行 N 次观察，得数据对 (x_i, y_i)，$i=1,2,\cdots,N$;

(2) 对 $i=1,2,\cdots,N$，若 $x_i>y_i$，记作“+”，若 $x_i<y_i$，记以“−”，若 $x_i=y_i$，记为“0”;

(3) 用 n_+ 和 n_- 分别表示 N 次观察中“+”和“−”的个数，且记 $n=n_++n_-$，$s=\min(n_+,n_-)$;

(4) 给定显著性水平 α，令 s_α 为满足 $\sum_{i=0}^{k}\binom{n}{i}\left(\frac{1}{2}\right)^n\leqslant\frac{\alpha}{2}$ 的最大的 k;

(5) 若 $s\leqslant s_\alpha$，则否定假设，认为 A，B 两种处理的效应不同，若 $s>s_\alpha$，则不否定假设.

书末列有专门的符号检验表(附表 6).在此只是介绍了符号检验法的原理.

习题 2−5

1. 某射手射击 5 次，每次命中的概率为 0.6，求下列事件的概率:

(1) 5 次中有 3 次中靶;

(2) 5 次中至少有 3 次中靶.

2. 一台设备由三个独立工作的基本元件组成，每个元件在时间 t 以前发生故障的概率为0.1，如果至少有两个元件发生故障，则整个设备也发生故障.求该设备在时间 t 以前不发生故障的概率.

3. 教科书出版的印数是 10000 本，每本装钉错误的概率等于 0.0001.求这批书中(1)正好 5 本有装钉错误; (2)至多 5 本有装钉错误的概率.

4. 在二项分布中记

$$b(k;n,p)=\binom{n}{k}p^k(1-p)^{n-k},k=0,1,2,\cdots,n,$$

求 k 使 $b(k;n,p)$ 达到最大值.

5. 为测定能降低血液中胆固醇含量的某种食品的有效性，让 100 个人吃这种食品. 经过一定时间，化验他们的胆固醇含量. 如果至少有 65 人吃了这种食品以后胆固醇含量降低，则营养学家就决定承认这种食品，如果这种食品事实上对降低胆固醇不起作用，试问营养学家承认它的概率是多少?

6. 对甲、乙两台车床生产的螺丝钉的直径同时进行抽样检查，测得 10 组数据如下(单位：厘米)：

x_i：2.01，2.06，2，2.04，1.95，1.94，1.97，2.03，2.01，1.98.

y_i：2.02，2.04，1.98，2.01，1.94，2，2.01，2.05，1.99，1.97.

给定显著性水平 $\alpha=0.05$，问是否可以认为两部车床生产的螺丝钉直径没有差异?

** §2.6 简单随机游动与有限马尔可夫链

一个质点沿着直线逐步移动，每一步它向上或向下移动一个单位距离，不受其它步如何移动的影响，而向上和向下移动的概率分别为 p 和 $q=1-p$. 假定 $0<p<1$. 若用 X_n 表示时刻 n 时质点所在的位置，于是质点就在 $(-\infty,\infty)$ 上整数格子点 Z 上游动，向上和向下，无休止地继续下去. 若用横坐标表示步数(或时刻)，纵坐标表示质点所在位置，(n,j_n) 表示 n 时刻质点处于位置 j_n，将点 $(0,j_0)$，$(1,j_1)$，…… 顺次用直线连接起来，就得到质点运动的一条轨道，如图 2－5.

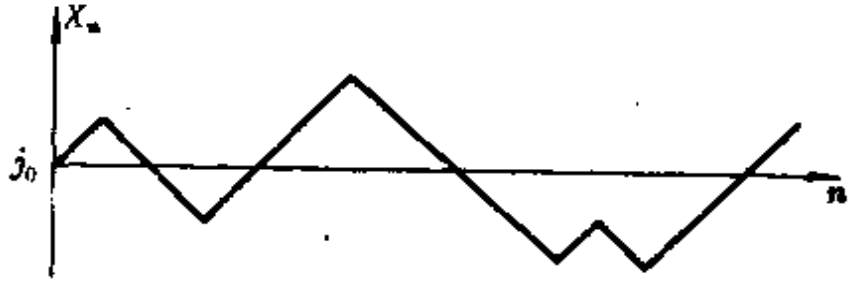


图 2－5

若 0 时刻质点位于 X_0 (可以是固定值，也可以是随机变量)，用ξ_n 表示第 n 步的位移，则 ξ_n 是取 $+1$，-1 为值的随机变量，且

$$P(\xi_n=1)=p\ ,P(\xi_n=-1)=q\ .$$

由于每一步向上、向下的概率不受其它任何步向上、向下的影响，因此 $X_0\ ,\xi_1\ ,\xi_2\ ,\cdots,\xi_n$ 独立．令

(2.6.1)　　$$X_n=X_0+\xi_1+\xi_2+\cdots+\xi_n\ ,n=0,1,2,\cdots$$

(这就是前述的 j_n)．我们称$\{X_n:n=0,1,2,\cdots\}$(或记作$\{X_n:n\in Z_+\}$)是一个**简单随机游动**，它是一个离散时间随机过程．我们不是把它看作一个个孤立的随机变量来研究，而是做为一个整体，研究它的整个进程．例如:我们不只是关心 $P(X_n\in[a,b])$有多大，而是关心类似于下面的一系列整体性质:

(1) 从 $(0,N)$ 中某一点 j 出发，质点在到达N 之前到达"0"的概率有多大? (请读者注意，如果将 ξ_n 认为是第 n 次赌博中甲赢的钱数,这就是赌徒输光问题，此时 $X_0=j$.)

(2) 从 $(0,N)$ 中某一点 j 出发，是否在有限时间内一定碰上"0"或"N"(在赌徒输光问题中已回答了这个问题，赌博不会无休止继续下去)?

(3) 质点从 $(0,N)$ 中某一点 j 出发，首次到达"0"或 N 的平均时间有多长 (即赌博持续时间问题，它的解决要用到下一章的随机变量的期望的概念)?

(4) 质点从 $j\geqslant 0$ 出发，迟早到达"0"的概率有多大? (有 j 元赌本的人与一个无限富有的人赌博，输光的概率有多大? 这可以通过在赌徒输光问题的解答中令 $N\to\infty$ 得到答案: 当 $p\leqslant q$ 时输光的概率为1，当 $p>q$ 时输光的概率为 $\left(\frac{q}{p}\right)^j$.) .

关于随机游动，有F.Spitzer 的专著[11]，我们这里只是介绍一

些极其简单的概念．由于篇幅的限制，不作深入讨论．

下面再简单介绍一下有限马尔可夫链．

在赌徒输光问题中，用 X_n 表示时刻 n 甲所有的钱数，已知 $X_n=i$ 的条件下，X_{n+1} 取值的概率，不受 n 时刻以前取值的影响，即

$$P(X_{n+1}=i+1 \mid X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=i)$$

$$=p, \quad 0<i<N.$$

$$P(X_{n+1}=i-1 \mid X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=i)$$

$$=q, \quad 0<i<N.$$

$$P(X_{n+1}=0 \mid X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=0)=1,$$

$$P(X_{n+1}=N \mid X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=N)=1.$$

系统的这种性质称为马尔可夫性或无后效性．直观上说，就是已知现在的条件下，将来与过去无关(或独立)．

设随机变量序列 $X_0, X_1, X_2, \cdots$ 只取 $\{0, 1, 2, \cdots, M\}$ 中的值，"$X_n=i$"表示系统在 n 时刻处于状态 i．如果存在 p_{ij}, $i, j \in \{0, 1, \cdots, M\}$，使得对一切的 n，一切的 $i_k \in \{0, 1, \cdots, M\}$，$k=0, 1, \cdots, n-1$，有

$$(2.6.2) \quad P(X_{n+1}=j \mid X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=i)$$

$$=p_{ij},$$

则称 $\{X_n, n \in Z_+\}$ 为**一齐次马尔可夫链**，称 $\{p_{ij}, i, j=0, 1, \cdots, M\}$ 为此马尔可夫链的**转移概率**．转移概率应满足

$$p_{ij} \geqslant 0, \quad i, j \in \{0, 1, \cdots, M\},$$

$$\sum_{j=0}^{M} p_{ij}=1, \quad i \in \{0, 1, \cdots, M\}.$$

为了方便，常把转移概率排成一个方阵

$$P=\begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0M} \\ p_{10} & p_{11} & \cdots & p_{1M} \\ \cdots & & & \\ p_{M0} & p_{M1} & \cdots & p_{MM} \end{bmatrix} \tag{2.6.3}$$

称为**转移概率矩阵**.

简单随机游动是马尔可夫链的特例. 设 X_n 表示 n 时刻赌徒甲所具有的钱数，由于当 $X_n=i$，$0<i<N$ 时 X_{n+1} 取 $i+1$ 为值的概率为 p，因而此时 $p_{i,i+1}=p$，而 X_{n+1} 取 $i-1$ 为值的概率为 q，因而 $p_{i,i-1}=q$. X_{n+1} 取其它值的概率为"0"，故 $p_{ij}=0, j\neq i+1, j\neq i-1$. 而当 $X_n=0$ 时，赌博停止，甲永远只有"0"元钱. 而当 $X_n=N$ 时，赌博也停止，甲永远拥有 N 元钱，此时 $X_{n+1}=X_{n+2}=\cdots=N$. 因而 $p_{00}=1, p_{NN}=1$. 根据这些分析，赌徒输光问题的转移概率矩阵为

$$P_1=\begin{array}{c} \begin{matrix} 0 & 1 & 2 & \cdots & N-1 & N \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & p & 0\cdots & 0 & 0 \\ 0 & q & 0 & p\cdots & 0 & 0 \\ \cdots & & & & & \\ 0 & 0 & 0 & 0\cdots q & 0 & p \\ 0 & 0 & 0 & 0\cdots 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{matrix} \tag{2.6.4}$$

这个**转移概率矩阵称为有吸收壁的随机游动**，即当 X_n 一旦到达 0 状态或 N 状态，则在此时刻以后永远停留在相应的状态.

设转移概率矩阵为

$$(2.6.5)\qquad P_2=\begin{array}{c}\begin{matrix}0 & 1 & 2 & 3 & \cdots\cdots & N\end{matrix}\\ \begin{bmatrix}0 & 1 & 0 & 0 & \cdots\cdots & 0\\ q & 0 & p & 0 & \cdots\cdots & 0\\ 0 & q & 0 & p & 0\ \cdots & 0\\ \cdots\cdots & & & & & \\ \cdots\cdots & & & & q\ 0 & p\\ 0 & 0 & 0 & 0\cdots 0\ 1 & 0\end{bmatrix}\end{array}\begin{matrix}0\\1\\2\\\vdots\\\\N\end{matrix}$$

这表示一个具有反射壁的随机游动，当质点走到 $[0, N]$ 的端点时，下一步一定回头走. 在赌徒输光问题中，如果不是有一赌徒输光就停止，而是当一个赌徒输光时，另一赌徒给他一元钱，使得赌博可以永远继续下去. 那么这个模型中赌徒甲(或乙)的钱数，就是一个具有反射壁的随机游动.

对于有限马尔可夫链，如果已知转移概率矩阵

$$P=(p_{ij})_{i,j=0}^{N},$$

可以证明 n 步转移概率矩阵

$$(2.6.6)\qquad (p_{ij}^{(n)})_{i,j=0}^{N}\triangleq(P(X_n=j|X_0=i))_{i,j=0}^{N}=P^n.$$

事实上，若 $n=2$，由全概率公式知

$$p_{ij}^{(2)}=P(X_2=j|X_0=i)=\sum_{k=0}^{N}P(X_2=j,X_1=k|X_0=i)$$

$$=\sum_{k=0}^{N}P(X_2=j|X_0=i,X_1=k)P(x_1=k|X_0=i)$$

$$=\sum_{k=0}^{N}P(X_2=j|X_1=k)P(X_1=k|X_0=i)$$

$$=\sum_{k=0}^{N}p_{ik}p_{kj},0\leqslant i,j\leqslant N.$$

即 $p_{ij}^{(2)}$ 为矩阵 $P\cdot P(=P^2)$ 中第 i 行第 j 列的元素. 对任何 $n\geqslant 0$，

$m \geqslant 0$，用归纳法可证：

$$(2.6.7) \qquad p_{ij}^{(m+n)} = \sum_{k=0}^{N} p_{ik}^{(m)} p_{kj}^{(n)} \ , 0 \leqslant i \ , j \leqslant N \ .$$

因此(2.6.6)成立. (2.6.7)称之为 K－C 方程.

最后我们再举一个例子来结束关于有限马尔可夫链的讨论.

例 2.6.8 艾伦费斯特(Ehrenfest)模型 一个坛子装有 N 个或者是红的或者是黑的球. 随机地从中摸出一球换以另一种颜色的球(如取出的球为黑色，换成红球再放入坛中……). 用 X_n 表示 n 次抽取后坛中的黑球数. 如果已知 $X_n = i$，则下一时刻 X_{n+1} 取什么值与 n 时刻以前的 $X_k (k < n)$ 取什么值无关，且

$$P(X_{n+1} = i+1 | X_n = i) = \frac{N-i}{N} \ ,$$

$$0 \leqslant i \leqslant N \ .$$

$$P\ (X_{n+1} = i-1 | X_n = i) = \frac{i}{N} \ ,$$

其转移概率矩阵为

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots\cdots & 0 & 0 & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & \cdots\cdots & 0 & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & 0 \cdots & 0 & 0 & 0 \\ & \cdots\cdots\cdots & & & & & & \\ 0 & 0 & 0 & 0 & \cdots\cdots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & 0 & 0 & 0 & \cdots\cdots & 0 & 1 & 0 \end{bmatrix} \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{matrix} .$$

有一些实际问题，可以归结为有限状态马尔可夫链. 因此马尔可夫链的研究，无论从理论或是实际应用来说都是一个重要的领域，不过限于本书的目的和篇幅，不再进一步介绍了.

习题 2－6

1. 设一质点在直线上作简单随机游动，每一步向右和向左运动的概率分别为 p 和 $q(p+q=1)$. 假定质点的初始位置为 X_0，它在第 n 步的位移为 ξ_n，以 X_n 表示质点在时刻 n 的位置，即 $X_n=X_0+\xi_1+\cdots+\xi_n$，证明对任意整数 $n\geqslant 0$ 及 $i, i_a, i_1, \cdots, i_{n-1}$，有

$$P(X_{n+1}=i+1|X_0=i_0, X_1=i_1,\cdots,X_{n-1}=i_{n-1}, X_n=i)$$
$$=P(X_{n+1}=i+1|X_n=i)=p,$$
$$P(X_{n+1}=i-1|X_0=i_0, X_1=i_1,\cdots,X_{n-1}=i_{n-1}, X_n=i)$$
$$=P(X_{n+1}=i-1|X_n=i)=q,$$

假定条件部分的事件概率大于零.

2. 质点沿着直线在 0,1,2,3 四个位置上作随机游动，如果在时刻 n 质点位于"0"，则在时刻 $n+1$ 它以概率 $p(0<p<1)$ 停留在 0，而以概率 $q=1-p$ 转移到3. 如果时刻 n 质点位于3，则它以后永远停留在 3. 如果时刻 n 质点位于1 或2，则在时刻 $n+1$ 它以概率 q 转移到右边的邻点，而以概率 p 转移到左边的邻点. 试求:

(1) 该马尔可夫链的转移概率矩阵;

(2) 二步转移概率矩阵.

复习题 2－A

1. 掷三颗骰子，求在已知没有两颗点数相同的条件下至少出现一个一点的概率

2. 箱中原有 5 个白球，7 个黑球，每次取出一球，加上两个与它颜色相同的球放回箱中，求在 4 次取球中前两次取得黑球，后两次取得白球的概率.

3. 从{1,2,3,4,5}中不放回地依次选出两数，求

(1) 第一次选到奇数的概率;

(2) 第二次选到奇数的概率;

(3) 两次都选到奇数的概率.

4. 抛一枚硬币，如果出现正面，就从 1 号箱取一球，反之就从 2 号箱取一球. 1 号箱放了 3 个红球，1 个白球；2 号箱放了 1 个红球，1 个白球. 试求:

(1) 取到红球的概率;

(2) 如果取到的是红球，它是从1号箱中取出的概率.

5. 箱中有 b 个黑球，r 个红球，随机取出一球，然后添上 c 个与它颜色相同的球一起放回箱中，再取一球，求第二次取得红球的条件下，第一次取得黑球的概率.

6. 第一只箱中有1个白球与9个黑球；第2只箱中有1个黑球与5个白球. 从这两个箱中各取一球，余下的球都倒入第三只空箱中，再从中取出一球，求取得白球的概率.

7. 某采购员购买10个一包的电气元件. 他的方法是:从一包中随机地抽查3个，如果这3个元件都是好的才买下这一包.假定含有4个次品的包数占30%，而其余的70%的包各含1个次品. 求这个采购员拒购的概率.

8. 装有 $m(m\geqslant 3)$ 个白球和 n 个黑球的罐中失去一球，但不知其颜色. 为了猜测它的颜色，随机地从罐中摸出两球，结果都得白球，问失去的是白球的概率是多少?

9. 袋中装有 a $(a\neq 0)$ 个黑球，$b(b\neq 0)$ 个白球，采用不放回方式摸球. 证明:"第一次摸得黑球"与"第二次摸得黑球"不独立.

10. 若随机变量 ξ 只取一个值 a，证明 ξ 与任意随机变量 η 独立.

11. 在矩形 $\{(x,y):\ a\leqslant x\leqslant b,\ c\leqslant y\leqslant d\}$ $(a<b,c<d)$ 中随机地取一点，分别以 ξ,η 表示它的横、纵坐标. 证明 ξ 与 η 独立.

12. 掷骰子六次. 求下述两事件的概率:

(1) 两次出现1，1次出现3，3次出现6;

(2) 三次出现偶数，两次出现5，一次出现1或3.

13. 圆内有一内接正方形，向圆内独立地投10个点.求4点落在正方形内，3点落在一弓形内，其余3点落在另外三个弓形内的概率.

14. 某厂产品的废品率为0.03，现在要把产品装箱.若要以不小于0.9的概率保证每箱中至少有100个正品，那么每箱至少要装多少个产品(用泊松逼近进行估计)?

15. 箱中放了1个黑球、2个红球与3个白球，从中取5次球，每次1个，取后放回. 求黑球和白球分别至少取到两次的概率.

16. 3个白球和3个黑球分别箱在两只箱中，每箱3个球，每次从两箱中各取一球交换，以 X_n 表示 n 次交换后第一箱中的白球数. 证明 $\{X_n,n\geqslant 0\}$ 是马尔可夫链，并求其转移概率矩阵.

复习题 2-B

1. 三只箱子分别放了白球与黑球. 第一只中有 2 个白球，3 个黑球；第二只中有 2 个白球，2 个黑球;第三只中有 3 个白球，1 个黑球. 从第一只箱子中取一球放入第二箱中，然后从第二只箱中取一球放入第三箱中，最后从第三只箱中取一球放入第一箱中.求最后各箱中球的组成不变的概率.

2. 若甲有 $n+1$个硬币，乙有n个，他们把硬币一一抛出. 证明甲得到正面比乙多的概率为 $\frac{1}{2}$.

3. 设一个家庭有 n 个小孩的概率为

$$p_n = \begin{cases} \alpha p^n , & n \geqslant 1 , \\ 1-\dfrac{\alpha p}{1-p} , & n=0 , \end{cases}$$

其中 $0<p<1, 0<\alpha<\dfrac{1-p}{p}$ ，若认为生男生女是等可能的，证明一个家庭有k 个 $(k \geqslant 1)$男孩的概率为$2\alpha p^k/(2-p)^{k+1}$.

4. 甲袋中有a 个白球 b 个黑球，乙袋中有 b 个白球 a 个黑球.采用有放回方式取球，先从甲袋中取出一球，以后的抽取按下述规则进行：若前一次取出的是白球，下一次就在同一袋中取球；若前一次取出的是黑球,则下一次在另一袋中取球.求第n 次取到白球的概率.

5. 甲、乙两名射手轮流对同一目标进行射击，命中率分别为 p_1 , p_2 . 甲先射，谁先命中谁得胜. 问甲、乙二人获胜的概率各为多少?

6. 设重复伯努利试验中每次成功的概率为 p, P_n 表示 n 重试验中有偶数次成功的概率. 试证

$$P_n = p\,(1-P_{n-1}) + (1-p)P_{n-1} ,$$

由此证明

$$P_n = \frac{1+(1-2p)^n}{2} .$$

7. 在重复伯努利试验中，假设每次成功的概率为p , 求在n 次成功前至少失败 m 次的概率.

8. 甲袋中有 $N-1$ 个白球 1 个黑球，乙袋中有 N 个白球. 每次从两袋中各取一球进行交换，这样经过了 n 次，求黑球在甲袋中的概率，并讨论 $n \to \infty$ 时的情形.

9. 两个赌徒在一局中获胜的概率都是 $\frac{1}{2}$，两人约定谁先赢得一定局数谁就获得全部赌本. 但赌博在中途因故中断，此时赌徒甲还需赢 m 局才胜，乙还需赢 n 局才胜.

(1) 应如何分配赌本才合理?

(2) 证明：对 $m \geqslant 1, n \geqslant 1$，

$$\sum_{k=0}^{n-1} \binom{m+k-1}{k} \left(\frac{1}{2}\right)^{m+k} = \frac{1}{2^{m+n-1}} \sum_{k=0}^{n-1} \binom{m+n-1}{k}.$$

第三章　分布与期望

在前两章中，我们已经多次提到了随机变量，并给出了它的数学定义，但是对于如何描述随机变量取值的概率规律并没有展开讨论．本章将对随机变量取值的概率规律作进一步研究，并给出在应用上和理论上都很重要的一些概率分布．

§3．1　离散型随机变量及其分布列

我们先研究一类比较特殊、易于掌握的随机变量．

定义 3．1．1　定义在样本空间 Ω 上只取有限个或可数个实数值的随机变量称为**离散型随机变量**．

在前两章中我们讨论过的在产品检验中检验 n 件产品中次品的个数 ξ，n 重伯努利试验中成功的次数 ξ 都是离散型随机变量．

对于离散型随机变量，只要知道它可能取的一切值以及取每个值的概率，就掌握了这个随机变量取值的概率规律．为此给出如下的

定义 3．1．2　设 ξ 是离散型随机变量，它可能取的一切值的集为 $\{a_i: i=1,2,\cdots\}$，称

(3．1．3)　　$$P(\xi=a_i)=P_i,\quad i=1,2,\cdots$$

或

(3．1．4)　　$$\begin{pmatrix} a_1 & a_2 \cdots a_n \cdots \\ p_1 & p_2 \cdots p_n \cdots \end{pmatrix}$$

为离散型随机变量 ξ 的**分布列**．

定理 3．1．5　离散型随机变量的分布列 (3．1．3)或 (3．1．4)满足下列性质：

(1) $p_i \geqslant 0$， $i=1, 2, \cdots$；

(2) $\sum_{i=1}^{\infty} p_i = 1$.

反之，任意一个具有 (1)、(2)两条性质的数列 $\{p_i\}$均可作为一个随机变量的分布列.

证 因为 $p_i = P(\xi = a_i)$，由概率的非负性知 $p_i \geqslant 0$. 再者，由于 $a_1, a_2, \cdots$ 是 ξ 所有可能取的值，所以

$$\bigcup_{i=1}^{\infty} \{\xi = a_i\} = \Omega .$$

再由 $\{\xi = a_i\} \bigcap \{\xi = a_j\} = \varnothing$，$i \neq j$ 及概率的可数可加性就有

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P(\xi = a_i) = P\left(\bigcup_{i=1}^{\infty} \{\xi = a_i\}\right)$$

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

反之，若 $p_i \geqslant 0$，$\sum_{i=1}^{\infty} p_i = 1$，令

$$\Omega = \{1, 2, 3, \cdots\},$$

$$\mathscr{F} = \{A : \quad A \subset \Omega\},$$

$$P(A) = \sum_{i \in A} p_i, \quad A \subset \Omega ,$$

则 $(\Omega, \mathscr{F}, P)$为概率空间. 又令

$$\xi(i) = i ,$$

则 ξ 为 $(\Omega, \mathscr{F}, P)$上的随机变量，且其分布列为

$$\begin{pmatrix} 1 & 2 & \cdots n \cdots \\ p_1 & p_2 & \cdots p_n \cdots \end{pmatrix} .$$

分布列不仅明显地给出了$\{\xi = a_i\}$的概率，而且对于任意的实数 $a < b$，事件$\{a \leqslant \xi \leqslant b\}$的概率也可由分布列算出，即

$$P(a \leqslant \xi \leqslant b) = \sum_{a \leqslant a_i \leqslant b} P(\xi = a_i)$$

$$=\sum_{i:a\leqslant a_i\leqslant b} p_i .$$

从这个意义上讲，离散型随机变量的分布列完全地描述了它取值的概率规律.

现在我们举出最有用的几类离散型随机变量，并讨论它们所具有的性质.

定义 3.1.6 设随机变量ξ的分布列为

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad 0<p<1,$$

称该随机变量服从**0－1 分布或两点分布**，并称上述分布列为以p为参数的 0－1 分布（或两点分布）.

掷一枚均匀硬币，令

$$\xi=\begin{cases} 1, & \text{若掷得正面}, \\ 0, & \text{若掷得反面}, \end{cases}$$

则ξ服从$p=\frac{1}{2}$的两点分布.

在第一章我们曾提到必然事件可以作为随机事件的特例. 若$P(\xi=a)=1$，则称ξ为服从**单点分布或退化分布**的随机变量.

定义 3.1.7 设随机变量ξ的分布列为

$$p_k=P(\xi=k)=\binom{n}{k}p^k\ q^{n-k},\ k=0,1,\cdots,n,$$

其中$q=1-p$，$0<p<1$，则称ξ为**二项随机变量**或称ξ服从参数为(n,p)的**二项分布**，记作$\xi\sim B(n,p)$，并称上述分布列是参数为(n,p)的二项分布，此外还记

$$b(k;n,p)=\binom{n}{k}p^k\,q^{n-k},\ k=0,1,\cdots,n.$$

容易验证：(1) $\binom{n}{k} p^k q^{n-k} > 0$，$k=0, 1, \cdots, n$，且

(2) $\sum_{k=0}^{n} b(k; n, p) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1$.

正如我们在第二章中所讨论过的，二项分布来源于n重伯努利试验. 在n重伯努利试验中，若每次成功的概率为p，则成功的次数ξ服从参数为(n, p)的二项分布.

定理 3.1.8 若随机变量$\xi \sim B(n, p)$，$\eta \sim B(m, p)$且ξ，η独立，则$\xi+\eta \sim B(m+n, p)$.

证 由ξ，η独立及$\{\xi+\eta=k\} = \bigcup_{j=0}^{k} \{\xi=j, \eta=k-j\}$，有

$$P(\xi+\eta=k) = \sum_{j=0}^{k} P(\xi=j, \eta=k-j)$$

$$= \sum_{j=0}^{k} P(\xi=j) P(\eta=k-j)$$

$$= \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j} \binom{m}{k-j} p^{k-j} q^{m-k+j}$$

$$= \sum_{j=0}^{k} \binom{n}{j} \binom{m}{k-j} p^k (1-p)^{n+m-k}.$$

由于$\sum_{j=0}^{k} \binom{n}{j} \binom{m}{k-j} = \binom{n+m}{k}$（参看习题 3－1，10），故得

$\xi+\eta \sim B(m+n, p)$.

推论 3.1.9 若随机变量$\xi_i \sim B(1, p)$，$i=1, 2, \cdots, n$独立，$\xi = \xi_1+\xi_2+\cdots+\xi_n$，则$\xi \sim B(n, p)$.

命题3．1．10 对任意给定的n，p，定义(3．1．7)中的$b(k;n,p)$具有性质：

(1)$b(k;n,p)$在$k=k_0$处达到最大值，其中

$$k_0=\begin{cases}[(n+1)p], & \text{当}(n+1)p\text{不是整数},\\ (n+1)p\text{或}(n+1)p-1, & \text{当}(n+1)p\text{是整数},\end{cases}$$

$[x]$表示不超过x的最大整数．

(2)当$k<k_0$时$b(k;n,p)$单调增，当$k>k_0$时$b(k;n,p)$单调减．

证 由于

$$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{\binom{n}{k}p^k q^{n-k}}{\binom{n}{k-1}p^{k-1}q^{n-k+1}}=\frac{(n-k+1)p}{k(1-p)}$$

及$(n-k+1)p/[k(1-p)]>1$，$=1$，<1分别等价于$k<$，$=$，$>(n+1)p$，所以

$$b(k;n,p)\begin{cases}>b(k-1;n,p), & \text{当}k<(n+1)p,\\ =b(k-1;n,p), & \text{当}k=(n+1)p,\\ <b(k-1;n,p), & \text{当}k>(n+1)p.\end{cases}$$

故$b(k;n,p)$先随k增加；达到最大值后，随k减少，且当$(n+1)p$为整数时，$b(k;n,p)$在$k=(n+1)p-1$及$(n+1)p$处达到最大值，而在$(n+1)p$不是整数时，$b(k;n,p)$在$k=[(n+1)p]$处达到最大值．

下面再正式给出在§1．2中已经详细讨论过的超几何分布的定义．

定义3．1．11 设$N\geqslant M$，$N\geqslant n$．若随机变量ξ的分布列为

$$p_k = P(\xi = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \cdots, n,$$

则称 ξ 为服从超几何分布的随机变量，并称上述分布列为超几何分布.

布实上，它是一个分布列. 因为 $p_k \geqslant 0$，$k = 0, 1, \cdots, n$；再者

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

由组合的意义，不难证明 $\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$ (参看习题 3－1，10)，因而 $\sum_{k=0}^{n} p_k = 1$.

在定理2．5．4 中作为二项分布的极限，已经导出并定义了泊松分布，并在定义2．5．5 中定义了服从泊松分布的随机变量 ξ. 为了完整起见，我们重述如下的

定义 3．1．12 若随机变量 ξ 的分布列为

$$P(\xi = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots,$$

其中 $\lambda > 0$，则称 ξ 为服从参数为 λ 的泊松分布的随机变量，简称为泊松随机变量，记作 $\xi \sim \mathcal{P}(\lambda)$，并称上述分布列为泊松分布. 事实上，它确实是一个分布列，因为 $e^{-\lambda}\lambda^k/k! > 0$，且

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

由泊松分布的导出过程可以看出：如果随机变量 ξ 表示为数众多的个体中出现的个数，每一个个体是否出现彼此独立，而每个个体出现的概率相同且值非常小，则 ξ 近似于服从泊松分布．例如：

(1) 某电话交换台在单位时间内收到用户的呼叫次数 ξ 服从泊松分布．因为一个交换台为它所管辖的千百个用户服务，而每个用户在给定的单位时间内是否呼叫彼此独立且每个用户在给定的单位时间内呼叫的概率又非常小(且彼此相同)．

(2)一本书中某一页(或某几页)上印刷错误的个数服从泊松分布．因为一页上有几百个印刷符号，每个符号是否印错彼此独立，每个符号印错的概率可以认为相同且非常小．

(3)在一个固定时间内，从某块放射性物质中发射出的 $\alpha-$ 粒子数．

(4)某地区居民中能活到 100 岁的人数．因为一个地区有成千上万人，每个人能否活到 100 岁彼此独立，且每个人能活到 100 岁的概率相同又比较小．

(5)一匹布上的疵点数．

下面我们将从另一个角度导出泊松分布．

*例 3.1.13　(泊松流)　设某种事件(例如收到用户呼叫)发生在一列随机的时间点列上，并设存在某个正常数 λ 使得下述条件成立：

1. 在任意长为 h 的时间区间内，正好发生一次“事件”的概率彼此相同且等于 $\lambda h+o(h)$，$o(h)$是当 $h\to 0$ 时比 h 更高阶的无穷小量；

2. 在任意长为 h 的区间内发生两次或两次以上的“事件”的概率为 $o(h)$ $(h\to 0)$；

3. 在互不相交的时间区间内“事件”发生的次数彼此独立;

在这些条件下，我们将证明在 $[0,t)$ 区间内“事件”发生的次数 $N(t)$ 服从以 λt 为参数的泊松分布.

将 $[0,t)$ 等分为 n 个互不相交的子区间(如图 3－1),

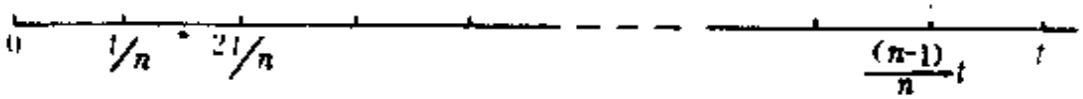


图 3－1

于是对于给定的 $k=0,1,2,\cdots$

$$P(N(t)=k)=P(\text{在 } n \text{ 个子区间中恰有 } k \text{ 个各含一个“事件”})+P(N(t)=k,\text{且至少有一个子区间内含有多于一个的“事件”}). \tag{3.1.14}$$

用 A_n 和 B_n 分别表示出现在(3.1.14)右端的两个事件. 则由条件 1，3 知

$$P(A_n)=\binom{n}{k}\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right]^k\left[1-\frac{\lambda t}{n}-o\left(\frac{t}{n}\right)\right]^{n-k}$$

因为当 $n\to\infty$ 时

$$n\left(\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right)=\lambda t+t\left[\frac{o(t/n)}{t/n}\right]\to\lambda t,$$

令 $p_n=\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)$，则 $np_n\to\lambda t\ (n\to\infty)$，从而二项分布

$$\binom{n}{k}p_n^k\,(1-p_n)^{n-k}\to e^{-\lambda t}\,\frac{(\lambda t)^k}{k!}$$

(其证明与定理 2.5.4 的证明类似). 于是

$$(3.1.15)\qquad P(A_n)\to e^{-\lambda t}\frac{(\lambda t)^k}{k!}\qquad (n\to\infty).$$

而由条件2知

$$(3.1.16)\qquad P(B_n)\leqslant P(n\text{个子区间中至少有一个子区间包含多于一个“事件”})$$

$$=P\left(\bigcup_{i=1}^{n}\text{“第}i\text{个子区间包含多于一个“事件””}\right)$$

$$\leqslant\sum_{i=1}^{n}P(\text{第}i\text{个子区间包含多于一个“事件”})$$

$$=n\ o\left(\frac{t}{n}\right)=t\frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}}\to 0\quad (n\to\infty).$$

由(3.1.14)—(3.1.16)可得

$$P(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\quad k=0,1,2,\cdots.$$

$\{N(t),t\geqslant 0\}$是一个具有上述分布的随时间t演变的随机变量族，称之为**泊松过程**．λ是单位时间内“事件”发生的“平均次数”(§3.3将会进一步阐明这一参数的意义)，可以由经验来确定．

泊松过程有着重要的应用．

泊松分布还具有下述性质：

命题3.1.17 若随机变量$\xi_i\sim\mathcal{P}(\lambda_i)$，$i=1,2$，且独立，$\xi=\xi_1+\xi_2$，则$\xi\sim\mathcal{P}(\lambda_1+\lambda_2)$．

请读者作为习题证明此命题．

定义3.1.18 设$0<p<1$，随机变量ξ的分布列为

$$P(\xi=n)=(1-p)^n p,\quad n=0,1,2,\cdots,$$

则称ξ为服从几何分布的随机变量并称上述分布列为几何分布.

几何分布由以下问题导出：考虑一个独立重复试验，每次试验成功的概率为p，ξ表示第一次成功前失败的次数，则ξ服从参数为p的几何分布．显然

$$\sum_{n=0}^{\infty}(1-p)^n\cdot p=p\cdot\frac{1}{p}=1.$$

这说明ξ以概率1取有限值．设一个射手每次射击命中率为p，$0<p<1$，则他永远不命中的概率为零.

一般地，设有一个独立重复试验序列，每次试验成功的概率为p，r为一给定的正整数，用ξ表示在第r次成功前失败的次数，即用$\{\xi=k\}$表示事件“第r次成功发生在第$r+k$次试验”，用$f(k;r,p)$表示$\{\xi=k\}$的概率，则

$$P(\xi=k)=f(k;r,p)=\binom{r+k-1}{k}p^r q^k,\tag{3.1.19}$$

$$k=0,1,2,\cdots.$$

这是因为$\{\xi=k\}$发生，当且仅当在前$r+k-1$次的试验中恰有k次失败$r-1$次成功且第$r+k$次也成功．它们的概率分别为$\binom{r+k-1}{k}p^{r-1}q^k$和$p$．因而(3.1.19)成立.

由于

$$\binom{r+k-1}{k}p^r q^k=\frac{(r+k-1)(r+k-2)\cdots(r+1)r}{k!}p^r q^k$$

$$=\frac{(-r)(-r-1)\cdots(-r-k+1)}{k!}p^r(-q)^k$$

$$=\binom{-r}{k}p^r(-q)^k,\quad k=0,1,2,\cdots,$$

我们给出下面的

定义 3.1.20 称分布列 (3.1.19)为参数为 (r,p)的**负二项分布**，称具有分布列 (3.1.19)的随机变量 ξ 服从参数为 (r,p) 的负二项分布.

由二项级数公式

$$\sum_{k=0}^{\infty}\binom{-r}{k}(-q)^k=(1-q)^{-r}=p^{-r}$$

知

$$\sum_{k=0}^{\infty}f(k;r,p)=\sum_{k=0}^{\infty}\binom{-r}{k}p^r(-q)^k$$

$$=p^r\sum_{k=0}^{\infty}\binom{-r}{k}(-q)^k=1.$$

因此 (3.1.19)满足分布列所应具备的两个条件. 它也表明第 r 次成功发生在有限次试验的概率为 1.

另一方面，若 $\xi_1,\xi_2,\cdots,\xi_r$ 是独立的、具有同样参数 p 的几何随机变量，$\xi=\xi_1+\xi_2+\cdots+\xi_r$，则 ξ 服从参数为 (r,p)的负二项分布 (可以用 §4.4 的方法证明).

和负二项分布布切相关的是下述的

定义 3.1.21 若随机变量 η 具有分布列

$$P(\eta=n)=\binom{n-1}{r-1}p^r q^{n-r},\quad n=r,r+1,\cdots,$$

则称 η 为服从参数为 (r,p)的**帕斯卡分布**，并称上述分布列为具有参数 (r,p)的**帕斯卡分布**.

若 $n=r+k$，则

$$P(\eta=n)=P(\eta=r+k)=\binom{r+k-1}{k}p^r q^k,\quad k=0,1,2,\cdots.$$

因此

$$P(\eta=r+k)=f(k;r,p),\quad k=0,1,2,\cdots.$$

故帕斯卡分布不过是负二项分布的变形．即ξ服从负二项分布当且仅当$\eta=\xi+r$服从帕斯卡分布．

现在我们利用负二项分布来计算历史上著名的巴拿赫(Banach)火柴问题．

例3.1.22 某人在左右两个口袋中各放一盒火柴，当他要用火柴时随机地选择一个口袋拿一根火柴．假定开始时两盒中都有N根火柴，问他第一次发现一盒空了，且另一盒还有r根火柴的概率u_r．

解 由对称性，所求事件可分解为两个互不相容的等概率事件B_1和B_2的并：

B_1="第一次发现左边口袋火柴盒空了，且右边口袋火柴盒中尚有r根火柴"，

B_2="第一次发现右边口袋火柴盒空了，且左边口袋火柴盒中尚有r根火柴"．

于是我们所要求的概率

$$u_r=P(B_1\cup B_2)=2P(B_1).$$

为求$P(B_1)$，我们把取一根火柴看作一次试验，从左口袋中取火柴视为成功，每次成功的概率p为$\frac{1}{2}$，则首次发现左口袋空时是第$N+1$次成功，右口袋中尚有r根火柴相当于$N+1$次成功以前有$N-r$次失败，故

$$P(B_1)=f\left(N-r,N+1,\frac{1}{2}\right)=\binom{2N-r}{N-r}\left(\frac{1}{2}\right)^{N-r}\left(\frac{1}{2}\right)^{N+1}$$

$$=\binom{2N-r}{N}2^{-2N+r-1}.$$

$$u_r = 2P(B_1) = \binom{2N-r}{N} 2^{r-2N},\quad r=0,1,\cdots,N.$$

前面我们所举的一切例子中的离散型随机变量都取整数值. 从理论上说，这并没有包括了所有的离散型随机变量. 并且不要误认为离散型随机变量取的值 $\{a_1, a_2, \cdots\}$ 一定是数轴上一个个孤立点. 为说明此问题，我们给出下面的

例3.1.23 设随机变量 ξ 取一切有理数为值. 按一定方式将有理数与自然数建立一一对应，记这样建立了顺序的全部有理数集为 $\{r_1, r_2, \cdots r_n, \cdots\}$，并令

$$P(\xi = r_k) = \frac{1}{2^k},\quad k=1,2,\cdots. \tag{3.1.24}$$

由于

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

故 ξ 为具分布列 (3.1.24)的随机变量. 值得注意的是由于有理数的稠密性，任意两个实数之间有无穷多个有理数，若 $a<b$，则 a, b 之间有无穷多个有理数，从而

$$P(a \leqslant \xi \leqslant b) = \sum_{k:\, a \leqslant r_k \leqslant b} \frac{1}{2^k}$$

仍为一个无穷级数的和.

在实际问题中，常要同时研究某一随机变量的函数. 例如某种货物每件价值 a 元，设废品数 ξ 为随机变量. 每出现一件废品要损失 a 元，因此常要考虑由废品引起的损失 $\eta = a\xi$. 这时 η 也是一个随机变量. 我们需要由 ξ 的分布列求出 η 的分布列来.

例3.1.25 若 ξ 的分布列为

$$\begin{pmatrix} a_1 & a_2, & \cdots a_n & \cdots \\ p_1 & p_2 & \cdots p_n & \cdots \end{pmatrix},$$

$\eta=a\xi$ $(a\neq 0)$，则 η 的分布列为

$$\begin{pmatrix} aa_1 & aa_2 \cdots aa_n \cdots \\ p_1 & p_2 \cdots \ p_n \ \cdots \end{pmatrix}.$$

例 3.1.26 若 ξ 的分布列为

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{4} & \frac{1}{12} \end{pmatrix},$$

$\eta=\xi^2$，则 η 可能取的值为 0，1，4 且

$$P(\eta=0)=P(\xi=0)=\frac{1}{3},$$

$$P(\eta=1)=P(\{\xi=-1\}\cup\{\xi=1\})=\frac{1}{3}+\frac{1}{4},$$

$$P(\eta=4)=P(\xi=2)=\frac{1}{12}.$$

于是 η 的分布列为

$$\begin{pmatrix} 0 & 1 & 4 \\ \frac{1}{3} & \frac{7}{12} & \frac{1}{12} \end{pmatrix}.$$

对于一般的函数 $\eta=f(\xi)$，我们有

定理 3.1.27 （离散型随机变量函数的分布列） 设 ξ 为离散型随机变量，其分布列为

$$\begin{pmatrix} a_1 & a_2 & \cdots a_n \cdots \\ p_1 & p_2 & \cdots p_n \ \cdots \end{pmatrix}.$$

设 $f(x)$ 为 R 上的实函数，若 $\eta=f(\xi)$，则 η 为离散型随机变量，η 可能取的值的集合为 $f(a_i)$，$i=1,2,\cdots$ 中不同值的全体，即

$$\{b_k:\text{存在}i,\text{使得}f(a_i)=b_k\}.$$

η 的分布列为

$$P(\eta=b_k)=\sum_{i\in I_k}P(\xi=a_i)=\sum_{i\in I_k}p_i$$

其中 $I_k=\{i:f(a_i)=b_k\}$

注意：正如例 3.1.26 所示，对应于 b_k 有时不只一个 a_i 使 $f(a_i)=b_k$，那么 $\{\eta=b_k\}=\bigcup_{i\in I_k}\{\xi=a_i\}$，计算 $\{\eta=b_k\}$ 的概率时应该把所有使得 $f(a_i)=b_k$ 的事件 $\{\xi=a_i\}$ 的概率都加在一起.

习题 3-1

1. 袋中有 a 个白球，b 个黑球，从中一次取出 r 个球，设其中黑球数为 ξ，求 ξ 的分布列.

2. 将一颗骰子掷两次，以 ξ 表示第一次出现的点数与第二次出现的点数之差，求 ξ 的分布列及事件 $\{2\leqslant\xi\leqslant4\}$ 的概率.

3. 设随机变量 ξ 的分布列为 $P(\xi=k)=\frac{k}{15}$，$k=1,2,3,4,5$. 求 (1) $P(\xi=1$ 或 $\xi=2)$；(2) $P(\frac{1}{2}<\xi<\frac{5}{2})$；(3) $P(\xi>3)$.

4. c 取何值时下列数列成为概率分布：

(1) $p_k=\frac{c}{N}$，$k=1,2,\cdots,N$；

(2) $p_k=c\frac{\lambda^k}{k!}$，$k=1,2,\cdots$；

(3) $p_k=c\left(\frac{2}{3}\right)^k$，$k=1,2,3$.

5. 设 ξ 服从几何分布，求下述随机变量的分布列：

(1) $\eta=\frac{1+(-1)^\xi}{2}$；

(2) $\eta=\sin\frac{\xi\pi}{2}$.

6 . 射手打靶，每次命中率为 0 .8 .他连续射击，直至脱靶为止 .

(1)求射击次数的分布列；

(2) 射击几次的可能性最大？

7 . 某零件次品率为 0 .1 ,

(1)在该零件的生产过程中抽取 4 个零件，设次品数为 ξ，求 ξ 的分布列 .

(2) 在该零件的生产过程中抽取 20 个零件，问次品数为几的可能性最大？

8 . 证明命题 3 . 1 .17 成立 .

9 . 若 η 服从几何分布，试证：对一切非负整数 k , n

$$P(\eta=k+n\mid\eta\geqslant k)=P(\eta=n).$$

这一性质称为几何分布的无记忆性 .

反之，若 η 为只取非负整值的离散型随机变量，且上述性质对一切非负整数 k 及 $n=0$ 成立，则 η 为几何型随机变量 .

10 . 设 k , n , m 是正整数且 $k\leqslant n+m$. (i)试将 $(1+x)^{m+n}=(1+x)^n(1+x)^m$ 按等式两端的不同方式展成 x 的多项式，比较 x^k 的系数，并由此证明等式

$$\binom{m+n}{k}=\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j};$$

(ii)试直接按组合数的定义证明上述恒等式 .

§3 .2 离散型随机变量的期望、方差与母函数

我们已经知道离散型随机变量的分布列完全描述了随机变量取值的概率规律.但是为了对随机变量有个概括的认识，我们还需要了解刻划随机变量的某些特征数值 . 正如我们想了解一个班级学生某门课程的考试成绩，虽然记分册上有考试成绩的全部情况，但是当我们对两个班级的考试成绩进行比较时，并不满足于两个记分册，而是要在记分册的基础上进行一些加工整理 . 例如我们可以把两个班级各自的平均分数计算出来，看哪个班平均成绩好一些 . 又如我们希望知道每个班成绩好的和成绩差

的悬殊大小，就要想办法求出表示分数好坏差别大小的数字特征的值来等等.

另外，当随机变量的分布列具有某些参数时，我们需要了解这些参数的意义是什么.

随机变量的特征数值中最重要的是表示它平均取值的期望(或称均值)和表示它取值的分散程度的方差.

1. 期望(或均值)

为了给出随机变量平均取值大小的合理定义，我们先看一个例子.

设某城市共有 n 个家庭，其中有 k 个小孩的家庭为 n_k 个 $(k=0,1,2,\cdots,m)$. 若用随机变量 ξ 表示从该城市中任取一个家庭的小孩数，则

$$P(\xi=k)=\frac{n_k}{n},\ k=0,1,\cdots,m.$$

自然可以认为

$$\xi\text{的平均取值}=\frac{\text{该城市的小孩总数}}{\text{该城市的家庭总数}}$$

$$=\frac{0\times n_0+1\times n_1+\cdots+m\cdot n_m}{n}=0\times\frac{n_0}{n}+1\times\frac{n_1}{n}+\cdots+m\times\frac{n_m}{n}$$

$$=\sum_{k=0}^{m}k\cdot P(\xi=k).$$

一般地，设随机变量 ξ 具有分布列

$$\begin{pmatrix}a_0 & a_1\cdots a_m\\ p_0 & p_1\cdots p_m\end{pmatrix},$$

对 ξ 取 N_1 次值 (N_1 较大)，若其中有 n_{1k} 次取值 a_k，$k=0,1,\cdots,m$，则 $n_{10}+n_{11}+\cdots+n_{1m}=N_1$，且对这 N_1 次取值，

$$\xi\text{的平均取值}=\frac{a_0n_{10}+a_1n_{11}+\cdots+a_mn_{1m}}{N_1}$$

$$= a_0 \frac{n_{10}}{N_1} + a_1 \frac{n_{11}}{N_1} + \cdots + a_m \frac{n_{1m}}{N_1} .$$

仿此，依次对 ξ 取 N_2，N_3，… 次值．设 ξ 取 N_l 次值的试验中，有 n_{lk} 次取值 a_k，$k=0$，1，…，m，则

$$n_{l0}+n_{l1}+\cdots+n_{lm}=N_l ,$$

$$\xi \text{的平均取值} = a_0 \frac{n_{l0}}{N_l} + a_1 \frac{n_{l1}}{N_l} + \cdots + a_m \frac{n_{lm}}{N_l} .$$

由频率的定义知 $\frac{n_{1k}}{N_1}$，$\frac{n_{2k}}{N_2}$，…，$\frac{n_{lk}}{N_l}$，… 是事件 $\{\xi=a_k\}$ 出现的频率，它们应该稳定在 $P(\xi=a_k)=p_k$ 附近，$k=0$，1，…，m．从而 ξ 的平均取值应该稳定在

$$\sum_{k=0}^{m} a_k P(\xi=a_k)=\sum_{k=0}^{m} a_k p_k$$

上．

对于取可数多个值的离散型随机变量 ξ，也可作类似的考察并得到类似的结论．我们在下面给出表示离散型随机变量平均取值大小的数字特征——期望的定义．

定义 3.2.1 设离散型随机变量 ξ 的分布列为

$$\begin{pmatrix} a_1 & a_2 \cdots a_n \cdots \\ p_1 & p_2 \cdots p_n \cdots \end{pmatrix} \tag{3.2.2}$$

且 $\sum_{n=1}^{\infty} |a_n| p_n < \infty$，则称 $\sum a_i p_i$ 为 ξ (或者分布列 (3.2.2))的**期望**（或均值），记作 $E\xi$．

这里要求级数 $\sum a_i p_i$ 绝对收敛的缘由是：如果级数 $\sum_{n=1}^{\infty} a_n p_n$ 是条件收敛，那么级数的和与级数的项的次序有关．而一个随机变量的分布级律只与它能取哪些值及取相应值的概率有关，和这些

值的排列次序无关，作为它的平均值不能在它的值的一种排列下是这个数，换一种排列又是另一个数．

例 3．2．3 二项分布 $B(n,p)$ 的期望．若随机变量 ξ 的分布列为

$$P(\xi=k)=\binom{n}{k}p^k q^{n-k},\ k=0,1,\cdots,n$$

则

$$E\xi=\sum_{k=0}^{n}k\binom{n}{k}p^k q^{n-k}$$

$$=\sum_{k=1}^{n}\frac{n(n-1)\cdots(n-k+1)}{(k-1)!}p^k q^{n-k}$$

$$=np\sum_{k=1}^{n}\frac{(n-1)(n-2)\cdots[(n-1)-(k-1)+1]}{(k-1)!}p^{k-1}q^{(n-1)-(k-1)}$$

$$=np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j q^{n-1-j}=np(p+q)^{n-1}$$

$$=np.$$

例3．2．4 泊松分布的期望．若随机变量 $\xi\sim\pi(\lambda)$，则

$$E\xi=\sum_{k=0}^{\infty}k\cdot e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{k=1}^{\infty}e^{-\lambda}\frac{\lambda^k}{(k-1)!}$$

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

这说明泊松分布中的参数 λ 是泊松分布的期望．

例 3．2．5 巴特凯维茨 (Bortkiowicz)根据普鲁士军队的报告，计算过 10 个连队中的骑兵在连续二十年的时间内因受马的践踏而致死亡的人数．将每年每队中因马的践踏致死的人数 ξ 作为

随机变量，它有 10×20 个观察值，得到频率如表3－1第二行所示．由表中数据算得每年每队因马践踏致死事故的平均数为

$$0\times 0.545+1\times 0.325+2\times 0.110+3\times 0.015+4\times 0.005=0.61.$$

表 3－1　马践踏致死事故的频率

ξ	0	1	2	3	4
频率	0.545	0.325	0.110	0.015	0.005
概率	0.544	0.331	0.101	0.021	0.003

从现象上看，一个连队的骑兵中在 $[0,t]$（单位：年）中“受马践踏致死”符合泊松流（例3.1.3）的假设，因此采用例3.1.3的记号，$\xi=N(1)\sim\pi(\lambda)$，现在用 $\lambda=0.61$，计算泊松分布的相应概率 $P(\xi=k)$，$k=0,1,2,3,4$ 得表中第三行（亦可通过查附表1得出）．从表中看出，这样算出的概率与相应的频率差别甚小．这说明我们从现象上分析得出的结论是正确的．

本例也告诉我们在实践中如何获得泊松分布的参数 λ．

例3.2.6　（超几何分布的期望）设 ξ 服从超几何分布，其分布列为

$$P(\xi=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,\cdots,n.$$

同二项分布类比，我们猜想它的期望可能是 $n\cdot\dfrac{M}{N}$．注意

$$k\binom{M}{k}=M\cdot\binom{M-1}{k-1},\quad \binom{N}{n}=\frac{N}{n}\binom{N-1}{n-1},$$

于是

$$E\xi=\sum_{k=0}^{n} k\cdot P(\xi=k)=\sum_{k=1}^{n}\frac{k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$=n\frac{M}{N}\sum_{k=1}^{n}\frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}}\quad(\text{令}k-1=j)$$

$$=n\frac{M}{N}\sum_{j=0}^{n-1}\frac{\binom{M-1}{j}\binom{(N-1)-(M-1)}{(n-1)-j}}{\binom{N-1}{n-1}}=n\cdot\frac{M}{N},$$

其中 $\dfrac{\binom{M-1}{j}\binom{(N-1)-(M-1)}{(n-1)-j}}{\binom{N-1}{n-1}}$ 相当于在具有 $M-1$ 个次品的 $N-1$ 个产品中任取 $n-1$ 个，其中恰有 j 个次品的概率，所以从 $j=0$ 到 $j=n-1$ 求和应该为 1.

2. 方差

随机变量的期望反映了随机变量取值的平均大小. 在有些情况下仅仅知道平均值是不够的. 例如有甲、乙两种牌子的手表，它们的日走时误差 (单位：秒) 分别为 ξ_1，ξ_2，各具有如下的分布列：

$$\begin{pmatrix}-1 & 0 & 1\\ 0.1 & 0.8 & 0.1\end{pmatrix},$$

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.1 & 0.2 & 0.4 & 0.2 & 0.1 \end{pmatrix}.$$

容易验证，$E\xi_1 = E\xi_2 = 0$，从日走时误差的平均情况来看，分不出这两种牌号的表的优劣．但是进一步观察，就会看到甲种牌号的表优于乙种牌号的表．这是因为：第一，甲种牌号的表只有20%分散在其平均值的两侧，而乙种牌号的表却有60%分散在其平均值的两侧．第二，甲种表误差只是 ± 1 秒，而乙种表却有20%误差在 ± 2 秒．总而言之，甲种表偏离平均值较小，乙种表偏离平均值较大．

如何给出一个随机变量偏离期望值的程度的度量呢？自然会想到用 $|\xi - E\xi|$ 表示 ξ 与 $E\xi$ 偏离的大小，用 $E|\xi - E\xi|$ 表示平均偏离的大小．但由于绝对值运算在数学处理上有许多不便，人们便用 $E(\xi - E\xi)^2$ 或 $\sqrt{E(\xi - E\xi)^2}$ 表衡量 ξ 与它的期望 $E\xi$ 偏离的大小．

定义3.2.7 设 ξ 为离散型随机变量，其期望 $E\xi$ 存在，如果 $E(\xi - E\xi)^2$ 也存在（有限），则称 $E(\xi - E\xi)^2$ 为随机变量 ξ 的方差，记作 $D\xi$ 或 $\mathrm{Var}\xi$．方差的算术平方根称为 ξ 的标准差或均方差．

注意，ξ 的方差实际上是 ξ 的函数 $(\xi - E\xi)^2$ 的期望，按照定义，应该首先计算出 $\eta = (\xi - E\xi)^2$ 的分布列，然后再来计算 $E\eta$．为了直接由 ξ 的分布列计算 ξ 的方差，我们来证明下述

定理3.2.8 若 ξ 为离散型随机变量，其分布列为

$$P(\xi = a_k) = p_k,\ k = 1, 2, \cdots.$$

f 为定义在 $\{a_k,\ k = 1, 2, \cdots\}$ 上的任意实函数，若

$$\sum_{k=1}^{\infty} |f(a_k)| p_k < \infty,$$

则 $Ef(\xi)$ 存在，且

$$Ef(\xi)=\sum_{k=1}^{\infty}f(a_k)p_k.$$

证　令

$$\{b_j:\text{存在 } a_k \text{ 使 } f(a_k)=b_j\}.$$

因为自变量集$\{a_j\}$至多为可数无限多个，所以$f(x)$的值集$\{b_j\}$至多可数．于是

$$P(f(\xi)=b_j)=\sum_{k:f(a_k)=b_j}P(\xi=a_k)$$

$$=\sum_{k:f(a_k)=b_j}p_k.$$

$$E|f(\xi)|=\sum_j|b_j|P(f(\xi)=b_j)$$

$$=\sum_j|b_j|\sum_{k:f(a_k)=b_j}p_k=\sum_j\sum_{k:f(a_k)=b_j}|b_j|p_k$$

$$=\sum_j\sum_{k:f(a_k)=b_j}|f(a_k)|p_k=\sum_k|f(a_k)|p_k<\infty.$$

因而$Ef(\xi)$存在，且采用与上面相同的算法可得

$$Ef(\xi)=\sum_j b_j\cdot P(f(\xi)=b_j)=\sum_j b_j\sum_{k:f(a_k)=b_j}p_k$$

$$=\sum_j\sum_{k:f(a_k)=b_j}b_jp_k$$

$$=\sum_j\sum_{k:f(a_k)=b_j}f(a_k)p_k$$

$$=\sum_k f(a_k)p_k.$$

注意，这里由于所有b_j是互不相同的，因此指标集$\{k:f(a_k)=b_j\}$是互不相交的，且它们的并就构成了$\{a_k\}$的全部指标集

$\{k: k=1, 2, \cdots\}$.

为了给出方差的计算公式，我们进而证明下述

命题 3.2.9 设 ξ 的分布列为

$$P(\xi=a_k)=p_k, k=1, 2, \cdots$$

且 $\sum_k a_k^2 p_k < \infty$，则 $D\xi$ 存在且

$$D\xi = E\xi^2 - (E\xi)^2 = \sum_k a_k^2 p_k - (E\xi)^2.$$

证 由定理 3.2.8 知，$D\xi$ 存在的条件是

$$\sum_k (a_k - E\xi)^2 p_k < \infty,$$

且 $E\xi$ 存在. 由施瓦兹不等式知

$$\left(\sum_k |a_k| p_k\right)^2 \leqslant \sum_k 1^2 p_k \cdot \sum_k a_k^2 p_k < \infty,$$

因而由题设条件知 $E\xi$ 存在，再由

$$\begin{aligned}\sum_k (a_k - E\xi)^2 p_k &\leqslant \sum_k [2a_k^2 + 2(E\xi)^2] p_k \\ &= 2\sum_k a_k^2 p_k + 2(E\xi)^2 \sum_k p_k < \infty,\end{aligned}$$

知方差存在. 于是由定理 3.2.8 得

$$\begin{aligned}D\xi &= \sum_k (a_k - E\xi)^2 p_k = \sum_k [a_k^2 - 2a_k E\xi + (E\xi)^2] p_k \\ &= \sum_k a_k^2 p_k - 2\sum_k a_k p_k \cdot E\xi + (E\xi)^2 \sum_k p_k \\ &= \sum_k a_k^2 p_k - (E\xi)^2 \\ &= E\xi^2 - (E\xi)^2.\end{aligned}$$

期望和方差是随机变量最重要的数字特征，下面再给出其它一些数字特征：

定义 3．2．10 若ξ为离散型随机变量，k为正整数，$E|\xi|^k$，$E\xi^k$和$E(\xi-E\xi)^k$存在，则它们分别称为ξ的k阶原点绝对矩、k阶原点矩和k阶中心矩．

关于它们存在性之间的关系，有下述

定理 3．2．11 若k为正整数，$E|\xi|^k$存在，则对一切$j \leqslant k$的正整数j，$E|\xi|^j$都存在，并且$E\xi^k$，$E(\xi-E\xi)^k$也存在．

证 由$E|\xi|^k$存在，$E\xi^k$显然存在．而当$1 \leqslant j \leqslant k$时

$$E|\xi|^j=\sum_{|a_i|\leqslant 1}|a_i|^j p_i+\sum_{|a_i|>1}|a_i|^j p_i$$

$$\leqslant 1+\sum_i |a_i|^k p_i=1+E|\xi|^k<\infty .$$

故由$E|\xi|^k$存在可知$E|\xi|^j$，$1\leqslant j\leqslant k$存在．再由$E|\xi|^k$存在，易知$E\xi$存在，由

$$E|\xi-E\xi|^k=\sum_j |a_j-E\xi|^k p_j$$

$$\leqslant \sum_j (|a_j|+|E\xi|)^k p_j$$

$$=\sum_j \sum_{i=0}^{k}\binom{k}{i}|a_j|^i |E\xi|^{k-i} p_j$$

$$=\sum_{i=0}^{k}\binom{k}{i}E|\xi|^i (|E\xi|)^{k-i}<\infty ,$$

可知$E(\xi-E\xi)^k$存在．

例 3．2．12 若$\xi \sim B(n,p)$，求$D\xi$．

解 例 3．2．3已经求出$E\xi=np$，利用定理 3．2．9，当

$n \geqslant 2$ 时，

$$D\xi = E\xi^2 - (E\xi)^2 = \sum_{k=0}^{n} k^2 b(k; n, p) - (np)^2$$

$$= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} - (np)^2$$

$$= \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k q^{n-k} + np - (np)^2$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)(n-1)\cdots(n-k+1)}{(k-2)!} p^{k-2} q^{(n-2)-(k-2)}$$

$$+ np - (np)^2$$

$$= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{(n-2)-j} + np - (np)^2$$

$$= n(n-1)p^2 + np - (np)^2 = npq.$$

当 $n=1$ 时，易见 $D\xi = pq$.

例 3.2.13 若 $\xi \sim \mathscr{P}(\lambda)$，求 $D\xi$.

解 $D\xi$ 的存在性将在计算 $E\xi^2$ 的过程中可见. 由例 3.2.4 知 $E\xi = \lambda$，故

$$D\xi = E\xi^2 - (E\xi)^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2$$

$$= \sum_{k=2}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2$$

$$= \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-2)!} + \lambda - \lambda^2$$

$$= \lambda^2 \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

下面给出一个求非负整值随机变量的期望和方差的工具.

3. 非负整值随机变量的母函数

定义 3.2.14 若ξ为只取非负整数值的随机变量，其分布列为

$$P(\xi=k)=p_k,\quad k=0,1,2,\cdots,$$

则称定义在$[-1,1]$上的函数

$$f(s)=\sum_{k=0}^{\infty}s^k p_k \tag{3.2.15}$$

为ξ的**母函数**.

由于$f(1)=\sum_k p_k=1$，因此(3.2.15)的右端在区间$[-1,1]$上绝对且一致收敛，所以母函数$f(s)$在$[-1,1]$上连续，在$(-1,1)$内有任意阶导数. 它可唯一确定随机变量ξ的概率分布，因为

$$p_n=\frac{f^{(n)}(0)}{n!}.$$

例 3.2.16 若$\xi\sim B(n,p)$则ξ的母函数

$$f(s)=\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}s^k=(q+ps)^n.$$

例 3.2.17 若$\xi\sim\mathcal{P}(\lambda)$，则$\xi$的母函数

$$f(s)=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}s^k=e^{\lambda(s-1)}.$$

由幂级数的性质知，假如$E\xi^n$存在，则我们可以用母函数在$s=1$处的导数来表示ξ的直到n阶原点矩. 即

$$f'(s)=\sum_{k=0}^{\infty}k\,s^{k-1}p_k,$$

$$f''(s)=\sum_{k=0}^{\infty} k(k-1)s^{k-2}p_k,$$

……

$$f^{(n)}(s)=\sum_{k=0}^{\infty} k(k-1)\cdots(k-n+1)s^{k-n}p_k$$

在 $s=1$ 处存在，且

$$f'(1)=E\xi,$$

$$f''(1)=\sum_k k(k-1)p_k=E\xi^2-E\xi,$$

$$f^{(3)}(1)=\sum_k k(k-1)(k-2)p_k=E\xi^3-3E\xi^2+2E\xi,$$

……

因而我们有

$$E\xi=f'(1), \tag{3.2.18}$$

$$E\xi^2=f''(1)+f'(1), \tag{3.2.19}$$

$$D\xi=E\xi^2-(E\xi)^2=f''(1)-[f'(1)]^2+f'(1), \tag{3.2.20}$$

……

下面我们举例说明如何用母函数计算随机变量的期望和方差.

例 3.2.21 若 ξ 服从参数为 p 的几何分布，则

$$p_k=P(\xi=k)=q^kp,\ k=0,1,2,\cdots.$$

ξ 的母函数

$$f(s)=\sum_{k=0}^{\infty} q^kps^k=\sum_{k=0}^{\infty}(sq)^kp=\frac{p}{1-qs}.$$

对母函数求导

$$f'(s)=\frac{pq}{(1-qs)^2},$$

$$f''(s)=\frac{2pq^2}{(1-qs)^3}.$$

由 (3.2.18)及 (3.2.20),

$$E\xi=f'(1)=\frac{q}{p},$$

$$D\xi=f''(1)-[f'(1)]^2+f'(1)=2\frac{q^2}{p^2}-\left(\frac{q}{p}\right)^2+\frac{q}{p}$$

$$=\frac{q^2}{p^2}+\frac{q}{p}=\frac{q^2+pq}{p^2}=\frac{q}{p^2}.$$

习题 3-2

1. 设 ξ 的分布列为

(1) $P(\xi=k)=\frac{1}{2^k}$, $k=1,2,\cdots$;

(2) $P(\xi=k)=\frac{a^k}{(1+a)^{k+1}}$, $k=0,1,2,\cdots$ $(a>0)$,求 $E\xi$ 和 $D\xi$.

2. 箱内有 5 个电器元件,其中有 2 个是次品,设每次从箱中随机地抽取一个进行检查,直到查出全部次品为止.求所需检验次数的数学期望与方差.

3. 5 把钥匙中只有 1 把能开锁,每次随意地试用一把,如果

(1) 把每次试过的钥匙分开;

(2) 把每次试过的钥匙又混杂进去.

求直到把锁打开的试用次数的期望与方差.

4. 从 $\{1,2,\cdots,N\}$ 这 N 个数中随机地取 n 个数,以 ξ 表示其中最小的数.若采用

(1) 有放回;

(2) 无放回.

求 $E\xi$ 和 $D\xi$.

5. 设 ξ，η 的母函数分别为

(1) $\frac{1}{4}(1+s)^2$;

(2) $\frac{1}{2-s}$.

求 $P(\xi>n)$ 及 $P(\eta=2n)$.

6. 试由 $\nu(\lambda)$ 的母函数，求它的期望与方差.

§3.3 连续型随机变量

在§3.1 我们讨论了离散型随机变量，它的可能值或者是有限个，或者是可数无穷多个. 但在实践中还有许多随机变量，它们可能取的值的集合不是可数的. 例如在会面问题中，我们用 ξ，η 分别表示甲、乙两人到达的时刻，它们可以取 $[0, T]$ 中任意实数，因此 ξ，η 的可能值集合是不可数. 在这类随机变量中最常见的是连续型随机变量，它的数学定义为：

定义 3.3.1 称随机变量 ξ 为连续型的，如果存在一个定义在 $(-\infty, \infty)$ 上的非负可积函数 $f(x)$，使得对任何的实数 $a \leqslant b$，有

$$P(a \leqslant \xi \leqslant b) = \int_a^b f(x)\,dx.$$

函数 f 称为随机变量 ξ 的**概率密度函数或分布布度**.

直观地看，连续型随机变量 ξ 在一个区间 $[a, b]$ 上取值的概率等于区间 $[a, b]$ 之上，概率分布密度曲线 $f(x)$ 之下曲边梯形 (即图 3－2 中阴影部份)的面积.

由定义易知，f 作为概率密度函数应具有以下性质：

(3.3.2) $f(x) \geqslant 0$，对一切 $x \in (-\infty, \infty)$;

(3.3.3) $\int_{-\infty}^{\infty} f(x)\,dx=1$.

反之，任何一个具有性质 (3.3.2)、(3.3.3)的实直线上的可积函数，都可以成为一个连续型随机变量的概率密度函数 (此事实是定理 3.5.2 的第二部份的直接推论).

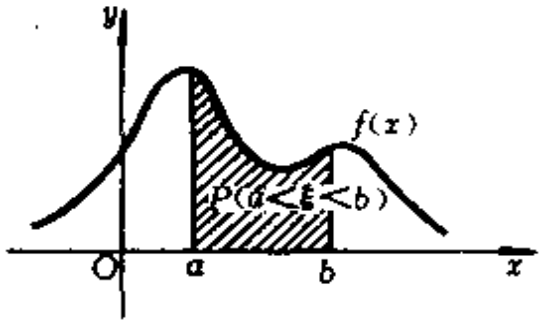


图 3－2

对于连续型随机变量 ξ 来说，它取任何一个点的概率皆为零. 即对任何实数 a，

$$P(\xi=a)=\int_a^a f(x)\,dx=0.$$

下面我们举出一些最常见的连续型随机变量.

定义 3.3.4 称 ξ 为区间 $[a,b]$ 上均匀分布的随机变量，如果它是连续型随机变量，具有概率密度函数

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant x\leqslant b;\\ 0, & \text{其它}.\end{cases}$$

记作 $\xi\sim U(a,b)$，$f(x)$ 如图 3－3 所示. 上述函数也称为在 $[a,b]$ 上均匀分布的密度函数.

均匀分布的随机变量来源于几何概率.若随机试验是在区间 $[a,b]$ 内随机取点,ξ 表示所取点位置的坐标,则ξ 为在$[a,b]$ 上均匀分布的随机变量.

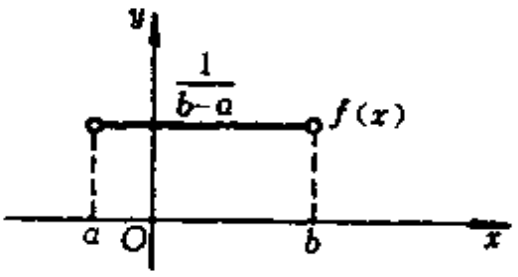


图 3-3

例 3.3.5 某公共汽车站从上午7时起,每 15 分钟开出一班车,即7 :00 , 7 : 15 , 7 : 30,…… 时刻汽车从此站开出,如果某乘客到达此站的时刻为 7 :00 到 7 :30 之间均匀分布的随机变量.试求他等候乘车的时间(1)少于 5 分钟;(2)超过 10 分钟的概率 .

例 设乘客 7 时过 ξ 分钟到达此站, 则 ξ 的分布密度为

$$f(x)=\begin{cases}\dfrac{1}{30}, & 0\leqslant x\leqslant 30,\\ 0, & \text{其它}.\end{cases}$$

(1) P (乘客等候乘车的时间少于 5 分钟)

$$=P(10<\xi\leqslant 15)+P(25<\xi\leqslant 30)$$

$$=\int_{10}^{15}\frac{1}{30}\,dx+\int_{25}^{30}\frac{1}{30}\,dx=\frac{1}{3};$$

(2) P (乘客等候乘车的时间超过 10 分钟)

$$=P(0<\xi<5)+P(15<\xi<20)$$

$$=\int_{0}^{5}\frac{1}{30}\,dx+\int_{15}^{20}\frac{1}{30}\,dx=\frac{1}{3}.$$

例 3.3.6 在长度为l的线段上随机地选取一点,把线段分成两段.试求较短的一段与较长的一段长度之比小于 $\frac{1}{4}$ 的概率.

解 以此线段所在直线为坐标轴,线段的左端点为坐标原点建立坐标系,在该线段上随机取点,就相当于在 $[0,l]$ 中随机取数.设

随机点的坐标为 ξ，则 ξ 为 $[0, l]$ 上均匀分布的随机变量．ξ 把 $[0, l]$ 分成两部分 $[0, \xi]$ 与 $[\xi, l]$．所求事件概率为

$$P\left(\left\{\xi<\frac{l}{2}, \frac{\xi}{l-\xi} \leqslant \frac{1}{4}\right\} \cup\left\{\xi>\frac{l}{2}, \frac{l-\xi}{\xi} \leqslant \frac{1}{4}\right\}\right)$$

$$=P\left(\left\{0 \leqslant \xi \leqslant \frac{l}{5}\right\} \cup\left\{\frac{4 l}{5} \leqslant \xi \leqslant l\right\}\right)$$

$$=\int_0^{l / 5} \frac{1}{l} d x+\int_{4 l / 5}^l \frac{1}{l} d x=\frac{2}{5} .$$

$[a, b]$ 上的均匀分布来源于在 $[a, b]$ 内随机取点．对于 G 是任意一个可求长的一维点集，也可以有类似的讨论：设 ξ 表示在 G 中随机取出的点，$\mu(G)$ 表示 G 的长度，令

$$f(x)=\begin{cases}\dfrac{1}{\mu(G)}, & x \in G, \\ 0, & \text{其它},\end{cases} \quad -\infty<x<\infty . \tag{3.3.7}$$

则对任何区间 $[a, b]$ 来说，

$$P(a \leqslant \xi \leqslant b)=\frac{\mu(G \cap[a, b])}{\mu(G)}=\int_a^b f(x) d x .$$

(3.3.7)是一分布密度，仿定义 3.3.4 可称之为 G 上的均匀分布的密度函数．因而可以称 ξ 为服从 G 上的均匀分布的随机变量，记作 $\xi \sim U(G)$．这是定义 3.3.4 的推广．

定义3.3.8 若ξ为连续型随机变量，其分布密度

$$f(x)=\frac{1}{\sqrt{2 \pi} \sigma} \exp \left\{-\frac{(x-\mu)^2}{2 \sigma^2}\right\},-\infty<x<\infty , \tag{3.3.9}$$

其中$-\infty<\mu<\infty$，$\sigma>0$．则称 ξ 为**正态随机变量**，或称 ξ 服从以 (μ, σ^2) 为参数的正态分布．记作 $\xi \sim N(\mu, \sigma^2)$．当 $\mu=$

$=0$，$\sigma^2=1$ 时称 ξ 为标准正态随机变量，即 $\xi\sim N(0,1)$.

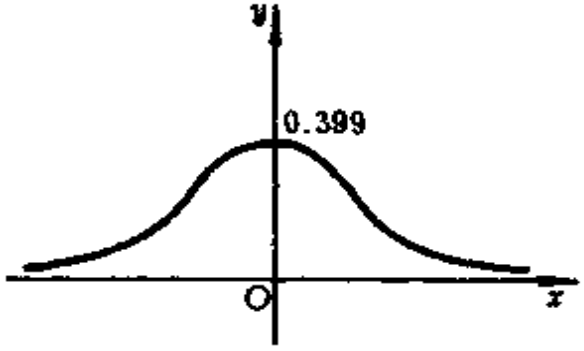


图 3-4

(3.3.9)所示的函数称为参数为 (μ,σ^2)的正态分布的密度函数．参数为 $(0,1)$的标准正态分布的密度函数如图 3－4 所示．

在 §3.4 中将会知道 μ 表示 ξ 的期望，它表征 ξ 的平均位置，σ^2 为 ξ 的方差，它表征 ξ 取值的分散程度．在图 3－5 中，我们给出了 μ 相同而 $\sigma=1$，$\sigma=2$ 及 $\sigma=\frac{1}{2}$ 的三个正态分布密度曲线．

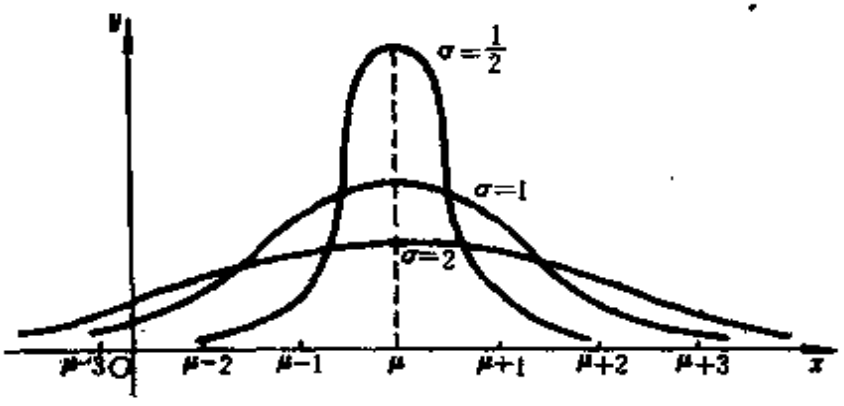


图 3-5

正态分布密度曲线具有如下性质：

(1) 曲线关于 $x=\mu$ 对称；

(2) 当 $x=\mu$ 时 $f(x)$达到最大值 $\dfrac{1}{\sqrt{2\pi}\ \sigma}$；

(3) 当 $x\to\pm\infty$ 时 $f(x)\to 0$，即曲线以 x 轴为渐近线；

(4) 曲线在 $x=\mu\pm\sigma$ 处有拐点．

为了证明(3．3．9)中的函数确为概率密度函数,还必须证明：

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

事实上，利用变量替换 $y=(x-\mu)/\sigma$，可得

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

于是只需证

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx=\sqrt{2\pi}\ .$$

为此，令 $I=\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx$，则

$$I^2=\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\ dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\ dx\,dy\ .$$

变换成极坐标，

$$x=r\cos\theta\ ,\quad y=r\sin\theta\ ,\quad 0\leqslant r<\infty\ ,\quad 0\leqslant\theta\leqslant 2\pi\ ,$$

即得

$$I^2=\int_0^{\infty}\int_0^{2\pi} e^{-\frac{r^2}{2}}\ r\,dr\,d\theta=2\pi\int_0^{\infty} r e^{-\frac{r^2}{2}}\ dr$$

$$=-2\pi\ e^{-\frac{r^2}{2}}\Big|_0^{\infty}=2\pi\ .$$

故 $I=\sqrt{2\pi}$，从而得证所需结论.

由定理 2.5.9 知，正态分布可以看作二项分布的极限，这可以看作是正态分布的理论来源.

在实践上，正态分布是一个有广泛应用的概率分布. 相当广泛的一类随机现象可以用正态分布或近似地用正态分布来刻画(参见 §5.6). 下面举两个例子来说明如何从观察数据来判断所观察的指标是服从正态分布的.

例 3.3.10 表3－2 中的数据是一个自动机床生产的 100 个零件的口径 (单位：mm)，最小值是 10.93，最大值是 11.08，将区间 [10.925，11.085] (左端点略小于 10.93，右端点略大于 11.08)等分成 8 个小区间 [10.925，10.945)，[10.945，10.965]，…，[11.065，11.085)，然后统计出这 100 个数据在每个小区间中的个数——频数，算出频数与数据总数的比——频率，制成表 3－3.

表 3－2

11.02	10.99	10.93	11.01	10.98	10.94	11.02
11.01	10.99	11.00	11.07	10.98	10.97	10.99
10.96	11.02	10.97	10.98	11.00	10.97	11.05
10.95	11.00	11.02	10.99	10.98	11.00	10.98
11.05	11.00	10.98	10.97	11.01	11.07	11.06
10.97	10.94	10.99	11.01	11.03	10.95	11.00
10.95	11.00	11.00	11.00	10.99	10.97	11.02
10.93	10.99	11.02	11.01	11.08	11.00	10.95
11.01	10.93	11.00	10.99	11.01	10.99	10.99
11.04	11.00	11.00	10.97	11.02	11.01	11.04
10.96	10.96	10.99	10.96	10.98	11.08	11.06
10.99	11.00	10.97	11.02	10.98	10.98	11.00
10.99	11.00	11.00	10.99	10.99	10.98	10.97
10.97	10.99	11.03	11.02	11.06	11.04	11.03
11.05	10.96					

表 3－3

序　号	区　间	频　数 n_i	频　率 $\frac{n_i}{100}$	概　率
1	[10.925，10.945]	5	0.05	0.04
2	[10.945，10.965]	9	0.09	0.11
3	[10.965，10.985]	20	0.20	0.19
4	[10.985，11.005]	33	0.33	0.24
5	[11.005，11.025]	17	0.17	0.21
6	[11.025，11.045]	6	0.06	0.13
7	[11.045，11.065]	6	0.06	0.05
8	[11.065，11.085]	4	0.04	0.02

表 3－2 中杂乱无章的数据经过分组制成表 3－3 后，就能看出一点规律性的东西．例如多数数据集中在区间 [10.965，11.025) 中，小于 10.945 或大于 11.065 的都不多．还可以用图形将各小区间内数据的频率表示出来，取一个直角坐标系，如图 3－6，在横坐标轴上标出各小区间的端点：$a_1=10.925$，$a_2=10.945$，…，$a_9=11.085$ 等等，然后以每个小区间为底，以相应的 (频率)/(小区间长) 为高做出 8 个矩形，这样，100 个数据中出现在某个小区间中的频率，就由这个小区间上矩形的面积表示出来了，这使我们能更直观地看到 100 个数据在 x 轴上是怎样分布的．通常称图 3－6 为**频率直方图**，或简称为**直方图**．

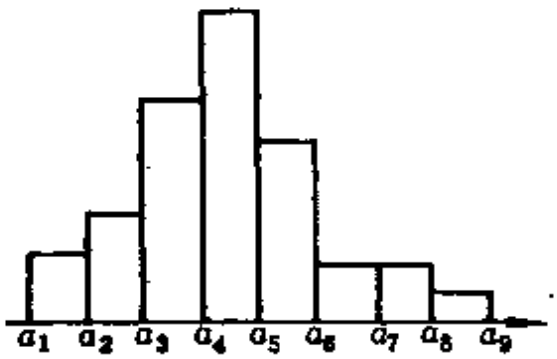


图 3－6

如果用 ξ 表示自动机床生产的一个零件的口径，那么表 3－2 中的 100 个数据就是取自 ξ 的 100 个值．ξ 在每个小区间内取值的频率，在一般情况下接近于 ξ 在那个区间中取值的概率．例如，ξ 在区间 [10.965，10.985] 中取值的频率一般接近于 $P(10.965 \leqslant \xi \leqslant 10.985)$，所以图 3－6 的直方图的外轮廓线一般接近于 ξ 的概率分布密度曲线．从直方图看，它有一个峰，比较对称，而且越远离中间值的数据出现的频率越小．所以我们有理由假设 ξ 的分布是正态分布或接近于正态分布．

为了进一步检查上述 100 个数据是否来自正态母体 (分布) $\xi \sim N(\mu, \sigma^2)$，用 x_i，$i=1, 2, \cdots, 100$ 表示这 100 个数据．由 §5．2 知，可以用 $\overline{x} = \frac{1}{100}\sum_{i=1}^{100} x_i$，$s^2 = \frac{1}{100}\sum_{i=1}^{100}(x_i - \overline{x})^2$ 作为 μ，σ^2 的估计值．经计算得知

$$\mu \approx \widetilde{x} = 10.9972, \quad \sigma \approx s = 0.03268.$$

如果这100 个就据的确来自 $\xi \sim N(\mu, \sigma^2)$，那么表 3－3 中所列各区间上的频率应该近似于 $N(10.9972, (0.03268)^2)$的相应区间上的概率．例如 [10.965，10.985] 上的频率 0.20 应近似于

$$\int_{10.965}^{10.985} \frac{1}{\sqrt{2\pi} \times 0.03268} \exp\left\{-\frac{(x-10.9972)^2}{2(0.03268)^2}\right\} dx \approx 0.19155.$$

表 3－3 的最末一行列出了 $N(10.9972, (0.03268)^2)$的相应区间上的概率．比较这些频率与概率，可以看出它们还是相当接近的，所以可以认为这 100 个就据来自正态母体 $\xi \sim N(10.9972, (0.03268)^2)$．

例 3．3．11　表3－4 中的 157 个数据是 1979 年某地区

157 名学生高考中的化学成绩*)，图 3－7 是这 157 个数据的频率直方图．若用 η 表示该地区一个考生的化学高考成绩，η 是一个随机变量．从直方图看，我们有理由假设 η 的分布是正态分布或接近于正态分布．

表 3－4

41.5	29.5	40.5	23.5	34.5	39.5	37
20.5	32.5	26	33	24	36	14.5
25	32	34.5	42	37	45	31
31	44.5	29.5	50.5	39	55	57
34	49.5	31.5	30	38	18	27
29.5	27	43.5	39	37.5	15	33
36.5	52.5	24	53	35	45.5	29.5
16	39.5	42.5	41.5	50.5	32.5	32.5
46.5	30.5	26	30.5	21.5	64	68
58	56	37.5	47.5	38	34.5	43.5
37	57	43	36.5	24	18.5	42.5
33.5	29.5	27.5	42	20	15.5	26
37.5	28.5	26.5	40.5	18	12.5	13
27.5	24.5	19.5	39	37.5	49.5	19
28.5	37.5	11.5	32.5	35	47.5	41.5
50.5	23.5	28	13.5	26	11	12
42.5	36.5	39.5	24	33.5	45.5	59.5
38	42.5	35.5	18	42.5	47.5	38.5
46.5	40	39.5	59.5	37.5	44.5	35
56.5	59.5	48.5	50.5	37.5	24	39.5
46.5	27.5	45	42.5	31	43.5	41.5
38.5	36	48.5	36	37.5	36.5	44.5
10.5	37	28				

*)这个材料取自叶佩华等编的《教育统计学》(人民教育出版社)．

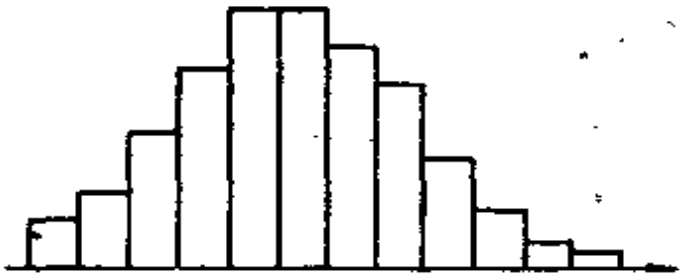


图 3-7

由于正态分布具有广泛的用途，而且我们经常需要计算概率 $P(a \leqslant \xi \leqslant b)$，但是由于不定积分

$$\int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

不是初等函数，因而必须用数值计算来求积分的近似值．本书附表 2 给出了标准正态分布函数

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

的数值表．表中列出了对应于 $x \geqslant 0$ 的函数值．利用这个表可以计算一切正态随机变量取值于某一区间的概率的近似值．

例 3．3．12 若随机变量 $\xi \sim N(0,1)$，则由附表 2 得

(1) $P(\xi \leqslant 2)=\Phi(2)=0.97725$；

(2) $P(\xi>2.34)=1-\Phi(2.34)=1-0.99036=0.00964$；

(3) $P(\xi<-2)=P(\xi>2)=1-\Phi(2)$

$=1-0.97725=0.02275$，

这里利用了正态分布的密度函数关于原点的对称性；

(4) $P(-0.5 \leqslant \xi \leqslant 2.65)=P(\xi \leqslant 2.65)-P(\xi<-0.5)$

$$=\Phi(2.65)-P(\xi>0.5)=\Phi(2.65)-[1-\Phi(0.5)]$$

$$=0.99598-(1-0.69146)=0.68744.$$

附表 2 中只有标准正态分布函数的数值，对于参数 $\mu\neq 0$ 或 $\sigma\neq 1$ 的正态分布，可以利用下述定理化成标准正态分布来计算.

定理 3.3.13 若随机变量 $\xi\sim N(\mu,\sigma^2)$，则

$$\eta=\frac{\xi-\mu}{\sigma}\sim N(0,1).$$

因而

$$P(\xi\leqslant x)=\Phi\left(\frac{x-\mu}{\sigma}\right).$$

证 对任何实数 y

$$P(\eta\leqslant y)=P\left(\frac{\xi-\mu}{\sigma}\leqslant y\right)=P(\xi\leqslant\sigma y+\mu)$$

$$=\int_{-\infty}^{\sigma y+\mu}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

作代换 $s=\dfrac{t-\mu}{\sigma}$，则当 $t=\sigma y+\mu$ 时，$s=y$，$dt=\sigma ds$，

$$P(\eta\leqslant y)=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds.$$

因而 $\eta\sim N(0,1)$. 又因为

$$P(\xi\leqslant x)=P\left(\frac{\xi-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right)=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

定理得证.

例 3.3.14 若随机变量 $\xi\sim N(\mu,\sigma^2)$，则由定理 3.3.13

及附表 2 得

$$(1)\,P\{|\xi-\mu|\leqslant\sigma\}=P\left(-1\leqslant\frac{\xi-\mu}{\sigma}\leqslant 1\right)=\Phi(1)-\Phi(-1)$$

$$=\Phi(1)-[1-\Phi(1)]=2\Phi(1)-1$$

$$=2\times 0.84134-1=0.68268;$$

$$(2)\,P\{|\xi-\mu|\leqslant 2\sigma\}=2\Phi(2)-1=0.95450;$$

$$(3)\,P\{|\xi-\mu|\leqslant 3\sigma\}=2\Phi(3)-1=0.99730.$$

此例告诉我们：(1)ξ 在以 μ 为中点，长度为 2σ 的区间 $[\mu-\sigma,\mu+\sigma]$ 内取值的概率约为 68.3%；(2)在 $[\mu-2\sigma,\mu+2\sigma]$ 内取值的概率大于 95%；(3)在 $[\mu-3\sigma,\mu+3\sigma]$ 内取值的概率大于 99.7%．也就是说 ξ 落在这个区间之外的概率小于 0.3%，这在实际上就认为是不可就发生的．这就是质量控制中的 3σ 原则．

例 3．3．15　质量控制的3σ 原则． 设在正常生产的情况下，某零件的尺寸 $\xi\sim N(\mu,\sigma^2)$，为了在生产过程中随时检查有无系统性误差 (即有异常情况)出现，人们画一个质量控制图(图3-8)，横轴表示时间，画三条虚线：$y=\mu$ 称为中线；$y=\mu+3\sigma$，称为上控制线；$y=\mu-3\sigma$ 称为下控制线．每隔一定时间，对产品尺寸进行检查，测量的产品尺寸应落在上、下控制线之内．如果点子超出控制线，则很可能是生产出现异常情况，应该暂停生产进行检查．当然也可能虚报，但虚报的可就性比较小，特别在不太长的时间内出现两次以上超出控制线，那就必须严加检查，直到查出原因，才能恢复生产．当然在质量控制图的具体使用上还有许多细节要考虑，而且常常不是抽检一个产品，而是每次抽检几只 (例如 5 只)，求出它们尺寸的平均值来制作控制图，而且有时还伴用其它的控制图．这里只是说明一下原理．此外，即或检查的产品按其尺寸所画的点子均在控制线内，如果出现大多数点子

均在中线以上(或以下)，也说明可能有系统的误差，有必要追查原因.

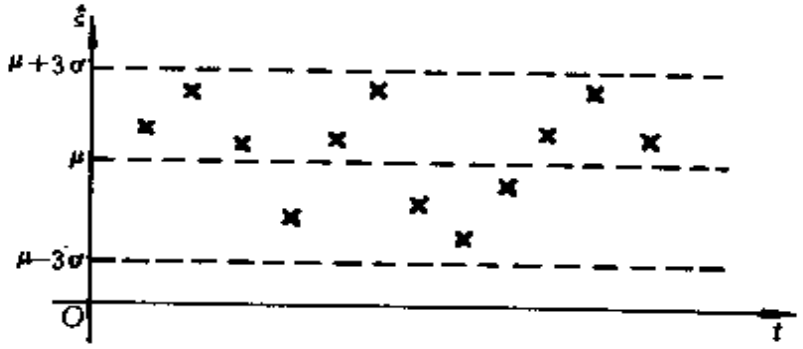


图 3-8

定义3.3.16 若连续型随机变量ξ的概率密度函数(如图3-9)

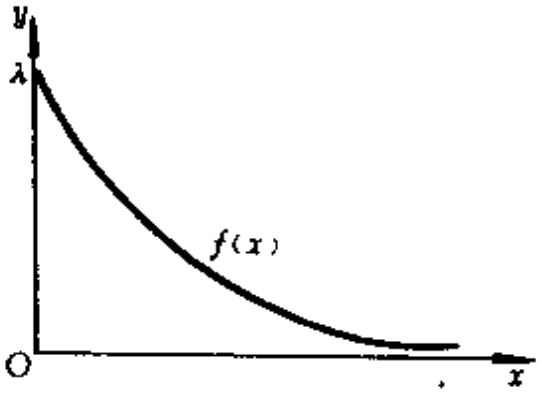


图 3-9

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

其中λ是正实数，则称ξ服从参数为λ的指数分布，或称ξ为具有参数λ的指数随机变量.而上述函数称为具参数λ的指数分布密度.

在例3.1.13中我们讲到泊松过程，即随时间演变的一个随机过程$\{N(t), t\geqslant 0\}$，其中$N(t)$表示直到时刻t“事件”发生的次数，它的分布列为

$$P(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k=0,1,2,\cdots.$$

如果我们用ξ表示第一次“事件”发生的时刻，或“事件”发生的等待时间，则

$$P(\xi>t)=P(N(t)=0)=e^{-\lambda t}$$

$$=\int_t^{+\infty}\lambda e^{-\lambda x}dx, \quad t\geqslant 0.$$

由此即知ξ服从参数为λ的指数分布.

在实际中，到某个特定“事件”发生所需的等待时间往往服从指数分布. 例如：从现在开始到一次地震发生的时间；某人到接了一次指指号码的电话所需要的等待时间；某些元件的寿命；某人打一个电话的持续时间；…… 往往可近似地看作服从指数分布.

指数分布具有下述性质.

定理3.3.17 若ξ服从参数为λ的指数分布，则对一切实数$s\geqslant 0$，$t\geqslant 0$

(3.3.18) $$P(\xi\geqslant s+t\mid\xi\geqslant s)=P(\xi\geqslant t)$$

恒成立.

证 $$P(\xi\geqslant s+t\mid\xi\geqslant s)=\frac{P(\xi\geqslant s+t,\ \xi\geqslant s)}{P(\xi\geqslant s)}$$

$$=\frac{P(\xi\geqslant s+t)}{P(\xi\geqslant s)}=\frac{\int_{s+t}^{\infty}\lambda e^{-x}dx}{\int_s^{\infty}\lambda e^{-x}dx}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(\xi \geqslant t).$$

称满足 (3.3.18)的随机变量为**无记忆的**. 如果某元件的寿命 ξ 服从指数分布，那么已知它用了 s 小时无损坏的条件下再使用 t 小时以上的概率和从一开始使用时算起它能工作 t 小时以上的概率相同. 所以可以形象地说指数分布表征“永远年青”. 下面我们证明指数分布是唯一具有无记忆性的连续型随机变量.

***定理 3.3.19** 若 ξ 为无记忆的连续型随机变量，则它必服从指数分布.

证 令 $G(t) = P(\xi \geqslant t)$，$t \geqslant 0$，则由 (3.3.18) 知

$$G(t)G(s) = G(t+s)$$

对一切 t，$s \geqslant 0$ 成立.

首先，对任何自然数 n

$$G(1) = G\left(\frac{1}{n} + \frac{n-1}{n}\right) = G\left(\frac{1}{n}\right)G\left(\frac{n-1}{n}\right) = G\left(\frac{1}{n}\right)G\left(\frac{1}{n}\right)G\left(\frac{n-2}{n}\right)$$

$$= \cdots = \left[G\left(\frac{1}{n}\right)\right]^n.$$

因而 $$G\left(\frac{1}{n}\right) = [G(1)]^{\frac{1}{n}}.$$

再者，对任何自然数 m，n 有

$$G\left(\frac{m}{n}\right) = \left[G\left(\frac{1}{n}\right)\right]^m = [G(1)]^{\frac{m}{n}}.$$

对任意 $t > 0$，存在一列有理数 $\{r_k\}$，$0 < r_1 < r_2 \cdots < r_k < \cdots$ 且 $r_k \to t$ $(k \to \infty)$. 因为

$$\{\xi \geqslant r_k\} \downarrow \bigcap_{k=1}^{\infty} \{\xi \geqslant r_k\} = \{\xi \geqslant t\},$$

由概率的上连续性得

$$G(t)=P(\xi\geqslant t)=P\left(\bigcap_{k=1}^{\infty}\{\xi\geqslant r_k\}\right)=\lim_{k\to\infty}P(\xi\geqslant r_k)$$

(3.3.20)

$$=\lim_{k\to\infty}G(r_k)=\lim_{k\to\infty}[G(1)]^{r_k}=[G(1)]^{t}.$$

条件(3.3..18)表明 $G(1)>0$，否则条件概率无意义．由于 ξ 为随机变量，故有

$$\lim_{t\to\infty}G(t)=\lim_{t\to\infty}P(\xi\geqslant t)=P(\xi=\infty)=0.$$

结合(3.3.20)有

$$G(1)<1.$$

令 $G(1)=e^{-\lambda}$，则 $\lambda>0$，且

$$G(t)=e^{-\lambda t},\qquad t\geqslant 0,$$

即

$$P(\xi\geqslant t)=\int_t^{\infty}\lambda e^{-\lambda s}\,ds,\ t\geqslant 0.$$

ξ 具有分布密度

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0,\\ 0, & \text{其它}.\end{cases}$$

例3.3.21 设某人打一次电话需用的时间 ξ 服从参数为 $\frac{1}{10}$ (单位：分)的指数分布，当你走进电话室需要打电话时，某人恰巧在你前面开始打电话．求以下几个事件的概率：

(1)你需要等待10分钟以上；

(2)你需要等待10—20分钟；

(3)你已经等了10分钟还需再等10分钟以上.

解 用ξ表示某人的通话时间，也就是你的等待时间. 则ξ的分布密度

$$f(x)=\begin{cases}\frac{1}{10}e^{-\frac{x}{10}}, & x\geqslant 0,\\ 0, & \text{其它}.\end{cases}$$

故所求概率分别为

$$(1)P(\xi>10)=\int_{10}^{\infty}\frac{1}{10}e^{-\frac{x}{10}}dx=-e^{-\frac{x}{10}}\Big|_{10}^{\infty}=e^{-1}\approx 0.368;$$

$$(2)P(10<\xi<20)=\int_{10}^{20}\frac{1}{10}e^{-\frac{x}{10}}dx=-e^{-\frac{x}{10}}\Big|_{10}^{20}$$

$$=e^{-1}-e^{-2}\approx 0.233;$$

$$(3)P(\xi>10+10\mid\xi>10)=P(\xi>10)\approx 0.368.$$

定义3.3.22 若连续型随机变量ξ的分布密度为

$$f(x)=\begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

其中$\Gamma(t)=\int_{0}^{\infty}e^{-y}y^{t-1}dy$，$t>0$，$\lambda>0$为伽玛函数，则称$\xi$服从以$(t,\lambda)$为参数的$\Gamma$－分布. 记作$\xi\sim\Gamma(t,\lambda)$，而称上述函数为具参数$(t,\lambda)$的$\Gamma$－分布的密度函数.

以$(1,\lambda)$为参数的Γ－分布，就是以λ为参数的指数分布.

又在泊松过程中，等待 n 个“事件”发生所需的时间就服从参数为 (n,λ) 的 Γ－分布（参见例3.5.8）. 或者说，n 个独立的参数为 λ 的指数分布的随机变量之和就是一个参数为 (n,λ) 的 Γ－分布的随机变量. 这一点，我们将在下一章§4.4予以证明.

还有一些在理论和应用中非常重要的随机变量的分布类型，我们将在以后引入.

习题 3－3

1. 对下列函数，求常数 c，使之成为概率密度函数：

(1) $f(x)=\dfrac{c}{1+x^2}$，　$-\infty<x<\infty$；

(2) $f(x)=\begin{cases} c\cos x, & -\dfrac{\pi}{2}\leqslant x\leqslant\dfrac{\pi}{2}, \\ 0, & \text{其它}; \end{cases}$

(3) $f(x)=\begin{cases} cx^2e^{-h^2x^2}, & x\geqslant 0, \\ 0, & x<0. \end{cases}$ $(h>0)$

2. 设 $\xi\sim N(3,9)$，求 a 使：

(1) $P(\xi<a)=0.95$；

(2) $P(|\xi-3|>3a)=0.01$.

3. 某城市每天用电量不超过百万度，以 ξ 表示每天的耗电率（即用电度数除以百万度），它具有密度函数

$$f(x)=\begin{cases} 12x(1-x)^2, & 0<x<1, \\ 0, & \text{其它}. \end{cases}$$

若该城市每天的供电量仅80万度，求供电量不足的概率，若每天供电量为90万度呢？

4. 从北郊某地乘车前往南区火车站搭火车有两条路线可走，第一条穿过市区，路程较短，但易受交通拥挤之影响，所需时间 $\xi\sim N(50,100)$；第二条路线沿环程公路走，路程较长，但意外阻塞少，所需时间 $\eta\sim N(60,16)$. (1)假如有70分钟可用，问应走哪条路线以较大的概率按时到达？(2)假如只有

65 分钟可用，又应走哪一条路线？

5．设工厂生产的轴承的直径 ξ 服从正态分布 $N(2.5, 0.0001)$．(1)轴承直径落在 (2.492，2.510)中的概率为多少？(2)ξ 以实际必然性 (即以 0.9973)的概率落在什么范围内？

6．如果 $\xi \sim U(0, 5)$，问方程 $4x^2+4\xi x+\xi+2=0$ 的两个根都是实根的概率为多少？

7．求下列分布密度中的规格化常数 c：

$$(1) f(x)=\begin{cases} \dfrac{c}{x}\exp\left\{-\dfrac{(\ln x-\mu)}{2\sigma^2}\right\}, & x>0, \\ 0, & x\leqslant 0 \end{cases}$$

(称此分布为对数正态分布)；

$$(2) f(x)=\begin{cases} c\,x^{p-1}\ (1-x)^{q-1}, & 0<x<1, \\ 0, & \text{其它}, \end{cases}$$

其中 $p>0$，$q>0$ 为常数 (称此分布为贝塔 (B)分布)；

$$(3) f(x)=c\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad -\infty<x<\infty.$$

§3.4 连续型随机变量的期望

为了给连续型随机变量的期望下一个合理的定义，以表示它取值的平均大小，我们先假定随机变量 ξ 的分布密度 $f(x)$ 只在 $[a, b]$ 上取非零值，并且是 $[a, b]$ 上的连续函数．取分点

$$a=a_0<a_1<a_2<\cdots<a_n=b,\ a_{i+1}-a_i=\frac{b-a}{n},$$

则随机变量 ξ 落在区间 $[a_i, a_{i+1})$ 中的概率为

$$P(\xi\in[a_i, a_{i+1}))=\int_{a_i}^{a_{i+1}} f(x)dx.$$

如果我们把 $\xi\in[a_i, a_{i+1})$ 近似地看成 $\xi=a_i$，则按离散型随机变量期望的定义知 ξ 取值的平均大小应近似于

$$\sum_{i=0}^{n-1} a_i \int_{a_i}^{a_{i+1}} f(x)dx = \sum_{i=0}^{n-1} a_i f(b_i)(a_{i+1}-a_i),$$

$$b_i \in (a_i, a_{i+1}).$$

当 n 越大，近似程度越好．当 $n \to \infty$ 时，上述和式的极限即为

$$\int_a^b x f(x)dx = \int_{-\infty}^{\infty} x f(x)dx.$$

因此我们给出下面的

定义 3.4.1 若连续型随机变量 ξ 的分布密度 $f(x)$，$-\infty < x < \infty$ 满足条件

$$\int_{-\infty}^{\infty} |x| f(x)dx < \infty$$

则称积分

$$\int_{-\infty}^{\infty} x f(x)dx \tag{3.4.2}$$

为随机变量 ξ (或分布密度 $f(x)$)的**期望或均值**，记作 $E\xi$．否则称 ξ 的期望不存在．

我们对连续型随机变量的分布及期望作一个力学比拟．若将连续型随机变量看成是一条总质量为 1 的、以密度 $f(x)$ 分布在 R.(即 $(-\infty, \infty)$)上的有限或无限长的棒，则 (3.4.2)定义的 $E\xi$ 表示此棒重心的坐标．在这个意义上 $E\xi$ 是随机变量 ξ 取值位置的中心．由这个直观解释，容易猜出某些随机变量的期望．例如对于服从在 $[a, b]$ 上均匀分布的随机变量 ξ，有 $E\xi = \frac{a+b}{2}$；对于服从正态分布 $N(\mu, \sigma^2)$的随机变量 η，有 $E\eta = \mu$，因为分

布密度关于 $x=\mu$ 对称．下面由计算得出这些结果．

例 3.4.3 若随机变量 $\xi \sim U(a, b)$，则

$$E\xi = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b+a}{2}.$$

例 3.4.4 若随机变量 $\xi \sim N(\mu, \sigma^2)$，则 $E\xi = \mu$．

解 首先需证 $E\xi$ 存在．由于

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

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

$$= |\mu| + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \left|\frac{x-\mu}{\sigma}\right| e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}\, d\frac{x-\mu}{\sigma}$$

$$= |\mu| + \frac{2\sigma}{\sqrt{2\pi}}\int_0^{\infty} u\, e^{-\frac{u^2}{2}}\, du$$

$$= |\mu| + \frac{2\sigma}{\sqrt{2\pi}}\left(-e^{-\frac{u^2}{2}}\right)\Big|_0^{\infty} = |\mu| + \frac{2\sigma}{\sqrt{2\pi}} < \infty,$$

因而 $E\xi$ 存在且

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

令 $u = \dfrac{x-\mu}{\sigma}$，则

$$E\xi=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma u+\mu)e^{-\frac{u^2}{2}}\,du$$

$$=\mu\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u\,e^{-\frac{u^2}{2}}\,du.$$

由于第二个积分的被积函数为奇函数，因而积分值为零，而第一项中 $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=1$，因而 $E\xi=\mu$.

例3.4.5 若 ξ 是参数为 λ 的指数随机变量，则

$$E\xi=\int_0^{\infty}x\cdot\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda}\int_0^{\infty}u\,e^{-u}\,du=\frac{1}{\lambda}.$$

其意义为：若 ξ 表示元件的寿命，服从参数为 λ 的指数分布，那么其平均寿命为 $\frac{1}{\lambda}$.

例3.4.6 若随机变量 $\xi\sim\Gamma(t,\lambda)$，即 ξ 具有分布密度

$$f(x)=\begin{cases}\dfrac{\lambda^t}{\Gamma(t)}x^{t-1}e^{-\lambda x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

则

$$E\xi=\int_0^{\infty}\frac{\lambda^t}{\Gamma(t)}x^t e^{-\lambda x}\,dx=\frac{1}{\Gamma(t)}\int_0^{\infty}(\lambda x)^t e^{-\lambda x}dx.$$

令 $\lambda x=u$，则由 $\Gamma(\alpha)=\int_0^{\infty}u^{\alpha-1}e^{-u}\,du$ 知

$$E\xi = \frac{1}{\lambda \Gamma(t)} \int_0^{\infty} u^t e^{-u} du = \frac{\Gamma(t+1)}{\lambda \Gamma(t)} .$$

再由 Γ − 函数的性质：$\Gamma(t+1) = t\Gamma(t)$知

$$E\xi = \frac{t}{\lambda} .$$

例 3.4.7 若ξ的概率分布密度为

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

这时因为

$$\int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx$$

$$= \lim_{A \to \infty} \frac{1}{\pi} \int_0^A \frac{d(x^2+1)}{x^2+1} = \lim_{A \to \infty} \frac{1}{\pi} \ln(1+x^2) \Big|_0^A$$

$$= \lim_{A \to \infty} \frac{1}{\pi} \ln(1+A^2) = \infty ,$$

所以 $E\xi$ 不存在.

注 若随机变量 ξ 具有分布密度

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

其中 $\lambda > 0, \mu$ 为任意实数. 则称 ξ 服从参数为(λ, μ)的柯西分布. 例 3.4.7 中的 ξ 服从参数为$(1,0)$的柯西分布. 对于参数为(λ, μ)的柯西分布，也可以仿照例 3.4.7 的方法证明它的期望不存在. 对于期望不存在的随机变量，为了用一个指标来刻划它取值的大小状况，在概率论与数理统计(特别是后者)中引进了中位数的概念，它的定义如下:

定义 3.4.8 称数μ为随机变量ξ的中位数，如果μ使

$$P(\xi \geqslant \mu) \geqslant \frac{1}{2}, \quad P(\xi \leqslant \mu) \geqslant \frac{1}{2}$$

同时成立.

随机变量的中位数总是存在的，但有时不唯一，例如ξ为离散型随机变量且分布列为

$$P(\xi=1)=\frac{1}{2}, \quad P(\xi=2)=\frac{1}{2},$$

这时一切$\mu \in [1,2]$都是它的中位数.

易知均匀分布、正态分布的期望与中位数相同，而柯西分布的中位数存在，但由前知它的期望不存在. 对期望存在的随机变量来说，中位数概念也是有意义的.

和离散型随机变量一样，我们可以用$E(\xi - E\xi)^2$来表征连续型随机变量ξ偏离平均值的程度，而把它称为ξ的方差. 关于它的存在性和计算公式，与离散型随机变量的情况类似(参见定义 3.2.7、定理 3.2.8 及命题 3.2.9). 这牵涉到连续型随机变量的函数的分布及期望的计算公式问题，它们将在 §3.6 加以讨论. 在那里我们将证明：设ξ为具有分布密度$f(x)$的连续型随机变量，$g(x), -\infty < x < \infty$是连续函数. 若$\int |g(x)| f(x)dx < \infty$，则$Eg(\xi)$存在且

$$Eg(\xi)=\int_{-\infty}^{\infty} g(x) f(x)\, dx. \tag{3.4.9}$$

因此我们给出

定义 3.4.10 设ξ为具有分布密度$f(x)$的连续型随机变量，若$E\xi$存在且

$$\int_{-\infty}^{\infty}(x-E\xi)^2 f(x)\,dx<\infty,$$

则称上式左端为 ξ 的**方差**，记作 $D\xi$ 或 $\mathrm{Var}\xi$. 由(3.4.9)知此时

$$D\xi=\mathrm{Var}\xi=E(\xi-E\xi)^2.$$

称 $\sqrt{D\xi}$ 为 ξ 的**标准差或均方差**.

若 $\int_{-\infty}^{\infty}|x|^n f(x)\,dx<\infty$ (n 为正整数)，则分别称

$$\int_{-\infty}^{\infty}x^n f(x)\,dx,$$

$$\int_{-\infty}^{\infty}(x-E\xi)^n f(x)\,dx,$$

$$\int_{-\infty}^{\infty}|x|^n f(x)\,dx$$

为 ξ 的 n 阶原点矩、n 阶中心矩、n 阶原点绝对矩，由假设条件知：此时 $E\xi$ 存在且上述三积分都绝对可积，且由(3.4.9)知它们分别为 $E\xi^n$、$E(\xi-E\xi)^n$、$E|\xi|^n$. 由 $D\xi$ 的定义易证

$$D\xi=E\xi^2-(E\xi)^2. \tag{3.4.11}$$

例3.4.12 若随机变量 $\xi\sim U(a,b)$，则由 $E\xi=\dfrac{a+b}{2}$，有

$$D\xi=\int_a^b x^2\,\frac{1}{b-a}\,dx-\left(\frac{a+b}{2}\right)^2$$

$$=\frac{b^3-a^3}{3(b-a)}-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}.$$

例 3.4.13 若随机变量 $\xi \sim N(\mu, \sigma^2)$，则 $D\xi = \sigma^2$.

证 因为 $E\xi = \mu$，故由定义知

$$D\xi = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\left(\frac{x-\mu}{\sigma}\right)$$

$$= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du.$$

由于

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = 2\int_{0}^{\infty} x^2 e^{-\frac{x^2}{2}} dx$$

$$= 2\left[x\left(-e^{-\frac{x^2}{2}}\right)\right]\Big|_0^{\infty} - \int_0^{\infty} \left(-e^{-\frac{x^2}{2}}\right) dx$$

$$= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi},$$

所以

$$D\xi = \sigma^2.$$

至此我们知道了正态分布 $N(\mu, \sigma^2)$ 中两个参数的意义，μ 表示期望，σ^2 表示方差.

例 3.4.14 若随机变量 ξ 服从参数为 λ 的指数分布，则

$$D\xi = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{1}{\lambda^2}\int_0^{\infty} u^2 e^{-u}\,du - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{1}{\lambda^2}\left(2\int_0^{\infty} u\,e^{-u}\,du\right) - \frac{1}{\lambda^2}\ ,$$

由于 $\int_0^{\infty} ue^{-u}\,du$ 等于参数为 1 的指数随机变量的期望，故等于 1．从而我们有

$$D\xi = \frac{1}{\lambda^2}\ .$$

由于柯西分布的期望不存在，因而方差也不存在．至于 $\Gamma-$ 分布的期望和方差，读者可以直接计算，也可以由 §3．7 期望与方差的性质来进行计算．

习题 3－4

1．设随机变量 ξ 具有密度函数：

(1) $f(x)=\begin{cases} x & 0<x\leqslant 1, \\ 2-x & 1\leqslant x<2, \\ 0 & \text{其它;} \end{cases}$ (称为三角形分布)

(2) $f(x)=\frac{1}{2a}e^{-\frac{|x-b|}{a}}\ ,-\infty<x<\infty;$ (称为拉普拉斯分布)

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

求 $E\xi$ 及 $D\xi$

2．设随机变量 ξ 的密度函数为

$$f(x)=\begin{cases} 6x(1-x), & 0\leqslant x\leqslant 1, \\ 0, & \text{其它.} \end{cases}$$

求 $P(|\xi-E\xi|<2\sqrt{D\xi}\ )$.

3．设随机变量 ξ 的分布密度函数为

$$f(x)=\begin{cases}\dfrac{x^n e^{-x}}{n!}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

求 $E\xi$ 与 $D\xi$．

4．求对数正态分布的随机变量 ξ 的期望与方差，其分布密度为：

$$f(x)=\begin{cases}\dfrac{1}{x\sqrt{2\pi}\,\sigma}\exp\left\{-\dfrac{1}{2\sigma^2}(\ln x-\mu)^2\right\}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

5．(1) 若随机变量 $\xi\sim U(-1,1)$，求 $E\xi^3$，$E\xi^4$；

(2) 若随机变量 $\xi\sim U(0,1)$，求 $E\xi^3$，$E\xi^4$．

6．求下列分布的期望与方差：

(1) 贝塔分布 $B(p,q)$，

$$f(x)=\begin{cases}\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, & 0<x<1,\\ 0, & \text{其它},\end{cases}$$

$p>0$，$q>0$ 为常数；

(2) 威布尔 (Weibull) 分布 $W(\lambda,a)$，

$$f(x)=\begin{cases}a\lambda x^{a-1}\exp(-\lambda x^a), & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

$\lambda>0$，$a>0$ 为常数

(注：威布尔分布在可靠性统计中常用)；

(3) 学生分布 $t(n)$ (亦称 $t-$ 分布)，

$$f(x)=\frac{1}{\sqrt{n\pi}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

§3.5 分布函数

对于离散型随机变量，我们用分布列来描述它取值的概率规律；对连续型随机变量，则用它的概率密度函数来描述它取值的概率规律．那么对一般的随机变量，如何描述它取值的概率规律呢？回忆随机变量的定义2.4.2知：若ξ是随机变量，则$\xi \leqslant x$为随机事件，它就应该有概率，如果对一切实数x，$P(\xi \leqslant x)$都知道了，那么对一切有限、无限的开、闭、半开半闭区间的概率也都能由概率的性质计算出来，因而ξ取值的概率规律也就完全掌握了．这就引导出下面的定义3.5.1．

1．分布函数

定义3.5.1 若ξ为随机变量，称函数

$$F(x)=P(\xi \leqslant x), \quad -\infty<x<\infty$$

为ξ的**概率分布函数**(简称**分布函数**)．

随机变量ξ的分布函数在b点的值$F(b)$表示ξ取小于或等于b的值的概率．概率分布函数具有以下特征性质．

定理3.5.2 若$F(x)$是随机变量ξ的分布函数，则

(i) 若$a<b$ 则$F(a) \leqslant F(b)$; (单调增性)

(ii) $\lim\limits_{x \to \infty} F(x)=1$，$\lim\limits_{x \to -\infty} F(x)=0$;

(iii) $\lim\limits_{x \to x_0^+} F(x)=F(x_0)$，对一切实数$x_0$．(右连续性)反之，若定义在$(-\infty,\infty)$上的实函数满足以上条件，则$F(x)$一定是某一随机变量的分布函数．

证 (i) 若$a<b$，则$\{\xi \leqslant a\} \subset \{\xi \leqslant b\}$，由概率的性质可知$P(\xi \leqslant a) \leqslant P(\xi \leqslant b)$．

(ii) 由(i)可知(ii)中的两个极限存在．因而由概率的下连续性，

$$\lim_{x\to\infty} F(x)=\lim_{n\to\infty} F(n)=\lim_{n\to\infty} P(\xi\leqslant n)$$

$$=P\left(\bigcup_{n=1}^{\infty}\{\xi\leqslant n\}\right)=P(\xi<\infty)=P(\Omega)=1$$

再由概率的上连续性

$$\lim_{x\to-\infty} F(x)=\lim_{n\to\infty} F(-n)=\lim_{n\to\infty} P(\xi\leqslant -n)$$

$$=P\left(\bigcap_{n=1}^{\infty}\{\xi\leqslant -n\}\right)=P(\xi=-\infty)=P(\phi)=0.$$

(iii)由 (i)

$$\lim_{x\to x_0^+} F(x)=\lim_{n\to\infty} F\left(x_0+\frac{1}{n}\right)=\lim_{n\to\infty} P\left(\xi\leqslant x_0+\frac{1}{n}\right).$$

由于$\left\{\xi\leqslant x_0+\frac{1}{n+1}\right\}\subset\left\{\xi\leqslant x_0+\frac{1}{n}\right\}$，$n=1,2,\cdots$，再由概率的上连续性，

$$\lim_{x\to x_0^+} F(x)=P\left(\bigcap_{n=1}^{\infty}\left\{\xi\leqslant x_0+\frac{1}{n}\right\}\right)=P(\xi\leqslant x_0)$$

$$=F(x_0).$$

定理的第二部分的证明，要用到概率空间的构造及测度论的一些知识，在此从略，有兴趣的读者可参看[15，第二章§3]．

为了更清楚地理解：ξ的分布函数描述了ξ取值的概率规律，再举几例说明如下：

(iv)对一切$a<b$，

$$P(a<\xi\leqslant b)=F(b)-F(a).$$

这一性质由$\{a<\xi\leqslant b\}=\{\xi\leqslant b\}\setminus\{\xi\leqslant a\}$，$\{\xi\leqslant a\}\subset\{\xi\leqslant b\}$及概率的性质而得．

$$(\mathrm{v})\ P(\xi<b)=P\left(\bigcup_{n=1}^{\infty}\left\{\xi\leqslant b-\frac{1}{n}\right\}\right)=\lim_{n\to\infty}P\left(\xi\leqslant b-\frac{1}{n}\right)$$

$$=\lim_{n\to\infty}F\left(b-\frac{1}{n}\right)=F(b-0).$$

一般地 $F(b)$不一定等于 $F(b-0)$，

$$F(b)=F(b-0)+P(\xi=b).$$

(vi) $P(\xi=b)=F(b)-F(b-0)$.

(vii) $P(a\leqslant\xi<b)=F(b-0)-F(a-0)$.

(viii) $P(a\leqslant\xi\leqslant b)=F(b)-F(a-0)$.

(ix) $P(a<\xi<b)=F(b-0)-F(a)$.

定理3.5.3 若ξ为离散型随机变量，具有分布列

$$P(\xi=x_k)=p_k,\ k=1,2,\cdots,$$

则ξ的分布函数

$$F(x)=\sum_{k:\,x_k\leqslant x}p_k,\quad -\infty<x<\infty.$$

证 对任意实数 x，ξ的分布函数在 x 处的值

$$F(x)=P(\xi\leqslant x)=P\left(\bigcup_{k:\,x_k\leqslant x}\{\xi=x_k\}\right)$$

$$=\sum_{k:\,x_k\leqslant x}P(\xi=x_k)=\sum_{k:\,x_k\leqslant x}p_k.$$

定理3.5.4 (1)ξ为连续型随机变量的充分必要条件是存在非负可积函数使得对一切实数 x

$$F(x)=\int_{-\infty}^{x}f(t)\,dt;$$

(2) 若$f(x)$在x_0点连续，则 $F(x)$在x_0点可导且

$F'(x_0)=f(x_0)$；

(3) 连续型随机变量的分布函数必连续.

证 若ξ为具有分布密度$f(x)$的连续型随机变量，则对一切$a\leqslant x$有

$$P(a<\xi\leqslant x)=\int_a^x f(t)dt.$$

令$a\to-\infty$，由概率的下连续性知

$$P(\xi\leqslant x)=\int_{-\infty}^{x} f(t)\,dt, \qquad \text{对一切 } x \text{ 成立}.$$

反之，若ξ的分布函数能表成上述积分，则对一切$a\leqslant b$都有

$$P(a<\xi\leqslant b)=F(b)-F(a)=\int_a^b f(x)dx.$$

因而ξ为具有分布密度$f(x)$的连续型随机变量.

(2)、(3) 是熟知的分析性质.

例3.5.5 设随机变量ξ的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ \dfrac{x}{2}, & 0\leqslant x<1,\\ \dfrac{2}{3}, & 1\leqslant x<2,\\ \dfrac{3}{4}, & 2\leqslant x<3,\\ 1, & x\geqslant 3.\end{cases}$$

$F(x)$的图形如图 3－10 所示．则

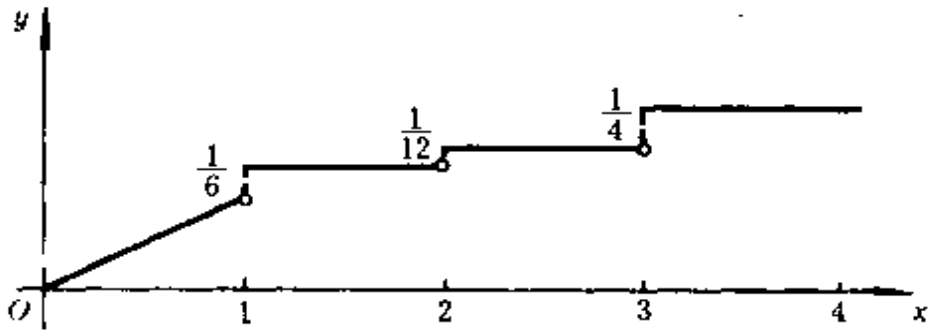


图 3－10

$$P(\xi<3)=\lim_{n\to\infty}P\left(\xi\leqslant 3-\frac{1}{n}\right)=\frac{3}{4},$$

$$P(\xi=\frac{1}{2})=F(\frac{1}{2})-F(\frac{1}{2}-0)=\frac{1}{4}-\lim_{x\to\frac{1}{2}}\frac{x}{2}=0,$$

$$P(2\leqslant\xi<4)=F(4-0)-F(2-0)=1-\frac{2}{3}=\frac{1}{3}.$$

注意：此例中的随机变量既非离散型，又非连续型．

由分布函数与分布密度的关系容易得到

例 3．5．6 若随机变量$\xi\sim U(a,b)$，则 ξ 的分布函数为

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

例 3．5．7 若随机变量ξ服从参数为 λ 的指数分布，则其分布函数

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

$$P(\xi>x)=e^{-\lambda x},\quad x>0.$$

现在我们阐述Γ－ 分布的一个来源.

*** 例3.5.8** 若$\{N(t):t\geqslant 0\}$为例3.1.13给出的参数是λ的泊松流，则对任何实数$t\geqslant 0$，

$$P(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\quad k=0,1,2,\cdots.$$

用η表示第n个“事件”发生的时间，即$\eta>0$，$N(\eta)=n$且当$t<n$时，$N(t)<n$. 由此易知$\{\eta\leqslant t\}=\{N(t)\geqslant n\}$. 于是当$t\leqslant 0$时$\eta$的分布函数$F_\eta(t)=P(\eta\leqslant t)=0$，当$t>0$时，

$$F_\eta(t)=P(\eta\leqslant t)=P(N(t)\geqslant n)$$

$$=\sum_{k=n}^{\infty}e^{-\lambda t}\frac{(\lambda t)^k}{k!}=e^{-\lambda t}\sum_{k=n}^{\infty}\frac{(\lambda t)^k}{k!}$$

$$=1-e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}$$

是t的连续函数，且由右端表达式可知$F_\eta(t)$可导且导数连续(当$t>0$时). 故

$$f(t)=F_\eta'(t)=\lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

$$=\frac{\lambda^n}{\Gamma(n)}t^{n-1}e^{-\lambda t},\quad t>0.$$

又由Γ－ 函数的定义易知

$$\int_0^{\infty}f(t)\,dt=1.$$

于是η为连续型随机变量，且服从参数为(n,λ)的Γ－ 分布.

即 $\eta \sim \Gamma(n,\lambda)$.

$$f(t)=\begin{cases}\dfrac{\lambda^n}{\Gamma(n)}t^{n-1}e^{-\lambda t}, & t\geqslant 0,\\ 0, & t<0.\end{cases}$$

*2. 期望

对于一般随机变量定义期望，要用到斯蒂尔杰斯 (Stieltjes) 积分，我们简单地介绍如下：

若$F(x)$是$[a,b]$上的单调函数，$g(x)$是定义在$[a,b]$上的函数，若对$[a,b]$的任一分割 $a=x_0<x_1<\cdots<x_n=b$ 及任意 $\xi_k\in(x_{k-1},x_k]$，当分割的最大子区间的长度 $\lambda\to 0$ 时

$$\sum_{k=1}^{n} g(\xi_k)[F(x_k)-F(x_{k-1})]\to I,$$

则称 I 为 g 关于 F 在 $(a,b]$ 上的斯蒂尔杰斯积分，记作

$$\int_{(a,b]} g(x)dF(x).$$

如果 $g(x)$，$F(x)$在 $(-\infty,\infty)$上有定义，对一切 $a<b$ 上述积分存在且

$$\lim_{\substack{b\to\infty\\ a\to-\infty}}\int_{(a,b]} g(x)dF(x)$$

存在，则称上述极限为 $g(x)$关于 $F(x)$在 $(-\infty,\infty)$上的斯蒂尔杰斯积分，记作

$$\int_{-\infty}^{\infty} g(x)dF(x).$$

定义 3.5.9 若ξ的分布函数为 $F(x)$，且

$$\int_{-\infty}^{\infty} |x| dF(x) < \infty$$

则称 $\int_{-\infty}^{\infty} x\, dF(x)$ 为 ξ 的期望，记作 $E\xi$．若 $E\xi$ 存在且

$$\int_{-\infty}^{\infty} (x - E\xi)^2\, dF(x) < \infty,$$

则称上式左端为 ξ 的方差，记作 $D\xi$ 或 $\mathrm{Var}\xi$．

习题 3－5

1．设离散型随机变量的分布列为

$$\begin{pmatrix} -2 & -1 & 0, & 1, & 2 \\ 0.1 & 0.2 & 0.2 & 0.4 & 0.1 \end{pmatrix},$$

求 ξ 的分布函数 $F(x)$，画出 $F(x)$ 的图象并求 $P(-2<\xi\leqslant 1)$ 及 $P(|\xi|<2)$．

2．进行 4 次独立射击，每次射击的命中率为 0.8．以 ξ 表示命中的次数．求 ξ 的分布函数，并求 $P(\xi<3)$，$P(1\leqslant\xi<4)$，$P(1\leqslant\xi\leqslant 3)$．

3．设 ξ 的密度函数为

(1) $f(x)=\begin{cases} cx^2, & 1\leqslant x\leqslant 2, \\ cx, & 2<x<3, \\ 0, & \text{其它}; \end{cases}$

(2) $f(x)=ce^{-\lambda|x|}$，　$-\infty<x<\infty$，　$\lambda>0$；

(3) $f(x)=\begin{cases} c\sin 3x, & \dfrac{\pi}{6}<x\leqslant\dfrac{\pi}{3}, \\ 0, & \text{其它}, \end{cases}$

求c及ξ的分布函数.

4. 若ξ的分布函数为

$$(1)\ F(x)=\begin{cases}1-(1+x)e^{-x}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$$

$$(2)\ F(x)=\begin{cases}0, & x\leqslant 0,\\ \sin x, & 0<x\leqslant \frac{\pi}{2},\\ 1, & x>\frac{\pi}{2};\end{cases}$$

$$(3)\ F(x)=\begin{cases}1-\dfrac{1}{(1+\frac{x}{2})^2}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

求ξ的分布密度.

5. 设$F_1(x)$, $F_2(x)$都是概率分布函数, $a>0$, $b>0$, $a+b=1$, 证明, $aF_1(x)+bF_2(x)$也是概率分布函数.

§3.6 随机变量的函数及其分布与期望

§3.2, §3.4中当我们讨论随机变量方差的时候, 遇到了随机变量ξ的函数$(\xi-E\xi)^2$, 并要求它的期望. 按定义, 应该首先计算$\eta=(\xi-E\xi)^2$的分布密度$f_\eta(x)$, 然后再计算$E\eta=\int xf_\eta(x)dx$. 而在那里, 我们用的是

$$E(\xi-E\xi)^2=\int_{-\infty}^{\infty}(x-E\xi)^2 f_\xi(x)dx,$$

这一节将说明这样定义的合理性.

若ξ是随机变量, $g(x)$是$(-\infty,\infty)$上的连续函数, 可以

证明 $g(\xi)$也是随机变量，即对一切实数 x，$\{g(\xi)\leqslant x\}$是随机事件．$g(\xi)$的分布函数可以由 ξ 的分布函数确定，即 $P(g(\xi)\leqslant x)=P(\xi\in\{y:g(y)\leqslant x\})$，对于一般的函数 $g(x)$，这并不是很容易的事，要用到测度论的知识．这里仅就一些特殊情形（也是最常用的情形）来研究这问题．

对于离散型随机变量 ξ，在 §3．3 我们已经看到，对任何函数 g，$g(\xi)$均为离散型随机变量，我们已经求出了它的分布列的表达式及 $Eg(\xi)$的表达式．现在我们集中讨论连续型随机变量，先从一些例子出发．

例 3．6．1 若 ξ 为具有分布密度 $f_\xi(x)$的连续型随机变量，$\eta=a\xi+b$，$a\neq 0$，试证 η 为连续型随机变量并求其分布密度．

解 若 $a>0$．

$$F_\eta(y)=P(a\xi+b\leqslant y)=P\left(\xi\leqslant\frac{y-b}{a}\right)=\int_{-\infty}^{\frac{y-b}{a}}f_\xi(x)dx.$$

作变量替换，令 $\frac{u-b}{a}=x$，则 $dx=\frac{du}{a}$，

$$F_\eta(y)=\int_{-\infty}^{y}\frac{1}{a}f_\xi\left(\frac{u-b}{a}\right)du$$

对一切 $y\in(-\infty,\infty)$都成立，因而 η 为连续型随机变量，且

$$f_\eta(x)=\frac{1}{a}f_\xi\left(\frac{x-b}{a}\right).$$

若 $a<0$，

$$F_\eta(y)=P(a\xi+b\leqslant y)=P\left(\xi\geqslant\frac{y-b}{a}\right)=\int_{\frac{y-b}{a}}^{\infty}f_\xi(x)\,dx$$

$$= \int_{y}^{-\infty} \frac{1}{a} f_{\xi}\left(\frac{u-b}{a}\right) du = -\int_{-\infty}^{y} \frac{1}{a} f_{\xi}\left(\frac{u-b}{a}\right) du$$

$$= \int_{-\infty}^{y} \frac{1}{|a|} f_{\xi}\left(\frac{u-b}{a}\right) du .$$

此时

$$f_{\eta}(x) = \frac{1}{|a|} f_{\xi}\left(\frac{x-b}{a}\right), \quad -\infty < x < \infty .$$

于是只要 $a \neq 0$，$\eta = a\xi + b$ 即为连续型随机变量，其分布密度为

$$f_{\eta}(x) = \frac{1}{|a|} f_{\xi}\left(\frac{x-b}{a}\right) .$$

当 $a = 0$ 时，η 为单点分布，$P(\eta = b) = 1$.

例 3.6.2 若随机变量 $\xi \sim N(\mu, \sigma^2)$，$a \neq 0$，$\eta = a\xi + b$，则 $\eta \sim N(a\mu + b, a^2\sigma^2)$. 特别，当 $\eta = \frac{\xi - \mu}{\sigma}$ 时，$\eta \sim N(0, 1)$.

证 由于

$$f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ,$$

则对 $\eta = a\xi + b$，由例 3.6.1 知

$$f_{\eta}(x) = \frac{1}{\sqrt{2\pi}\,|a|\sigma} e^{-\frac{\left(\frac{x-b}{a} - \mu\right)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}\,|a|\sigma} e^{-\frac{[x-(a\mu+b)]^2}{2a^2\sigma^2}} ,$$

故 $\eta \sim N(a\mu + b, a^2\sigma^2)$. 当 $\eta = \frac{\xi - \mu}{\sigma} = \frac{1}{\sigma}\xi - \frac{\mu}{\sigma}$ 时，

$a\mu+b=0$，$a^2\sigma^2=\dfrac{\sigma^2}{\sigma^2}=1$，即 $\eta\sim N(0,1)$.

例3.6.3 若ξ是连续型随机变量，分布密度为$f_\xi(x)$，且$f_\xi(x)$连续，试证$\eta=\xi^2$为连续型随机变量并求其分布密度.

解 由于$\eta=\xi^2\geqslant 0$，故当$y<0$时$F_\eta(y)=0$，当$y\geqslant 0$时，

$$F_\eta(y)=P(\xi^2\leqslant y)=P(-\sqrt{y}\leqslant\xi\leqslant\sqrt{y})$$

$$=\int_{-\sqrt{y}}^{\sqrt{y}}f_\xi(x)\,dx.$$

考虑到$f_\xi(x)$在$(-\infty,\infty)$上连续，对y求导即得

$$f_\eta(y)=F_\eta'(y)=\frac{1}{2}y^{-\frac{1}{2}}(f_\xi(\sqrt{y})+f_\xi(-\sqrt{y})),\ y>0.$$

故η为具分布密度

$$f_\eta(x)=\begin{cases}\dfrac{1}{2}x^{-\frac{1}{2}}(f_\xi(\sqrt{x})+f_\xi(-\sqrt{x})), & x>0,\\ 0, & x\leqslant 0\end{cases}$$

的连续型随机变量.

下面给出两个较为一般的定理.

定理3.6.4 若ξ是具有分布密度$f_\xi(x)$的连续型随机变量，$g(x)$为严格单调(增或减)、可微且$g'(x)\neq 0$的函数，$\eta=g(\xi)$，则η为连续型随机变量，且有分布密度

$$f_\eta(y)=\begin{cases}f_\xi(g^{-1}(y))|g^{-1}(y)]'|, & \text{若存在}x\text{使}y=g(x),\\ 0, & \text{若对一切}x,y\neq g(x).\end{cases}$$

证 若$g(x)$单调增，且可微，$g'(x)>0$，设y_1，y_2都是$g(x)$的值域中的数，则

$$P(y_1<\eta<y_2)=P(y_1<g(\xi)<y_2)$$

$$=P(g^{-1}(y_1)<\xi<g^{-1}(y_2))$$

$$= \int_{g^{-1}(y_1)}^{g^{-1}(y_2)} f_\xi(x)\,dx .$$

令 $x = g^{-1}(y)$，则 $dx = [g^{-1}(y)]' dy$ 且

$$P(y_1 < \eta < y_2) = \int_{y_1}^{y_2} f_\xi(g^{-1}(y))[g^{-1}(y)]' dy .$$

由于 $g(x)$ 单调连续，因而值域是区间，设为 (α, β)，于是

$$f_\eta(y) = \begin{cases} f_\xi(g^{-1}(y))\,[g^{-1}(y)]', & y \in (\alpha, \beta), \\ 0, & \text{其它}. \end{cases}$$

对于 $g(x)$ 单调减的情形留给读者证明.

定理 3.6.5 若 $g(x)$ 是分段单调可导的函数，即存在有限个或可数个区间 $[a_i, a_{i+1}]$，$i = \cdots, -n, \cdots, -1, 0, 1, 2, \cdots$ 使得在 $[a_i, a_{i+1}]$ 上 $g(x)$ 单调增或单调减，$g'(x) \neq 0$ 且将在此区间内函数 $y = g(x)$ 的反函数记为 $x = g_i^{-1}(y)$，相应地 y 的区间记为 $[\alpha_i, \beta_i]$，$\alpha_i = \min\{g(a_i), g(a_{i+1})\}$，$\beta_i = \max\{g(a_i), g(a_{i+1})\}$，则 $\eta = g(\xi)$ 的概率密度函数为

$$f_\eta(y) = \sum_{i=-\infty}^{\infty} f(g_i^{-1}(y))\,|[g_i^{-1}(y)]'|\, I_{(\alpha_i, \beta_i)}(y),$$

其中

$$I_{(a,b)}(x) = \begin{cases} 1, & a < x < b, \\ 0, & \text{其它}. \end{cases}$$

***证** 首先，我们证明

$$P(a_i \leqslant \xi < a_{i+1}, g(\xi) \leqslant y) = \begin{cases} 0, & y \leqslant \alpha_i, \\ \displaystyle\int_{\alpha_i}^{y} f_\xi(g_i^{-1}(y))\,|[g_i^{-1}(y)]'|\,dy, & \alpha_i < y < \beta_i, \\ P(a_i \leqslant \xi \leqslant a_{i+1}), & y \geqslant \beta_i . \end{cases}$$

等式中 $y \leqslant \alpha_i$ 及 $y \geqslant \beta_i$ 的情形是显然的，对于等式中 $\alpha_i < y < \beta_i$ 的部分，我们只证 $g'(x) I_{[a_i, a_{i+1})}(x) \geqslant 0$ 的情形 ($g'(x) I_{[a_i, a_{i+1})}(x) \leqslant 0$ 的情形可类似证明)．此时

$$P(a_i \leqslant \xi \leqslant a_{i+1}, g(\xi) \leqslant y) = P(\alpha_i \leqslant g(\xi) \leqslant \beta_i, g(\xi) \leqslant y)$$

$$= P(\alpha_i \leqslant g(\xi) \leqslant y) = P(a_i \leqslant \xi \leqslant g_i^{-1}(y))$$

$$= \int_{a_i}^{g_i^{-1}(y)} f_\xi(x)\,dx = \int_{\alpha_i}^{y} f_\xi(g_i^{-1}(g))[g_i^{-1}(y)]'\,dy,$$

上式中最后一步的证明与定理 3.6.4 的有关部分相同，因此我们有

$$F_\eta(y) = P(g(\xi) \leqslant y) = \sum_{i=-\infty}^{\infty} P(a_i \leqslant \xi \leqslant a_{i+1}, g(\xi) \leqslant y)$$

$$= \sum_{i=-\infty}^{\infty} P(a_i \leqslant \xi \leqslant a_{i+1}, \alpha_i \leqslant g(\xi) \leqslant \min\{y, \beta_i\})$$

$$= \sum_{i=-\infty}^{\infty} \int_{\alpha_i}^{\min\{y, \beta_i\}} f_\xi(g_i^{-1}(y))|(g_i^{-1}(y))'|\,dy\, I_{[a_i, \infty)}(y)$$

对 y 求导，有

$$F_\eta'(y) = \sum_{i: y \in (\alpha_i, \beta_i)} f_\xi(g_i^{-1}(y))|[g_i^{-1}(y)]'|$$

$$= \sum_i f_\xi(g_i^{-1}(y))|[g_i^{-1}(y)]'|\, I_{(\alpha_i, \beta_i)}(y).$$

例 3.6.6 已知随机变量 $\xi \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$，即

$$f_\xi(x) = \begin{cases} \frac{1}{\pi}, & -\frac{\pi}{2} < x < \frac{\pi}{2}, \\ 0, & \text{其它}. \end{cases}$$

求 $\eta=\cos\xi$ 的分布密度.

解 此题可以应用定理 3.6.5 求解，但是通过下述的直接计算方法更简洁明了些.

$$F_\eta(y)=P\left(-\frac{\pi}{2}<\xi<\frac{\pi}{2},\cos\xi\leqslant y\right)$$

$$=P\left(-\frac{\pi}{2}<\xi<0,\cos\xi\leqslant y\right)+P\left(0<\xi<\frac{\pi}{2},\cos\xi\leqslant y\right)$$

$$=\begin{cases}0, & y<0,\\ P\left(-\frac{\pi}{2}<\xi<-\cos^{-1}y\right)+P\left(\cos^{-1}y<\xi<\frac{\pi}{2}\right), & 0<y<1,\\ 1, & y\geqslant 1\end{cases}$$

$$=\begin{cases}0, & y\leqslant 0,\\ 1-\frac{2}{\pi}\cos^{-1}y, & 0<y<1,\\ 1, & y\geqslant 1.\end{cases}$$

故 η 的分布密度

$$f_\eta(y)=F_\eta'(y)=\begin{cases}-\frac{2}{\pi}(\cos^{-1}y)', & 0<y<1,\\ 0, & \text{其它}\end{cases}$$

$$=\begin{cases}\frac{2}{\pi}\frac{1}{\sqrt{1-y^2}}, & 0<y<1,\\ 0, & \text{其它}.\end{cases}$$

例 3.6.7 已知 $\xi\sim N(0,1)$，$\eta=\cos\xi$，求 η 的分布密度.

解 $y=\cos x$ 的单调区间为 $(2\pi i,\pi(2i+1))$，$i=\cdots,-1,$

0，1，…．若 $y\leqslant -1$，则 $F_\eta(y)=0$，$y\geqslant 1$，$F_\eta(y)=1$．若 $-1<y<1$ 则由定理 3．6．5 知

$$f_\eta(y)=\sum_{k=-\infty}^{\infty}\{f_\xi(\cos^{-1}y+2\pi k)\,|\,(\cos^{-1}y+2\pi k)'|$$
$$+f_\xi(2\pi(k+1)-\cos^{-1}y)|\,(2\pi(k+1)-\cos^{-1}y)'|\}$$

$$=\sum_{k=-\infty}^{\infty}\left\{f_\xi(\cos^{-1}y+2\pi k)\frac{1}{\sqrt{1-y^2}}+f_\xi(2\pi(k+1)\right.$$

$$\left.-\cos^{-1}y)\frac{1}{\sqrt{1-y^2}}\right\}$$

$$=\sum_{k=-\infty}^{\infty}\frac{1}{\sqrt{1-y^2}}[f_\xi(2\pi(k+1)-\cos^{-1}y)+f_\xi(2\pi k+\cos^{-1}y)]$$

$$=\sum_{k=-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-y^2}}\left(\exp\left\{-\frac{1}{2}(2\pi(k+1)-\cos^{-1}y)^2\right\}\right.$$

$$\left.+\exp\left\{-\frac{1}{2}(2\pi k+\cos^{-1}y)^2\right\}\right).$$

若 $y\notin(-1,1)$，则 $f_\eta(y)=0$．

对于连续型随机变量 ξ 的函数 $g(\xi)$的期望 $Eg(\xi)$，我们有类似于离散型随机变量的下述结果：

定理 3．6．8 若ξ的概率密度函数为$f_\xi(x)$，$\eta=g(\xi)$，$g(x)$是 $(-\infty,\infty)$上的连续函数，且

$$\int_{-\infty}^{\infty}|g(x)|f_\xi(x)\,dx<\infty,$$

则 $Eg(\xi)$存在且

$$Eg(\xi)=\int g(x)f_{\xi}(x)dx.$$

证 我们只对满足定理 3.6.5 条件的 $g(x)$ 证明，一般情形定理仍成立，只是要用到更深的理论.首先我们证明 $E\eta$ 存在：

$$\int_{-\infty}^{\infty}|y|f_{\eta}(y)dy=\sum_{i}\int_{\alpha_i}^{\beta_i}|y|f_{\xi}(g_i^{-1}(y))|(g_i^{-1}(y))'|dy.$$

分别对 $(g_i^{-1}(y))'>0$ 或 <0 进行变量替换 $x=g_i^{-1}(y)$，则当 $(g_i^{-1})(y))'>0$ 时，

$$\int_{\alpha_i}^{\beta_i}|y|f_{\xi}(g_i^{-1}(y))\ [g_i^{-1}(y)]'dy$$

$$=\int_{a_i}^{a_{i+1}}|g(x)|f_{\xi}(x)\,dx.$$

当 $[g_i^{-1}(y)]'<0$ 时

$$\int_{\alpha_i}^{\beta_i}|y|f_{\xi}(g_i^{-1}(y))|[g_i^{-1}(y)]\,|dy$$

$$=\int_{\beta_i}^{\alpha_i}|y|f_{\xi}(g_i^{-1}(y))\ (g_i^{-1}(y))'dy$$

$$=\int_{a_i}^{a_{i+1}}|g(x)|f_{\xi}(x)\,dx.$$

因为此时 $g(a_i)=\beta_i$，$g(a_{i+1})=\alpha_i$. 故

$$\int_{-\infty}^{\infty}|y|f_{\eta}(y)\,dy=\sum_{i}\int_{a_i}^{a_{i+1}}|g(x)|f_{\xi}(x)dx$$

$$= \int_{-\infty}^{\infty} |g(x)| f_\xi(x)dx < \infty,$$

即 $Eg(\xi)$存在．而证明

$$Eg(\xi) = E\eta = \int_{-\infty}^{\infty} y f_\eta(y)dy = \int_{-\infty}^{\infty} g(x) f_\xi(x)dx$$

的方法是完全类似的．

对于一般的具有分布函数 $F(x)$的随机变量 ξ，如果 $g(x)$是连续函数（甚至更一般的函数），$\eta = g(\xi)$．那么在

$$\int_{-\infty}^{\infty} |g(x)| dF_\xi(x) < \infty$$

的条件下，应用测度论的知识，可以得到

$$Eg(\xi) = E\eta = \int_{-\infty}^{\infty} g(x)dF_\xi(x). \tag{3.6.9}$$

此事实的证明已超出本书的范围，有兴趣的读者可以参看[15，第三章§4 定理4]．

习题 3－6

1. 设随机变量 ξ 的分布列为

$$\begin{pmatrix} 0 & \frac{\pi}{2} & \pi \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix},$$

求 $\eta_1 = \frac{2}{3}\xi + 2$ 与 $\eta_2 = \cos\xi$ 的分布列．

2. 设 ξ 服从参数为 λ 的指数分布，求下列随机变量的密度函数：

(1) $\eta = \sqrt{\xi}$；

(2) $\eta = \frac{1}{\lambda}\ln\xi$；

(3) $\eta=e^{-\xi}$.

3. 设圆的半径 $\xi\sim N(1000,0.25)$，求圆周长 η 及圆面积 ζ 的概率密度函数.

4. 设 $\xi\sim U(0,1)$，η 与 ξ 满足 $\operatorname{tg}(\frac{\pi}{2}\eta)=e^{\xi}$，$\eta\in(0,1)$，求 η 的密度函数.

5. 设 ξ 的密度函数为

$$f(x)=\begin{cases}\cos x, & x\in(0,\frac{\pi}{2}),\\ 0, & \text{其它}.\end{cases}$$

求 $\eta=\xi^2$ 的期望与方差.

6. 设 $\xi\sim N(\mu,\sigma^2)$，求 $\eta=\xi^3$ 的期望与方差.

7. 设 $\xi\sim U(0,1)$，$\Phi_{\mu,\sigma}(x)$ 为正态分布 $N(\mu,\sigma^2)$ 的分布函数，试证：若随机变量 η 满足 $\Phi_{\mu,\sigma}(\eta)=\xi$，则 $\eta\sim N(\mu,\sigma^2)$.

8. 设 $F(x)$ 连续且严格增，则 η 的分布函数为 $F(x)$ 当且仅当 $\xi=F(\eta)\sim U(0,1)$.

§3.7 期望和方差的性质、切比雪夫不等式

这一节讲述期望和方差的性质，因为这些性质大多涉及几个随机变量，需要用到多维随机变量的知识，暂且不能严格证明，因此许多性质我们述而不证，详细证明将在下一章给出. 另外，由于期望和方差的概念只是就离散型和连续型随机变量两种情形给出的，所以对于它们有关性质的证明，也只就这两种情形给出. 一般情形的证明，要用到进一步的知识，超出了本书的范围.

命题 3.7.1 若随机变量 $\xi=c$，即恒为常数，则 $E\xi=c$，$D\xi=0$.

证 $\xi=c$，则 ξ 服从单点分布 $P(\xi=c)=1$，按离散型随机变量期望的定义 $E\xi=c\cdot P(\xi=c)=c$，而 $D\xi=E\xi^2-(E\xi)^2=c^2\cdot P(\xi=c)-c^2=0$.

命题 3.7.2 若随机变量 ξ 的期望存在，c 为常数，则 $c\xi$ 的期望存在且 $E(c\xi)=c\cdot E\xi$；若 $D\xi$ 存在，则 $D(c\xi)=c^2D\xi$.

证 当 $c=0$ 时，$c\xi=0$，$c\cdot E\xi=c^2D\xi=0$，因而结论显然

成立．故以下设$c\neq 0$．

若ξ是离散型随机变量，其分布列为

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$$

则$c\xi$的分布列为

$$\begin{pmatrix} ca_1, & ca_2 & \cdots \\ p_1, & p_2 & \cdots \end{pmatrix}$$

由$E\xi$存在知$\sum_i |a_i|p_i<\infty$，故$\sum_i |c\,a_i|p_i=|c|\sum_i |a_i|p_i<\infty$，从而$E(c\xi)$存在且

$$E(c\xi)=\sum_i ca_i p_i=c\sum_i a_i p_i=cE\xi .$$

若$D\xi$存在，则$\sum_i (a_i-E\xi)^2 p_i<\infty$，从而

$$\sum_i (ca_i-cE\xi)^2 p_i=c^2\sum_i (a_i-E\xi)^2 p_i<\infty ,$$

即$E[c\xi-E(c\xi)]^2$存在，此即$D(c\xi)$且

$$D(c\xi)=c^2 D\xi .$$

若ξ是连续型随机变量，具分布密度$f(x)$，由$E\xi$存在知$\int |x|f(x)dx<\infty$，再由定理3．6．8及$\int |cx|f(x)dx=|c|\int_{-\infty}^{\infty}|x|f(x)dx<\infty$知$E(c\xi)$存在且

$$E(c\xi)=\int_{-\infty}^{\infty} cxf(x)dx=c\int_{-\infty}^{\infty} xf(x)\,dx=c\cdot E\xi .$$

由$D\xi$存在知$\int_{-\infty}^{\infty}(x-E\xi)^2 f(x)\,dx<\infty$，从而

$$\int_{-\infty}^{\infty}(cx-cE\xi)^2 f(x)\,dx<\infty .$$

再由定理 3.6.8 知 $D(c\xi)=E(c\xi-cE\xi)^2$ 存在且

$$D(c\xi)=\int_{-\infty}^{\infty}(cx-cE\xi)^2 f(x)\,dx=c^2 D\xi .$$

命题 3.7.3 若 ξ_1，ξ_2 为同一概率空间上的两个随机变量，且 $E\xi_i$ 存在，$i=1,2$. 则 $E(\xi_1\pm\xi_2)$ 存在且

$$E(\xi_1\pm\xi_2)=E\xi_1\pm E\xi_2 .$$

命题 3.7.4 若随机变量 $\xi_1,\xi_2,\cdots,\xi_n$ 独立且 $E\xi_i$ 存在，$i=1,2,\cdots,n$，则 $E(\xi_1\xi_2\cdots\xi_n)$ 存在且

$$E\xi_1\xi_2\cdots\xi_n=E\xi_1\cdot E\xi_2\cdots E\xi_n .$$

若 $\xi_1,\xi_2,\cdots,\xi_n$ 两两独立且 $D\xi_i$ 存在 $i=1,2,\cdots,n$，则

$$D(\xi_1+\xi_2+\cdots+\xi_n)=D\xi_1+D\xi_2+\cdots+D\xi_n .$$

这两个命题留待下一章证明.

命题 3.7.5 若 ξ 是非负随机变量，$E\xi$ 存在，则对任何 $\varepsilon>0$，

$$P(\xi\geqslant\varepsilon)\leqslant\frac{E\xi}{\varepsilon} .$$

证 若 ξ 为离散型随机变量，ξ 可取的值为 $\{a_i\}$，将 $\{a_i: a_i\geqslant\varepsilon\}$ 记成 $\{a_{i_1},a_{i_2},\cdots\}$ 则

$$\begin{aligned}P(\xi\geqslant\varepsilon)&=P(\xi=a_{i_1})+P(\xi=a_{i_2})+\cdots\\&\leqslant\frac{a_{i_1}}{\varepsilon}P(\xi=a_{i_1})+\frac{a_{i_2}}{\varepsilon}P(\xi=a_{i_2})+\cdots\\&=\frac{1}{\varepsilon}\sum_k a_{i_k}P(\xi=a_{i_k})\leqslant\frac{E\xi}{\varepsilon} .\end{aligned}$$

若 ξ 为连续型随机变量，分布密度为 $f(x)$，则

$$\begin{aligned}P(\xi\geqslant\varepsilon)&=\int_{\varepsilon}^{\infty}f(x)dx\leqslant\int_{\varepsilon}^{\infty}\frac{x}{\varepsilon}f(x)dx\\&\leqslant\frac{1}{\varepsilon}\int_0^{\infty}xf(x)dx=\frac{E\xi}{\varepsilon} .\end{aligned}$$

对于一般情形，也可类似地用斯蒂尔杰斯积分证明，细节从略．

定理 3．7．6　切比雪夫不等式　对于随机变量 ξ，若 $D\xi$ 存在，则

$$P(|\xi-E\xi|\geqslant\varepsilon)\leqslant\frac{D\xi}{\varepsilon^2}\ .$$

证　因为 $P(|\xi-E\xi|\geqslant\varepsilon)=P(|\xi-E\xi|^2\geqslant\varepsilon^2)$，$|\xi-E\xi|^2$ 是非负随机变量，故由命题 3．7．5 知切比雪夫不等式成立．

切比雪夫不等式是个重要的不等式，它在证明大数定律时起了重要的作用．

命题 3．7．7　(1) 若 ξ 是非负随机变量且 $E\xi=0$，则 $P(\xi=0)=1$；(2) 若 $D\xi=0$ 则 $P(\xi=E\xi)=1$．

证　由于 $D\xi=E(\xi-E\xi)^2$，从而 (2) 是 (1) 的推论．今证 (1)．

$$P(\xi=0)=1-P(\xi>0)\ ,$$

$$0\leqslant P(\xi>0)=P\left(\bigcup_n\left\{\xi\geqslant\frac{1}{n}\right\}\right)=\lim_{n\to\infty}P\left(\xi\geqslant\frac{1}{n}\right).$$

但由命题 3．7．5，$P\left(\xi\geqslant\frac{1}{n}\right)\leqslant\frac{E\xi}{\frac{1}{n}}=0$，$n=1,2,\cdots$．故

$$P(\xi>0)=0\ .$$

从而

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

命题 3．7．8　对于随机变量 ξ，$D\xi=E\xi^2-(E\xi)^2$．

证　由方差的定义知

$$D\xi=E(\xi-E\xi)^2=E(\xi^2-2\xi\ .\ E\xi+(E\xi)^2)\ .$$

$D\xi$ 存在的充分必要条件是 $E\xi^2$ 存在，且 $E\xi^2$ 存在时，$E\xi$ 也存在．

这时由期望的性质 (命题 3．7．3)

$$D\xi = E\xi^2 - 2E\xi \cdot E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2 .$$

例 3．7．9 若随机变量$\xi_1, \xi_2, \cdots, \xi_n$独立，且都服从同一 0—1 分布，即每一 ξ_i的分布列为

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

则 $\xi = \xi_1 + \xi_2 + \cdots + \xi_n \sim B(n, p)$，且由于

$$E\xi_i = p, \quad D\xi_i = p(1-p),$$

于是利用期望与方差的性质 (命题 3．7．3 及命题 3．7．4)

$$E\xi = E(\xi_1 + \xi_2 + \cdots + \xi_n) = E\xi_1 + E\xi_2 + \cdots + E\xi_n = np,$$

$$D\xi = D(\xi_1 + \xi_2 + \cdots + \xi_n) = D\xi_1 + D\xi_2 + \cdots + D\xi_n = np(1-p).$$

这比以前求 $B(n, p)$的期望与方差的方法要简单一些．

习题 3－7

1．若 $\xi \sim N(1, 2)$，$\eta \sim U(0, 2)$且 ξ 与 η 独立，求 $E(\xi - \eta^2)$与 $D(\xi - \eta^2)$．

2．设 $\xi_1, \xi_2, \cdots, \xi_n$独立，且有相同的期望 a 与相同的方差 σ^2，求

$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ 的期望与方差．

3．若事件 A 在第 i 次出现的概率为 p_i，以 η 表示 A 在前 n 次独立试验中出现的次数，求 $E\eta$ 及 $D\eta$．

4．从$\{1, 2, \cdots, n\}$这 n 个数中有放回地取出 k 个，求所得数之和的期望与方差．

5．每次试验中事件 A 发生的概率为 0.5，如果做 100 次独立试验，以 ξ 表示 A 发生的次数．试用切比雪夫不等式估计 $P(40 \leqslant \xi \leqslant 60)$．

6．设 $\xi \sim N(\mu, \sigma^2)$，试证：对任何 $\varepsilon > 0$，

$$P(|\xi-\mu| \geqslant \varepsilon\sigma) \leqslant \sqrt{\frac{2}{\pi}}\ \varepsilon^{-1} e^{-\frac{\varepsilon^2}{2}},$$

并比较此估计与切比雪夫不等式估计的优劣．

复习题 3－A

1．设 ξ 为取非负整数值的随机变量，若 ξ 的分布满足条件：对任何非负整数值 m、k 有

$$P(\xi = m+k \mid \xi \geqslant m) = P(\xi = k),$$

则称 ξ 的分布具有无记忆性．试证：取非负整数值的随机变量的分布具有无记忆性当且仅当它服从几何分布．

2．商店中出售某种商品，据历史记录分析，每月销售量服从参数为 7 的泊松分布．问在月初进货时要库存多少件此种商品，才能以 0.999 的概率满足顾客的需要．

3．设 ξ 只能取正整数值，相应的概率按等比数列递减．为了使 $E\xi = 10$，数列的首项与公比应取为多少？在这些条件下计算 $P(\xi \leqslant 10)$．

4．某装置在 $(0, t)$ 时间间隔内发生故障的次数是泊松过程，若已知此装置在 1000 小时工作中发生故障的次数的期望为 10，求它在 100 小时中发生故障的概率．

5．求负二项分布 $f(k; r, p)$ 的母函数，并由此求该分布的期望与方差．

6．一个部件正常运行的时间(单位为月)，是参数为 $\lambda = \frac{\alpha}{2}$ 的指数随机变量．今有一个由一个部件及相同的备件组成的系统．(1)问这一系统能正常运行的时间 ξ 应服从什么样的分布？(2)求这一系统至少能正常运行 5 个月的概率；(3)四个这样的独立系统中至少有一个的运行时间不少于 10 个月的概率是多少？

7．已知某公共汽车到达车站的时刻在10 点到 10 点半之间均匀分布．如果你 10 点钟到达车站，你将等候 10 分钟以上的概率是多少？如果到 10 点 15 分公共汽车还没来，你至少还要再等 10 分钟的概率是多少？

8．在半径为 R 的圆中投一点，它落在圆内任一区域的概率与该区域的面积成正比．设 ξ 表示这点到圆心的距离，求 ξ 的分布函数与分布密度，并求 $P\left(\frac{R}{3}<\xi\leqslant\frac{R}{2}\right)$．

9．设 ξ 的密度函数的曲线如图 3－11 所示，其中 $a>0$．

(1) 写出密度函数的表达式，求出 h；

(2) 求分布函数 $F(x)$；

(3) 求 $P\left(\frac{a}{2}<\xi\leqslant a\right)$．

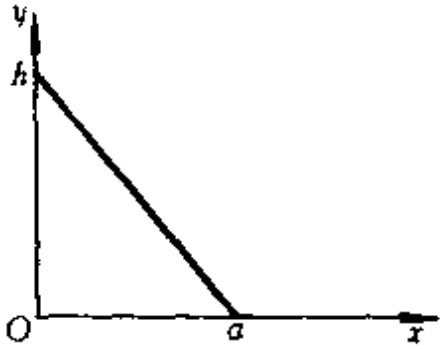


图 3－11

10．在半径为 R，中心在坐标原点的圆周上任取一点 (即该点的极角服从 $(-\pi,\pi)$ 上的均匀分布)．求

(1) 该点的横坐标的密度函数；

(2) 连接该点与 $(-R,0)$ 所成弦的长度的密度函数．

11．连接半径为 R 的圆周上一已知点与圆周上任一点 (即该点的极角服从 $(-\pi,\pi)$ 上的均匀分布)的弦长的期望值是什么？

12．(1) 证明：若对于 ξ，$Ee^{a\xi}$ 存在 ($a>0$ 为常数)，则对任意 $\varepsilon>0$，

$$P(\xi\geqslant\varepsilon)\leqslant\frac{Ee^{a\xi}}{e^{a\varepsilon}}\ ;$$

(2) 若 $g(x)$ 为 $(0,\infty)$ 上正的非降函数，且 $Eg(|\xi-E\xi|)$ 存在，则对任意 $\varepsilon>0$，

$$P(|\xi-E\xi|\geqslant\varepsilon)\leqslant\frac{Eg(|\xi-E\xi|)}{g(\varepsilon)}\ .$$

13．(1) 设 $\xi_1,\xi_2,\cdots,\xi_n$ 独立同分布，$E\xi_1=a$，$D\xi_1=\sigma^2$．证明对任意 $\varepsilon>0$，

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|<\varepsilon\right)\geqslant1-\frac{\sigma^2}{n\varepsilon^2}\ ;$$

(2) 利用上式估计：掷一枚硬币，要掷多少次才能保证正面出现的频率在 0.4 至 0.6 之间的概率不小于 90% (假设硬币是均匀的)？

复习题 3－B

1．设 ξ 是取非负整数值的随机变量，证明：

(1) $E\xi=\sum_{n=1}^{\infty} P(\xi\geqslant n)$；

(2) $D\xi=2\sum_{n=1}^{\infty} n\,P(\xi\geqslant n)-E\xi\,(E\xi+1)$.

2．袋中有 N 个球，其中白球数 ξ 为随机变量，其数学期望为 n.

(1) 求从袋中摸一球得到白球的概率；

(2) 若 ξ 的方差为 σ^2，从袋中一次摸出两球，求恰含一个白球的概率.

3．数量 N 很大的一群人接受一种血液化验，这种化验可以采用两种办法进行．一种是每个人的血液单独化验，这时总共需化验 N 次．第二种是把 N 个人随机地分成若干组，每组 k 个人，把这 k 个人每人的血液取出一半混在一起进行化验．如果这混合血液呈阴性，就说明每个人的血液都呈阴性，这样对这 k 个人只要数一次化验，就能断定他们都没有这种病；如果呈阳性，那就说明这 k 个人中至少有一个人的血液呈阳性．这时对每个人余下的一半血液逐个再化验一次以确定哪些人是患者，在这种情况下，对这 k 个人共数 $k+1$ 次化验．假定对每个人来说，化验呈阳性的概率皆为 p，且各人的化验结果在概率意义上是彼此独立的．问：

(1) 指定的 k 个人的混合血液(假定无交互作用)化验呈阳性的概率是多少？

(2) 采用第二种化验办法时，N 个人总共所需的化验次数 ξ 的期望是多少(假定 N 是 k 的倍数)？

(3) k 取什么值时所需化验次数的期望达到最小？

4．(1) 设 $D\xi$ 存在，证明对任意常数 c，$D\xi\leqslant E(\xi-c)^2$；(2) 若 ξ 只在有限区间 $[a,b]$ 中取值．证明：$a\leqslant E\xi\leqslant b$，$D\xi\leqslant\left(\frac{b-a}{2}\right)^2$ (对离散型和连续型随机变量证明)．

5．在半圆的直径上任取一点 P，过点 P 作直径的垂线交圆周于 Q．设圆的半径为 R，试证 PQ 的期望为 $\frac{\pi}{4}R$．但若在半圆的圆周上任取一点 P，

过 P 作直径的垂线交直径于 Q，则 PQ 的期望为 $\frac{2}{\pi}R$.

*6. 设$\{F(x), F_n(x), n\geqslant 1\}$是一列概率分布函数，以 D 表示 $F(x)$的连续点作成的集合，若对任意 $a, b\in D$ 有

$$\lim_{n\to\infty}[F_n(b)-F_n(a)]=F(b)-F(a),$$

证明对任意 $\varepsilon>0$，存在 $c, d\in D$ 及正整数 N，使当 $n\geqslant N$ 时

$$F_n(d)-F_n(c)>1-\varepsilon.$$

7. 设$\{\xi_n, n\geqslant 1\}$为一列随机变量，有相同的期望 a，$D\xi_n=\sigma_n^2$，且 $\lim_{n\to\infty}\sigma_n=0$. 试用切比雪夫不等式证明对 $(-\infty, \infty)$上任意有界连续函数 $f(x)$有

$$\lim_{n\to\infty}Ef(\xi_n)=f(a).$$

第四章　多维随机变量

当我们研究客观世界中的随机现象时，往往关心的不是一个指标，而常常是相互联系着的多个指标．例如在打靶试验中，有时只关心弹着点与目标中心位置的距离，而有时为了有效地矫正试验者的射击技术，我们要考察弹着点是上下偏离目标还是左右偏离目标，这就要同时考察弹着点的横坐标 ξ 与纵坐标 η．它们都是同一随机试验的样本空间上的函数，即同一概率空间中的随机变量．又如要考察一个班级的健康状况，随机地选取一个学生，测量他的身高 ξ，体重 η，那么 ξ，η 就是在这一随机试验中同时考虑的两个随机变量．再如要考察花粉在水中的随机运动，要同时考察它在某一时刻 t 在直角坐标系中三个坐标轴方向上的速度等．我们不仅要讨论这些随机变量本身，还要讨论它们之间的相互关系．事实上，§2.4 所讨论的几个独立随机变量，就是在同一概率空间上具有特殊相互关系的随机变量．

§4.1　多维随机变量的定义及其概率分布

首先讨论多维随机变量的概念及其性质．

定义 4.1.1　设 ξ_1，ξ_2，$\cdots$，ξ_n 是同一概率空间 $(\Omega, \mathscr{F}, P)$ 中的 n 个随机变量，则称由它们构成的向量 $(\xi_1, \xi_2, \cdots, \xi_n)$ 为 n 维随机变量．

例 4.1.2　设 (ξ, η) 表示打靶时弹着点的位置，其中 ξ 为某一平面直角坐标系中的横坐标，η 为纵坐标，则 (ξ, η) 为二维随机变量．

例 4.1.3　设 ξ，η，ζ 为花粉在液体中运动时在三个坐标轴方向上的速度，由于花粉受液体分子的随机碰撞，(ξ, η, ζ) 为一

三维随机变量.

定义 4.1.4 若 $(\xi_1,\xi_2,\cdots,\xi_n)$为某一概率空间中的 n 维随机变量，称 n 元函数

$$F(x_1,x_2,\cdots x_n)=P\{\xi_1\leqslant x_1,\xi_2\leqslant x_2,\cdots,\xi_n\leqslant x_n\},-\infty<x_i<\infty,i=1,2,\cdots,n$$

为 n 维随机变量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的**联合概率分布函数**，简称为**分布函数**.

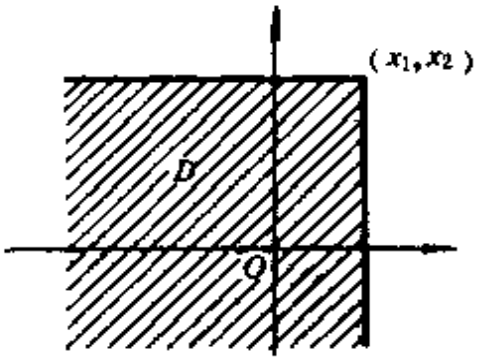


图 4-1

如图 4-1 所示，$F(x_1,x_2)$就是 (ξ_1,ξ_2)在区域 D 中取值的概率.

例 4.1.5 将一枚硬币投掷两次，ξ_1 表示第一次投掷时正面向上的次数，ξ_2 表示第二次投掷时正面向上的次数，试求二维随机变量(ξ_1,ξ_2)的分布函数.

解 在此随机试验中 (ξ_1,ξ_2)可能取的值是 (0,0)，(0,1)(1,0)，(1,1)，取每个值相应的概率是

$$P(\xi_1=0,\xi_2=0)=\frac{1}{4},\quad P(\xi_1=0,\xi_2=1)=\frac{1}{4},$$

$$P(\xi_1=1,\xi_2=0)=\frac{1}{4},\quad P(\xi_1=1,\xi_2=1)=\frac{1}{4}.$$

当 x_1,x_2 中有一个小于 0 时，如图 4-2，

$$P(\xi_1\leqslant x_1,\xi_2\leqslant x_2)=P(\varnothing)=0.$$

当 x_1,x_2 都大于或等于 1 时，如图 4-3，

$$P(\xi_1\leqslant x_1,\xi_2\leqslant x_2)=P(\Omega)=1.$$

当 $0\leqslant x_1<1,0\leqslant x_2<1$ 时，如图 4-4，

$$P(\xi_1\leqslant x_1,\xi_2\leqslant x_2)=P(\xi_1=0,\xi_2=0)=\frac{1}{4}.$$

当 $0\leqslant x_1<1,x_2\geqslant 1$ 时，如图 4-5，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(\{\xi_1 = 0, \xi_2 = 0\} \cup \{\xi_1 = 0, \xi_2 = 1\})$$
$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

当 $x_1 \geqslant 1$，$0 \leqslant x_2 < 1$ 时，如图 4－6，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(\{\xi_1 = 0, \xi_2 = 0\} \cup \{\xi_1 = 1, \xi_2 = 0\})$$
$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

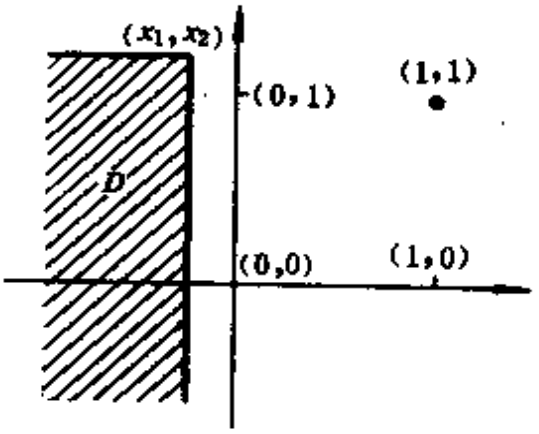


图 4－2

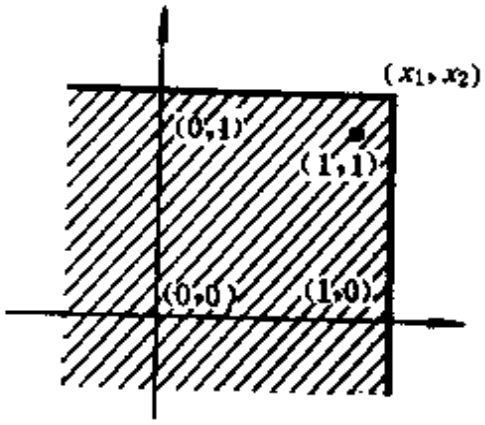


图 4－3

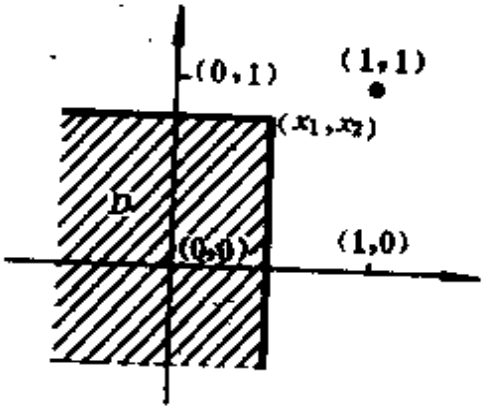


图 4－4

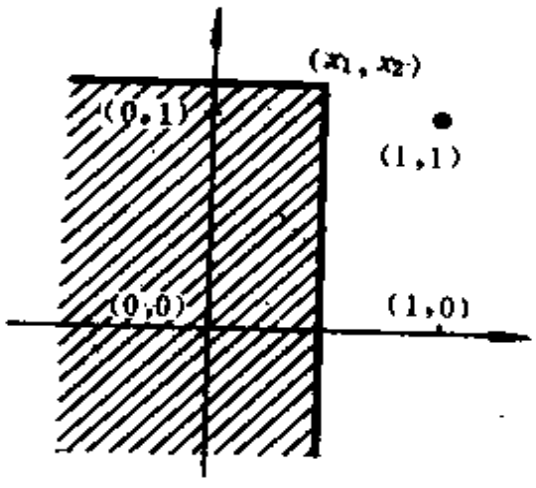


图 4－5

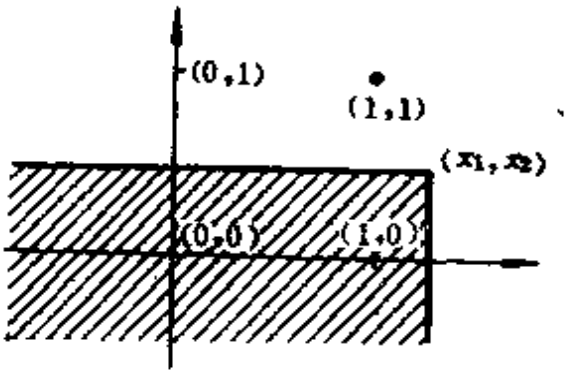


图 4-6

所以

$$F(x_1, x_2) = \begin{cases} 0, \text{当 } x_1<0 \text{ 或 } x_2<0, \\ \frac{1}{4}, \text{当 } 0\leqslant x_1<1, 0\leqslant x_2<1, \\ \frac{1}{2}, \text{当 } 0\leqslant x_1<1, x_2\geqslant 1, \\ \frac{1}{2}, \text{当 } x_1\geqslant 1, 0\leqslant x_2<1, \\ 1, \text{当 } x_1\geqslant 1, x_2\geqslant 1. \end{cases}$$

例 4.1.6 甲、乙两只船在时间间隔[0，24)内独立到达某个码头，每个船到达的时间是任意的，用 ξ_1 表示甲船到达的时刻，ξ_2 表示乙船到达的时刻．试求二维随机变量 (ξ_1, ξ_2) 的联合分布函数．

解 这一随机试验相当于在区域

$$G=\{(g_1, g_2): 0\leqslant g_1, g_2<24\}$$

内随机地取点，即概率 P 为 G 上的几何概率，$\xi_1(g_1, g_2)=g_1$，$\xi_2(g_1, g_2)=g_2$．

当 $x_1<0$ 或 $x_2<0$ 时，

$$P(\xi_1<x_1, \xi_2<x_2)=P(\{(g_1, g_2)\in G: g_1<0, g_2<0\})$$

$$= P(\varnothing) = 0 .$$

当 $0 \leqslant x_1 < 24$，$0 \leqslant x_2 < 24$ 时，如图 4－7，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(\{(g_1, g_2) \in G : g_1 \leqslant x_1, g_2 \leqslant x_2\})$$

$$= \frac{\text{区域}\{(g_1, g_2) : 0 \leqslant g_1 \leqslant x_1, 0 \leqslant g_2 \leqslant x_2\}\text{的面积}}{\text{区域 } G \text{ 的面积}} = \frac{x_1 x_2}{24^2} .$$

当 $0 \leqslant x_1 < 24$，$x_2 \geqslant 24$ 时，如图 4－8，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(\{(g_1, g_2) \in G : 0 \leqslant g_1 \leqslant x_1, 0 \leqslant g_2 < 24\})$$

$$= \frac{\text{区域}\{(g_1, g_2) : 0 \leqslant g_1 \leqslant x_1, 0 \leqslant g_2 < 24\}\text{的面积}}{\text{区域 } G \text{ 的面积}} = \frac{x_1}{24} .$$

同理当 $x_1 \geqslant 24$，$0 \leqslant x_2 < 24$ 时，如图 4－9，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(\{(g_1, g_2) : 0 \leqslant g_1 < 24, 0 \leqslant g_2 \leqslant x_2\})$$

$$= \frac{x_2}{24} .$$

当 $x_1 \geqslant 24$，$x_2 \geqslant 24$ 时，

$$P(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = P(G) = 1 .$$

所以 (ξ_1, ξ_2) 的分布函数

$$F(x_1, x_2) = \begin{cases} 0, \text{当 } x_1 \leqslant 0 \text{ 或 } x_2 < 0; \\ \dfrac{x_1 x_2}{24^2}, \text{当 } 0 \leqslant x_1 < 24, 0 \leqslant x_2 < 24; \\ \dfrac{x_1}{24}, \text{当 } 0 \leqslant x_1 < 24, x_2 \geqslant 24; \\ \dfrac{x_2}{24}, \text{当 } x_1 \geqslant 24, 0 \leqslant x_2 < 24; \\ 1, \text{当 } x_1 \geqslant 24, x_2 \geqslant 24. \end{cases}$$

为表述方便，以下我们主要研究二维随机变量分布函数的性质，读者不难把它们推广到任意 n 维随机变量的分布函数上．

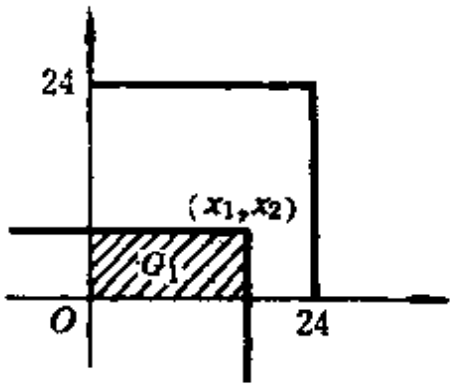


图 4－7

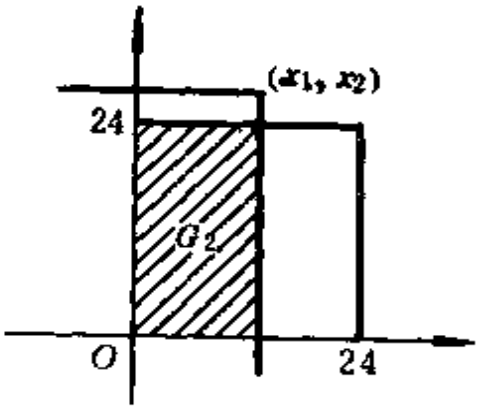


图 4－8

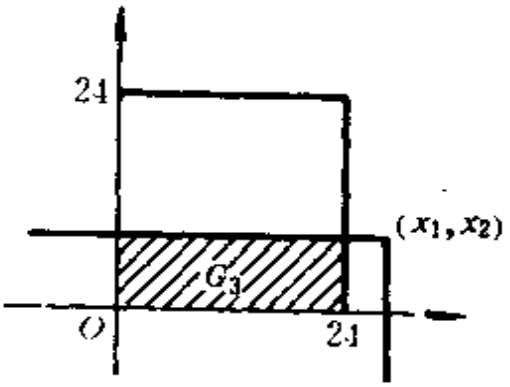


图 4－9

定理 4.1.7 二维随机变量 (ξ, η)的分布函数 $F(x, y)$具有以下性质：

(1) $F(x, y)$分别是变量 x 和 y 的非减函数，即对任意固定的实数 y，当 $x_2 > x_1$ 时 $F(x_2, y) \geqslant F(x_1, y)$；对于任意固定的实数 x，当 $y_2 > y_1$ 时 $F(x, y_2) \geqslant F(x, y_1)$；

(2) 对一切二维向量 $a = (a_1, a_2)$，$b = (b_1, b_2)$，$a_1 \leqslant b_1$，$a_2 \leqslant b_2$ 都有

$$\Delta_{b,a}F \overset{*)}{=} F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geqslant 0;$$

*). $\Delta_{b,a}F$ 表示函数 $F(x, y)$的二阶差分，即 $\Delta_{b,a}F = \Delta^x_{b_1,a_1}\Delta^y_{b_2,a_2}F(x, y)$
$= \Delta^x_{b_1,a_1}(F(x, b_2) - F(x, a_2)) = [F(b_1, b_2) - F(a_1, b_2)] - [F(b_1, a_2) - F(a_1, a_2)]$.

(3) $F(-\infty,-\infty)\hat{=}\lim\limits_{\substack{x\to-\infty\\y\to-\infty}}F(x,y)=0$;

$F(x,-\infty)\hat{=}\lim\limits_{y\to-\infty}F(x,y)=0$，对一切实数 x；

$F(-\infty,y)\hat{=}\lim\limits_{x\to-\infty}F(x,y)=0$，对一切实数 y；

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

(4) $F(x,y)$对每个自变量右连续，即对任意的实数 x，y，

$$F(x+0,y)\hat{=}\lim_{\Delta x\to0+}F(x+\Delta x,y)=F(x,y),$$

$$F(x,y+0)\hat{=}\lim_{\Delta y\to0+}F(x,y+\Delta y)=F(x,y).$$

反之，任一满足上述性质的 R^2 上的函数 $F(x,y)$都可以作为某一概率空间上二维随机变量的分布函数.

证 (1)由命题 1.5.17，概率的非负性及 $x_2>x_1$ 知$\{\xi\leqslant x_2,\eta\leqslant y\}\supset\{\xi\leqslant x_1,\eta\leqslant y\}$，

$$\begin{aligned}F(x_2,y)-F(x_1,y)&=P(\xi\leqslant x_2,\eta\leqslant y)-P(\xi\leqslant x_1,\eta\leqslant y)\\&=P(x_1<\xi\leqslant x_2,\eta\leqslant y)\geqslant0,\end{aligned}$$

因而对于任意固定的 y，$F(x,y)$关于 x 单调非减. 同理可证对任意固定的 x，$F(x,y)$关于 y 单调非减.

(2) 由(1)的证明及 $b_2\geqslant a_2$，$b_1\geqslant a_1$ 可知

$$\begin{aligned}\Delta_{b,a}F&\hat{=}F(b_1,b_2)-F(b_1,a_2)-F(a_1,b_2)+F(a_1,a_2)\\&=[F(b_1,b_2)-F(b_1,a_2)]-[F(a_1,b_2)-F(a_1,a_2)]\\&=P(\xi\leqslant b_1,a_2<\eta\leqslant b_2)-P(\xi<a_1,a_2<\eta\leqslant b_2)\\&=P(a_1<\xi\leqslant b_1,a_2<\eta\leqslant b_2)\geqslant0\end{aligned}$$

分布函数$F(x,y)$的二阶差分 $\Delta_{b,a}F$ 表示随机点 (ξ,η)取自以 (a_1,a_2)为左下端点，(b_1,b_2)为右上端点的半开半闭矩形区域的概率 (见图 4－10).

(3)由于 $F(x,y)$对每个自变量是单调非减的，因而(3)中

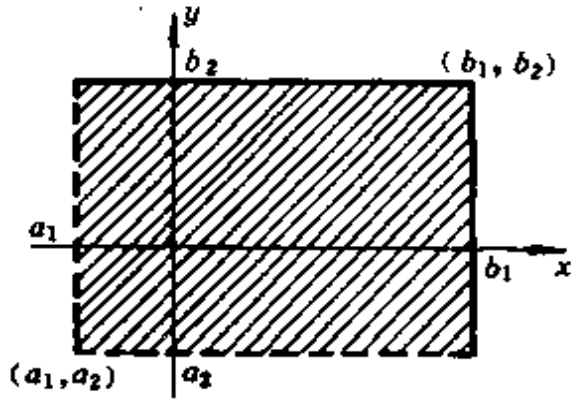


图 4－10

的极限都存在且极限非负，故可以取序列$\{-n\}$使

$$F(-\infty, y)=\lim_{n\to\infty} F(-n, y)=\lim_{n\to\infty} P(\xi\leqslant -n, \eta\leqslant y).$$

由于$\{\xi\leqslant -n\}\downarrow\bigcap_{n=1}^{\infty}\{\xi\leqslant -n\}=\varnothing$，再由概率的上连续性知

$$0\leqslant F(-\infty, y)\leqslant\lim_{n\to\infty} P(\xi\leqslant -n)=P\left(\bigcap_{n=1}^{\infty}\{\xi\leqslant -n\}\right)$$

$$=P(\varnothing)=0.$$

同理可证$F(x, -\infty)=0$及$F(-\infty, -\infty)=0$.

由于$F(x, y)$是每个自变量的单调非减函数且有上界，因而当$x\to+\infty$，$y\to+\infty$时极限存在且

$$F(+\infty, +\infty)=\lim_{\substack{x\to+\infty\\ y\to+\infty}} F(x, y)=\lim_{n\to+\infty} F(n, n)$$

$$=\lim_{n\to\infty} P(\xi\leqslant n, \eta\leqslant n).$$

由于

$$\{\xi\leqslant n, \eta\leqslant n\}\uparrow\bigcup_{n=1}^{\infty}\{\xi\leqslant n, \eta\leqslant n\}=\{\xi<\infty, \eta<\infty\}=\Omega,$$

故由概率的下连续性有

$$F(+\infty, +\infty)=P\left(\bigcup_{n=1}^{\infty}\{\xi\leqslant n, \eta\leqslant n\}\right)=P(\Omega)=1.$$

(4) $F(x+\Delta x, y)$当$\Delta x \downarrow 0$时是不增的且有下界$F(x, y)$，因而极限存在．取序列$\{\frac{1}{n}\}$，则

$$F(x, y) \leqslant \lim_{n\to\infty} F\left(x+\frac{1}{n}, y\right) = \lim_{n\to\infty} P\left(\xi \leqslant x+\frac{1}{n}, \eta \leqslant y\right).$$

由于事件$\{\xi \leqslant x+\frac{1}{n}, \eta \leqslant y\} \downarrow \bigcap_{n=1}^{\infty} \{\xi \leqslant x+\frac{1}{n}, \eta \leqslant y\} = \{\xi \leqslant x, \eta \leqslant y\}$及概率的上连续性，上式右端的极限为$P(\xi \leqslant x, \eta \leqslant y)$，即

$$F(x, y) \leqslant \lim_{n\to\infty} F\left(x+\frac{1}{n}, y\right) = F(x, y).$$

此即$F(x+0, y) = F(x, y)$．同样可得$F(x, y)$关于y右连续．

反之，若$F(x, y)$满足性质(1)—(4)，可以造一个概率空间并在其上定义二维随机变量(ξ, η)，使$F(x, y)$为其分布函数(参看[15]，第二章§3.2)．

注意：性质(2)是作为概率分布函数的必要的特征，它不能由非减性等导出．例如函数

$$F(x, y) = \begin{cases} 0, & x+y<0, \\ 1, & x+y \geqslant 0 \end{cases}.$$

满足性质(1)、(3)、(4)，但不满足性质(2)．取$a=(-1, -1)$ $b=(1, 1)$，则

$$\begin{aligned}\Delta_{b,a}F &= F(1,1) - F(-1,1) - F(1,-1) + F(-1,-1) \\ &= 1-1-1+0 = -1<0,\end{aligned}$$

它不满足性质(2)，因而不能成为概率分布函数．性质(2)是随机点取自矩形区域内概率非负性的一种反映．

定义4.1.8 设$F(x, y)$是(ξ, η)的分布函数，分别称ξ, η的分布函数$F_\xi(x), F_\eta(y)$为$F(x, y)$的关于ξ、η的边缘分布函数．可以证明：

$$F_\xi(x) = F(x, +\infty) \triangleq \lim_{v\to+\infty} F(x, v),$$

$$F_\eta(y) = F(+\infty, y) = \lim_{u\to+\infty} F(u, y).$$

事实上，由定理 4.1.7 性质(2)知 $F(x,+\infty)$，$F(+\infty,y)$存在，再由概率的下连续性知

$$F(x,+\infty)=\lim_{n\to+\infty}F(x,n)=\lim_{n\to+\infty}P(\xi\leqslant x,\eta\leqslant n)$$

$$=P\left(\bigcup_{n=1}^{\infty}\{\xi\leqslant x,\eta\leqslant n\}\right)=P(\xi\leqslant x)=F_{\xi}(x).$$

同法可证另一式.

反过来，若已知 ξ，η 各自的分布函数，未必能确定(ξ,η)的联合分布函数(参看例 4.2.4，4.2.5 及例 4.3.16 和它后面的解释). 当 $\xi_1,\xi_2,\cdots,\xi_n$ 独立时，边缘分布可以决定联合分布.

定理 4.1.9 随机变量 $\xi_1,\xi_2,\cdots,\xi_n$ 独立的充分必要条件是$(\xi_1,\xi_2,\cdots,\xi_n)$的联合分布函数等于它关于 $\xi_1,\cdots,\xi_n$ 的边缘分布函数的乘积. 即 $(\xi_1,\xi_2,\cdots,\xi_n)$的联合分布函数

(4.1.10) $$F(x_1,x_2,\cdots,x_n)=F_{\xi_1}(x_1)F_{\xi_2}(x_2)\cdots F_{\xi_n}(x_n)$$

对一切 $x_k\in(-\infty,\infty)$，$k=1,2,\cdots,n$ 成立，其中 $F_{\xi_k}(x_k)$为 ξ_k 的分布函数，$k=1,2,\cdots,n$.

证 若 $\xi_1,\xi_2,\cdots,\xi_n$ 独立，由定义 2.4.5，取 $B_k=(-\infty,x_k]$ 即得(4.1.10)，反之若(4.1.10)对一切 x_k，$k=1,2,\cdots,n$ 成立，则由分布函数的性质，用数学归纳法可证对一切 B_k，$k=1,2,\cdots,n$，是有限、无限的开、闭、半开半闭的区间均有

(4.1.11)

$$P(\xi_1\in B_1,\xi_2\in B_2,\cdots,\xi_n\in B_n)=P(\xi_1\in B_1)P(\xi_2\in B_2)\cdots P(\xi_n\in B_n)$$

成立. 读者可证明其中一些特殊情形. 在此我们证明当 $n=2$ 时若(4.1.10)成立，则对 $B_1=[a,b]$，$B_2=[c,d]$，(4.1.11)成立.

事实上，由概率的上连续性及定理 4.1.7 性质(2)知

$$P(\xi_1\in[a,b],\xi_2\in[c,d])=\lim_{n\to\infty}P\left(\xi_1\in\left(a-\frac{1}{n},b\right],\xi_2\in\left(c-\frac{1}{n},d\right]\right)$$

$$=\lim_{n\to\infty}P\left(a-\frac{1}{n}<\xi_1\leqslant b,c-\frac{1}{n}<\xi_2\leqslant d\right)$$

$$= \lim_{n\to\infty}\left[F(b,d)-F\left(b,c-\frac{1}{n}\right)-F\left(a-\frac{1}{n},d\right)+F\left(a-\frac{1}{n},c-\frac{1}{n}\right)\right]$$

$$= \lim_{n\to\infty}\left[F_{\xi_1}(b)F_{\xi_2}(d)-F_{\xi_1}(b)F_{\xi_2}\left(c-\frac{1}{n}\right)-F_{\xi_1}\left(a-\frac{1}{n}\right)F_{\xi_2}(d)\right.$$

$$\left.+F_{\xi_1}\left(a-\frac{1}{n}\right)F_{\xi_2}\left(c-\frac{1}{n}\right)\right]$$

$$= \lim_{n\to\infty}\left\{F_{\xi_1}(b)\left[F_{\xi_2}(d)-F_{\xi_2}\left(c-\frac{1}{n}\right)\right]\right.$$

$$\left.-F_{\xi_1}\left(a-\frac{1}{n}\right)\left[F_{\xi_2}(d)-F_{\xi_2}\left(c-\frac{1}{n}\right)\right]\right\}$$

$$= \lim_{n\to\infty}\left[F_{\xi_1}(b)-F_{\xi_1}\left(a-\frac{1}{n}\right)\right]\left[F_{\xi_2}(d)-F_{\xi_2}\left(c-\frac{1}{n}\right)\right]$$

$$= \lim_{n\to\infty}P\left(a-\frac{1}{n}<\xi_1\leqslant b\right)P\left(c-\frac{1}{n}<\xi_2\leqslant d\right)$$

$$= P\left(\bigcap_{n=1}^{\infty}\left\{a-\frac{1}{n}<\xi_1\leqslant b\right\}\right)P\left(\bigcap_{n=1}^{\infty}\left\{c-\frac{1}{n}<\xi_2\leqslant d\right\}\right).$$

$$= P(a\leqslant\xi_1\leqslant b)\,P\,(c\leqslant\xi_2\leqslant d) = P(\xi_1\in[a,b])P(\xi_2\in[c,d]).$$

这就说明定义 2.4.5 中关于随机变量独立的定义与(4.1.10)等价. 验证(4.1.10)要比验证定义 2.4.5 容易得多.

不仅如此，我们还可以给出更容易判别独立的条件.

定理 4.1.12 若$F(x_1, x_2, \cdots, x_n)$是 n 维随机变量$(\xi_1, \xi_2, \cdots, \xi_n)$的联合分布函数，则 $\xi_1, \xi_2, \cdots, \xi_n$ 独立的充分必要条件是存在函数 $G_k(x_k)$，$k=1, 2, \cdots, n$，使得对一切 $x_k\in(-\infty,\infty)$，

$k=1,2,\cdots,n$，有

(4.1.13) $$F(x_1,x_2,\cdots,x_n)=G_1(x_1)G_2(x_2)\cdots G_n(x_n).$$

且当 (4.1.13)成立时关于 ξ_k 的边缘分布

$$F_{\xi_k}(x_k)=G_k(x_k)/G_k(+\infty),k=1,2,\cdots,n,$$

其中 $G_k(+\infty)=\lim\limits_{x_k\to+\infty}G_k(x_k)$必存在.

证 我们只对 $n=2$ 的情形证明.条件的必要性由定理 4.1.9 给出,此时只要取 $G_k(x_k)=F_{\xi k}(x_k)$就够了.以下证条件的充分性:若 (4.1.13)成立,则由于 $F(x,y)$为概率分布函数,$F(+\infty,+\infty)=1$,因而存在 x_1,x_2 使 $F(x_1,x_2)>0$,从而 $G_1(x_1)\neq 0$,

$\lim\limits_{y\to+\infty}G_1(x_1)G_2(y)=\lim\limits_{y\to+\infty}F(x_1,y)=F(x_1,+\infty)$存在,因而 $\lim\limits_{x_2\to+\infty}G_2(x_2)$存在,记为 $G_2(+\infty)$.同理可证 $G_1(+\infty)$也存在,且由 $F(+\infty,+\infty)=1$ 可知 $G_1(+\infty)G_2(+\infty)=1$.所以

$$F(x_1,x_2)=\frac{G_1(x_1)G_2(x_2)}{G_1(+\infty)G_2(+\infty)}.$$

再分别令 $x_1,x_2\to+\infty$ 得

$$F_{\xi_1}(x_1)=F(x_1,+\infty)=\frac{G_1(x_1)}{G_1(+\infty)},$$

$$F_{\xi_2}(x_2)=F(+\infty,x_2)=\frac{G_2(x_2)}{G_2(+\infty)}.$$

且

$$F(x_1,x_2)=F_{\xi_1}(x_1)F_{\xi_2}(x_2),x_1,x_2\in(-\infty,\infty)$$

由定理 4.1.9 知 ξ_1,ξ_2 独立.

现在利用定理 4.1.9 重新证明命题 2.4.7.

定理 4.1.14 若随机变量 $\xi_1,\xi_2,\cdots,\xi_n$ 独立,则对任意的 k,$2\leqslant k\leqslant n$,$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$,$\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_k}$ 也独立.

证 我们只证 $n=3$,$k=2$,$i_1=1$,$i_2=2$ 的情形.即由 ξ_1,ξ_2,ξ_3 独立来证 ξ_1,ξ_2 独立.

由于事件 $\{\xi_3\leqslant m\},m=1,2,\cdots$ 是一个增事件序列且

$$\Omega=\bigcup_{m=1}^{\infty}\{\xi_3\leqslant m\}.$$

因此对任意实数 x_1，x_2，

$$\{\xi_1\leqslant x_1,\ \xi_2\leqslant x_2,\ \xi_3\leqslant m\},\ m=1,2,\cdots$$

也是一个增事件序列且

$$\bigcup_{m=1}^{\infty}\{\xi_1\leqslant x_1\ \xi_2\leqslant x_2,\ \xi_3\leqslant m\}=\{\xi_1\leqslant x_1,\ \xi_2\leqslant x_2\}.$$

故由概率的下连续性及 ξ_1，ξ_2，ξ_3 的独立性知对任意实数 x_1，x_2，

$$\begin{aligned}P(\xi_1\leqslant x_1,\ \xi_2\leqslant x_2)&=\lim_{m\to\infty}P(\xi_1\leqslant x_1,\ \xi_2\leqslant x_2,\ \xi_3\leqslant m)\\&=\lim_{m\to\infty}F(x_1,x_2,n)=\lim_{m\to\infty}F_{\xi_1}(x_1)F_{\xi_2}(x_2)F_{\xi_3}(m)\\&=F_{\xi_1}(x_1)F_{\xi_2}(x_2)\lim_{m\to\infty}P(\xi_3\leqslant m)\\&=F_{\xi_1}(x_1)F_{\xi_2}(x_2)P\left(\bigcup_{m=1}^{\infty}\{\xi_3\leqslant m\}\right)=F_{\xi_1}(x_1)F_{\xi_2}(x_2).\end{aligned}$$

这就证明了 ξ_1，ξ_2，独立.

在第五、六两章还要用到多维随机变量独立的概念及有关性质，现在介绍于下. 由于定理 4.1.9 已经告诉我们：随机变量的独立可以用它们的分布函数来刻划，因此有下面的

定义 4.1.15 设$(\xi_{11},\cdots,\xi_{1n_1})$，$(\xi_{21},\cdots,\xi_{2n_2})$，…，$(\xi_{k1},\cdots,\xi_{kn_k})$是 k 个多维随机变量，若对任意实数 $x_{11},\cdots,x_{1n_1}$，$x_{21},\cdots,x_{2n_2},\cdots,x_{k1},\cdots,x_{kn_k}$ 有

$$\begin{aligned}&P\left(\bigcap_{l=1}^{k}\{\xi_{l1}\leqslant x_{l1},\cdots,\xi_{ln_l}\leqslant x_{ln_l}\}\right)\\&=\prod_{l=1}^{k}P(\xi_{l1}\leqslant x_{l1},\cdots,\xi_{ln_l}\leqslant x_{ln_l}),\end{aligned}$$

则称 $(\xi_{11},\cdots,\xi_{1n_1})$，$(\xi_{21},\cdots,\xi_{2n_2})$，…，$(\xi_{k1},\cdots,\xi_{kn_k})$为独立的多维随机变量.

定理 4.1.16 若 $\xi_{11},\cdots,\xi_{1n_1},\xi_{21}\cdots,\xi_{2n_2},\cdots,\xi_{k1},\cdots,\xi_{kn_k}$ 是

独立的随机变量，则$(\xi_{11},\cdots,\xi_{1n_1})$，$(\xi_{21},\cdots,\xi_{2n_2})$，$\cdots$，$(\xi_{k1},\cdots,\xi_{kn_k})$是独立的多维随机变量

证 对任意的实数$x_{11},\cdots,x_{1n_1},x_{21},\cdots,x_{2n_2},\cdots,x_{k1},\cdots,x_{kn_k}$，由定理4.1.9及$\xi_{11},\cdots,\xi_{1n_1},\cdots,\xi_{k1},\cdots,\xi_{kn_k}$独立知

$$P\left(\bigcap_{l=1}^{k}\{\xi_{l1}\leqslant x_{l1},\cdots,\xi_{kn_l}\leqslant x_{ln_l}\}\right)=P\left(\bigcap_{l=1}^{k}\bigcap_{m=1}^{n_l}\{\xi_{lm}\leqslant x_{lm}\}\right)$$

$$=\prod_{l=1}^{k}\prod_{m=1}^{n_l}P(\xi_{lm}\leqslant x_{lm}).$$

再由假设及定理4.1.14知，对任意的l，$\xi_{l1},\cdots,\xi_{ln_l}$独立，所以再由定理4.1.9知

$$\prod_{m=1}^{n_l}P(\xi_{lm}\leqslant x_{lm})=P(\xi_{l1}\leqslant x_{l1},\cdots,\xi_{ln_l}\leqslant x_{ln_l})$$

代入上式右端即知定理的结论成立.

定理4.1.17 若$(\xi_{11},\cdots,\xi_{1n_1})$，$(\xi_{21},\cdots,\xi_{2n_2})$，$\cdots$，$(\xi_{k1},\cdots,\xi_{kn_k})$是$k$个独立的多维随机变量，$g_l(x_{l1},\cdots,x_{ln_l}),l=1,\cdots,k$是使$g_l(\xi_{l1},\cdots,\xi_{ln_l})$成为随机变量的函数，则$g_1(\xi_{11},\cdots,\xi_{1n_1})$，$g_2(\xi_{21},\cdots,\xi_{2n_2})$，$\cdots$，$g_k(\xi_{k1},\cdots,\xi_{kn_k})$是独立的随机变量.

关于定理中的$g_l(x_{l1},\cdots,x_{ln_l})$需要满足的一般条件的叙述及定理的证明超出了本书的范围，有兴趣的读者可参看[15，第二章§4推论3]．但是在此我们应该指出，在本书今后用到此定理时，相应的函数$g_l(x_{l1},\cdots,x_{ln_l})$是满足要求的.

习题 4－1

1. 设二维随机变量(ξ,η)的分布函数为$F(x,y)$，试用它来表示下列概率：(1)$P(a\leqslant\xi<b,\eta\leqslant y)$；(2)$P(\xi=a,\eta<y)$.(3)$P(a\leqslant\xi<b,c\leqslant\eta<d)$.

2. 设二维随机变量(ξ,η)的分布函数为

$$F(x,y)=\begin{cases}c-2^{-x}-2^{-y}+2^{-x-y},&\text{当 }x\geqslant 0,y\geqslant 0\text{时，}\\0,&\text{当 }x<0\text{或}y<0\text{时，}\end{cases}$$

(1) 求 c；(2)求 (ξ,η)落在区域$\{(x,y):1<x<2,3<y<5\}$中的概率；(3)求 ξ 与 η 的边缘分布函数.

3. 设(ξ,η)的分布函数为

$$F(x,y)=A\left(B+\operatorname{arc\,tg}\frac{x}{2}\right)\left(C+\operatorname{arc\,tg}\frac{y}{3}\right)$$

求常数 A、B、C、及 ξ,η 的边缘分布，试问 ξ,η 独立吗?

4. 设三维随机变量 (ξ_1,ξ_2,ξ_3)的分布函数为 $F(x_1,x_2,x_3)$，证明：若对一切 $x_k\in(-\infty,\infty)$，$k=1,2,3$ 有

$$F(x_1,x_2,x_3)=F_{\xi_1}(x_1)F_{\xi_2}(x_2)F_{\xi_3}(x_3),$$

则对一切有限左开右闭区间 B_1,B_2,B_3 有

$$F(\xi_1\in B_1,\xi_2\in B_2,\xi_3\in B_3)=P(\xi_1\in B_1)P(\xi_2\in B_2)P(\xi_3\in B_3).$$

§4.2 多维离散型随机变量

现在来讨论多维随机变量中每个随机变量都是离散型随机变量的情形.

定义 4.2.1 若 n 维随机变量 $(\xi_1,\xi_2,\cdots,\xi_n)$的每一分量 ξ_i 是离散型随机变量，$i=1,2,\cdots n$，则称 $(\xi_1,\xi_2,\cdots,\xi_n)$为 n 维离散型随机变量.

定义 4.2.2 若二维离散型随机变量(ξ,η)可能取的一切值为 (x_i,y_j)，$i,j=1,2,\cdots$，则称

$$p_{ij}=p(\xi=x_i,\eta=y_j),\ i,j=1,2,\cdots$$

为 (ξ,η)的**联合分布列**，简称**分布列**.

定理 4.2.3 二维离散型随机变量(ξ,η)的联合分布列具有下述性质：

(1) $p_{ij}\geqslant 0$，$i,j=1,2,\cdots$；

(2) $\sum\limits_i\sum\limits_j p_{ij}=1$；

(3) $p_{i\cdot}\triangleq\sum\limits_j p_{ij}=p(\xi=x_i)$，$i=1,2,\cdots$；

$p_{\cdot j}\triangleq\sum\limits_i p_{ij}=p(\eta=y_j)$，$j=1,2,\cdots$.

并且具有性质(1),(2)的任何数集$\{p_{ij}:i,j=1,2,\cdots\}$均可作为

某一二维离散型随机变量的分布列.

证 性质(1)、(2)是显然的，现在证性质(3)成立.

$$\sum_j p_{ij}=\sum_j P(\xi=x_i,\eta=y_j)=P\left(\bigcup_j\{\xi=x_i,\eta=y_j\}\right),$$

由于

$$\bigcup_j\{\xi=x_i,\eta=y_j\}=\{\xi=x_i\}\bigcap\left(\bigcup_j\{\eta=y_j\}\right)=\{\xi=x_i\}\bigcap\Omega=\{\xi=x_i\},$$

上式右端等于 $P(\xi=x_i)$，故(3)的第一式获证. 同样可证第二式.

反之，若数集$\{p_{ij}:i,j=1,2,\cdots\}$具有性质(1),(2),则令 $\Omega\hat{=}\{(i,j):i,j=1,2,\cdots\}$,$\mathscr{F}$ 为 Ω 的一切子集作成的集类，则易证$\mathscr{F}$ 是一个事件 $\sigma-$ 域；对于任何 $A\in\mathscr{F}$，定义 $P(A)\hat{=}\sum_{(i,j)\in A}p_{ij}$，则易证 $P(A)$，$A\in\mathscr{F}$ 是$\mathscr{F}$ 上的一个概率，因而$(\Omega,\mathscr{F},P)$是一概率空间. 最后定义 $\xi(i,j)\hat{=}i$，$\eta(i,j)\hat{=}j$，$i,j=1,2,\cdots$，则(ξ,η)是取值于 Ω 并以$\{p_{ij}:i,j=1,2,\cdots\}$为分布列的二维离散型随机变量.

与一维情形类似，人们常常把二维离散型随机变量的联合分布及边缘分布用阵列的形式表示出来.

表 4-1

ξ \ η	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$					$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

表 4-1 中最右边的一列是 ξ 的边缘分布列，对每一 i，$p_{i\cdot}=\sum_j p_{ij}$，

即表中第 i 行所有 p_{ij} 之和；表 4－1 中最下边的一行是 η 的边缘分布列，$p_{\cdot j}$ 是第 j 列所有的 p_{ij} 之和，即 $p_{\cdot j}=\sum_i p_{ij}$，$j=1,2,\cdots$.

例 4.1.5 中的 (ξ,η) 就是一个二维离散型随机变量，其分布列

$$p_{ij}=P(\xi=i,\eta=j)=\frac{1}{4},\ i,j=0,1$$

现在再举一些多维离散型分布的例.

例 4.2.4 把3 个球等可能地放入编号为 1，2，3 的三个盒中，记落入第 1 号盒中的球的个数为 ξ，落入第 2 号盒中球的个数为 η，则 (ξ,η) 为二维离散型随机变量，其分布列

$$p_{ij}=P(\xi=i,\eta=j)=P(\xi=i)P(\eta=j\mid\xi=i),$$
$$i,j=0,1,2,3.$$

此时 ξ 为参数为 $\left(3,\frac{1}{3}\right)$ 的二项随机变量，

$$P(\xi=i)=\binom{3}{i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{3-i},\ i=0,1,2,3.$$

若 $\xi=i$，则还剩 $3-i$ 个球等可能地落入第 2、3 号盒中，因而在 $\xi=i$ 的条件下 η 是参数为 $\left(3-i,\frac{1}{2}\right)$ 的二项随机变量，即

$$P(\eta=j\mid\xi=i)=\binom{3-i}{j}\left(\frac{1}{2}\right)^j\left(\frac{1}{2}\right)^{3-i-j}=\binom{3-i}{j}\left(\frac{1}{2}\right)^{3-i},$$
$$j=0,1,2,3$$

于是

$$p_{ij}=\binom{3}{i}\binom{3-i}{j}\left(\frac{1}{2}\right)^{3-i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{3-i}$$

$$= \frac{3!}{i!\,j!\,(3-i-j)!}\left(\frac{1}{3}\right)^3, \quad i,j=0,1,2,3, \\ i+j\leqslant 3.$$

现将 (ξ,η)的联合分布及边缘分布列成表 4－2.

表 4－2

ξ \ η	0	1	2	3	p_i
0	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{27}$	$\frac{8}{27}$
1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{4}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	0	0	$\frac{2}{9}$
3	$\frac{1}{27}$	0	0	0	$\frac{1}{27}$
p_j	$\frac{8}{27}$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{1}{27}$	

(ξ,η)的联合分布列可以完全确定 ξ，η 各自的(边缘)分布列；反过来，ξ，η 的分布列如果已知，能否由它们确定(ξ,η)的联合分布列呢？答案是否定的，请看下例.

例 4.2.5 把3 个红球和 3 个白球等可能地放入编号为 1，2，3 的三个盒子中，记落入第 1 号盒中的白球数为 ξ，落入第 2 号盒中的红球数为 η. 试求(ξ,η)的联合分布及边缘分布.

解 由问题的提法我们知道，已知落入第 1 号盒中的白球数 $\xi=i$ 不影响落入第 2 号盒中红球数 $\eta=j$ 的概率，于是对于 i，$j=0,1,2,3$，

$$P(\xi=i,\eta=j)=P(\xi=i)P(\eta=j\mid\xi=i) \\ =P(\xi=i)P(\eta=j) \\ =\binom{3}{i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{3-i}\binom{3}{j}\left(\frac{1}{3}\right)^j\left(\frac{2}{3}\right)^{3-j}.$$

将(ξ,η)的联合分布及边缘分布列成表4－3.

表 4－3

ξ \ η	0	1	2	3	$p_{i\cdot}$
0	$\frac{8}{27}\cdot\frac{8}{27}$	$\frac{4}{9}\cdot\frac{8}{27}$	$\frac{2}{9}\cdot\frac{8}{27}$	$\frac{1}{27}\cdot\frac{8}{27}$	$\frac{8}{27}$
1	$\frac{8}{27}\cdot\frac{4}{9}$	$\frac{4}{9}\cdot\frac{4}{9}$	$\frac{2}{9}\cdot\frac{4}{9}$	$\frac{1}{27}\cdot\frac{4}{9}$	$\frac{4}{9}$
2	$\frac{8}{27}\cdot\frac{2}{9}$	$\frac{4}{9}\cdot\frac{2}{9}$	$\frac{2}{9}\cdot\frac{2}{9}$	$\frac{1}{27}\cdot\frac{2}{9}$	$\frac{2}{9}$
3	$\frac{8}{27}\cdot\frac{1}{27}$	$\frac{4}{9}\cdot\frac{1}{27}$	$\frac{2}{9}\cdot\frac{1}{27}$	$\frac{1}{27}\cdot\frac{1}{27}$	$\frac{1}{27}$
$p_{\cdot j}$	$\frac{8}{27}$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{1}{27}$	

例4.2.4与例4.2.5中(ξ,η)有着相同的边缘分布，但(ξ,η)的联合分布却不相同．这说明联合分布不仅包含了ξ,η的分布规律，还包含了它们之间的相互联系．

例4.2.4的一般化为

例4.2.6 若在n次重复独立试验中，每次试验出现的互不相容的结果（事件）有$k+1$个，记为$A_1,\cdots,A_k,A_{k+1}$，且$\bigcup_{j=1}^{k+1}A_j=\Omega$．记

$$P(A_i)=p_i,\ i=1,\cdots,k+1,$$

因而

$$\sum_{i=1}^{k+1}p_i=1.$$

在n次试验中，令ξ_i表示A_i出现的次数，$i=1,2,\cdots,k$．则

$$(4.2.7)\quad P(\xi_1=n_1,\cdots,\xi_k=n_k)=\frac{n!}{n_1!\cdots n_k!\,n_{k+1}!}p_1^{n_1}\cdots p_k^{n_k}p_{k+1}^{n_{k+1}},$$

其中$\sum_{i=1}^{k+1}n_i=n$，$n_i\geqslant 0$，$i=1,\cdots,k,k+1$（注意n_{k+1}是事件A_{k+1}出现的次数）．

证 在n次试验中，事件A_i $(i=1,\cdots,k)$出现n_i次，则事

件 A_{k+1} 恰好出现 $n_{k+1}=n-\sum_{i=1}^{k} n_i \geqslant 0$ 次．由试验的独立性假设知，任一 A_i $(i=1, \cdots, k+1)$出现 n_i 次的结果序列，其发生的概率等于 $p_1^{n_1} \cdots p_k^{n_k} p_{k+1}^{n_{k+1}}$ ．因此为了证明(4.2.7)，只需算出 A_i $(i=1, 2, \cdots, k+1)$出现 n_i $(n_1+\cdots+n_{k+1}=n)$次的结果序列的个数．

由组合公式知，n 个结果其中 A_1 恰好出现 n_1 次的组合数是 $\binom{n}{n_1}$，而在 A_1 出现 n_1 次后 (即剩下有 $n-n_1$ 个结果)，A_2 恰好出现 n_2 次的组合数是 $\binom{n-n_1}{n_2}$，依此，在 A_i $(i=1, \cdots, l)$恰好出现 n_i 次后(即剩下有 $n-(n_1+\cdots+n_l)$个结果)，A_{l+1} 恰好出现 n_{l+1} 次的组合数是 $\binom{n-(n_1+\cdots+n_l)}{n_{l+1}}$．所以 A_i $(i=1, \cdots, k+1)$ 出现 n_i $(n_1+\cdots+n_{k+1}=n)$次的结果序列数是

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-(n_1+\cdots+n_{k-1})}{n_k}\binom{n-(n_1+\cdots+n_k)}{n_{k+1}}$$

$$=\frac{n!}{n_1!\cdots n_k!\, n_{k+1}!},$$

(4.2.7)获证．

定义 4.2.8 在例4.2.6的假设和记号下，称具有分布列(4.2.7)的 k 维随机变量 $(\xi_1, \cdots, \xi_k)$ 为服从多 $(k+1)$项分布的随机变量，而称(4.2.7)为 $k+1$ 项分布列，记作 $M(n; p_1, \cdots, p_k)$(因为 $p_{k+1}=1-(p_1+\cdots+p_k)$)．

显然当 $k=1$时，$p_2=1-p_1$，$n_2=n-n_1$，因而(4.2.7)就是二项分布．所以多项分布是二项分布的推广．

多项分布有广泛的应用，例如在产品质量检查中，已知一、二、三等品出现的概率分别为 p_1，p_2，p_3，$p_1+p_2+p_3=1$．有放回地抽取 n 件，用 ξ_1，ξ_2，ξ_3 分别表示其中一、二、三等品的件数，则 (ξ_1, ξ_2)服从 $M(n; p_1, p_2)$；(ξ_1, ξ_3)服从 $M(n; p_1, p_3)$；(ξ_2, ξ_3)服从 $M(n; p_2, p_3)$．

定理 4.2.9 设 (ξ, η)为二维离散型随机变量，其分布列为

$$P(\xi=x_i, \eta=y_j)=p_{ij},\ i, j=1, 2, \cdots,$$

则 ξ，η 独立的充分必要条件是

(4.2.10) $\qquad p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$，$i,j=1,2,\cdots$，

其中$\{p_{i\cdot}\}$，$\{p_{\cdot j}\}$分别为 (ξ,η)关于 ξ，η 的边缘分布，即 $P(\xi=x_i)=p_{i\cdot}$，$P(\eta=y_j)=p_{\cdot j}$.

证 若 ξ，η 独立，对定义 2.4.5，令 $n=2$，$B_1=\{x_i\}=[x_i,x_i]$，$B_2=\{y_j\}=[y_j,y_j]$，则得(4.2.10).

反之若(4.2.10)成立，设 $F(x,y)$为(ξ,η)的联合分布函数，则对任意实数 x，y，有

$$F(x,y)=P(\xi\leqslant x,\eta\leqslant y)=P\Big(\bigcup_{x_i\leqslant x}\bigcup_{y_j\leqslant y}\{\xi=x_i,\eta=y_j\}\Big)$$

$$=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}P(\xi=x_i,\eta=y_j)=\sum_{i:x_i\leqslant x}\sum_{j:y_j\leqslant y}p_{ij}$$

$$=\sum_{i:x_i\leqslant x}\sum_{j:y_j\leqslant y}p_{i\cdot}p_{\cdot j}=\sum_{i:x_i\leqslant x}p_{i\cdot}\sum_{j:y_j\leqslant y}p_{\cdot j}$$

$$=P(\xi\leqslant x)\,P(\eta\leqslant y)\,,$$

故由定理 4.1.2 知 ξ，η 独立.

为了符号和书写简单起见，上面着重讨论了二维离散型随机变量的有关结论. 至于一般的 n 维离散型随机变量的情形，结论和证明都是很类似的. 现在将它们叙述成下面的定理，以便读者了解. 证明就不再叙述了，有兴趣的读者，不妨自己补证.

定理 4.2.10 设 n 维离散型随机变量 $(\xi_1,\cdots,\xi_n)$可能取的一切值为 $(x_{1i_1},\cdots,x_{ni_n})$，$i_1,\cdots,i_n=1,2,\cdots$，称

$$p_{i_1\cdots i_n}\triangleq P(\xi_1=x_{1i_1},\cdots,\xi_n=x_{ni_n}),\ i_1,\cdots,i_n=1,2,\cdots$$

为$(\xi_1,\cdots,\xi_n)$的联合分布列. 则下列结论成立：

Ⅰ. 联合分布列具有下述性质：

(1) $p_{i_1\cdots i_n}\geqslant 0$，$i_1,\cdots,i_n=1,2,\cdots$；

(2) $\sum_{i_1}\cdots\sum_{i_n}p_{i_1\cdots i_n}=1$；

(3) 对任意的 $1\leqslant k\leqslant n$，任意的 $1\leqslant l_1<l_2<\cdots<l_k\leqslant n$，记 1，2，

$\cdots, n$ 中除 $l_1, \cdots, l_k$ 外的数码为 $m_1, \cdots, m_{n-k}$，则 $(\xi_{l_1}, \cdots, \xi_{l_k})$ 的联合分布列

$$P(\xi_{l_1}=x_{l_1 i_{l_1}}, \cdots, \xi_{l_k}=x_{l_k i_{l_k}})=\sum_{i_{m_1}} \cdots \sum_{i_{m_{n-k}}} p_{i_1, \cdots, i_n},$$

其中 $i_{l_1}, \cdots, i_{l_k}=1, 2, \cdots$；

(4)具有性质(1)、(2)的任何数集$\{p_{i_1 \cdots i_n}: i_1, \cdots, i_n=1, 2, \cdots\}$均可作为一$n$维离散型随机变量的联合分布列.

Ⅱ. $(\xi_1, \cdots, \xi_n)$ 独立的充分必要条件是

$$p_{i_1 \cdots i_n}=P(\xi_1=x_{1 i_1}) \cdots P(\xi_n=x_{n i_n}), i_1, \cdots, i_n=1, 2, \cdots.$$

例 4.2.11 由表4－3及定理4.2.9知例4.2.5中随机变量 ξ，η 是相互独立的.

例 4.2.12 设(ξ, η)为例4.2.4中的二维离散型随机变量，则由表4－2及定理4.2.9知它们是不独立的，事实上，

$$P(\xi=0, \eta=0)=\frac{1}{27} \neq P(\xi=0)P(\eta=0)=\frac{8}{27} \cdot \frac{8}{27}.$$

进一步，设 $(\xi_1, \cdots, \xi_k)$为服从多项分布 $M(n; p_1, \cdots, p_k)$(例4.2.6中的(4.2.7))，则由定理4.2.10知它们一般是不独立的. 事实上，当 $p_1, p_2, \cdots, p_k$ 中至少有两数不等于零时，则由初等不等式

$$(1-a)(1-b) \geqslant 1-(a+b), 0 \leqslant a \leqslant 1, 0 \leqslant b \leqslant 1$$

中等号不成立当且仅当 a, b 都不等于零这一事实知

$$(1-p_1) \cdots (1-p_k)>1-(p_1+\cdots+p_k).$$

因此由(4.2.7)知

$$P(\xi_1=0, \cdots, \xi_k=0)=p_{k+1}^n=[1-(p_1+\cdots+p_k)]^n$$
$$<(1-p_1)^n \cdots (1-p_k)^n=P(\xi_1=0) \cdots P(\xi_k=0),$$

再由定理4.2.10知 $\xi_1, \cdots, \xi_k$ 不独立.

例 4.2.13 在 n 重伯努利试验概型(§2.5)中，用 ξ_i 表示第 i 次试验成功的次数，用 A_i 表示第 i 次成功，则

$$\{\xi_i=\delta_i\}=\begin{cases}A_i\ ,\text{当}\delta_i=1\text{时},\\ A_i^c\ ,\text{当}\delta_i=0\text{时}.\end{cases}$$

于是对于任意的 $\delta_1,\cdots,\delta_n=0$ 或 1，都有

$$P(\xi_1=\delta_1,\cdots,\xi_n=\delta_n)=P(\xi_1=\delta_1)\cdots P(\xi_n=\delta_n).$$

故由定理 4.2.10 II 知 $\xi_1,\cdots,\xi_n$ 独立.

习题 4－2

1. 设离散型随机向量 (ξ,η)的联合分布为

ξ \ η	−1	0	1
0	0.1	0.2	0
1	0.2	0.3	0.2

(1)求 ξ 与 η 的边缘分布；(2)求 $P(\xi=1,\eta\geqslant 0)$；(3)ξ,η 独立吗?

2. 小盒中有 4 支红铅笔、2 支蓝铅笔、3 支绿铅笔，从中任取 3 支，设 ξ,η 分别是取出的红、蓝铅笔数．求 (1) (ξ,η)的联合分布；(2)ξ,η 的边缘分布；(3) $P(\xi<3,\eta=2)$；(4)ξ,η 独立吗?

3. 掷一枚均匀硬币 3 次，设 ξ 表示头两次出现正面的次数，η 表示在这 3 次中出现正面的总次数．(1)求 (ξ,η)的联合分布列；(2)求 ξ 与 η 的边缘分布列；(3)ξ 与 η 独立吗?

4. 掷两枚均匀骰子．求 (ξ,η)的联合分布列，其中(1)ξ 是两个点数的较大者，η 是两个点数之和；(2)ξ 是第一枚骰子出现的点数，η 是两个点数中的较大者；(3)ξ,η 分别是两个点数中的较小、较大者．

5. 两名射手对着同一个靶子相互独立地各自射击 2 次，第一、二射手命中的概率分别为 0.8 与 0.6．设 ξ、η 分别是第一、二射手命中的次数．求(1) ξ,η 的分布列；(2) (ξ,η)的联合分布列；(3) $P(\xi\leqslant 1,\eta=1)$.

6. 在上题中，若以 ζ 表示两人中靶的总次数．求(1)(ξ,ζ)的联合分布列；(2)ζ 的分布列；(3)ξ 与 ζ 是否独立?

§4.3 多维连续型随机变量

本节讨论多维随机变量的分布函数具有“密度”的情形，先给出

定义 4.3.1 设 $(\xi_1, \cdots, \xi_n)$是 n 维随机变量，如果对于它的分布函数 $F(x_1, \cdots, x_n)$存在一个非负可积函数 $f(x_1, \cdots, x_n)$使得对任意实数 $x_1, \cdots, x_n$，

$$(4.3.2)\ F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \cdots, u_n)\, du_1 \cdots du_n,$$

则称 $(\xi_1, \cdots, \xi_n)$为 n **维连续型随机变量**，称 $f(x_1, \cdots, x_n)$为 $(\xi_1, \cdots, \xi_n)$(或它的分布函数 $F(x_1, \cdots, x_n)$)的**联合分布密度**.

注意，对于 n 维连续型随机变量，我们不是像 n 维离散型随机变量那样，定义成"每一 ξ_i 都是连续型随机变量". 因为每一 ξ_i 是连续型随机变量并不能保证联合分布密度的存在. 而按照定义 4.3.1，n 维连续型随机变量 $(\xi_1, \cdots, \xi_n)$的每一分量仍然是连续型的(定理 4.3.17).

首先介绍一个结果，它告诉我们什么样的函数可以是联合分布密度，即

定理 4.3.3 n 维连续型随机变量 $(\xi_1, \cdots \xi_n)$的分布密度 $f(x_1, \cdots, x_n)$具有下列性质：

(1) 对任意的实数 $x_1, \cdots, x_n$，有 $f(x_1, \cdots, x_n) \geqslant 0$；

(2) $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1.$$

反之，若 $f(x_1, \cdots, x_n)$具有性质 (1)，(2)，则它必为某一 n 维连续型随机变量的联合分布密度.

证 若 $f(x_1, \cdots, x_n)$是 $(\xi_1, \cdots, \xi_n)$的联合分布密度，则由定义 4.3.1 知(1)成立. 其次由分布函数的性质 $F(+\infty, \cdots, +\infty) = 1$(其证明与定理 4.1.7 性质(3)的第四式的证明完全一样)知

$$1 = F(+\infty, \cdots, +\infty) \triangleq \lim_{\substack{x_i \to +\infty \\ i=1, \cdots, n}} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \cdots, u_n) du_1 \cdots du_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_n.$$

对于逆定理，为了符号简单起见，我们只证明 $n=2$ 的情形．设 $f(x_1, x_2)$ 是满足性质(1)，(2)的任一函数，则

(4.3.4)
$$F(x, y) \triangleq \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)dudv$$

对任何实数 x, y 都存在，且显然使定理 4.1.7 中的性质(1)(3)，(4)成立，甚至 $F(x, y)$ 对每一自变量连续．还有对任意给定的二维向量 $a=(a_1, a_2)$，$b=(b_1, b_2)$，$a_1 \leqslant b_1$，$a_2 \leqslant b_2$，由 $\Delta_{b,a}F$ 的定义、(4.3.4)及性质(2)有

$$\Delta_{b,a}F = \int_{-\infty}^{b_1}\int_{-\infty}^{b_2} f(x, y)dxdy - \int_{-\infty}^{b_1}\int_{-\infty}^{a_2} f(x, y)dxdy$$

(4.3.5)
$$- \int_{-\infty}^{a_1}\int_{-\infty}^{b_2} f(x, y)dxdy + \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} f(x, y)dxdy$$

$$= \int_{-\infty}^{b_1}\int_{a_2}^{b_2} f(x, y)dxdy - \int_{-\infty}^{a_1}\int_{a_2}^{b_2} f(x, y)dxdy$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x, y)dxdy \geqslant 0.$$

故 $F(x, y)$ 使定理 4.1.7 的性质(2)成立．因而由定理 4.1.7 的后一部份知由 (4.3.4)定义的 $F(x, y)$ 是某一二维随机变量的分布函数，所以 $f(x, y)$ 是某一二维连续型随机变量的联合分布密度．

例 4.3.6．均匀分布．设 G 为平面上可求积区域(例如圆、长方形、三角形及其他各种常见的有限图形)，$\mu(G)$ 为 G 的面积，$\mu(G)>0$．令

$$(4.3.7)\qquad f(x,y)=\begin{cases}\dfrac{1}{\mu(G)}, & (x,y)\in G,\\ 0, & \text{其它}.\end{cases}$$

显然 $f(x,y)$ 具有定理 4.3.3 的性质(1)，(2)，所以由定理 4.3.3 的后一部份知它是某一二维连续型随机变量 (ξ,η) 的联合分布密度．我们称此随机变量为在 G 上均匀分布的随机变量，称分布密度(4.3.7)是 G 上的均匀分布的密度．

若 G 为以原点为圆心，以 R 为半径的圆，则 G 上的均匀分布密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi R^2}, & x^2+y^2\leqslant R^2,\\ 0, & x^2+y^2>R^2,\end{cases}$$

若 $G=\{(x,y):a_1\leqslant x\leqslant b_1,\ a_2\leqslant y\leqslant b_2\}$ (矩形)，则相应的分布密度为

$$f(x,y)=\begin{cases}\dfrac{1}{(b_1-a_1)(b_2-a_2)}, & \text{当 } a_1\leqslant x\leqslant b_1,\ a_2\leqslant y\leqslant b_2,\\ 0, & \text{其它}.\end{cases}$$

实际上，G 不一定必须是平面上的可求积区域，而可以是任一 n 维空间中一个可求积的几何体．只要将 $\mu(G)$ 理解为相应的体积，(4.3.7)右边就定义了相应的几何体上的均匀分布．例如设 G 为三维空间中的球体 $B_R^3=\{(x,y,z):x^2+y^2+z^2\leqslant R^2\}$，

$$(4.3.8)\qquad f(x,y,z)=\begin{cases}\dfrac{4}{3\pi R^3}, & x^2+y^2+z^2\leqslant R^2,\\ 0, & \text{其它}.\end{cases}$$

为 B_R^3 上的均匀分布的密度．

应该指出：均匀分布就是几何概率的随机变量语言的表述．

例 4.3.8 多维正态分布．首先我们注意一维正态分布 $N(\mu,\sigma^2)$ 的密度函数可以写成

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

注意其中幂指数为一元正定二次型 $(x-\mu)(\sigma^2)^{-1}(x-\mu)$ 的 $-\frac{1}{2}$ 倍．因此，可以想象将此二次型换成一个 n 元正定二次型，以把一维正态分布密度推广成 n 维正态分布密度．即令

$$f(x)=Ce^{-\frac{1}{2}(x-\mu)'B^{-1}(x-\mu)},$$

其中 B 为 n 阶正定方阵，x、μ 为 n 维列向量，$(x-\mu)'$表示列向量 $x-\mu$ 的转置，C 是一待定的正常数．为了使 (4.3.9)定义的 $f(x)$是一 n 维连续型随机变量的分布密度，由定理 4.3.3 知 C 必需满足条件

$$C\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}(x-\mu)'B^{-1}(x-\mu)}dx_1\cdots dx_n=1,$$

其中 $x'=(x_1,\cdots,x_n)$．应用数学分析中积分变元法经过计算可得 $C=1/((2\pi)^{n/2}\sqrt{|B|}\ )$*)，其中 $|B|$ 为方阵 B 的行列式．所以我们称

$$(4.3.9)\quad f(x)=f(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}\sqrt{|B|}}e^{-\frac{1}{2}(x-\mu)'B^{-1}(x-\mu)}$$

为 n 维正态分布的密度函数，以它为联合分布密度的概率分布称为以 μ，B 为参数的 n 维正态分布，记作 $N(\mu,B)$．相应的 n 维连续型随机变量称为服从 $N(\mu,B)$的 n 维随机变量．

下面我们来导出二维正态分布密度的另一种常见的形式．令 $B=\begin{pmatrix}b_{11}&b_{12}\\b_{12}&b_{22}\end{pmatrix}$，由于 B 为正定方阵，所以 $b_{11}>0$，$b_{22}>0$，$b_{12}^2<b_{11}b_{22}$，因此可令 $b_{ii}\triangleq\sigma_i^2$，$\sigma_i>0$，$i=1,2$，$\rho\triangleq b_{12}/(\sigma_1\sigma_2)$．于是 $|B|=\sigma_1^2\sigma_2^2(1-\rho^2)$，$|\rho|<1$，

*)C 值的计算完全是纯分析的，在此略去．有兴趣的读者可参看 [15；第 129 页—131 页]．

$$B^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\sigma_1^{-2} & -\rho/(\sigma_1\sigma_2)\\ -\rho/(\sigma_1\sigma_2) & \sigma_2^{-2}\end{pmatrix}$$

二维正态分布密度

(4.3.10)

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right.\right.$$

$$\left.\left.-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

所以有时也将二维正态分布记作$N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$.

二维正态分布密度的图象是三维空间中的一个曲面，如图4－11所示．其形状像一顶"帽子"，用平行于X_1OX_2的平面截此曲面，所得截口都是椭圆；用平行于X_1OX_3或X_2OX_3且通过(μ,μ)的平面截曲面的截口是正态曲线．

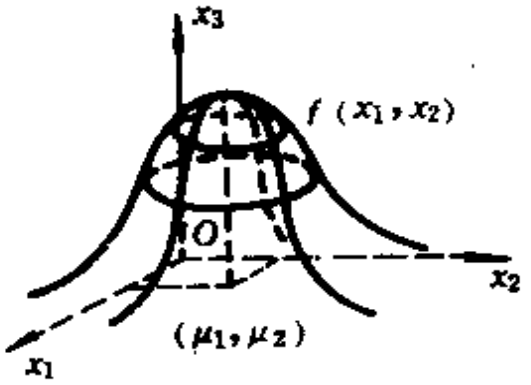


图 4－11

下面再给出联合分布密度的一些性质，为简单起见，有时结果叙述成一般情形以便应用，而证明只就二维情形讨论，有时则两者都是二维的．

定理4.3.11 设$f(x_1,\cdots,x_n)$是n维连续型随机变量的联合分布密度，G是n维空间中任可求积区域（或可数个可求积区域

之并），则有

(4.3.12) $P((\xi_1,\cdots,\xi_n)\in G)=\int\cdots\int_G f(x_1,\cdots,x_n)dx_1\cdots dx_n$.

证 为了书写简单，我们仅就二维情形证明.首先若 $G=\{(x,y):a_1<x\leqslant b_1,\ a_2<y\leqslant b_2\}$ (左开右闭矩形)，则

$$P((\xi_1,\xi_2)\in G)=P(a_1<\xi_1\leqslant b_1,\ a_2<\xi_2\leqslant b_2)$$

$$=P(\xi_1\leqslant b_1,a_2<\xi_2\leqslant b_2)-P(\xi_1\leqslant a_1,\ a_2<\xi_2\leqslant b_2)$$

$$=F(b_1,b_2)-F(b_1,a_2)-F(a_1,b_2)+F(a_1,a_2)$$

$$=\int_{-\infty}^{b_1}\int_{-\infty}^{b_2}f(x_1,x_2)dx_1dx_2-\int_{-\infty}^{b_1}\int_{-\infty}^{a_2}f(x_1,x_2)dx_1dx_2$$

$$-\int_{-\infty}^{a_1}\int_{-\infty}^{b_2}f(x_1,x_2)dx_1dx_2+\int_{-\infty}^{a_1}\int_{-\infty}^{a_2}f(x_1,x_2)dx_1dx_2$$

$$=\int_{-\infty}^{b_1}\int_{a_2}^{b_2}f(x_1,x_2)dx_1dx_2-\int_{-\infty}^{a_1}\int_{a_2}^{b_2}f(x_1,x_2)dx_1dx_2$$

$$=\int_{a_1}^{b_1}\int_{a_2}^{b_2}f(x_1,x_2)dx_1dx_2=\int_G\int f(x_1,x_2)dx_1dx_2.$$

即(4.3.12)对任何左开右闭矩形成立．其次由概率的上连续性

$$P(a_1\leqslant\xi_1\leqslant b_1,\ a_2\leqslant\xi_2\leqslant b_2)$$

$$=\lim_{n\to\infty}P\left(a_1-\frac{1}{n_1}<\xi_1\leqslant b_1,\ a_2-\frac{1}{n}<\xi_2\leqslant b_2\right)$$

$$=\lim_{n\to\infty}\int_{a_1-\frac{1}{n}}^{b_1}\int_{a_2-\frac{1}{n}}^{b_2}f(x_1,x_2)dx_1dx_2$$

$$=\int_{a_1}^{b_1}\int_{a_2}^{b_2}f(x_1,x_2)dx_1dx_2.$$

即 (4.3.12)对任何矩形成立.而在任何矩形边界上，(ξ_1, ξ_2)的概率为零，所以由概率的可加性及积分对积分区域的可加性知(4.3.12)对 G 是任何有限个除边界外两两不相交 (即任何两个都没有公共内点)的矩形的和成立. 再由于可求积区域可以由有限个矩形之并逼近，所以由概率及积分的上、下连续性知(4.3.12)对任何可求积区域成立*). 最后再由积分对积分区域的可加性及概率的可列可加性知 G 为可求积区域的可数并时，(4.3.12)成立.

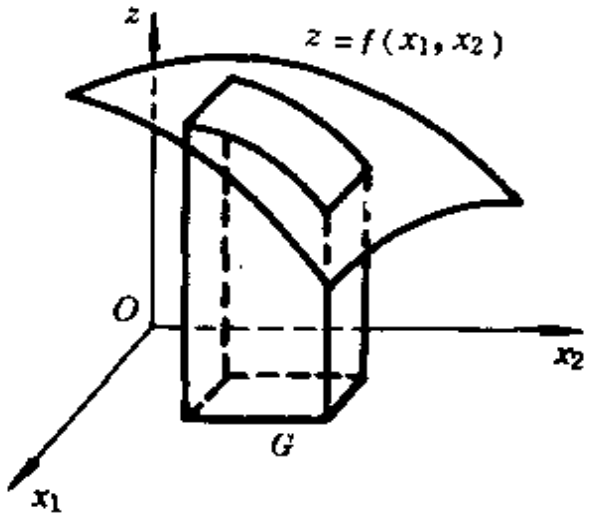


图 4－12

几何上，$f(x_1, x_2)$表示空间的一个曲面，(4.3.12)($n=2$ 的情形)表示(ξ_1, ξ_2)落入 G 的概率等于以 G 为底以 $z=f(x_1, x_2)$ 为顶的曲顶柱体的体积.

例 4.3.13 设(ξ_1, ξ_2)为具有分布密度

$$f(x_1, x_2)=\frac{1}{2\pi} e^{-(x_1^2+x_2^2)/2},\ x_1, x_2\in R$$

的二维正态随机变量，则对任意的正实数 t，由(4.3.12)有

$$P(\xi_1^2+\xi_2^2\leqslant t)=\frac{1}{2\pi}\iint\limits_{x_1^2+x_2^2\leqslant t} e^{-(x_1^2+x_2^2)/2}\,dx_1dx_2.$$

*) 此处的事实如确切说明需要一定的篇幅，超出了本书范围，有兴趣的读者可参看[15，第三章 §5.3] 有关部分.

将右端积分变量转换成极坐标，$x_1=r\cos\theta$，$x_2=r\sin\theta$，则

$$P(\xi_1^2+\xi_2^2\leqslant t)=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\sqrt{t}}e^{-r^2/2}\,rdrd\theta$$

$$=-e^{-r^2/2}\Big|_0^{\sqrt{t}}=1-e^{-t/2}=\int_0^t\frac{1}{2}e^{-u/2}du.$$

即 $\eta=\xi_1^2+\xi_2^2$ 服从参数为 $\frac{1}{2}$ 的指数分布.

定理 4.3.14 设(ξ,η)为具有联合分布密度$f(x,y)$的二维连续型随机变量，则 ξ，η 都是连续型随机变量，它们分别有分布密度

$$f_\xi(x)=\int_{-\infty}^{\infty}f(x,v)dv,\qquad x\in R^1,$$

$$f_\eta(y)=\int_{-\infty}^{\infty}f(u,y)du,\qquad y\in R^1.$$

证 ξ的分布函数

$$F_\xi(x)=P(\xi\leqslant x)=P((\xi,\eta)\in G),$$

其中$G=\{(u,v):u\leqslant x\}$显然为可数个可求积区域的并，所以由定理4.3.11知

$$F_\xi(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty}f(u,v)dudv=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty}f(u,v)dv\right]du.$$

且由$f(u,v)$非负知

$$\int_{-\infty}^{\infty}f(u,v)dv\geqslant 0.$$

所以ξ是连续型随机变量，且$\int_{-\infty}^{\infty}f(x,v)dv$是它的分布密度.

同样可证关于 η 的有关结论.

例 4.3.15 设 (ξ, η) 为在以原点为圆心，以 R 为半径的圆上均匀分布的二维随机变量，由例 4.3.6 知它是连续型随机变量. 今求 $f_\xi(x)$, $f_\eta(y)$ (注: 今后若在同一问题中涉及几个随机变量，则用 $F_\xi(\cdot)$, $f_\xi(\cdot)$ 分别表示 ξ 的分布函数、分布密度等).

解 由例 4.3.6 知 (ξ, η) 的分布密度

$$f(x,y)=\begin{cases}\dfrac{1}{\pi R^2}, & x^2+y^2\leqslant R^2,\\ 0, & x^2+y^2>R^2.\end{cases}$$

于是由定理 4.3.14 知: ξ 有分布密度

$$f_\xi(x)=\int_{-\infty}^{\infty} f(x,v)\mathrm{d}v=\begin{cases}\displaystyle\int_{x^2+y^2\leqslant R^2}\frac{1}{\pi R^2}\,dv=\frac{2\sqrt{R^2-x^2}}{\pi R^2}, & |x|\leqslant R,\\ 0, & |x|>R.\end{cases}$$

由对称性可得 η 有分布密度

$$f_\eta(y)=\begin{cases}\dfrac{2\sqrt{R^2-y^2}}{\pi R^2}, & |y|\leqslant R,\\ 0, & |y|>R.\end{cases}$$

解 4.3.16 二维正态分布的边缘分布. 设随机变量 (ξ, η) 服从二维正态分布 $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$，则由 (4.3.10) 知 (令 $u=\dfrac{x_1-\mu_1}{\sigma_1}$, $v=\dfrac{x_2-\mu_2}{\sigma_2}$)

$$f_\xi(x_1)=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left.(u^2-2\rho uv+v^2)\right\}dv$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left.[u^2-\rho^2u^2+(v-\rho u)^2]\right\}dv$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\,e^{-\frac{u^2}{2}}\int_{-\infty}^{\infty}e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}}dv$$

$$= \frac{1}{\sqrt{2\pi}\ \sigma_1}\,e^{-(x_1-\mu_1)^2/(2\sigma_1^2)}.$$

即 ξ 服从 $N(\mu_1,\sigma_1^2)$. 同法可证 η 服从 $N(\mu_2,\sigma_2^2)$，即

$$f_\eta(x_2)=\frac{1}{\sqrt{2\pi}\ \sigma_2}\,e^{-(x_2-\mu_2)^2/(2\sigma_2^2)}.$$

由此例可见，二维正态分布的边缘分布仍然是正态分布，但都不依赖于参数 ρ. 也就是说对于给定的 μ_1，μ_2，σ_1^2，σ_2^2 取不同的 ρ，得到不同的二维正态分布，但它们的边缘分布却是相同的，这又一次表明边缘分布不能确定联合分布.

定赖 4.3.14 可以推广成下列有用的结果.

定理 4.3.17 设 n 推连续型随机变量 $(\xi_1,\cdots,\xi_n)$ 有分布密度 $f(x_1,\cdots,x_n)$，$\xi_{i_1},\cdots,\xi_{i_m}$ $(1\leqslant m<n)$ 是 $(\xi_1,\cdots,\xi_n)$ 的 m 个分量，则 $(\xi_{i_1},\cdots,\xi_{i_m})$ 是 m 维连续型随机变量，它有分布密度

$$f_{(i_1,\cdots,i_m)}(x_{i_1},\cdots,x_{i_m})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)dx_{i_{m+1}}\cdots dx_{i_n},$$

其中 $\{i_{m+1},\cdots,i_n\}=\{1,2,\cdots,n\}\backslash\{i_1,\cdots,i_m\}$.

证明与定理 4.3.14 相同.

下面我们给出连续型随机变量独立的判别条件.

定理 4.3.18 (1) 若 $\xi_1,\cdots,\xi_n$ 是独立的连续型随机变量，它们分别具有分布密度 $f_{\xi_1}(x_1),\cdots,f_{\xi_n}(x_n)$，则 $(\xi_1,\cdots,\xi_n)$ 是 n 维

连续型随机变量，$f_{\xi_1}(x_1)\cdots f_{\xi_n}(x_n)$是它的联合分布密度．

(2) 若 n 维连续型随机变量 $(\xi_1, \cdots, \xi_n)$的联合分布密度$f(x_1, \cdots, x_n)$可以写成函数$f_1(x_1), \cdots f_n(x_n)$的乘积，其中$f_i(x_i)$只与变量 x_i 有关，$i=1, 2, \cdots, n$，则 $\xi_1, \cdots, \xi_n$ 为独立的连续型随机变量，而且$f_i(x_i)/\int_{-\infty}^{\infty} f_i(u_i)du_i$ 是 ξ_i 的分布密度．

证 先证(1)：因为 $\xi_1, \cdots, \xi_n$ 独立．所以由定理 4.1.9 知 $(\xi_1, \cdots, \xi_n)$的分布函数

$$F(x_1, \cdots, x_n) = F_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n).$$

由于 ξ_i 有分布密度$f_{\xi_i}(x_i)$，

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} f_{\xi_1}(u_1)du_1 \cdots \int_{-\infty}^{x_n} f_{\xi_n}(u_n)du_n$$

$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\xi_1}(u_1)\cdots f_{\xi_n}(u_n)du_1\cdots du_n.$$

故由联合分布密度的定义知$f_{\xi_1}(x_1)\cdots f_{\xi_n}(x_n)$是$(\xi_1, \cdots, \xi_n)$的联合分布密度，$(\xi_1, \cdots, \xi_n)$是 n 维连续型随机变量．

再证(2)：若(2)的假设成立，则$(\xi_1, \cdots, \xi_n)$的分布函数

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \cdots, u_n)du_1\cdots du_n$$

$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(u_1)\cdots f_n(u_n)du_1\cdots du_n$$

$$= \int_{-\infty}^{x_1} f_1(u_1)du_1 \int_{-\infty}^{x_2} f_2(u_2)du_2 \cdots \int_{-\infty}^{x_n} f_n(u_n)du_n$$

$$= G_1(x_1)\cdots G_n(x_n),$$

其中$G_i(x_i)=\int_{-\infty}^{x_i} f_i(u_i)du_i$，$i=1,2,\cdots,n$. 于是由定理 4.1.12 知 $\xi_1,\cdots,\xi_n$ 为独立的随机变量，而且 ξ_i 的分布函数

$$F_{\xi_i}(x_i)=\frac{G_i(x_i)}{G_i(+\infty)}=\int_{-\infty}^{x_i}\frac{f_i(u_i)}{\int_{-\infty}^{\infty}f_i(v)dv}du_i.$$

因而 ξ_i 是连续型随机变量，有分布密度 $f_i(x)\Big/\int_{-\infty}^{\infty}f_i(u)du$.

用定理 4.3.18 的(2)来判别连续型随机变量是否独立是非常方便的，因为它不需要验证 $f_i(x_i)$ 是不是 ξ_i 的分布密度.

例如，已知随机变量 (ξ,η) 具有分布密度

$$f(x,y)=\begin{cases}4xye^{-(x^2+y^2)}, & x>0,y>0,\\ 0, & \text{其它}.\end{cases}$$

则 ξ,η 为独立的连续型随机变量.

事实上，令

$$f_i(x)=\begin{cases}2xe^{-x^2}, & x>0,\\ 0, & x\leqslant 0,\end{cases}\quad i=1,2.$$

则 $f(x,y)=f_1(x)f_2(y)$. 所以由定理 4.3.18 知 ξ,η 独立且为连续型随机变量. 由于 $\int_{-\infty}^{\infty}f_i(x)dx=\int_0^{\infty}2xe^{-x^2}dx=-e^{-x^2}\Big|_0^{\infty}=1$，所以 $f_\xi(x)=f_\eta(x)=f_1(x)=f_2(x)$.

例 4.3.19 若随机变量 (ξ,η) 服从 $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，则 ξ,η 独立的充分必要条件是 $\rho=0$.

一般地，若 $(\xi_1,\cdots,\xi_n)$ 服从 $N(\mu,B)$，则 $\xi_1,\cdots,\xi_n$ 独立的充分必要条件是 B 为 n 阶正定对角形方阵. 在此条件下，ξ_i 服从

$N(\mu_i, b_{ii})$，$i=1, 2, \cdots, n$，其中$(\mu_1, \cdots, \mu_n)=\mu'$（$\mu$的转置），$B=(b_{ij})_{i,j=1}^n$.

证 若$\rho=0$，则(4.3.10)成为

$$f(x_1, x_2)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right\}$$

$$\cdot\frac{1}{\sqrt{2\pi}\ \sigma_2}\exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right\},$$

所以由定理4.3.18 (2)知(ξ, η)独立. 反之若ξ, η独立，且ξ, η分别服从$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$，则由定理4.3.18(1)知(ξ, η)有分布密度

$$\frac{1}{\sqrt{2\pi}\ \sigma_1}\exp\left\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right\}$$

$$\cdot\frac{1}{\sqrt{2\pi}\ \sigma_2}\exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_1^2}\right\}.$$

与(4.3.10)比较，即得$\rho=0$.

一般地，若B为对角方阵，即$b_{ij}=0$，当$i\neq j$，则二次型$(x-\mu)'B^{-1}(x-\mu)=\sum_{i=1}^n\frac{(x_i-\mu_i)^2}{b_{ii}}$，于是$N(\mu, B)$的分布密度(4.3.9)就可写成

$$f(x_1, \cdots, x_n)=\frac{1}{\sqrt{2\pi b_{11}}}e^{-(x_1-\mu_1)^2/(2b_{11})}$$

$$\cdots\frac{1}{\sqrt{2\pi b_{nn}}}e^{-(x_n-\mu_n)^2/(2b_{nn})}.$$

故由定理4.3.18(2)知$\xi_1, \cdots, \xi_n$独立. 再因

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi b_{ii}}}e^{-(x_i-\mu_i)^2/(2b_{ii})}dx_i=1,$$

ξ_i 服从 $N(\mu_i, b_{ii})$，$i=1, 2, \cdots, n$.

反之，若 ξ_i 服从 $N(\mu_i, b_{ii})$，$i=1, 2, \cdots, n$，独立，则由定理 4.3.18(1)知 $(\xi_1, \cdots, \xi_n)$ 有分布密度

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi b_{ii}}} e^{-(x_i-\mu_i)^2/(2b_{ii})}$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{b_{11}\cdots b_{nn}}} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{b_{ii}}}.$$

与(4.3.9)比较即知 $b_{ij}=0$，$i\neq j$.

习题 4－3

1. 设 (ξ, η) 的联合分布密度 $f(x, y)=ce^{-(x+1)^2-|y|}$. (1)求 c；(2)求关于 ξ、η 的边缘分布密度；(3)求 (ξ, η) 的联合分布函数；(4)ξ 与 η 独立吗?

2. 设 (ξ, η) 的联合分布密度

$$f(x, y)=\begin{cases} \cos x \cdot \cos y, & 0<x, y<\dfrac{\pi}{2}, \\ 0, & \text{其它}. \end{cases}$$

(1)求 (ξ, η) 的联合分布函数；(2)求关于 ξ，η 的边缘分布密度；(3)ξ 与 η 独立吗? (4)求 $P(\eta<2\xi)$.

3. 设 (ξ, η) 在顶点为 $(0, 0)$，$(a, 0)$，(a, b)，$(0, b)$ 的矩形上均匀分布. (1)求 (ξ, η) 的联合分布密度；(2)求 (ξ, η) 的联合分布函数；(3)用以下两种方法求 $P\left(0\leqslant\xi\leqslant\dfrac{a}{2}, 0\leqslant\eta\leqslant\dfrac{b}{3}\right)$：第一种方法借助于分布密度，第二种方法借助于分布函数.

4. 设独立随机变量的密度函数分别为

$$f_\xi(x)=\begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}}, & |x|<1, \\ 0, & |x|\geqslant 1; \end{cases}$$

$$f_\eta(y)=\frac{1}{\pi(1+y^2)}.$$

求：(1)(ξ, η) 的分布密度；(2)联合分布函数.

5. 设 (ξ,η)相互独立且都服从 $[0,1]$ 上的均匀分布，求 $x^2+\xi x+\eta=0$ 有实根的概率.

6. 设 (ξ,η)在被直线 $x=2$，$y=0$，$y=x$ 所围成的三角形上均匀分布. (1)求联合分布密度及联合分布函数；(2)求边缘分布密度及边缘分布函数；(3)求 (ξ,η)落在上述三角形的内切圆中的概率；(4)ξ 与 η 独立吗?

7. 设 (ξ,η,ζ)的联合分布密度为

$$f(x,y,z)=\begin{cases}\dfrac{1}{8\pi^2}(1-\sin x\sin y\sin z), & 0\leqslant x,y,z\leqslant 2\pi,\\ 0, & \text{其它}.\end{cases}$$

证明 ξ,η,ζ 两两独立但不独立.

§4.4 多维随机变量函数的分布

在一些实际问题中，有时需要研究随机变量的函数. 例如悬浮在液体中的花粉在某一坐标系下速度的三个分量 v_1,v_2,v_3 都是随机变量，已知它们的联合分布，要求它的动能 $E=\frac{1}{2}m(v_1^2+v_2^2+v_3^2)$的分布. 又如在数理统计中，我们经常需要对作为总体的随机变量 ξ 的期望、方差等通过样本 $\xi_1,\cdots,\xi_n$ 来进行估计 (有关概念参看命题 2.4.7 后面的一段介绍)，例如用样本的平均$\overline{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$来估计 $E\xi$，用样本的方差 $S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2$ 来估计 $D\xi$ (参看第五章). 这时很自然要研究$\overline{\xi}$，S^2 的分布，也就是 $\xi_1,\cdots,\xi_n$ 的函数的分布.

设 $\xi=(\xi_1,\cdots,\xi_n)$是一个 n 维随机变量，且 $g(x_1,\cdots,x_n)$是 R^n 上的函数使得 $\eta(\omega)=g(\xi_1(\omega),\cdots,\xi_n(\omega))$作为 Ω 上的函数也是一个随机变量. 若已知$(\xi_1,\cdots,\xi_n)$的联合分布，怎样求出 $\eta=g(\xi_1,\cdots,\xi_n)$的分布呢? 和一维情形一样，从原则上讲，我们可以如下表示 η 的分布函数

(4.4.1)
$$F_\eta(y)=P(g(\xi_1,\cdots,\xi_n)\leqslant y)$$
$$=P((\xi_1,\cdots,\xi_n)\in\{(x_1,\cdots,x_n):g(x_1,\cdots,x_n)\leqslant y\}),$$

但是只有在一些比较简单的情形才能求出 η 的分布函数，本节的目的便是讨论一些具体情形（着重离散型和连续型随机变量）.

定理 4.4.2 和的分布. (1) 设 (ξ, η) 为二维离散型随机变量，分布列为

$$p_{ij}=P(\xi=x_i, \eta=y_j), i, j=1, 2, \cdots.$$

令 $\zeta=\xi+\eta$. 则 ζ 可能取的值的集合为 x_i+y_j, $i, j=1, 2, \cdots$ 中不同的数的集合，记作 $\{z_k: k=1, 2, \cdots\}$ 且对任何 z_k,

(4.4.3)
$$P(\zeta=z_k)=\sum_i P(\xi=x_i, \eta=z_k-x_i)$$
$$=\sum_j P(\xi=z_k-y_j, \eta=y_j),$$

若 (ξ, η) 独立，则

(4.4.4)
$$P(\zeta=z_k)=\sum_i P(\xi=x_i)P(\eta=z_k-x_i)$$
$$=\sum_j P(\xi=z_k-y_j)P(\eta=y_j).$$

(2) 设 (ξ, η) 为连续型随机变量，分布密度为 $f(x, y)$. 令 $\zeta=\xi+\eta$，则 ζ 为连续型随机变量，且有分布密度

(4.4.5)
$$f_\zeta(z)=\int_{-\infty}^{\infty} f(x, z-x)dx=\int_{-\infty}^{\infty} f(z-y, y)dy.$$

若 ξ, η 还独立，则 ζ 有分布密度

(4.4.6)
$$f_\zeta(z)=\int_{-\infty}^{\infty} f_\xi(x)f_\eta(z-x)dx=\int_{-\infty}^{\infty} f_\xi(z-y)f_\eta(y)dy,$$

即两个独立的连续型随机变量 ξ, η 的和的分布密定是 ξ, η 的密定的卷积.

证 我们只对连续型情形证明，离散型的证明留给读者.

由定理 4.3.11 知，对任意实数 z,

$$P(\zeta\leqslant z)=P(\xi+\eta\leqslant z)$$
$$=P((\xi, \eta)\in G_z)=\iint_{G_z} f(x, y)dxdy,$$

其中 $G_z \triangleq \{(x,y): x+y \leqslant z\}$. 将重积分化为累次积分并使用变量替换 $u-x=y$，则由上式得知，对任意实数 z 有

$$F_\zeta(z)=\int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f(x,y)dy=\int_{-\infty}^{\infty} dx \int_{-\infty}^{z} f(x,u-x)du$$

$$=\int_{-\infty}^{z} du \int_{-\infty}^{\infty} f(x,u-x)dx .$$

因而 ζ 是一具有分布密度 $\int_{-\infty}^{\infty} f(x,z-x)dx$ 的连续型随机变量，在上述表达式中作变量替换 $y=z-x$，即得 (4.4.5)的第二个等式.

当 ξ，η 独立时，可取 $f_\xi(x)f_\eta(y)$ 为 $f(x,y)$，于是 ζ 有分布密度 (4.4.6).

下面举一些例子说明上述定理的应用.

例 4.4.7 设随机变量 ξ，η 独立，且分别服从参数为 λ_1，λ_2 的普哇松分布. 求 $\zeta=\xi+\eta$ 的分布列.

解 由于 ξ，η 可取一切非负整值. 所以 $\zeta=\xi+\eta$ 也可取一切非负整值. 由 ξ，η 独立及 (4.4.4)知对任何非负整数 l，

$$P(\zeta=l)=\sum_{k=0}^{l} P(\xi=k)P(\eta=l-k)$$

$$=\sum_{k=0}^{l} e^{-\lambda_1}\cdot\frac{\lambda_1^k}{k!}\cdot e^{-\lambda_2}\cdot\frac{\lambda_2^{l-k}}{(l-k)!}$$

$$=e^{-(\lambda_1+\lambda_2)}\frac{1}{l!}\sum_{k=0}^{l}\frac{l!}{k!\,(l-k)!}\lambda_1^k\lambda_2^{l-k}$$

$$=e^{-(\lambda_1+\lambda_2)}\cdot\frac{(\lambda_1+\lambda_2)^l}{l!},$$

这说明 ζ 服从参数为 $\lambda_1+\lambda_2$ 的普哇松分布.

例 4.4.8 设随机变量 ξ_1, ξ_2 独立，且分别服从参数为 λ_1, λ_2 的指数分布，求 $\zeta=\xi_1+\xi_2$ 的分布.

解 由于

$$f_{\xi_i}(x)=\begin{cases}\lambda_i e^{-\lambda_i x}, & x\geqslant 0,\\ 0, & \text{其它},\end{cases}\qquad i=1,2.$$

于是由 ξ_1, ξ_2 的独立及定理 4.4.2 的 (4.4.6)知当 $z\geqslant 0$ 时

$$f_\zeta(z)=\int_0^z \lambda_1\lambda_2 e^{-\lambda_1 x}\cdot e^{-\lambda_2(z-x)}dx$$

$$=\begin{cases}\dfrac{\lambda_1\lambda_2}{\lambda_1-\lambda_2}\left[e^{-\lambda_2 z}-e^{-\lambda_1 z}\right], & \lambda_1\neq\lambda_2,\\ \lambda_1\lambda_2 z e^{-\lambda_2 z}, & \lambda_1=\lambda_2.\end{cases}$$

而当 $z<0$ 时，$f_\zeta(z)=0$. 故有以下两种情况：

(1) 当 $\lambda_1\neq\lambda_2$ 时，$f_\zeta(z)=\dfrac{\lambda_1\lambda_2}{\lambda_1-\lambda_2}\left[e^{-\lambda_2 z}-e^{-\lambda_1 z}\right]$, $z\geqslant 0$; $f_\zeta(z)=0$, $z<0$.

(2) 当 $\lambda_1=\lambda_2$ 时，$f_\zeta(z)=\lambda_1^2 z e^{-\lambda_1 z}$, $z\geqslant 0$; $f_\zeta(z)=0$, $z<0$. 由定义 3.3.22 及 $\Gamma(2)=1$ 知，此时 ζ 服从参数为 $(2,\lambda_1)$ 的 Γ 分布. 而且由该定义知参数为 $(1,\lambda_1)$ 的 Γ 分布就是参数为 λ_1 的指数分布，所以此事实说明：若 $\xi_i\sim\Gamma(1,\lambda)$, $i=1,2$, 独立，则 $\xi_1+\xi_2\sim\Gamma(2,\lambda)$.

更一般地有

例 4.4.9 设随机变量 $\xi_i\sim\Gamma(t_i,\lambda)$, $t_i>0$, $\lambda>0$, $i=1,2$, 且 ξ_1, ξ_2 独立，则 $\xi_1+\xi_2\sim\Gamma(t_1+t_2,\lambda)$.

证 由定义 3.3.22 知

$$f_{\xi_i}(x)=\begin{cases}0, & x<0,\\ \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t_i-1}}{\Gamma(t_i)}, & x\geqslant 0,\end{cases}\qquad i=1,2,\ x\in R^1.$$

再由定理 4.4.2 的 (4.4.6)知，当 $z<0$ 时 $f_{\xi_1+\xi_2}(z)=0$，当 $z\geqslant 0$ 时，

$$f_{\xi_1+\xi_2}(z)=\int_0^z \frac{\lambda e^{-\lambda x}(\lambda x)^{t_1-1}}{\Gamma(t_1)}\cdot\frac{\lambda e^{-\lambda(z-x)}[\lambda(z-x)]^{t_2-1}}{\Gamma(t_2)}dx$$

$$=\frac{e^{-\lambda z}\lambda^{t_1+t_2}}{\Gamma(t_1)\Gamma(t_2)}\int_0^z x^{t_1-1}(z-x)^{t_2-1}dx \qquad (x=zu)$$

$$=\frac{e^{-\lambda z}z^{t_1+t_2-1}\lambda^{t_1+t_2}}{\Gamma(t_1)\Gamma(t_2)}\int_0^1 u^{t_1-1}(1-u)^{t_2-1}du$$

$$=\frac{\lambda e^{-\lambda z}(\lambda z)^{t_1+t_2-1}}{\Gamma(t_1+t_2)}.$$

最后一步用到“数学分析”中的下列公式：

$$\int_0^1 u^{t_1-1}(1-u)^{t_2-1}du=B(t_1,t_2)=\frac{\Gamma(t_1)\Gamma(t_2)}{\Gamma(t_1+t_2)},$$

其中 $B(t_1,t_2)$表示贝塔 (B)函数 [19，第 654 页，B 函数]。

例 4.4.10 设二维随机变量 (ξ,η)服从 $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，求 $\zeta=\xi+\eta$ 的分布。

解 由定理 4.4.2 的 (4.4.5)及(4.3.10)知

$$f_\zeta(z)=\int_{-\infty}^{\infty} f(x,z-x)dx$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left.\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(z-x-\mu_2)}{\sigma_1\sigma_2} + \frac{(z-x-\mu_2)^2}{\sigma_2^2}\right]\right\} dx .$$

令 $u=x-\mu_1$，$v=z-\mu_1-\mu_2$，$\sigma^2=\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2$，则上式右端的方括号中的式子等于

$$\frac{u^2}{\sigma_1^2} - \frac{2\rho u(v-u)}{\sigma_1\sigma_2} + \frac{(v-u)^2}{\sigma_2^2}$$

$$= \left[\frac{\sigma}{\sigma_1\sigma_2}u - \frac{(\sigma_1+\rho\sigma_2)}{\sigma_2\sigma}v\right]^2 + \frac{(1-\rho^2)}{\sigma^2}v^2 .$$

再令 $t=\frac{1}{\sqrt{1-\rho^2}}\left[\frac{\sigma}{\sigma_1\sigma_2}u - \frac{(\sigma_1+\rho\sigma_2)}{\sigma_2\sigma}v\right]$，则 $dx=du$

$$= \frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\sigma}dt ,$$

$$f_\zeta(z) = \frac{1}{2\pi\sigma}\exp\left\{-\frac{v^2}{2\sigma^2}\right\}\int_{-\infty}^{\infty} e^{-t^2/2}dt$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(z-\mu_1-\mu_2)^2}{2\sigma^2}\right\}$$

即 $\zeta \sim N(\mu_1+\mu_2, \sigma^2) = N(\mu_1+\mu_2, \sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2)$．特别当 $\rho=0$ 时（即 ξ,η 独立），$\xi+\eta \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$．

定理 4.4.11　（商的分布）．　设 (ξ,η) 为连续型随机变量，分布密度为 $f(x,y)$，则 $\zeta=\xi/\eta$ 为连续型随机变量，且有分布密度

$$f_\zeta(z)=\int_{-\infty}^{\infty} f(yz,y)|y|dy, \tag{4.4.12}$$

特别若 ξ，η 独立，则 ζ 有分布密度

$$f_\zeta(z)=\int_{-\infty}^{\infty} f_\xi(yz)f_\eta(y)|y|dy. \tag{4.4.13}$$

证 首先注意 ζ 在 $\eta=0$ 时没有定义，但由于 η 为连续型，$P(\eta=0)=0$，所以 ζ 在 $\eta=0$ 时可以随意给定一值而不影响它的分布．其次由定理 4.3.18 知 ξ，η 独立时，$f_\xi(x)f_\eta(y)$ 是 (ξ,η) 的分布密度．所以由 (4.4.12) 立得 (4.4.13)．综上所述，只需证 (4.4.12)．

由 (4.4.1′) 及定理 4.3.11 知

$$F_\zeta(z)=P\left((\xi,\eta)\in\left\{(x,y):\ \frac{x}{y}\leqslant z\right\}\right)$$

$$=\int_{\frac{x}{y}\leqslant z}\int f(x,y)dxdy,$$

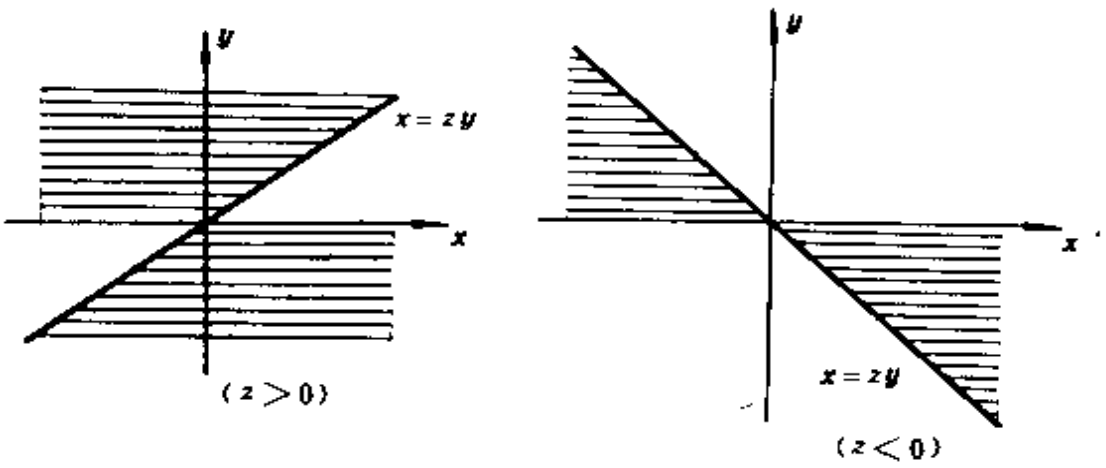


图 4－13

其中积分区域如图 4－13 中的阴影部分．故

$$F_\zeta(z)=\int_0^{\infty}dy\int_{-\infty}^{zy}f(x,y)dx+\int_{-\infty}^{0}dy\int_{zy}^{\infty}f(x,y)dx.$$

作变量替换 $x=yu$，并交换积分顺序得

$$F_\zeta(z)=\int_0^{\infty}dy\int_{-\infty}^{z}f(yu,y)ydu+\int_{-\infty}^{0}dy\int_{z}^{-\infty}f(yu,y)ydu$$

$$=\int_{-\infty}^{z}du\int_{-\infty}^{\infty}f(yu,y)|y|\,dy.$$

因此 (4.4.12)是 ζ 的分布密度．

类似地还有

定理 4.4.14 (积的分布)． 设(ξ,η)为连续型随机变量，分布密度为$f(x,y)$，则 $\zeta=\xi\eta$ 为连续型随机变量，且有分布密度

(4.4.15)． $$f_\zeta(z)=\int_{-\infty}^{\infty}f\left(\frac{z}{y},y\right)\left|\frac{1}{y}\right|dy.$$

若 ξ,η 独立，则

(4.4.16) $$f_\zeta(z)=\int_{-\infty}^{\infty}f_\xi\left(\frac{z}{y}\right)f_\eta(y)\frac{1}{|y|}dy.$$

证明与定理 4.4.11 类似，留给读者．

一般地，若 $\xi=(\xi_1,\cdots,\xi_n)$为连续型随机变量，其分布密度为 $f(x_1,\cdots,x_n)$，$g_i(x_1,\cdots,x_n)(i=1,\cdots,m)$为使得 $g_i(\xi_1,\cdots,\xi_n)$ 为随机变量的 R^n 上的函数 (例如 R^n 上的连续函数)，令

$$\eta=(\eta_1,\cdots,\eta_m)=(g_1(\xi_1,\cdots,\xi_n),\cdots,g_m(\xi_1,\cdots,\xi_n)),$$

则在定理 4.3.11 中取 $G=\{(x_1,\cdots,x_n):g_i(x_1,\cdots,x_n)\leqslant y_i,i=1,\cdots,m\}$即得 η 的分布函数

(4.4.17) $$F_\eta(y_1,\cdots,y_m)=P(\eta_1\leqslant y_1,\cdots,\eta_m\leqslant y_m)$$
$$=P(g_1(\xi_1,\cdots,\xi_n)\leqslant y_1,\cdots,g_m(\xi_1,\cdots,\xi_n)\leqslant y_m)$$
$$=\int\cdots\int_{\substack{g_i(x_1,\cdots,x_n)\leqslant y_i\\ i=1,2,\cdots,m}} f(x_1,\cdots,x_n)dx_1\cdots dx_n.$$

由此对于具体的函数$g_i(x_1,\cdots,x_n)$即可求出η的分布函数.

例 4.4.18 若ξ，η为独立随机变量，且具有相同的指数分布密度

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

试求$\zeta_1=\xi+\eta$，$\zeta_2=\xi/\eta$的密度函数$f_{(\zeta_1,\zeta_2)}(y_1,y_2)$.

解 由定理 4.3.18 知(ξ,η)有分布密度

$$f_{(\xi,\eta)}(x,y)=f(x)f(y)=\begin{cases}e^{-(x+y)}, & x>0,y>0,\\ 0, & \text{其它}.\end{cases}$$

由 (4.4.17)知

$$F_{(\zeta_1,\zeta_2)}(y_1,y_2)=\int\int_{\substack{x+y\leqslant y_1,\ x/y\leqslant y_2\\ 0<x,y<\infty}} e^{-(x+y)}dxdy.$$

作变量替换$u=x+y$，$v=\dfrac{x}{y}$，则$x=\dfrac{uv}{1+v}$，$y=\dfrac{u}{1+v}$，雅可比行列式为

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}\dfrac{v}{1+v} & \dfrac{u}{(1+v)^2}\\ \dfrac{1}{1+v} & \dfrac{-u}{(1+\ )^2}\end{vmatrix}=\frac{-u}{(1+v)^2}$$

于是当$y_1, y_2>0$时，

$$F_{(\zeta_1,\zeta_2)}(y_1, y_2)=\int_0^{y_1} du\int_0^{y_2} e^{-u}\cdot\frac{u}{(1+v)^2}\,dv.$$

当y_1, y_2不全为正时，$F_{(\zeta_1,\zeta_2)}(y_1, y_2)=0$. 故

$$f_{(\zeta_1,\zeta_2)}(y_1, y_2)=\begin{cases} y_1e^{-y_1}\cdot\dfrac{1}{(1+y_2)^2}, & y_1, y_2>0,\\ 0, & \text{其它}. \end{cases}$$

由独立性判别准则（定理 4.3.18 (2)）知 ζ_1 与 ζ_2 独立，再由 $\int_0^{\infty}\frac{dx}{(1+x)^2}=1$ 知ζ_1, ζ_2的分布密度为

$$f_{\zeta_1}(x)=\begin{cases} xe^{-x}, & x>0,\\ 0, & x\leqslant 0, \end{cases}$$

$$f_{\zeta_2}(x)=\begin{cases} \dfrac{1}{(1+x)^2}, & x>0,\\ 0, & x\leqslant 0. \end{cases}$$

下面我们应用上面的一般理论，讨论数理统计及实际中有用的一些分布和结论.

定理 4.4.19 设$\xi_1, \xi_2, \cdots, \xi_n$为 n 个独立的标准正态分布的随机变量，$C=(c_{ij})$为 n 阶正交方阵. 即 $C\ \tilde{C}=CC'=I$, C' 为 C 的转置. 若 $\eta_k=\sum_{l=1}^{n} c_{kl}\xi_l$, $k=1, 2, \cdots, n$, 则 $\eta_1, \cdots, \eta_n$ 仍为 n 个独立的标准正态分布的随机变量.

证 由定理 4.3.18 (1) 知 n 维随机变量 $(\xi_1, \cdots, \xi_n)$有分布密度 $(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n} x_k^2\right\}$，故由(4.4.17)知 $\eta=(\eta_1, \cdots, \eta_n)$的分布函数

$$F_\eta(y_1,\cdots,y_n)=\int\cdots\int_{\substack{\sum_{l=1}^{n}c_{kl}x_l\leqslant y_k\\ k=1,\cdots,n}}\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}(x_1^2+\cdots+x_n^2)}dx_1\cdots dx_n.$$

作变量替换 $t_k=\sum_{l=1}^{n}c_{kl}x_l$，$k=1,\cdots,n$，则 $x_k=\sum_{l=1}^{n}c_{lk}t_l$，$k=1,\cdots,n$，替换的雅可比行列式

$$J=\left|\frac{\partial x_k}{\partial t_l}\right|=|C'|,\ |J|=1.$$

且由 $CC'=I$ 知 $\sum_{k=1}^{n}c_{lk}c_{mk}=\begin{cases}1, & l=m\\ 0, & l\neq m\end{cases}$，于是

$$\sum_{k=1}^{n}x_k^2=\sum_{k=1}^{n}\left(\sum_{l=1}^{n}c_{lk}t_l\right)^2=\sum_{k=1}^{n}\sum_{l,m=1}^{n}c_{lk}c_{mk}t_lt_m$$

$$=\sum_{l,m=1}^{n}t_lt_m\sum_{k=1}^{n}c_{lk}c_{mk}=\sum_{l=1}^{n}t_l^2.$$

故

$$F_\eta(y_1,\cdots,y_n)=\int_{-\infty}^{y_1}\cdots\int_{-\infty}^{y_n}\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}(t_1^2+\cdots+t_n^2)}dt_1\cdots dt_n$$

$$=\prod_{l=1}^{n}\int_{-\infty}^{y_l}\frac{1}{\sqrt{2\pi}}e^{-t_l^2/2}dt_l,$$

因此由定理 4.1.9 知 $\eta_1,\cdots,\eta_n$ 都是标准正态分布且彼此独立.

例 4.4.20 若随机变量 $\xi_1,\cdots,\xi_n$ 独立且 $\xi_k\sim N(\mu_k,\sigma_k^2)$，$k=1,\cdots,n$，$\eta=\xi_1+\cdots+\xi_n$，则 $\eta\sim N\left(\sum_{k=1}^{n}\mu_k,\sum_{k=1}^{n}\sigma_k^2\right)$.

证法 1 由例4.4.10 中 $\rho=0$ 的特例知 $\xi_1+\xi_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$. 假设 $\xi_1+\cdots+\xi_k\sim N(\mu_1+\cdots+\mu_k,\sigma_1^2+\cdots+\sigma_k^2)$，则由定理 4.1.14 及定理 4.1.16 知 $(\xi_1,\cdots,\xi_k)$，ξ_{k+1} 独立，再由定理 4.1.17 知 $\xi_1+\cdots+\xi_k$ 与 ξ_{k+1} 独立. 由上述假设及例 4.4.10 中的特例知 $(\xi_1+\cdots+\xi_k)+\xi_{k+1}\sim N((\mu_1+\cdots+\mu_k)+\mu_{k+1},(\sigma_1^2+\cdots+\sigma_k^2)+\sigma_{k+1}^2)$，故由归纳法知所说结论成立.

***证法 2** 应用上述定理，容易计算出

$$\eta=\sqrt{\sum_{k=1}^{n}\sigma_k^2}\left(\sum_{l=1}^{n}\frac{\sigma_l}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}}\cdot\frac{\xi_l-\mu_l}{\sigma_l}\right)+\sum_{k=1}^{n}\mu_k.$$

于是只需证：$\sum\limits_{l=1}^{n}\dfrac{\sigma_l}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}}\cdot\dfrac{\xi_l-\mu_l}{\sigma_l}$ 为标准正态分布.

由定理 4.1.17 知 $\dfrac{\xi_1-\mu_1}{\sigma_1},\cdots,\dfrac{\xi_n-\mu_n}{\sigma_n}$ 为独立的标准正态随机变量（也不难直接证明），又

$$\left(\frac{\sigma_1}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}}\right)^2+\cdots+\left(\frac{\sigma_n}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}}\right)^2=1,$$

由线性代数知存在 n 阶正交方阵 $C=(c_{kl})$ 使 $\left(\dfrac{\sigma_1}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}},\cdots,\dfrac{\sigma_n}{\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}}\right)$ 为它的第一行，令 $\eta_k=\sum\limits_{l=1}^{n}c_{kl}\dfrac{\xi_l-\mu_l}{\sigma_l}$，$k=1,\cdots,n$，则由定理 4.4.19 知 $\eta_1,\cdots,\eta_n$ 为独立的标准正态随机变量，特别，

$$\eta_1=\sum_{i=1}^{n}c_{1i}\frac{\xi_i-\mu_i}{\sigma_i}=\sum_{i=1}^{n}\frac{\sigma_i}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}}\cdot\frac{\xi_i-\mu_i}{\sigma_i}\sim N(0,1).$$

所以

$$\eta=\sqrt{\sum_{k=1}^{n}\sigma_k^2}\ \eta_1+\sum_{k=1}^{n}\mu_k\sim N\left(\sum_{k=1}^{n}\mu_k,\sum_{k=1}^{n}\sigma_k^2\right).$$

注:例 4.4.20 的证法 2 实际上证明了:n 个独立正态随机变量 $\xi_k\sim N(\mu_k,\sigma_k^2)$,$k=1,2,\cdots,n$ 的任意线性组合 $a_1\xi_1+a_2\xi_2+\cdots+a_n\xi_n$ 为正态变量 $N\left(\sum_{k=1}^{n}a_k\mu_k,\sum_{k=1}^{n}a_k^2\sigma_k^2\right)$.这样若 $\xi_1,\cdots,\xi_n$ 独立且都服从 $N(\mu,\sigma^2)$,则

$$\overline{\xi}=\frac{1}{n}\sum_{k=1}^{n}\xi_k\sim N\left(\mu,\frac{\sigma^2}{n}\right).$$

这个结论在数理统计中常常用到.

例 4.4.21 若分子运动的速度的三个分量 v_1,v_2,v_3 独立,且都服从 $N(0,\sigma^2)$,试求速度(绝对值)$\sqrt{v_1^2+v_2^2+v_3^2}$ 的分布.

解 令 $\zeta=\sqrt{v_1^2+v_2^2+v_3^2}$,于是当 $x\leqslant 0$ 时,$F_\zeta(x)=0$.当 $x>0$ 时,由于 (v_1,v_2,v_3) 具有分布密度

$$f(x_1,x_2,x_3)=\frac{1}{(2\pi\sigma^2)^{3/2}}e^{-\frac{1}{2\sigma^2}(x_1^2+x_2^2+x_3^2)},\ x_1,x_2,x_3\in R^1,$$

$$F_\zeta(x)=P(v_1^2+v_2^2+v_3^2\leqslant x^2)\iiint\limits_{\sum_{i=1}^{3}x_i^2\leqslant x^2}\frac{1}{(2\pi\sigma^2)^{3/2}}\cdot e^{-\frac{1}{2\sigma^2}(x_1^2+x_2^2+x_3^2)}dx_1dx_2dx_3.$$

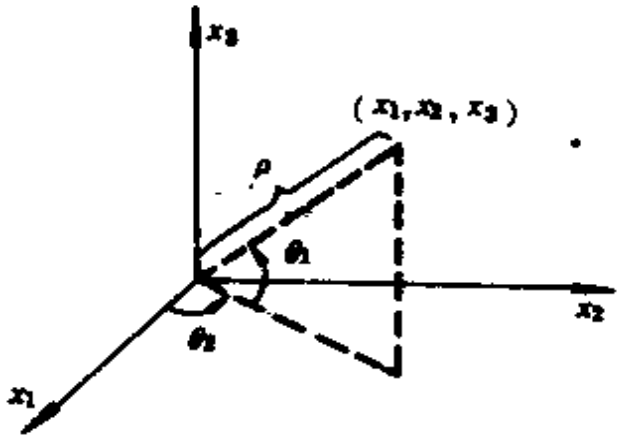


图 4－14

变换成球坐标：

$$x_1 = \rho \cos\theta_1 \cos\theta_2 ,$$
$$x_2 = \rho \cos\theta_1 \sin\theta_2 ,$$
$$x_3 = \rho \sin\theta_1 ,$$

$$0 \leqslant \rho < \infty , \ -\frac{\pi}{2} \leqslant \theta_1 \leqslant \frac{\pi}{2} , \ -\pi \leqslant \theta_2 \leqslant \pi .$$

则雅可比行列式

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial \rho} & \dfrac{\partial x_1}{\partial \theta_1} & \dfrac{\partial x_1}{\partial \theta_2} \\ \dfrac{\partial x_2}{\partial \rho} & \dfrac{\partial x_2}{\partial \theta_1} & \dfrac{\partial x_2}{\partial \theta_2} \\ \dfrac{\partial x_3}{\partial \rho} & \dfrac{\partial x_3}{\partial \theta_1} & \dfrac{\partial x_3}{\partial \theta_2} \end{vmatrix} = -\rho^2 \cos\theta_1 .$$

故有

$$F_\zeta(x) = \int_0^x d\rho \int_{-\pi}^{\pi} d\theta_2 \int_{-\pi/2}^{\pi/2} \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\rho^2/(2\sigma^2)} \rho^2 \cos\theta_1 d\theta_1$$

$$= \frac{1}{\sigma^3}\sqrt{\frac{2}{\pi}}\int_0^x \rho^2 e^{-\rho^2/(2\sigma^2)}d\rho \ .$$

因而 ζ 有分布密度

$$f_\zeta(x)=\begin{cases}\dfrac{1}{\sigma^3}\sqrt{\dfrac{2}{\pi}}\ x^2 e^{-x^2/(2\sigma^2)}, & x>0\ ,\\ 0\ , & x\leqslant 0\ .\end{cases}$$

这一分布在统计物理中称为马克斯威尔(Maxwell)分布.

一般地，用类似的方法有

例 4.4.22 若 $\xi_1,\cdots,\xi_n$ 为独立的标准正态随机变量，$\chi_n \triangleq \sqrt{\sum_{k=1}^n \xi_k^2}$ ，则 χ_n 有分布密度

$$(4.4.23)\qquad f_{\chi_n}(x)=\begin{cases}\dfrac{1}{2^{n/2-1}\Gamma\left(\dfrac{n}{2}\right)}\ x^{n-1}e^{-x^2/2}\ , & x>0\ ,\\ 0\ , & x\leqslant 0\ ,\end{cases} \qquad x\in R^1.$$

***证** 仿例 4.4.21 当 $x\leqslant 0$ 时，$F_{\chi_n}(x)=0$，当 $x>0$ 时，

$$F_{\chi_n}(x)=\int\cdots\int_{\sum_{l=1}^n x_l^2\leqslant x^2}\frac{1}{(2\pi)^{n/2}}e^{-\sum_{l=1}^n x_l^2/2}dx_1\cdots dx_n .$$

作坐标变换：

$$\begin{aligned}x_1&=\rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\cos\theta_{n-1}\ ,\\ x_2&=\rho\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\sin\theta_{n-1}\ ,\\ &\cdots\cdots\\ x_{n-1}&=\rho\cos\theta_1\sin\theta_2\ ,\\ x_n&=\rho\sin\theta_1\ ,\end{aligned}$$

$$0\leqslant\rho<\infty\ ,\ -\frac{\pi}{2}\leqslant\theta_1,\cdots,\theta_{n-2}\leqslant\frac{\pi}{2}\ ,\ -\pi\leqslant\theta_{n-1}\leqslant\pi\ .$$

按照例4.4.21，应该先算出上述坐标变换的雅可比行列式．但是也可以按下述办法利用 $F_{\chi_n}(+\infty)=1$ 不算出雅可比行列式亦可求出 χ_n 的密度．由上述变换可见

$$\frac{1}{\rho^{n-1}}\begin{vmatrix} \frac{\partial x_1}{\partial \rho} & \frac{\partial x_1}{\partial \theta_1} & \cdots & \frac{\partial x_1}{\partial \theta_{n-1}} \\ & \cdots\cdots\cdots & & \\ \frac{\partial x_n}{\partial \rho} & \frac{\partial x_n}{\partial \theta_1} & \cdots & \frac{\partial x_n}{\partial \theta_{11-1}} \end{vmatrix}$$

与 ρ 无关，因而可将它的绝对值记成 $D(\theta_1, \cdots, \theta_{n-1})$，则

$$F_{\chi_n}(x)=\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta_1 \cdots \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} d\theta_{n-2} \int_{-\pi}^{\pi} d\theta_{n-1} D(\theta_1, \cdots, \theta_{n-1})$$

$$\cdot \int_0^x \left(\frac{1}{2\pi}\right)^{n/2} \cdot e^{-\frac{\rho^2}{2}} \rho^{n-1} d\rho$$

$$\triangleq C_n \int_0^x \rho^{n-1} e^{-\frac{\rho^2}{2}} d\rho$$

为求 C_n，令 $x \to \infty$，然后作变换 $t=\frac{\rho^2}{2}$（因而 $\rho=\sqrt{2t}$，$d\rho=\frac{dt}{\sqrt{2t}}$），即得

$$1=F_{\chi_n}(+\infty)=C_n \int_0^{\infty} \rho^{n-1} e^{-\rho^2/2} d\rho$$

$$=C_n 2^{\frac{n}{2}-1} \int_0^{\infty} e^{-t} \cdot t^{\frac{n}{2}-1} dt = C_n 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right),$$

其中 $\Gamma\left(\frac{n}{2}\right)$ 为 $\Gamma-$ 函数 $\Gamma(p)=\int_0^{\infty} t^{p-1}e^{-t}dt$ 在 $p=\frac{n}{2}$ 处的值.

于是得 $C_n=\left[2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)\right]^{-1}$，而

$$f_{\chi_n}(x)=\begin{cases}0, & x\leqslant 0,\\ \dfrac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}x^{n-1}e^{-x^2/2}, & x>0.\end{cases}$$

例 4.4.24 若 $\xi_1,\cdots,\xi_n$ 是独立的标准正态随机变量，则

$$\chi_n^2=\xi_1^2+\cdots+\xi_n^2$$

有分布密度

$$(4.4.25)\quad f_{\chi_n^2}(x)=\begin{cases}0, & x\leqslant 0,\\ \dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0,\end{cases}$$

即 $\Gamma\left(\frac{n}{2},\frac{1}{2}\right)$ 分布 (定义 3.3.22).

证 当 $x<0$ 时 $F_{\chi_n^2}(x)=P(\chi_n^2\leqslant x)=0$. 当 $x>0$ 时，由例 4.4.22 并作变量替换 $t^2=u$，即知

$$F_{\chi_n^2}(x)=P(\chi_n^2\leqslant x)=P(\chi_n\leqslant\sqrt{x})=\int_0^{\sqrt{x}}\left[2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)\right]^{-1}t^{n-1}e^{-t^2/2}dt$$

$$=\int_0^x\left[2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\right]^{-1}u^{\frac{n}{2}-1}e^{-\frac{u}{2}}du,$$

故得 (4.4.25).

注：(4.4.25)也可仿照例 4.4.20 的办法用归纳法求出.

定义 4.4.26 以(4.4.25)为密度的概率分布称为自由度为 n 的 χ^2 分布，简记作 $\chi^2(n)$分布. 具有分布密度 (4.4.25)的随机变量 ξ 称为自由度为 n 的 χ^2 变量，记作 $\xi \sim \chi^2(n)$. 有时直接用 $\chi^2(n)$表示自由度为 n 的 χ^2 变量. 为了应用，书末的附表 3 给出了满足 $P(\chi^2(n) > \chi_\alpha^2) = \alpha$ 的 χ_α^2 值. 其图形见附表 3 的表头.

χ^2 分布是数理统计中常用的分布.

例 4.4.27 若 $\chi_{n_1}^2, \cdots, \chi_{n_k}^2$ 是独立的随机变量，$\chi_{n_l}^2 \sim \chi^2(n_l)$，$l = 1, \cdots, k$. 则 $\chi_{n_1}^2 + \cdots + \chi_{n_k}^2 \sim \chi^2(n_1 + \cdots + n_k)$.

证 由例 4.4.24 及定义 4.4.26 知自由度为 n_l 的 χ^2 分布即 $\Gamma\left(\frac{n_l}{2}, \frac{1}{2}\right)$分布. 现在用归纳法来证明上述结论.

当 $k=1$ 时自然成立. 假设对 l 结论成立，由定理 4.1.14，4.1.16 及 4.1.17 知 $\chi_{n_1}^2 + \cdots + \chi_{n_l}^2$，$\chi_{n_{l+1}}^2$ 独立，且分别服从自由度为 $n_1 + \cdots + n_l$，n_{l+1} 的 χ^2 分布，亦即分别服从 $\Gamma\left(\frac{1}{2}(n_1 + \cdots + n_l), \frac{1}{2}\right)$，$\Gamma\left(\frac{n_{l+1}}{2}, \frac{1}{2}\right)$ 分布，由例 4.4.9 知 $\sum_{i=1}^{l+1} \chi_{n_i}^2 \sim \Gamma\left(\frac{1}{2}\sum_{i=1}^{l+1} n_i, \frac{1}{2}\right)$即服从自由度为 $n_1 + \cdots + n_l + n_{l+1}$ 的 χ^2 分布.

例 4.4.28 设随机变量 ξ，η 独立，$\xi \sim N\left(0, \frac{1}{n}\right)$ 且 $\eta = \chi_n/\sqrt{n}$，χ_n 由例 4.4.22 定义，则 $\zeta = \xi/\eta$ 有分布密度

$$(4.4.29)\quad f_\zeta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

证 由 (4.4.23)知 η 有分布密度

$$f_\eta(x)=\begin{cases}0, & x\leqslant 0,\\ \dfrac{\sqrt{2n}}{\Gamma\left(\dfrac{n}{2}\right)}\left(\dfrac{x\sqrt{n}}{\sqrt{2}}\right)^{n-1}e^{-\frac{nx^2}{2}}, & x>0.\end{cases}$$

于是由定理 4.4.11 的(4.4.13)知 ζ 有分布密度

$$f_\zeta(x)=\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2x^2}{2}}\frac{\sqrt{2n}}{\Gamma\left(\dfrac{n}{2}\right)}\left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1}e^{-\frac{ny^2}{2}}y\,dy$$

$$=\frac{1}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\int_0^\infty\left(\frac{y\sqrt{n}}{\sqrt{2}}\right)^{n-1}e^{-\frac{ny^2}{2}\left(1+\frac{x^2}{n}\right)}ny\,dy.$$

作积分变换 $u=\dfrac{ny^2}{2}\left(1+\dfrac{x^2}{n}\right)$，即得

$$f_\zeta(x)=\frac{\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\int_0^\infty u^{\frac{n-1}{2}}e^{-u}du$$

$$=\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

定义 4.4.30 具有分布密度 (4.4.29)的概率分布称为自由度为 n 的 t 分布，简记作 $t(n)$ 分布；具有分布密度 (4.4.29) 的随机变量 ξ 称为服从 $t(n)-$ 分布的随机变量或自由度为 n 的 $t-$ 变量，记作 $\xi \sim t(n)$，有时直接用 $t(n)$表示自由度为 n 的 t 变量.

由 (4.4.29)知 $t(n)$ 分布的分布密度 f 的曲线关于 y 轴对称，且当 $x \to \pm\infty$ 时，它的值趋于零. 于是对任意的实数 a 有

$$P(t(n)>a)=P(t(n)<-a).$$

$t(n)$ 分布在数理统计中常常用到，为了应用，书末附表 4 给出了满足

$$P(|t(n)|>t_\alpha)=\alpha$$

的 t_α 值. 其图形见附表 4 的表头.

例 4.4.31 (i)求 a 使 $P(|t(7)|>a)=0.05$；

(ii) 求 b 使 $P(t(10)>b)=0.05$.

解 (i) 查附表 4 中 $n=7$ 那一行，对着 0.05 的数 2.37 就是所求的 a.

(ii) 因为

$$\begin{aligned}0.10&=2P(t(10)>b)=P(t(10)>b)+P(t(10)<-b)\\&=P(|t(10)|>b),\end{aligned}$$

所以查附表 4 中 $n=10$ 那一行，对着 0.10 的数 1.81 就是所求的 b.

例 4.4.32 设随机变量 ξ，η 独立，且 $\xi \sim \chi^2(m)$，$\eta \sim \chi^2(n)$，则 $\zeta=(\xi/m)/(\eta/n)$有分布密度

(4.4.33)

$$f_\zeta(x)=\begin{cases}0, & x\leqslant 0,\\ \dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}(mx+n)^{-\frac{m+n}{2}}, & x>0.\end{cases}$$

证 先求$\overline{\zeta}=\xi/\eta$的分布密度，再求$\zeta=\frac{n}{m}\overline{\zeta}$的分布密度.

由于ξ，η独立，应用定理4.4.11中的(4.4.13)及(4.4.25)知

$$f_{\overline{\zeta}}(z)=\int_{-\infty}^{\infty}f_{\chi^2(m)}(yz)f_{\chi^2(n)}(y)\,|y|\,dy$$

$$=\int_0^{\infty}yf_{\chi^2(m)}(yz)f_{\chi^2(n)}(y)\,dy.$$

当$z\leqslant 0$时在$y\in(0,+\infty)$上$f_{\chi^2(m)}(yz)=0$，所以$f_{\overline{\zeta}}(z)=0$；当$z>0$时

$$f_{\overline{\zeta}}(z)=\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$\cdot\int_0^{\infty}y\,e^{-\frac{1}{2}yz}(yz)^{\frac{m}{2}-1}e^{-\frac{y}{2}}y^{\frac{n}{2}-1}dy$$

$$=\frac{z^{\frac{m}{2}-1}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty}y^{\frac{m+n}{2}-1}e^{-\frac{1}{2}(z+1)y}\,dy.$$

作变量替换，$t=\frac{1}{2}(z+1)y$，则$y=\frac{2}{z+1}t$，

$$f_{\overline{\zeta}}(z)=\frac{z^{\frac{m}{2}-1}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

$$\cdot\int_0^{\infty}\left(\frac{2}{z+1}\right)^{\frac{m+n}{2}}t^{\frac{m+n}{2}-1}e^{-t}dt$$

$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} z^{\frac{m}{2}-1}(z+1)^{-\frac{m+n}{2}} .$$

应用例 3.6.1 知

$$f_\zeta(x) = f_\xi\left(\frac{mx}{n}\right)\cdot\frac{m}{n} ,$$

由此即得 (4.4.33).

定义 4.4.34 称具有分布密度 (4.4.33)的概率分布是第一、二自由度分别为 m, n 的 F 分布，简记作 $F(m,n)$分布．若随机变量 ξ 具有分布密度 (4.4.33)，则称 ξ 服从 $F(m,n)$分布，记作 $\xi \sim F(m,n)$. 此时也称 ξ 是第一、二自由度分别是 m, n 的 F 变量，有时迳直用 $F(m,n)$表示．

$F(m,n)$分布在数理统计中经常用到．为了应用，书末附表 5 给出了满足

$$P(F(m,n) > F_\alpha(m,n)) = \alpha$$

的 $F_\alpha(m,n)$值．其中 $\alpha=0.25$, $\alpha=0.05$, $\alpha=0.01$. 例如，$F_{0.25}(12,4)=2.08$, $F_{0.05}(7,20)=2.51$, $F_{0.01}(4,3)=28.71$. F 分布的密度函数图形见附表 5 的表头．

在应用时，有时需要数值 $F_{0.75}(m,n)$, $F_{0.95}(m,n)$, $F_{0.99}(m,n)$，但是这些随在附表 5 中是查不到的．应用下面的命题及附表 5 可以求出这些值．

命题 4.4.35 若随机变量 $\xi_1 \sim F(m,n)$, $\xi_2 \sim F(n,m)$，则对任何正实数 x 有

(4.4.36) $$P(\xi_1 \leqslant x) = P\left(\xi_2 \geqslant \frac{1}{x}\right) .$$

证 设随机变量 $\eta_1 \sim \chi^2(m)$, $\eta_2 \sim \chi^2(n)$，则

$$P(\xi_1 \leqslant x) = P\left(\frac{n\eta_1}{m\eta_2} \leqslant x\right)$$

$$= P\left(\frac{m\eta_2}{n\eta_1} \geqslant \frac{1}{x}\right) = P\left(\xi_2 \geqslant \frac{1}{x}\right).$$

例 4.4.37 对任何正整数 m，n，实数 $\alpha \in (0,1)$ 有

(4.4.38) $$F_{1-\alpha}(m,n) = [F_\alpha(n,m)]^{-1}.$$

证 在 (4.4.36) 中，令 $x = [F_\alpha(n,m)]^{-1}$，则有

$$P(F(m,n) \leqslant [F_\alpha(n,m)]^{-1})$$

$$= P(F(n,m) \geqslant F_\alpha(n,m)) = \alpha.$$

由此即得 (4.4.38)。

由 (4.4.38) 即知

$$F_{0.75}(m,n) = \frac{1}{F_{0.25}(n,m)},$$

$$F_{0.95}(m,n) = \frac{1}{F_{0.05}(n,m)},$$

$$F_{0.99}(m,n) = \frac{1}{F_{0.01}(n,m)}.$$

习题 4－4

1. 设随机变量 ξ 与 η 独立，分布列分别为

ξ	0	1	2
p	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

η	0	1
p	$\frac{1}{3}$	$\frac{2}{3}$

求随机变量 ζ 的分布列，其中 (1) $\zeta = \xi + \eta$；(2) $\zeta = \xi\eta$；(3) $\zeta = \xi^2 + \eta^2$。

2. 掷一枚均匀硬币 3 次，设 ξ，η 分别为正、反面出现的次数。求下列随机变量的分布列：(1) $\zeta = \xi + \eta$；(2) $\zeta = \xi - \eta$；(3) $\zeta = \xi\eta$。

3. 设随机变量 $\xi_i \sim B(n_i, p)$，$i = 1, 2$，且独立，求 $\zeta = \xi + \eta$ 的分布列。

4.设随机变量 ξ，η 独立，ξ 服从 $[0,1]$ 上的均匀分布，η 服从参数为 1 的指数分布．求 $\xi+\eta$ 及 ξ/η 的概率密度．

5.设随机向量 (ξ,η) 的分布密度为

$$f(x,y)=\begin{cases}120y(x-y)(1-x), & 0\leqslant y\leqslant x\leqslant 1,\\ 0, & \text{其它},\end{cases}$$

求 $\eta/\sqrt{\xi}$ 的分布密度．

6.设随机变量 ξ_i，$i=1,2,3,4,5$ 为独立的且都服从参数为 λ 的指数分布．对于 $a>0$，求 (1) $P(\min(\xi_1,\cdots,\xi_5)\leqslant a)$；(2) $P(\max(\xi_1,\cdots,\xi_5)\leqslant a)$.

§4.5 多维随机变量函数的数字特征

在 §3.7 中讲述随机变量的期望和方差的性质时，提到它们大都涉及几个随机变量，需要用到关于多维随机变量的知识，并且遗留下一些性质在第四章证明．在这一节里我们首先解决这些问题，然后进一步讨论涉及到多个随机变量的数字特征，如随机变量的协方差、相关系数等等．不过，和以前一样，我们也只讨论离散型和连续型两种随机变量的情形．一般情形的讨论超出了本书的范围，有兴趣者可参看 [15] 第三章 §4，§5.

现在证明命题 3.7.3 及命题 3.7.4.

定理 4.5.1 若 ξ_1，ξ_2 为同一概率空间上的两个随机变量，且 $E\xi_i$，$i=1,2$ 存在，则 (i) $E(\xi_1+\xi_2)$ 存在且

$$E(\xi_1+\xi_2)=E\xi_1+E\xi_2;$$

(ii) 对任意实数 α_1，α_2，有

$$E(\alpha_1\xi_1+\alpha_2\xi_2)=\alpha_1E\xi_1+\alpha_2E\xi_2.$$

证 若 (i) 成立，则由命题 3.7.2 知 $E(\alpha_i\xi_i)=\alpha_iE\xi_i$，$i=1,2$ 存在，因而 (ii) 成立．现证 (i).

若 (ξ_1,ξ_2) 为离散型随机变量，其分布列为

$$p_{ij}=P(\xi_1=x_i,\ \xi_2=y_j),\ i,j=1,2,\cdots.$$

则由定理 4.4.2 知 $\zeta=\xi_1+\xi_2$ 有分布列

$$P(\zeta=z_k)=\sum_i P(\xi_1=x_i,\ \xi_2=z_k-x_i),$$

其中 $\{z_k;\ k=1,2,\cdots\}$为 x_i+y_j，$i,j=1,2,\cdots$ 中一切两两不同的数组成的集．于是

$$E|\zeta|=\sum_k |z_k|P(\zeta=z_k)=\sum_k\sum_i |z_k|P(\xi_1=x_i,\xi_2=z_k-x_i)$$

$$=\sum_i\sum_j |x_i+y_j|P(\xi_1=x_i,\xi_2=y_j)$$

$$\leqslant\sum_i |x_i|\sum_j P(\xi_1=x_i,\xi_2=y_j)+\sum_j |y_j|\sum_i P(\xi_1=x_i,\xi_2=y_j)$$

$$=\sum_i |x_i|P(\xi_1=x_i)+\sum_j |y_j|P(\xi_2=y_j)=E|\xi_1|+E|\xi_2|<\infty.$$

故 $E(\xi_1+\xi_2)$存在，采用类似的推导和变推方法即得

$$E(\xi_1+\xi_2)=E\zeta=\sum_k z_kP(\xi_1=x_i,\xi_2=z_k-x_i)$$

$$=\sum_i\sum_j (x_i+y_j)P(\xi_1=x_i,\xi_2=y_j)$$

$$=\sum_i x_i\sum_j P(\xi_1=x_i,\xi_2=y_j)+\sum_j y_j\sum_i P(\xi_1=x_i,\xi_2=y_j)$$

$$=\sum_i x_iP(\xi_1=x_i)+\sum y_jP(\xi_2=y_j)=E\xi_1+E\xi_2.$$

若 (ξ_1,ξ_2)有分布密度 $f(x,y)$，则由定理 4.4.2 知 $\zeta=\xi_1+\xi_2$ 有分布密度 $\int_{-\infty}^{\infty} f(x,z-x)dx$，于是

$$E|\zeta|=\int_{-\infty}^{\infty}\left(|z|\int_{-\infty}^{\infty} f(x,z-x)dx\right)dz.$$

作变量替推 $y=z-x$，$x=x$，然后应用定理 4.3.14 即得

$$E|\zeta|=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x+y|f(x,y)dxdy$$

$$\leqslant \int_{-\infty}^{\infty}\left(|x|\int_{-\infty}^{\infty} f(x,y)dy\right)dx$$

$$+\int_{-\infty}^{\infty}\left(|y|\int_{-\infty}^{\infty} f(x,y)dx\right)dy$$

$$=\int_{-\infty}^{\infty}|x|f_{\xi_1}(x)dx+\int_{-\infty}^{\infty}|y|f_{\xi_2}(y)dy=E|\xi_1|+E|\xi_2|<\infty,$$

故 $E(\xi_1+\xi_2)$存在．再采用类似的积分换顺序及变量替换，即得

$$E(\xi_1+\xi_2)=E\zeta=\int_{-\infty}^{\infty}\left(z\int_{-\infty}^{\infty} f(x,z-x)dx\right)dz$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy=E\xi_1+E\xi_2.$$

推论 4.5.2 若$\xi_1,\xi_2,\cdots\xi_n$为同一概率空间上的随机变量，$E\xi_k$，$k=1,2,\cdots,n$存在，$\alpha_1,\cdots,\alpha_n$为任意实数，则 $E(\alpha_1\xi_1+\cdots+\alpha_n\xi_n)$存在且

$$E(\alpha_1\xi_1+\cdots+\alpha_n\xi_n)=\alpha_1E\xi_1+\cdots+\alpha_nE\xi_n.$$

应用归纳法及定理 4.5.1 即可证得此结果．

定理 4.5.3 若$\xi_1,\xi_2,\cdots,\xi_n$为独立的随机变量，且 $E\xi_k$ 存在，$k=1,2,\cdots,n$则 $E(\xi_1\cdots\xi_n)$存在，且

$$E(\xi_1\cdots\xi_n)=E\xi_1\cdots\cdots E\xi_n.$$

证 由于离散型情形较易，而且方法类似，我们只对连续型情形证明．先证二维情形．

由定理 4.3.18 (1)知 (ξ_1,ξ_2)有分布密度$f_{\xi_1}(x_1)f_{\xi_2}(x_2)$，再由定理 4.4.14 知$\zeta=\xi_1\xi_2$有分布密度 $\int_{-\infty}^{\infty} f_{\xi_1}\left(\frac{x}{y}\right)f_{\xi_2}(y)\frac{1}{|y|}dy$.

于是

$$E|\zeta|=\int_{-\infty}^{\infty}\left(|z|\int_{-\infty}^{\infty} f_{\xi_1}\left(\frac{z}{y}\right)f_{\xi_2}(y)\frac{1}{|y|}dy\right)dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\frac{z}{y}\right| f_{\xi_1}\left(\frac{z}{y}\right) f_{\xi_2}(y)\,dydz ,$$

作变量替换 $z = xy$，$y = y$，则雅可比行列式的绝对值为 $|y|$，

$$E|\zeta| = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |xy| f_{\xi_1}(x) f_{\xi_2}(y)\,dxdy$$

$$= \int_{-\infty}^{\infty} |x| f_{\xi_1}(x)\,dx \int_{-\infty}^{\infty} |y| f_{\xi_2}(y)\,dy = E|\xi_1| \cdot E|\xi_2| < \infty ,$$

故 $E(\xi_1\xi_2)$存在，且用相同的方法可得

$$E(\xi_1\xi_2) = E\zeta = \int_{-\infty}^{\infty} \frac{z}{|y|} f_{\xi_1}\left(\frac{z}{y}\right) f_{\xi_2}(y)\,dydz$$

$$= \int_{-\infty}^{\infty} x f_{\xi_1}(x)\,dx \int_{-\infty}^{\infty} y f_{\xi_2}(y)\,dy = E\xi_1 \cdot E\xi_2 .$$

下面用归纳法证明一般情形：首先定理对 $n=2$ 成立，设定理对 $n=m-1$ 成立，证定理对 $n=m$ 成立．由定理 4.1.16 及 4.1.17 知 $\xi_1\cdots\xi_{m-1}$ 与 ξ_m 独立，再由定理 4.1.14 知 ξ_1，…，ξ_{m-1} 独立，因而 $E(\xi_1\cdots\xi_{m-1})$存在且 $E(\xi_1\cdots\xi_{m-1}) = E\xi_1\cdots E\xi_{m-1}$．于是由第一部分的证明 $E[(\xi_1\cdots\xi_{m-1})\xi_m]$ 存在且

$$E[(\xi_1\cdots\xi_{m-1})\xi_m] = E(\xi_1\cdots\xi_{m-1}) \cdot E\xi_m = E\xi_1\cdots E\xi_{m-1}E\xi_m .$$

定理 4.5.4 若 ξ_1，…，ξ_n 两两独立且 $D\xi_k$ 存在，$k=1$，2，…，n，则

$$D(\xi_1 + \cdots + \xi_n) = D\xi_1 + \cdots + D\xi_n .$$

证 由于方法的类似，我们只讨论连续型随机变量的情形．首先注意 $D\xi$ 存在当且仅当 $E\xi$ 及 $E\xi^2$ 存在．事实上，由定义 3.4.10 知若 $D\xi$ 存在，则 $E\xi$ 存在且

$$\int_{-\infty}^{\infty}(x-E\xi)^2 f_\xi(x)<\infty,$$

由此即得

$$E\xi^2=\int_{-\infty}^{\infty}x^2 f_\xi(x)dx<\infty.$$

反之，若 $E\xi$，$E\xi^2$ 存在，则

$$\int_{-\infty}^{\infty}(x-E\xi)^2 f_\xi(x)dx=\int_{-\infty}^{\infty}x^2 f_\xi(x)dx-(E\xi)^2<\infty,$$

因此 $D\xi$ 存在．

再证 $D(\xi_1+\cdots+\xi_n)$存在．因为 $D\xi_k$ 存在，$k=1,2,\cdots,n$，由上一段知 $E\xi_k$，$E\xi_k^2$，$k=1,2,\cdots,n$ 存在．于是由推论 4.5.2 知 $E(\xi_1+\cdots+\xi_k)$存在，由定理 4.5.3 及 ξ_k，ξ_l $(k\neq l)$独立知 $E(\xi_k\xi_l)$，$k,l=1,\cdots,n$ 存在，因而再由推论 4.5.2 知

$$E\left[\sum_{k=1}^{n}\sum_{l=1}^{n}\xi_k\xi_l\right]=E[(\xi_1+\cdots+\xi_n)^2] \tag{4.5.5}$$

存在，所以 $D(\xi_1+\cdots+\xi_n)$存在．

由于

$$D\xi=\int_{-\infty}^{\infty}(x-E\xi)^2 f_\xi(x)dx=\int_{-\infty}^{\infty}x^2 f_\xi(x)dx-(E\xi)^2$$

$$=E(\xi^2)-(E\xi)^2$$

及推论 4.5.2，定理 4.5.3，

$$D(\xi_1+\cdots+\xi_n)=E[(\xi_1+\cdots+\xi_n)^2]-[E(\xi_1+\cdots+\xi_n)]^2$$

$$=\sum_{k=1}^{n}\sum_{l=1}^{n}E(\xi_k\xi_l)-\sum_{k=1}^{n}\sum_{l=1}^{n}E\xi_kE\xi_l$$

$$=\sum_{k=1}^{n}[E(\xi_k^2)-(E\xi_k)^2]+\sum_{\substack{k=1\\k\neq l}}^{n}\sum_{l=1}^{n}[E(\xi_k\xi_l)-E\xi_k\cdot E\xi_l]$$

$$=\sum_{k=1}^{n}[E(\xi_k^2)-(E\xi_k)^2]=\sum_{k=1}^{n}D\xi_k.$$

下面我们讨论的内容涉及到多个随机变量的数字特征．为此我们先回忆：一维随机变量的期望和方差来源于刻画它取值的大小状况及分散程度．对于多维随机变量，除了每个分量取值的大小及分散程度以外，很自然地想到是考查它们所取的值之间是否存在什么"函数"关系，最简单的是：是否存在某种近似的线性关系？如果有，它们的相关程度如何？这些问题的研究，导出了相关系数和协方差的概念．下面便来进行具体的研究，为了使讨论简单起见，我们只讨论二个随机变量的情形．

设 (ξ,η) 是二维随机变量，$D\xi>0$，$D\eta>0$．一般来说，由于取值的随机性，ξ，η 之间不一定有线性函数关系．例如 ξ，η 表示一个 15 岁的男学生的身高与体重．一个男孩在发育到 15 岁时要受到许多不确定因素的影响，所以 15 岁时的身高与体重是不能准确预言的，都是随机变量，而且身高与体重之间很难说有确定的函数关系，当然更谈不到存在线性函数关系了．但是也有身材高的体重也大，而且一般有随着身材增高体重有一种"线性"增加的趋势．表 4－4 列出了任意挑选的 10 个 15 岁的男学生的身高（单位：厘米）与体重（单位：公斤）．

表 4－4

x（身高）	157	167	165	158	155	156	164	160	158	163
y（体重）	46	55	52	46	42	45	49	47	44	49

把身高 x 作为横坐标，体重 y 作为纵坐标，在坐标纸上画出 10 个学生（身高，体重）对应的点（见图 4－15），就会看到随着身高增长体重呈直线增加的大致趋势．这种关系我们认为是 ξ，η 间的一种线性相关关系．为此我们设法找一条能反映 ξ，η 之间

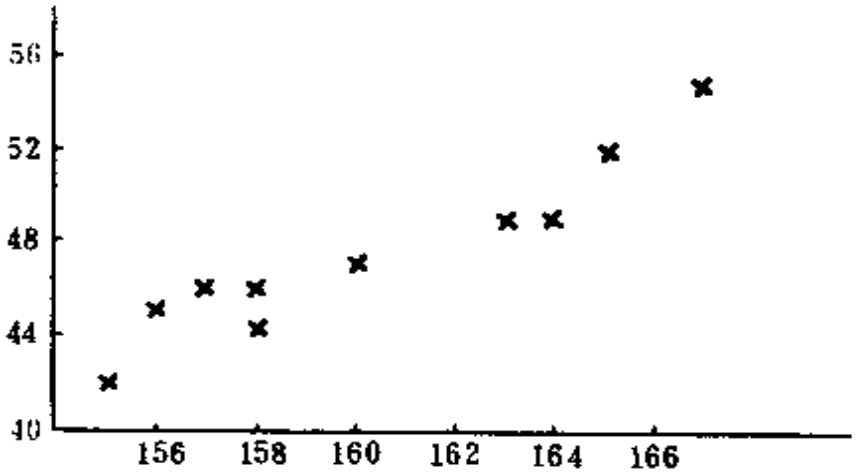


图 4-15

的线性相关关系的直线．一个自然的想法是要找的直线 $y=\alpha+\beta x$ ($x=\gamma+\delta y$)能使得 η 与 $\alpha+\beta\xi$ (ξ 与 $\gamma+\delta\eta$)尽量地"接近"．按照以前考虑取值的偏离期望的程度的想法，这种接近的程度用 $E[\eta-(\alpha+\beta\xi)]^2$($E[\xi-(\gamma+\delta\eta)]^2$)表示．于是我们要找的直线就是

定义 4.5.6 设 (ξ,η)为一二维随机变量，$D\xi>0$，$D\eta>0$，如果 α，β 满足条件

$$(4.5.7)\qquad E[\eta-(\alpha+\beta\xi)]^2=\min_{a,b}E[\eta-(a+b\xi)]^2\ ,$$

则称直线 $y=\alpha+\beta x$ 为 η 关于 ξ 的**回归直线**；若 γ，δ 满足

$$(4.5.7)'\qquad E[\xi-(\gamma+\delta\eta)]^2=\min_{c,d}E[\xi-(c+d\eta)]^2\ ,$$

则称 $x=\gamma+\delta y$ 为 ξ 关于 η 的回归直线．我们用回归直线来反映 ξ，η 间的线性相关关系．

由于 ξ，η 的对称性，我们只需讨论 η 关于 ξ 的回归直线．

定理 4.5.8 设 (ξ,η)为一二线随机变量，$D\xi>0$，$D\eta>0$．则

$$(4.5.8)\qquad \beta\triangleq\frac{E(\xi-E\xi)(\eta-E\eta)}{D\xi}\ ,\qquad \alpha\triangleq E\eta-\beta E\xi$$

满足条件 (4.5.7)，而且

(4.5.9)

$$E[\eta-(\alpha+\beta\xi)]^2=D\eta\left[1-\left(\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi D\eta}}\right)^2\right].$$

将 ξ，η 在上述公式中的位置对调即得 ξ 关于 η 的回归直线的有关结论.

证 将 (4.5.7)右端的数学期望展开 (记 $\eta_0\hat{=}\eta-E\eta$，$\xi_0\hat{=}\xi-E\xi$)：

$$\begin{aligned}&E[\eta-(a+b\xi)]^2\\&=E[\eta_0-b\xi_0+(E\eta-a-bE\xi)]^2\\&=b^2E(\xi_0^2)-2bE(\eta_0\xi_0)+E(\eta_0^2)+(E\eta-a-bE\xi)^2\\&\quad+2E(\eta_0-b\xi_0)(E\eta-a-bE\xi).\end{aligned}$$

由于 $E(\eta_0-b\xi_0)=E(\eta-E\eta)-bE(\xi-E\xi)=0$，$E(\xi_0^2)=E(\xi-E\xi)^2=D\xi$，$E\eta_0^2=D\eta$，

所以

$$\begin{aligned}&E[\eta-(a+b\xi)]^2\\&=b^2D\xi-2bE(\eta_0\xi_0)+D\eta+(E\eta-a-bE\xi)^2\\&=\left(b\sqrt{D\xi}-\frac{E(\eta_0\xi_0)}{\sqrt{D\xi}}\right)^2+(E\eta-a-bE\xi)^2\\&\quad+D\eta\left[1-\frac{[E(\xi_0\eta_0)]^2}{D\xi D\eta}\right].\end{aligned}$$

上式右端的第三项与 a，b 无关，所以要使 $E[\eta-(a+b\xi)]^2$ 达到最小，只需上式右端前两项等于零即可. 故 β，α 由(4.5.8)确定，而且此时

$$\begin{aligned}E[\eta-(\alpha+\beta\xi)]^2&=D\eta\left[1-\frac{(E(\xi_0\eta_0))^2}{D\xi D\eta}\right]\\&=D\eta\left[1-\left(\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi D\eta}}\right)^2\right].\end{aligned}$$

由 ξ，η 的对称性，即得定理的另一结论．

对于任何二维随机变量 (ξ,η)，只要 $D\xi>0$，都可以用 (4.5.8)的 α，β 确定 η 对 ξ 的回归直线，来反映 ξ，η 间的线性相关关系．这时可以根据 $E[\eta-(\alpha+\beta\xi)]^2$ 的大小来刻划 η 与 ξ 的线性相关程度．但是如果考虑 ξ 对 η 的回归直线，则又应该用 $E[\xi-(\gamma+\delta\eta)]^2=D\xi\left[1-\left(\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi D\eta}}\right)^2\right]$ 来刻划 η 与 ξ 的线性相关程度．这样用两个数字来刻划同一件事情是不合适的．但是又可以由 (4.5.9)及上式看出它们的大小都决定于

$$\rho_{\xi,\eta}\triangleq\frac{E(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi D\eta}}. \tag{4.5.10}$$

于是有

定义 4.5.11 对于二维随机变量 (ξ,η)，设 $D\xi>0$，$D\eta>0$，则由(4.5.10)定义的 $\rho_{\xi,\eta}$ 称为 ξ，η 的**相关系数**，而称

$$\mathrm{Cov}(\xi,\eta)\triangleq E(\xi-E\xi)(\eta-E\eta)$$

为 ξ，η 的**协方差**，称

$$\beta\triangleq\frac{E(\xi-E\xi)(\eta-E\eta)}{D\xi}=\sqrt{\frac{D\eta}{D\xi}}\,\rho_{\xi,\eta}\left(\delta\triangleq\sqrt{\frac{D\xi}{D\eta}}\,\rho_{\xi,\eta}\right)$$

为 η 对 ξ(ξ 对 η)的**回归系数**．

对于二维随机变量 (ξ,η)，设 $D\xi>0$，$D\eta>0$，则由(4.5.9)及(4.5.10)知 $|\rho_{\xi,\eta}|\leqslant 1$，即

$$-1\leqslant\rho_{\xi,\eta}\leqslant 1, \tag{4.5.12}$$

而且当 $|\rho_{\xi,\eta}|$ 越大时，$E[\eta-(\alpha+\beta\xi)]^2$ 及 $E[\xi-(\gamma+\delta\xi)]^2$ 越小，因而 ξ，η 的线性相关程度越好．在极端的情形 $|\rho_{\xi,\eta}|=1$ 时，由命题 3.7.7 知 $P(\eta=\alpha+\beta\xi)=1$，而此时，

$$\beta=\begin{cases}\sqrt{D\eta/D\xi}, & \rho_{\xi,\eta}=1,\\ -\sqrt{D\eta/D\xi}, & \rho_{\xi,\eta}=-1.\end{cases}$$

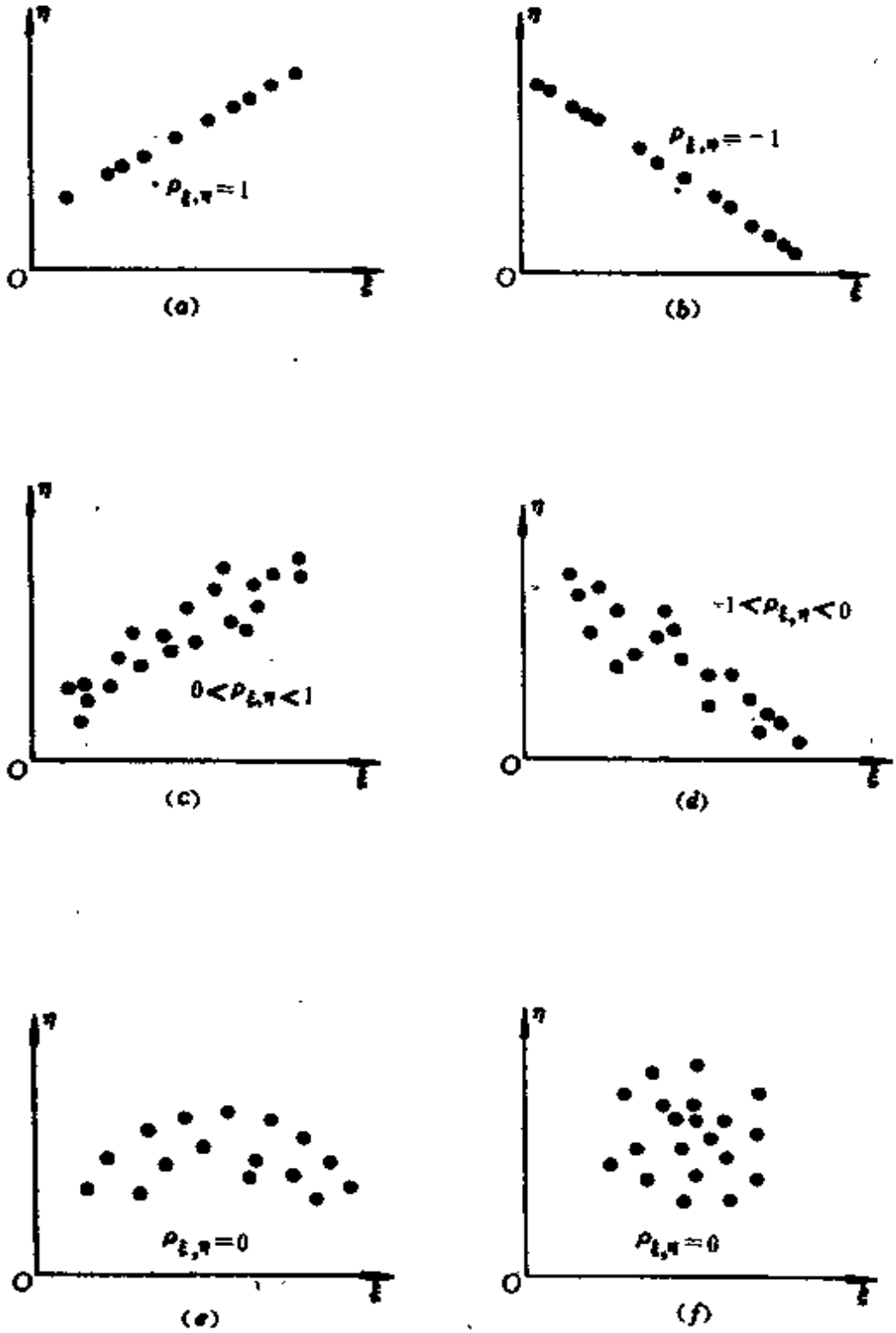

η
O
ξ
$\rho_{\xi,\eta}=1$
(a)
$\rho_{\xi,\eta}=-1$
(b)
$0<\rho_{\xi,\eta}<1$
(c)
$-1<\rho_{\xi,\eta}<0$
(d)
$\rho_{\xi,\eta}=0$
(e)
$\rho_{\xi,\eta}=0$
(f)

图 4－16

故由 (4.5.8)知：当 $\rho_{\xi,\eta}=1$ 时，$P\left(\frac{\eta-E\eta}{\sqrt{D\eta}}=\frac{\xi-E\eta}{\sqrt{D\xi}}\right)=1$，

当 $\rho_{\xi,\eta}=-1$ 时，$P\left(\frac{\eta-E\eta}{\sqrt{D\eta}}=-\frac{\xi-E\xi}{\sqrt{D\xi}}\right)=1$. 另一极端情形 $\rho_{\xi,\eta}=0$ 时，$\beta=\delta=0$，$\alpha=E\eta$，$\gamma=E\xi$，则 η，ξ 没有线性关系. 此外当 $0<\rho_{\xi,\eta}\leqslant 1$ 时，$\beta>0$，$\delta>0$，此时 η 取值有随 ξ 取值的增加而增加的趋势；当 $-1\leqslant\rho_{\xi,\eta}<0$ 时，$\beta<0$，$\delta<0$，则 η 取值有随 ξ 取值的增加而减少的趋势. 上述各种情形下，ξ，η 取值的大致关系如图 4－16 所示.

例 4.5.13 多维正态分布中参数的概率意义. 为简单起见，着重考虑二维情形.

设二维随机变量 $(\xi,\eta)\sim N(\mu_1,\mu_2,;\sigma_1^2,\sigma_2^2;\rho)$，由例4.3.16知 $\xi\sim N(\mu_1,\sigma_1^2)$，$\eta\sim N(\mu_2,\sigma_2^2)$，因此 μ_1，μ_2 分别是 ξ，η 的期望，σ_1^2，σ_2^2 分别是 ξ，η 的方差. 现在我们来证明 $\rho=\rho_{\xi,\eta}$.

* 首先我们注意：若 (ξ,η)具有分布密度 $f(x,y)$，则由定理 4.4.14 的 (4.4.15)知

$$E(\xi\eta)=\int_{-\infty}^{\infty}\left(z\int_{-\infty}^{\infty}f\left(\frac{z}{y},y\right)\frac{1}{|y|}dy\right)dz\ ,$$

作积分变元替换 $z=xy$，$y=y$，注意其雅可比行列式的绝对值为 $|y|$，于是

$$E(\xi\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy\ .$$

其次，用 $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$的分布密度 (4.3.10)代入上式并注意

$$\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}$$

$$= \frac{1}{\sigma_1^2}\left[x-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]^2+(1-\rho^2)\frac{(y-\mu_2)^2}{\sigma_2^2},$$

即得

$$E\,\xi\eta=\frac{1}{\sqrt{2\pi}\,\sigma_2}\int_{-\infty}^{\infty}dy\cdot y\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}$$

$$\cdot\frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}x\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\right.$$

$$\left.\cdot\left[x-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]^2\right\}dx$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_2}\int_{-\infty}^{\infty}y\,(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2))\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}dy$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_2}\left[\mu_1\int_{-\infty}^{\infty}y\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}dy\right.$$

$$+\rho\frac{\sigma_1}{\sigma_2}\int_{-\infty}^{\infty}(y-\mu_2)^2\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}dy$$

$$\left.+\mu_2\rho\frac{\sigma_1}{\sigma_2}\int_{-\infty}^{\infty}(y-\mu_2)\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}dy\right]$$

$$=[\mu_1\mu_2+\rho\sigma_1\sigma_2].$$

故

$$\rho=\frac{1}{\sigma_1\sigma_2}(E\,(\xi\eta)-\mu_1\mu_2)=\frac{E\,(\xi-E\xi)(\eta-E\eta)}{\sqrt{D\xi D\eta}}=\rho_{\xi,\eta}.$$

这样，我们就完全了解了 $N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$中五个参数的概率意义.

注：在这一节中，我们主要是讨论和计算了随机变量的和与积的数学期望. 办法是应用期望的定义及和、积的分布密度的公式写出积分表达式，用积分变元替换得到公式

$$E(\xi+\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy\quad,$$

$$E\xi\eta=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy\,.$$

然后再加以应用或计算. 事实上，对于连续型随机变量 $(\xi_1,\cdots,\xi_n)$的函数 $g(\xi_1,\cdots,\xi_n)$的数学期望有下列更一般的公式：若

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|g(x_1,\cdots,x_n)|f(x_1,\cdots,x_n)dx_1\cdots dx_n<\infty,$$

则 $Eg(\xi_1,\cdots,\xi_n)$存在且

$$Eg(\xi_1,\cdots,\xi_n)$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,\cdots,x_n)f(x_1,\cdots,x_n)dx_1\cdots dx_n\,.$$

不过这个公式的证明超出了本书的范围，这里不再证明.
有兴趣的读者可参看 [15] 的第三章 §5 定理 6.

习题 4－5

1. 设随机变量 ξ 与 η 独立，分布密度分别为

$$f_\xi(x)=\begin{cases}\frac{1}{2}, & 0\leqslant x\leqslant 2,\\ 0, & \text{其它};\end{cases}\qquad f_\eta(y)=\begin{cases}e^{-y}, & y\geqslant 0,\\ 0, & y<0.\end{cases}$$

求 $E(\xi\eta)$及 $D(\xi\eta)$.

2. 设随机变量 ξ 与 η 独立，ξ 服从正态分布 $N(1,4)$，η 服从 $[0,2]$ 上的均匀分布．求 $E(\xi\eta)$，$E(\xi-\eta)^2$ 及 $D(\xi+\eta)$.

3. 设独立随机变量 ξ 与 η 有数学期望 $E\xi=2$，$E\eta=-3$，方差 $D\xi=1$，$D\eta=2$，求 $E[3\xi^2\eta+2\eta^2+1]$.

4. 设二维随机变量 (ξ,η)具有分布密度

$$f(x,y)=\begin{cases}\dfrac{1}{4}\sin x\sin y, & 0\leqslant x,y\leqslant\pi,\\ 0, & \text{其它}.\end{cases}$$

求 ξ 与 η 的数学期望、方差及协方差.

5.设二维随机变量 (ξ,η)的分布密度为 $f(x,y)=\dfrac{1}{\pi(1+x^2+y^2)^2}$，求相关系数.

6.设 (ξ,η)为二维随机变量．已知 $D\xi=1$，$D\eta=4$，$\mathrm{Cov}(\xi,\eta)=1$．求 $\xi_1\triangleq\xi-2\eta$，$\eta_1\triangleq 2\xi-\eta$ 的相关系数.

7.掷一枚均匀骰子两次，设 ξ 表示出现的点数之和，η 表示第一次出现的点数减去第二次出现的点数．求 ξ 与 η 的协方差.

复习题 4－A

1. 设在 $\triangle ABC$ 中，$AB=l$，$BC=k$，$\angle B=90°$，在 $\triangle ABC$ 内任取一点(即 M 服从在 $\triangle ABC$ 中均匀分布)，记 M 到 AB 的距离为 ρ、$\angle MAB=\varphi$，求 (ρ,φ)的联合分布函数.

2.设一包晶体管共有 5 支，其中 2 支是次品．每次从中取出一支进行检查(不放回)，直到所有的次品都被检查出来为止．记 ξ 为直到第一次查出次品所需的检查次数，η 为第一次查出次品后直到第二次查出次品所需的检查次数．求 (ξ,η)的分布列，并求给定 $\eta=1$ 的条件下，ξ 的分布列.

3. 设一批产品中，一、二、三等品分别占 50%，30%，20%．从中抽取 n 件(每次抽取后放回)，设抽到一、二、三等品的件数分别为 ξ，η，ζ．求 (ξ,η,ζ)的联合分布列及各边缘分布列.

4. 考虑独立贝努利试验序列，其中每次试验成功概率为 p．令 ξ 表示第一次成功前的失败次数，η 为头两次成功之间的失败次数．求 (ξ,η)的联合分布列.

5. 设随机变量 ξ，η 独立，且 $P(\xi=1)=P(\eta=1)=p>0$，$P(\xi=0)=P(\eta=0)=1-p>0$，定义随机变量

$$\zeta=\begin{cases}1, & \xi+\eta \text{ 为偶数},\\ 0, & \xi+\eta \text{ 为奇数}.\end{cases}$$

问 p 取什么值时，ξ，ζ 独立?

6. 在{0，1，…，9}中采用 (1)有放回，(2)不放回两种方式任取两数 ξ，η. 求在 $\eta=k$ $(0\leqslant k\leqslant 9)$的条件下 ξ 的分布列 (即求 $P(\xi=i|\eta=k)$，$i=0,\cdots,9$).

7. 甲、乙两人约定中午 12 点 30 分在某地见面. 如果甲来到的时间在 12 点 15 分到 12 点 45 分之间均匀分布；乙独立地到达，且到达的时间在 12 点至 13 点之间均匀分布. 试求：(1)先到的人等待另一人到达的时间不超过 t 的概率；(2)甲先到的概率.

8. 在长度为 L 的线段上随机选取两点 ξ，η 使 ξ 与 η 独立，且 ξ 与 η 分别服从 $\left(0,\frac{L}{2}\right)$，$\left(\frac{L}{2},L\right)$ 上的均匀分布. 求由 0 到 ξ、由 ξ 到 η 及由 η 到 L 的三条线段能构成三角形的三条边的概率.

9. 设二维随机变量 (ξ,η) 的分布密度 $f(x,y)$ 为连续函数，定义

$$F(y|x)\triangleq P(\eta\leqslant y|\xi=x)=\lim_{\Delta x\to 0}P(\eta\leqslant y|x-\Delta x\leqslant\xi\leqslant x),$$

并称之为在 $\xi=x$ 的条件下，η 的**条件分布函数**. 证明：若 $f_\xi(x)>0$，则

$$P(\eta\leqslant y|\xi=x)=\int_{-\infty}^{y}\frac{f(x,u)}{f_\xi(x)}du,\quad y\in R^1.$$

称 $f(y|x)\triangleq f(x,y)/f_\xi(x)$ 为给定 $\xi=x$ 的条件下，η 的条件分布密度函数. 同样，当 $f_\eta(y)>0$ 时，称 $f(x|y)\triangleq f(x,y)/f_\eta(y)$ 为给定 $\eta=y$ 的条件下，ξ 的条件分布密度.

10. 设二维随机变量 (ξ,η) 服从单位圆 $x^2+y^2\leqslant 1$ 内的均匀分布，$f(x,y)$ 是其分布密度，求 $f(y|x)$，$f(x|y)$.

11. 设二维随机变量 (ξ,η) 的分布密度为

$$f(x,y)=\begin{cases}c(x^2-y^2)e^{-x}, & 0\leqslant x<\infty,\ -x\leqslant y\leqslant x,\\ 0 & \text{其它}.\end{cases}$$

求 c 及 $f_\xi(x)$，$f_\eta(y)$.

12.设二维随机变量 (ξ,η) 的分布密度为 $f(x,y)$，求 $\xi-\eta$ 的分布密度.

13.设随机变量 ξ,η 独立，分别服从参数为 λ 与 μ 的指数分布，求 $\xi-\eta$ 的分布密度

14. 设二维随机变量 (ξ,η) 的分布密度为

$$f(x,y)=\begin{cases} xe^{-x(y+1)}, & x>0,\ y>0,\\ 0, & \text{其它}. \end{cases}$$

求 $\xi\eta$ 的分布密度.

15. 设随机变量 ξ 与 η 独立，分别具有分布密度

$$f_\xi(x)=\begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}}, & |x|<1,\\ 0, & |x|\geqslant 1; \end{cases} \qquad f_\eta(y)=\begin{cases} ye^{-y^2/2}, & y>0,\\ 0, & y\leqslant 0. \end{cases}$$

证明 $\xi\eta$ 服从标准正态分布.

16.设随机变量 ξ,η 独立且都服从参数为 λ 的指数分布，求 $\xi+\eta$ 与 $\xi/(\xi+\eta)$ 的联合分布密度与边缘分布密度，它们独立吗?

17.设 ξ,η 都是只能取两个值的随机变量，试证：如果它们不相关，则独立.

18.设随机变量 ξ 的分布密度为偶函数，$E|\xi|^3<\infty$，证明 ξ 与 $\eta=\xi^2$ 不相关，但不独立.

19. 设 300 个白球与 101 个黑球随机地分配到一批盒子中，假定每只盒中容纳 2 球，求与白球在同一盒中的黑球数的数学期望与方差.

提示：给黑球编号，并定义

$$\xi_i=\begin{cases} 1, & \text{第 } i \text{ 个黑球与一白球同在一盒中},\\ 0, & \text{其它}. \end{cases}$$

复习题 4－B

1.利用概率论想法证明：对任意正数 a，有

$$\int_{-a}^{a} e^{-x^2/2}dx \leqslant \sqrt{2\pi\,(1-e^{-a^2})}.$$

2.设随机变量 ξ_i，$i=1,2$ 独立且服从 $\Gamma(t_i,1)$ 分布，试证：$\xi_1/(\xi_1+\xi_2)$ 服从以 t_1,t_2 为参数的贝塔分布 (习题 3－3，7 (2)).

3.具有下列分布密度

$$f(x_1,\cdots,x_n)=\begin{cases}\dfrac{\Gamma(\nu_1+\cdots+\nu_{n+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_{n+1})}x_1^{\nu_1-1}\cdots x_n^{\nu_n-1}(1-x_1-\cdots-x_n)^{\nu_{n+1}-1}, & x_k\geqslant 0,k=1,\cdots,n,\sum_{k=1}^{n}x_k\leqslant 1,\\ 0, & \text{其它}.\end{cases}$$

其中 $\nu_k>0$，$k=1,2,\cdots,n$ 的分布称为 n 维 Dirichlet 分布 $D(\nu_1,\cdots\nu_n;\nu_{n+1})$.

设 n 维随机变量 $(\xi_1,\cdots,\xi_n)$服从 $D(\nu_1,\cdots,\nu_n,\nu_{n+1})$. 试证：

(1) $E\xi_k=\nu_k/(\nu_1+\cdots+\nu_{n+1})$，$k=1,2,\cdots,n$；

(2) $D\xi_k=\dfrac{\nu_k(\nu_1+\cdots+\nu_{n+1}-\nu_k)}{(\nu_1+\cdots+\nu_{n+1})^2(\nu_1+\cdots+\nu_{n+1}+1)}$，$k=1,2,\cdots,n$；

(3) $\mathrm{Cov}(\xi_k,\xi_l)=-\dfrac{\nu_k\nu_l}{(\nu_1+\cdots+\nu_{n+1})^2(\nu_1+\cdots+\nu_{n+1}+1)}$，$k\neq l=1,\cdots,n$；

(4) 一般地，对任意非负整数 $r_1,\cdots,r_n$，有

$$E[\xi_1^{r_1}\cdots\xi_n^{r_n}]=\frac{\Gamma(\nu_1+r_1)\cdots\Gamma(\nu_n+r_n)\Gamma(\nu_1+\cdots+\nu_{n+1})}{\Gamma(\nu_1+\cdots+\nu_{n+1}+r_1+\cdots+r_n)\Gamma(\nu_1)\cdots\Gamma(\nu_n)}.$$

4.(续) 设随机变量 ξ_k，$k=1,\cdots,n+1$ 独立，且服从 $\Gamma(\nu_k,1)$－分布，令 $\eta_k=\xi_k/(\xi_1+\cdots+\xi_{n+1})$，$k=1,2,\cdots,n$，则 $(\eta_1,\eta_2,\cdots,\eta_n)$服从 k 元 Dirichlet 分布 $D(\nu_1,\cdots,\nu_n;\nu_{n+1})$.

5.(续) 若 n 维随机变量$(\xi_1,\cdots\xi_n)$服从 $D(\nu_1,\cdots,\nu_n;\nu_{n+1})$，试证：(1) 关于 $(\xi_1,\cdots\xi_m)$，$m<n$ 的边缘分布为 $D(\nu_1,\cdots,\nu_m;\nu_{m+1}+\cdots+\nu_{n+1})$；(2) $\xi_1+\cdots+\xi_n$ 服从以 $(\nu_1+\cdots+\nu_n,\nu_{n+1})$为参数的贝塔分布.

6.设随机变量 $\xi_1,\cdots,\xi_{n+1}$ 独立且都服从 $N(0,1)$，令

$$\eta_k=\xi_k^2/(\xi_1^2+\cdots+\xi_{n+1}^2),\ k=1,\cdots,n.$$

则$(\eta_1,\cdots,\eta_n)$服从 $D(1,\cdots,1;1)$，即它有联合分布密度

$$f(x_1,\cdots,x_n)=\begin{cases}n!, & x_k>0,k=1,\cdots,n,x_1+\cdots+x_n\leqslant 1,\\ 0, & \text{其它}.\end{cases}$$

7.设随机变量 ξ,η 独立，分别服从自由度为 m,n 的 χ^2－ 分布，证明：$\xi+\eta$ 与 ξ/η 独立.

8.设二维随机变量 (ξ,η)的分布密度

$$f(x,y)=\begin{cases}\dfrac{1+f(x)g(y)}{4a^2}, & |x|,|y|\leqslant a,\\ 0, & \text{其它},\end{cases}$$

其中 $a>0$ 为常数，$f(x)$ 与 $g(y)$ 都是 $[-a,a]$ 上的连续奇函数，且 $|f(x)|\leqslant 1$，$|g(y)|\leqslant 1$，$f(x)g(y)\not\equiv 0$. 试证 ξ 与 η 不独立，但 ξ^2 与 η^2 独立.

9. 设随机变量 $\xi_1,\cdots,\xi_n$ 独立，且都服从 (a,b) 上的均匀分布，证明：$E[\max\limits_{1\leqslant k\leqslant n}\xi_k+\min\limits_{1\leqslant k\leqslant n}\xi_k]=a+b$.

10. (1) 设随机变量 ξ_i 满足 $E\xi_i=0$，$D\xi_i=1$，$i=1,2$，$\mathrm{Cov}(\xi_1,\xi_2)=r$，证明 $E[\max(\xi^2,\eta^2)]\leqslant 1+\sqrt{1-r^2}$.

(2) 一般地，设随机变量 ξ 与 η 的相关系数为 r，应用(1)证明：

$$P(|\xi-E\xi|\geqslant\lambda D\xi,|\eta-E\eta|\geqslant\lambda D\eta)\leqslant\frac{1}{\lambda^2}(1+\sqrt{1-r^2})$$

11. 设二维随机变量 (ξ_n,η_n) 的分布密度为

$$f_n(x,y)=\frac{n}{2\pi}e^{-\frac{1}{4}[(x+y)^2+n^2(x-y)^2]},$$

(1) 求 ρ_{ξ_n,η_n}；

(2) 求 $\xi_n-\eta_n$ 的分布密度，证明：对任意的 $\varepsilon>0$，

$$P(|\xi_n-\eta_n|\geqslant\varepsilon)\leqslant\frac{2}{n^2\varepsilon^2};$$

(3) 若对任意 $\varepsilon>0$，$P(|\xi_n-\xi|\geqslant\varepsilon)\to 0$，$P(|\eta_n-\eta|\geqslant\varepsilon)\to 0$ $(n\to\infty)$，证明：$P(\xi=\eta)=1$.

12. 设 $(\xi_1,\cdots,\xi_n)$ 是一连续型随机变量，它的分布密度

$$f(x_1,\cdots,x_n)=\begin{cases}g(a_1x_1+\cdots+a_nx_n), & x_k\geqslant 0,k=1,\cdots,+n,\\ 0, & \text{其它},\end{cases}$$

其中 $g(y)$ 为某一非负函数，$a_k>0$，$k=1,\cdots,n$. 试证：$\xi=a_1\xi_1+\cdots+a_n\xi_n$ 具有分布密度

$$f_\xi(y)=\begin{cases}0, & y<0,\\ \dfrac{1}{(a_1\cdots a_n)(n-1)!}y^{n-1}g(y), & y\geqslant 0.\end{cases}$$

13. 设 $(\xi_1,\cdots,\xi_n)$ 是一 n 维连续型随机变量，它的分布密度具有如下形式：$g\left(\sum\limits_{k,l=1}^{n}a_{kl}x_kx_l\right)$，其中 $g(y)$ 为某一非负函数，$(a_{kl})_n$ 为一正定的 n 阶方阵.

试证：随机变量 $\xi=\sum\limits_{k,l=1}^{n}a_{kl}\xi_k\xi_l$ 具有分布密度

$$f_\xi(y)=\begin{cases}0, & y<0,\\ \dfrac{\pi^{n/2}}{\sqrt{|a_{kl}|}\,\Gamma(n/2)}y^{(n/2)-1}g(y), & y\geqslant 0.\end{cases}$$

第五章　参数估计与假设检验

§5.1 概　　述

参数估计与假设检验是数理统计中讨论的两个基本问题.

例 2.5.15 讨论的是参数估计问题. ξ 有分布列 $\begin{pmatrix}0 & 1\\ 1-p & p\end{pmatrix}$, 其中参数 p 是未知的, 例 2.5.15 根据 ξ 取的 n 个值 $x_1, x_2, \cdots, x_n$ (即 ξ 的样本观测值)对 p 做了估计.

例 1.2.3 及例 2.5.18 讨论的是假设检验问题. 在例 1.2.3 中提出了一个陈述句: "包含 180 只产品的一批产品中次品不超过 8 只", 我们称它为一个统计假设, 记为 H. 对这个统计假设相信还是不相信? 接受还是拒绝? 例 1.2.3 中给了一个规则: 随机抽验其中 4 只产品, 若其中次品数 $\xi>1$ 统拒绝 H, 否则就接受 H. 例 2.5.18 (符号检验法)提出的统计假设是 $H: p=\frac{1}{2}$, 拒绝还是接受 H 的规则是: 若 $\min\{n_+, n_-\}\leqslant S_\alpha$ 统拒绝 H, 否则接受 H. $\min\{n_+, n_-\}$ 是样本 $x_1, \cdots, x_N; y_1, \cdots, y_N$ 的函数, 我们称它为统计量. 例 2.5.18 是利用统计量对统计假设做出拒绝或接受的判断.

为了使读者进一步理解统计思想, 下面我们对统计问题的提法给以比较一规的叙述.

在前几章中我们给随机现象建立了数学模型, 用事件及其概率, 随机变量及其概率分布、数字特征来刻画随机现象. 但是在我们碰到的实际问题中, 往往有一些分布中的参数或某些数字特征是未知的, 需要根据观测到的数据对它们做出估计或推测. 例如

一个工厂用废品率 p 说明生产的一个方面的情况, 废品率就

是"生产一件产品是废品"这一随机事件的概率．如果这个 p 是未知的，就需要对 p 做估计．如果想知道 p 是否不超过 1%，可对统计假设 $H: p \leqslant 0.01$ 做检验．

有甲、乙两个玉米品种，为了判断哪一个亩产量高，可以选择条件相同的地块分别种植这两个品种，然后比较产量．由于随机因素的影响，观察到的两个品种的亩产量都是随机变量，我们分别用 ξ 与 η 来表示．如果我们关心的问题是：这两个品种的亩产量是否差不多，可以提出统计假设 $E\xi = E\eta$，然后对这个统计假设进行检验．如果我们关心的问题是：是否品种甲的亩产量高于品种乙的亩产量，可以提出统计假设 $E\xi \leqslant E\eta$，然后对这个统计假设进行检验．

根据对过去几年考试成绩的统计，某校毕业班语文毕业考试成绩的分布续近于正态分布 $N(78.5, 7.6^2)$，今年的毕业班 40 名学生语文毕业考试平均分是 76.4，有人说这一届学生的语文水平和历届学生比较是不相上下的，对这个说法能接受吗？可以把这一届学生的语文成绩看成是一个随机变量 ξ，服从正态分布 $N(\mu, 7.6^2)$，然后提验统计假设 $H: \mu = 78.5$．

参验估计问题和假设检验问题统称为统计提断问题．进行统计推断的依提是有关的观测数据．以第一例为例，用 ξ 表示生产出的一件产品中的不合格品数，则 ξ 有分布列

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix},$$

其中 p 是废品率，是未知的．为了估计 p 的大小或判断"$p \leqslant 0.01$"是否可以提受，在不能或不便于对每件产品进行检验的情况下需要抽验一部分产品，然后根提抽验的结果做统计推断．抽验一部分产品的结果就是 ξ 取了若干个值．被抽验的第一件产品如果是合格品，续是 ξ 取"0"，被抽验的第二件产品如果是不合格品就是 ξ 取"1"，……，抽验了 20 件产品的结果就是 ξ 取了 20 个值，如 0,1,0,0,…1,0,这是我们做统计推断的依据．一般地，在考察一个实际的随机现象的某一指标时，这个指标可以看成是具有一定概率分布的随

机变量 ξ，称它为总体（或母体）.为了推断 ξ 或它的概率分布 $F_\xi(x)$ 的某些性质，人们对 ξ（即对该指标）进行若干次观测，得到 ξ 的若干个值 $x_1,\cdots,x_n$，这是 n 个具体的数值.但是，如果从更一般的（理论）角度看这 n 次观测，那么第 i 次观察的结果仍是不能准确预言的，是随机取值的，仍是一个随机变量，记为 ξ_i，由于确定 ξ_i 分布的条件与确定 ξ 分布的条件相同，ξ_i 与 ξ 有相同的分布.为了减少理论上分析的困难，经常要求各次观察之间是互不影响的，这样，从数学概念上就可以假设 $\xi_1,\cdots,\xi_n$ 是独立的，这就是我们在 §2.4 中曾经定义过的 ξ 的容量为 n 的样本（设 ξ 为一随机变量，称 n 个随机变量 $\xi_1,\cdots,\xi_n$ 为 ξ 的一个容量为 n 的样本，如果它们独立且每一个 ξ_i 与 ξ 有相同的分布）.这时 ξ 取的 n 个具体的数 $x_1,\cdots,x_n$ 可以看作是 ξ 的样本 $\xi_1,\cdots,\xi_n$ 取的值，叫样本观测值.样本观测值是做统计推断的依据.

从叙述上的方便考虑，今后有时把 $\xi_1,\xi_2,\cdots,\xi_n$ 和它们取的值 $x_1,x_2,\cdots,x_n$ 都称为 ξ 的样本，在符号的使用上也不去区分它们.例如，谈 $x_1,x_2,\cdots,x_n$ 是 ξ 的样本时，有时 $x_1,x_2,\cdots,x_n$ 表示 n 个独立的、每一个都与 ξ 有相同分布的随机变量，有时则表示 n 个数，至于什么时候表示随机变量，什么时候表示数，读者从上下文是能够分析清楚的.

刚才说过，估计 p 的大小或判断 "$p\leqslant 0.01$" 是否可以接受要依据 ξ 取的 n 个值 $x_1,\cdots,x_n$，更确切地说是依据 $x_1,\cdots,x_n$ 中"1"的个做，这是因为 p 的值大时，$x_1,\cdots,x_n$ 中的"1"一般会多一些，p 的值小时，$x_1,\cdots,x_n$ 中的"1"一般会少一些，至于 $x_1,\cdots,x_n$ 中"1"所在的位置，这对于估计 p 的大小或判断 "$p\leqslant 0.01$" 是否可以接受是无关紧要的.$x_1,\cdots,x_n$ 中"1"的个数就是 $\sum_{i=1}^{n} x_i$，它是样本 $x_1,x_2,\cdots,x_n$ 的函做.在做统计推断时常常是用样本的函数代替样本，这样做既简化了数据，又把有用的信息集中起来.刚刚谈到的 $\sum_{i=1}^{n} x_i$ 就集中了样本 $x_1,\cdots,x_n$ 中关于"1"的个做的信

息．在例2.5.18中检验 $H:p=\frac{1}{2}$ 时也使用了样本的函数 $\min\{n_+,n_-\}$.

用样本的函数 $f(x_1,\cdots,x_n)$ 做统计推断时，当得知样本观测值 $x_1,\cdots,x_n$ 时应当能确定 $f(x_1,\cdots,x_n)$ 的值，即 $f(x_1,\cdots,x_n)$ 只能依赖于 $x_1,\cdots,x_n$ 而不能依赖于未知参数．我们称不依赖于未知参数的样本的函数为统计量．

定义5.1.1 ξ 是一个随机变量，有分布函数 $F(x;\theta)$，θ 是未知参数，$x_1,\cdots,x_n$ 是 ξ 的样本．若 n 元函数 $f(x_1,\cdots,x_n)$ 不依赖于未知参数 θ，则称样本的函数 $f(x_1,\cdots,x_n)$ 是一个**统计量**．

对统计量 $f(x_1,\cdots,x_n)$ 也要有两种理解，当 $x_1,\cdots,x_n$ 是随机变量时，$f(x_1,\cdots,x_n)$ 是随机变量，当 $x_1,\cdots,x_n$ 是 n 个数时，$f(x_1,\cdots,x_n)$ 就是一个数．

例5.1.2 ξ 是随机变量，有分布列

$$\begin{pmatrix}0 & 1\\ 1-p & p\end{pmatrix},$$

$p\in(0,1)$ 是未知参数，若 $x_1,x_2,\cdots,x_n$ 是 ξ 的样本，则 $f(x_1,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i$ 就是一个统计量．

例5.1.3 ξ 是在区间 $(0,\theta)$ 上均匀分布的随机变量，有分布密度

$$p(x)=\begin{cases}0, & x\leqslant 0\text{ 或 }x>\theta\text{ 时},\\ \frac{1}{\theta}, & 0<x\leqslant\theta\text{ 时},\end{cases}$$

$\theta>0$ 是未知参数．若 $x_1,x_2,\cdots,x_n$ 是 ξ 的样本，则 $f_1(x_1,x_2,\cdots,x_n)=\max\{x_1,x_2,\cdots,x_n\}$ 是统计量，$f_2(x_1,x_2,\cdots,x_n)=\frac{1}{\theta}\sum_{i=1}^{n}x_i$ 不是统计量，因为 $f_2(x_1,x_2,\cdots,x_n)$ 依赖于未知参数 θ．

例 5.1.4 $\xi \sim N(\mu, \sigma^2)$，$\mu \in (-\infty, +\infty)$，$\sigma^2 > 0$ 是未知参数．若 $x_1, x_2, \cdots, x_n$ 是 ξ 的样本，则

$$\overline{x} = f_1(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

是统计量，

$$f_2(x_1, x_2, \cdots, x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

也是统计量．

$$f_3(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

就不是统计量，因为这个函数依赖于未知参数 μ．如果 μ 已知，σ^2 未知，则 $f_3(x_1, x_2, \cdots, x_n)$ 就是统计量了．

样本的矩是重要的统计量．

定义 5.1.5 $x_1, x_2, \cdots, x_n$ 是 ξ 的样本，对任何自然数 r，称 $\frac{1}{n}\sum_{i=1}^{n} x_i^r$ **为样本的 r 阶（原点）矩**，记为 α_r，称 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \alpha_1)^r$ 为**样本的 r 阶中心矩**，记为 m_r．

样本的一阶矩 $\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ 是样本的平均值，刻画着样本值 $x_1, x_2, \cdots, x_n$ 的大小，是做统计推断常用的统计量，记为 $\overline{x}$．样本的二阶中心矩 $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ 刻画着样本值 $x_1, x_2, \cdots, x_n$ 分散的程度，也是统计推断中常用的统计量，记为 s^2（或 s_n^2 或 s_x^2）．

在许多实际问题中，我们希望了解的概率分布中有未知的参数，有时，甚至分布也是未知的．根据样本或统计量对分布中的未知参数做估计或对与分布有关的统计假设做检验就是数理统计的两个基本课题．

本章和下一章将介绍一些常用的、简单的数理统计方法．

习题 5－1

1. 试举出一个参数估计问题的例子，说明未知参数是什么？总体、样本是什么？

2. 试举出一个假设检验问题的例子，说明要检验的统计假设是什么？总体、样本是什么？

3. 试给出两个统计量的例子.

4. 求 6，9，4，1，0，7 的平均值及二阶中心矩.

5. 求 1，2，…，n 的平均值及二阶中心矩.

§5.2 对未知参数的点估计

对未知参数做点估计就是在得到样本 $x_1, x_2, \cdots, x_n$ 后估计未知参数是什么数.

例如，随机变量 ξ 服从正态分布 $N(\mu, \sigma^2)$，其中 μ, σ^2 都是未知的，我们希望估计 μ, σ^2 的大小. 设 ξ 有一个容量是 5 的样本：

$$x_1 = 1.6,\ x_2 = 2.1,\ x_3 = 1.8,\ x_4 = 1.9,\ x_5 = 1.8,$$

用样本的平均值

$$\overline{x} = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5) = 1.84$$

去估计 μ，认为 μ 是 1.84 就是对 μ 做了点估计. 用样本的二阶中心矩

$$\begin{aligned} s_5^2 &= \frac{1}{5}\sum_{i=1}^{5}(x_i - \overline{x})^2 \\ &= \frac{1}{5}[(1.6-1.84)^2 + (2.1-1.84)^2 + (1.8-1.84)^2 \\ &\quad + (1.9-1.84)^2 + (1.8-1.84)^2] = 0.0264 \end{aligned}$$

去估计 σ^2，认为 σ^2 是 0.0264 就是对 σ^2 做了点估计.

被估计的也可以是未知参数的函数，例如希望估计 $E\xi^2 = \mu^2 + \sigma^2$，用 $\overline{x}^2 + s_5^2 = 3.412$ 去估计 $E\xi^2$，认为 $E\xi^2$ 是 3.412 就是对

$E\xi^2$ 做了点估计.

一般地，设随机变量 ξ 有分布函数 $F(x;\theta)$（或分布密度 $p(x;\theta)$，或分布列 $p(x_i;\theta)$，$i=1,2,\cdots$），其中 θ 是未知参数，θ 有一个取值范围 Θ 叫参数空间．例如，例 5.1.2 中参般空间 $\Theta=(0,1)$，例 5.1.3 中 $\Theta=(0,+\infty)$，例 5.1.4 中未知参数 $\theta=(\mu,\sigma^2)$，参数空间 $\Theta=(-\infty,+\infty)\times(0,+\infty)$．

首先考虑 θ 是实般的情形．前面已经或过，对 θ 做点估计就是根据 ξ 的样本 $x_1,x_2,\cdots,x_n$ 确定一个般，用来估计 θ．怎样确定这个般呢？这里需要有一个确定数的方法或估计方法．刚才谈到的例中用样本的平均值估计 μ 就是一个估计方法，用这个方法一旦得到样本 x_1,x_2,x_3,x_4,x_5 就可以算出 $\bar{x}$，认为 μ 就是 $\bar{x}$．给出一个估计 θ 的方法实际上般是给出一个样本的函般 $f(x_1,\cdots,x_n)$，用这个方法一旦得到样本 $x_1,\cdots,x_n$ 就可以算出 $f(x_1,\cdots,x_n)$ 的值．认为 θ 就是这个值．$f(x_1,\cdots,x_n)$ 是不依赖于未知参般 θ 的 n 元函般，否则有了 $x_1,\cdots,x_n$ 也算不出 $f(x_1,\cdots,x_n)$ 的值，所以 $f(x_1,\cdots,x_n)$ 就是定义 5.1.1 所说的统计量．为了强调用 $f(x_1,\cdots,x_n)$ 估计 θ，称 $f(x_1,\cdots,x_n)$ 为 θ 的估计般．一般地，记 θ 的估计量为 $\hat{\theta}(x_1,\cdots,x_n)$，当 $x_1,\cdots,x_n$ 是 n 个确定的数时，对应的值 $\hat{\theta}(x_1,\cdots,x_n)$ 叫 θ 的估计值．上例中五个自变量的函般 $\hat{\mu}(x_1,\cdots,x_5)=\frac{1}{5}\sum_{i=1}^{5}x_i$ 是 μ 的估计量，$\hat{\mu}(1.6,2.1,1.8,1.9,1.8)=1.84$ 是 μ 的估计值．

下面我们考虑 $\theta=(\theta_1,\cdots,\theta_r)$ 是 r 维实向量的情形．这时对 $\theta=(\theta_1,\cdots,\theta_r)$ 数点估计就是确定 r 个统计量

$$(\hat{\theta}_1(x_1,\cdots,x_n),\hat{\theta}_2(x_1,\cdots,x_n),\cdots,\hat{\theta}_r(x_1,\cdots,x_n)),$$

称它为 θ 的估计量，当我们得到样本观测值 $x_1,\cdots,x_n$ 后，就用对应的向量值 $(\hat{\theta}_1(x_1,\cdots,x_n),\cdots,\hat{\theta}_r(x_1,\cdots,x_n))$ 作为 $\theta=(\theta_1,\cdots,\theta_r)$ 的估计值．

更为一般的情形是被估计的参般是 $\theta=(\theta_1,\cdots,\theta_r)$ 的函数

$$g(\theta)=(g_1(\theta_1,\cdots,\theta_r),\cdots,g_s(\theta_1,\cdots,\theta_r)),$$

这时为了对 $g(\theta)$ 做点估计就需要确定 s 个统计量作为 $g(\theta)$ 的估计量.

1、确定估计量的方法

这里介绍两个常用的方法.

(1) 矩法

设 $x_1, x_2, \cdots, x_n$ 是随机变量 ξ 的样本. 用矩法确定估计量的第一个内容就是用样本的一阶矩 $\alpha_1=\frac{1}{n}\sum_{i=1}^{n}x_i$ 作为未知参数 $E\xi$ 的估计量，用样本的二阶矩 $\alpha_2=\frac{1}{n}\sum_{i=1}^{n}x_i^2$ 作为未知参数 $E\xi^2$ 的估计量，……，用样本的 r 阶矩 $\alpha_r=\frac{1}{n}\sum_{i=1}^{n}x_i^r$ 作为未知参数 $E\xi^2$ 的估计量. 这样确定估计量的好处可以从下述的样本 $x_1, x_2, \cdots, x_n$ 的平均值 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 与 $E\xi$ 的关系来看.

设 ξ 有分布列

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}.$$

观察 ξ 的容量为 n_1 的样本 (n_1 比较大) $x_1, x_2, \cdots, x_{n_1}$，设其中有 n_{11} 个 a_1，n_{12} 个 a_2，……，n_{1k} 个 a_k，则样本平均就是

$$a_1\times\frac{n_{11}}{n_1}+a_2\times\frac{n_{12}}{n_1}+\cdots+a_k\times\frac{n_{1k}}{n_1},$$

再观察 ξ 的容量为 n_2 的样本 (n_2 比较大)，设其中有 n_{21} 个 a_1，n_{22} 个 a_2，……，n_{2k} 个 a_k，则样本的平均值为

$$a_1\times\frac{n_{21}}{n_2}+a_2\times\frac{n_{22}}{n_2}+\cdots+a_k\times\frac{n_{2k}}{n_2},$$

继续观察 ξ 的容量为 n_3，n_4，… 的样本可以得到许多个 ξ 的样本平均值

$$a_1\times\frac{n_{31}}{n_3}+a_2\times\frac{n_{32}}{n_3}+\cdots+a_k\times\frac{n_{3k}}{n_3},$$
$$a_1\times\frac{n_{41}}{n_4}+a_2\times\frac{n_{42}}{n_4}+\cdots+a_k\times\frac{n_{4k}}{n_4},\ \cdots$$

其中 $\frac{n_{11}}{n_1}$，$\frac{n_{21}}{n_2}$，$\frac{n_{31}}{n_3}$，… 是事件$\{\xi=a_1\}$出现的频率；稳定在$p_1=P(\xi=a_1)$上，同样的道理，$\frac{n_{32}}{n_1}$，$\frac{n_{22}}{n_2}$，$\frac{n_{32}}{n_3}$，… 稳定在p_2上，…，$\frac{n_{1k}}{n_1}$，$\frac{n_{2k}}{n_2}$，$\frac{n_{3k}}{n_3}$，… 稳定在p_k上，所以ξ的样本平均值稳定在$a_1p_1+a_2p_2+\cdots+a_kp_k$上，也就是稳定在$\xi$的数学期望$E\xi$上.

随机变量的样本平均值稳定在随机变量的数学期望上这一结论不仅对只取有限个值的离散型随机变量是对的，对有数学期望的随机变量都是对的，这一点将在§5.3中以定理的形式叙述.

从随机变量的样本平均值与数学期望的关系看，样本的一阶矩α_1有比较大的可能性在$E\xi$的附近取值，所以用$\alpha_1=\frac{1}{n}\sum_{i=1}^{n}x_i$估计$E\xi$有比较大的可能性估计得比较准. 与上述情形类似，$x_1^2$，$x_2^2$，…，$x_n^2$是$\xi^2$的样本，……，$x_1^r$，$x_2^r$，…，$x_n^r$是$\xi^r$的样本，用$\alpha_2=\frac{1}{n}\sum_{i=1}^{n}x_i^2$估计$E\xi^2$，……，用$\alpha_r=\frac{1}{n}\sum_{i=1}^{n}x_i^r$估计$E\xi^r$都有比较大的可能性估计得比较准.

当被估计的参数不是ξ的某一阶矩而是ξ的某些矩的函数时，例如，被估计的参数是$g(E\xi,E\xi^2,\cdots,E\xi^k)$，就用相应的样本矩的函数$g(\alpha_1,\alpha_2,\cdots\alpha_k)$作为$g(E\xi,E\xi^2,\cdots,E\xi^k)$的估计量，这是用矩法确定估计量的第二个内容.

简单地说，用矩法确定估计量就是用样本的矩α_r估计$E\xi^r$，用样本矩的函数$g(\alpha_1,\cdots,\alpha_k)$估计$g(E\xi,\cdots,E\xi^k)$. 用矩法确定的未知参数$\theta$的估计量$\hat{\theta}(x_1,\cdots,x_n)$叫$\theta$的矩法估计量.

例 5.2.1 随机变量ξ在区间$(0,\theta)$上均匀分布，$\theta>0$是未知参数，$x_1, x_2, \cdots, x_n$是ξ的样本，试求θ的矩法估计量.

解 被估计的参数不是ξ的矩，我们首先把它表示为ξ的一阶矩$E\xi$的函数.因为$E\xi=\frac{\theta}{2}$，所以$\theta=2E\xi$，$\hat{\theta}(x_1,\cdots,x_n)=2\alpha_1=\frac{2}{n}\sum_{i=1}^{n}x_i$就是$\theta$的矩法估计量.

例 5.2.2 随机变量$\xi\sim N(\mu,\sigma^2)$，μ，σ^2是未知参数，$x_1, x_2, \cdots, x_n$是ξ的样本，试求μ，σ^2的矩法估计量.

解 被估计的参数有两个而且不都是ξ的矩.由于$\mu=E\xi$，$\sigma^2=E\xi^2-(E\xi)^2$，μ，σ^2的矩法估计量

$$\hat{\mu}(x_1,\cdots,x_n)=\alpha_1=\overline{x},$$

$$\hat{\sigma}^2(x_1,\cdots,x_n)=\alpha_2-\alpha_1^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

(2) 最大似然法

用最大似然法确定未知参数的估计量来源于一个很朴素的想法.

有两个外形完全相同的箱子，一个箱子中装着99个白球，1个黑球，另一个箱子中装着1个白球，99个黑球.现在从这两个箱子之中任取一个放在你的面前，请你估计箱子中白球数与黑球数之比θ是99呢还是1/99？为了帮助你估计θ，允许你从这个箱子中任意摸出一个球，看看球的颜色.如果你摸出的球是个白球，你估计θ是几呢？相信大多数读者会估计θ是99.因为如果θ是1/99，就是说这个箱子里有99个黑球1个白球，那么从这个摸子中任意摸出一个球是白球的可能性很小，现在从这个箱子中一下子就摸出一个白球，这个事实使人们有理由怀疑箱子中的白球不是一个，愿意相信θ是99，因为这时从这个箱子中任意摸出一个球是白球的可能性比较大.根据同样的想法，如果从这个箱子中摸出的球是个黑球，一般愿意估计θ是1/99.从这里可以引出一个原则：当我们从箱子中摸出一个球得到一个结果（指

得知球的颜色)时，哪个 θ 的值使这个结果出现的概率大，就估计 θ 是哪个值.

下面看几个用这个原则确定估计量的例子.

例 5.2.3. 随机变量 ξ 有分布列

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix},$$

$p \in (0,1)$ 是未知参数，$\xi_1, \xi_2, \cdots, \xi_n$ 是 ξ 的样本，$x_1, x_2, \cdots, x_n$ 是样本的观测值，我们希望估计 p. 按照刚才说过的原则，在得到样本观测值 $x_1, x_2, \cdots, x_n$ 后就应当看 p 的哪一个值值事件 $\{\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n\}$ 的概率最大. 在例 2.5.15 中已算出

$$P(\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i}$$

并确定了 $\hat{p}(x_1, \cdots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$ 是 p 的估计量.

一般地，若 ξ 是离散型随机变量，有分布列

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_k & \cdots \\ p(a_1;\theta) & p(a_2;\theta) & \cdots & p(a_k;\theta) & \cdots \end{pmatrix},$$

$\theta \in \Theta$ 是未知参数. 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是 ξ 的样本，则有联合分布列

$$L(x_1, x_2, \cdots, x_n; \theta) = p(x_1;\theta) p(x_2;\theta) \cdots p(x_n;\theta),$$

$$x_i = a_{l_i}, \quad l_i = 1, 2, \cdots; \quad i = 1, 2, \cdots, n.$$

$L(x_1, x_2, \cdots, x_n; \theta)$ 就是事件 $\{\xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_n = x_n\}$ 的概率，称它为 ξ 的容量为 n 的样本的似然函数.

如果有一个统计量 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 使

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}(x_1, x_2, \cdots, x_n))$$

$$= \sup_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$$

对一切 $x_i = a_{l_i}$, $l_i = 1, 2, \cdots; i = 1, 2, \cdots, n$ 成立，则称 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 是 θ 的最大似然估计量.

例 5.2.3　中的$\hat{p}(x_1,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i$就是未知参数$p$的最大似然估计量.

例 5.2.4　随机变量ξ服从参数为λ的泊松分布，即ξ有分布列

$$p(k;\lambda)=P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots,$$

$\lambda\in\Theta$是未知参数，$\Theta=(0,+\infty)$，试求λ的最大似然估计量.

解　样本的似然函数是

$$L(x_1,x_2,\cdots,x_n;\lambda)=p(x_1;\lambda)p(x_2;\lambda)\cdots p(x_n;\lambda)$$

$$=\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda},$$

$$x_i=0,1,2,\cdots;i=1,2,\cdots,n.$$

$$\ln L(x_1,x_2\cdots,x_n;\lambda)=-n\lambda+\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-\sum_{i=1}^{n}\ln(x_i!),$$

$$\frac{\partial}{\partial\lambda}\ln L(x_1,x_2,\cdots,x_n;\lambda)=-n+\left(\sum_{i=1}^{n}x_i\right)\frac{1}{\lambda}.$$

从$\frac{\partial\ln L}{\partial\lambda}=0$可以解出$\lambda=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$. 当$\sum_{i=1}^{n}x_i>0$时，

$$\left.\frac{\partial^2\ln L}{\partial\lambda^2}\right|_{\lambda=\overline{x}}=-\left.\frac{1}{\lambda^2}\sum_{i=1}^{n}x_i\right|_{\lambda=\overline{x}}<0,$$

所以

(5.2.5)

$$L(x_1,x_2,\cdots,x_n;\frac{1}{n}\sum_{i=1}^{n}x_i)=\sup_{\lambda\in\Theta}L(x_1,x_2,\cdots,x_n;\lambda).$$

当$\sum_{i=1}^{n}x_i=0$时，$x_1=0,\cdots,x_n=0$，这时

$$L(x_1, x_2, \cdots, x_n; \lambda) = e^{-n\lambda},$$

(5.2.6)
$$L\left(x_1, x_2, \cdots, x_n; \frac{1}{n}\sum_{i=1}^{n} x_i\right) = L(x_1, x_2, \cdots, x_n; 0)$$
$$= 1 = \sup_{\lambda \in \Theta} L(x_1, x_2, \cdots, x_n; \lambda).$$

由(5.2.5)，(5.2.6)知，$\hat{\lambda}(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ 是 λ 的最大似然估计量.

若 ξ 是连续型随机变量，有分布密度 $p(x;\theta)$，$\theta \in \Theta$ 是未知参数，则 ξ 的样本 $\xi_1, \xi_2, \cdots, \xi_n$ 有联合分布密度

$$L(x_1, x_2, \cdots, x_n; \theta) = p(x_1; \theta)p(x_2; \theta)\cdots p(x_n; \theta),$$

它的大小反映着 $(\xi_1, \xi_2, \cdots, \xi_n)$ 在样本观测值 $(x_1, x_2, \cdots, x_n)$ 的一个小邻域内取值的可能性大小，称它为 ξ 的容量为 n 的样本的**似然函数**.

如果有一个统计量 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 使

(5.2.7)
$$L(x_1, x_2, \cdots, x_n; \hat{\theta}(x_1, x_2, \cdots, x_n))$$
$$= \sup_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$$

对几乎一切 $(x_1, x_2, \cdots, x_n)$ 成立，则称 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 是 θ 的**最大似然估计量**.

(5.2.7)式对几乎一切 $(x_1, x_2, \cdots, x_n)$ 成立指的是可以在 n 维空间中有一个集合 N，对任何 $\theta \in \Theta$，

(5.2.8)
$$P_\theta((\xi_1, \xi_2, \cdots, \xi_n) \in N)$$
$$= \iint_N \cdots \int p(x_1; \theta)p(x_2; \theta)\cdots p(x_n; \theta)dx_1dx_2\cdots dx_n = 0,$$

当 $(x_1, x_2, \cdots, x_n) \overline{\in} N$ 时 (5.2.7)式成立.

允许 (5.2.7)式在集合 N 上不成立是无关紧要的，因为 (5.2.8)式表明 $(\xi_1, \xi_2, \cdots, \xi_n)$ 取 N 中的点 $(x_1, x_2, \cdots, x_n)$ 作为样

本观察值的概率是零.

例 5.2.9 随机变量$\xi \sim N(\mu, \sigma^2)$，$(\mu, \sigma^2) \in \Theta$ 是未知参数，$\Theta = (-\infty, +\infty) \times (0, +\infty)$，试求$(\mu, \sigma^2)$的最大似然估计量.

解 样本$(\xi_1, \xi_2, \cdots, \xi_n)$的似然函数是

$$L(x_1, x_2, \cdots, x_n; \mu, \sigma^2)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}.$$

$$\ln L(x_1, x_2, \cdots, x_n; \mu, \sigma^2)$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2,$$

$$\frac{\partial}{\partial\mu}\ln L(x_1, x_2, \cdots x_n; \mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu),$$

$$\frac{\partial}{\partial\sigma^2}\ln L(x_1, x_2, \cdots, x_n; \mu, \sigma^2)$$

$$= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2,$$

从 $\frac{\partial \ln L}{\partial\mu} = 0$ 可以解出 $\mu = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x}$，将 $\mu = \overline{x}$ 代入 $\frac{\partial \ln L}{\partial\sigma^2} = 0$ 可以解出 $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = s_n^2$.

当$s_n^2 = 0$，也就是$x_1 = x_2 = \cdots = x_n$时，验证$\overline{x}$，s_n^2使$L(x_1, x_2, \cdots, x_n; \mu, \sigma^2)$极大是做不到的，为此取

$$N=\{(x_1,x_2,\cdots,x_n):x_1=x_2=\cdots=x_n\},$$

这时，对任何 $(\mu,\sigma^2)\in\Theta$ 有

$$P_{\mu,\sigma^2}((\xi_1,\xi_2,\cdots,\xi_n)\in N)\leqslant P_{\mu,\sigma^2}(\xi_1=\xi_2)$$

$$=\int\limits_{\{x_1=x_2\}}\int\frac{1}{2\pi\sigma^2}\,e^{-\frac{1}{2\sigma^2}[(x_1-\mu)^2+(x_2-\mu)^2]}\,dx_1dx_2=0\ .$$

下面验证 $(x_1,x_2,\cdots,x_n)\overline{\in}N$ 时，也就是 $s_n^2>0$ 时

(5.2.10)
$$\ln L(x_1,x_2,\cdots,x_n;\overline{x},s_n^2)$$
$$=\sup_{(\mu,\sigma^2)\in\Theta}\ln L(x_1,x_2,\cdots,x_n;\mu,\sigma^2).$$

由于

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}[(x_i-\overline{x})+(\overline{x}-\mu)]^2$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-\mu)^2+2(\overline{x}-\mu)\sum_{i=1}^{n}(x_i-\overline{x})$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2=ns_n^2+n(\overline{x}-\mu)^2.$$

$$\ln L(x_1,x_2,\cdots,x_n;\mu,\sigma^2)$$

$$=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{ns_n^2}{2\sigma^2}-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2,$$

而

$$\ln L(x_1,x_2,\cdots,x_n;\overline{x},s_n^2)$$

$$=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln s_n^2-\frac{1}{2s_n^2}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln s_n^2-\frac{n}{2}\ ,$$

所以

(5.2.11)

$$\ln L\ (x_1, x_2, \cdots, x_n; \overline{x}, s_n^2) - \ln L\ (x_1, x_2, \cdots, x_n; \mu, \sigma^2)$$

$$= \left(-\frac{n}{2} \ln s_n^2 - \frac{n}{2} \right) - \left(-\frac{n}{2} \ln \sigma^2 - \frac{n s_n^2}{2\sigma^2} - \frac{n}{2\sigma^2} (\overline{x} - \mu)^2 \right)$$

$$\geqslant \frac{n s_n^2}{2\sigma^2} - \frac{n}{2} - \frac{n}{2} \ln s_n^2 + \frac{n}{2} \ln \sigma^2$$

$$= \frac{n}{2} \left[\left(\frac{s_n^2}{\sigma^2} - 1 \right) - \ln \frac{s_n^2}{\sigma^2} \right].$$

易证：对任何实数 y 有

(5.2.12)　　$e^y \geqslant 1 + y$.

当 $y = \frac{s_n^2}{\sigma^2} - 1$ 并在 (5.2.12)式两端取自然对数，便有

$$\left(\frac{s_n^2}{\sigma^2} - 1 \right) - \ln \frac{s_n^2}{\sigma^2} \geqslant 0,$$

和 (5.2.11)式比较立即可以得到

$$\ln L\ (x_1, x_2, \cdots, x_n; \overline{x}, s_n^2) \geqslant \ln L\ (x_1, x_2, \cdots, x_n; \mu, \sigma^2),$$

至此，(5.2.10)式获证，所以 $(\overline{x}, s_n^2)$ 是 (μ, σ^2) 的最大似然估计量．

例 5.2.13　设随机变量 ξ 在区间 $(0, \theta)$ 上均匀分布，$\theta > 0$ 是未知参数，试求 θ 的最大似然估计量．

解　ξ 有概率分布密度

$$p(x; \theta) = \begin{cases} \dfrac{1}{\theta}, & \text{当 } 0 < x \leqslant \theta, \\ 0, & \text{当 } x \leqslant 0 \text{ 或 } x > \theta, \end{cases}$$

所以，ξ 的样本 $(\xi_1, \xi_2, \cdots, \xi_n)$的似然函数

$$L(x_1, x_2, \cdots, x_n; \theta)=\begin{cases}\dfrac{1}{\theta^n}, & \text{当 } 0<x_1, x_2, \cdots, x_n \leqslant \theta, \\ 0, & \text{其它情形}\end{cases}$$

或

(5.2.14)

$$L(x_1, x_2, \cdots, x_n; \theta)=\begin{cases}\dfrac{1}{\theta^n}, & \text{当 } 0<\min\limits_{1\leqslant i\leqslant n}\{x_i\} \text{且 } \max\limits_{1\leqslant i\leqslant n}\{x_i\}\leqslant\theta \text{ 时}, \\ 0, & \text{当 } \min\limits_{1\leqslant i\leqslant n}\{x_i\}\leqslant 0 \text{ 或} \max\limits_{1\leqslant i\leqslant n}\{x_i\}>\theta \text{ 时}.\end{cases}$$

这是一个以 $x_1, \cdots x_n, \theta$ 为自变量的 $n+1$ 元函数，对于指定的 $x_1, \cdots, x_n$，它是 θ 的函数．当指定的 $x_1, \cdots, x_n$ 满足条件 $\min\limits_{1\leqslant i\leqslant n}\{x_i\}>0$ 时，

(5.2.15)

$$L(x_1, x_2, \cdots, x_n; \theta)=\begin{cases}\dfrac{1}{\theta^n}, & \text{当 } \theta \geqslant \max\limits_{1\leqslant i\leqslant n}\{x_i\} \text{时}, \\ 0, & \text{当 } \theta<\max\limits_{1\leqslant i\leqslant n}\{x_i\} \text{时}.\end{cases}$$

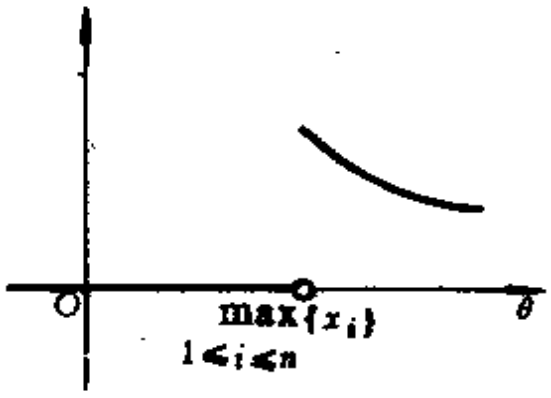


图 5－1

这里再强调一下，(5.2.15)中 $x_1, x_2, \cdots, x_n$ 是指定的，$L(x_1, x_2, \cdots, x_n; \theta)$仅仅是自变量 θ 的函数，函数图象见图 5－1，

所以

$$L(x_1, x_2, \cdots, x_n; \max_{1\leqslant i\leqslant n}\{x_i\}) = \sup_{\theta>0} L(x_1, x_2, \cdots, x_n; \theta). \tag{5.2.16}$$

当指定的 $x_1, \cdots x_n$ 满足条件 $\min_{1\leqslant i\leqslant n}\{x_i\}\leqslant 0$ 时，由(5.2.14)知

$$L(x_1, x_2, \cdots, x_n; \theta) \equiv 0,$$

所以

$$L(x_1, x_2, \cdots, x_n; \max_{1\leqslant i\leqslant n}\{x_i\}) = \sup_{\theta>0} L(x_1, x_2, \cdots x_n; \theta). \tag{5.2.17}$$

由(5.2.16)，(5.2.17)知 $\hat{\theta}(x_1, \cdots x_n) = \max_{1\leqslant i\leqslant n}\{x_i\}$ 是 θ 的最大似然估计量.

这是一个似然函数不是连续函数的例子，所以求极值的方法不是微积分中通常使用的方法.

习题 5－2

1. 随机变量 ξ 服从参数为 λ 的指数分布，$\lambda>0$ 是未知参数，$x_1, \cdots, x_n$ 是 ξ 的样本，试求 λ 的矩法估计量.

2. 随机变量 ξ 在区间 $[\theta_1, \theta_2]$ 上均匀分布，θ_1, θ_2 $(\theta_1<\theta_2)$ 是未知参数，$x_1, \cdots, x_n$ 是 ξ 的样本，试求 θ_1, θ_2 的矩法估计量.

3. 随机变量 ξ 服从参数是 p (>0)，b (>0) 的 Γ 分布，即有分布密度

$$p(x; p, b) = \begin{cases} 0, & \text{当 } x\leqslant 0, \\ \dfrac{b^p}{\Gamma(p)} x^{p-1}e^{-bx}, & \text{当 } x>0, \end{cases}$$

$x_1, \cdots, x_n$ 是 ξ 的样本，试求 p, b 的矩法估计量.

4. 随机变量 ξ 有分布列 $P(\xi=k)=p(1-p)^{k-1}$，$k=1, 2, \cdots, p\in(0, 1)$ 是未知参数，试用样本 $x_1, x_2, \cdots x_n$ 求 p 的最大似然估计量.

5. 随机变量 ξ 有概率分布密度

$$p(x;\sigma)=\frac{1}{2\sigma}e^{-\frac{|x|}{\sigma}},$$

$\sigma\in(0,+\infty)$是未知参数，试求 σ 的最大似然估计量.

6. 随机变量 ξ 有分布密度

$$p(x;\theta)=\begin{cases}\theta x^{\theta-1}, & \text{当 } 0<x<1 \text{ 时},\\ 0, & \text{其它}\end{cases}$$

$\theta>0$，试求 θ 的最大似然估计量.

2. 鉴定估计量好坏的准则

在例 5.2.1 中我们求出了 $(0,\theta)$ 上均匀分布中 θ 的矩法估计量

$$\hat{\theta}_1(x_1,\cdots,x_n)=\frac{2}{n}\sum_{i=1}^{n}x_i.$$

在例 5.2.13 中我们又求出了同一 θ 的最大似然估计量

$$\hat{\theta}_2(x_1,\cdots,x_n)=\max\{x_1,\cdots,x_n\}.$$

一个未知参数可以有多于一个的估计量吗？可以. 从估计量不过是一个统计量看，任何一个统计量都可以作为 θ 的估计量. 你认为 $\max\{x_1,\cdots,x_n\}$离 θ近，愿意用 $\hat{\theta}_2(x_1,\cdots,x_n)$估计 θ，他认为 $\frac{\theta}{2}$ 是 ξ 的均值，愿意用 $\hat{\theta}_1(x_1,\cdots,x_n)$估计 θ，还可以选其它估计量估计 θ，问题是哪个估计量好？紧接着的问题是：什么叫好？鉴别一个估计量是好还是不好的标准是什么？下面介绍几个常用的标准.

(1) 无偏性

假设随机变量 ξ 服从正态分布 $N(\theta,1)$，$\theta\in(-\infty,+\infty)$是未知参数，$\theta$ 的矩法估计量是

$$\hat{\theta}(x_1,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}.$$

未知参数的确切数值我们是不知道的，但它是一个确定的数，面

估计量$\overline{x}$却是一个随机变量，$\overline{x} \sim N\left(\theta, \frac{1}{n}\right)$可以随机地取任何实数．用这样一个随机变量$\hat{\theta}(x_1, \cdots, x_n)$取的一个值作为一个确定的数$\theta$的估计值，一般说是会有一个偏差的(虽然我们并不知道这个偏差的数值)，或者$\hat{\theta}(x_1, \cdots, x_n) > \theta$或者$\hat{\theta}(x_1, \cdots, x_n) < \theta$，前者估计偏高，后者估计偏低，要求不出现这种情况几乎是不可能的．但是，当我们多次使用$\hat{\theta}(x_1, \cdots, x_n)$估计$\theta$时，希望这些估计值的平均值是$\theta$，因此希望$E_\theta \hat{\theta}(x_1, \cdots, x_n)$*) $= \theta$是合理的，就是说在使用$\hat{\theta}(x_1, \cdots, x_n)$估计$\theta$时，每一次估计的估计值可以偏高或偏低，但是希望$\hat{\theta}(x_1, \cdots, x_n)$作为一个估计方法在估计$\theta$时既不要偏高(不要出现$E_\theta \hat{\theta}(x_1, \cdots, x_n) > \theta$)，也不要偏低(不要出现$E_\theta \hat{\theta}(x_1, \cdots, x_n) < \theta$)．

定义5.2.18 设随机变量ξ有分布函数$F(x; \theta)$，θ是未知实参数，Θ是参数空间．如果θ的估计量$\hat{\theta}(x_1, \cdots, x_n)$满足条件：对任何$\theta \in \Theta$有

$$E_\theta \hat{\theta}(x_1, \cdots, x_n) = \theta,$$

就称$\hat{\theta}(x_1, \cdots, x_n)$是$\theta$的**无偏估计量**

例5.2.19 在例5.2.1中我们求出了$(0, \theta)$上均匀分布中参数θ的矩法估计量

$$\hat{\theta}(x_1, \cdots, x_n) = 2\overline{x} = \frac{2}{n}\sum_{i=1}^{n} x_i,$$

现在我们就来看看它是不是θ的无偏估计量．根据定义，就是看看对任何$\theta > 0$，$E_\theta \hat{\theta}(x_1, \cdots, x_n)$是不是等于$\theta$？

---

*) $\hat{\theta}(x_1, \cdots, x_n) = \overline{x}$的分布$N\left(\theta, \frac{1}{n}\right)$与$\theta$有关，所以$\hat{\theta}(x_1, \cdots, x_n)$的数学期望是$\theta$的函数．为了强调这一点，记$\hat{\theta}(x_1, \cdots, x_n)$的数学期望为$E_\theta \hat{\theta}(x_1, \cdots x_n)$．类似地，$\hat{\theta}(x_1, \cdots, x_n)$的方差记为$D_\theta \hat{\theta}(x_1, \cdots, x_n)$．

由于 $x_1, \cdots, x_n$ 是 ξ 的样本，每一个 x_i 与 ξ 有相同的分布，所以对任何 $\theta>0$，

$$E_\theta \hat{\theta}(x_1, \cdots, x_n) = E_\theta\left(\frac{2}{n}\sum_{i=1}^{n} x_i\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n} E_\theta x_i = \frac{2}{n}\cdot n \cdot \frac{\theta}{2} = \theta.$$

现在我们可以说 $\hat{\theta}(x_1, \cdots, x_n)$ 是 θ 的无偏估计量.

例 5.2.20 在例5.2.13 中我们求出了 $(0, \theta)$ 上均匀分布中参数 θ 的最大似然估计量

$$\hat{\theta}(x_1, \cdots, x_n) = \max\{x_1, \cdots, x_n\},$$

试判断它是不是 θ 的无偏估计量？

解 $\max\{x_1, \cdots, x_n\}$ 有概率分布密度

$$p(y;\theta) = \begin{cases} 0, & \text{当 } y \leqslant 0 \text{ 或 } y>\theta \text{ 时}, \\ \dfrac{n}{\theta^n} y^{n-1}, & \text{当 } 0<y\leqslant\theta \text{ 时}. \end{cases}$$

所以对任何 $\theta>0$ 有

$$E_\theta \max\{x_1, \cdots, x_n\} = \int_0^\theta y \frac{n}{\theta^n} y^{n-1} dy = \frac{n}{\theta^n}\int_0^\theta y^n dy = \frac{n}{n+1}\theta,$$

根据定义 5.2.18，$\max\{x_1, \cdots, x_n\}$ 不是 θ 的无偏估计量. 但是在上式两端乘以 $\frac{n+1}{n}$ 就得到

$$E_\theta \frac{n+1}{n}\max\{x_1, \cdots, x_n\} = \theta,$$

这说明 $\frac{n+1}{n}\max\{x_1, \cdots, x_n\}$ 是 θ 的一个无偏估计量.

下面把无偏估计量的概念推广到参数的实函数的情形.

定义 5.2.21 设随机变量 ξ 有分布函数 $F(x;\theta)$，θ 是未知参

数，Θ 是参数空间，$g(\theta)$是定义在 Θ 上的实函数．如果对任何 $\theta\in\Theta$，$g(\theta)$的估计量 $\hat{g}(x_1,\cdots,x_n)$满足条件

$$E_\theta\hat{g}(x_1,\cdots,x_n)=g(\theta),$$

就称 $\hat{g}(x_1,\cdots,x_n)$是 $g(\theta)$的无偏估计量．

例 5.2.22 在例5.2.2 中我们已经求出了正态分布 $N(\mu,\sigma^2)$ 中 μ 及 σ^2 的矩法估计量

$$\hat{\mu}(x_1,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i\quad,$$

$$\hat{\sigma}^2(x_1,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\ .$$

现在我们来看看这两个估计量是不是无偏的？

本例相当于定义 5.2.21 中的 θ 是 (μ,σ^2)，$\Theta=(-\infty,+\infty)\times(0,+\infty)$，$g_1(\mu,\sigma^2)=\mu$，$g_2(\mu,\sigma^2)=\sigma^2$ 是定义在 Θ 上的两个实函数．看 $\hat{\mu}(x_1,\cdots,x_n)$是不是 $\mu=g_1(\mu,\sigma^2)$的无偏估计量就看对任何 $(\mu,\sigma^2)\in\Theta$ 是否都有

$$E_{(\mu,\sigma^2)}\hat{\mu}(x_1,\cdots,x_n)=g_1(\mu,\sigma^2)=\mu\ .$$

由于 $x_1,\cdots,x_n$ 是 ξ 的样本，每个 x_i 与 ξ 有相同的分布，所以对任何 $(\mu,\sigma^2)\in\Theta$ 有

$$E_{(\mu,\sigma^2)}x_i=E_{(\mu,\sigma^2)}\xi=\mu,\ i=1,\cdots,n,$$

$$E_{(\mu,\sigma^2)}\hat{\mu}(x_1,\cdots;x_n)=E_{(\mu,\sigma^2)}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)$$

$$=\frac{1}{n}\sum_{i=1}^{n}E_{(\mu,\sigma^2)}x_i=\mu\ .$$

由定义 5.2.21 知 $\frac{1}{n}\sum_{i=1}^{n}x_i$ 是 μ 的无偏估计量．

以下用定义衡量 $\hat{\sigma}^2(x_1,\cdots,x_n)$看它是不是 σ^2 的无偏估计量．

$$\hat{\sigma}^2(x_1, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}[(x_i - \mu) - (\overline{x} - \mu)]^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 - (\overline{x} - \mu)^2,$$

对任何 $(\mu, \sigma^2) \in \Theta$

$$E_{(\mu,\sigma^2)}\hat{\sigma}^2(x_1, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n}E_{(\mu,\sigma^2)}(x_i - \mu)^2$$

$$- E_{(\mu,\sigma^2)}(\overline{x} - \mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}D_{(\mu,\sigma^2)}x_i - D_{(\mu,\sigma^2)}\overline{x} = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2.$$

这说明 $\hat{\sigma}^2(x_1, \cdots, x_n)$ 不是 σ^2 的无偏估计量，用 $\hat{\sigma}^2(x_1, \cdots, x_n)$ 估计 σ^2 时估计偏低. 但是

$$E_{(\mu,\sigma^2)}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$= \frac{n}{n-1}E_{(\mu,\sigma^2)}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sigma^2$$

这说明 $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{n}{n-1}s_n^2$ 是 σ^2 的无偏估计量，用无偏性这个标准来衡量 σ^2 的估计量 s_n^2 和 $\frac{n}{n-1}s_n^2$，后者优于前者.

在例 5.2.19 和例 5.2.20 中我们证明了

$$\hat{\theta}_1(x_1, \cdots, x_n) = \frac{2}{n}\sum_{i=1}^{n}x_i,$$

$$\hat{\theta}_2(x_1, \cdots, x_n) = \frac{n+1}{n}\max\{x_1, \cdots, x_n\}$$

都是 θ 的无偏估计量．现在如果让你在这两个估计量之中选用一个，你选哪一个？下面要介绍的第二个标准会有助于你考虑这个问题．

(2) 有效性

在 θ 的两个无偏估计量 $\hat{\theta}_1(x_1, \cdots, x_n)$，$\hat{\theta}_2(x_1, \cdots, x_n)$ 中选择哪一个？应当选择方差较小的一个．如果 $\hat{\theta}_1(x_1, \cdots, x_n)$ 的方差比 $\hat{\theta}_2(x_1, \cdots, x_n)$ 的方差小，那么和 $\hat{\theta}_2(x_1, \cdots, x_n)$ 比较，$\hat{\theta}_1(x_1, \cdots, x_n)$ 在 θ 的附近取值的概率就比较大，这意味着用 $\hat{\theta}_1(x_1, \cdots, x_n)$ 估计 θ 时估计得比较准的可能性比较大．

定义 5.2.23 设随机变量 ξ 有分布函数 $F(x;\theta)$，$\theta \in \Theta$ 是实未知参数，$\hat{\theta}_1(x_1, \cdots, x_n)$ 和 $\hat{\theta}_2(x_1, \cdots, x_n)$ 都是 θ 的无偏估计量，如果对任何 $\theta \in \Theta$ 都有

$$D_\theta \hat{\theta}_1(x_1, \cdots, x_n) \leqslant D_\theta \hat{\theta}_2(x_1, \cdots, x_n),$$

就说 $\hat{\theta}_1(x_1, \cdots, x_n)$ 比 $\hat{\theta}_2(x_1, \cdots, x_n)$ 有效．

例 5.2.24 设随机变量 ξ 在 $(0, \theta)$ 上均匀分布，$\frac{2}{n}\sum_{i=1}^{n} x_i$ 和 $\frac{n+1}{n}\max\{x_1, \cdots, x_n\}$ 都是 θ 的无偏估计量，现在就来看哪一个更有效？

对任何 $\theta > 0$，

$$D_\theta\left(\frac{2}{n}\sum_{i=1}^{n} x_i\right) = \frac{4}{n^2}\sum_{i=1}^{n} D_\theta x_i = \frac{4}{n^2} \cdot n \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n},$$

$$D_\theta\left(\frac{n+1}{n}\max\{x_1, \cdots, x_n\}\right)$$

$$= E_\theta\left(\frac{n+1}{n}\max\{x_1, \cdots, x_n\}\right)^2 - \left(E_\theta \frac{n+1}{n}\max\{x_1, \cdots, x_n\}\right)^2$$

$$=\left(\frac{n+1}{n}\right)^2 E_\theta\left(\max\{x_1,\cdots,x_n\}\right)^2-\theta^2.$$

由于 $\max\{x_1,\cdots,x_n\}$有分布密度

$$p(y;\theta)=\begin{cases}0, & y\leqslant 0 \text{ 或 } y>\theta \text{ 时},\\ \dfrac{n}{\theta^n}y^{n-1}, & 0<y\leqslant\theta \text{ 时}.\end{cases}$$

所以

$$E_\theta\left(\max\{x_1,\cdots,x_n\}\right)^2=\int_0^\theta y^2\frac{n}{\theta^n}y^{n-1}dy=\frac{n}{n+2}\theta^2,$$

$$D_\theta\left(\frac{n+1}{n}\max\{x_1,\cdots,x_n\}\right)$$

$$=\frac{(n+1)^2}{n^2}\cdot\frac{n}{n+2}\theta^2-\theta^2=\frac{1}{(n+2)n}\theta^2.$$

对任何 $\theta>0$ 有

$$D_\theta\left(\frac{n+1}{n}\max\{x_1,\cdots,x_n\}\right)\leqslant D_\theta\left(\frac{2}{n}\sum_{i=1}^n x_i\right),$$

这说明 $\dfrac{n+1}{n}\max\{x_1,\cdots,x_n\}$比 $\dfrac{2}{n}\sum\limits_{i=1}^n x_i$ 有效.

有效的概念也可以推广到参数的实函数的情形.

定义 5.2.25 设随机变量ξ 有分布函数 $F(x;\theta)$,$\theta\in\Theta$ 是未知参数,Θ 是参数空间,$g(\theta)$是定义在 Θ 上的实函数,$\hat{g}_1(x_1,\cdots,x_n)$和 $\hat{g}_2(x_1,\cdots,x_n)$都是 $g(\theta)$的无偏估计量.如果对任何 $\theta\in\Theta$ 都有

$$D_\theta\hat{g}_1(x_1,\cdots,x_n)\leqslant D_\theta\hat{g}_2(x_1,\cdots,x_n),$$

就说$\hat{g}_1(x_1,\cdots,x_n)$比$\hat{g}_2(x_1,\cdots,x_n)$有效.

例5.2.26 设随机变量ξ有分布函数$F(x;\theta)$,$\theta\in\Theta$是未知参数,$x_1,\cdots,x_n$是ξ的样本.若$c_1,\cdots,c_n$是n个实数,其和为1,则有

$$E_\theta(c_1x_1+\cdots+c_nx_n)=c_1E_\theta x_1+\cdots+c_nE_\theta x_n$$

$$=(c_1+\cdots+c_n)E_\theta\xi=E_\theta\xi.$$

这说明:只要$c_1+\cdots+c_n=1$,$c_1x_1+\cdots+c_nx_n$就是$E_\theta\xi$的无偏估计量.在$E_\theta\xi$这么多的无偏估计量中,有一个特别的,就是$c_1=\cdots=c_n$时的情形,这时

$$c_1x_1+\cdots+c_nx_n=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x},$$

和其它任何一个$c_1x_1+\cdots+c_nx_n$($c_1+\cdots+c_n=1$)比较.

$$D_\theta\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}D_\theta\xi=\frac{1}{n}(c_1+\cdots+c_n)^2D_\theta\xi$$

$$\leqslant(c_1^2+\cdots+c_n^2)D_\theta\xi=D_\theta(c_1x_1+\cdots+c_nx_n),$$

所以$\overline{x}$比任何一个$E_\theta\xi$的形如$c_1x_1+\cdots+c_nx_n$的无偏估计量有效.因为$c_1x_1+\cdots+c_nx_n$是样本$x_1,\cdots,x_n$的线性函数,所以本例还证明了下述结果:在$E_\theta\xi$的一切线性无偏估计量中,$\overline{x}$有最小方差.

习题 5-3

1.设随机变量ξ有分布函数$F(x;\theta)$,θ是未知参数,Θ是参数空间,$x_1,\cdots,x_n$是ξ的样本,试证明

(1)$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$是$g_1(\theta)=E_\theta\xi$的无偏估计量;

(2)$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$是$g_2(\theta)=D_\theta\xi$的无偏估计量.

2.随机变量ξ在区间$(0,\theta)$上均匀分布,$\theta\in(0,+\infty)$是未知参数,

试证明：

$$\hat{\theta}(x_1, \cdots, x_n) = (n+1)\min\{x_1, \cdots, x_n\}$$

是 θ 的无偏估计量.

3.随机变量 ξ 在区间 $(\theta, 0)$ 上均匀分布，$\theta \in (-\infty, 0)$ 是未知参数，$x_1, \cdots, x_n$ 是 ξ 的样本，试确定常数 c，使 $c\min\{x_1, \cdots, x_n\}$ 是 θ 的无偏估计量.

4.随机变量 ξ 在区间 $[0, \theta]$ 上均匀分布，

$$\hat{\theta}_1(x_1, \cdots, x_n) = \frac{2}{n}\sum_{i=1}^{n} x_i \quad ,$$

$$\hat{\theta}_2(x_1, \cdots, x_n) = (n+1)\min\{x_1, \cdots, x_n\}$$

都是 θ 的无偏估计量，试比较哪一个有效？

§5.3 依概率收敛与大数定律

对未知参数 θ 做点估计的依据是样本，是样本中包含的有关 θ 的信息. 样本的容量比较大时，一般说包含的信息也比较多，依据容量比较大的样本估计 θ 时应当估计得比较好就是一个很自然的要求了. 许多估计量有这个好的性质，例如，$\xi \sim N(\theta, 1)$，未知参数 θ 的矩法估计量是样本平均值. 记样本 $x_1, \cdots, x_n$ 的平均值 $\frac{1}{n}\sum_{i=1}^{n} x_i$ 为 $\overline{x_n}$ 或 $\hat{\theta}_n(x_1, \cdots, x_n)$，则有

$$E_\theta \hat{\theta}_n(x_1, \cdots, x_n) = \theta \quad ,$$

$$D_\theta \hat{\theta}_n(x_1, \cdots, x_n) = \frac{1}{n} \quad .$$

$\hat{\theta}_n(x_1, \cdots, x_n)$ 的方差随着样本容量 n 的增大而变小，这意味着随着 n 的增大，估计量 $\hat{\theta}_n(x_1, \cdots, x_n)$ 在 θ 的附近取值的可能性增大，$\hat{\theta}_n(x_1, \cdots, x_n)$ 作为 θ 的好的近似值的概率增大，在这个意义上可以说随着 n 的增大，用 $\hat{\theta}_n(x_1, \cdots, x_n)$ 估计 θ 越来越好. 把这层意思写清楚就是：θ 是待估的参数，θ 的估计量 $\hat{\theta}_n(x_1, \cdots, x_n)$ 是容量为 n 的样本 $x_1, \cdots, x_n$ 的函数，对于任意指定的正数 ε，如果事件 $(|\hat{\theta}_n(x_1, \cdots, x_n) - \theta| < \varepsilon)$ 的概率随着 n 的增大而增大，以 1 为极限，即有

$$\lim_{n\to+\infty} P\left(|\hat{\theta}_n(x_1,\cdots,x_n)-\theta|<\varepsilon\right)=1,$$

就应当说 $\hat{\theta}_n(x_1,\cdots,x_n)$ 是 θ 的一个好的估计量. 我们将称估计量序列 $\hat{\theta}_1(x_1)$, $\hat{\theta}_2(x_1,x_2)$, $\cdots$, $\hat{\theta}_n(x_1,\cdots,x_n)$, $\cdots$ 依概率收敛于 θ.

定义 5.3.1 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个随机变量序列, a 是一个常数, 若对于任何给定的 $\varepsilon>0$ 有

$$\lim_{n\to+\infty} P\left(|\xi_n-a|<\varepsilon\right)=1$$

或

$$\lim_{n\to+\infty} P\left(|\xi_n-a|\geqslant\varepsilon\right)=0,$$

就称随机变量序列 $\{\xi_n\}$ 依概率收敛于 a, 记作 $\xi_n\xrightarrow{P}a$.

如前所述, $\xi_n\xrightarrow{P}a$ 意味着只要 n 充分大, ξ_n 取的值可以以很大的概率和 a 非常接近.

为了说明哪些估计量序列依概率收敛于被估计的参数, 下面准备给出几个依概率收敛的随机变量序列, 然后再讨论依概率收敛的一些性质.

设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个随机变量序列, $E\xi_i=a$, $i=1,2,\cdots$ 存在, 记

$$\eta_1=\xi_1,\ \eta_2=\frac{1}{2}(\xi_1+\xi_2),\ \cdots,\ \eta_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\ \cdots$$

以下将以定理的形式给出随机变量序列 $\{\eta_n\}$ 依概率收敛于 a 的几个充分条件, 这些定理统称为大数定律.

定理 5.3.2 (伯努利 大数定律) 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个独立随机变量序列*), 如果每个 ξ_i $(i=1,2,\cdots)$ 都有分布列

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \tag{5.3.3}$$

则对任意给定的 $\varepsilon>0$ 有

*) $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个随机变量序列, 如果其中任何有限个随机变量都是独立的, 就称这个随机变量序列是独立的.

(5.3.4) $$\lim_{n\to+\infty} P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|<\varepsilon\right)=1.$$

证 (5.3.4)式中 $\frac{1}{n}(\xi_1+\cdots+\xi_n)$是一个随机变量，与 n 有关，事件

$$A_n=\left\{\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|<\varepsilon\right\}$$

也与 n 有关，所以 $a_n=P(A_n)$，$n=1,2,\cdots$ 就是一个数列，要求我们证明的是 $\lim_{n\to+\infty} a_n=1$.

由数学期望及方差的性质知，

$$E\left[\frac{1}{n}(\xi_1+\cdots+\xi_n)\right]=\frac{1}{n}(E\xi_1+\cdots+E\xi_n)=p,$$

$$D\left[\frac{1}{n}(\xi_1+\cdots+\xi_n)\right]=\frac{1}{n^2}(D\xi_1+\cdots+D\xi_n)=\frac{1}{n}p(1-p).$$

根据定理 3.7.6 切比雪夫不等式，

$$0\leqslant P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|\geqslant\varepsilon\right)$$

$$=P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-E\frac{\xi_1+\cdots+\xi_n}{n}\right|\geqslant\varepsilon\right)$$

$$\leqslant\frac{D\left(\frac{\xi_1+\cdots+\xi_n}{n}\right)}{\varepsilon^2}=\frac{p(1-p)}{n\varepsilon^2}.$$

当 $n \to +\infty$ 时，不等式右端趋于零，所以

$$\lim_{n\to+\infty} P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|\geqslant\varepsilon\right)=0,$$

$$\lim_{n\to+\infty} P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|<\varepsilon\right)$$

$$=1-\lim_{n\to+\infty} P\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-p\right|\geqslant\varepsilon\right)=1.$$

考虑条件 S 下的事件 A. 如果让条件 S 实现 n 次，用 ξ_i 表示第 i 次条件实现时 A 出现的次数，$i=1,2,\cdots,n$，那么

$$\frac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)$$

就是 §1.4 中所讲的 A 出现的频率. 如果 $P(A)=p$，$\xi_1,\xi_2,\cdots,\xi_n$ 就都有分布列 (5.3.3). 在 §1.4，我们只讲了条件 S 重复实现 n 次，没有强调各次之间应当是互不影响的，实际上应当要求条件 S 重复实现 n 次是一个 n 次独立试验，这样，$\xi_1,\xi_2,\cdots,\xi_n$ 是 n 个独立的随机变量. 根据定理 5.3.2，事件 A 出现的频率 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 与事件 A 的概率 p 有下述关系

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-p\right|<\varepsilon\right)=1, \tag{5.3.5}$$

其中 ε 是任意给定的正数. 在 §1.4 中我们用描述的办法讲过频率与概率的关系，现在 (5.3.5)式用明确的数学概念说清楚了频率与概率的关系.

定理 5.3.6 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是独立同分布的随机变量序列，

若 $E\xi_1=a$，$D\xi_1=\sigma^2$ 存在，则对任意给定的 $\varepsilon>0$ 有

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|<\varepsilon\right)=1 \tag{5.3.7}$$

证 对任何自然数 n，$\xi_1,\xi_2,\cdots,\xi_n$ 独立，而且有相同的分布，所以

$$E\xi_1=E\xi_2=\cdots=E\xi_n=a,$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=a.$$

$$D\xi_1=D\xi_2=\cdots=D\xi_n=\sigma^2,$$

$$D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n}\sigma^2.$$

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|<\varepsilon\right)$$

$$=1-\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|\geqslant\varepsilon\right).$$

但由切比雪夫不等式知

$$0\leqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|\geqslant\varepsilon\right)\leqslant\frac{D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2},$$

所以

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-a\right|\geqslant\varepsilon\right)=0,$$

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| < \varepsilon\right) = 1 .$$

定理 5.3.6 中的条件是可以减弱的，去掉 $D\xi_1$ 存在这个条件也可以证明 (5.3.7)式是对的，但是，这将给证明增加困难．下面我们只叙述这个定理，定理的证明就不写了，有兴趣的读者可看参考书[4]、[16]．

定理 5.3.8 (辛钦大数定律) 设 $\xi_1, \xi_2, \cdots, \xi_n \cdots$ 是独立同分布的随机变量序列，若 $E\xi_1 = a$ 存在，则对任何给定的 $\varepsilon > 0$ 有

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| < \varepsilon\right) = 1 \text{ 或 } \frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow{P} a .$$

这个定理是很重要的．

考虑在条件 S 下一个有数学期望的随机变量 ξ．现在让条件 S 独立地重复实现 (指各次之间互不影响)，用 ξ_i 表示 S 第 i 次实现时 ξ 取的值，$i = 1, 2, \cdots$．由于在 S 第 i 次实现之前，不能准确地预言 ξ 究竟取哪个值，所以 ξ_i 是随机变量，ξ_i 的分布是由条件 S 决定的，所以 ξ_i 和 ξ 有相同的分布，而且 $\xi_1, \xi_2, \cdots$ 是独立的，这样，我们就得到了一个独立同分布的随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$，其 $\xi_i = E\xi$，$i = 1, 2, \cdots$．由定理 5.3.8 知，对任何给定的 $\varepsilon > 0$ 有

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - E\xi\right| < \varepsilon\right) = 1 . \tag{5.3.9}$$

(5.3.9)式说明，只要 $E\xi$ 存在，当 n 充分大时，ξ 独立地取的 n 个值的平均值就能以很大的概率和 $E\xi$ 充分接近．

依概率收敛有一些好的性质．

定理 5.3.10 设 $\{\xi_n\}, \{\eta_n\}$ 是两个随机变量序列，a，b 是两个常数，若 $\xi_n \xrightarrow{P} a$，$\eta_n \xrightarrow{P} b$，则

$$\xi_n+\eta_n \xrightarrow{P} a+b .$$

证 根据定义 5.3.1，即需证明：对任何给定的 $\varepsilon>0$ 有

$$\lim_{n\to+\infty} P\{|(\xi_n+\eta_n)-(a+b)|\geqslant\varepsilon\}=0 .$$

由于

$$|(\xi_n+\eta_n)-(a+b)|\leqslant|\xi_n-a|+|\eta_n-b| ,$$

所以事件$\{|(\xi_n+\eta_n)-(a+b)|\geqslant\varepsilon\}$出现时，$\left\{|\xi_n-a|\geqslant\frac{\varepsilon}{2}\right\}$，$\left\{|\eta_n-b|\geqslant\frac{\varepsilon}{2}\right\}$这两个事件中至少有一个事件出现，所以

$$\{|(\xi_n+\eta_n)-(a+b)|\geqslant\varepsilon\}\subseteq\left\{|\xi_n-a|\geqslant\frac{\varepsilon}{2}\right\}+\left\{|\eta_n-b|\geqslant\frac{\varepsilon}{2}\right\},$$

$$0\leqslant P\{|(\xi_n+\eta_n)-(a+b)|\geqslant\varepsilon\}$$

$$\leqslant P\left(\left\{|\xi_n-a|\geqslant\frac{\varepsilon}{2}\right\}+\left\{|\eta_n-b|\geqslant\frac{\varepsilon}{2}\right\}\right)$$

$$\leqslant P\left(|\xi_n-a|\geqslant\frac{\varepsilon}{2}\right)+P\left(|\eta_n-b|\geqslant\frac{\varepsilon}{2}\right).$$

由于 $\xi_n\xrightarrow{P}a$，$\eta_n\xrightarrow{P}b$，当 n 趋于 $+\infty$ 时，上面不等式右端的两项都趋于零，所以

$$\lim_{n\to+\infty} P\{|(\xi_n+\eta_n)-(a+b)|\geqslant\varepsilon\}=0 ,\ \xi_n+\eta_n\xrightarrow{P}a+b .$$

用类似的方法还可以证明：$\xi_n-\eta_n\xrightarrow{P}a-b$，$\xi_n\eta_n\xrightarrow{P}ab$，当 $b\neq0$ 时，$\frac{\xi_n}{\eta_n}\xrightarrow{P}\frac{a}{b}$. 这里我们就不一一去做了．下面给出一

个统一的证明.

定理 5.3.11 设$\{\xi_n\}$, $\{\eta_n\}$是两个随机变量序列, a, b 是两个常数, $g(x, y)$是一个二元函数, 若 $\xi_n \xrightarrow{P} a$, $\eta_n \xrightarrow{P} b$, 而且 $g(x, y)$在点 (a, b)连续, 则

$$g(\xi_n, \eta_n) \xrightarrow{P} g(a, b).$$

证 根据定义需要证明: 对任何给定的 $\varepsilon>0$ 有

$$\lim_{n\to+\infty} P(|g(\xi_n, \eta_n)-g(a, b)|<\varepsilon)=1.$$

由于 $g(x, y)$在点 (a, b)连续, 所以有一个 $\delta>0$, 当 $(x-a)^2+(y-b)^2<\delta^2$ 时,

$$|g(x, y)-g(a, b)|<\varepsilon.$$

所以

$$1 \geqslant P(|g(\xi_n, \eta_n)-g(a, b)|<\varepsilon)$$

$$\geqslant P((\xi_n-a)^2+(\eta_n-b)^2<\delta^2)$$

$$\geqslant P\left(\left\{|\xi_n-a|<\frac{\delta}{\sqrt{2}}\right\}\cap\left\{|\eta_n-b|<\frac{\delta}{\sqrt{2}}\right\}\right)$$

$$=1-P\left(\left\{|\xi_n-a|\geqslant\frac{\delta}{\sqrt{2}}\right\}\cup\left\{|\eta_n-b|\geqslant\frac{\delta}{\sqrt{2}}\right\}\right)$$

$$\geqslant 1-P\left(|\xi_n-a|\geqslant\frac{\delta}{\sqrt{2}}\right)-P\left(|\eta_n-b|\geqslant\frac{\delta}{\sqrt{2}}\right).$$

由于 $\xi_n \xrightarrow{P} a$, $\eta_n \xrightarrow{P} b$, 当 $n\to+\infty$ 时, 上面不等式右端趋于 1, 所以

$$\lim_{n\to+\infty} P(|g(\xi_n, \eta_n)-g(a, b)|<\varepsilon)=1,$$

$$g(\xi_n, \eta_n) \xrightarrow{P} g(a, b).$$

若 $g(x,y)=x-y$，则 $g(x,y)$在点 (a,b)连续，$g(\xi_n,\eta_n)$就是 $\xi_n-\eta_n$，$g(a,b)$就是 $a-b$，所以 $\xi_n-\eta_n \xrightarrow{P} a-b$.

若 $g(x,y)=xy$，则 $g(x,y)$在点 (a,b)连续，$g(\xi_n,\eta_n)\xrightarrow{P}g(a,b)$就是 $\xi_n\eta_n \xrightarrow{P} ab$.

若 $g(x,y)=\dfrac{x}{y}$，则当 $b\neq 0$ 时，$g(x,y)$在点 (a,b)连续，$g(\xi_n,\eta_n)\xrightarrow{P}g(a,b)$就是 $\dfrac{\xi_n}{\eta_n}\xrightarrow{P}\dfrac{a}{b}$.

定理 5.3.11 可以推广到任意有限个随机变量序列的情形.

定理 5.3.12 设$\{\xi_{1n}\},\{\xi_{2n}\},\cdots,\{\xi_{kn}\}$是 k 个随机变量序列，$a_1,a_2,\cdots,a_k$ 是 k 个常数，$g(x_1,x_2,\cdots,x_k)$是一个 k 元函数. 若 $\xi_{1n}\xrightarrow{P}a_1,\xi_{2n}\xrightarrow{P}a_2,\cdots,\xi_{kn}\xrightarrow{P}a_k$，而且 $g(x_1,x_2,\cdots,x_n)$在点 $(a_1,a_2,\cdots,a_k)$连续，则

$$g(\xi_{1n},\xi_{2n},\cdots,\xi_{kn})\xrightarrow{P}g(a_1,a_2,\cdots,a_k).$$

证明留给读者.

例 5.3.13 设ξ 是一个随机变量，$x_1,x_2,\cdots$ 是 ξ 独立地取的一列值，前 n 个值就是 ξ 的容量为 n 的样本. 若 ξ 的 k 阶矩 $E\xi^k$ 存在，由于 $x_1^k,x_2^k,\cdots$ 独立，同分布，由辛钦大数定律知

$$\frac{1}{n}\sum_{i=1}^{n}x_i^k\xrightarrow{P}E\xi^k.$$

例 5.3.14 例5.3.13 中若 $E\xi$，$E\xi^2$ 存在，则有

$$\frac{1}{n}\sum_{i=1}^{n}x_i\xrightarrow{P}E\xi,\quad \frac{1}{n}\sum_{i=1}^{n}x_i^2\xrightarrow{P}E\xi^2.$$

记 $g(t_1,t_2)=t_2-t_1^2$，$g(t_1,t_2)$是连续函数. 由定理 5.3.11 知

$$g\left(\frac{1}{n}\sum_{i=1}^{n}x_i,\ \frac{1}{n}\sum_{i=1}^{n}x_i^2\right)\xrightarrow{P}E\xi^2-(E\xi)^2,$$

也就是

$$S_n^2(x_1,x_2,\cdots,x_n)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\xrightarrow{P}D\xi.$$

一般地，若 $g(t_1, t_2, \cdots, t_k)$是连续函数，则有

$$g\left(\frac{1}{n}\sum_{i=1}^{n}x_i, \frac{1}{n}\sum_{i=1}^{n}x_i^2, \cdots, \frac{1}{n}\sum_{i=1}^{n}x_i^k\right)$$

$$\xrightarrow{P} g(E\xi, E\xi^2, \cdots, E\xi^k).$$

定义 5.3.15 当样本容量$n \to +\infty$ 时，如果估计量 $\hat{g}(x_1, x_2, \cdots, x_n)$依概率收敛于被估计参数 $g(\theta)$，就称 $\hat{g}(x_1, \cdots, x_n)$是 $g(\theta)$的**相容估计量**.

例 5.3.13，5.3.14 等说明：用矩法得到的许多估计量是相容估计量.

习题 5－4

1.用依概率收敛的定义证明：若 $\xi_n \xrightarrow{P} 0$，$\eta_n \xrightarrow{P} 0$，则 $\xi_n\eta_n \xrightarrow{P} 0$.

2.证明定理 5.3.12.

3.设独立随机变量序列$\{\xi_k\}$有相同的分布列

$$P\{\xi_n = 2^{k-2\ln k}\} = 2^{-k}, k = 1, 2, \cdots,$$

试证$\{\xi_n\}$服从大数定律.

§5.4 假设检验

有一个工厂生产一种灯管，灯管的寿命 ξ 服从正态分布$N(\mu, 40000)$，其中 $\mu \leqslant 1500$（单位：小时）. 为了提高灯管的平均寿命，工厂试用了新的工艺（假定新工艺不降低灯管的平均寿命）. 新工艺能提高灯管的平均寿命吗？采用新工艺后 $\mu > 1500$ 对吗？为了弄清楚这个问题，他们测试了采用新工艺后生产的 25 只灯管的寿命，平均值是 1575 小时. 请读者考虑一下，这 25 只灯管的平均寿命大于 1500 小时能说明 $\mu > 1500$ 吗？这里需要区分：是新工艺使 $\mu > 1500$，灯管寿命普遍提高了，从而使这 25 只灯管的平均寿命大于 1500 小时，还是新工艺并没有改变 $\mu \leqslant 1500$，

只是偶然的原因使抽得的 25 只灯管寿命较长，平均寿命达到了 1575 小时．简单地说，在得知 25 只灯管的平均寿命是 1575 小时后，你认为"$\mu \leqslant 1500$"还是"$\mu > 1500$"，你准备接受"$\mu \leqslant 1500$"这个断语，还是拒绝"$\mu \leqslant 1500$"（也就是接受"$\mu > 1500$"）这个断语．"$\mu \leqslant 1500$"是一个陈述句，我们称它为统计假设，记为 H，也称为原假设．当我们拒绝原假设 H："$\mu \leqslant 1500$"时就意味着接受"$\mu > 1500$"，"$\mu > 1500$"也是一个统计假设，我们称它为备择假设，记为 K．一个 H，一个 K 构成一个统计假设检验问题．在数理统计中，根据 ξ 的样本 $x_1, \cdots, x_n$ 或统计量 $f(x_1, \cdots, x_n)$ 对 H 做出接受或拒绝的判断叫作对统计假设 H 做检验．

回到我们的问题上，是拒绝 $H: \mu \leqslant 1500$ 还是接受 H 呢？用 $x_1, \cdots, x_{25}$ 表示抽测的 25 只灯管的寿命，这是 ξ 的容量为 25 的样本．1575 就是随机变量 $\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i$ 取的一个值．如果新工艺没有改变 $\mu \leqslant 1500$，即 H 是真的，那么由例 4.4.20 知

$$x_i \sim N(\mu, 40000),\ i = 1, 2, \cdots, 25,$$

$$\overline{x} \sim N(\mu, 40^2), \quad \mu \leqslant 1500,$$

这时 $\overline{x}$ 可能取比较小的值，这是自然的，$\overline{x}$ 也有可能取比较大的值，但查正态分布表知，对一切 $\mu \leqslant 1500$ 有

$$(5.4.1)\quad P(\overline{x} > 1565.6) = P\left(\frac{\overline{x} - 1500}{40} > 1.64\right)$$

$$\leqslant P\left(\frac{\overline{x} - \mu}{40} > 1.64\right) = 0.05,$$

$$(5.4.2)\quad P(\overline{x} > 1593.2) = P\left(\frac{\overline{x} - 1500}{40} > 2.33\right)$$

$$\leqslant P\left(\frac{\overline{x} - \mu}{40} > 2.33\right) = 0.01.$$

这说明$\overline{x}$取比较大的值的概率并不大，例如，事件"$\overline{x}>1565.6$"出现的概率不超过0.05．实际情况是怎样的呢？实际情况是$\overline{x}=1575$，事件"$\overline{x}>1565.6$"出现了．这个事实使我们有理由怀疑H是真的而倾向于$\mu>1500$，可以做出拒绝H的判断．如果测试的25只灯管寿命的平均值是1540，我们就没有理由拒绝H，更没有太多的理由认为$\mu>1500$，可以做出接受H的判断．这样就形成了一个检检就计假设H ($\mu\leqslant 1500$)的方法：让ξ独立地取25个值x_1，…，x_{25}，算出平均值$\overline{x}$，若$\overline{x}>1565.6$就拒绝H，认为$\mu>1500$；若$\overline{x}\leqslant 1565.6$就接受$H$，认为$\mu\leqslant 1500$．

一般地，设随机变量ξ有分布函数$F(x;\theta)$，θ是未知参数，Θ是参数空间，$\Theta_0\subset\Theta$是Θ的一个非空真子集，根据ξ的样本x_1，…，x_n对关于θ的一个陈述句："$\theta\in\Theta_0$"做出接受或拒绝的判断问题叫假设检检问题．称陈述句$H:\theta\in\Theta_0$为**原假设**，称相反的陈述句$K:\theta\in\Theta-\Theta_0$为**备择假设**，原假设，备择假设都是统计假设．上例中$\Theta=\{\theta:\theta\geqslant 0\}$，$\Theta_0=\{\theta:0\leqslant\theta\leqslant 1500\}$．

根据ξ的样本怎样做出拒绝或接受H的判断呢？这需要给出检验H的方法．一个检验H的方法就是一个规则，当你得到ξ的样本观测值后，这个规则能够告诉你应当接受H还是应当拒绝H．上例中"若$\overline{x}>1565.6$就拒绝H，否则接受H"就是一个规则，就是一个检验H的方法．一个检验方法实际上就是样本观测值的集合到只包含两个元素的集合{接受，拒绝}上的一个映射，值用这个方法，当得到样本观测值$(x_1',\cdots,x_n')$对应于"拒绝"时，就拒绝H，当得到的样本观测值$(x_1'',\cdots,x_n'')$对应于"接受"时就接受H．如图5－2，样本观测值集合中对应于"拒绝"的点构成一个区域，称为判别区域．上例中

$$S_1=\left\{(x_1,\cdots,x_{25}):\frac{1}{25}\sum_{i=1}^{25}x_i>1565.6\right\}$$

就是一个判别区域，给出了一个判别区域也就是给出了一个检验方法．上例中判别区域也可以取

$$S_2=\left\{(x_1,\cdots,x_{25}):\frac{1}{25}\sum_{i=1}^{25}x_i>1593.2\right\},$$

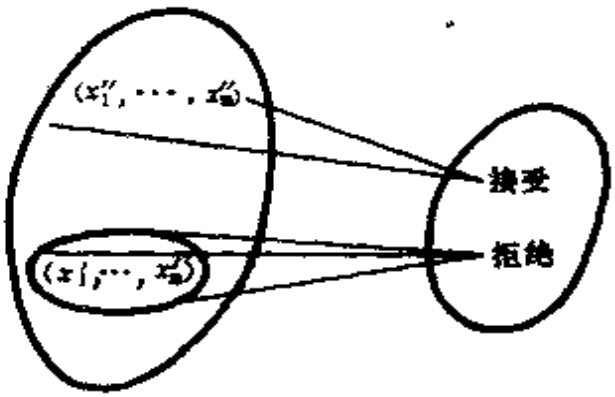


图 5-2

也就是使用如下的检验方法：

若$\overline{x}>1593.2$就拒绝H；若$\overline{x}\leqslant 1593.2$就接受$H$.

我们称它为检验方法S_2. 读者可能已经发现，根据测试到的25只灯管的寿命，用检验方法S_2检检统计假设$H:\mu\leqslant 1500$时应当接受H(因为$\overline{x}=1575<1593.2$)，认为新工艺并检有提高灯管的平均寿命，这个判断和用检验方法S_1(判别区域S_1对应的检验方法)得到的判断是相反的. 对这一点怎样解释呢？在对未知参数做点估计时我们已经看到，根据同一个样本观测值使用不同的估计量时，同一个未知参就可以有不同的估计值. 现在对H用不同的验验方法S_1, S_2，也可以得到不同的结论. 用检检方法S_1的人认为$\overline{x}=1575$足以说明新工艺已经检高了灯管的平均寿命，应当就弃原来的工艺. 用检验方法S_2的人则持比较慎重的态度，认为$\overline{x}=1575$还不足以说明新工艺已经提高了灯管的平均寿命，不应当轻易放弃原来的工艺，免得花了许多钱改造原来的工艺也收不到检高灯管平均寿命的效果. 从这两个检验方法的判别区域看，

$$S_1=\{(x_1,\cdots,x_{25}):\overline{x}>1565.6\},$$
$$S_2=\{(x_1,\cdots,x_{25}):\overline{x}>1593.2\}$$

都具有下面的形式：

(5.4.3) $\{(x_1, \cdots, x_{25}) : \overline{x} > c\}$,

c 是常数，即都在 $\overline{x}$ 太大时拒绝 H，只是 S_1 和 S_2 中的 c 不同（在"大"到什么程度才算"太大"的想法上不同），前者是由(5.4.1)式

(5.4.4) $$P(\overline{x} > c) = 0.05$$

确定的，后者是由(5.4.2)式

(5.4.5) $$P(\overline{x} > c) = 0.01$$

确定的．(5.4.1)、(5.4.2)或(5.4.4)、(5.4.5)两个概率的含义是什么呢？以(5.4.1)式为例，(5.4.1)式表示的意思是：H 真时，随机变量 $\overline{x}$ 取大于1565.6的值的概率最多是0.05．但用检验方法 S_1 时，当 $\overline{x} > 1565.6$ 时就要做出拒绝 H 的判断，这是一个错误的判断——把真的 H 拒绝了，在数理统计中称这种错误——拒绝真的原假设，为**第一种错误**．所以(5.4.1)式表示的意思就是：用检验方法 S_1 时犯第一种错误的概率最多是0.05．同样道理，(5.4.2)式表示的意思是：用检验方法 S_2 时犯第一种错误的概率最多是0.01．一句话，两个判别区域 S_1、S_2 形式相同，都在 $\overline{x}$ 太大时拒绝 H，只是犯第一种错误的概率不同．在检验 H：$\mu \leqslant 1500$ 时究竟使用哪一个检验方法，这取决于允许犯第一种错误的概率可以达到的数值，这个数值叫做**信度**，记为 α．在打算采用抽验一部分灯管寿命的方法检验 H：$\mu \leqslant 1500$ 时，一要确定样本的容量（本例是25），二要明确信度 α 的值．通常使用的 α 是0.05，也就是使用检验方法 S_1．如果觉得使用 $\alpha = 0.05$ 的检验方法做出的拒绝 H 的判断不够慎重，通常取 $\alpha = 0.01$，也就是使用检验方法 S_2．如果觉得使用 $\alpha = 0.05$ 的检验方法已过于慎重，新工艺难于被接受，可取 $\alpha = 0.10$．

以上我们结合一个例子讲了怎样从实际需要中提出假设检验问题，什么叫对一个统计假设做检验，什么叫检验方法，什么叫判别区域，什么叫信度等等．最后我们还是结合这个例子归纳出确定一个检验方法，也就是确定一个判别区域的步骤：

第一，从对假设检验问题的分析得出判别区域的形式．本例的假设检验问题是

$$H:\mu\leqslant 1500,\ K:\mu>1500.$$

通过分析知道，样本平均值$\overline{x}$太大时应当拒绝H，所以判别区域取(5.4.3)的形式，其中常数c是待定的．

第二，由给定的信度α确定(5.4.3)式中的常数c．

为了帮助读者理解怎样得出判别区域的形式，下面再看一个例子．某工厂一台自动车床加工一种标准件，其长度$\xi\sim N(\mu,\sigma^2)$，σ^2是未知参数，它的大小能说明加工的精度．根据以往测试的数据知$\sigma^2\geqslant 0.20$．在改进工艺后可以把了解加工精度是否有改进作为下面的假设检验问题来处理：

$$H:\sigma^2\geqslant 0.20,\ K:\sigma^2<0.20.$$

取ξ的样本$x_1,x_2,\cdots,x_n$，由于σ^2的大小可以由它的无偏估计量$\frac{n}{n-1}S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$的大小来反映，在$H$真时，$\frac{n}{n-1}S_n$取比较大的值是自然的，如果$\frac{n}{n-1}S_n^2$取的值比较小，我们就有理由怀疑$H$是真的因而拒绝$H$，所以判别区域的形式是

$$S=\left\{(x_1,\cdots,x_n):\frac{n}{n-1}S_n^2<c\right\}.$$

至于怎样确定常数c将在后面讨论．

下面将介绍几个常用的检验方法．为了使读者对这些方法有一定程度的理解，我们将按照上面的步骤介绍怎样确定这些检验方法的判别区域．

1. U检验法

随机变量ξ服从正态分布$N(\mu,\sigma_0^2)$，$\sigma_0>0$是已知数，μ是未知参数．考虑下面的假设检验问题：

$$H:\mu=\mu_0,\quad K:\mu\neq\mu_0,$$

μ_0是指定的值．当H真时，ξ的样本$x_1,x_2,\cdots,x_n$的平均值

$$\bar{x} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right) ,$$

$\bar{x}$ 取的值接近于 μ_0 是自然的，$\bar{x}$ 取的值比 μ_0 大很多或小很多的可能性比较小，所以判别区域应当具有下面的形式：

(5.4.6) $$S=\{(x_1, \cdots, x_n): |\bar{x}-\mu_0|>c\},$$

c 是待定的常数，由信度 α 来确定．如果 $\alpha=0.05$，那么 c 就是这样一个常数，当 H 真时，也就是 $\mu=\mu_0$ 时，

$$P(|\bar{x}-\mu_0|>c)=0.05$$

或

$$P\left(\left|\frac{\bar{x}-\mu_0}{\sqrt{\sigma_0^2/n}}\right|>\frac{c}{\sqrt{\sigma_0^2/n}}\right)=0.05 .$$

但 H 真时，

$$\frac{\bar{x}-\mu_0}{\sqrt{\sigma_0^2/n}} \sim N(0,1) ,$$

查标准正态分布表

$$\frac{c}{\sqrt{\sigma_0^2/n}}=1.96 ,$$

判别区域就是

$$S=\left\{(x_1, \cdots, x_n): |\bar{x}-\mu_0|>1.96\sqrt{\frac{\sigma_0^2}{n}}\right\} .$$

对于一般的 α，常数 c 由下式确定

$$P\left(\left|\frac{\bar{x}-\mu_0}{\sqrt{\sigma_0^2/n}}\right|>\frac{c}{\sqrt{\sigma_0^2/n}}\right)=\alpha .$$

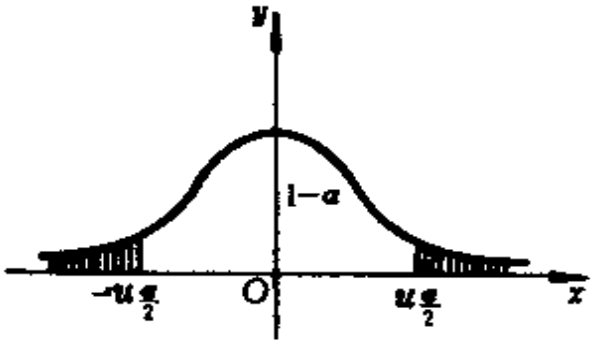


图 5-3

从正态分布表可以查到数 $u_{\frac{\alpha}{2}}$ 使得

$$\int_{-u_{\frac{\alpha}{2}}}^{u_{\frac{\alpha}{2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\mathrm{d}x = 1-\alpha$$

(参见图 5－3)，所以 $\frac{c}{\sqrt{\sigma_0^2/n}} = u_{\frac{\alpha}{2}}$，判别区域是

$$S=\left\{ (x_1, \cdots, x_n) : |\overline{x}-\mu_0| > u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}} \right\},$$

检验 $H: \mu=\mu_0$ 的方法是：

若 $|\overline{x}-\mu_0| > u_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}$，就拒绝 H；

若 $|\overline{x}-\mu_0| \leqslant u_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}$，就接受 H.

称这个检验正态分布 $N(\mu, \sigma_0^2)$ 中参数 μ 是否是 μ_0 的方法为 U 检验法.

例 5.4.7 根据对过去几年考试成绩的统计，某校毕业班语文毕业考试成域的分布检近于正态分布 $N(78.5, 7.6^2)$. 今年毕业班 40 名学生的成绩 (语文)如下：

81	77	70	65	79	77	71	64	79	74
73	65	79	76	74	60	80	86	88	81
85	86	78	83	86	79	83	88	77	83
68	74	85	67	74	83	66	74	73	66

平均分数是 76.4. 有人说这一届学生的语文水平和历届学生比较是不相上下的，对这个说法能接受吗 (信度 α 取 0.05)?

解 可以把这一届学生的语文成绩看成是一个随机变量 ξ，

服从正态分布 $N(\mu, 7.6^2)$，40 名学生的考试成绩 $x_1, \cdots, x_{40}$ 是 ξ 的一个样本，这届学生的语文水平反映在 μ 的大小上，判断这届学生的语文水平是否和历届的学生语文水平不相上下可以作为如下的一个假设检验问题来处理．

$$H: \mu = 78.5 \ , \ K: \mu \neq 78.5 \ .$$

用 U 检验法，本例

$$\overline{x} = 76.4 \ , \ \mu_0 = 78.5 \ , \ \sigma_0 = 7.6 \ , \ n = 40 \ , \ \alpha = 0.05 \ , \ u_{\frac{\alpha}{2}} = 1.96 \ ,$$

$$u_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} = 1.96 \times \frac{7.6}{\sqrt{40}} = 2.35 \ ,$$

$$|\overline{x} - \mu_0| = |76.4 - 78.5| = 2.1 \ .$$

所以 $|\overline{x} - \mu_0| < u_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}$，应当接受 H，认为这届学生和历届学生的语文水平不相上下．

2．t **检验法**（检验正态分布的数学期望是不是指定的数）

随机变量 ξ **服从正态分布** $N(\mu, \sigma^2)$，μ 和 σ^2 都是未知参数，考虑下面的假设检验问题：

$$H: \mu = \mu_0 \ , \ K: \mu \neq \mu_0 \ ,$$

μ_0 是指定的数．

初看起来，这里考虑的问题与 U 检验法考虑的问题似乎是一样的，其实是有差别的，差别就是这里的 σ^2 是未知的．如果我们仍用判别区域 (5.4.6)

$$\left\{ (x_1, \cdots, x_n) : \left| \frac{\overline{x} - \mu_0}{\sqrt{\sigma^2/n}} \right| > \frac{c}{\sigma/\sqrt{n}} \right\} ,$$

在给定信度 α 后，虽然能确定 $\frac{c}{\sigma/\sqrt{n}}$，但不能确定 c．下面我们用 σ^2 的无偏估计量

$$\frac{n}{n-1}S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

代替 $\frac{\overline{x}-\mu_0}{\sqrt{\sigma^2/n}}$ 中的 σ^2 得

(5.4.8)
$$\frac{\overline{x}-\mu_0}{\sqrt{S_n^2/(n-1)}}=\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}\quad,$$

判别区域取

(5.4.9)
$$S=\left\{(x_1,\cdots,x_n):\left|\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}\right|>c\right\},$$

在给定信度 α 后，为了能求出 c，必须知道 $H:\mu=\mu_0$ 真时 $\frac{\overline{x}-\mu}{S_n}\sqrt{n-1}$ 的分布，这需要下面的定理及推论.

定理 5.4.10 设 $x_1, x_2, \cdots, x_n$ 是独立随机变量，$x_i\sim N(0,1)$，$i=1,2,\cdots,n$，记

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad S_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\quad,$$

则 $\overline{x}\sim N\left(0,\frac{1}{n}\right)$，$nS_n^2\sim\chi^2(n-1)$，并且 $\overline{x}$ 与 nS_n^2 独立.

证 $nS_n^2=\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i=1}^{n}x_i^2-n\overline{x}^2$

$$=\sum_{i=1}^{n}x_i^2-\left[\frac{1}{\sqrt{n}}(x_1+\cdots+x_n)\right]^2.$$

令

$$y_1=\frac{1}{\sqrt{2}}x_1-\frac{1}{\sqrt{2}}x_2\quad,$$

$$y_2=\frac{1}{\sqrt{6}}x_1+\frac{1}{\sqrt{6}}x_2-\frac{2}{\sqrt{6}}x_3\quad,$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$y_{n-1}=\frac{1}{\sqrt{n(n-1)}}x_1+\cdots$$

$$+\frac{1}{\sqrt{n(n-1)}}x_{n-1}-\frac{n-1}{\sqrt{n(n-1)}}x_n\quad,$$

$$y_n=\frac{1}{\sqrt{n}}x_1+\frac{1}{\sqrt{n}}x_2+\cdots+\frac{1}{\sqrt{n}}x_n\quad.$$

这个线性变换是正交变换，所以有 $\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}y_i^2$ ，

$$nS_n^2=\sum_{i=1}^{n}y_i^2-y_n^2=\sum_{i=1}^{n-1}y_i^2\quad.$$

又由定理 4.4.19 知独立标准正态随机变量 $x_1, x_2, \cdots, x_n$ 经过正交变换得到的 $y_1, y_2, \cdots, y_n$ 仍是独立的标准正态随机变量，nS_n^2 恰好是 $y_1, \cdots, y_{n-1}$ 的平方和，所以

$$nS_n^2\sim\chi^2(n-1)\,.$$

$\overline{x}\sim N\left(0,\frac{1}{n}\right)$ 是我们熟悉的结论．现在

$$\overline{x}=\frac{1}{n}(x_1+\cdots+x_n)$$

$$=\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}x_1+\cdots+\frac{1}{\sqrt{n}}x_n\right)=\frac{1}{\sqrt{n}}y_n$$

是 y_n 的函数，nS_n^2 是 $(y_1, \cdots, y_{n-1})$ 的函数．由定理 4.1.16，4.1.17 知 $\overline{x}$ 与 nS_n^2 独立．

推论 5.4.11 设 $x_1, x_2, \cdots, x_n$ 是独立随机变量，$x_i \sim N(\mu, \sigma^2)$，$i=1,2,\cdots,n$，则

$$\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \qquad \frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$$

并且 $\overline{x}$ 与 $\frac{nS_n^2}{\sigma^2}$ 独立．

证 令 $z_i = \frac{x_i-\mu}{\sigma}$，$i=1,2,\cdots,n$，则 $z_1, z_2, \cdots, z_n$ 独立，$z_i \sim N(0,1)$，$i=1,2,\cdots,n$．由定理 5.4.10 知

$$\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \sim N\left(0, \frac{1}{n}\right),$$

$$\sum_{i=1}^{n} (z_i-\overline{z})^2 \sim \chi^2(n-1).$$

并且 $\overline{z}$ 与 $\sum_{i=1}^{n} (z_i-\overline{z})^2$ 独立．但 $\overline{z} = \frac{\overline{x}-\mu}{\sigma}$，

$$\sum_{i=1}^{n} (z_i-\overline{z})^2 = \sum_{i=1}^{n} \left(\frac{x_i-\mu}{\sigma} - \frac{\overline{x}-\mu}{\sigma}\right)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n} (x_i-\overline{x})^2 = \frac{nS_n^2}{\sigma^2},$$

所以 $\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$ 与 $\overline{z}$ 独立，因而还与 $\overline{z}$ 的函数 $\overline{x} = \sigma\overline{z} + \mu$ 独立，$\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 是熟知的结论．

现在回来讨论统计量(5.4.8)在 $H:\mu=\mu_0$ 真时的分布．由推论 5.4.11 知，这时

$$\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$$

并且二者独立，统计量 (5.4.8)

$$\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}=\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\Big/\sqrt{\frac{nS_n^2}{\sigma^2}/(n-1)},$$

分子是一个标准正态随机变量，分母是一个 χ^2 变量除以它的自由度 $n-1$ 后的算术平方根而且分子、分母两个随机变量是独立的，由例 4.4.28 及定义 4.4.30 知，$\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}$ 服从自由度是 $n-1$ 的 t 分布．

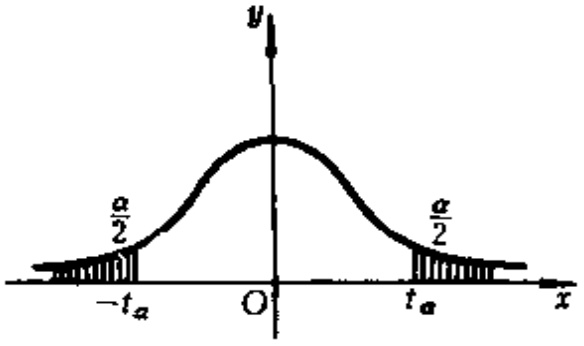


图 5-4

在得知统计量 (5.4.8)在 $H:\mu=\mu_0$ 真时服从自由度是 $n-1$ 的 t 分布后判别区域 (5.4.9)中的常数 c 就容易确定了．例如 $n=10$，$\alpha=0.05$ 时，c 是这样一个数，当 $H:\mu=\mu_0$ 真时，

$$P\left(\left|\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}\right|>c\right)=0.05.$$

查附表4 t分布表知 $c=2.26$.一般地,对给定的信度 α，常数 c 是这样一个数，当 $H:\mu=\mu_0$ 真时，

$$P\left(\left|\frac{\overline{x}-\mu_0}{S_n}\sqrt{n-1}\right|>c\right)=\alpha\ .$$

查 t 分布表，c 等于自由度是 $n-1$ 的 t_α (见图 5－4). 这样，当信度 α 给定后，判别区域就是

$$(5.4.12)\qquad S=\left\{(x_1,\cdots,x_n):|\overline{x}-\mu_0|>t_\alpha\frac{S_n}{\sqrt{n-1}}\right\}\ ,$$

检验 H 的方法就是

若 $|\overline{x}-\mu_0|>t_\alpha\dfrac{S_n}{\sqrt{n-1}}$ 就拒绝 H；

若 $|\overline{x}-\mu_0|\leqslant t_\alpha\dfrac{S_n}{\sqrt{n-1}}$ 就接受 H.

例 5.4.13 在例5.4.7 中，如果知道历届学生语文成绩的分布近似于正态分布 $N(78.5,\sigma^2)$,本届学生语文成绩的分布近似于正态分布 $N(\mu,\sigma^2)$，但不知道 σ^2 的大小，这时可算出 $S_{40}=\sqrt{\dfrac{1}{40}\sum_{i=1}^{40}(x_i-\overline{x})^2}$,然后用 t 检验法对假设检验问题 $H:\mu=78.5$，$K:\mu\neq78.5$ 做检验 .经过计算 $S_{40}=7.30$. 在 t 分布表中没有自由度是 39 的 t_α，可用自由度是 40 的 t_α，$t_\alpha=2.02$.由于

$$|\overline{x}-\mu_0|=|76.4-78.5|=2.1\ ,$$

$$t_\alpha\frac{S_{40}}{\sqrt{40-1}}=2.02\times\frac{7.30}{\sqrt{39}}=2.36\ ,$$

$$|\overline{x}-\mu_0|<t_\alpha\frac{S_n}{\sqrt{n-1}}\ ,$$

应当接受 H，认为这届学生和历届学生的语文水平不相上下.

3. t 检验法 (检验两个正态分布的数学期望是否相等)

设随机变量 $\xi \sim N(\mu, \sigma^2)$，$\eta \sim N(v, \sigma^2)$，$\mu$，$v$，$\sigma^2$ 是未知参数. $x_1, x_2, \cdots, x_m$ 是 ξ 的样本，$y_1, y_2, \cdots, y_n$ 是 η 的样本，$x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n$ 独立. 考虑下面的假设检验问题：

$$H: \mu = v, \quad K: \mu \neq v.$$

记

$$\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \quad , \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_j \ ,$$

$$S_x^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \overline{x})^2 \ , \quad S_y^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \overline{y})^2 \ ,$$

则

$$\overline{x} \sim N\left(\mu, \frac{1}{m}\sigma^2\right), \quad \overline{y} \sim N\left(v, \frac{1}{n}\sigma^2\right),$$

$$\frac{mS_x^2}{\sigma^2} \sim \chi^2(m-1), \qquad \frac{nS_y^2}{\sigma^2} \sim \chi^2(n-1).$$

由于 $\overline{x}$，$\overline{y}$ 独立，

$$\overline{x} - \overline{y} \sim N\left(\mu - v, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right),$$

当 $H: \mu = v$ 真时，

$$\overline{x} - \overline{y} \sim N\left(0, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right),$$

$$\frac{\overline{x} - \overline{y}}{\sqrt{\sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim N(0,1). \tag{5.4.14}$$

如果 σ^2 是未知的，我们可以按照 U 检验法那样确定判别区域. 现在 σ^2 是未知的，我们用 σ^2 的无偏估计量

$$\frac{mS_x^2+nS_y^2}{m+n-2}$$

代替 (5.4.14)中的 σ^2 得到统计量

$$\frac{\overline{x}-\overline{y}}{\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}, \tag{5.4.15}$$

判别区域取

$$S=\left\{(x_1,\cdots,x_m;y_1,\cdots,y_n):\ \frac{|\overline{x}-\overline{y}|}{\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}>c\right\}. \tag{5.4.16}$$

为了能在给定信度 α 后确定常数 c ,需要知道 $H:\mu=v$ 真时统计量 (5.4.15)的分布 .

引理 5.4.17 设 $x_1,x_2,\cdots,x_m;y_1,y_2,\cdots,y_n$ 是独立随机变量 ,$x_i\sim N(\mu,\sigma^2)$, $y_j\sim N(v,\sigma^2)$, $i=1,2,\cdots,m;j=1,2,\cdots,n$. 记

$$\overline{x}=\frac{1}{m}\sum_{i=1}^{m}x_i,\quad \overline{y}=\frac{1}{n}\sum_{j=1}^{n}y_j,$$

$$S_x^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\overline{x})^2,\quad S_y^2=\frac{1}{n}\sum_{j=1}^{n}(y_j-\overline{y})^2,$$

则 $\overline{x}$, $\overline{y}$, S_x^2 , S_y^2 是独立随机变量 .

证明留给读者 .

定理 5.4.18 设 $x_1,x_2,\cdots,x_m;y_1,y_2,\cdots,y_n$ 是独立随机变量 ,$x_i\sim N(\mu,\sigma^2)$, $y_j\sim N(v,\sigma^2)$, $i=1,2,\cdots,m$; $j=1,2,\cdots,n$,则

$$(5.4.19)\quad \frac{(\bar{x}-\bar{y})-(\mu-v)}{\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim t(m+n-2).$$

证 由引理 5.4.17 知$\bar{x}$，$\bar{y}$，S_x^2，S_y^2 独立，又由推论 5.4.11 知

$$\bar{x} \sim N\left(\mu, \frac{1}{m}\sigma^2\right), \quad \bar{y} \sim N\left(v, \frac{1}{n}\sigma^2\right),$$

$$\frac{mS_x^2}{\sigma^2} \sim \chi^2(m-1), \qquad \frac{nS_y^2}{\sigma^2} \sim \chi^2(n-1),$$

所以

$$\frac{(\bar{x}-\bar{y})-(\mu-v)}{\sqrt{\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim N(0,1),$$

$$\frac{mS_x^2+nS_y^2}{\sigma^2} \sim \chi^2(m+n-2),$$

并且这两个随机变量是独立的，所以

$$\frac{(\bar{x}-\bar{y})-(\mu-v)}{\sqrt{\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)}} \Bigg/ \sqrt{\frac{mS_x^2+nS_y^2}{\sigma^2(m+n-2)}} \sim t(m+n-2).$$

(5.4.19)式获证.

现在回到统计量 (5.4.15)的分布问题上来，当 $H:\mu=v$ 真时，统计量 (5.4.15)就是统计量 (5.4.19)，所以服从自由度是 $m+n-2$ 的 t 分布，由于它取绝对值太大的值可能性较小，所以

判别区域取 (5.4.16)的形式．给定信度 α 后，常数 c 是自由度为 $m+n-2$ 的 t_α，这个判别区域确定的检验方法是：

若 $|\overline{x}-\overline{y}|>t_\alpha\sqrt{\dfrac{mS_x^2+nS_y^2}{m+n-2}\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}$ 就拒绝 H，否则接受 H．

例 5.4.20 某地某年高考后随机抽得15 名男生、12 名女生的物理考试成绩如下：

男生	49	48	47	53	51	43	39	57	56	46
	42	44	55	44	40					
女生	46	40	47	51	43	36	43	38	48	54
	48	34								

从这 27 名学生的成绩能说明这个地区男女生的物理水平不相上下吗 (信度取 0.05)？

解 把这个地区男生物理考试成绩看作是服从正态分布 $N(\mu,\sigma^2)$的随机变量 ξ，女生物理考试成绩看作是服从正态分布 $N(v,\sigma^2)$的随机变量 η，15 名男生的成绩 $x_1,\cdots,x_{15}$ 就是 ξ 的一个样本，12 名女生的成绩 $y_1,\cdots,y_{12}$ 就是 η 的一个样本．本例的问题可以作为下面的假设检验问题来处理：

$$H:\mu=v\ ,\qquad K:\mu\neq v\ .$$

用 t 检验法．

$$m=15\ ,\ n=12\ ,\ m+n-2=25\ ,$$

$$\overline{x}=47.6\ ,\ \overline{y}=44\ ,\ |\overline{x}-\overline{y}|=3.6\ ,$$

$$15S_x^2=\sum_{i=1}^{15}(x_i-\overline{x})^2=469.6\ ,\ 12S_y^2=\sum_{j=1}^{12}(y_j-\overline{y})^2=412\ ,$$

查 t 分布表，自由度是 25 的 $t_{0.05}=2.06$，

$$t_{0.05}\sqrt{\frac{15S_x^2+12S_y^2}{25}\left(\frac{1}{15}+\frac{1}{12}\right)}\doteq 2.06\times 2.30$$

$$\doteq 4.74>|\overline{x}-\overline{y}|\ ,$$

所以应当接受 H，认为这个地区男女生物理水平不相上下．

4. F 接验法 (检验两个正态分布的方差是否相等)

设随机变量 $\xi \sim N(\mu, \sigma_1^2)$，$\eta \sim N(\nu, \sigma_2^2)$，$\mu$，$\nu$，$\sigma_1^2$，$\sigma_2^2$ 是未知参数，$x_1, x_2, \cdots, x_m$ 是 ξ 的样本，$y_1, y_2, \cdots, y_n$ 是 η 的样本，考虑假设检验问题：

$$H: \sigma_1^2 = \sigma_2^2, \quad K: \sigma_1^2 \neq \sigma_2^2.$$

σ_1^2，σ_2^2 的大小一般可以由它们的无偏估计

$$\hat{\sigma}_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \overline{x})^2 = \frac{mS_x^2}{m-1},$$

$$\hat{\sigma}_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2 = \frac{nS_y^2}{n-1}$$

来反映，所以 H 真时，一般 $\hat{\sigma}_1^2$ 与 $\hat{\sigma}_2^2$ 相差不多，$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ 太大或太小时应当拒绝 H，判别区域取

$$S = \left\{(x_1, \cdots, x_m; y_1, \cdots, y_n): \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < c_1 \text{ 或 } \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > c_2\right\}.$$

把 $x_1, \cdots, x_m; y_1, \cdots, y_n$ 看作独立的随机变量，由推论 5.4.11 及引理 5.4.17 知

$$\frac{mS_x^2}{\sigma_1^2} = \frac{\sum_{i=1}^{m}(x_i - \overline{x})^2}{\sigma_1^2} \sim \chi^2(m-1),$$

$$\frac{nS_y^2}{\sigma_2^2} = \frac{\sum_{j=1}^{n}(y_j - \overline{y})^2}{\sigma_2^2} \sim \chi^2(n-1)$$

并且这两个随机变量是独立的．当 H 真时，$\sigma_1^2 = \sigma_2^2$，

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}=\frac{\dfrac{mS_x^2}{\sigma_1^2}\Big/(m-1)}{\dfrac{nS_y^2}{\sigma_2^2}\Big/(n-1)}$$

在给定信度 α 后为了确定 c_1，c_2 使

$$P\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}<c_1 \text{ 或 } \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}>c_2\right)=\alpha \tag{5.4.21}$$

必须知道 $\hat{\sigma}_1^2/\hat{\sigma}_2^2$ 的概率分布．在 H 真时，$\hat{\sigma}_1^2/\hat{\sigma}_2^2$ 是两个独立的 χ^2 变量分别除以其自由度后的比，由例 4.4.32 及定义 4.4.34 知，$\hat{\sigma}_1^2/\hat{\sigma}_2^2$ 服从第一自由度是 $m-1$、第二自由度是 $n-1$ 的 F 分布．

现在可以确定 (5.4.21)式中的常数 c_1，c_2 了．习惯上取 c_1，c_2 使

$$P\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}<c_1\right)=\frac{\alpha}{2},$$

$$P\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}>c_2\right)=\frac{\alpha}{2}.$$

参见图 5－5，也就是取

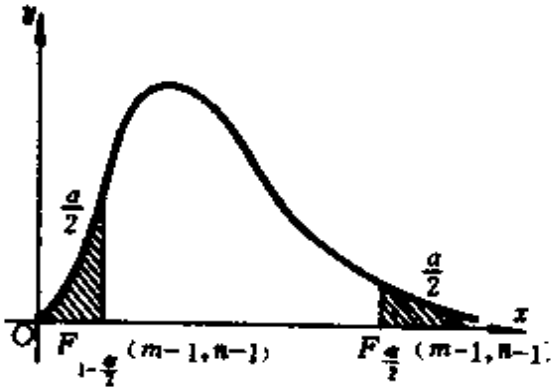


图 5－5

$$c_1 = F_{1-\frac{\alpha}{2}}(m-1, n-1), \quad c_2 = F_{\frac{\alpha}{2}}(m-1, n-1).$$

这样，判别区域就是

$$S = \left\{ (x_1, \cdots, x_m; y_1, \cdots, y_n): \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < F_{1-\frac{\alpha}{2}}(m-1, n-1) \right.$$

$$\left. \text{或} \ \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > F_{\frac{\alpha}{2}}(m-1, n-1) \right\},$$

对应的检验方法是：

$$\text{若} \quad \frac{(n-1)\sum_{i=1}^{m}(x_i-\overline{x})^2}{(m-1)\sum_{j=1}^{n}(y_j-\overline{y})^2} < F_{1-\frac{\alpha}{2}}(m-1, n-1) \ \text{或}$$

$$\frac{(n-1)\sum_{i=1}^{m}(x_i-\overline{x})^2}{(m-1)\sum_{j=1}^{n}(y_j-\overline{y})^2} > F_{\frac{\alpha}{2}}(m-1, n-1),$$

就拒绝 H，否则接受 H.

从附表 5 F 分布表中可以查到数值 $F_{0.25}(m, n)$，$F_{0.05}(m, n)$，$F_{0.01}(m, n)$，但 $F_{0.75}(m, n)$，$F_{0.95}(m, n)$，$F_{0.99}(m, n)$查不到，这时可以利用公式 (4.4.38)得到

$$F_{0.95}(m, n) = 1/F_{0.05}(n, m), \quad F_{0.99}(m, n) = 1/F_{0.01}(n, m)$$

来计算.

例 5.4.22 在例5.4.20 中我们曾假设 ξ 和 η 的分布有相同的方差 σ^2. 如果对这一点有怀疑，例如，会不会是 $\xi \sim N(\mu, \sigma_1^2)$，$\eta \sim N(\mu, \sigma_2^2)$，但 $\sigma_1^2 \neq \sigma_2^2$ 呢？这个问题可以作为下面的假设检

验问题来处理：

$$H:\sigma_1^2=\sigma_2^2\ ,\quad K:\sigma_1^2\neq\sigma_2^2\ .$$

下面根据例 5.4.20 中的样本，用F 检验法检验 H．本例 $m=15$，$n=12$，

$$\sum_{i=1}^{m}(x_i-\overline{x})^2=469.6\ ,\ \sum_{j=1}^{n}(y_j-\overline{y})^2=412\ .$$

如信度 α 取 0.10，则查 F 分布表可得

$$F_{\frac{\alpha}{2}}(m-1,n-1)=F_{0.05}(14,11)>2.72\ ,$$

$$F_{1-\frac{\alpha}{2}}(m-1,n-1)=F_{0.95}(14,11)$$

$$=\frac{1}{F_{0.05}(11,14)}<\frac{1}{2.53}\doteq 0.39\ .$$

而

$$\frac{(n-1)\sum_{i=1}^{m}(x_i-\overline{x})^2}{(m-1)\sum_{j=1}^{n}(y_j-\overline{y})^2}=\frac{11\times 469.6}{14\times 412}\doteq 0.90\ ,$$

所以

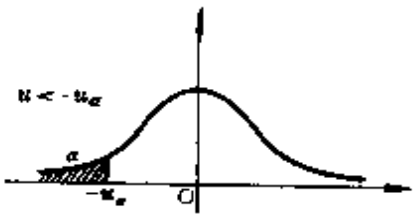
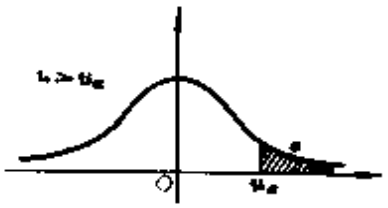
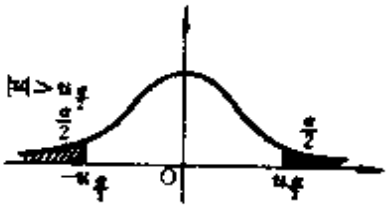
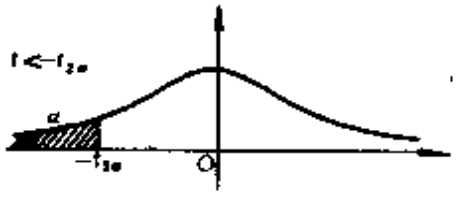
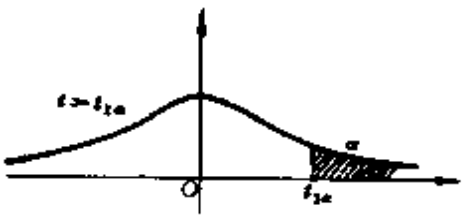
$$F_{1-\frac{\alpha}{2}}(m-1,n-1)<\frac{(n-1)\sum_{i=1}^{m}(x_i-\overline{x})^2}{(m-1)\sum_{j=1}^{n}(y_j-\overline{y})^2}$$

$$<F_{\frac{\alpha}{2}}(m-1,n-1)\ ,$$

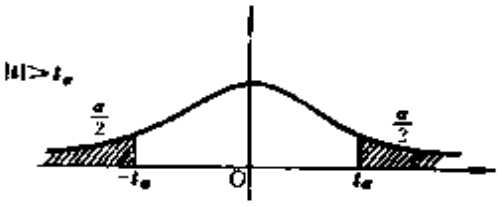
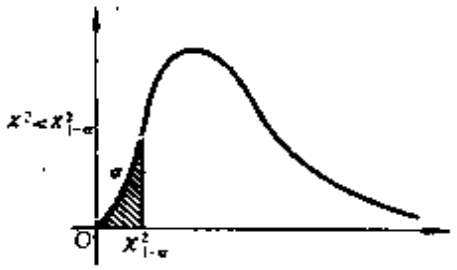
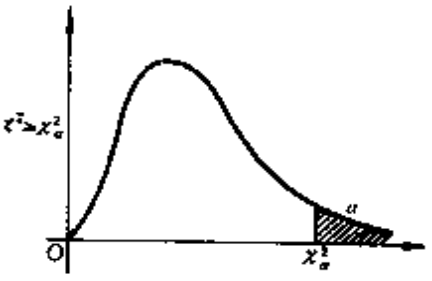
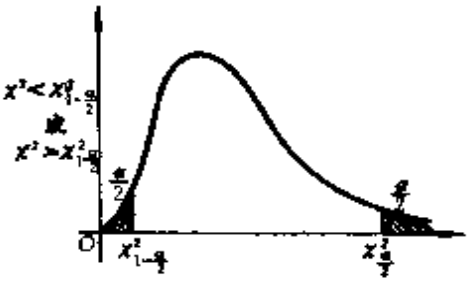
应当接受 H，认为 σ_1^2 和 σ_2^2 是相同的．

和正态分布的参数有关的假设检验问题还可以提出很多，这里就不一一地叙述了，下面列出一张表(表5－1)供读者使用．

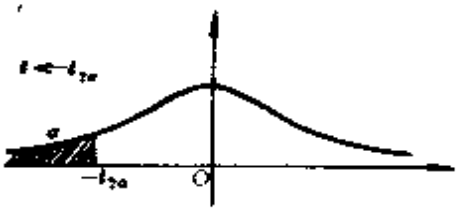
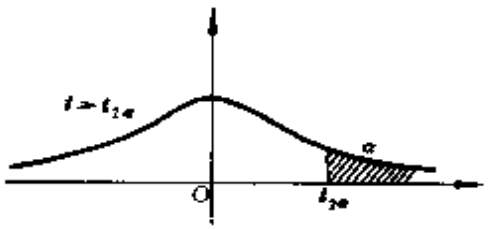
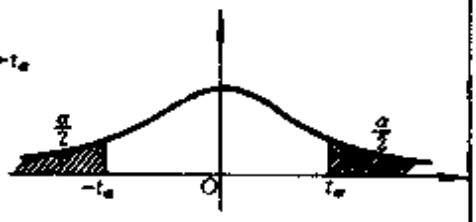
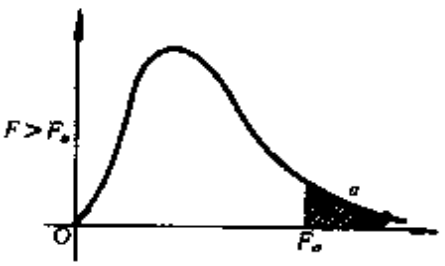
表 5－1

已知条件	检验问题 H　K	统计量	判别区域	统计表
$\xi\sim N(E\xi,\sigma^2)$，σ^2 已知，a 是指定的数，$x_1,x_2,\cdots,x_n$ 是 ξ 的样本	$E\xi\geqslant a$，$E\xi<a$	$u=\dfrac{\overline{x}-a}{\sigma/\sqrt{n}}$，其中 $\overline{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$		正态分布表
	$E\xi\leqslant a$，$E\xi>a$			
	$E\xi=a$，$E\xi\neq a$			
$\xi\sim N(E\xi,\sigma^2)$，σ^2 未知，a 是指定的数，$x_1,x_2,\cdots,x_n$ 是 ξ 的样本	$E\xi\geqslant a$，$E\xi<a$	$t=\dfrac{\overline{x}-a}{s/\sqrt{n-1}}$，其中 $s=\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$		自由度为 $n-1$ 的 t 分布表
	$E\xi\leqslant a$，$E\xi>a$			

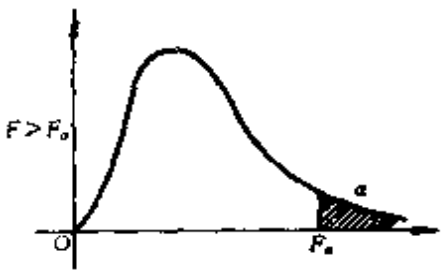
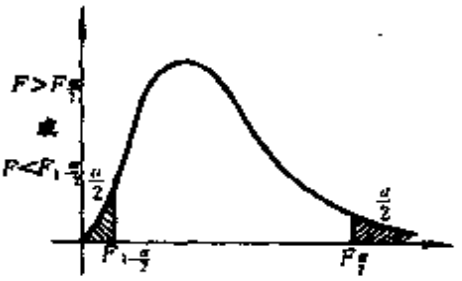
续表

已知条件	检验问题 H K	统计量	判别区域	统计表
	$E\xi=a$， $E\xi\neq a$		$\|t\|>t_\alpha$ 	
$\xi\sim N(E\xi, D\xi)$，σ_0^2 是指定的数，x_1，…，x_n 是 ξ 的样本	$D\xi\geqslant\sigma_0^2$， $D\xi<\sigma_0$	$\chi^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sigma_0^2}$	$\chi^2<\chi^2_{1-\alpha}$ 	自由度为 $n-1$ 的 χ^2 分布表
	$D\xi\leqslant\sigma_0^2$， $D\xi>\sigma_0$		$\chi^2>\chi^2_\alpha$ 	
	$D\xi=\sigma_0^2$， $D\xi\neq\sigma_0^2$		$\chi^2<\chi^2_{1-\frac{\alpha}{2}}$ 或 $\chi^2>\chi^2_{\frac{\alpha}{2}}$ 	

续表

已知条件	检验问题 H K	统计量	判别区域	统计表
$\xi\sim N(E\xi,\sigma^2),\eta\sim N(E\eta,\sigma^2),\sigma^2$ 未知 $x_1,\cdots,x_m$ 是 ξ 的样本 $y_1,\cdots,y_n$ 是 η 的样本	$E\xi\geqslant E\eta$, $E\xi<E\eta$	$t=\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{mS_x^2+nS_y^2}{m+n-2}}}\cdot\dfrac{1}{\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$, 其中 $mS_x^2=\sum_{i=1}^{m}(x_i-\bar{x})^2$, $nS_y^2=\sum_{i=1}^{n}(y_i-\bar{y})^2$	$t<-t_{2\alpha}$  	自由度为 $m+n-2$ 的 t 分布表
	$E\xi\leqslant E\eta$, $E\xi>E\eta$		$t>t_{2\alpha}$  	
	$E\xi=E\eta$ $E\xi\neq E\eta$		$\lvert t\rvert>t_\alpha$  	
$\xi\sim N(E\xi,D\xi),\eta\sim N(E\eta,D\eta),x_1,\cdots,x_m$ 是 ξ 的样本 $y_1,\cdots,y_n$ 是 η 的样本	$D\xi\geqslant D\eta$, $D\xi<D\eta$	$F=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{m}(x_i-\bar{x})^2}\cdot\dfrac{m-1}{n-1}$	$F>F_\alpha$  	第一自由度为 $n-1$ 第二自由度为 $m-1$ 的 F 分布表

续表

已知条件	检验问题 H K	统计量	判别区域	统计表
	$D\xi \leqslant D\eta$, $D\xi > D\eta$	$F=\dfrac{\sum_{i=1}^{m}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}\cdot\dfrac{n-1}{m-1}$		第一自由度为 $m-1$ 第二自由度为 $n-1$ 的 F 分布表
	$D\xi = D\eta$, $D\xi \neq D\eta$			

习题 5－5

1. 为比较甲、乙两个小麦品种在产量上有无显著差异，现在选择了 20 块条件相同的地块，甲、乙两个品种各种 10 块得到产量如下 (单位：公斤)：

甲　16.2　18.5　22.8　19.6　20.7　20.7　23.3　19.9　20.5　18.8

乙　20.8　22.6　19.8　25.5　19.5　21.8　23.8　19.2　18.7　19.1

假设甲、乙两个品种在小地块上的产量分别服从正态分布 $N(\mu, \sigma^2)$，$N(\nu, \sigma^2)$，试检验统计假设 $H: \mu=\nu$ (信度取 0.05)。

2. 有一种安眠剂，据说，在一定剂量下能比某种旧安眠剂平均增加睡眠 3 小时，已知使用旧安眠剂的睡眠时间 (单位：小时) 服从正态分布 $N(20.8, 1.8^2)$。为了检验对新安眠剂的这种说法是否正确，收集了一组使用新安眠剂的睡眠时间 (单位：小时)：

26.7　22.0　24.1　21.0　27.2　25.0　23.4

试问这组数据能否说明新安眠剂确有新的疗效(信度取0.05)?

3. 一台自动机床加工的零件长度服从正态分布 $N(\mu,\sigma^2)$,加工精度 $\sigma^2\leqslant 0.1$.在工作一段时间后,需要检验这一机床是否还保持这一加工精度,即考虑假设检验问题

$$H:\sigma^2\leqslant 0.1\ ,\quad K:\sigma^2>0.1\ .$$

为此抽取了这台机床加工的20个零件,测得长度如下:

4.81 4.89 4.92 5.04 4.99 4.98 4.94 5.03 5.03 5.16

5.28 5.14 5.13 5.04 4.96 4.94 4.85 5.09 5.09 4.96

试根据这个样本检验 H,信度取0.05.

4. 证明引理5.4.17.

5. 证明(5.4.15)式中 $\dfrac{mS_x^2+nS_y^2}{m+n-2}$ 是 σ^2 的无偏估计量.

§5.5 区间估计

在§5.2中我们讲过怎样对未知参数做点估计.例如随机变量 $\xi\sim N(\mu,1)$,μ 是未知参数,对 μ 做点估计就是选定一个估计量(也就是选定一个估计方法),如 $\overline{x}=\dfrac{1}{5}\sum\limits_{i=1}^{5}x_i$,然后取得样本观测值,如

(5.5.1) $x_1=1.9$,$x_2=1.8$,$x_3=2.1$,$x_4=1.8$,$x_5=1.7$,

算出 μ 的估计值

$$\hat{\mu}=(1.9+1.8+2.1+1.8+1.7)\times\frac{1}{5}=1.86\ .$$

请读者注意,这里是用随机变量 $\overline{x}$ 随机取得的一个值1.86作为 μ 的估计值.我们这样做是因为 $\overline{x}$ 有一些好的性质,例如,$E\overline{x}=\mu$,在 μ 的一切线性无偏估计中,$\overline{x}$ 的方差最小等等,这些意味着 $\overline{x}$ 在 μ 的附近取值的概率比较大.所以,用 $\overline{x}$ 随机取得的一个值做为 μ 的估计值,估计得比较准确的概率也就比较大.然而,在使用点估计时,有时也会感到有不十分令人满意的地方.例如,只知道 $\overline{x}$ 在 μ 的附近取值的概率比较大,但"附近"和"比

较大"的含意是不清楚的．下面将要讲的区间估计是给 μ 一个估计区间（不象点估计那样给 μ 一个估计值），当使用一个区间估计方法时，能知道估计成功的概率有多大．

以下用 $X_1, X_2, \cdots, X_n$ 表示 ξ 的样本，用 $x_1, x_2, \cdots, x_n$ 表示样本的观测值．

上例中，ξ 的容量为 5 的样本平均值

$$\overline{X} \sim N\left(\mu, \frac{1}{5}\right),$$

从这里可以得到

$$P\left(\overline{X}-1.96\times\frac{1}{\sqrt{5}}<\mu<\overline{X}+1.96\times\frac{1}{\sqrt{5}}\right)=0.95,$$

或

$$P(\overline{X}-0.877<\mu<\overline{X}+0.877)=0.95 \quad (5.5.2)$$

$(\overline{X}-0.877, \overline{X}+0.877)$是一个区间，区间的端点是随机变量，所以是随机区间．在得到样本观测值后，$\overline{X}$ 就取定某一个值$\overline{x}$，$(\overline{x}-0.877, \overline{x}+0.877)$就是一个普通的区间，它可能套住了 μ（如图 5－6 (a)），也可能没有套住 μ（如图 5－6 (b)）．

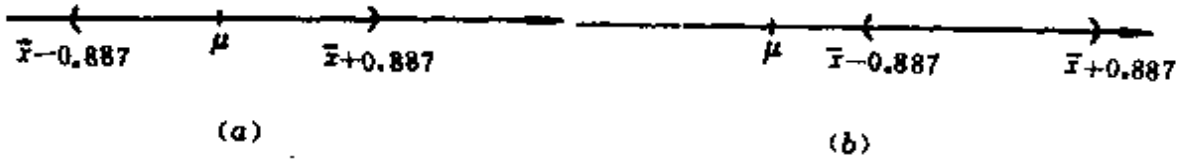


图 5－6

(5.5.2)式表明：随机区间 $(\overline{X}-0.877, \overline{X}+0.877)$套住 μ 的概率是 0.95．如果我们估计 μ 就在区间 $(\overline{X}-0.877, \overline{X}+0.877)$内，也就是在得到样本观测值 x_1, x_2, x_3, x_4, x_5 时，估计 μ 就在区间 $(\overline{x}-0.877, \overline{x}+0.877)$内，例如得到样本 (5.5.1)时，估计 μ 就在 $(1.86-0.877, 1.86+0.877)$内，则估计成功的概率是 0.95．

象上面所说的，给未知参数 μ 选定一个随机区间 $(\overline{X}-0.877,\ \overline{X}+0.877)$，在得到样本观测值 x_1, x_2, x_3, x_4, x_5 后估计 μ 就在区间 $(\overline{x}-0.877,\ \overline{x}+0.877)$内，就叫对 μ 做区间估计.

一般的，设 $\xi \sim F(x;\theta)$，$\theta \in \Theta$ 是未知的. 所谓对未知参数 $g(\theta)$做区间估计就是选定两个统计量 $f_1(X_1, \cdots, X_n) < f_2(X_1, \cdots, X_n)$，用它们构成一个随机区间 $(f_1(X_1, \cdots, X_n), f_2(X_1, \cdots, X_n))$使

$$P\{f_1(X_1, \cdots, X_n) < g(\theta) < f_2(X_1, \cdots, X_n)\} = \beta,$$

$0<\beta<1$ 是预先给定的数，叫置信系数，当得到样本观测值 $x_1, x_2, \cdots, x_n$ 时就估计 $g(\theta)$在区间

$$(f_1(x_1, \cdots, x_n), f_2(x_1, \cdots, x_n))$$

内.

例 5.5.3 从一批钉子中抽取16枚，测得长度 (单位: cm)为

2.14 2.10 2.13 2.15 2.13 2.12 2.13 2.10
2.15 2.12 2.14 2.10 2.13 2.11 2.14 2.11

设钉长 $\xi \sim N(\mu, 0.01^2)$，试对 μ 做区间估计 (置信系数 β 取 0.95).

解 ξ 的容量为 16 的样本平均值

$$\overline{X} \sim N\left(\mu, \frac{1}{16} \times 0.01^2\right),$$

$$P(\overline{X} - 1.96 \times 0.0025 < \mu < \overline{X} + 1.96 \times 0.0025) = 0.95.$$

取 $f_1(X_1, \cdots, X_{16}) = \overline{X} - 0.0049$，$f_2(X_1, \cdots, X_{16}) = \overline{X} + 0.0049$. 由 16 个样本观测值可以算出

$$\overline{x} = \frac{1}{16} \times (2.14 + 2.10 + \cdots + 2.11) = 2.125,$$

$$f_1(x_1, \cdots, x_{16}) = \overline{x} - 0.0049 = 2.125 - 0.0049 \doteq 2.12,$$

$$f_2(x_1, \cdots, x_{16}) = \overline{x} + 0.0049 = 2.125 + 0.0049 \doteq 2.13,$$

结论：估计 μ 在区间 (2.12, 2.13)内.

当 $\xi \sim N(\mu, \sigma^2)$，$\mu$，$\sigma^2$ 未知时，利用

$$\frac{\overline{X}-\mu}{S_n}\sqrt{n-1} \sim t(n-1),$$

从 t 分布表 (附表 4) 查自由度是 $n-1$ 的 t_α 表可以得到

$$P\left(\overline{X}-t_\alpha\frac{S_n}{\sqrt{n-1}}<\mu<\overline{X}+t_\alpha\frac{S_n}{\sqrt{n-1}}\right)=1-\alpha,$$

$\alpha \in (0,1)$，我们可以用随机区间

$$\left(\overline{X}-t_\alpha\frac{S_n}{\sqrt{n-1}}, \overline{X}+t_\alpha\frac{S_n}{\sqrt{n-1}}\right)$$

对 μ 做区间估计，置信系数是 $1-\alpha$.

例 5.5.4 某糖厂用自动打包机包装糖，每包重量 ξ 服从正态分布 $N(\mu,\sigma^2)$. 某日开工后测得 10 包糖重量 (单位:g) 是

x_1	x_2	x_3	x_4	x_5	x_6	x_7
1001.8	1002.4	999.5	998.7	997.7	1005.4	997.1

x_8	x_9	x_{10}
995.5	1004.1	1003.2

试用这 10 包糖的重量对 μ 做区间估计 (置信系数取 0.95).

解 $\overline{x}=\frac{1}{10}\times(1001.8+1002.4+\cdots+1003.2)$

$$=\frac{1}{10}\sum_{i=1}^{10}x_i=1000.54,$$

$$S_{10}=\sqrt{\frac{1}{10}\sum_{i=1}^{10}(x_i-\overline{x})^2}=3.13.$$

查 t 分布表，自由度是 9 的 $t_{0.05}=2.26$，

$$\overline{x}-t_{0.05}\frac{S_{10}}{\sqrt{10-1}}=1000.54-2.26\times\frac{3.13}{3}$$

$$=1000.54-2.36=998.2\ ,$$

$$\overline{x}+t_{0.05}\frac{S_{10}}{\sqrt{10-1}}=1000.54+2.26\times\frac{3.13}{3}$$

$$=1000.54+2.36=1002.9\ .$$

结论：估计 μ 在 998.2 和 1002.9 克之间.

对两个正态总体的数学期望之差也可以做区间估计. 设 $\xi\sim N(\mu,\sigma^2)$，$\eta\sim N(v,\sigma^2)$，有相同的方差 σ^2，$X_1,\cdots,X_m$ 是 ξ 的样本，$Y_1,\cdots,Y_n$ 是 η 的样本，利用

$$\frac{(\overline{X}-\overline{Y})-(\mu-v)}{\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}\sim t(m+n-2)$$

从 t 分布表查出自由度是 $m+n+2$ 的 t_α，使

$$P\left(\overline{X}-\overline{Y}-t_\alpha\sqrt{\frac{mS_x^2+mS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}<\mu-v<\right.$$

$$\left.\overline{X}-\overline{Y}+t_\alpha\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}\right)=1-\alpha\ .$$

所以可以用随机区间

$$\left(\overline{X}-\overline{Y}-t_\alpha\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}\right.,$$

$$\left.\overline{X}-\overline{Y}+t_\alpha\sqrt{\frac{mS_x^2+nS_y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}\right)$$

对 $\mu-v$ 做区间估计.

习题 5－6

1. 随机抽测某市 100 名 7 岁男孩的血压，收缩压 $x_1, \cdots, x_{100}$ 的平均值 $\overline{x}=96.4$ (单位：毫米汞柱)，二阶中心矩 $\frac{1}{100}\sum_{i=1}^{100}(x_i-\overline{x})^2=66.75$；舒张压 $y_1, \cdots, y_{100}$ 的平均值 $\overline{y}=59.8$，二阶中心矩 $\frac{1}{100}\sum_{i=1}^{100}(y_i-\overline{y})^2=69.85$．设该市 7 岁男孩的收缩压和舒张压分布分别接近于 $N(\mu, \sigma_1^2)$，$N(v, \sigma_2^2)$，试对 μ，v 做区间估计 (置信系数取 0.90)．

2. 将一个班的 30 个学生随机地分成两组，对这两组学生用不同的方法进行某种技能训练，结果每个学生需要的训练时间如下：

第一组

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
10	12	8	7	9	13	9	12
x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	
13	7	6	12	10	8	10	

第二组

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
5	8	11	6	5	10	4	9
y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	
4	6	10	12	4	7	8	

若用第一、二种方法训练学生所需时间的分布分别接近于 $N(\mu, \sigma^2)$，$N(v, \sigma^2)$，试对统计假设 $H: \mu=v$ 做检验，信度取 0.05．若 $\mu \neq v$，试给 $\mu-v$ 做一个区间估计，置信系数取 0.95．

3. 试给例 5.5.4 中参数 σ^2 一个区间估计，置信系数取 0.90．

§5.6 中心极限定理与大样本统计推断

在 § 5.4，5.5 中，我们介绍了有关正态分布参数的假设检验方法和区间估计方法，这是因为在工农业生产和科学试验中许多随机现象是可以用正态随机变量来刻画 (或近似刻画) 的．因而

人们对正态随机变量的研究较多．对正态随机变量的研究有两个重要的内容，一是推导了许多和正态随机变量有关的统计量的分布．一是从理论上探索为什么许多随机变量呈现出正态分布或其分布接近于正态分布．

在例 3.3.10 中曾经看到，一个自动机床生产的零件的口径 ξ 是一个随机变量，它的分布是正态分布或接近于正态分布．让我们分析一下影响其分布的一些因素，一方面是生产者按设计要求调整机床使生产出的零件口径应为 μ（单位：mm），例如 $\mu=11$，另一方面是在生产过程中有许多随机因素使生产的口径偏离 11mm，例如材料不均匀使口径产生偏离 η_1，机床微小的振动使口径产生偏离 η_2，温度的微小变化使口径产生偏离 η_3 等等．这些 η_i 都是随机变量，而且可以认为是独立的随机变量，ξ 正是由 μ 与这些众多独立随机变量决定的，为简单起见，假设看到的是 $\xi=\mu+\eta$，其中 η 是总的偏离，它是许多独立的随机变量之和．人们探索的问题是在什么条件下，η 会呈现正态分布或其分布接近于正态分布，得到的条件是很一般的，其大意是 $\sum_{i=1}^{n}\eta_i$ 的每一项对总和的影响都不大．我们将要介绍的中心极限定理中将用数学的语言确切地叙述这个条件，在此以前读者可以考察一些自己熟悉的例子，那些用正态随机变量刻画的随机现象是不是受许多微小的随机因素独立影响形成的．这种条件的一般性也从一个方面说明许多实际的随机现象可以用正态分布来刻画或近似描述．

设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是随机变量序列，一类关于随机变量序列

$$\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}},\quad n=1,2,\cdots$$

对应的分布函数序列收敛于标准正态分布函数的定理叫中心极限

定理．下面我们只列出其中的一个定理，对定理证明有兴趣的读者可参看[4，第一册，第五章]或[15，第八章，§2 推论 1]．

定理 5.6.1 设ξ_1，ξ_2，…，ξ_n，… 是一个独立随机变量序列，$D\xi_k$ (>0)有限，$k=1,2,\cdots$，记 $B_n=\sqrt{\sum_{k=1}^{n} D\xi_k}$，若对任意的 $\varepsilon>0$ 有

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-E\xi_k|>\varepsilon B_n}(x-E\xi_k)^2\,dF_k(x)=0,\tag{5.6.2}$$

其中 $F_k(x)$是 ξ_k 的分布函数，则对任何实数 x，有

$$\lim_{n\to\infty}P\left(\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}}<x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt.\tag{5.6.3}$$

记 $\eta_{nk}=\dfrac{\xi_k-E\xi_k}{B_n}$，$k=1,2,\cdots,n$；$n=1,2,\cdots$．

定理 5.6.1 的结论是 n 充分大时，独立随机变量 η_{n1}，η_{n2}，…，η_{nn} 的和

$$\sum_{k=1}^{n}\eta_{nk}=\sum_{k=1}^{n}\frac{\xi_k-E\xi_k}{B_n}=\frac{\sum_{k=1}^{n}\xi_k-\sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}}$$

的分布接近于标准正态分布．由于对任何 $\varepsilon>0$，

$$P\left(\max_{1\leqslant k\leqslant n}|\eta_{nk}|>\varepsilon\right)\leqslant\sum_{k=1}^{n}P\left(|\eta_{nk}|>\varepsilon\right)$$

$$=\sum_{k=1}^{n}P\left(|\xi_k-E\xi_k|>\varepsilon B_n\right)$$

$$= \sum_{k=1}^{n} \int_{|x-E\xi_k|>\tau B_n} dF_k(x)$$

$$\leqslant \frac{1}{(\varepsilon B_n)^2} \sum_{k=1}^{n} \int_{|x-E\xi_k|>\tau B_n} (x-E\xi_k)^2 dF_k(x).$$

条件 (5.6.2)表明，在 n 充分大时，$|\eta_{nk}|$ 有一项大于 ε 的概率很小，这就是说，以很大的概率 $|\eta_{nk}|$ 同时小于 ε，即 η_{nk}，$k=1,2,\cdots,n$ 均匀地小. 条件 (5.6.2)就是对前面所说的 $\sum_{k=1}^{n}\eta_{nk}$ 的每一项对总和影响不大的比较清楚的叙述，条件 (5.6.2)叫林德伯格(Linderberg)条件. 在增加对 $D\xi_k$，$k=1,2,\cdots$ 的一点约束条件后，林德伯格条件还是 (5.6.3)成立的必要条件.

当 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个独立同分布的随机变量序列，$b=D\xi_1(>0)$有限时，

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-E\xi_k|>\varepsilon B_n} (x-E\xi_1)^2 dF_1(x)$$

$$= \lim_{n\to\infty} \frac{1}{nb^2} \cdot n \int_{|x-E\xi_1|>\varepsilon b\sqrt{n}} (x-E\xi_1)^2 dF_1(x) = 0,$$

林德伯格条件 (5.6.2)成立，所以有

推论 5.6.4 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一个独立同分布的随机变量序列，若 $D\xi_1(>0)$有限，则对任何实数 x 有

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n}\xi_k - \sum_{k=1}^{n}E\xi_k}{\sqrt{D\left(\sum_{k=1}^{n}\xi_k\right)}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

推论 5.6.4 是大样本统计推断的理论依据之一. 在 §5.4,5.5 中对正态分布的未知参数做的统计推断属于小样本统计推断，那时我们精确地知道所关心的一些统计量的分布，正是因为知道了这些分布，才有可能对给定的有关正态分布参数的假设检验问题，确定出适当的判别区域，对正态分布中的未知参数做出区间估计. 然而，在实际问题中也常常会遇到不是正态随机变量的情形，甚至随机变量的分布是未知的，这时和这个随机变量有关的统计量，如样本平均值的精确分布也是未知的，无法根据统计量的精确分布对某一假设检验问题确定适当的判别区域或对某一未知参数做出区间估计. 但是当样本容量无限增大时，有时某些统计量的极限分布是知道的，这使我们有可能用容量比较大的样本解决某些统计推断问题. 作为例子，下面我们介绍数学期望的大样本统计检验的方法.

根据推论 5.6.4，若 ξ 是一个方差有限的随机变量，$D\xi=\sigma_0^2$. 已知，$x_1, x_2, \cdots, x_n, \cdots$ 是 ξ 的一个容量为可数无限的样本，则对任何实数 x 有

$$\lim_{n\to+\infty} P\left(\frac{\sum_{i=1}^{n} x_i - nE\xi}{\sqrt{n\sigma_0^2}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

或

$$\lim_{n\to+\infty} P\left(\frac{\frac{1}{n}\sum_{i=1}^{n} x_i - E\xi}{\sigma_0/\sqrt{n}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \tag{5.6.5}$$

当我们考虑假设检验问题

$$H: E\xi=\mu_0 \quad , \quad K: E\xi\neq\mu_0$$

时，因为我们不知道 ξ 的分布，我们不能像 U 检验法中那样利用 H 真时 $\overline{x}\sim N(\mu_0, \sigma_0^2/n)$ 确定判别区域. 但是从 (5.6.5) 式中知

道，只要 n 比较大，在 H 真时 $\frac{\overline{x}-\mu_0}{\sigma_0/\sqrt{n}}$ 的分布接近于标准正态分布，仍就可以用判别区域 (5.4.6)。例如已知某种化学纤维强力 ξ 的方差是 1.19^2，希望检验统计假设 $H: E\xi=6.2$。取一个容量为 100 的样本测得强力平均值为 $\overline{x}=6.35$，当信度 α 取 0.05 时 (5.4.6) 中

$$c=1.96\times\frac{1.19}{10}=0.23\ ,$$

因为

$$|\overline{x}-\mu_0|=6.35-6.2=0.15<0.23\ ,$$

应当接受 H。在实际问题中如果 $D\xi$ 也是未知的，因为样本容量比较大，可以用样本的二阶中心矩 $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$ 代替。

习题 5-7

1. 在一家保险公司里有 10000 个人参加保险，每人每年付 12 元保险费，若在一年内一个人死亡的概率为 0.006，死亡时其家属可向保险公司领得 1000 元，问：

(1) 保险公司亏本的概率有多大?

(2) 保险公司一年的利润不少于 40000 元、60000 元、80000 元的概率有多大?

2. 为估计某市 17 岁男学生身高，抽测了其中 200 人得身高平均值 $\overline{x}=166.3$ (单位 cm)，二阶中心矩 $s^2=37.22$ (cm^2)，试对该市 17 岁男学生平均身高做区间估计。

第六章　回归分析与因子试验的方差分析

§6.1 方差分析

在§5.4中我们曾经介绍过 t 检验法，用这个检验方法可以检验方差相同的两个正态总体的数学期望是否相同．但是有许多场合需要我们同时比较几个方差相同的正态总体的数学期望是否相同．

有一个灯泡厂用四种不同材料的灯丝 A_1，A_2，A_3，A_4 制成四批灯泡．现在从这四批灯泡中分别随机抽取几只进行使用寿命(一只灯泡能发光的时间)的试验，试验结果见表 6－1 (单位：小时)．

表　6－1

A_1	A_2	A_3	A_4
1600	1580	1460	1510
1610	1640	1550	1520
1650	1640	1600	1530
1680	1700	1620	1570
1700	1750	1640	1600
1720		1660	1680
1800		1740	
		1820	

表中四组灯泡的寿命平均值分别是 1680、1662、1636.25 和 1568.3．从这四个平均值的大小能不能说用 A_1 制成的灯泡寿命最长，用 A_4 制成的灯泡寿命最短呢？对这个问题需要慎重地考虑．这是因为用同一灯丝材料制成的灯泡其寿命也有长有短．例如第一组中长的可达 1800 小时，短的只有 1600 小时，而第四组中短

的虽只有1510小时，但长的也有超过1600小时的，而且寿命是长还是短也有偶然性．这里需要区分这四个平均寿命的差别是随机因素造成的，还是由使用灯丝材料的不同造成的．实际上，用同一种灯丝材料制成的灯泡其寿命是一个随机变量，我们用 ξ_1，ξ_2，ξ_3，ξ_4 分别表示四种灯丝材料制成的灯泡的寿命，这样，表6－1中的四列数据分别是 ξ_1，ξ_2，ξ_3，ξ_4 的样本，应当比较的是 $E\xi_1$，$E\xi_2$，$E\xi_3$，$E\xi_4$ 的大小．我们可以把这个问题做为检验统计假设 $H: E\xi_1=E\xi_2=E\xi_3=E\xi_4$ 的问题来处理．

这个问题的一般提法是：设有 p 个独 的随机变量 ξ_1，ξ_2，…，ξ_p，$\xi_i\sim N(\mu_i,\sigma^2)$，$i=1,\cdots,p$，其中 $\mu_1,\cdots,\mu_p,\sigma^2$ 是未知的．ξ_1 有一个样本 $y_{11},y_{12},\cdots,y_{1n_1}$；$\xi_2$ 有一个样本 $y_{21},y_{22},\cdots,y_{2n_2}$；……；$\xi_p$ 有一个样本 $y_{p1},y_{p2},\cdots,y_{pn_p}$，怎样利用联合样本

(6．1．1) $(y_{11},\cdots,y_{1n_1};y_{21},\cdots,y_{2n_2};\cdots;y_{p1},\cdots,y_{pn_p})$

检验统计假设

$$H:\mu_1=\mu_2=\cdots=\mu_p$$

我们从分析离差

$$S_T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$$

入手，其中 $\bar{y}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}y_{ij}$，$n=\sum_{i=1}^{p}n_i$．$S_T$ 取的值反映着样本(6．1．1)取的值的分散程度．例如表6－1中26个数据有大有小，其分散的程度可以用数值

$$S_T=(1600-1637.3)^2+(1610-1637.3)^2+\cdots+(1680-1637.3)^2$$
$$=195711.5$$

来刻画，其中1637.3是表6－1中26个数值的平均值．

分散的原因可能有两个，一个是 $\mu_1,\mu_2,\cdots,\mu_p$ 不全相同，ξ_1，ξ_2，…，ξ_p 的数学期望不全相同自然会影响到其样本观测值的大小不一；另一个原因可能是 $\mu_1,\mu_2,\cdots,\mu_p$ 完全相同，但由于 ξ_1，ξ_2，…，ξ_p 是随机变量，其样本观测值也会大小不一．为了区分这

两种情况，我们把离差 S_T 做一个分解．记

$$y_{i\cdot}=\frac{1}{n_i}\sum_{j=1}^{n_i}y_{ij},i=1,2,\cdots,p.$$

$y_{1\cdot},y_{2\cdot},\cdots,y_{p\cdot}$ 分别是 $\xi_1,\xi_2,\cdots,\xi_p$ 的样本平均值．

$$(6.1.2)\ S_T=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}[(y_{ij}-y_{i\cdot})+(y_{i\cdot}-\bar{y})]^2$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2+\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{i\cdot}-\bar{y})^2+2\sum_{i=1}^{p}[(y_{i\cdot}-\bar{y})\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})]$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2+\sum_{i=1}^{p}n_i(y_{i\cdot}-\bar{y})^2,$$

最后一步利用了 $\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})=0,i=1,2,\cdots,p$．记

$$S_E=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2.$$

S_E 是 S_T 分解的第一项，其中 $\sum_{j=1}^{n_1}(y_{1j}-y_{1\cdot})^2$ 反映着 ξ_1 的样本的分散程度，这个分散完全是随机性造成的；同样 $\sum_{j=1}^{n_2}(y_{2j}-y_{2\cdot})^2$ 反映着 ξ_2 的样本的分散程度，也是随机性造成的；……，所以从 S_T 分解出来的第一部分 S_E 完全是随机性造成的．它反映了各个组内的偏差总和（有时称为组内偏差平方和）．基于这一点，如果 μ_1，μ_2，…，μ_p 不完全相同，那么由此带来的样本 (6．1．1)的分散将在 S_T 分解的第二部分

$$S_A=\sum_{i=1}^{p}n_i(y_{i\cdot}-\bar{y})^2$$

上反映，它是组间的偏差平方和．因此，和 S_E 比较，如果 S_A 不太大，我们只能认为 S_A 也是随机性造成的，从而接受 H．和 S_E 比较，如果 S_A 太大，我们就有理由怀疑 S_A 不完全是随机性造成的，从而拒绝 H．所以，我们将取形如

(6 . 1 . 3) $\left\{\frac{S_A}{S_E} > c\right\}$

的判别区域，其中常数 c 将由信度 α 来确定．为了确定 c，需要知道 H 真时，统计量 S_A/S_E 的分布．为此我们将使用下面的定理．

定理6 . 1 . 4*) 设随机变量 $Y_i \sim N(0,1)$，$i=1,2,\cdots,n$ 独立，$A_1,\cdots,A_k$ 都是 n 阶对称方阵．记 $Y=(Y_1,Y_2,\cdots,Y_n)'$．若 $Y'A_iY \sim \chi^2(r_i)$，$i=1,2,\cdots,k-1$，$Y'A_kY$ 是非负二次型，并且

$$Y'AY=\sum_{i=1}^{k} Y'A_iY \sim \chi^2(r),$$

则 $Y'A_kY \sim \chi^2(r_k)$，$r_k=r-\sum_{i=1}^{k-1} r_i$，并且 $Y'A_iY$，$i=1,2,\cdots,k$ 独立．

当 H 真时，$\mu_1=\cdots=\mu_p \triangleq \mu$．

在(6 . 1 . 2)式的两端同除以 σ^2，并记

$$z_{ij}=\frac{y_{ij}-\mu}{\sigma},\quad z_i=\frac{1}{n_i}\sum_{j=1}^{n_i} z_{ij},\ j=1,\cdots,n_i;\ i=1,\cdots,p,$$

$$\bar{z}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} z_{ij},\quad Z=(z_{11},\cdots,z_{1n_1};\cdots;z_{p1},\cdots,z_{pn_p})',$$

则 $z_{ij} \sim N(0,1)$，$j=1,\cdots,n_i$；$i=1,\cdots,p$ 独立．由定理 5·4·10 知(6 . 1 . 2)式的左端除以 σ^2 后

$$\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2}{\sigma^2}=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij}-\bar{z})^2 \sim \chi^2(n-1),$$

是 Z 的二次型．(6 . 1 . 2)式右端第一项除以 σ_2 后

$$\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2}{\sigma^2}=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(z_{ij}-z_i)^2.$$

---

*)证明见参考书目[8]，[16]．

因为 $\sum_{j=1}^{n_i}(z_{ij}-z_{i\cdot})^2 \sim \chi^2(n_i-1)$，$i=1,\cdots,p$ 独立，由例 4.4.27 知

$$\frac{S_E}{\sigma^2} \sim \chi^2(n-p).$$

(6.1.2)式右端第二项除以 σ^2 后

$$\frac{S_A}{\sigma^2} = \sum_{i=1}^{p} n_i(z_{i\cdot}-\bar{z})^2$$

是 Z 的非负二次型，根据定理 6·1·4，应当有 $\frac{S_A}{\sigma^2} \sim \chi^2(p-1)$，并且 $\frac{S_A}{\sigma^2}$ 与 $\frac{S_E}{\sigma^2}$ 独立. 所以

$$(6.1.5)\quad \frac{n-p}{p-1}\frac{S_A}{S_E} = \frac{\frac{S_A}{\sigma^2}\Big/p-1}{\frac{S_E}{\sigma^2}\Big/n-p} \sim F(p-1,n-p).$$

请读者注意：(6.1.5)是在 H 真时得到的，在给定信度 α 后从 F 分布表查到 $F_\alpha(p-1,n-p)$便有

$$P\left(\frac{n-p}{p-1}\cdot\frac{S_A}{S_E} > F_\alpha(p-1,n-p)\right)=\alpha$$

这样，在得知 H 真时统计量 $(n-p)S_A/(p-1)S_E$ 的分布后，(6.1.3)的 c 由信度 α 确定了，判别区域是

$$\left\{(y_{11};\cdots,y_{1n_1};\cdots;y_{p1},\cdots,y_{pn_p}):\frac{n-p}{p-1}\frac{S_A}{S_E} > F_\alpha(p-1,n-p)\right\},$$

对应的检验方法是：

若 $\frac{n-p}{p-1}\frac{S_A}{S_E} > F_\alpha(p-1,n-p)$，就拒绝 H，否则接受 H.

例 6.1.6 前面提到的用四种不同的灯丝材料制做的灯泡的寿命 ξ_1，ξ_2，ξ_3，ξ_4 是独立随机变量，他们的分布都接近于正态

分布．可以假定 $\xi_i \sim N(\mu_i, \sigma^2)$，$i=1,2,3,4$．现在检验统计假设

$$H: \mu_1=\mu_2=\mu_3=\mu_4,$$

信度取 0.05．

表 6－1 已经给出了联合样本的观测值，$n_1=7$，$n_2=5$，$n_3=8$，$n_4=6$，$p=4$，$n=\sum_{i=1}^{4} n_i=26$．查 F 分布表 $F_{0.05}(3,22)=3.05$．下面计算 $\frac{n-p}{p-1}\frac{S_A}{S_E}$．首先给出计算 S_A，S_E 的简化计算公式，记

$$T_i=\sum_{j=1}^{n_i} y_{ij},$$

则

$$y_{i\cdot}=T_i/n_i,\ i=1,2;\cdots,p;\ \bar{y}=\frac{1}{n}\sum_{i=1}^{p} T_i.$$

$$S_E=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2=\sum_{i=1}^{p}\left(\sum_{j=1}^{n_i} y_{ij}^2-n y_{i\cdot}^2\right)=\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-\sum_{i=1}^{p}\frac{T_i^2}{n_i},$$

$$S_A=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2-S_E=\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-n\bar{y}^2-S_E$$

$$=\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-\frac{\left(\sum_{i=1}^{p} T_i\right)^2}{n}\right)-\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2-\sum_{i=1}^{p}\frac{T_i^2}{n_i}\right)$$

$$=\sum_{i=1}^{p}\frac{T_i^2}{n_i}-\frac{\left(\sum_{i=1}^{p} T_i\right)^2}{n}.$$

在计算 S_A，S_E 时，用表 6－2 的计算表格是方便的．

$$S_E=69895900-69744549.2=151350.8,$$

$$S_A=69744549.2-\frac{42570^2}{26}=44360.7,$$

$$\frac{n-p}{p-1}\frac{S_A}{S_E}=\frac{22}{3}\times\frac{44360.7}{151350.8}=2.15\ .$$

表 6－2

灯泡种类	A_1	A_2	A_3	A_4	总　　和
n_i	7	5	8	6	$n=\sum_{i=1}^{4}n_i=26$
T_i	11760	8310	13090	9410	$\sum_{i=1}^{4}T_i=42570$
T_i^2	138297600	69056100	171348100	88548100	
T_i^2/n_i	19756800	13811220	21418512.5	14758016.7	$\sum_{i=1}^{4}\frac{T_i^2}{n_i}=69744549.2$
$\sum_{j=1}^{n_i}y_{ij}^2$	19785400	13828100	21503700	14778700	$\sum_{i=1}^{4}\sum_{j=1}^{n_i}y_{ij}^2=69895900$

由于 $2.15<3.05$，应当接受 H，认为四种灯丝材料制做的灯泡在寿命上没有差别．

这个例子说明，如果只比较四个样本的平均值的大小，就会做出用 A_1 制成的灯泡使用寿命最长的结论．但经过方差分析，我们可以做出四种灯丝材料制做的灯泡在寿命上没有差别的判断，这样，灯泡厂在选择灯丝材料时就可以从其它方面去考虑，例如，在这四种灯丝材料中选择能使灯泡成本较低的材料等等．

例 6．1．7　从一所学校的三个年级中随机地各取一班学生进行某项测验，结果如下：

表　6－3

年　级	人　数	平均分数	标准差 *)
一	48	78.2	4.11
二	50	80.5	3.18
三	45	79.8	4.20

*) 样本 $x_1, x_2, \cdots, x_n$ 的标准差指 $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$．

根据这个结果能不能说这三个年级的成绩没有差别？

这个问题可以用方差分析的方法来解决．用 ξ_1，ξ_2，ξ_3 分别表示一、二、三年级学生的成绩，$\xi_i \sim N(\mu_i, \sigma^2)$，$i=1,2,3$ 独立．下面我们就来检验统计假设

$$H: \mu_1 = \mu_2 = \mu_3 .$$

ξ_1 有一个容量为 $n_1=48$ 的样本 $y_{11}, y_{12}, \cdots, y_{1n_1}$，表 6－3 提供了它们的平均值及标准差

$$y_{1\cdot}=78.2,\ \sqrt{\frac{1}{48}\sum_{j=1}^{48}(y_{1j}-y_{1\cdot})^2}=4.11$$

ξ_2 有一个容量为 $n_2=50$ 的样本 $y_{21}, y_{22}, \cdots, y_{2n_2}$，其平均值与标准差

$$y_{2\cdot}=80.5,\ \sqrt{\frac{1}{50}\sum_{j=1}^{50}(y_{2j}-y_{2\cdot})^2}=3.18 .$$

ξ_3 有一个容量为 $n_3=45$ 的样本 $y_{31}, y_{32}, \cdots, y_{3n_3}$，其平均值与标准差

$$y_{3\cdot}=79.8,\ \sqrt{\frac{1}{45}\sum_{j=1}^{45}(y_{3j}-y_{3\cdot})^2}=4.20 .$$

由这些数据可以计算出

$$S_E=\sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij}-y_{i\cdot})^2=4.11^2\times48+3.18^2\times50+4.20^2\times45$$

$$=2110.24,$$

$$n=n_1+n_2+n_3=48+50+45=143,$$

$$\bar{y}=\frac{1}{n}(n_1y_{1\cdot}+n_2y_{2\cdot}+n_3y_{3\cdot})$$

$$=\frac{1}{143}(78.2\times48+80.5\times50+79.8\times45)=79.5,$$

$$S_A=(78.2-79.5)^2\times48+(80.5-79.5)^2\times50+(79.8-79.5)^2\times45$$

$$=135.17,$$

$$\frac{n-p}{p-1}\frac{S_A}{S_E}=\frac{143-3}{3-1}\times\frac{135.17}{2110.24}=4.48.$$

取信度 $\alpha=0.05$，查 F 分布表，$F_{0.05}(2,140)=3$．由于 $4.48>3$，应当拒绝 H，认为三个年级学生的成绩是有差别的．

习题 6－1

1．将4个玉米品种 A_1，A_2，A_3，A_4 分别种在条件相同的16小块地上，每个品种种4块，得到的产量如下表（单位：公斤）．设 A_i 在一小块地上的产量 $\xi_i\sim N(\mu_i,\sigma^2)$，$i=1,2,3,4$ 独立，试检验统计假设 $H:\mu_1=\mu_2=\mu_3=\mu_4$，信度取0.05．

A_1	A_2	A_3	A_4
1.7	1.5	1.8	1.6
1.6	1.3	1.6	1.6
1.7	1.4	1.5	1.4
1.5	1.8	1.7	1.8

2．在例6．1．7中用 $\xi_1\sim N(\mu_1,\sigma^2)$、$\xi_2\sim N(\mu_2,\sigma^2)$、$\xi_3\sim N(\mu_3,\sigma^2)$ 分别表示一年级、二年级和三年级学生的成绩．

(1) 检验统计假设 $H_1:\mu_1=\mu_2$；

(2) 检验统计假设 $H_2:\mu_2=\mu_3$，

信度取0.05．

§6.2 线性回归分析

现实世界中的量之间有些有明确的函数关系，如正方形的边长 x 与面积 y 之间就有函数关系 $y=x^2$，做匀速直线运动的物体的速度 v、运动时间 t 及在时间间隔 t 内所走的路程 s 之间有函数关系 $s=vt$．另外，也有一些量，它们之间没有明确的函数关系，如人的身高与体重，年龄与血压等等．

为了了解身高与体重的关系，随机抽测了身高分别为155，157，159，161，163，165，167（单位是厘米）的15岁男学生各1至2名的体重（单位是公斤），列在表6－4中．

表 6－4

x (身高)	155	157	157	159	161	163	163	165	167
y (体重)	42	47	44	47	50	49	53	52	55

如果把身高 x 作为横坐标，体重 y 作为纵坐标，在坐标纸上画出 9 个学生身高、体重对应的点 (图 6－1)，我们就会看到随着身高的增长体重呈直线增加的趋势．

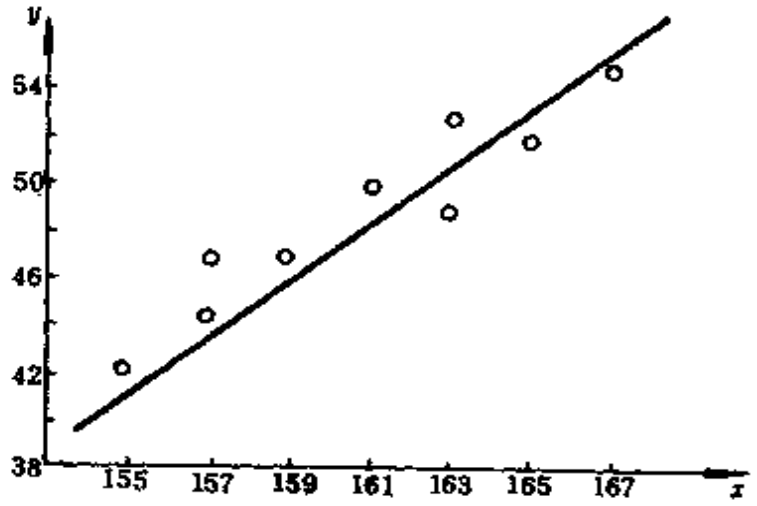


图 6－1

怎样刻画身高与体重之间的这种关系呢？一名 15 岁的男学生身高长到 157 厘米时，其体重可能是 44 公斤，可能是 44.2 公斤，……可能是某个范围内的任何一个值，而且无法预言一定是哪个值，所以我们可以把身高为 157 厘米、15 岁的男学生体重看作是一个随机变量．一般的，我们可以把身高为 x 厘米的、15 岁男学生的体重看作是一个随机变量 Y_x，体重随身高的增长而直线增长的趋势用下面的式子来描述：

$$EY_x=\beta_0+\beta_1 x\ .$$

对应于身高为 x 的体重的期望值随 x 的增大而线性地增大．记 $\varepsilon_x=Y_x-EY_x$，则有

$$Y_x=\beta_0+\beta_1 x+\varepsilon_x\ .$$

ε_x 是随机变量，$E\varepsilon_x=0$．我们的任务是设法求出 β_0，β_1 的估计值 $\hat{\beta}_0$，$\hat{\beta}_1$，这样我们就对身高与体重的关系有一个比较明确的认识．如果有人问你身高为 162 厘米的 15 岁男学生的体重是多少，你就可以回答：估计平均体重是 $\hat{\beta}_0+162\hat{\beta}_1$ 公斤．

估计 β_0，β_1 需要 x 及 Y_x 的观测值．表 6－4 中列出的就是 9 组 x 及 Y_x 的观测值．对这 9 组值可以这样理解：第一个学生是从身高为 $x_1=155$ 厘米的学生中任意选出的一个，现在记身高为 x_1 的学生的体重（是一个随机变量）为 y_1．按照前面的分析

$$y_1=\beta_0+\beta_1x_1+\varepsilon_1,\ E\varepsilon_1=0.$$

42 是随机变量 y_1 取的一个值．类似的，第二个学生是从身高为 $x_2=157$ 厘米的学生中任意选出的一个，记身高为 x_2 的学生体重为 y_2．

$$y_2=\beta_0+\beta_1x_2+\varepsilon_2,\ E\varepsilon_2=0.$$

47 是随机变量 y_2 取的一个值．

第六个、第七个学生身高都是 163 厘米．第六个学生是从身高为 $x_6=163$ 厘米的学生中任意选出的第一个学生，其体重在选出之前不能精确预言，是一个随机变量，记为 y_6．

$$y_6=\beta_0+\beta_1x_6+\varepsilon_6,\ E\varepsilon_6=0.$$

49 是 y_6 取的一个值．第七个学生是从身高为 $x_7=163$ 厘米的学生中任意选出的第二个学生，表示其体重的随机变量记为 y_7．

$$y_7=\beta_0+\beta_1x_7+\varepsilon_7,\ E\varepsilon_7=0,$$

53 是 y_7 取的一个值．

按照上面的理解，对应于 x_1，x_2，…，x_9 有 9 个随机变量 y_1，y_2，…，y_9，它们具有下面的结构

$$\begin{cases} y_1=\beta_0+\beta_1x_1+\varepsilon_1, \\ y_2=\beta_0+\beta_1x_2+\varepsilon_2, \\ \quad\cdots\cdots \\ y_9=\beta_0+\beta_1x_9+\varepsilon_9. \end{cases}$$

由于 9 个学生是任意挑选出来的，y_1，y_2，…，y_9 可以看作是独立

随机变量，因而 ε_1，ε_2，$\cdots$，ε_9 是独立的随机变量．还可以进一步假定 ε_1，ε_2，$\cdots$，ε_9 都服从 $N(0,\sigma^2)$分布，σ^2 可以是未知的．我们将根据 x_1，x_2，$\cdots$，x_9 及 y_1，y_2，$\cdots$，y_9 取的值对 β_0，β_1 做估计．

上面通过一个关于身高体重的例子引出了一个对 β_0，β_1 的估计问题．这个问题的一般提法是：随机变量 Y_x 的数学期望 EY_x 与变量 x 之间有线性关系 $EY_x=\beta_0+\beta_1 x$．记 $\varepsilon_x=Y_x-EY_x$，则

$$Y_x=\beta_0+\beta_1 x+\varepsilon_x, \tag{6.2.1}$$

假设 $\varepsilon_x\sim N(0,\sigma^2)$．对应于 x 的 n 个不完全相同的值 x_1，x_2，$\cdots$，x_n 有 n 个随机变量 y_1，y_2，$\cdots$，y_n，它们具有下面的结构．

$$\begin{cases} y_1=\beta_0+\beta_1 x_1+\varepsilon_1, \\ y_2=\beta_0+\beta_1 x_2+\varepsilon_2, \\ \cdots\cdots \\ y_n=\beta_0+\beta_1 x_n+\varepsilon_n, \end{cases} \tag{6.2.2}$$

其中 ε_1，ε_2，$\cdots$，ε_n 是独立的随机变量，每一个 ε_i 都服从分布 $N(0,\sigma^2)$*)，β_0，β_1，σ^2 是未知参数．称 (6.2.2)为一元线性回归模型．我们将根据 x_1，x_2，$\cdots$，x_n 及 y_1，y_2，$\cdots$，y_n 取的值对 β_0，β_1，σ^2 做出估计．为了叙述上的方便，以下 y_1，y_2，$\cdots$，y_n 取的值也用 y_1，y_2，$\cdots$，y_n 表示，读者从上下文可以区分 y_1，y_2，$\cdots$，y_n 是随机变量还是具体的数值．

1．β_0，β_1 的最小二乘估计与回归方程

β_0，β_1 的估计 $\hat{\beta}_0$，$\hat{\beta}_1$ 应当使 $\hat{\beta}_0+\hat{\beta}_1 x_i$ 与 y_i 比较接近，$i=1$，2，$\cdots$，n．我们将取使

$$S(b_0,b_1)=\sum_{i=1}^{n}[y_i-(b_0+b_1 x_i)]^2$$

达到极小的 b_0，b_1 作为 β_0，β_1 的估计值．为此先对 b_0，b_1 的函数 $S(b_0,b_1)$的形式做一些变换．记

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i.$$

*)细心的读者会发现 ε_i $(i=1,2,\cdots,n)$服从正态分布这一假设条件只在本节几节用到，更多的地方只用到条件：$E\varepsilon_i=0$，$D\varepsilon_i=\sigma^2$，$i=1,2,\cdots,n$．

$$(6.2.3)\quad S(b_0,b_1)=\sum_{i=1}^{n}[(y_i-\bar{y})-b_1(x_i-\bar{x})+(\bar{y}-b_0-b_1\bar{x})]^2$$

$$=b_1^2\sum_{i=1}^{n}(x_i-\bar{x})^2-2b_1\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-b_0-b_1\bar{x})^2$$

$$+2(\bar{y}-b_0-b_1\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})-2b(\bar{y}-b_0-b_1\bar{x})\sum_{i=1}^{n}(x_i-\bar{x}).$$

由于$\sum_{i=1}^{n}(x_i-\bar{x})=0$，$\sum_{i=1}^{n}(y_i-\bar{y})=0$，上式最后两项是零．$S(b_0,b_1)$的前三项是一个$b_1$的二次三项式，第四项是非负的，如能取到$b_1$使前三项达到极小，再取$b_0$使第四项为零，那么，使$S(b_0,b_1)$达到极小的$b_0$，$b_1$就找到了．注意，$x_1,\cdots,x_n$是不全相同的$n$个实数，所以$\sum_{i=1}^{n}(x_i-\bar{x})^2>0$，$S(b_0,b_1)$的前三项在

$$b_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

达到极小，再取

$$b_0=\bar{y}-b_1\bar{x}=\bar{y}-\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\bar{x}$$

(b_0,b_1)的第四项就是零．所以β_0，β_1的估计值

$$(6.2.4)\quad \hat{\beta}_1(y_1,\cdots,y_n)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2},$$

(6 .2 .5)　　$\hat{\beta}_0(y_1,\cdots,y_n)=\bar{y}-\bar{x}\hat{\beta}_1(y_1,\cdots,y_n)$.

把$\hat{\beta}_0(y_1,\cdots,y_n)$,$\hat{\beta}_1(y_1,\cdots,y_n)$看作变量$y_1,\cdots,y_n$的函数，它们就是$\beta_0$ β_1的估计量，称为β_0,β_1的**最小二乘估计量**.

例 6 .2 .6　试用表6－4的 9 组数据求线性回归模型(6 .2 .2)中β_0,β_1的最小二乘估计.

β_0,β_1的最小二乘估计已有表达式((6 .2 .4)与(6 .2 .5))，所以只要把表 6－4 的 9 组数代进去，就可以求出β_0,β_1的最小二乘估计.为了减少计算步骤，我们先对表达式(6 .2 .4)，(6 .2 .5)做一些变形，然后给出一计算表格，计算出β_0,β_1，见表 6－5.

$$\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})=\sum_{t=1}^{n}(x_ty_t-\bar{x}y_t-\bar{y}x_t+\bar{x}\bar{y})$$

$$=\sum_{t=1}^{n}x_ty_t-\bar{x}\sum_{t=1}^{n}y_t-\bar{y}\sum_{t=1}^{n}x_t+n\bar{x}\bar{y}$$

$$=\sum_{t=1}^{n}x_ty_t-\bar{x}\cdot n\bar{y}-\bar{y}\cdot n\bar{x}+n\bar{x}\bar{y}$$

$$=\sum_{t=1}^{n}x_ty_t-\frac{1}{n}\left(\sum_{t=1}^{n}x_t\right)\left(\sum_{t=1}^{n}y_t\right),$$

$$\sum_{t=1}^{n}(x_t-\bar{x})^2=\sum_{t=1}^{n}x_t^2-\frac{1}{n}\left(\sum_{t=1}^{n}x_t\right)^2.$$

表 6－5

$\sum_{t=1}^{9}x_t=1447$	$\sum_{t=1}^{9}y_t=439$	$n=9$
$\bar{x}=160.8$	$\bar{y}=48.8$	
$\sum_{t=1}^{9}x_t^2=232777$	$\sum_{t=1}^{9}y_t^2=21557$	$\sum_{t=1}^{9}x_ty_t=70711$
$\frac{1}{9}\left(\sum_{t=1}^{9}x_t\right)^2=232645.44$	$\frac{1}{9}\left(\sum_{t=1}^{9}y_t\right)^2=21413.44$	$\frac{1}{9}\left(\sum_{t=1}^{9}x_t\right)\left(\sum_{t=1}^{9}y_t\right)=70581.44$
$\sum_{t=1}^{9}(x_t-\bar{x})^2=131.56$	$\sum_{t=1}^{9}(y_t-\bar{y})^2=143.56$	$\sum_{t=1}^{9}(x_t-\bar{x})(y_t-\bar{y})=129.56$
	$\hat{\beta}_1=\frac{129.56}{131.56}=0.98$	
	$\hat{\beta}_0=48.8-0.98\times160.8=-108.78$	

我们将称

(6.2.7) $\hat{y}=-108.78+0.98x$

为15岁的男生体重关于身高的回归方程.利用回归方程可以进行预测,例如身高是162厘米的15岁男生体重大体上是$-108.78+0.98\times162=50$公斤.

在一般的一元线性回归模型(6.2.2)中,称

(6.2.8) $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$

为Y_x关于x的回归方程.其中β_0,β_1是β_0,β_1的最小二乘估计.

在实际工作中也常常遇到这样的情形:(x,y)是二维随机变量,(x,y)任意取了n组值$(x_1,y_1),\cdots,(x_n,y_n)$.如果将这$n$组值对应的点在坐标纸上标出,也有类似于图6-1的情况,y_i与x_i有某种线性关系的趋势,这时,我们也可以利用这节的公式(6.2.4),(6.2.5)计算$\hat{\beta}_1$,$\hat{\beta}_0$进而得到方程(6.2.8),将它作为表示x,y的线性关系的方程,也称为y关于x的回归方程.

习题 6-2

1.为了探讨高中一年级男生跳高成绩与下肢爆发力的关系,随机测得12名学生的立定跳远和跳高成绩见下表(单位m):

立定跳远 (x)	2.08	2.30	2.12	1.95	2.10	2.35
跳　高 (y)	1.20	1.28	1.15	1.08	1.22	1.30
立定跳远 (x)	2.15	2.25	1.80	2.06	2.04	2.26
跳　高 (y)	1.16	1.24	1.06	1.12	1.10	1.26

若跳高成绩$Y_x=\beta_0+\beta_1x+\varepsilon,\varepsilon\sim N(0,\sigma^2)$,试求$\beta_0,\beta_1$的最小二乘估计.

2.从1911年到1975年一英里跑的创纪录时间的部分资料如下:

x(年)	1911	1915	1923	1931	1937	1942
y	4:15.4	4:12.6	4:10.4	4:09.2	4:06.4	4:04.6
x(年)	1945	1954	1958	1964	1967	1975
y	4:01.4	3:59.4	3:54.5	3:54.1	3:51.1	3:49.4

假设$y=\beta_0+\beta_1x+\varepsilon,E\varepsilon=0$.

(1) 试求β_0,β_1的最小二乘估计;

(2) 预测2000年一英里跑创纪录时间.

2．σ^2 的估计

在算出例 6．2．6 中 β_0，β_1 的最小二乘估计后，我们可以算出身高为 162 厘米的学生平均体重的估计值为 50 公斤．由于身高为 162 厘米的学生体重

$$Y=\beta_0+162\beta_1+\varepsilon\sim N(\beta_0+162\beta_1,\sigma^2),$$

如果 σ^2 是已知的，那么大体上有

$$P(50-1.96\sigma<Y<50+1.96\sigma)=0.95,$$

$$P(50-3\sigma<Y<50+3\sigma)=0.997$$

等等，这样就对身高为 162 厘米的学生中百分之九十几的体重在哪一个范围内有一个估计．但是 σ^2 一般是未知的，因此，对 σ^2 也要进行估计．在求 β_0，β_1 的最小二乘估计时，我们已经看到，$S(b_0,b_1)$在 $b_0=\hat{\beta}_0$，$b_1=\hat{\beta}_1$ 处取极小值，我们将用这个极小值 $S(\hat{\beta}_0,\hat{\beta}_1)$构造 σ_2的估计量，我们将证明

(6．2．9)　$ES(\hat{\beta}_0,\hat{\beta}_1)=(n-2)\sigma^2$

这样，用

(6．2．10)　$$\hat{\sigma}^2=\frac{1}{n-2}S(\hat{\beta}_0,\hat{\beta}_1)$$

作为 σ^2 的估计量是无偏的．

现在来证明 (6．2．9)．

将 (6．2．4)，(6．2．5)代入 (6．2．3)得

(6．2．11)　$$S(\hat{\beta}_0,\hat{\beta}_1)=\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-x)^2}\right]^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$+\sum_{i=1}^{n}(y_i-\bar{y})^2-2\frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\sum_{t=1}^{n}(y_t-\bar{y})^2-\frac{\left[\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})\right]^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2}$$

记$\bar{\varepsilon}=\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t$，由(6.2.2)式知

$$\bar{y}=\frac{1}{n}\sum_{t=1}^{n}y_t=\frac{1}{n}\sum_{t=1}^{n}(\beta_0+\beta_1x_t+\varepsilon_t)=\beta_0+\beta_1\bar{x}+\bar{\varepsilon},$$

$$\sum_{t=1}^{n}(y_t-\bar{y})^2=\sum_{t=1}^{n}[(\beta_0+\beta_1x_t+\varepsilon_t)-(\beta_0+\beta_1\bar{x}+\bar{\varepsilon})]^2$$

$$=\sum_{t=1}^{n}[\beta_1(x_t-\bar{x})+(\varepsilon_t-\bar{\varepsilon})]^2$$

$$=\beta_1^2\sum_{t=1}^{n}(x_t-\bar{x})^2+\sum_{t=1}^{n}\varepsilon_t^2-n\bar{\varepsilon}^2+2\beta_1\sum_{t=1}^{n}(x_t-\bar{x})\varepsilon_t.$$

利用$\sum_{t=1}^{n}(x_t-\bar{x})\bar{y}=0$，$\sum_{t=1}^{n}(x_t-\bar{x})\bar{x}=0$可以得到

$$\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})=\sum_{t=1}^{n}(x_t-\bar{x})y_t=\sum_{t=1}^{n}(x_t-\bar{x})(\beta_1x_t+\varepsilon_t)$$

$$=\beta_1\sum_{t=1}^{n}(x_t-\bar{x})^2+\sum_{t=1}^{n}(x_t-\bar{x})\varepsilon_t,$$

(6.2.11)的第二项

$$\frac{\left[\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})\right]^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2}=\beta_1^2\sum_{t=1}^{n}(x_t-\bar{x}\ )^2+2\beta_1\sum_{t=1}^{n}(x_t-\bar{x})\varepsilon_t$$

$$+\left[\sum_{t=1}^{n}\frac{(x_t-\bar{x})\varepsilon_t}{\sqrt{\sum_{t=1}^{n}(x_s-\bar{x})^2}}\right]^2,$$

系，只是由于偶然性，挑选的9名学生的体重与身高呈现出图6－1的趋势呢？如果提问人的怀疑是对的，那么用回归方程 $\hat{y}=-108.78+0.98x$ 做预测就没有什么意义了．一般地，在求得回归方程(6.2.8)后，要判断一下 β_1 是否是零，我们将把判断 β_1 是否是零作为检验统计假设 $H:\beta_1=0$ 来处理．为了确定检验 H 的判别区域，先证明一条定理

定理6.2.14 $\dfrac{S(\hat{\beta}_0,\hat{\beta}_1)}{\sigma^2}\sim\chi^2(n-2)$，

$$\hat{\beta}_1\sim N\left(\beta_1,\sigma^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$$

并且 $\hat{\beta}_1$ 与 $S(\hat{\beta}_0,\hat{\beta}_1)$ 独立．

证 取正交方阵

$$C=\begin{pmatrix}\dfrac{x_1-\bar{x}}{\sqrt{\sum_{s=1}^{n}(x_s-\bar{x})^2}} & \dfrac{x_2-\bar{x}}{\sqrt{\sum_{s=1}^{n}(x_s-\bar{x})^2}} & \cdots & \dfrac{x_n-\bar{x}}{\sqrt{\sum_{s=1}^{n}(x_s-\bar{x})^2}}\\ \dfrac{1}{\sqrt{n}} & \dfrac{1}{\sqrt{n}} & \cdots & \dfrac{1}{\sqrt{n}}\\ c_{31} & c_{32} & \cdots & c_{3n}\\ \vdots & \vdots & & \\ c_{n1} & c_{n2} & \cdots & c_{nn}\end{pmatrix}$$

并令

$$\begin{pmatrix}z_1\\z_2\\\vdots\\z_n\end{pmatrix}=C\begin{pmatrix}\varepsilon_1/\sigma\\\varepsilon_2/\sigma\\\vdots\\\varepsilon_n/\sigma\end{pmatrix},$$

由于 $\dfrac{\varepsilon_1}{\sigma}$，$\dfrac{\varepsilon_2}{\sigma}$，…，$\dfrac{\varepsilon_n}{\sigma}$ 独立，都服从标准正态分布，由定理4.4.19知，$z_1,z_2,\cdots,z_n$ 独立，都服从标准正态分布．由(6.2.12)知

$$\frac{S(\hat{\beta}_0,\hat{\beta}_1)}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\varepsilon_i}{\sigma}\right)^2-\left(\sum_{i=1}^{n}\frac{x_i-\bar{x}}{\sqrt{\sum_{s=1}^{n}(x_s-\bar{x})^2}}\cdot\frac{\varepsilon_i}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}z_i^2-z_2^2-z_1^2=\sum_{i=3}^{n}z_i^2$$

是 $n-2$ 个独立标准正态随机变量的平方和，所以

$$\frac{S(\hat{\beta}_0,\hat{\beta}_1)}{\sigma^2}\sim\chi^2(n-2).$$

又

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_s-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1x_i+\varepsilon_i)}{\sum_{s=1}^{n}(x_s-\bar{x})^2}$$

$$=\frac{\beta_1\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i}{\sum_{s=1}^{n}(x_s-\bar{x})^2}$$

$$=\beta_1+\frac{\sigma}{\sqrt{\sum_{s=1}^{n}(x_s-\bar{x})^2}}z_1,$$

所以 $\hat{\beta}_1\sim N\left(\beta_1,\sigma^2\Big/\sum_{s=1}^{n}(x_s-x)^2\right)$. $\hat{\beta}_1$ 是 z_1 的函数，$S(\hat{\beta}_0,\hat{\beta}_1)$ 是 $z_3,\cdots,z_n$ 的函数，所以 $\hat{\beta}_1$ 与 $S(\hat{\beta}_0,\hat{\beta}_1)$ 独立.

现在回过来考虑对 $H:\hat{\beta}_1=0$ 的检验问题. 由于 β_1 是 β_1 的无偏估计，当 H 真时，$\hat{\beta}_1$ 一般取绝对值较小的值. 由定理 6.2.14 及例 4.4.28 知，H 真时

$$T=\frac{\hat{\beta}_1}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum_{s=1}^{n}(x_s-\bar{x})^2}}}=\frac{\hat{\beta}_1}{\sqrt{\dfrac{\sigma^2}{\sum_{s=1}^{n}(x_s-\bar{x})^2}}}\Bigg/\sqrt{\frac{S(\hat{\beta}_0,\hat{\beta}_1)}{\sigma^2(n-2)}}$$

$\sim t(n-2)$.

T 的绝对值太大时应当拒绝 H. 给定信度 α 后，检验 $H:\beta_1=0$ 的方法是：

(1) 从自由度是 $n-2$ 的 t 分布表中查出 t_α.

(2) 计算 T；

(3) 做判断：若 $|T|>t_\alpha$ 就拒绝 H，否则接受 H.

例 6.2.15 对于例6.2.6已用表 6-4 的数据算出 β_0，β_1 的最小二乘估计，得到了回归方程. 现在就来检验统计假设 H：$\beta_1=0$，取信度 $\alpha=0.05$. 本例 $n=9$.

(1) 查自由度是 7 的 t 分布表，得 $t_{0.05}=2.37$；

(2) 在例 6.2.6 中已算得

$$\hat{\beta}_1=0.98,\ \sum_{i=1}^{n}(x_i-\bar{x})^2=131.56;$$

在例 6.2.13 中已算得 $\hat{\sigma}^2=2.28$，所以

$$T=\frac{0.98}{\sqrt{\dfrac{2.28}{131.56}}}=7.88;$$

(3) $|T|=7.88>2.37$，应拒绝 H，认为 $\beta_1\neq 0$.

习题 6-4

在习题 6-2 第 1 题的问题中可以提出统计假设 $H:\beta_1=0$，试用这一小节的方法检验 H，取信度 $\alpha=0.05$.

4．多个变量的情形

当随机变量 Y 的数学期望 EY 与变量 $x_1, x_2, \cdots, x_p$ 之间有线性关系

$$EY = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

时，

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon .$$

由 $(x_1, x_2, \cdots, x_p)$ 取的 n 组值 $(x_{t1}, x_{t2}, \cdots, x_{tp})$，$t = 1, 2, \cdots, n$ 可以产生一个线性回归模型

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} + \varepsilon_2 \\ &\cdots\cdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} + \varepsilon_n , \end{aligned}$$

其中 $\varepsilon_t \sim N(0, \sigma^2)$，$t = 1, 2, \cdots, n$；$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 独立，$\beta_0, \beta_1, \cdots, \beta_p, \sigma^2$ 是未知参数．根据 n 组值 $(x_{t1}, x_{t2}, \cdots, x_{tp}; y_t)$，$t = 1, 2, \cdots, n$ 用与前面同样的方法可以确定 $\beta_0, \beta_1, \cdots, \beta_p, \sigma^2$ 的估计量．

$\beta_0, \beta_1, \cdots, \beta_p$ 的最小二乘估计是指使

$$S(b_0, b_1, \cdots, b_p) = \sum_{t=1}^{n} [y_t - (b_0 + b_1 x_{t1} + \cdots + b_p x_{tp})]^2$$

达到极小的 $b_0, b_1, \cdots, b_p$，它们是下面方程组的解．

$$\begin{cases} l_{11} b_1 + l_{12} b_2 + \cdots + l_{1p} b_p = l_{1y} \\ l_{21} b_1 + l_{22} b_2 + \cdots + l_{2p} b_p = l_{2y} \\ \quad \cdots\cdots \\ l_{p1} b_1 + l_{p2} b_2 + \cdots + l_{pp} b_p = l_{py} \\ b_0 = \bar{y} - (b_1 \bar{x}_1 + \cdots + b_p \bar{x}_p) \end{cases}$$

其中

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t , \bar{x}_1 = \frac{1}{n} \sum_{t=1}^{n} x_{t1} , \cdots , \bar{x}_p = \frac{1}{n} \sum_{t=1}^{n} x_{tp} ;$$

$$l_{ij} = \sum_{t=1}^{n} (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j) , i, j = 1, 2, \cdots, p ;$$

$$l_{iy}=\sum_{t=1}^{n}(x_{ti}-\bar{x}_i)(y_t-\bar{y})\quad i=1,2,\cdots,p.$$

$\beta_0,\beta_1,\cdots,\beta_p$的最小二乘估计分别记为$\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p$.

称

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_px_p$$

为Y关于变量$x_1,\cdots,x_p$的回归方程. σ_2的估计量是

$$\hat{\sigma}^2=\frac{1}{n-p-1}S(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p),$$

其中

$$S(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p)=\sum_{t=1}^{n}(y_t-\bar{y})^2-(\hat{\beta}_1l_{1y}+\hat{\beta}_2l_{2y}+\cdots+\hat{\beta}_pl_{py}).$$

例6.2.16 为了研究中学生男生跳远成绩和速度、弹跳力两项素质的关系，测试了15名学生的25米行进间跑x_1(单位：秒)、纵跳x_2(单位:厘米)和跳远y(单位:厘米)的成绩，如表6－6.

表 6－6

x_1	2.9	3.1	3.1	3.3	3.4	3.5	3.4	3.4	3.7	3.2
x_2	64	45	51	51	58	46	41	44	43	48
y	515	440	480	435	475	438	435	440	445	502
x_1	3.0	3.4	3.5	3.2	3.2					
x_2	57	62	45	51	54					
y	510	493	450	472	478					

试求回归方程$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$及$\hat{\sigma}^2$.

解 以下用$\sum_t$表示$\sum_{t=1}^{15}$.

$$\sum_t x_{t1}=49.3,\quad \sum_t x_{t2}=760,\quad \sum_t y_t=7008,$$

$$\bar{x}_1=3.29,\quad \bar{x}_2=50.67,\quad \bar{y}=467.2;$$

$$\sum_t x_{t1}^2=162.67,$$

$$l_{11}=\sum_t x_{t1}^2-\frac{1}{15}\left(\sum_t x_t\right)^2=0.637;$$

$$\sum_t x_{t1}x_{t2}=2486.7,$$

$$l_{12}=\sum_t x_{t1}x_{t2}-\frac{1}{15}\left(\sum_t x_{t1}\right)\left(\sum_t x_{t2}\right)=-11.17;$$

$$\sum_t x_{t2}^2=39208,$$

$$l_{22}=\sum_t x_{t2}^2-\frac{1}{15}\left(\sum_t x_{t2}\right)^2=701.3;$$

$$\sum_t x_{t1}y_t=22978.1,$$

$$l_{1y}=\sum_t x_{t1}y_t-\frac{1}{15}\left(\sum_t x_{t1}\right)\left(\sum_t y_t\right)=-54.86;$$

$$\sum_t x_{t2}y_t=357319,$$

$$l_{2y}=\sum_t x_{t2}y_t-\frac{1}{15}\left(\sum_t x_{t2}\right)\left(\sum_t y_t\right)=2247;$$

$$\sum_t y_t^2=3285690,$$

$$l_{yy}=\sum_t y_t^2-\frac{1}{15}\left(\sum_t y_t\right)^2=11552.4.$$

解方程组

$$\begin{cases}0.637b_1-11.17b_2=-54.86,\\-11.17b_1+701.3b_2=2247,\end{cases}$$

得 $\hat{\beta}_1=-41.54,\quad \hat{\beta}_2=2.54.$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2=467.2-(-41.54)\times3.29-2.54\times50.67$$
$$=475.16.$$

所以回归方程是

$$\hat{y}=475.16-41.54x_1+2.54x_2$$

$$\hat{\sigma}^2=\frac{1}{15-2-1}\left(l_{yy}-\hat{\beta}_1l_{1y}-\hat{\beta}_2l_{2y}\right)$$
$$=\frac{1}{12}[11552.4-(-41.54)\times(-54.86)-2.54\times2247]$$
$$=297.18.$$

习题 6－5

1．为了探讨高一男生跳高成绩与下肢爆发力、速度耐力两个专项素质的关系，随机测得12名学生的立定跳远 x_1（米）、1500米跑 x_2（秒）和跳高 y（米）成绩如下表，试求回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$ 及 $\hat{\sigma}^2$．

x_1	2.08	2.30	2.12	1.95	2.10	2.35	2.15	2.25
x_2	5.83	6.07	5.97	5.75	6.20	6.05	5.93	5.75
y	1.20	1.28	1.15	1.08	1.22	1.30	1.16	1.24
x_1	1.80	2.06	2.04	2.26				
x_2	6.23	6.13	5.83	6.35				
y	1.06	1.12	1.10	1.26				

2．有一架天平，称重时有随机误差 ε，$E\varepsilon=0$，$D\varepsilon=\sigma^2$．现对实重分别为 β_1，β_2，β_3，β_4 的4个物品 A_1，A_2，A_3，A_4（β_1，β_2，β_3，β_4 未知），按下述方法称量4次：

第一次：A_1，A_2，A_3，A_4 都放在天平左盘中，右盘砝码读数为 y_1；

第 k 次：A_1，A_k 放在天平左盘中，其余两个放在右盘中，右盘砝码读数为 y_k，$k=2,3,4$（如砝码在左盘中，右盘砝码读数为负）．这样

$$y_1=\beta_1+\beta_2+\beta_3+\beta_4+\varepsilon_1,$$

$$y_2=\beta_1+\beta_2-\beta_3-\beta_4+\varepsilon_2,$$

$$y_3=\beta_1-\beta_2+\beta_3-\beta_4+\varepsilon_3,$$

$$y_4=\beta_1-\beta_2-\beta_3+\beta_4+\varepsilon_4.$$

试求 β_1，β_2，β_3，β_4 的最小二乘估计．

5．可以化为线性回归的非线性回归问题

例6.2.17 出钢时使用的盛钢水的钢包，由于钢水对耐火材料的浸蚀，容积不断增大．为了找出使用次数与增大的容积之间的关系，对使用过2次，3次，…，16次的钢包进行了观测，结果见表6－7．

表 6－7

使用次数 x	2	3	4	5	6	7	8	
增大容积 y	6.42	8.20	9.58	9.50	9.70	10.00	9.93	
使用次数 x	9	10	11	12	13	14	15	16
增大容积 y	9.99	10.49	10.59	10.60	10.80	10.60	10.90	10.76

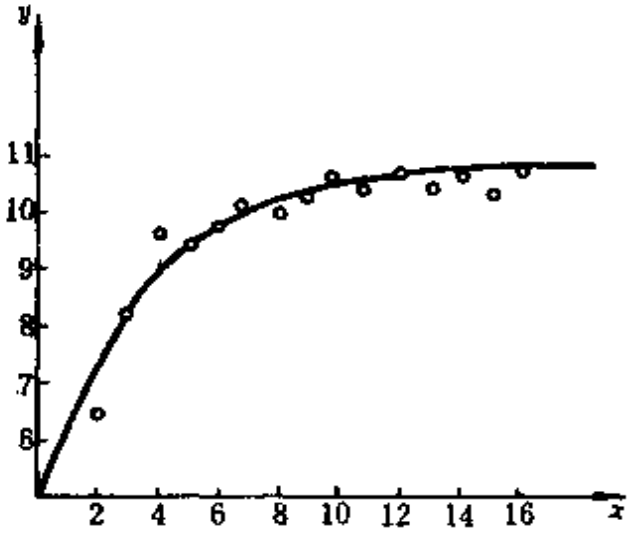


图 6－2

图6－2上标出了表6－7的15组数 (x_i, y_i)对应的点及大体上反映 Ey 与 x 之间关系的一条曲线，从图上看，这条曲线不是直线，但近似于一段对数函数曲线，因此可以考虑寻求一个如下形式的回归方程：

(6.2.18)

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1\log x$$

如果将 $\log x$ 看作是一个变量 z，那么求(6.2.18)中 $\hat{\beta}_0$，$\hat{\beta}_1$ 的问题就变为求线性回归模型中 β_0，β_1 的估计问题，以下就用15组数据 (z_i, y_i)，$i=1,2,\cdots,15$ 求 β_0，β_1 的最小二乘估计。

表 6－8

x	2	3	4	5	6	7	8	9
$z=\log x$	0.301	0.477	0.602	0.699	0.778	0.845	0.903	0.954
y	6.42	8.20	9.58	9.50	9.70	10.00	9.93	9.99
x	10	11	12	13	14	15	16	
$z=\log x$	1.000	1.041	1.079	1.114	1.146	1.176	1.204	
y	10.49	10.59	10.60	10.80	10.60	10.90	10.76	

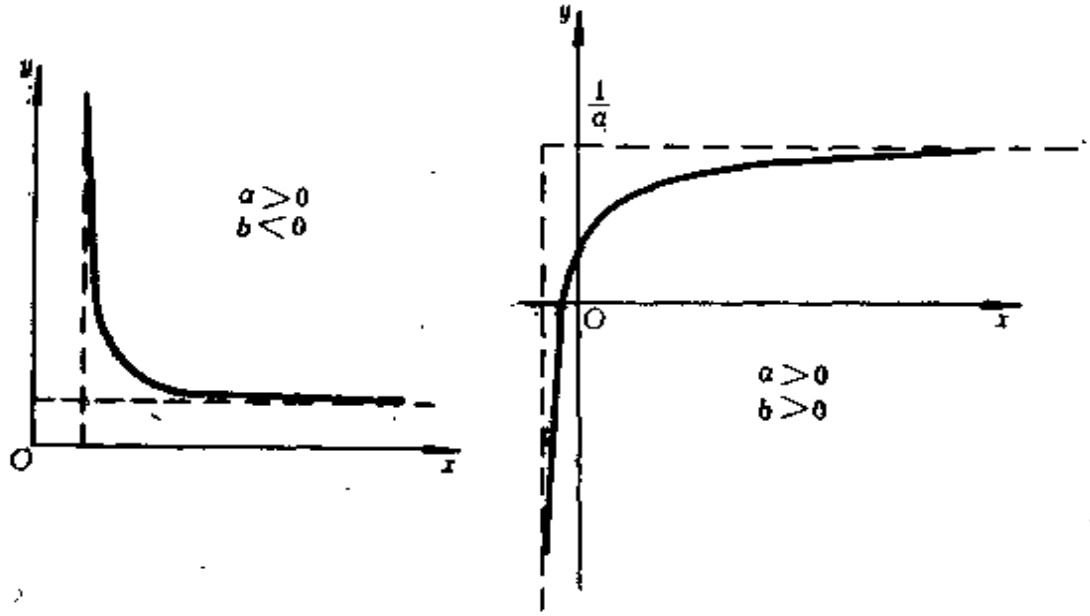
表 6-9

$\sum_{t=1}^{15} z_t = 13.319$	$\sum_{t=1}^{15} y_t = 148.06$	$n = 15$
$\bar{z} = 0.888$	$\bar{y} = 9.87$	
$\sum_{t=1}^{15} z_t^2 = 12.849$	$\sum_{t=1}^{15} y_t^2 = 1481.114$	$\sum_{t=1}^{15} z_t y_t = 135.649$
$\frac{1}{15}\left(\sum_{t=1}^{15} z_t\right)^2 = 11.826$	$\frac{1}{15}\left(\sum_{t=1}^{15} y_t\right)^2 = 1461.451$	$\frac{1}{15}\left(\sum_{t=1}^{15} z_t\right)\left(\sum_{t=1}^{15} y_t\right) = 131.467$
$\sum_{t=1}^{15} (z_t - z)^2 = 1.023$	$\sum_{t=1}^{15} (y_t - \bar{y})^2 = 19.66$	$\sum_{t=1}^{15} (z_t - \bar{z})(y_t - \bar{y}) = 4.18$
	$\hat{\beta}_1 = \frac{4.18}{1.023} = 4.084$	
	$\hat{\beta}_0 = 9.87 - 4.086 \times 0.888 = 6.24$	

这样就得到了回归方程

$$\hat{y} = 6.24 + 4.084 \log x$$

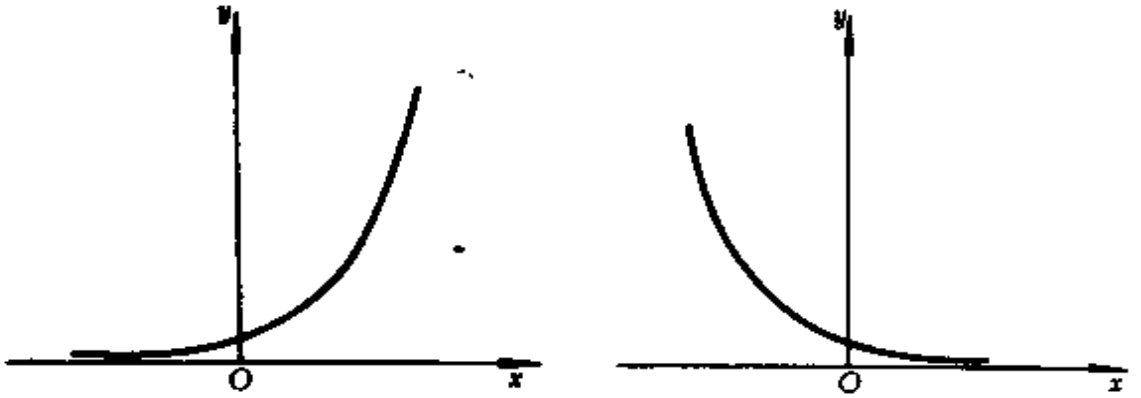
在本例中，根据直观的考虑求得了一个形如 (6.2.18)的回归方程．一般说，曲线的选择首先应当靠有关的专业知识，如不能根



$$\frac{1}{y} = a + \frac{b}{x}$$

图 6-3

据有关的专业知识确定 Ey 与 x 的函数关系，就可以从数学上考虑选择什么样的函数曲线．为了便于选择，图 6－3 到图 6－8 列出几种常用的函数曲线供参考．



$y=ae^{bx}\ (b>0)$　　$y=ae^{-bx}\ (b>0)$

图　6－4

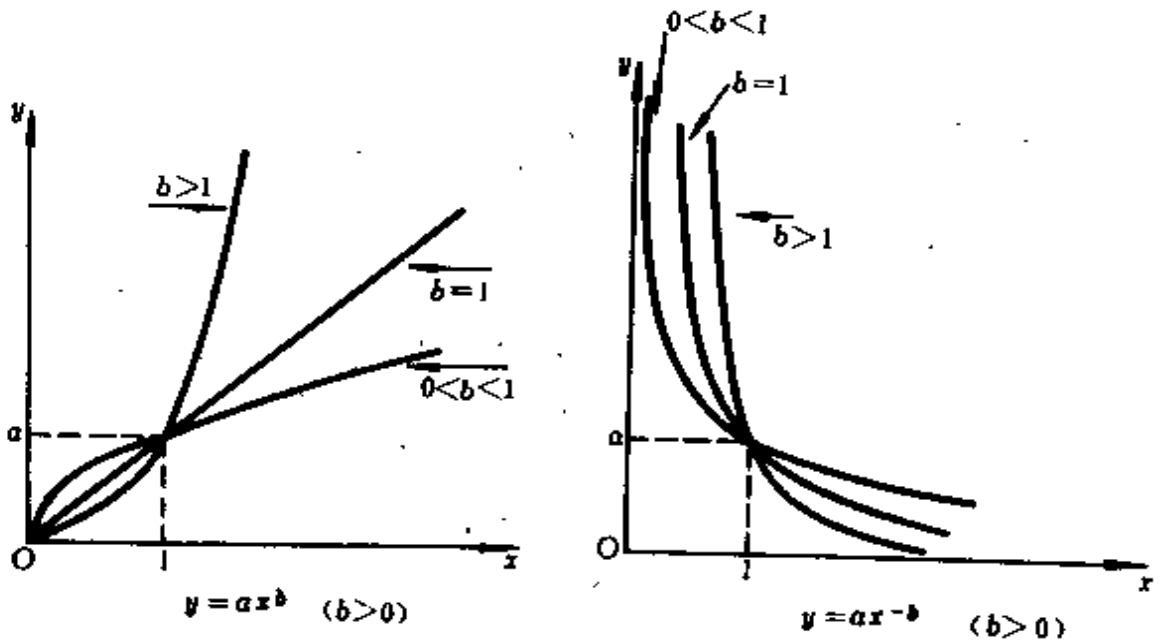


图　6－5

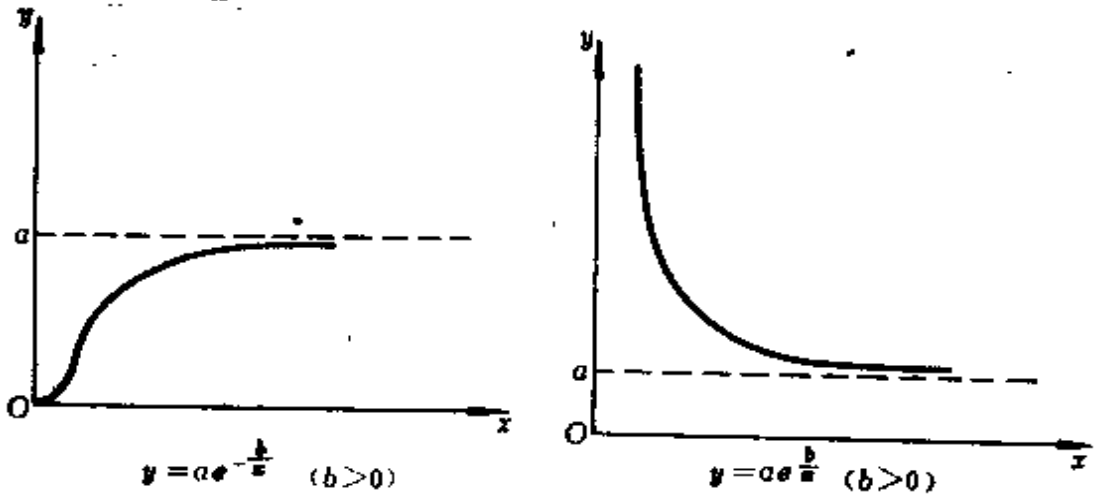

y
a
O
x
$y=ae^{-\frac{b}{x}}$ $(b>0)$
y
a
O
x
$y=ae^{\frac{b}{x}}$ $(b>0)$

图 6-6

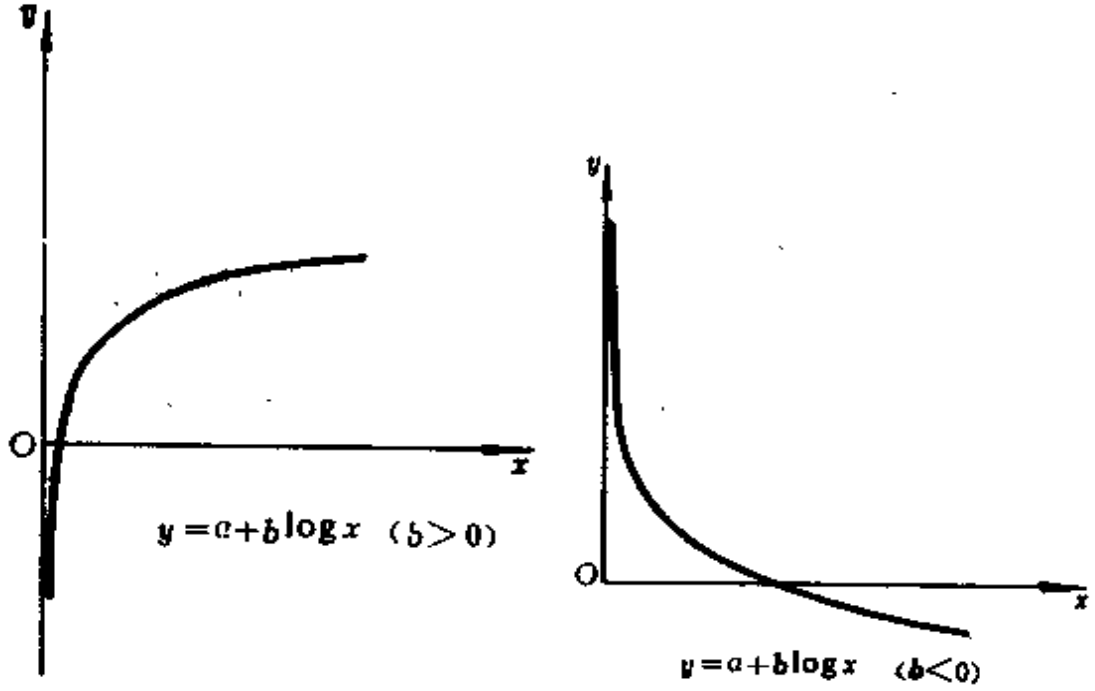

y
O
x
$y=a+b\log x$ $(b>0)$
y
O
x
$y=a+b\log x$ $(b<0)$

图 6-7

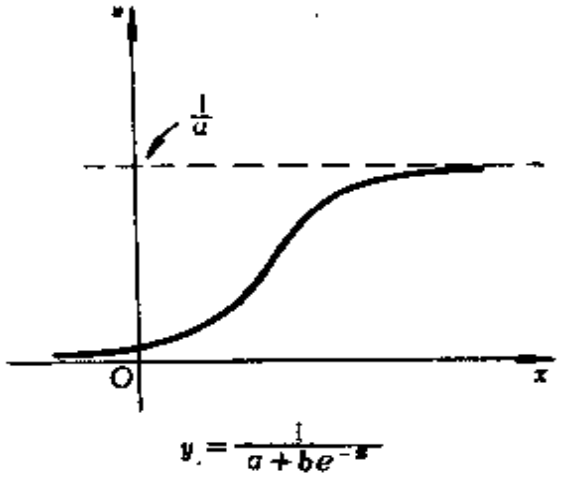


图 6－8

从数学上考虑选择曲线，例 6．2．17 还可以考虑如下形式的回归方程

(1) $\hat{y}=a+\frac{b}{x^2}$；

(2) $\hat{y}=ae^{\frac{b}{x}}$；

(3) $\hat{y}=ae^{\frac{b}{x^2}}$；

(4) $\hat{y}=ae^{\frac{b}{x+c}}$；

(5) $\frac{1}{\hat{y}}=a+\frac{b}{x}$．

究竟选用哪一个为好呢？一个自然的考虑是选用能使 $\sum_{i=1}^{15}(y_i-\hat{y}_i)^2$ 小的曲线为好，$\sum_{i=1}^{15}(y_i-\hat{y}_i)^2$ 小，说明曲线与 15 个点拟合的比较好．例 (6．2．17)选用了对数函数曲线，

$$\sum_{i=1}^{15}(y_i-\hat{y}_i)^2=\sum_{i=1}^{15}[y_i-(6.24+4.084\log x_i)]=2.58\ .$$

读者还可以试试选择其它曲线是否更好．

例 6．2．19　测得13 岁，14 岁，…，19 岁中学生男生的胸围

如下表

表 6－10

x (年龄)	13	14	15	16	17	18	19
y (胸围)	68.4	71.4	75.5	78.9	81.1	82.4	83.6

将 (x_i, y_i)对应的点在图 6－9 标出后可以看到 Ey 与 x 的关系可以用一段二次曲线来反映

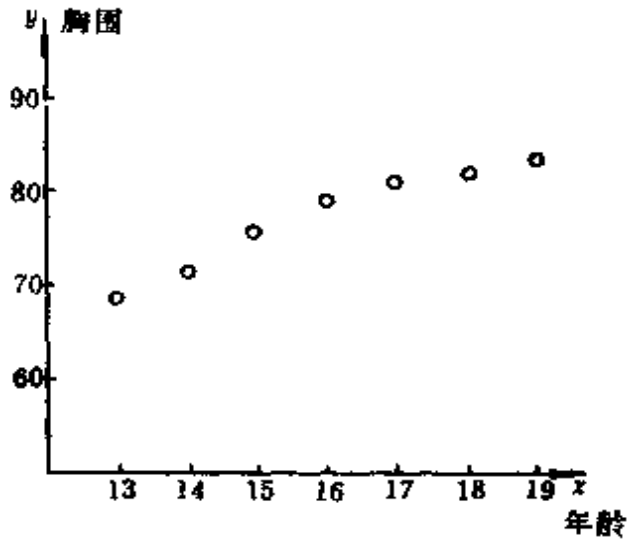


图 6－9

所以可以考虑求一个如下形式的回归方程

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x+\hat{\beta}_2 x^2. \tag{6.2.20}$$

将 x^2 看作一个变量，记为 x_2，再记 x 为 x_1，这样求 $\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\beta}_2$ 的问题就变为根据 7 组数据 $(x_{t1}, x_{t2}; y_t)$，$t=1, 2, \cdots, 7$ 求 β_0，β_1，β_2 的最小二乘估计的问题．

x_1	13	14	15	16	17	18	19
$x_2=x_1^2$	169	196	225	256	289	324	361
y	68.4	71.4	75.5	78.9	81.1	82.4	83.6

以下的计算步骤与例 6.2.16 完全相同，请读者完成．

习题 6－6

1．用例6．2．17中表6－7的15组数据求形如下形式的回归方程

$$\hat{y}=ae^{\frac{b}{x}}$$

并与回归方程$\hat{y}=a+b\log x$比较好坏．

2．完成例6．2．19的计算．

§6．3 正交试验

在§6．1讲方差分析方法时，用做试验及对试验结果做方差分析的方法，考察了灯丝材料对灯泡的使用寿命是否有影响．称灯丝材料为试验因子，四种不同的灯丝材料A_1，A_2，A_3，A_4为试验因子的四个水平．在这个例子中只有一个试验因子，是因子试验中最简单的情形．

有许多场合需要同时考察几个试验因子对某个指标y是否有显著影响，这时也可以让每个试验因子选择若干个水平，然后让每个因子的各个水平搭配做试验，必要时还要使同一个试验重复做几次，最后对试验结果做方差分析．但这样安排需要做的试验很多，有时需要耗费大量的人力、物力、时间以至使试验不可能进行．本节将要介绍的正交试验法是一种安排有多个试验因子的试验方法，在一定的条件下，用这种方法安排试验可以用较少的试验达到试验的目的．

例6．3．1 一个农场希望知道某一个玉米品种的高产栽培条件，他们选择了三个试验因子：种植密度、施化肥量、施化肥时间，每个试验因子选三个水平如表6－11．

表 6－11

因子 / 水平	A 施化肥量 (公斤／亩)	B 密 度 (株／亩)	C 施化肥时间
1	50	3700	按1：2：1分三次施完
2	40	3200	一次施完
3	30	2800	按3：5分两次施完

正交试验法是用正交表安排试验，表 6－12 就是一个正交表，

表 6－12

试验号 \ 列号	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

这个表有 9 行，4 列，每列都由“1”，“2”，“3”3 个数字构成，在每一列中，“1”，“2”，“3”的个数相同，都是 3 个，在任何两列中，同一行上的搭配(1，1)，(1，2)，(1，3)，(2，1)，(2，2)，(2，3)，(3，1)，(3，2)，(3，3)出现的次数也相同，都是 1 次．记这个表为 $L_9(3^4)$，这个记号表示：这是一个正交表，这个表有 9 行，4 列，每列由“1”，“2”，“3”3 个数字构成．

农场用 $L_9(3^4)$表安排试验就是先在表的 4 列中任选 3 列，例如选择第 1，2，3 列，然后将 A，B，C 三个试验因子分别放在这 3 列上见表 6－13．按照这个安排需要做 9 个试验，为此，农场选择了9 块大小、条件尽可能相同的地块种植该玉米品种．第 1 个试验就是按照表的第 1 行 1，2，3 列的数字(1，1，1)在一个地块上种植密度为每亩 3700 株(B_1)，施化肥量为每亩 50 公斤(A_1)，分三次施完(C_1)；第 2 个试验就是按照表的第 2 行 1，2，3 列的数字(1，2，2)在一个地块上种植密度为每亩 3200 株(B_2)，施化肥量为每亩 50 公斤(A_1)，一次施完(C_2)等等．经过试验，试验的结果——折合为每亩的产量已列在表中．

下面谈谈怎样分析这些试验结果以求得好的管理条件．

表 6－13

列号 / 试验号	A 1	B 2	C 3	4	试验结果 亩产量y_i(公斤)
1	1 (50)	1 (3700)	1 (三次)	1	463.5
2	1 (50)	2 (3200)	2 (一次)	2	409
3	1 (50)	3 (2800)	3 (二次)	3	428.5
4	2 (40)	1 (3700)	2 (一次)	3	451.5
5	2 (40)	2 (3200)	3 (二次)	1	435.5
6	2 (40)	3 (2800)	1 (三次)	2	429.5
7	3 (30)	1 (3700)	3 (二次)	2	453.5
8	3 (30)	2 (3200)	1 (三次)	3	426.5
9	3 (30)	3 (2800)	2 (一次)	1	394
K_{1q}	1301	1368.5	1319.5	1293	$G=3891.5$
K_{2q}	1316.5	1271	1254.5	1292	
K_{3q}	1274	1252	1317.5	1306.5	
R_q	42.5	116.5	65	14.5	
S_q	308.39	2604.39	910.89	43.72	

用 K_{pq} 表示 $L_9(3^4)$表中对应于第 q 列中数字 p 的试验结果之和 $(q=1,2,3,4;p=1,2,3)$，例如

$$K_{21}=451.5+435.5+429.5=1316.5,$$

$$K_{34}=428.5+451.5+426.5=1306.5$$

等等．记

$$R_q=\max\{K_{1q},K_{2q},K_{3q}\}-\min\{K_{1q},K_{2q},K_{3q}\},$$

$$q=1,2,3,4.$$

K_{11} 是在 A_1 条件下的 3 个试验结果之和，虽然这 3 个试验也搭配了 B，C 的水平，但 B 的 3 个水平都搭配过 1 次，B 的不同水平对 y 的影响在求和时也就抵消掉了，同样，C 的不同水平对 y 的影响在求和时也被抵消掉了，因此，K_{11} 的大小反映着 A_1 对 y 的影响．同理，K_{21}，K_{31} 的大小分别反映着 A_2，A_3 对 y 的影响．R_1 是 K_{11}，K_{21}，K_{31} 的极差，R_1 大说明 A 的 3 个不同水平使 y 产生

了大的变化，R_1 小说明 A 的 3 个不同水平使 y 的变化不大，所以 R_1 的大小反映了因子 A 对 y 影响的大小．同样，R_2，R_3 的大小分别反映了因子 B，C 对 y 影响的大小．对试验结果做分析，首先可以做直观分析，就是比较 R_1，R_2，R_3 的大小．本例 $R_2 > R_3 > R_1$，这说明种植密度 (B) 是主要因素，其次是施化肥次数 (C)，最后是施化肥量 (A)．就因子 B 的 3 个水平对 y 的影响看，由于 K_{12} 最大，所以找好的就培条件时在 B_1，B_2，B_3 中应选择 B_1．同理，在 C_1，C_2，C_3 中应选择 C_1，在 A_1，A_2，A_3 中应选择 A_2，所以 (B_1，C_1，A_2) 是好的栽培条件．

在上述的直观分析中将极差产生的原因归于试验因子的不同水平．极差会不会只是随机因素引起的呢？在实际工作中如认为有必要，应当对试验结果做进一步的方差分析．

用 G 表示 9 个试验结果的和，本例 $G = 3891.5$．记

$$S_q = \frac{K_{1q}^2 + K_{2q}^2 + K_{3q}^2}{3} - \frac{G^2}{9}, \quad q = 1, 2, 3, 4,$$

分母 3 是每列中每个数字出现的次做，9 是试验次数．将算出的 K_{pq}，G，S_q ($q = 1, 2, 3, 4$，$p = 1, 2, 3$) 填写在表中，每个 S_i 的自由度为 2 (水平数减 1)．

记 $\bar{y} = \frac{1}{9} G$，由于

$$\begin{aligned} S_1 &= \frac{K_{11}^2 + K_{21}^2 + K_{31}^2}{3} - \frac{G^2}{9} \\ &= 3\left[\left(\frac{y_1 + y_2 + y_3}{3}\right)^2 + \left(\frac{y_4 + y_5 + y_6}{3}\right)^2 + \left(\frac{y_7 + y_8 + y_9}{3}\right)^2 \right. \\ &\quad \left. - 3\left(\frac{y_1 + y_2 + \cdots + y_9}{9}\right)^2\right] \end{aligned}$$

$$=3\left[\left(\frac{y_1+y_2+y_3}{3}-\bar{y}\right)^2+\left(\frac{y_4+y_5+y_6}{3}-\bar{y}\right)^2+\left(\frac{y_7+y_8+y_9}{3}-\bar{y}\right)^2\right],$$

所以 S_1 表示试验因子 A 对产量 y 的影响，记为 S_A，在安排试验时，$L_9(3^4)$表中空着的第 4 列对应的 S_4 就是随机误差，仍记为 S_E. 如果试验因子 A 对产量 y 有影响，那么 S_A 相对 S_E 就应该较大. 给定信度 α 后从 F 分布表查出 $F_\alpha(2,2)$，若 $S_A/S_E>F_\alpha(2,2)$，就认为因子 A 对产量 y 有显著影响，否则认为没有显著影响. 类似地，令 $S_B=S_2$，$S_C=S_3$，若 $S_B/S_E>F_\alpha(2,2)$，就认为因子 B 对产量 y 有显著影响，否则认为没有显著影响. 若 $S_C/S_E>F_\alpha(2,2)$，就认为因子 C 对产量 y 有显著影响，否则认为没有显著影响.

本例取 $\alpha=0.05$，$F_{0.05}(2,2)=19$，

$$F_A=\frac{S_A}{S_E}=\frac{308.39}{43.72}\doteq 7.05<F_{0.05}(2,2),$$

$$F_B=\frac{S_B}{S_E}=\frac{2604.39}{43.72}\doteq 59.57>F_{0.05}(2,2),$$

$$F_C=\frac{S_C}{S_E}=\frac{910.89}{43.72}\doteq 20.83>F_{0.05}(2,2).$$

从方差分析的结果可以认为：

(1) 对产量来说，施化肥量的三个水平没有明显的差别.

(2) 种植密度的三个水平是有明显差别的，对应于三个水平 B_1，B_2，B_3 的平均亩产量分别是 456.2，423.7，417.3，B_1 最好.

(3) 施化肥时间的三个水平有明显的差别，对应于三个水平 C_1，C_2，C_3 的平均亩产量分别是 439.8，418.2，439.2，C_1 最好.

所以从获取高产量考虑，应取密度 3700 株/亩，化肥分三次就完，从降低成本考虑每亩施化肥 30 公斤.

在这个例中因子 A，B，C 分别放在 $L_9(3^4)$表的 1，2，3 列上了，放在 2，3，4 列或 2，1，3 列也可以.

本例如果安排完全试验，即让三个试验因子的每个水平都搭

配做试验，需要做27个试验．现在用 $L_9(3^4)$ 表安排试验只做了其中的9个试验就找到了好的栽培条件：A_3，B_1，C_1，而且这个条件下的试验并没有做．

例6．3．2 制备某种胶粘剂需要考虑三个试验因子 A，B，C 对其转化率的影响．现在每个因子取4个水平，用正交表 $L_{16}(4^5)$ 安排试验，见表6－14．

表 6－14

试验号 \ 列号	A 1	B 2	C 3	4	5	试验结果 转化率y_i (%)
1	1	1	1	1	1	90.0
2	1	2	2	2	2	83.6
3	1	3	3	3	3	89.0
4	1	4	4	4	4	83.2
5	2	1	2	3	4	87.0
6	2	2	1	4	3	92.7
7	2	3	4	1	2	86.1
8	2	4	3	2	1	90.6
9	3	1	3	4	2	88.2
10	3	2	4	3	1	83.1
11	3	3	1	2	4	89.7
12	3	4	2	1	3	83.2
13	4	1	4	2	3	81.6
14	4	2	3	1	4	86.6
15	4	3	2	4	1	82.2
16	4	4	1	3	2	89.1
K_{1q}	345.8	346.8	361.5	345.9	345.9	$G=1385.9$
K_{2q}	356.4	346.0	336.0	345.5	347.0	$\frac{G^2}{16}=120044.9$
K_{3q}	344.2	347.0	334.4	348.2	346.5	
K_{4q}	339.5	346.1	334.0	346.3	346.5	
S_q	38.22	0.21	138.50	1.10	0.18	

正交表 $L_{16}(4^5)$ 有16行，5列，每列都由“1”，“2”，“3”，“4”4个数字构成，在每一列中，“1”，“2”，“3”，“4”的个数相同，都是4

个，在任何两列中同一行上的(1，1)，(1，2)，(1，3)，(1，4)，(2，1)，…，(4，3)，(4，4)等16个搭配出现的次数也相同，都是1次.

用$L_{16}(4^5)$表安排试验只要将试验因子A，B，C安排在表的三个列上试验就安排好了，如表6－14，本例选择了第1，2，3列安排试验因子，需要做16个试验，第1个试验在条件A_1(A的第1水平)，B_1(B的第1水平)，C_1(C的第1水平)之下进行，第2个试验在条件A_1，B_2，C_2之下进行等等，16个试验的结果y_i列在表的最后一列，和例6.3.1一样，用K_{pq}表示对应于第q列数字p的试验结果之和($q=1,2,3,4,5$；$p=1,2,3,4$)，G表示16个试验结果之和，

$$S_q=\frac{K_{1q}^2+K_{2q}^2+K_{3q}^2+K_{4q}^2}{4}-\frac{G^2}{16},\quad q=1,2,3,4,5,$$

其中分母4是每列中每个数字出现的次数，16是试验次数，每个S_i的自由度是3(水平数4减去1)，随机误差$S_E=S_4+S_5$，自由度是$3+3=6$，$S_A=S_1$，$S_B=S_2$，$S_C=S_3$，给定信度$\alpha=0.05$，查F分布表知$F_{0.05}(3,6)=4.76$，由于

$$F_A=\frac{6}{3}\cdot\frac{S_A}{S_E}=59.72>4.76,$$

$$F_B=\frac{6}{3}\cdot\frac{S_B}{S_E}=0.39<4.76,$$

$$F_C=\frac{6}{3}\cdot\frac{S_C}{S_E}=216.41>4.76,$$

由方差分析的结果可以认为因子B对转化率没有显著影响，因子A，C对转化率有显著影响.

例6.3.3 一个工厂想考察补强剂种类(试验因子A)，促进剂种类(试验因子B)，防老剂种类(试验因子C)，补强剂用量(试验因子D)和促进剂用量(试验因子H)对胶料冲击弹性y的影响，每个试验因子选两个水平，用正交表$L_8(2^7)$安排试验见表

6－15，试验结果 y_i 列在表的最后一列．表中 K_{1q} 表示对应于第 q 列数字 1 的试验结果之和，K_{2q} 表示对应于第 q 列数字 2 的试验结果之和，G 表示 8 个试验结果之和，

$$S_q=\frac{K_{1q}^2+K_{2q}^2}{4}-\frac{G^2}{8}=\frac{K_{1q}^2+K_{2q}^2}{4}-\frac{(K_{1q}+K_{2q})^2}{8}=\frac{(K_{1q}-K_{2q})^2}{8},$$

$$q=1,2,\cdots,7,$$

其中分母 4 是每列中每个数字出现的次数，8 是试验次数，每个 S_i 的自由度是 1（水平数 2 减去 1）．随机误差 $S_E=S_3+S_7$，自由度是 2，$S_A=S_1$，$S_B=S_2$，$S_C=S_4$，$S_D=S_5$，

表　6－15

列号 / 试验号	A 1	B 2		C 4	D 5	H 6	7	试验结果 y_i
1	1	1	1	1	1	1	1	63
2	1	1	1	2	2	2	2	48
3	1	2	2	1	1	2	2	62
4	1	2	2	2	2	1	1	46
5	2	1	2	1	2	1	2	45
6	2	1	2	2	1	2	1	44
7	2	2	1	1	2	2	1	44
8	2	2	1	2	1	1	2	45
K_{1q}	219	200	200	214	214	199	197	$G=397$
K_{2q}	178	197	197	183	183	198	200	
S_q	210.125	1.125	1.125	120.125	120.125	0.125	1.125	

$S_H=S_6$，给定信度 $\alpha=0.05$，查 F 分布表知 $F_{0.05}(1,2)=18.51$．由于

$$F_A=\frac{2S_A}{S_E}=\frac{420.25}{2.25}\doteq 186.8>F_{0.05}(1,2),$$

$$F_B=\frac{2S_B}{S_E}=\frac{2.25}{2.25}=1<F_{0.05}(1,2),$$

$$F_C=\frac{2S_C}{S_E}=\frac{240.25}{2.25}\doteq 106.8>F_{0.05}(1,2),$$

$$F_D = \frac{2S_D}{S_E} = \frac{240.25}{2.25} \doteq 106.8 > F_{0.05}(1,2),$$

$$F_H = \frac{2S_H}{S_E} = \frac{0.25}{2.25} \doteq 0.11 < F_{0.05}(1,2)$$

可以认为试验因子 B 及 H 对胶料冲击弹性没有显著影响，试验因子 A，C 及 D 对胶料冲击弹性有显著影响.

这个例子只用了 8 次试验就做出了五个试验因子中有三个对 y 有影响，两个对 y 无影响的判断，原因是设计试验时每个试验因子只选了两个水平.试验因子选用的水平少，总的试验次数也就少了.但是要从这个试验中找一个好的水平组合可能会感到 A，C，D 三个试验因子水平选的太少了.因此，如果需要可以再安排包含 A，C，D 三个试验因子的试验，而且每个因子的水平可以多一点.由于要考察的试验因子已集中到三个，可以用比较少的试验找到好的水平组合，例如，A，C，D 各选四个水平用 $L_{16}(4^5)$ 表安排试验.

以上我们讲了什么叫正交表，怎样用正交表安排试验，对用正交试验得到的试验结果怎样做直观分析及方差分析等等.在结束这一节时再说明两点：

第一，正交试验法的内容很丰富，用途很广泛，我们只介绍了试验因子之间没有交互作用时怎样安排试验，因子之间没有交互作用指每个因子只对指标 y 单独起作用，对于有交互作用的情形和其它情形（例如因子的水平数不同）如何安排试验，有兴趣的读者可参看专门讲正交试验法的书，例如参考书[10].

第二，本节对正交试验的方差分析只讲了方法，没有讲方法的证明.方法的证明涉及到正交表的构造，介绍这部分内容将超出本书的范围，有兴趣的读者可参看试验的设计与分析方面的书，例如[10].

习题 6－6

某厂为了提高某种化工产品的转化率，用 $L_9(3^4)$表安排了包含三个试验因子的试验．试验因子是反应温度 A，反应时间 B，用碱量 C．各因子选取的水平见表 6－16．

表 6－16

试验因子 / 水平	A(℃)	B(分)	C(%)
1	80	90	5
2	85	120	6
3	90	150	7

用 $L_9(3^4)$表安排试验的情况及试验结果见表 6－17．

表 6－17

试验号 / 列号	A 1	B 2	C 3	4	转化率
1	1	1	1	1	31
2	1	2	2	2	54
3	1	3	3	3	38
4	2	1	2	3	53
5	2	2	3	1	49
6	2	3	1	2	42
7	3	1	3	2	57
8	3	2	1	3	62
9	3	3	2	1	64

试用方差分析方法推断试验因子 A，B，C 对转化率有无显著影响，信度取 0.05．

习 题 答 案

第 一 章

习题 1-1 1.(1){1,2,4,5,6,7,8,10,11,12,13,16,17,19,22,23,24,26}; (2){2,4,5,6,10,11,12,16,…,27}; (3){10,…,15,22,…,27}. 2.(1)三个盒中各有一球; (2)第二个盒中没有球; (3)第三个盒中至少有两个球. 3. $\Omega=\{(a_1,a_2),(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,a_1),(a_2,b_1),(a_2,b_2),(a_2,b_3),(b_1,a_1),(b_1,a_2),(b_1,b_2),(b_1,b_3),(b_2,a_1),(b_2,a_2),(b_2,b_1),(b_2,b_3),(b_3,a_1),(b_3,a_2),(b_3,b_1),(b_3,b_2)\}$; $A_0=\{(b_1,b_2),(b_1,b_3),(b_2,b_1),(b_2,b_3),(b_3,b_1),(b_3,b_2)\}$; $P(A_0)=\frac{6}{20}=\frac{3}{10}$; $A_1=\{(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,b_1),(a_2,b_2),(a_2,b_3),(b_1,a_1),(b_2,a_1),(b_3,a_1),(b_1,a_2),(b_2,a_2),(b_3,a_2)\}$; $P(A_1)=\frac{12}{20}=\frac{3}{5}$; $A_2=\{(a_1,a_2),(a_2,a_1)\}$; $P(A_2)=\frac{2}{20}=\frac{1}{10}$; $B_1=\{(a_1,a_2),(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,a_1),(a_2,b_1),(a_2,b_2),(a_2,b_3)\}$; $P(B_1)=\frac{8}{20}=\frac{2}{5}$. 4. $\Omega=\{(a_1,a_1),(a_1,a_2),(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,a_1),(a_2,a_2),(a_2,b_1),(a_2,b_2),(a_2,b_3),(b_1,a_1),(b_1,a_2),(b_1,b_1),(b_1,b_2),(b_1,b_3),(b_2,a_1),(b_2,a_2),(b_2,b_1),(b_2,b_2),(b_2,b_3),(b_3,a_1),(b_3,a_2),(b_3,b_1),(b_3,b_2),(b_3,b_3)\}$; $A_0=\{(b_1,b_1),(b_1,b_2),(b_1,b_3),(b_2,b_1),(b_2,b_2),(b_2,b_3),(b_3,b_1),(b_3,b_2),(b_3,b_3)\}$; $P(A_0)=\frac{9}{25}$; $A_1=\{(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,b_1),(a_2,b_2),(a_2,b_3),(b_1,a_1),(b_2,a_1),(b_3,a_1),(b_1,a_2),(b_2,a_2),(b_3,a_2)\}$; $P(A_1)=12/25$; $A_2=\{(a_1,a_1),(a_1,a_2),(a_2,a_1),(a_2,a_2)\}$; $P(A_2)=4/25$; $B_1=\{(a_1,a_1),(a_1,a_2),(a_1,b_1),(a_1,b_2),(a_1,b_3),(a_2,a_1),(a_2,a_2),(a_2,b_1),(a_2,b_2),(a_2,b_3)\}$; $P(B_1)=\frac{10}{25}=\frac{2}{5}$; $C=\{(a_1,a_1),(a_2,a_2),(b_1,b_1),(b_2,b_2),(b_3,b_3)\}$; $P(C)=\frac{5}{25}=\frac{1}{5}$. 5. 用$a_1,a_2$表示两件次品,

b_1, b_2, b_3 表示三件正品，$\Omega=\{\{a_1,a_2\},\{a_1,b_1\},\{a_1,b_2\},\{a_1,b_3\},\{a_2,b_1\},\{a_2,b_2\},\{a_2,b_3\},\{b_1,b_2\},\{b_1,b_3\},\{b_2,b_3\}\}$；$A_0=\{\{b_1,b_2\},\{b_1,b_3\},\{b_2,b_3\}\}$；$P(A_0)=3/10$；$A_1=\{\{a_1,b_1\},\{a_1,b_2\},\{a_1,b_3\},\{a_2,b_1\},\{a_2,b_2\},\{a_2,b_3\}\}$；$P(A_1)=\frac{6}{10}=\frac{3}{5}$；$A_2=\{\{a_1,a_2\}\}$；$P(A_2)=1/10$. 6. $\Omega=\{123,124,132,134,142,143,213,214,231,234,241,243,312,314,321,324,341,342,412,413,421,423,431,432\}$. (1)1/2；(2)1/4.

习题1－2 1. 1/15. 2. 5/33. 3. 41/90. 4. $\binom{M-1}{m-1}\binom{N-M}{n-m}\Big/\binom{N}{n}$. 5. $60/7^6$. 6. 207/625. 7. (1)0.504；(2)0.432；(3)0.036；(4)0.001；(5)0.063.

习题1－3 1. $(k-1)(n-k)\Big/\binom{n}{2}$. 2. 0.218. 3. (1)0.0048；(2)k 不超过2；(3)$N=12$. 4. 2.26×10^{-7}. 5. 0.0039.

习题1－4 1. 2/5. 2. (1)3/4；(2)1/4；(3)7/8；(4)0.5966. 3. (1)1/3；(2)1/2. 4. $(a-r)/a$. 5. 0.121.

习题1－6 1. (1)Ⅰ盒中至少有一球；(2)两球都在Ⅱ盒中；(3)Ⅲ盒中有红球没有白球；(4)白球不在Ⅱ盒中；(5)红球在Ⅰ盒中，白球不在Ⅱ盒中. 2. (1)A_3B_3；(2)$B_2\setminus A_2$；(3)A_3；(4)$A_1^cB_1^c$. 3. (1)(2)(4)(5)正确；(3)(6)不正确. 4. $B_1=A_1\cup A_2A_3\cup A_4$；$B_2=[A_1(A_2A_3\cup A_5)\cup A_6(A_3\cup A_2A_5)]A_4$. 5. (1)$ABC^c\cup AB^cC\cup A^cBC$；(2)$(ABC)^c$.

习题1－7 1. (1)5/6；(2)0；(3)2/3；(4)1/2；(5)1/6. 3. 0.26.

复习题1－A 1. $m/(n+m)$. 2. (1)$[k^n-(k-1)^n]/N^n$；(2)$[(N-k+1)^n-(N-k)^n]/N^n$. 3. (1)$p_1=1/\binom{6}{2}$；(2)$p_2=(n-3)/n$；(3)$p_3=1-\binom{5}{3}/\binom{6}{2}$. 4. 7/33. 5. 16/33. 6. $P(\xi=2)=\frac{1}{10}$；$P(\xi=3)=\frac{1}{5}$；$P(\xi=4)=\frac{3}{10}$；$P(\xi=5)=\frac{2}{5}$. 7. 若规定三天接见来访，12次恰巧发生在其中指定的两天的概率为0.008，如规定每天都接见来访，12次都不在星期日的概率为0.157.

8. $1-\frac{1}{2}\left[\left(1-\frac{t}{T}\right)^2+\left(1-\frac{\Delta}{T}\right)^2\right]$. 9. (2) $a(2-a)$; a^2; $a^2/2$. 10. (1) $l^2/\pi ab$; (2) $[2(a+b)l-l^2]/\pi ab$.

复习题1-B 1. (1) $\binom{N}{n}\Big/N^n$; (2) $\binom{n+N-1}{n}\Big/N^n$. 2. $P(\xi=12)=\frac{25}{216}$, $P(\xi=11)=\frac{27}{216}$. 3. $\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}\left(\frac{n-k}{n}\right)^N$. 5. $\frac{a+b+c}{\pi d}$.

6 $\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^n\frac{1}{n!}$. 7. $\frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}$. 8. (1) $1-\frac{4(3^n+1)+6\times 2^n}{4^n}$;

(2) $\left(\frac{3}{4}\right)^{n-1}-3\left(\frac{1}{2}\right)^{n-1}+3\left(\frac{1}{4}\right)^{n-1}$.

第 二 章

习题2-1 1. 1/8. 2. (1) $\frac{m-1}{2M-m-1}$; (2) $\frac{2m}{M+m-1}$;

(3) $\frac{m(2M-m-1)}{M(m-1)}$. 4. $\left(\frac{b}{a+b}\right)^N\left(\frac{b+1}{a+b+1}\right)^{N-1}\cdots\left(\frac{b+N-2}{a+b+N-2}\right)^2\cdot\left(\frac{b+N-1}{a+b+N-1}\right)$. 5. $\frac{1}{6}$.

习题2-2 1. $\frac{9}{14}$. 2. $\frac{13}{20}$. 3. 0.26. 4. $\frac{1}{3}$. 5. 0.9978.

习题2-3 1. 0.5. 2. (1) $\prod_{i=1}^{n}(1-p_i)$; (2) $1-\prod_{i=1}^{n}(1-p_i)$; (3) $\sum_{j=1}^{n}p_j\prod_{i\neq j}(1-p_i)$. 3. ≈ 0.93 4. (1) ≈ 0.02; (2) ≈ 0.16; (3) ≈ 0.86.

习题2-4 1. ξ能取的值是2,3,…,12; $P(\xi=i)=P(\xi=14-i)=\frac{i-1}{36}$, $i=2,3,4,5,6$; $P(\xi=7)=\frac{1}{6}$. 2. $P(\xi=k)=\frac{(5-k)(5-k-1)}{20}$, $k=1,2,3$. 4. $P(\xi=i)=P(\eta=i)=\frac{1}{5}$, $i=1,2,3,4,5$.

ξ	1	2	3	4	5	6	8	9	10	12	15	16	20	25
$P(\xi=k)$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{1}{25}$

习题 2－5 1.(1)0.35; (2)0.68. 2.0.972. 3.(1)0.003; (2)0.9994. 4. $[(n+1)p]$ $\left((n+1)p\text{不是整数时}\right)$; $(n+1)p$, $(n+1)p-1$ $\left((n+1)p\text{是整数时}\right)$. 5.提示:假定该食品对降低胆固醇含量不起作用,则吃食品后降低含量的概率为 0.5,然后使用二项分布的正态近似得 0.0013.

习题 2－6 1.提示:利用条件概率的定义及事件 $\{X_{n+1}=i+1, X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}, X_n=i\}=\{\xi_{n+1}=1, X_0=i_0, \xi_1=i_1-i_0, \cdots, \xi_{n-1}=i_{n-1}-i_{n-2}, \xi_n=i-i_{n-1}\}$及 $X_0, \xi_1, \cdots, \xi_n, \xi_{n+1}$ 的独立性.

复习题 2－A 1. $\frac{1}{2}$. 2. $\frac{35}{768}$. 3.(1) $\frac{3}{5}$; (2) $\frac{3}{5}$; (3) $\frac{3}{10}$. 4.(1) $\frac{5}{8}$; (2) $\frac{3}{5}$. 5. $\frac{b}{b+r+c}$. 6. $\frac{38}{105}$. 7. $\frac{23}{50}$. 8. $\frac{m(m-1)(m-2)}{(m+n)(m+n-1)(m+n-2)}$. 12.(1) $\frac{5}{3888}$; (2) $\frac{5}{72}$. 13.0.0093. 14.105. 15. $\frac{10}{27}$. 16.

$$P=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 & 1 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} & 2 \\ 0 & 0 & 1 & 0 & 3 \end{pmatrix}.$$

复习题 2－B 1. $\frac{42}{125}$. 2.若 ξ_n, η_n 分别表示前 n 次抛掷中甲、乙二人得到的正面次数,A 表示甲抛出的第 $n+1$ 枚硬币出现正面,则所求概率为 $P(A\cap\{\xi_n=\eta_n\}\cup\{\xi_n>\eta_n\})$. 3.应用公式 $\sum_{i=0}^{\infty}\binom{k+i}{k}\left(\frac{p}{2}\right)^i=\left(1-\frac{p}{2}\right)^{-(k+1)}$.

4.设 $B_n=$"第 n 次取到白球",则 $P(B_n)=\left(\frac{a}{a+b}\right)^2+\left(\frac{b}{a+b}\right)^2$, $n>1$; $P(B_1)=\frac{a}{a+b}$. 5.甲获胜的概率为 $\frac{p_1}{p_1+p_2-p_1p_2}$,乙获胜的概率为 $\frac{p_2-p_1p_2}{p_1+p_2-p_1p_2}$. 7. $\sum_{k=m}^{\infty}\binom{n+k-1}{k}p^n(1-p)^k$. 8. $\frac{1}{2}\left[1+\left(1-\frac{2}{N}\right)^n\right]$; $\frac{1}{2}$.

9.(1)甲、乙赌本分配的比应为 $\sum_{k=0}^{n-1}\binom{m+n-1}{k}\Big/\sum_{k=0}^{m-1}\binom{m+n-1}{k}$.

第 三 章

习题 3-1 1. $P(\xi=k)=\binom{b}{k}\binom{a}{r-k}\Big/\binom{a+b}{r}$, $k=0,1,\cdots,r$. 2. $P(\xi=k)=(6-|k|)/36$; k 取 $[-5,5]$ 内的一切整数值; $P(2\leqslant\xi\leqslant4)=\frac{1}{4}$. 3. (1)$\frac{1}{5}$; (2)$\frac{1}{5}$; (3)$\frac{3}{5}$. 4. (1)$c=1$; (2)$c=(e^{\lambda}-1)^{-1}$; (3)$\frac{27}{38}$. 5. (1)$P(\eta=0)=\frac{q}{1+q}$, $P(\eta=1)=\frac{1}{1+q}$; (2)$P(\eta=-1)=q^3/[(1+q)\cdot(1+q^3)]$, $P(\eta=0)=(1+q)^{-1}$, $P(\eta=1)=q/(1+q)(1+q^2)$. 6. (1)$P(\xi=k)=0.8^{k-1}\times0.2$, $k=1,2,\cdots$; (2)射击一次的可能性最大. 7. (1)$P(\xi=k)=\binom{4}{k}\times0.1^k\times0.9^{4-k}$, $k=0,1,2,3,4$; (2)1.

习题 3-2 1. (1)$E\xi=2$, $D\xi=2$; (2)$E\xi=a$, $D\xi=a+a^2$. 2. $E\xi=4$; $D\xi=1$. 3. (1)$E\xi=3$, $D\xi=2$; (2)$E\xi=5$, $D\xi=20$. 4. (1)$E\xi=\sum_{k=1}^{N}\left(\frac{k}{N}\right)^n$, $D\xi=\sum_{k=1}^{N}[2(N-k)+1]\left(\frac{k}{N}\right)^n-\left[\sum_{k=1}^{N}\left(\frac{k}{N}\right)^n\right]^2$; (2)$E\xi=\sum_{k=n}^{N}\binom{k}{n}\Big/\binom{N}{n}$; $D\xi=\sum_{k=0}^{N-n}(2k+1)\binom{N-k}{n}\Big/\binom{N}{n}-\left[\sum_{k=n}^{N}\binom{k}{n}\Big/\binom{N}{n}\right]^2$.

5. (1) (2)$2^{-(2n+1)}$.

$$P(\xi>n)=\begin{cases}1 & n<0\\ \frac{3}{4} & n=0\\ \frac{1}{4} & n=1\\ 0 & n\geqslant2;\end{cases}$$

习题 3-3 1. (1)$c=\frac{1}{\pi}$; (2)$c=\frac{1}{2}$; (3)$c=\frac{4}{\sqrt{\pi}}h^3$. 2. (1)$a=7.95$; (2)$a=2.58$. 3. $P(\xi>80)=0.0272$; $P(\xi>90)=0.0037$. 4. (1)$P(\xi<$

$70)=\Phi(2)<\Phi(2.5)=P(\eta<70)$; (2) $P(\xi<65)=\Phi(1.5)>\Phi(1.25)=P(\eta<65)$. **5**. (1) 0.6294; (2) ξ以0.9973的概率落在(2.47, 2.53)内.

6. $\frac{3}{5}$. **7**. (1) $c=\frac{1}{\sqrt{2\pi}\,\sigma}$; (2) $c=\frac{1}{B(p,q)}=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}$;

(3) $c=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}$.

习题 3-4 **1**. (1) $E\xi=1$, $D\xi=\frac{1}{6}$; (2) $E\xi=b$, $D\xi=2a^2$; (3) $E\xi=0$, $D\xi=\frac{\pi^2}{12}-\frac{1}{2}$. **2**. $\frac{11}{5\sqrt{5}}$. **3**. $E\xi=n+1$, $D\xi=n+1$. **4**. $E\xi=e^{\mu+\frac{\sigma^2}{2}}$; $D\xi=\exp\{2\mu+\sigma^2(e^{\sigma^2}-1)\}$. **5**. (1) $E\xi^3=0$, $E\xi^4=\frac{1}{5}$; (2) $E\xi^3=\frac{1}{4}$, $E\xi^4=\frac{1}{5}$. **6**. (1) $E\xi=\frac{p}{p+q}$, $D\xi=\frac{pq}{(p+q)^2(p+q+1)}$; (2) $E\xi=\lambda^{-\frac{1}{a}}\Gamma(1+\frac{1}{a})$, $D\xi=\lambda^{-\frac{2}{a}}\left[\Gamma(1+\frac{2}{a})-\Gamma^2(1+\frac{1}{a})\right]$; $E\xi=0\ (n>1)$, $D\xi=\frac{n}{n-2}\ (n>2)$.

习题 3-5 **1**. $P(-2<\xi\leqslant 1)=0.8$; $P(|\xi|<2)=0.8$.

2.
$$F(x)=\begin{cases}0, & x<0,\\ 0.0016, & 0\leqslant x<1,\\ 0.0272, & 1\leqslant x<2,\\ 0.1808, & 2\leqslant x<3,\\ 0.5904, & 3\leqslant x<4,\\ 1, & x\geqslant 4,\end{cases}$$
$P(\xi<3)=0.1808$, $P(1\leqslant\xi<4)=0.5888$, $P(1\leqslant\xi\leqslant 3)=0.5888$.

3. (1) $c=\frac{6}{29}$, $F(x)=\begin{cases}0, & x<1,\\ \frac{2}{29}(x^3-1), & 1\leqslant x<2,\\ \frac{1}{29}(3x^2+2), & 2\leqslant x<3,\\ 1, & x\geqslant 3.\end{cases}$

(2) $c=\frac{\lambda}{2}$，$F(x)=\begin{cases}\frac{1}{2}e^{\lambda x}, & x\leqslant 0,\\ 1-\frac{1}{2}e^{-\lambda x}, & x>0.\end{cases}$

(3) $c=3$；$F(x)=\begin{cases}0, & x<\frac{\pi}{6},\\ -\cos 3x, & \frac{\pi}{6}\leqslant x<\frac{\pi}{3},\\ 1, & x\geqslant\frac{\pi}{3}.\end{cases}$

4. (1) $f(x)=\begin{cases}0, & x\leqslant 0,\\ xe^{-x}, & x>0;\end{cases}$ (2) $f(x)=\begin{cases}0, & x\leqslant 0 \text{ 或 } x>\frac{\pi}{2},\\ \cos x, & 0<x\leqslant\frac{\pi}{2};\end{cases}$

(3) $f(x)=\begin{cases}0, & x\leqslant 0,\\ (1+\frac{x}{2})^{-3}, & x>0;\end{cases}$

习题 3-6 2. (1) $f(x)=\begin{cases}2\lambda x e^{-\lambda x^2}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$

(2) $f(x)=\lambda^2 e^{\lambda x}e^{-\lambda e^{\lambda x}}$；(3) $f(x)=\begin{cases}\lambda x^{\lambda-1}, & 0<x<1,\\ 0, & \text{其它}.\end{cases}$

3. 设圆周长为 η、圆面积为 ζ，则 $f_\eta(x)=\frac{1}{\pi\sqrt{2\pi}}e^{-2(\frac{x}{2\pi}-1000)^2}$；

$$f_\zeta(x)=\begin{cases}\frac{1}{2\pi\sqrt{x}}\left[e^{-2(\sqrt{\frac{x}{\pi}}+1000)^2}+e^{-2(\sqrt{\frac{x}{\pi}}-1000)^2}\right], & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

4. $f(x)=\begin{cases}\frac{\pi}{\sin\pi x}, & \frac{1}{2}<x<\frac{2}{\pi}\operatorname{arctg} e,\\ 0, & \text{其它}.\end{cases}$

5. $E\eta=\frac{\pi^2}{4}-2$，$D\eta=20-2\pi^2$.

6. $E\xi^3=3\sigma^2\mu+\mu^3$，$D\xi^3=\sigma^6+6\sigma^4\mu^2+9\sigma^2\mu^4$.

习题 3-7　1. $E(\xi-\eta^2)=-\frac{1}{3}$，$D(\xi-\eta^2)=\frac{154}{45}$．2. $E\overline{\xi}=a$；$D\overline{\xi}=\frac{\sigma^2}{n}$；3. $E\eta=\sum_{i=1}^{n}p_i$；$D\eta=\sum_{i=1}^{n}p_i(1-p_i)$．4. $E\eta=\frac{k(n+1)}{2}$；$D\eta=\frac{k(n^2-1)}{12}$．5. $P(40\leqslant\xi\leqslant60)\geqslant0.75$．

复习题 3-A　2. 16．3. 首项$a=\frac{1}{10}$，公比$b=\frac{9}{10}$；$1-\left(\frac{9}{10}\right)^{10}$

4. $1-e^{-1}$．5. 负二项分布的母函数$f(x)=p^r(1-xq)^{-r}$；$E\xi=\frac{rq}{p}$；$D\xi=\frac{rq}{p^2}$．

6. (1) $\xi\sim\Gamma(2,\frac{\alpha}{2})$，

$$f_\xi(x)=\begin{cases}\frac{\alpha^2}{4}xe^{-\frac{\alpha}{2}x}, & x\geqslant0,\\ 0, & x<0;\end{cases}$$

(2) $(1+\frac{5}{2}\alpha)e^{-\frac{5}{2}\alpha}$；(3) $1-[1-(1+5\alpha)e^{-5\alpha}]^4$．7. $\frac{2}{3}$；$\frac{1}{3}$．8. $f_\xi(x)=\begin{cases}\frac{2x}{R^2}, & 0\leqslant x\leqslant R,\\ 0, & \text{其他};\end{cases}$

$$F_\xi(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{R^2}, & 0\leqslant x<R,\\ 1, & x\geqslant R;\end{cases}\qquad P(\frac{R}{3}<\xi\leqslant\frac{R}{2})=\frac{5}{36}．$$

9. (1) $h=\frac{2}{a}$；

$$f(x)=\begin{cases}\frac{2}{a}-\frac{2}{a^2}x, & 0\leqslant x\leqslant a,\\ 0, & \text{其它};\end{cases}\qquad (2)\ F(x)=\begin{cases}0, & x<0,\\ \frac{2}{a}x-\frac{x^2}{a^2}, & 0\leqslant x<a;\\ 1, & x\geqslant a;\end{cases}$$

(3) $P(\frac{a}{2}<\xi\leqslant a)=\frac{1}{4}$．

10. (1) $f(x)=\begin{cases}\frac{1}{\pi\sqrt{R^2-x^2}}, & -R<x<R,\\ 0, & \text{其它};\end{cases}$

(2) $f(x)=\begin{cases}\frac{2}{\pi}\frac{1}{\sqrt{(2R)^2-x^2}}, & 0<x<2R,\\ 0, & \text{其它};\end{cases}$

11. $\frac{4R}{\pi}$. 13. (2)不少于250次.

复习题3-B 2.(1) $\frac{n}{N}$; (2) $\frac{2}{N(N-1)}(Nn-\sigma^2-n^2)$. 3.(1) $1-(1-p)^k$; (2) $\frac{N}{k}+N(1-q^k)$; (3)k取使 $q^k p \geqslant \frac{1}{k(k+1)}$ 的最小正整数.

第四章

习题4-1 1.(1) $F(b-,y)-F(a-,y)$; (2) $F(a,y-)-F(a-,y)$; (3) $F(b-,d-)-F(b-,c-)-F(a-,d-)+F(a-,c-)$. 2.(1)1;

(2) $\frac{3}{128}$; (3) $F_\xi(x)=\begin{cases}0, & x<0,\\ 1-2^{-x}, & x\geqslant 0,\end{cases}$ $F_\eta(y)=\begin{cases}0, & y<0,\\ 1-2^{-y}, & y\geqslant 0.\end{cases}$

3. $A=\frac{1}{\pi^2}$; $B=C=\frac{\pi}{2}$; $F_\xi(x)=\frac{1}{\pi}\operatorname{arctg}\frac{x}{2}+\frac{1}{2}$; $F_\eta(y)=\frac{1}{\pi}\operatorname{arctg}\frac{y}{3}+\frac{1}{2}$.

习题4-2 1.(1) $P(\xi=0)=0.3$, $P(\xi=1)=0.7$; $P(\eta=-1)=0.3$, $P(\eta=0)=0.5$, $P(\eta=1)=0.2$; (2)0.5; (3)不独立.

2.(1),(2)

ξ \ η	0	1	2	$P(\xi=i)$
0	1/84	1/14	1/28	5/42
1	1/7	2/7	1/21	10/21
2	3/14	1/7	0	5/14
3	1/21	0	0	1/21
$P(\eta=j)$	5/12	1/2	1/12	

(3)1/12; (4)不独立.

3.(1),(2)

ξ \ η	0	1	2	3	$P(\xi=i)$
0	1/8	1/8	0	0	1/4
1	0	1/4	1/4	0	1/2
2	0	0	1/8	1/8	1/4
$P(\eta=j)$	1/8	3/8	3/8	1/8	

(3)不独立.

4.(1) $P(\xi=i,\eta=j)=\begin{cases}\frac{1}{36}, & 1\leqslant i\leqslant 6,\ j=2i,\\ \frac{1}{18}, & 2\leqslant i\leqslant 6,\ i+1\leqslant j<2i,\\ 0, & \text{其它};\end{cases}$

(2) $P(\xi=i,\eta=j)=\begin{cases}\frac{i}{36}, & 1\leqslant i\leqslant 6,\ j=i,\\ \frac{1}{36}, & 1\leqslant i<j\leqslant 6,\\ 0, & \text{其它};\end{cases}$

(3) $P(\xi=i,\eta=j)=\begin{cases}\frac{1}{36}, & 1\leqslant i=j\leqslant 6.\\ \frac{1}{18}, & 1\leqslant i<j\leqslant 6.\\ 0, & \text{其它}.\end{cases}$

5.

ξ \ η	0	1	2	$P(\xi=i)$
0	0.006	0.019	0.014	0.04
1	0.051	0.154	0.115	0.32
2	0.102	0.307	0.230	0.64
$P(\eta=j)$	0.16	0.48	0.36	

$P(\xi\leqslant 1,\eta=1)=0.173$

6.

ζ \ ξ	0	1	2	3	4
0	0.0064	0.0192	0.0144	0	0
1	0	0.0512	0.1536	0.1152	0
2	0	0	0.1024	0.3072	0.2304
$P(\zeta=j)$	0.0064	0.0704	0.2704	0.4224	0.2304

ξ 与 ζ 不独立.

习题4-3 1.(1) $1/(2\sqrt{\pi})$; (2) $\frac{1}{\sqrt{\pi}}e^{-(x+1)^2}$; $\frac{1}{2}e^{-|y|}$;

(3) $F(x,y)=\begin{cases}\frac{1}{2\sqrt{\pi}}e^{y}\int_{-\infty}^{x}e^{-(u+1)^2}du, & y\leqslant 0,\\ \frac{1}{2\sqrt{\pi}}(2-e^{-y})\int_{-\infty}^{x}e^{-(u+1)^2}du, & y>0;\end{cases}$

(4) ξ,η 独立.

2.(1) $F(x,y)=\begin{cases}0, & x<0 \text{ 或 } y<0,\\ \sin x\sin y, & 0\leqslant x,y<\frac{\pi}{2},\\ \sin x, & 0\leqslant x<\frac{\pi}{2}, y\geqslant\frac{\pi}{2},\\ \sin y, & x\geqslant\frac{\pi}{2}, 0\leqslant y<\frac{\pi}{2},\\ 1, & x,y\geqslant\frac{\pi}{2};\end{cases}$

(2) $f_\xi(x)=f_\eta(x)=\begin{cases}\cos x, & 0<x<\frac{\pi}{2},\\ 0, & \text{其它;}\end{cases}$ (3)独立； (4) $\frac{5-2\sqrt{2}}{3}$.

3.(1) $f(x,y)=\begin{cases}\frac{1}{ab}, & 0\leqslant x\leqslant a, 0\leqslant y\leqslant a.\\ 0, & \text{其它;}\end{cases}$

(2) $F(x,y)=\begin{cases}0, & x<0 \text{ 或 } y<0,\\ \frac{xy}{ab}, & 0\leqslant x\leqslant a, 0\leqslant y\leqslant b,\\ \frac{y}{b}, & x\geqslant a, 0\leqslant y\leqslant b,\\ \frac{x}{a}, & 0\leqslant x<a, y\geqslant b,\\ 1, & x>a \quad y>b.\end{cases}$

(3) $\frac{1}{6}$;

4.(1) $f(x,y)=\begin{cases}(\pi^2(1+y^2)\sqrt{1-x^2})^{-1}, & |x|<1,\\ 0, & \text{其它;}\end{cases}$

(2) $F(x,y)=\begin{cases}0, & x\leqslant -1,\\ \frac{1}{\pi^2}\left(\frac{\pi}{2}+\arcsin x\right)\left(\frac{\pi}{2}+\operatorname{arctg} y\right), & |x|<1,\\ \frac{1}{\pi}\left(\frac{\pi}{2}+\operatorname{arctg} y\right), & x\geqslant 1.\end{cases}$

5. $\frac{1}{12}$. **6.** (1) $f(x,y)=\begin{cases}\frac{1}{2}, & 0\leqslant y\leqslant x\leqslant 2,\\ 0, & \text{其它};\end{cases}$

$$F(x,y)=\begin{cases}0 & x<0 \text{ 或 } y<0,\\ \frac{xy}{2}-\frac{y^2}{4} & 0\leqslant y\leqslant x\leqslant 2,\\ \frac{x^2}{4}, & 0\leqslant x\leqslant 2,\ y>x,\\ y-\frac{y^2}{4}, & x\geqslant 2,\ 0\leqslant y\leqslant 2,\\ 1, & x\geqslant 2,\ y\geqslant 2;\end{cases}$$

(2) $f_\xi(x)=f_\eta(x)=\begin{cases}\frac{x}{2}, & 0\leqslant x\leqslant 2,\\ 0, & \text{其它};\end{cases}$

$$F_\xi(x)=F_\eta(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{4}, & 0\leqslant x\leqslant 2,\\ 1, & x>2;\end{cases}$$

(3) $(3-2\sqrt{2})\pi$; (4) 不独立.

习题 4-4 **1.** (1)

$\zeta=i$	0	1	2	3
$P(\zeta=i)$	$\frac{1}{6}$	$\frac{11}{24}$	$\frac{7}{24}$	$\frac{1}{12}$

(2)

$\zeta=i$	0	1	2
$P(\zeta=i)$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{1}{12}$

(3)

$\zeta=i$	0	1	2	3	4	5
$P(\zeta=i)$	$\frac{1}{6}$	$\frac{1}{24}$	$\frac{1}{4}$	0	$\frac{1}{24}$	$\frac{1}{12}$

2. (1) $P(\zeta=3)=1$;

(2)

$\zeta=i$	-3	-1	1	3
$P(\zeta=i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$\zeta=i$	0	2
$P(\zeta=i)$	$\frac{1}{4}$	$\frac{3}{4}$

3. $\binom{n_1+n_2}{k}p^k(1-p)^{n_1+n_2-k}$; $k=0,1,\cdots,(n_1+n_2)$.

4. $f_{\xi+\eta}(x)=\begin{cases}0, & x<0,\\ 1-e^{-x}, & 0\leqslant x\leqslant 1,\\ e^{1-x}-e^{-x}, & x>1;\end{cases}$

$$f_{\xi/\eta}(x)=\begin{cases}0, & x\leqslant 0,\\ 1-\frac{1}{x}e^{-\frac{1}{x}}-e^{-\frac{1}{x}}, & x>0.\end{cases}$$

5. $f(x)=\begin{cases} 10x-\frac{96}{7}x^2+8x^7-\frac{30}{7}x^9, & 0\leqslant x\leqslant 1, \\ 0, & \text{其它}. \end{cases}$

6. (1) $1-e^{-5\lambda a}$; (2) $(1-e^{-\lambda a})^5$.

习题 4-5 1. 1; $\frac{5}{3}$. 2. 1; $4\frac{1}{3}$; $4\frac{1}{3}$. 3. -22. 4. $\frac{\pi}{2}$; $\frac{\pi}{2}$; $\frac{\pi^2}{4}-2$; $\frac{\pi^2}{4}-2$; 0. 5. 0. 6. $5/\sqrt{52}$. 7. 0.

复习题 4-A

1. $F(x,y)=\begin{cases} 0, & x<0 \text{ 或 } y<0, \\ \frac{x(2l-x\operatorname{ctg}y)}{kl}, & 0\leqslant x<k,\ 0\leqslant y<\operatorname{arctg}\frac{k}{l}, \\ \frac{x}{k}\left(2-\frac{x}{k}\right), & 0\leqslant x<k,\ y\geqslant\operatorname{arctg}\frac{k}{l}, \\ l\operatorname{tg}y, & x\geqslant k,\ 0\leqslant y<\operatorname{arctg}\frac{k}{l}, \\ 1, & x\geqslant k,\ y\geqslant\operatorname{arctg}\frac{k}{l}. \end{cases}$

2. $P(\xi=i,\eta=j)=\frac{1}{10}$, $1\leqslant i\leqslant 4$, $1\leqslant j\leqslant 5-i$;

$P(\xi=i\mid\eta=1)=\frac{1}{4}$, $i=1,2,3,4$.

3. $P(\xi=i,\eta=j,\zeta=k)=\frac{n!}{i!k!j!}0.5^i\times 0.3^j\times 0.2^k$, $0\leqslant i,j,k\leqslant n$, $i+j+k=n$.

$P(\zeta=i)=\binom{n}{i}0.5^n$, $i=0,1,\cdots,n$; $P(\eta=j)=\binom{n}{j}0.3^j\times 0.7^{n-j}$, $j=0,1,\cdots,n$; $P(\zeta=k)=\binom{n}{k}0.2^k\times 0.8^{n-k}$, $k=0,1,\cdots n$. 4. $P(\xi=i,\eta=j)=(1-p)^{i+j}\cdot p^2$, $i,j=0,1,2,\cdots$ 5. $\frac{1}{2}$. 6. $P(\xi=i\mid\eta=k)=\frac{1}{9}$, 当 $i\neq k$; 0, 当 $i=k$.

7. (1) $P(|\xi-\eta|\leqslant t)=\begin{cases} 0, & t<0, \\ \frac{t}{30}, & 0\leqslant t<15, \\ \frac{90t-t^2-225}{1800}, & 15\leqslant t<45, \\ 1, & t\geqslant 45. \end{cases}$ (2) $\frac{1}{2}$.

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

10. 当 $|x|<1$ 时 $f(y|x)=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & |y|<\sqrt{1-x^2},\\ 0, & \text{其它};\end{cases}$

当 $|y|<1$ 时 $f(x|y)=\begin{cases}\dfrac{1}{2\sqrt{1-y^2}}, & |x|<\sqrt{1-y^2},\\ 0, & \text{其它}.\end{cases}$

11. $c=\dfrac{1}{8}$；$f_\xi(x)=\begin{cases}\dfrac{1}{6}x^3e^{-x}, & 0\leqslant x<\infty,\\ 0, & x<0;\end{cases}$ $f_\eta(y)=\dfrac{1}{4}(|y|+1)e^{-|y|}$.

12 $f_{\xi-\eta}(x)=\int_{-\infty}^{\infty}f(y+x,y)dy=\int_{-\infty}^{\infty}f(y,y-x)dy$.

13. $\dfrac{\lambda\mu}{\lambda+\mu}e^{-\lambda|x|}$. **14** $f_{\xi\eta}(x)=\begin{cases}e^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$

16 $f_{\xi+\eta,\frac{\xi}{\xi+\eta}}(x,y)=\begin{cases}\lambda^2xe^{-\lambda x}, & x\geqslant 0,0\leqslant y\leqslant 1,\\ 0, & \text{其它};\end{cases}$

$f_{\xi+\eta}(x)=\begin{cases}\lambda^2xe^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0;\end{cases}$ $f_{\frac{\xi}{\xi+\eta}}(y)=\begin{cases}1, & 0\leqslant y\leqslant 1,\\ 0, & \text{其它}.\end{cases}$

$\xi+\eta$ 与 $\dfrac{\xi}{\xi+\eta}$ 独立. **19**. 75.75；28.45.

复习题 4－B **11**. (1) $\dfrac{n^2-1}{n^2+1}$；(2) $\dfrac{n}{2\sqrt{\pi}}e^{-\frac{n^2x^2}{4}}$.

第 五 章

习题 5－1 **4**. 4.5，10.25. **5**. $(n+1)/2$，$(n^2-1)/12$. 提示：利用公式 $\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$.

习题 5－2 1．$n\Big/\sum_{i=1}^{n}x_i$．　2．$\hat{\theta}_1=\overline{X}-\sqrt{3S_n^2}$，　$\hat{\theta}_2=\overline{X}+\sqrt{3S_n^2}$

3．$\hat{p}=\overline{x^2}/s_n^2$，$\hat{b}=\overline{x}/s_n^2$．　4．$\hat{P}=1/\overline{x}$．　5．$\hat{\sigma}=\sum_{i=1}^{n}|x_i|/n$．　6．$\hat{\theta}=n\Big/\Big|\sum_{i=1}^{n}\ln x_i\Big|$．

习题 5－3 3．$c=(n+1)/n$．　4．$\hat{\theta}_1(x_1,\cdots,x_n)$．

习题 5－4 3．提示：证明 $E\xi_n$ 存在，然后利用辛钦定理证明．

习题 5－5 1．接受 H．2．能说明新安眠剂已达到新的疗效．3．接受 H．

习题 5－6 1．估计 μ 在 (95，97.8)中，v 在 (58.4，61.2)中．　2．拒绝 H，估计 $\mu-v$ 在 (0.6，4.3)中．　3．估计 σ^2 在 (4.99，38.78)之中．

习题 5－7 1．(1) 0．(2)0.995，0.5，0.005．2．估计在 (165.5，167.2) 之中．

第 六 章

习题 6－1 1．接受 H．2．拒绝 H_1，接受 H_2．

习题 6－2 1．$\hat{\beta}_0=0.16$，$\hat{\beta}_1=0.48$．　2．(1)$\hat{\beta}_0=1049.7$ (秒)，$\hat{\beta}_1=-04154$．(2)3：38.9．

习题 6－3 $\hat{\sigma}^2=0.0011$．

习题 6－4 拒绝 H．

习题 6－5 1．$\hat{y}=-0.24+0.48x_1+0.067x_2$，$\hat{\sigma}^2=0.00092$．

2．$\hat{\beta}_1=\frac{1}{4}(y_1+y_2+y_3+y_4)$，$\hat{\beta}_2=\frac{1}{4}(y_1+y_2-y_3-y_4)$，

$\hat{\beta}_3=\frac{1}{4}(y_1-y_2+y_3-y_4)$，　$\hat{\beta}_4=\frac{1}{4}(y_1-y_2-y_3+y_4)$．

参 考 书 目

[1] 钟开莱 (魏宗舒等译),初等概率附随机过程,人民教育出版社 (1980)

[2] W.费勒 (上册胡迪鹤、林向清译,下册刘文译),概率论及其应用,科学出版社 (1964,1979).

[3] J.L.福尔克斯 (魏宗舒、吕乃刚译),统计思想,上海翻译出版公司 (1987).

[4] 复旦大学编,概率论,人民教育出版社 (1979).

[5] Б.В.格涅坚科 (丁寿田译),概率论教程,高等教育出版社(1956).

[6] 魏宗舒等编,概率论与数理统计教程,高等教育出版社(1983).

[7] 华东师范大学数学系编,概率论与数理统计习题集,高等教育出版社(1982).

[8] C.R.劳 (张燮等译),线性统计及其应用,科学出版社 (1987).

[9] S.罗斯 (李漳南、杨振明译),概率论初级教程,人民教育出版社 (1981).

[10] 上海市科学技术交流站组编,正交试验设计法,上海人民出版社 (1975).

[11] F.Spitzer,Principles of Random Walk,Springer－Verlag,New York,(1976).

[12] 夏道行、严绍宗、吴卓人、舒五昌编著,实变函数论与泛函分析,人民教育出版社(1979).

[13] A.Я.辛钦 (张里千、殷涌泉译),公用事业理论的数学方法,科学出版社 (1958).

[14] 徐承彝,概率论与数理统计,北京师范学院出版社 (1987).

[15] 严士健、王隽骧、刘秀芳,概率论基础,科学出版社 (1982).

[16] 严士健、王隽骧、徐承彝,概率论与数理统计基础,上海科学技术出版社 (1982).

[17] 郑维行、王声望,实变函数论与泛函分析概要,人民教育出版社 (1980).

[18] 中国科学院数学研究所统计组，常用数理统计方法，科学出版社(1979)。

[19] 中国大百科全书，数学，中国大百科全书出版社(1988)。

[20] 体育统计编写组编，体育统计，高等教育出版社(1987)。

附　　表

附表 1　泊松分布

本表给出了

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

的数值.

k \ λ	0.1	0.2	0.3	0.4	0.5	0.6
0	0.904837	0.818731	0.740818	0.670320	0.606531	0.548812
1	0.090484	0.163746	0.222245	0.268128	0.303265	0.329287
2	0.004524	0.016375	0.033337	0.053626	0.075816	0.098786
3	0.000151	0.001091	0.003334	0.007150	0.012636	0.019757
4	0.000004	0.000055	0.000250	0.000715	0.001580	0.002964
5		0.000002	0.000015	0.000057	0.000158	0.000356
6			0.000001	0.000004	0.000013	0.000035
7					0.000001	0.000003

k \ λ	0.7	0.8	0.9	1.0	2.0	3.0
0	0.496585	0.449329	0.406570	0.367879	0.135335	0.049787
1	0.347610	0.359463	0.365913	0.367879	0.270671	0.149361
2	0.121663	0.143785	0.164661	0.183940	0.270671	0.224042
3	0.028388	0.038343	0.049398	0.061313	0.180447	0.224042
4	0.004968	0.007669	0.011115	0.015328	0.090224	0.168031
5	0.000695	0.001227	0.002001	0.003066	0.036089	0.100819
6	0.000081	0.000164	0.000300	0.000511	0.012030	0.050409
7	0.000008	0.000019	0.000039	0.000073	0.003437	0.021604
8		0.000002	0.000004	0.000009	0.000859	0.008101
9				0.000001	0.000191	0.002701
10					0.000038	0.000810
11					0.000007	0.000221
12					0.000001	0.000055
13						0.000013
14						0.000003
15						0.000001

续表

k \ λ	4.0	5.0	6.0	7.0	8.0	9.0
0	0.018316	0.006738	0.002479	0.000912	0.000335	0.000123
1	0.073263	0.033690	0.014873	0.006383	0.002684	0.001111
2	0.146525	0.084224	0.044618	0.022341	0.010735	0.014998
3	0.195367	0.140374	0.089235	0.052129	0.028626	0.014994
4	0.195367	0.175467	0.133853	0.091226	0.057252	0.033737
5	0.156293	0.175467	0.160623	0.127717	0.091604	0.060727
6	0.104194	0.146223	0.160623	0.149003	0.122138	0.091090
7	0.059540	0.104445	0.137677	0.149003	0.139587	0.117116
8	0.029770	0.065278	0.103258	0.130377	0.139587	0.131756
9	0.013231	0.036266	0.068838	0.101405	0.124077	0.131756
10	0.005292	0.018133	0.041303	0.070983	0.099262	0.118580
11	0.001925	0.008242	0.022529	0.045171	0.072190	0.097020
12	0.000642	0.003434	0.011262	0.026350	0.048127	0.072765
13	0.000197	0.001321	0.005199	0.014188	0.029616	0.050376
14	0.000056	0.000472	0.002228	0.007094	0.016924	0.032384
15	0.000015	0.000157	0.000891	0.003311	0.009026	0.019431
16	0.000004	0.000049	0.000334	0.001448	0.004513	0.010930
17	0.000001	0.000014	0.000118	0.000596	0.002124	0.005786
18		0.000004	0.000039	0.000232	0.000944	0.002893
19		0.000001	0.000012	0.000085	0.000397	0.001370
20			0.000004	0.000030	0.000159	0.000617
21			0.000001	0.000010	0.000061	0.000264
22				0.000003	0.000022	0.000108
23				0.000001	0.000008	0.000042
24					0.000003	0.000016
25					0.000001	0.000006
26						0.000002
27						0.000001

附表 2　标准正态分布

本表给出了

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt\ (x\geqslant 0)$$



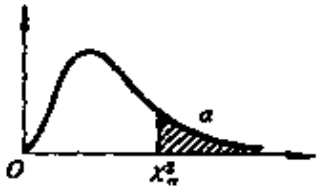
的数值.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0.1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0.2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0.3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0.4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0.5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0.6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0.7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0.8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0.9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1.0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1.1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1.2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1.3	90320	90490	90653	90824	90988	91149	91309	91466	91621	91774
1.4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1.5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1.6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1.7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1.8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1.9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2.0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2.1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2.2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2.3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2.4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2.5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2.6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2.7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2.8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2.9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	0.99865	99903	99931	99952	99966	99977	99984	99989	99993	99995

附表3 χ^2 分布

本表给出了满足

$$P(\chi^2>\chi_\alpha^2)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_{\chi_\alpha^2}^{+\infty}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx$$



的 χ_α^2 值 (n 是自由度).

n	α									
	0.99	0.98	0.95	0.90	0.50	0.10	0.05	0.02	0.01	0.001
	χ_α^2									
1	0.000	0.001	0.004	0.016	0.455	2.71	3.84	5.41	6.64	10.83
2	0.020	0.040	0.103	0.211	1.386	4.61	5.99	7.82	9.21	13.82
3	0.115	0.185	0.352	0.584	2.366	6.25	7.82	9.84	11.34	16.27
4	0.297	0.429	0.771	1.064	3.357	7.78	9.49	11.67	13.28	18.47
5	0.554	0.752	1.145	1.610	4.351	9.24	11.07	13.39	15.09	20.52
6	0.872	1.134	1.635	2.204	5.35	10.65	12.59	15.03	16.81	22.46
7	1.239	1.564	2.167	2.833	6.35	12.02	14.07	16.62	18.48	24.32
8	1.646	2.032	2.733	3.490	7.34	13.36	15.51	18.17	20.09	26.13
9	2.088	2.532	3.325	4.168	8.34	14.68	16.92	19.68	21.67	27.88
10	2.558	3.059	3.940	4.865	9.34	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.57	5.58	10.34	17.28	19.63	22.62	24.73	31.26
12	3.57	4.18	5.23	6.30	11.34	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	12.34	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	13.34	21.06	23.69	26.87	29.14	36.12
15	5.23	5.99	7.26	8.55	14.34	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	15.34	23.54	26.30	29.63	32.00	39.25
17	6.41	7.26	8.67	10.09	16.34	24.77	27.59	31.00	33.41	40.79
18	7.02	7.91	9.39	10.87	17.34	25.99	28.87	32.35	34.81	42.31
19	7.63	8.57	10.12	11.65	18.34	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	19.34	28.41	31.41	35.02	37.57	45.32
21	8.90	9.91	11.59	13.24	20.34	29.61	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	21.34	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	22.34	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	23.34	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	24.34	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	25.34	35.56	38.89	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	26.34	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	27.34	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	28.34	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	29.34	40.26	43.77	47.96	50.89	59.70

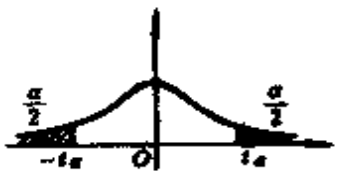
附表4 t 分 布

本表给出了满足

$$P(|t|>t_\alpha)=1-\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}$$

$$\times\int_{-t_\alpha}^{t_\alpha}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx=\alpha$$

的t_α值（n是自由度）.



n	α				
	0.10	0.05	0.02	0.01	0.001
	t_α				
1	6.31	12.71	31.82	63.66	636.62
2	2.92	4.30	6.97	9.93	31.60
3	2.35	3.18	4.54	5.84	12.94
4	2.13	2.78	3.75	4.60	8.61
5	2.02	2.57	3.37	4.03	6.86
6	1.94	2.45	3.14	3.71	5.96
7	1.90	2.37	3.00	3.50	5.41
8	1.86	2.31	2.90	3.36	5.04
9	1.83	2.26	2.82	3.25	4.78
10	1.81	2.23	2.76	3.17	4.59

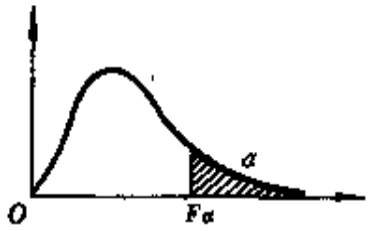
续表

n	α				
	0.10	0.05	0.02	0.01	0.001
	t_α				
11	1.80	2.20	2.72	3.11	4.44
12	1.78	2.18	2.68	3.06	4.32
13	1.77	2.16	2.65	3.01	4.22
14	1.76	2.15	2.62	2.98	4.14
15	1.75	2.13	2.60	2.95	4.07
16	1.75	2.12	2.58	2.92	4.02
17	1.74	2.11	2.57	2.90	3.97
18	1.73	2.10	2.55	2.88	3.92
19	1.73	2.09	2.54	2.86	3.88
20	1.73	2.09	2.53	2.85	3.85
21	1.72	2.08	2.52	2.83	3.82
22	1.72	2.07	2.51	2.82	3.79
23	1.71	2.07	2.50	2.81	3.77
24	1.71	2.06	2.49	2.80	3.75
25	1.71	2.06	2.48	2.79	3.73
26	1.71	2.06	2.48	2.78	3.71
27	1.70	2.05	2.47	2.77	3.69
28	1.70	2.05	2.47	2.76	3.67
29	1.70	2.04	2.46	2.76	3.66
30	1.70	2.04	2.46	2.75	3.65
40	1.68	2.02	2.42	2.70	3.55
60	1.67	2.00	2.39	2.66	3.46
120	1.66	1.98	2.36	2.62	3.37
∞	1.65	1.96	2.33	2.58	3.29

附表5　F分　布

本表给出了满足

$$P(F>F_\alpha)=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\int_{F_\alpha}^{+\infty}x^{\frac{m-2}{2}}(n+mx)^{-\frac{m+n}{2}}dx=\alpha$$



的 F_α 值 (m 是第一自由度，n 是第二自由度)，

$\alpha=0.25$

n	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	5.83	7.50	8.20	8.58	8.82	8.93	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.76	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.46	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.64	1.63	1.62	1.62	1.61	1.59	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.54	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.50	1.48

11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.47	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.44	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.42	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.40	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.38	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.36	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.35	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.34	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.33	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.32	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.31	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.30	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.30	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.29	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.28	1.25
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.26	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.22	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.19	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.16	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.12	1.00

(续表)

$\alpha=0.05$

n	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	252.2	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.63	4.62	4.56	4.43	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.63	3.64	3.57	3.51	3.44	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.62	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.49	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.38	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.30	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.22	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.16	2.07

16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.11	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.06	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.02	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	1.98	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	1.95	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	1.92	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	1.89	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	1.86	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.84	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.82	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.74	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.64	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.53	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.32	1.00

续表

n	$\alpha=0.01$														
	m														
	1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
	F_α														
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6313	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.32	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.65	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.20	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.06	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.82	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.03	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.48	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.08	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	3.78	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.10	3.86	3.54	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.34	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.18	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.05	2.87

16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	2.93	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	2.83	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.03	2.75	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.67	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.61	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.55	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.50	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.45	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.40	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.36	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.21	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.02	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	1.84	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.66	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.47	1.00

附表6　符号检验表

N	(%) 10	25	N	(%) 1	5	10	25	N	(%) 1	5
1			31	7	9	10	11	61	20	
2			32	8	9	10	12	62	20	
3		0	33	8	10	11	12	63	20	23
4		0	34	9	10	11	13	64	21	23
5	0	0	35	9	11	12	13	65	21	24
6	0	1	36	9	11	12	14	66	22	24
7	0	1	37	10	12	13	14	67	22	25
8	0	1	38	10	12	13	14	68	22	25
9	0	2	39	11	12	13	15	69	23	25
10	0	2	40	11	13	14	15	70	23	26
11	0	3	41	11	13	14	16	71	24	26
12	1	3	42	12	14	15	16	72	24	27
13	1	3	43	12	14	15	17	73	25	27
14	1	4	44	13	15	16	17	74	25	28
15	2	4	45	13	15	16	18	75	25	28
16	2	5	46	13	15	16	18	76	26	28
17	2	5	47	14	16	17	19	77	26	29
18	3	6	48	14	16	17	19	78	27	29
19	3	6	49	15	17	18	19	79	27	30
20	3	6	50	15	17	18	20	80	28	30
21	4	7	51	15	18	19	20	81	28	31
22	4	7	52	16	18	19	21	82	28	31
23	4	8	53	16	18	20	21	83	29	32
24	5	8	54	17	19	20	22	84	29	32
25	5	9	55	17	19	20	22	85	30	32
26	6	9	56	17	20	21	23	86	30	33
27	6	10	57	18	20	21	23	87	31	33
28	6	10	58	18	21	22	24	88	31	34
29	7	10	59	19	21	22	24	89	31	34
30	7	11	60	19	21	23	25	90	32	35

附表6　符号检验表

N	(%) 10	25	N	(%) 1	5	10	25	N	(%) 1	5
1			31	7	9	10	11	61	20	
2			32	8	9	10	12	62	20	
3		0	33	8	10	11	12	63	20	23
4		0	34	9	10	11	13	64	21	23
5	0	0	35	9	11	12	13	65	21	24
6	0	1	36	9	11	12	14	66	22	24
7	0	1	37	10	12	13	14	67	22	25
8	0	1	38	10	12	13	14	68	22	25
9	0	2	39	11	12	13	15	69	23	25
10	0	2	40	11	13	14	15	70	23	26
11	0	3	41	11	13	14	16	71	24	26
12	1	3	42	12	14	15	16	72	24	27
13	1	3	43	12	14	15	17	73	25	27
14	1	4	44	13	15	16	17	74	25	28
15	2	4	45	13	15	16	18	75	25	28
16	2	5	46	13	15	16	18	76	26	28
17	2	5	47	14	16	17	19	77	26	29
18	3	6	48	14	16	17	19	78	27	29
19	3	6	49	15	17	18	19	79	27	30
20	3	6	50	15	17	18	20	80	28	30
21	4	7	51	15	18	19	20	81	28	31
22	4	7	52	16	18	19	21	82	28	31
23	4	8	53	16	18	20	21	83	29	32
24	5	8	54	17	19	20	22	84	29	32
25	5	9	55	17	19	20	22	85	30	32
26	6	9	56	17	20	21	23	86	30	33
27	6	10	57	18	20	21	23	87	31	33
28	6	10	58	18	21	22	24	88	31	34
29	7	10	59	19	21	22	24	89	31	34
30	7	11	60	19	21	23	25	90	32	35